

M 现代计算机

XIANDAI JISUANJI

第30卷第4期（总第796期）

半月刊（1984年创刊）

2024年2月25日出版

主管单位 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 梁嘉璐
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
电子邮箱 tougao@moderncomputer.cn

ISSN 1007-1423
CN 44-1415/TP

邮发代码：46-121
定价：30.00元



邮局订刊二维码



现代计算机
官方网站二维码



M

现代计算机

第30卷 第4期（总第796期）

2024年2月

2024年2月 第30卷

第4期（总第796期）



ISSN 1007-1423
CN 44-1415/TP

MODERN COMPUTER

现代计算机



中山大学出版社 主办

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。

目次

研究与开发

面向可迁移跨域自适应学习的航空发动机剩余寿命预测方法

..... 李文骁, 李勇成, 李 鹏, 马浩统, 雷印杰 (1)

基于FaceNet的犬只鼻纹识别..... 李梦晗, 张晨蕾, 田存伟 (9)

基于距离视图表示与逐点细化结合的点云语义分割方法..... 陈 晗, 何鸿添, 雷印杰 (16)

基于改进点线特征的多源融合SLAM算法

..... 林嘉洁, 徐 胜, 苏成悦, 陈元电, 刘力斌, 施振华 (23)

多尺度注意力机制下的人脸表情识别算法设计..... 蒋文豪 (29)

基于LightGBM和蚁群算法优化的特征选择方法..... 别春洋, 陶贻勇 (34)

混叠场景下的轻量级口罩佩戴检测算法..... 安祯阳 (39)

基于自适应阈值的ORB特征点提取算法研究..... 谭帅奇, 帅鹏举 (43)

实践与经验

基于注意力机制的三维模型特征提取..... 王欢欢, 李舒晴 (48)

LSTM-Prophet混合模型在物料储备需求预测中的应用..... 辛 唯 (53)

机械完工尾项管理系统在FPSO建造项目的应用

..... 邢同超, 石振华, 赵西磊, 王 亮, 伍江勇 (58)

基于数据交换格式标准的通用一致性测试框架..... 雷海申, 徐 巍, 裴 安, 董 楠 (62)

开发案例

基于学习体验的中医药文化数字展示系统开发研究..... 祝敏娇 (66)

基于微信小程序的鸟鸣声识别系统的研究与实现

..... 陈凌芳, 周 雁, 王庆娟, 林佳皓, 湛业恒 (75)

基于Spring Boot的蒿草智能识别小程序的研究与实现..... 赵文欣, 苏 峰 (83)

基于Web的异常数据检测系统设计与实现..... 常 鑫, 田立勤 (88)

基于工业互联网的能碳数字化和智能化建设..... 张 伟, 安振文, 刘子强 (94)

嵌入式实验室门禁人脸识别系统设计..... 翟 丽, 邹琦璋 (99)

基于三维点云建模的操作驾驶舱设计..... 宋亚楠, 姜 吟, 吕厚稳, 葛楚琳, 曹纪慧 (104)

基于CNN的人脸年龄与性别检测系统设计..... 孙小广, 万若楠, 余光正 (109)

三维虚拟直播间的设计与制作..... 陈 军, 陈俊汶 (113)

基于OpenStack云架构下Hadoop分布式集群部署实训教学过程研究

..... 周少珂, 王 雷, 石磊磊, 李 凯 (117)

Modern Computer

(Vol. 30, No. 4; Feb. 25, 2024)

CONTENTS

Research and Development

Transferable cross-domain adaptive learning for aircraft engine remaining useful life prediction	(1)
Canine nose print recognition based on FaceNet	(9)
Point cloud semantic segmentation based on range view representation and point-wise refinement	(16)
Multi-source fusion SLAM algorithm based on improved point-Line features	(23)
Design of face expression recognition algorithm under the multi-scale attention mechanism.....	(29)
Feature selection method based on LightGBM and ant colony optimization	(34)
Lightweight mask wearing detection algorithm in aliasing scenario	(39)
Research on ORB feature point extraction algorithm based on adaptive threshold	(43)

Practice and Experience

Three-dimensional model feature extraction based on attention mechanism	(48)
Application of LSTM-PROPHET mixed model in material reserve demand forecasting	(53)
Application of mechanical completion punch management system in FPSO construction project	(58)
Universal conformance testing framework based on data exchange format standards.....	(62)

Development Solution

Research on the development of a digital display system for traditional Chinese medicine culture based on learning experience	(66)
Research and implementation of birdsong recognition system based on WeChat mini program	(75)
Research and implementation of wormwood intelligent recognition applet based on Spring Boot	(83)
Design and implementation of abnormal data detection system based on Web	(88)
Digital and intelligent construction of energy and carbon based on industrial internet	(94)
Embedded laboratory access control face recognition system design	(99)
Design of operating cockpit based on 3D point cloud modeling	(104)
Design of facial age and gender detection system based on CNN	(109)
Design and production of 3D virtual broadcasting room	(113)
Research on the training and teaching process of Hadoop distributed cluster deployment based on OpenStack cloud architecture	(117)

研究与开发

文章编号: 1007-1423(2024)04-0001-08

DOI: 10.3969/j.issn.1007-1423.2024.04.001

面向可迁移跨域自适应学习的航空发动机剩余寿命预测方法

李文骁¹, 李勇成¹, 李 鹏², 马浩统³, 雷印杰^{1*}

(1. 四川大学电子信息学院, 成都 610065; 2. 中国电子科技集团公司第十研究所航空电子信息系统技术重点实验室, 成都 610036; 3. 中国科学院光电技术研究所中国科学院光束控制重点实验室, 成都 610209)

摘要: 航空发动机剩余寿命(RUL)预测任务中数据集标签较少且工况多变, 导致传感器时间序列之间存在明显分布差异, 限制了RUL预测方法的泛化能力。跨域学习的提出为该任务提供了一种可行的解决方案。传统跨域学习通过域自适应方法最小化源域和目标域特征之间的分布差异, 得到跨域对齐特征, 实现跨域知识迁移。但随着航空发动机的退化, 前后时间步的语义信息也发生变化, 导致原先对齐特征的局部语义不匹配, 影响模型性能。针对该问题, 提出方法基于可迁移对抗方法对跨域RUL预测方法展开研究, 通过优化局部域鉴别器输出的概率熵, 使得对齐特征在局部上难以区分。利用模型在RUL预测过程中的目标互信息进行语义约束, 得到同时具有局部可迁移性和目标语义重要性的域不变特征, 提升模型的泛化能力。在CMAPSS航空发动机数据集上进行的实验表明, 该方法在RMSE和SCORE两个指标上均超过现有的其他跨域自适应方法, 证实了其有效性。

关键词: 剩余寿命预测; 跨域学习; 域自适应; 可迁移对抗

0 引言

作为预测与健康管理(PHM)中的关键技术之一, 剩余寿命(remaining useful life, RUL)预测技术可以预测航空发动机故障前的剩余时间^[1]。目前, 基于深度学习的方法, 以端到端的方式进行RUL预测, 受到研究人员广泛关注。然而, 深度学习模型的良好预测性能通常是在标签充足、运行条件恒定情况下实现的。在实际项目中, 机械设备的工作环境可能发生“领域偏移”^[2], 导致跨域场景下的模型失去准确性^[3]。跨域学习通过领域自适应(DA)方法为上述问题提供了一种可行的解决方案^[4]。

现有基于域自适应的跨域RUL预测方法^[5-6], 大多直接通过最小化源域和目标域之间的分布差异实现跨域对齐。然而, 这种方法假

定了对齐特征在局部上具有同等的相似程度和退化程度, 这在实际任务中难以满足。

针对以上问题, 本文提出了基于可迁移对抗的跨域自适应寿命预测(TADA)方法, 为航空发动机样本设置了退化信息和退化趋势上的局部域鉴别器, 增强跨域迁移能力。并且, TADA中引入目标语义约束, 进一步保留目标语义信息, 使域不变特征同时具有局部可迁移性和目标语义重要性。

1 相关工作

在本节中, 将重点介绍基于数据驱动的剩余寿命预测和领域自适应方面相关工作。

1.1 基于深度学习的寿命预测方法

当前基于深度学习(DL)的剩余寿命(RUL)

收稿日期: 2023-09-25 修稿日期: 2023-10-31

基金项目: 国家自然科学基金面上项目(62276176); 装发预研项目(32102040403)

作者简介: 李文骁(1999—), 男, 四川达州人, 硕士研究生, 研究方向为剩余寿命预测; 李勇成(1999—), 男, 湖北恩施人, 硕士研究生, 研究方向为剩余寿命预测; 李鹏(1984—), 男, 河南焦作人, 高级工程师, 博士, 研究方向为航空电子系统架构和智能计算技术; 马浩统(1983—), 男, 河南襄城人, 研究员, 博士, 研究方向为光束控制技术和新型光电成像技术; *通信作者: 雷印杰(1983—), 男, 四川成都人, 教授, 博导, 博士, 研究方向为人工智能与计算机视觉、迁移学习、剩余寿命预测等, E-mail: yinjie@scu.edu.cn

预测方法,可分为深度卷积神经网络、循环神经网络和前馈神经网络^[7]。全航等^[8]提出基于卷积神经网络的寿命预测模型。Guo等^[9]基于循环神经网络(RNN)进行RUL预测。Huan等^[10]通过两个分离的双向LSTM分别处理退化信号辅助输入信息。Zhang等^[11]使用双向GRU模型,合并LSTM中的遗忘门和输入门,提高了网络性能。但上述循环神经网络及其变体,在处理长序列时依然存在梯度消失问题。Mo等^[12]基于注意力机制和Transformer架构进行寿命预测,提高了模型捕获长距离依赖关系的能力。

1.2 领域自适应方法

目前关于域自适应(DA)^[13]方法的研究主要集中在跨域图像分类问题上,针对跨域RUL预测这样的多元时间序列回归问题的研究较少。da Costa等^[14]将LSTM与对抗域自适应框架结合,得到域鉴别器难以区分的域不变特征。Ragab等^[15-16]通过双向LSTM,使用类似生成对抗网络(GAN)的损失函数进行对抗训练。Shi等^[17]提出了一种基于Wasserstein距离的多尺度对抗领域自适应方法,通过多尺度卷积进行特征提取。然而,这些方法大多仅关注源域和目标域特征的全局对齐,忽略了域不变特征在局部的可迁移性和目标语义重要性,本文的TADA正是针对此问题产生。

2 本文方法

2.1 问题定义

在剩余寿命(RUL)预测问题中,令 $X = \{x_1, x_2, \dots, x_n\}$,其中 $x_i \in \mathbf{R}^d$ 表示时间步 i 的 d 维输入特征, $Y = \{y_1, y_2, \dots, y_n\}$,其中 y_i 是表示时间步 i 的RUL标签。任务是找到一个映射 $f: f(X) = Y$ 。

在领域自适应中,定义源域 $D_s = \{X_s, Y_s\}$,其中 X_s 是属于源分布 $P(X_s)$ 的输入数据, Y_s 是相应的标签。目标域 $D_T = \{X_T\}$,其中 X_T 是属于目标分布 $P(X_T)$ 的无标签训练数据。对于域自适应RUL预测任务,通过时间序列建模,得到源域和目标域输入数据,即 $X_s \in \mathbf{R}^{q \times d}$, $X_T \in \mathbf{R}^{q \times d}$,其中 q 为训练样本长度。源域中的标签 Y_s ,是相应长度为 q 的RUL序列。假设 $P(X_s) \neq P(X_T)$,目标

是将知识从源域 D_s 迁移至目标域 D_T ,以此学习 D_T 上的预测函数 $f_T: f_T(X_T) \rightarrow Y_T$,确定标签 Y_T 。

2.2 模型结构

本文提出的基于可迁移对抗的跨域自适应(TADA)剩余寿命预测方法整体结构如图1所示,主要由特征提取器(FE)、可迁移性评估模块(TEM)、互信息模块(MI)、域鉴别器模块(D1)和剩余寿命预测器(RP)组成。其中TEM引入了时间步级别的退化信息域鉴别器(D2)和RUL预测级别的退化趋势域鉴别器(D3),D1和RP由简单的全连接层实现。

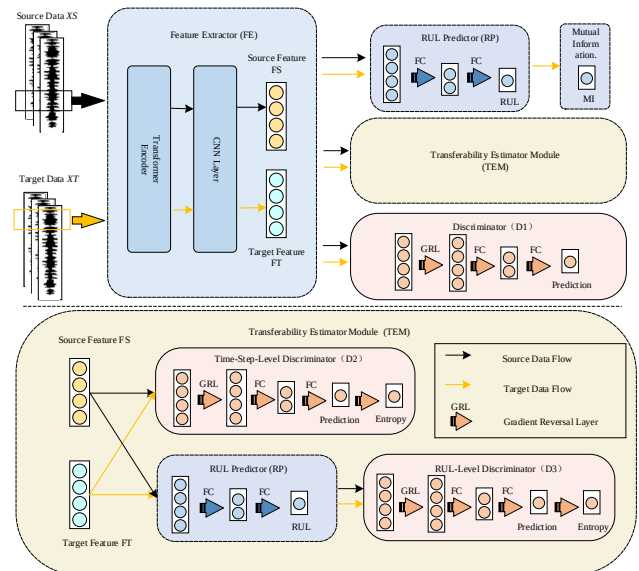


图1 TADA结构

2.3 特征提取

本文方法的FE整体结构如图2所示,由于涡轮发动机退化数据具有明显退化趋势,故选用与文献[12]相同的Transformer编码器,解决长序列预测中存在的上下文梯度消失问题,使模型聚焦全局特征中更重要的退化信息。

随着时间的变化,复杂设备退化程度不断增加,其退化特征的分布也随之变化。信号中包含的特征比早期阶段更多,这使得很难平衡全局和局部特征。针对这个问题,本文基于CNN的局部感知性^[18],引入不同卷积核大小的CNN提取不同尺度的局部特征,在保持全局上下文信息捕获能力的基础上,对局部特征敏感。

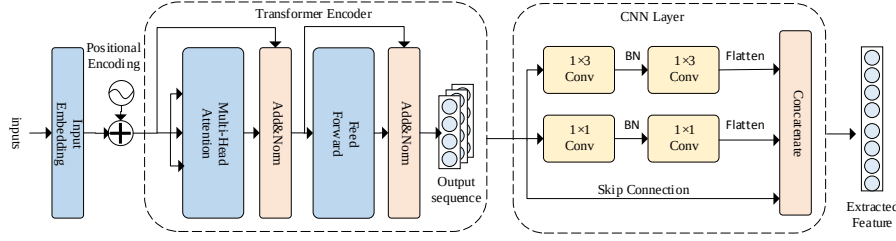


图2 特征提取器结构

特征提取器的最终输出，为后续的域鉴别器和RUL预测器提供输入。RUL回归的损失函数如下：

$$L_{\text{RUL}} = \frac{1}{N} \sum_{n=1}^N (y_n - \hat{y}_n)^2 \quad (1)$$

其中： $\hat{y}_n$ 为RUL预测值， y_n 为真实标签。

2.4 基于对抗的域自适应

在领域自适应问题中，所有的对抗性损失都使用标准分类损失 L_D 来训练对抗性鉴别器，公式如下：

$$L_D = -[y \log \hat{y} + (1 - y) \log (1 - \hat{y})] \quad (2)$$

其中： $\hat{y}$ 为鉴别器预测值， y 为真实标签。

然而，它们在用于训练生成器映射的损失函数 L_G 上有所不同。本文使用的梯度反转层(GRL)^[19]，通过梯度反转优化映射，得到生成器损失函数，公式如下：

$$L_G = -L_D \quad (3)$$

2.5 可迁移性估计

由于模型在局部的可迁移性是未知的，本文设计了一种可迁移性评估模块(TEM)，鼓励学习到的特征在局部具有同样的退化信息和退化趋势，结构如图1所示。

该模块基于文献[20]中对抗的思想，采用了一个基于时间步长级别的退化信息域鉴别器(D2)和一个基于RUL预测级别的退化趋势域鉴别器(D3)，使用交叉熵 L_{Cross} 作为损失函数，并通过梯度反转层实现对抗，公式如下：

$$L_{\text{Cross}}(y, \hat{y}) = -[y \log \hat{y} + (1 - y) \log (1 - \hat{y})] \quad (4)$$

$$L_{\text{Time}} = -\frac{1}{R} \sum_{r=1}^R L_{\text{Cross}}(y, D_{\text{Time}}(G_f(x_r^d))) \quad (5)$$

$$L_{\text{Trend}} = L_{\text{Cross}}(y, D_{\text{Trend}}(P(G_f(x^d)))) \quad (6)$$

其中： $D_{\text{Time}}(G_f(x_r^d))$ ， $d \in [s, t]$ 表示在时间步长 r 上特征的退化信息属于源域的概率； $D_{\text{Trend}}(P(G_f(x^d)))$ ， $d \in [s, t]$ 表示当前退化趋势属于源域的概率， y 为真实标签。

域鉴别器 D_{Time} 和 D_{Trend} 尝试为源域数据分配1，为目标域数据分配0，而特征提取器则通过梯度反转对抗这种情况。若某个特征能欺骗 D_{Time} 和 D_{Trend} ，使其为该特征分配0.5，则其在局部具有同等的退化信息和退化趋势。通过熵函数评价鉴别器预测值的不确定程度，代表局部的可迁移性，设计了迁移评估损失函数 L_{Trans} ，公式如下：

$$L_{\text{Trans}} = -\frac{1}{2} (H(D_{\text{Time}}(\cdot)) + H(D_{\text{Trend}}(\cdot))) \quad (7)$$

其中： $H(\cdot)$ 为标准熵函数； $D_{\text{Time}}(\cdot)$ 和 $D_{\text{Trend}}(\cdot)$ 为域鉴别器预测值。

在鉴别器和特征提取器的对抗过程中，预测值的熵值越高表示特征对齐过程中局部信息在源域和目标域上不可分。通过优化此损失函数，增强域不变特征在局部的可迁移性。

2.6 目标语义约束

强制使两个域的特征相似可能会影响RUL预测性能。因此，本文基于分类任务中通过互信息模拟有监督学习输入输出强相关性的思想^[21]，设计了互信息损失函数 L_{mi} ，公式如下：

$$L_{\text{mi}} = I(p^t; x^t) = H(p^t) - H(p^t | x^t) \quad (8)$$

$$H(p^t) = h(\bar{p}^t) = -\sum_{k=1}^K \bar{p}_k^t \log(\bar{p}_k^t) \quad (9)$$

$$H(p^t | x^t) = \frac{1}{N} \sum_{n=1}^N h(p_n^t) = \frac{1}{N} \sum_{n=1}^N \sum_{k=1}^K p_{kn}^t \log(p_{kn}^t) \quad (10)$$

其中 $p_{kn}^t = \text{softmax}(\hat{y}_{kn}^t)$ ， $\bar{p}_k^t = E_{x_n^t}[p_k^t]$ ， $\hat{y}_{kn}^t$ 为目标

域样本 n 在时间步长 k 时刻的 RUL 预测值, $H(p^i)$ 和 $H(p^i|x^i)$ 为熵和条件熵。

互信息 $I(p^i; x^i)$ 最大化的目标等价于两个部分: 最大化信息熵 $H(p^i)$ 和最小化条件熵 $H(p^i|x^i)$ 。即 $H(p^i|x^i)$ 和 $H(p^i)$ 分别评价了目标域上 RUL 预测值的聚集性^[22] 和多样性^[23]。同时最大化 $H(p^i)$ 和最小化 $H(p^i|x^i)$, 满足了聚类假设^[24], 能将决策边界远离语义空间中的相似特征, 得到更具退化趋势的 RUL 映射, 如图 3 所示。

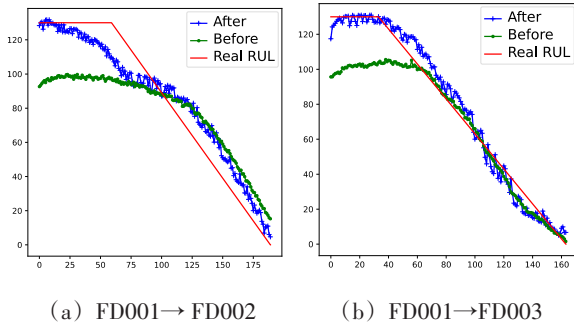


图 3 目标约束对比可视化结果

最终, 模型的损失函数如下:

$$L = \lambda_{\text{RUL}} L_{\text{RUL}} + \lambda_{\text{D}} L_{\text{D}} + \lambda_{\text{Time}} L_{\text{Time}} + \lambda_{\text{Trend}} L_{\text{Trend}} - (\lambda_{\text{Trans}} L_{\text{Trans}} + \lambda_{\text{mi}} L_{\text{mi}}) \quad (11)$$

其中: λ 为各项损失的平衡, 经过多次实验选取最佳值, 分别设置为 1、0.05、0.05、0.05、0.5 和 0.01。

3 实验相关设置

3.1 实验环境

本文实验在 Ubuntu18.04 系统下进行, 采用 PyTorch 1.6 深度学习框架, 硬件配置为 NVIDIA GeForce GTX 1080 Ti, 11 GB 显存。训练过程中参数设置如下: 采用 SGD 优化器, 初始学习率为 0.02, 权重衰减系数为 0.001, 训练 240 个 epoch, 批量大小为 128。

3.2 数据预处理

本文选择了航空发动机领域的通用数据集 C-MAPSS 来验证 TADA 方法的有效性, 其中包括 4 个子集, FD001 运行条件和故障种类较为简

单、FD002 和 FD004 运行条件更多、FD003 和 FD004 故障种类更多。

由于不同运行条件具有明显差异, 传感器读数具有不同的尺度。通过归一化将原始数据重新缩放到 $[0, 1]$ 的范围内使训练过程更平稳, 公式如下:

$$X^i = \frac{x^i - x_{\min}^i}{x_{\max}^i - x_{\min}^i} \quad (12)$$

其中, $x_{\max}^i$ 和 $x_{\min}^i$ 是第 i 个传感器数据中最大值和最小值。

但归一化处理后的原始传感器数据是一系列长度不同的时间序列数据。为了便于模型训练, 根据先前研究^[25], 使用滑动窗口算法将长序列分割成相同大小的短样本, 窗口长度设置为 70, 滑动步长设置为 1, 并且每个窗口中最后一个周期的 RUL 被用作 RUL 标签。每台发动机的生命周期可分为健康阶段和线性退化阶段^[26]。在健康阶段, 发动机处于恒定的最大寿命值, 健康阶段过后, 引擎开始线性退化至寿命结束。为与其他方法进行比较, 我们将每个引擎的最大寿命值设置为 130 个周期。

3.3 评价指标

为评估所有模型的预测性能, 本文采用了两个常用的评估指标, 分别是均方根误差 (RMSE) 和误差得分 (SCORE) 函数^[27]。RMSE 指标用以评估模型 RUL 预测的准确性, 公式如下:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{Y}_i - Y_i)^2} \quad (13)$$

在 RUL 预测应用中, RMSE 指标对早期和晚期的 RUL 预测平等对待, 然而晚期 RUL 的预测可能对系统更有害。因此, 引入了 SCORE 指标, 对晚期 RUL 的预测进行更严厉的惩罚, 公式如下:

$$SCORE = \sum_{i=1}^N \begin{cases} e^{-\frac{\hat{Y}_i - Y_i}{13}} - 1, & \hat{Y}_i < Y_i \\ e^{-\frac{\hat{Y}_i - Y_i}{10}} - 1, & \hat{Y}_i > Y_i \end{cases} \quad (14)$$

其中: N 为测试集样本的个数; Y_i 为第 i 个样本时刻涡轮发动机实际 RUL 值; $\hat{Y}_i$ 为根据模型输出计算出的 RUL 预测值。

3.4 实验结果分析

3.4.1 对比实验

为验证TADA方法的有效性,将部分实验结果可视化,如图4所示。并将其与其他域自适应方法进行对比实验,实验设置了12个复杂工况下跨域RUL预测任务,其中1个子数据集(例如FD001)被视为源域,其他子数据集(例如FD002、FD003、FD004)被视为目标域,结果见表1和表2。

表1展示了对比方法在跨域任务中的均方根误差值(RMSE)。根据数据分析,除了3个较简单的跨域任务外,TADA在9个目标域比源域复杂的

跨域任务中均取得最佳结果(最低RMSE值)。尤其当目标域包含比源域更多的操作条件时,跨域差异更大,TADA的性能明显优于其他所有方法,例如“FD001→FD004”和“FD003→FD004”。

表2展示了TADA与其他方法在RUL预测中的SCORE得分。低分表示模型有更好的早期预测能力,TADA在12个跨域场景中有8个展现出更好的性能。结合表1和表2,本文提出的TADA方法在12个跨域场景中, RMSE和SCORE的平均值这两个指标,相较对比WADA方法的18.76和1968,分别降低18.4%和4.3%,达到了15.31和1883。表明了TADA在差异显著的复杂跨域任务上的优势。

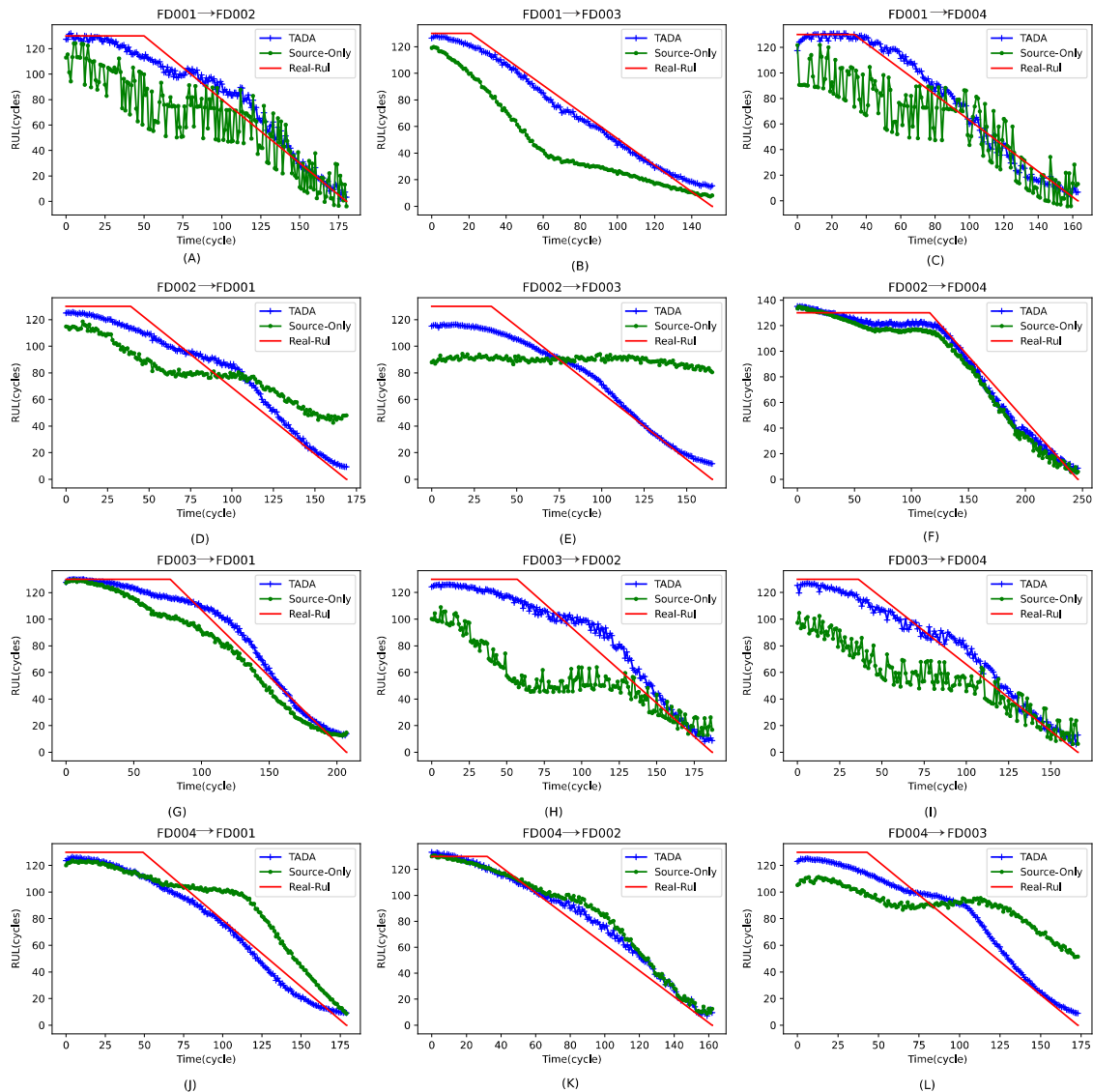


图4 跨域场景下的TADA可视化结果

表 1 对比实验结果(RMSE)

Method	Source only	DANN ^[14]	ADDA ^[15]	CADA ^[16]	WADA ^[17]	TADA
FD001→FD002	43.25	48.62	19.87	19.52	18.60	<u>15.15</u>
FD001→FD003	29.29	45.87	39.74	39.58	<u>19.13</u>	19.59
FD001→FD004	47.27	43.82	31.78	31.23	21.93	<u>16.79</u>
FD002→FD001	32.79	28.10	14.33	13.88	15.09	<u>12.38</u>
FD002→FD003	34.50	37.46	32.60	33.53	19.13	<u>17.14</u>
FD002→FD004	14.99	31.85	34.35	33.71	23.38	<u>13.03</u>
FD003→FD001	11.66	31.74	19.97	19.54	17.64	<u>11.59</u>
FD003→FD002	36.34	44.62	23.47	19.33	<u>16.38</u>	16.96
FD003→FD004	37.31	47.94	26.33	20.61	19.49	<u>17.47</u>
FD004→FD001	20.17	31.54	37.89	20.10	19.15	<u>16.11</u>
FD004→FD002	14.74	24.93	28.77	18.50	20.30	<u>11.49</u>
FD004→FD003	29.98	27.84	<u>14.14</u>	14.49	14.86	16.02
Mean	29.36	37.01	26.93	23.669	18.76	<u>15.31</u>

表 2 对比实验结果(SCORE)

Method	Source only	DANN	ADDA	CADA	WADA	TADA
FD001→FD002	22209	93841	4426	2122	2043	<u>1090</u>
FD001→FD003	9929	27005	10711	8415	<u>1411</u>	4328
FD001→FD004	33846	57044	12369	11577	3260	<u>1408</u>
FD002→FD001	9861	8411	455	351	<u>485</u>	640
FD002→FD003	10653	17406	3449	5213	<u>1394</u>	5113
FD002→FD004	1146	66305	12227	15106	3279	<u>1506</u>
FD003→FD001	669	5113	1391	1451	1191	<u>890</u>
FD003→FD002	10903	37297	20764	5257	2156	<u>1672</u>
FD003→FD004	11502	141117	48748	3219	2594	<u>1919</u>
FD004→FD001	2827	7586	121677	1840	1427	<u>1251</u>
FD004→FD002	1376	17001	62215	4460	3963	<u>771</u>
FD004→FD003	7569	5941	537	682	<u>413</u>	2003
Mean	10207	40339	24914	4974	1968	<u>1883</u>

3.4.2 消融实验

为进一步验证TADA中各个模块的有效性，针对TEM和MI模块，以RMSE和SCORE作为评价指标，在C-MAPSS数据集上进行了五组消融实验，结果见表3。在TEM模块中，1为时间步级别的局部退化信息评估模块，2为RUL预测级别的局部退化趋势评估模块。

由1、2组实验和1、3组实验可知，TEM模块通过让跨域特征的局部退化信息和退化趋势在源域和目标域上难以区分，增强了模型的可迁移性；由4、5组实验可知，MI模块引入目标约束，保留了更多目标域的线性退化信息，进一步提高了RUL预测性能。

表 3 消融实验结果

Number	MI	TEM		RMSE	SCORE
		1	2		
1	×	×	×	25.72	6370
2	×	×	√	18.35	3018
3	×	√	×	17.83	2851
4	×	√	√	15.81	2325
5	√	√	√	<u>15.31</u>	<u>1883</u>

4 结语

本文对跨域场景下的航空发动机剩余寿命预测任务展开研究，提出了一种基于可迁移对抗的跨域自适应寿命预测方法(TADA)，为时间序列设置不同级别的局部域鉴别器，通过对抗

的方式评估了特征的局部可迁移性,且在跨域自适应中引入目标语义约束,进一步保留目标域退化信息,从而得到既具有局部可迁移性,又具有目标语义重要性的特征。为了证实TADA的有效性,本文在数据集CMAPSS上和其他域自适应方法进行对比实验,实验结果表明,TADA在RMSE和SCORE两个指标上均优于对比方法,证实了TADA的有效性。

TADA对于跨域迁移的分析,均是在时域场景下进行。在未来的研究中,可以考虑增加频域语义重要性和可迁移性的分析,进一步提升跨域自适应学习的泛化能力。

参考文献:

- [1] KHAN S, YAIRI T. A review on the application of deep learning in system health management[J]. Mechanical Systems and Signal Processing, 2018, 107: 241-265.
- [2] FARAHANI A, VOGHOEI S, RASHEED K, et al. A brief review of domain adaptation[C]//Advances in Data Science and Information Engineering: Proceedings from ICDATA 2020 and IKE 2020, 2021: 877-894.
- [3] MAO W, HE J, ZUO M J. Predicting remaining useful life of rolling bearings based on deep feature representation and transfer learning[J]. IEEE Transactions on Instrumentation and Measurement, 2019, 69(4): 1594-1608.
- [4] CHEN Z, HE G, LI J, et al. Domain adversarial transfer network for cross-domain fault diagnosis of rotary machinery[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(11): 8702-8712.
- [5] HUANG C G, ZHU J, HAN Y, et al. A novel Bayesian deep dual network with unsupervised domain adaptation for transfer fault prognosis across different machines[J]. IEEE Sensors Journal, 2021, 22(8): 7855-7867.
- [6] ZHANG J, JIANG Y, LI X, et al. An adaptive remaining useful life prediction approach for single battery with unlabeled small sample data and parameter uncertainty[J]. Reliability Engineering & System Safety, 2022, 222: 108357.
- [7] ZHAO R, YAN R, CHEN Z, et al. Deep learning and its applications to machine health monitoring[J]. Mechanical Systems and Signal Processing, 2019, 115: 213-237.
- [8] 全航, 张强, 邵思羽, 等. 基于CNN-WaveNet的滚动轴承剩余寿命预测[J]. 计算机应用研究, 2021, 38(10): 3098-3103.
- [9] GUO L, LI N, JIA F, et al. A recurrent neural network based health indicator for remaining useful life prediction of bearings[J]. Neurocomputing, 2017, 240: 98-109.
- [10] HUANG C G, HUANG H Z, LI Y F. A bidirectional LSTM prognostics method under multiple operational conditions[J]. IEEE Transactions on Industrial Electronics, 2019, 66(11): 8792-8802.
- [11] ZHANG Y, XIN Y, LIU Z, et al. Health status assessment and remaining useful life prediction of aero-engine based on BiGRU and MMoE[J]. Reliability Engineering & System Safety, 2022, 220: 108263.
- [12] MO Y, WU Q, LI X, et al. Remaining useful life estimation via transformer encoder enhanced by a gated convolutional unit[J]. Journal of Intelligent Manufacturing, 2021, 32: 1997-2006.
- [13] 李晶晶, 孟利超, 张可, 等. 领域自适应研究综述[J]. 计算机工程, 2021, 47(6): 1-13.
- [14] DA COSTA P R O, AKÇAY A, ZHANG Y, et al. Remaining useful lifetime prediction via deep domain adaptation[J]. Reliability Engineering & System Safety, 2020, 195: 106682.
- [15] RAGAB M, CHEN Z, WU M, et al. Adversarial transfer learning for machine remaining useful life prediction[C]//Proceedings of the 2020 IEEE International Conference on Prognostics and Health Management (ICPHM), Detroit, MI, USA: IEEE Press, 2020: 1-7.
- [16] RAGAB M, CHEN Z, WU M, et al. Contrastive adversarial domain adaptation for machine remaining useful life prediction[J]. IEEE Transactions on Industrial Informatics, 2020, 17(8): 5239-5249.
- [17] SHI H, HUANG C, ZHANG X, et al. Wasserstein distance based multi-scale adversarial domain adaptation method for remaining useful life prediction[J]. Applied Intelligence, 2023, 53(3): 3622-3637.
- [18] LI X, DING Q, SUN J Q. Remaining useful life estimation in prognostics using deep convolution neural networks[J]. Reliability Engineering & System Safety, 2018, 172: 1-11.
- [19] GANIN Y, USTINOVA E, AJAKAN H, et al. Domain-adversarial training of neural networks[J].

- The Journal of Machine Learning Research, 2016, 17 (1):2096-2030.
- [20] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C] //Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, Canada, 2014:2672-2680.
- [21] HU W, MIYATO T, TOKUI S, et al. Learning discrete representations via information maximizing self-augmented training[C]//Proceedings of the 34th International Conference on Machine Learning. Sydney NSW Australia:MIT Press, 2017:1558-1567.
- [22] GRANDVALET Y, BENGIO Y. Semi-supervised learning by entropy minimization [C] //Proceedings of the 17th International Conference on Neural Information Processing Systems, 2004:529-536.
- [23] CUI S, WANG S, ZHUO J, et al. Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations [C] //Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE Press, 2020:3941-3950.
- [24] CHAPELLE O, ZIEN A. Semi-supervised classification by low density separation [C] //Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics, 2005:57-64.
- [25] HUANG C G, HUANG H Z, LI Y F. A bidirectional LSTM prognostics method under multiple operational conditions [J]. IEEE Transactions on Industrial Electronics, 2019, 66(11):8792-8802.
- [26] XIANG S, QIN Y, LUO J, et al. Multicellular LSTM-based deep learning model for aero-engine remaining useful life prediction [J]. Reliability Engineering & System Safety, 2021, 216:107927.
- [27] HUANG C G, HUANG H Z, LI Y F. A bidirectional LSTM prognostics method under multiple operational conditions [J]. IEEE Transactions on Industrial Electronics, 2019, 66(11):8792-8802.

Transferable cross-domain adaptive learning for aircraft engine remaining useful life prediction

Li Wenxiao¹, Li Yongcheng¹, Li Peng², Ma Haotong³, Lei Yinjie^{1*}

(1. College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China; 2. CETC Key Laboratory of Avionic Information System Technology, The 10th Research Institute of China Electronics Technology Group Corporation, Chengdu 610036, China; 3. Key Laboratory of Optical Engineering, Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu 610036, China)

Abstract: In the task of predicting the remaining useful life (RUL) of aviation engines, the scarcity of labeled data and the variability of operating conditions result in significant distribution differences among sensor time series, hampering the generalization ability of RUL prediction methods. Cross-domain learning offers a feasible solution to this challenge. Traditional cross-domain learning minimizes the distribution discrepancy between the source and target domains to obtain aligned features, facilitating cross-domain knowledge transfer. However, as aviation engines degrade, the semantic information between consecutive time steps changes, causing local semantic mismatches in the previously aligned features, which adversely affects model performance. To address this issue, this paper explores a transfer adversarial approach for cross-domain RUL prediction, optimizing the probability entropy of local domain discriminator outputs to make aligned features indistinguishable at the local level. The method utilizes the target mutual information during RUL prediction to impose semantic constraints, resulting in domain-invariant features with both local transferability and target semantic importance, thus enhancing the model's generalization ability. Experimental results on the CMAPSS aviation engine dataset demonstrate the effectiveness of this approach, outperforming existing cross-domain adaptation methods in terms of RMSE and SCORE metrics.

Keywords: remaining useful life prediction; cross-domain learning; domain adaptation; transferable adversarial

文章编号: 1007-1423(2024)04-0009-08

DOI: 10.3969/j.issn.1007-1423.2024.04.002

基于 FaceNet 的犬只鼻纹识别

李梦晗, 张晨蕾, 田存伟*

(聊城大学物理科学与信息工程学院, 聊城 252000)

摘要: 随着养宠家庭的增加, 犬只身份识别成为一个重要问题, 犬鼻纹独一无二, 可以用来准确地辨认犬只身份。为了实现犬鼻纹识别, 设计了一种基于 FaceNet 网络的犬只鼻纹识别方法。FaceNet 的 Backbone 采用 ResNet18, 损失函数采用 Triplet loss。并通过 SRGAN 网络对犬鼻纹图像预处理, 放大图像分辨率, 使鼻纹图像更加清晰。基于 FaceNet 网络的犬只鼻纹识别方法为犬只身份识别和鉴定提供了一种有效可行的解决方案。

关键词: FaceNet; 犬鼻纹识别; ResNet18

0 引言

近年来, 越来越多的人选择独居或者小家庭化, 在这种情况下, 宠物成为了陪伴和情感支持的重要角色, 而且与宠物相处可以降低人们在生活中产生的压力, 缓解焦虑和孤独感, 增加人们的幸福感。随着越来越多的家庭选择养宠物犬陪伴自己, 宠物在人们生活中的重要性不断增加。然而养宠也带来了一系列问题, 包括宠物店的监管和喂养、宠物的丢失和寻找, 以及宠物管理、交易、保险和医疗等方面都需要更智能和系统化的解决方案。

传统的宠物识别一般采用耳洞、烙印、微芯片植入、标签等方法, 据各种研究显示, 微芯片已被证实某些情况下对动物产生副作用^[1]。这些方法不仅会对动物造成精神和身体上的伤害, 还容易被复制和伪造, 这就会导致识别不出来宠物的身份。所以基于生物识别的方法变得越来越受欢迎, 比如: 基于鼻纹、虹膜、视网膜血管、毛发图案和面部图像的特征识别。其中面部图像识别是最广泛使用的, 但是宠物的脸会受到各种光照变化、姿势和大规模纹理变化的影响。然而, 犬只鼻纹认证是一个具有挑

战性的问题, 因为缺乏大量的标记数据^[2]。

如果两只犬属于同一品种的话, 它们会有许多相似之处, 不同品种的犬只也有惊人的相似之处, 而持续的杂交进一步模糊了犬之间的区别。宠物犬的鼻纹拥有独特不可复制的特性, 此外, 犬的鼻纹在成长过程中也不会改变, 所以鼻纹就成为了辨认犬类的最重要的特征。

本文基于 FaceNet 和 ResNet18 模型对犬鼻纹进行识别, 该模型在 FaceNet 的基础上, 对其头部网络进行改进, 采用 ResNet18 网络对犬鼻纹图片进行特征提取, 并使用 Triplet loss 损失函数, 在对数据集进行处理时通过 SRGAN 网络对犬鼻纹图像进行分辨率重建, 鼻纹图像更加清晰。经过实验验证, 本文提出的模型能较为准确地识别出犬只身份, 在数据集上准确率达到 94.1%。

1 相关理论

1.1 FaceNet 原理

FaceNet 是由 Google 提出的一种用于人脸识别和人脸验证的卷积神经网络。FaceNet 是对人脸识别、验证、聚类所有人脸问题进行系统解决的一个技术框架^[3]。对于整个 FaceNet 结构来说, 特征提取的部分可以使用不同的网

收稿日期: 2023-10-09 修稿日期: 2023-10-31

作者简介: 李梦晗(1998—), 女, 山东菏泽人, 硕士研究生, 研究方向为人工智能; 张晨蕾(1998—), 女, 山东淄博人, 硕士研究生, 研究方向为人工智能; *通信作者: 田存伟(1981—), 男, 山东聊城人, 硕士, 副教授, 研究方向为智能控制与信息处理, E-mail: tiancw2008@163.com

络结构，最初使用了两种深度卷积网络：Zeiler&Fergus 和 Google 的 Inception v1。FaceNet 网络结构如图 1 所示，采用深度卷积神经网络学习输入图像的欧氏空间特征，将得到的图像特征映射到多维空间中，得到对应的特征向量。对特征向量进行 L2 正则化，筛选出有效特征，使用正则化后的特征向量，最后输入 Triplets Loss 损失函数进行计算。

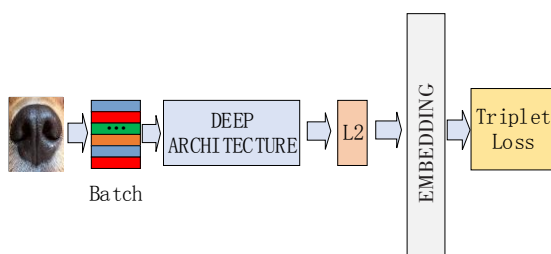


图 1 FaceNet 结构

FaceNet 模型通过将人脸图像特征映射到多维空间，采用计算欧氏空间距离的方法进行人脸相似度的判断^[4]。在 FaceNet 中空间距离的远近直接影响人脸图像的相似度大小，同一个人的图像在欧氏空间的距离比较小，不同人的图片特征之间的欧氏距离较大。在进行人脸验证时，为了确定两张人脸照片是否来自同一人，需要计算并设定一个欧氏距离的阈值，以便判定特征向量之间的距离。

1.2 ResNet18

本文采用了经典的卷积神经网络模型之一的 ResNet18 残差网络，其主要的特点是引入了残差块 (ResIDual Block)。残差块结构如图 2 所示。残差块下方的连接为“快捷连接”，可以跳过一层或多层，将网络信息传输到更深层^[5]，从而解决了卷积神经网络在增加层数时可能出

现的梯度消失和梯度爆炸问题，使深度神经网络不再退化，更容易训练，实现了恒等映射的功能。

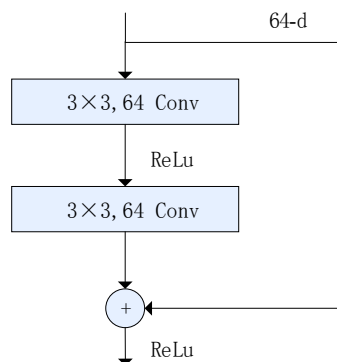


图 2 残差块结构

ResNet18 结构如图 3 所示，输入大小为 224×224 的 RGB 图像，经过一个卷积层，然后经过归一化处理 (batch normalization, BN) 和激活函数 (ReLU)，通过最大池化层将数据减半，从而减少存储空间，通道数不变。接着将得到的数据送进四个残差模块，每个残差模块由两个残差块组成，每个残差块由两个 3×3 的卷积层组成^[6]。最后对得到的特征图进行全局平均池化，将其转化为一维向量送入全连接层中分类输出。

1.3 Triplet loss

Triplet loss (三元损失) 是一种用于训练差异性小的样本的损失函数，与交叉熵相比，三元组损失对于带有偏倚标签信息的影响较小，并广泛用于细粒度视觉任务^[7]。Triplet loss 的优势在于它能够处理输入向量非常相似的情况，有助于区分那些差异较小的输入向量，捕捉微小细节的差异，同时也可以学习到更加细微的特征。图 4 为 Triplet loss 的结构示意图。

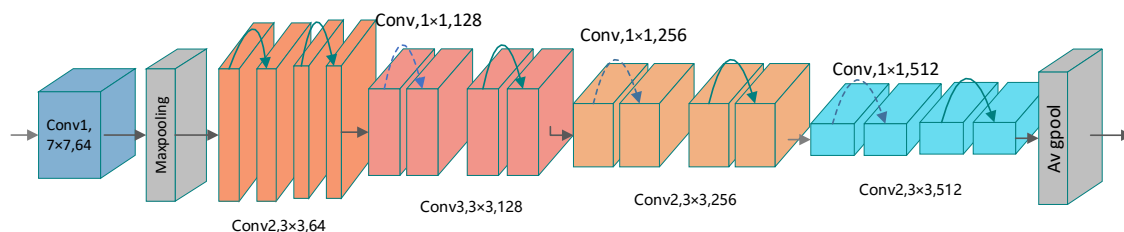


图 3 ResNet18 结构图

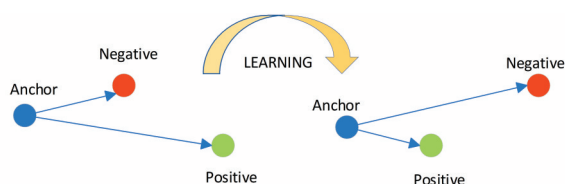


图4 Triplet loss 结构图

Triplet loss 中的三元对应了上图的锚点(Anchor)、负样本(Negative)、正样本(Positive)。其中：①锚点(A)：选择某类中任意一幅图像作为锚样本^[8]；②正样本(P)：表示与锚点属于同一个个体的其他图像；③负样本(N)：表示与锚点不同个体的图像。

Triplet loss 损失函数如式(1)：

$$\sum_i^N \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right] \quad (1)$$

式中： x_i^a 表示 Anchor； x_i^p 表示 Positive； x_i^n 表示

Negative； α 表示间隔值。

每次从训练样本中随机抽取一个 Anchor、一个与 Anchor 同一类的 Positive 以及和 Anchor 不同类的 Negative。使用 Triplet loss 优化调整三个样本之间的距离，使 Anchor 与 Positive 之间的距离变小，Anchor 与 Negative 之间的距离变大。

2 网络总体模型搭建

本文使用 FaceNet 网络作为主体网络架构，采用卷积神经网络 ResNet18 为头部网络将输入图像特征映射到多维空间中，提取到特征将其转化为特征向量。对得到的特征向量进行 L2 标准化。L2 标准化就是每个数据与 L2 范数相除，这个过程也就是将特征向量映射到同一个数量级上^[9]。最后 Triplet loss 优化调整图像特征向量的欧氏距离，从而更准确地区分不同犬鼻纹图像。FaceNet 整体网络结构如图 5 所示。

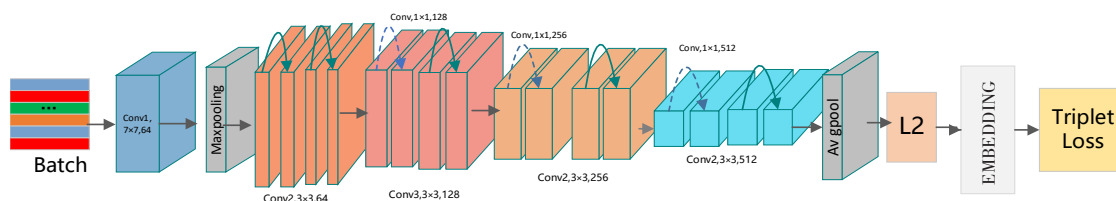


图5 总体网络结构

3 实验结果与分析

3.1 实验环境

实验环境见表 1。

表 1 实验环境

名称	配置
CPU	Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz 2.59 GHz
GPU	NVIDIA GeForce RTX 3090
GPU 加速库	CUDA 11.6.134、CUDNN 8.2.1
操作系统	Windows 10(64 位)
软件环境	Python3.8、PyTorch 1.10.2

3.2 建立犬鼻纹数据集

实验采用的数据集主要来自网络搜集、手机拍摄，一共为 6000 只犬，38437 张犬鼻纹图

像，每只犬至少有两张不同的犬鼻纹图片，每张犬鼻纹图片的模糊程度、明暗程度或者角度都不一样。每一只犬有属于自己的 ID，将其中的 5900 只犬的鼻纹图片用于模型训练，剩余的 100 只犬的鼻纹图像用来测试。部分数据集样本图片如图 6 所示。



图6 犬鼻纹数据集部分样本图片

3.3 使用 SRGAN 提升图像分辨率

3.3.1 SRGAN

在获取犬鼻纹的过程中往往会得到分辨率

较低或者不清晰的图像。针对该问题,本文提出采用SRGAN重建图像的分辨率,得到更清晰的犬鼻纹图像,清晰的图像提供了更高分辨率的视觉信息,有助于提高后续实验的可重复性和数据准确性。

超分辨率生成对抗网络(super-resolution generative adversarial network, SRGAN)^[10]结合了生成对抗网络(GAN)^[11]和超分辨率技术,提出了一个判别网络和一个生成网络来一起训练图像,以及一个感知损失函数,使重建的图形纹理更清晰,增强图像视觉效果^[12]。基于GAN的超分辨率网络训练中可能存在模式崩溃的情况^[13],使用均方差(mean squared error, MSE)为损失函数的普通超分辨模型在训练时会丢失一些高频细节。

3.3.2 SRGAN模型结构

SRGAN由生成器网络(generator network)和判别器网络(discriminator network)两部分组成,生成网络可以将图像作为 n 维向量输入来生成重建的图像输出^[14],主要作用是将低分辨率图片转换为高分辨率图片,低分辨率图像先经过卷积层和多个残差模块进行特征提取,然后进入

上采样层将特征图的分辨率逐渐提高。图7为生成器网络结构。

判别器网络用于判别生成的超分辨率图像和真实高分辨率图像,其目标是判断输入图像是否为生成器生成。图8为判别器网络结构,是由不断重复的Conv层、LR层和BN层所组成^[15]。输入高分辨率图像,多个卷积层进行特征提取,每个卷积层都添加了激活函数Leaky ReLU来引入非线性性质。全连接层将特征图转化为单一的概率分数,判断输入图像的真伪,输出层采用Sigmoid激活函数,将概率值设置为0~1之间,输出图像的概率。

3.3.3 SRGAN网络训练

训练SRGAN模型时,输入低分辨率图像到模型,生成的高分辨率图像可能会包含错误细节。生成器损耗可以通过比较从错误的高分辨率图像中获得的判别结果与实际图像来确定^[16]。将真实高分辨率图像和生成的高分辨率图像分别输入到VGG网络提取特征,比较两幅图像的特征得到判别损失。图9展示了生成损失和判别损失的变化曲线。

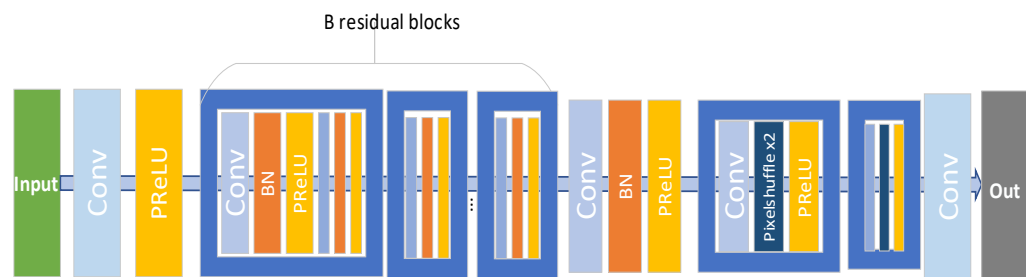


图7 生成器网络结构

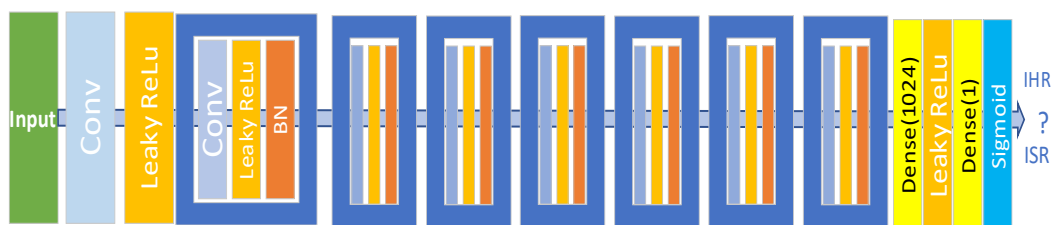


图8 判别器网络结构

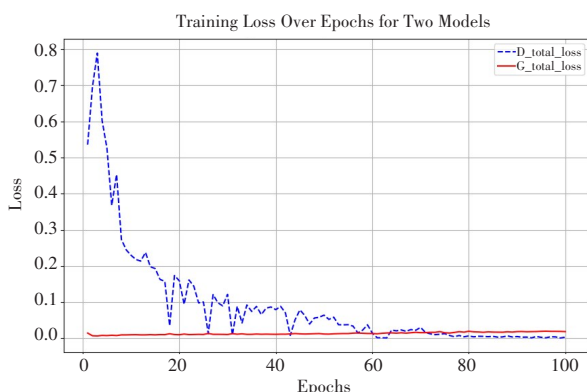


图9 生成损失和判别损失曲线图

3.3.4 SRGAN模型评估与实验结果

本实验使用PSNR(峰值信噪比)和SSIM(结构相似性指数)性能指标对模型进行评价。计算方式是通过测量图像的峰值信号和噪音直接按比例评估图像的失真程度, PSNR的数值越高表示图像质量越高。PSNR公式如下:

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX(2/I)}{MSE} \right) \quad (2)$$

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2 \quad (3)$$

SSIM是一种比较两幅图像之间结构相似性的指标, 其计算考虑了亮度、对比度和结构三个方面的相似性, 来综合评估图像之间的相似性, 取值在-1~1之间, 两幅图像相似性越高, 其数值越接近1。SSIM计算公式如下:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (4)$$

图10和图11为本实验的PSNR和SSIM曲线图, 最终模型的PSNR为29.56, SSIM为0.85。

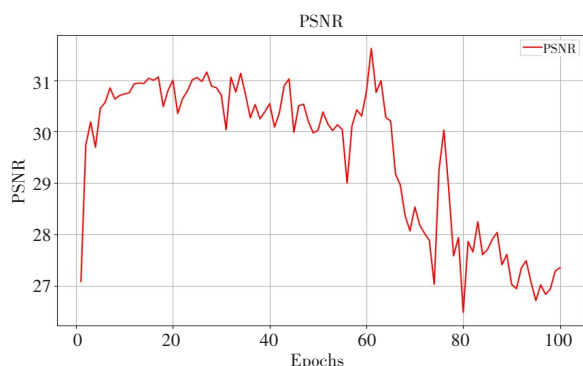


图10 PSNR变化曲线

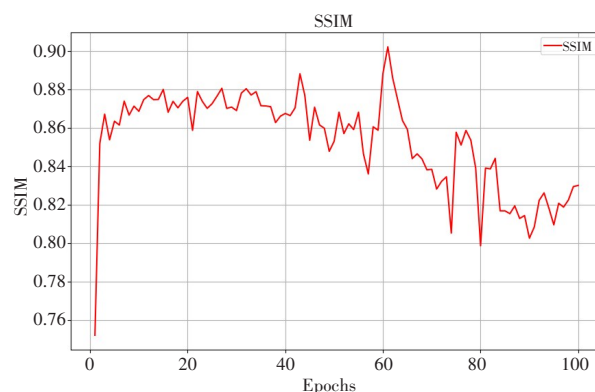


图11 SSIM变化曲线

经过SRGAN处理的犬鼻纹图像对比如图12所示, 从图中可以看出图像清晰度有明显的提升, 鼻子纹路更加清晰, 为下一步实验提供更可靠、清晰的数据集。

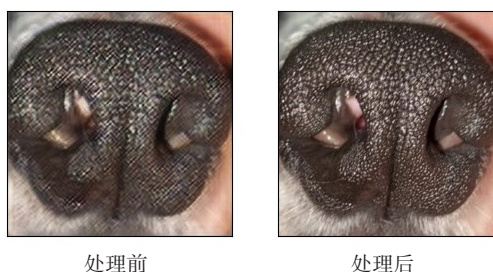


图12 经过SRGAN网络处理的犬鼻纹图像对比

3.4 网络训练及性能分析

基于FaceNet的犬只鼻纹识别模型的主干特征提取网络为ResNet18, 本实验采用SGD优化器, SGD算法一次只进行随机单个样本的迭代更新, 使得SGD并不是每次迭代都向着整体最优化方向^[17]。选择合适的学习率是本实验的重要一步, 过小的学习率会导致训练进展缓慢, 收敛速度减慢; 学习率较大会导致训练不稳定, 使其在最优解周围震荡。为了选择一个合适的学习率, 分别选取0.01、0.001对模型进行训练, 在选取学习率时训练用max_steps控制训练循环, max_steps可以控制训练时长, 防止过拟合。图13、图14是max_steps设置为30000时不同学习率下的训练损失曲线。

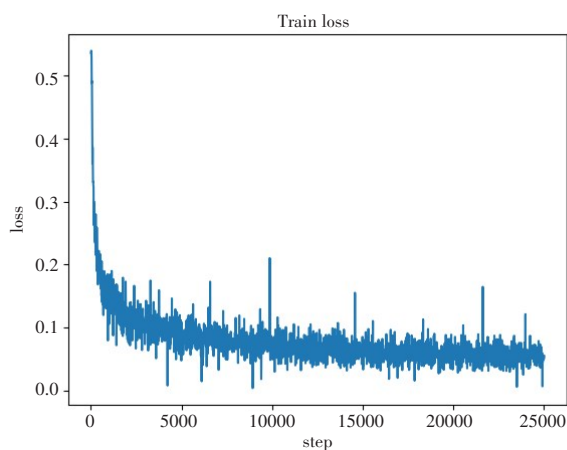


图 13 学习率为 0.001 的 loss 曲线

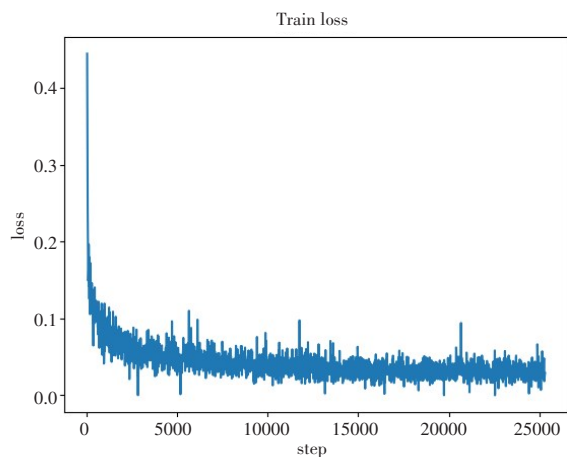


图 14 学习率为 0.01 的 loss 曲线

通过图 13 和图 14 对比分析得知, loss 越小, 模型的检测性能越好; loss 越大, 模型的检测性能越差。学习率为 0.01 时模型更加稳定, 所以本文选用 0.01 学习率训练测试模型。

本实验建立一个 common 网络与 ResNet18 网络对比, 该网络由三层卷积层组成, 每个卷积层后面加入一个最大池化层, 是一个简单的卷积神经网络。图 15 为两个网络训练损失的对比图。可以看出, 采用 ResNet18 为特征提取网络的模型 loss 曲线更稳定。

3.5 模型评价指标

本实验的评价指标主要有: 精度(Precision)、召回率(Recall)与准确度(Accuracy)。它们主要由 TP、FP、TN、FN 计算得出。其中 TP 是指将正样本预测为正样本, FP 是指将负样本预测为正样本, FN 是将正样本标签预测为负样本, TN

是负样本预测正确。其中精确度、召回率、准确率的公式如下:

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

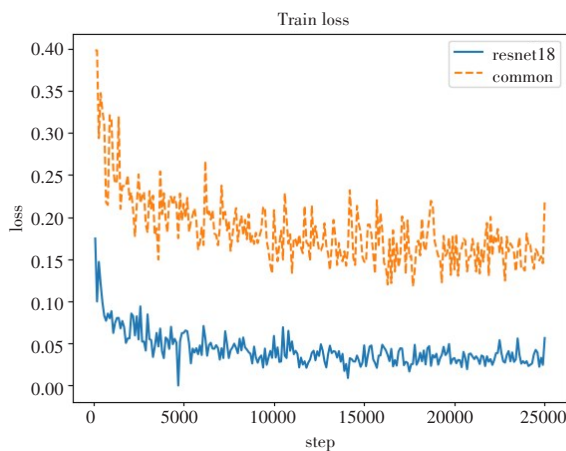


图 15 不同特征提取网络 loss 对比曲线

为了评估模型性能, 该实验设计了基于 FaceNet 的两种特征提取网络 ResNet18 和 common 网络, 分别在数据集上进行训练, 结果见表 2。

表 2 不同特征提取网络性能对比

网络	准确率/%	召回率/%	精确率/%	时间/s
ResNet18	94.1	97.8	95.7	72
common	84.2	92.9	88.8	91

通过实验可以看出, 采用 ResNet18 为特征提取网络的准确率、召回率和精确率都比较高, 测试速度也优于 common 网络。

3.6 预测结果

本实验将测试目录中的图片都进行了预测, 也可以只选取一张图片预测。

如图 16 所示, 这三张图片为同一 ID 的不同预测结果。该犬只鼻子被栏杆遮挡一部分, 所以未能准确提取全部正确的特征, 所以识别结果出现了多种可能。还有一张测试图像未能被成功识别。

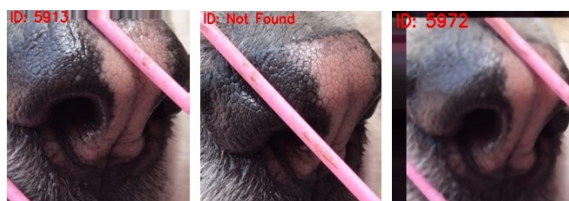


图 16 遮挡的鼻纹图像

图 17 中三张图片是部分识别正确的犬鼻纹图像。

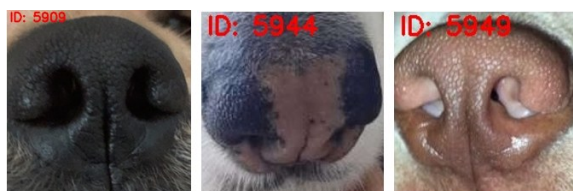


图 17 正确识别的结果

4 结语

本文提出一种基于FaceNet的犬只鼻纹识别模型，通过ResNet18网络提取犬鼻纹的细节特征，ResNet18网络中残差块利用跳跃连接能够训练更深层次的卷积神经网络，并且解决了梯度消失和梯度爆炸的问题，提高了模型的训练效率，使网络能够更好地学习复杂的犬鼻纹特征。实验结果表明，该模型的准确率能达到94.1%，可以较为准确地识别犬只鼻纹，基本满足了用于识别犬只身份的需求，但是对遮挡鼻纹的预测效果还不是很理想，后续还会继续优化，以提高识别精度，并使其更适用于各种实际场景。

参考文献：

- [1] IN H C, LEE Y S, SHIN H J, et al. The formation and invariance of canine nose pattern of beagle dogs from early puppy to young adult periods[J]. *Animals*, 2021, 11(9): 2664-2664.
- [2] SHEN F, WANG Z, WANG Z J, et al. A competitive method for dog nose-print re-identification[EB/OL]. arXiv:2205.15934, 2022.
- [3] 张颖. 基于FaceNet改进的人脸识别算法[J]. *电子技术*, 2022, 51(3): 25-27.
- [4] 李志华, 张见雨, 魏忠诚. 基于MTCNN和Facenet的人脸识别系统设计[J]. *现代电子技术*, 2022, 45(4): 139-143.
- [5] XUE G H, LI S X, HOU P, et al. Research on light-weight YOLO coal gangue detection algorithm based on resnet18 backbone feature network[J]. *Internet of Things*, 2023, 22: 100762.
- [6] 刘志, 高东明. 不同深度学习模型在餐厨垃圾类型识别中的应用与比较[J/OL]. *环境工程*, 2023: 1-9. <https://kns.cnki.net/kcms/detail/11.2097.X.2023-0829.1802.006.html>.
- [7] LU Z Y, LU Y G. A balanced triplet loss for person re-identification[J]. *International Journal of Pattern Recognition and Artificial Intelligence*, 2023, 37(1): 22-30.
- [8] 刘世锋, 常蕊, 李斌, 等. 基于RGB-D图像的牛只个体识别方法[J/OL]. *农业机械学报*, 2023: 1-10. <http://kns.cnki.net/kcms/detail/11.1964.s.20230914.1340.019.html>.
- [9] 郑倩. 牛唇纹身份识别应用系统的设计与实现[D]. 内蒙古: 内蒙古科技大学, 2022.
- [10] LEDIG C, THEIS L, HUSZÁR F. Photo-realistic single image super-resolution using a generative adversarial network[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, NJ, USA, 2017: 105-114.
- [11] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M. Generative adversarial nets[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems, MA, USA, 2014: 2672-2680.
- [12] LIU Z H, HE P F, WANG F F. PCB defect images super-resolution reconstruction based on improved SRGAN[J]. *Applied Sciences*, 2023, 13(11): 6786.
- [13] 刘慧, 卢云志, 张雷. 基于Dropout改进的SRGAN网络DrSRGAN[J]. *科学技术与工程*, 2023, 23(23): 10015-10022.
- [14] LI W H, HE D, LIU Y Q, et al. Super-resolution reconstruction, recognition, and evaluation of laser confocal images of hyperaccumulator *Solanum nigrum* endocytosis vesicles based on deep learning: comparative study of SRGAN and SRResNet[J]. *Frontiers in Plant Science*, 2023, 14(14): 1146485-1146485.
- [15] 李昊, 赵光哲. DRUSR: 面向效果的图像超级分辨率重建[J]. *计算机工程与应用*, 2023, 59(24): 165-175.
- [16] ZHANG J, WANG X X, LIU J Y, et al. Multispectral drone imagery and SRGAN for rapid phenotypic mapping of individual Chinese cabbage plants[J]. *Plant Phenomics*, 2022, 2022: 0007.
- [17] 杨波, 何金平, 张立娜. 基于改进SPP-x的YOLOv5神经网络水稻叶片病害识别检测[J]. *中国农机化学报*, 2023, 44(9): 190-197.

(下转第28页)

文章编号: 1007-1423(2024)04-0016-07

DOI: 10.3969/j.issn.1007-1423.2024.04.003

基于距离视图表示与逐点细化结合的点云语义分割方法

陈 晗, 何鸿添, 雷印杰*

(四川大学电子信息学院, 成都 610065)

摘要: 三维点云语义分割是机器实现环境感知的重要途径。在现有的研究中, 基于体素的算法和基于点的算法在面对大规模的点云数据时计算效率低下。而基于距离视图的算法, 在对点云进行投影和反投影时会不可避免地造成精度损失。针对上述问题, 提出了基于距离视图表示与逐点细化结合的点云语义分割新框架 RpNet。充分的实验表明, 所提出的方法在三维点云室外场景数据集 SemanticKITTI 上的平均交并比达到 64.2%, 推理速度达到 58 帧/秒, 兼顾了高精度和高速度。

关键词: 语义分割; 三维点云; 距离视图; 逐点细化

0 引言

随着激光雷达传感器的快速发展, 三维视觉技术在自动驾驶、虚拟现实、机器人导航等领域得到了广泛的应用。与视觉相机相比, 激光雷达不受光照条件和天气状况的影响, 更具鲁棒性。同时, 激光雷达采集到的点云与二维图像相比, 蕴含更加丰富的几何结构信息。因此, 基于三维点云的环境感知技术, 特别是逐点分类的语义分割技术, 成为人们研究的重点。

近年来, 相关研究人员提出了诸多方案来解决三维点云的语义分割问题。基于点的算法^[1-5]需要对点云进行多次复杂的逐点操作, 基于体素的算法^[6-9]则需要使用三维卷积, 这在面对大规模的点云数据时计算效率受限。基于距离视图的算法^[10-17]首先将点云投影到二维平面, 再利用成熟高效的二维算法实现语义分割, 最终将结果反投影回三维空间。这类方案在速度上具有明显优势, 但是在投影和反投影时存在量化误差和多对一映射的问题, 限制了计算精度。基于距离视图的方案还存在另外一个问题,

即点云投影至二维平面后并不能完全占满所有的像素, 这就导致距离视图中存在许多空值, 这些空值产生的噪声也会对分割性能造成一定的影响。

针对上述问题, 本文提出了一种基于距离视图表示与逐点细化结合的三维点云语义分割方法 RpNet。该模型对原始距离视图中的每个空值点进行信息补全, 同时结合了基于距离视图的编码器和基于逐点细化的解码器, 在准确度和推理速度上均具有优越性。

1 本文方法

RpNet 的整体框架结构如图 1 所示。对输入的原始点云数据, 首先使用球面投影将其转化成距离视图表示, 然后利用距离视图补全算法将空值像素补全。再通过基于距离视图的编码器提取多尺度特征, 最后送入基于逐点细化的解码器中, 得到最终的语义分割结果。

1.1 球面投影

对于获得的非结构化点云数据, 通过球面

收稿日期: 2023-10-03 修稿日期: 2023-10-26

基金项目: 国家自然科学基金面上项目(62276176)

作者简介: 陈晗(1998—), 男, 湖南常德人, 硕士研究生, 研究方向为计算机视觉; 何鸿添(1997—), 男, 广东佛山人, 硕士研究生, 研究方向为计算机视觉; *通信作者: 雷印杰(1983—), 男, 四川成都人, 教授, 博导, 博士, 研究方向为人工智能与计算机视觉, E-mail: yinjie@scu.edu.cn

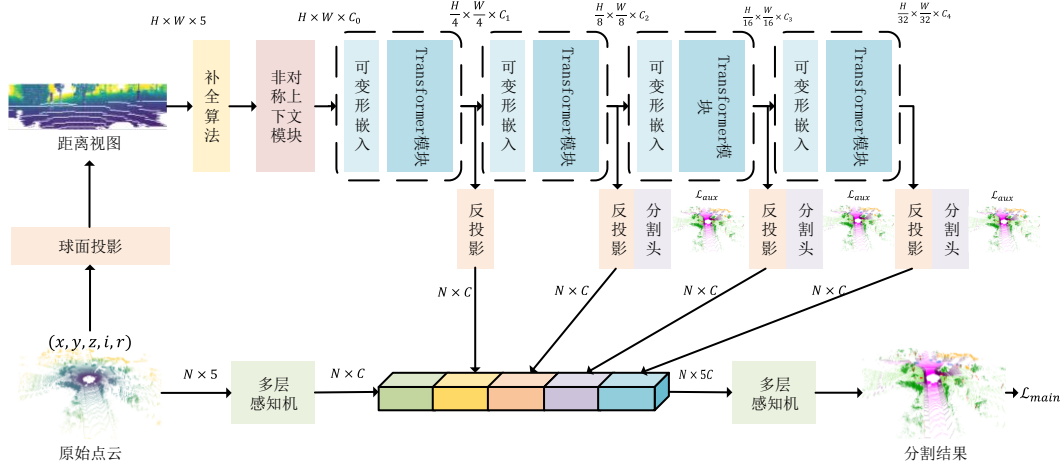


图1 RPNet的整体框架结构

投影将其转化成有序的距离视图表示，将点云坐标 (x, y, z) 映射成图像坐标 (u, v) 的计算公式如下：

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{1}{2} [1 - \arctan(y, x) \pi^{-1}] W \\ [1 - (\arcsin(z, r^{-1}) + f_{\text{down}}) f^{-1}] H \end{pmatrix}$$

其中： $f = |f_{\text{down}}| + |f_{\text{up}}|$ 表示激光雷达的垂直视场大小， $r = \sqrt{x^2 + y^2 + z^2}$ 表示每个点的距离。将点云坐标 (x, y, z) ，反射强度 i 和距离 r 共同组成大小为 $[5, H, W]$ 的张量作为输入。

1.2 距离视图补全算法

通过球面投影将点云投影到二维平面后，由于点的数量小于二维像素数量，这将导致投影后的距离视图中存在许多空值，从而影响语义分割的性能。本文提出了一种距离视图补全算法，首先根据周围坐标 x_i 和掩码 m_i ，计算均值 $x_c = \sum_i x_i / \sum_i m_i$ 作为该空值的坐标，然后计算各相邻坐标 x_i 与空值坐标 x_c 的距离 $d_i = \|x_i - x_c\|_2$ ，再得到相邻距离 d_i 中最小距离处的索引 $\text{index} = \text{argmin}(d_i)$ ，最后根据该索引从相邻特征中得到空值处的特征值 $f = \text{Gather}(f_i, \text{index})$ 。算法流程如图2所示。

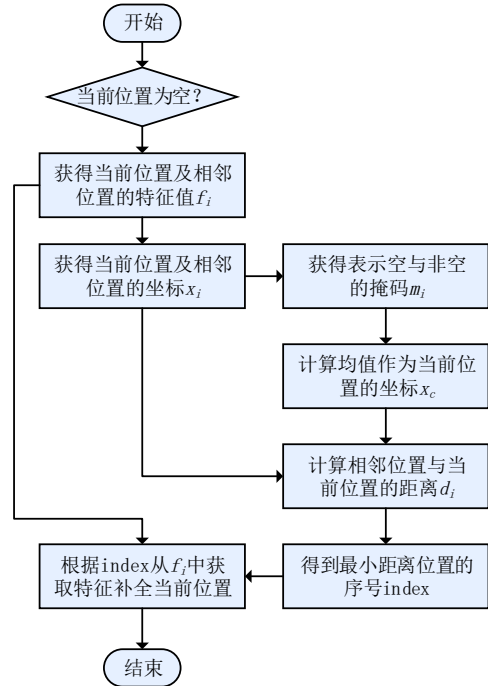


图2 距离视图补全算法流程图

1.3 基于距离视图的编码器

1.3.1 编码器基本结构

编码器采用经典的金字塔结构来提取多尺度特征。模型总共分为四个阶段，每个阶段由块嵌入层和Transformer^[18]模块组成。块嵌入层利用跨步卷积将图像分块嵌入到特征空间并转

化成序列,同时可以实现下采样。Transformer模块使用Segformer^[19]中的自注意力层和前馈层,同时为了缓解模型的过拟合现象,本文还在自注意力层中引入了DropKey^[20]策略。

1.3.2 非对称上下文模块

在自动驾驶场景中广泛存在长宽差异明显的各向异性物体,如行人、电杆等。受文献[8]启发,本文提出了非对称上下文模块(asymmetrical context module, ACM)来提取浅层上下文信息,以更好地匹配不同形状的物体,结构如图3所示。

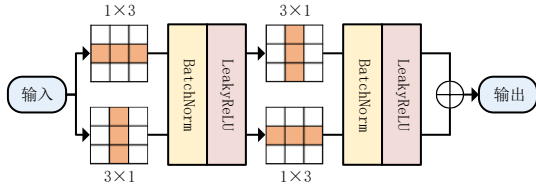


图3 非对称上下文模块示意图

模块分为两个分支,其中一个分支先后经过 3×1 和 1×3 的卷积核,另一个分支则先后经过 1×3 和 3×1 的卷积核,最后再将二者相加得到最终的特征。通过这种设计,增强了神经网络的垂直响应和水平响应。

1.3.3 可变形嵌入模块

将点云投影到二维平面后产生了严重的扭曲变形,当使用传统的正方形块嵌入处理距离视图时,每个图像块中将不可避免地包含许多不相关区域。为了解决这个问题,本文提出了可变形嵌入模块(deformable patch embedding, DPE)。可变形嵌入模块将块嵌入层中的常规卷积替换成带有调制机制的可变形卷积^[21],从而能够自适应嵌入不同形状的图像块。可变形卷积的计算如图4所示。

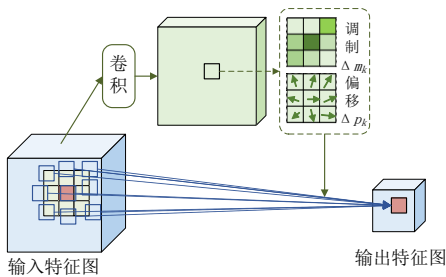


图4 调制可变形卷积示意图

给定 K 个位置采样的卷积核,记 w_k 和 p_k 分别是第 k 个位置的权重和偏移量,则对于 3×3 的卷积核来说, $K=9$, $p_k \in \{(-1, -1), (-1, 0), \dots, (1, 1)\}$, $(-1, -1)$ 表示左上角的位置, $(1, 1)$ 表示右下角的位置。记 $y(\cdot)$ 和 $x(\cdot)$ 为当前位置的特征值,对于位置 p 来说,传统卷积的计算过程为

$$y(p) = \sum_{k=1}^K w_k \cdot x(p + p_k) \quad (2)$$

调制可变形卷积通过一个额外的卷积获取每个位置的偏移量 Δp_k 和调制量 Δm_k ,前者实现不同形状的卷积核,后者对不同位置赋予不同的感知强度,计算过程如下:

$$y(p) = \sum_{k=1}^K w_k \cdot x(p + p_k + \Delta p_k) \cdot \Delta m_k \quad (3)$$

1.4 基于逐点细化的解码器

本文提出的基于逐点细化的解码器结构如图1下半部分所示。从基于距离视图的编码器中获得4个不同层次的特征张量 $F_1 \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times C_1}$, $F_2 \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times C_2}$, $F_3 \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times C_3}$, $F_4 \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times C_4}$ 。使用一个线性层将其统一到相同大小的特征通道,然后通过反投影将不同尺度的二维特征图还原到三维空间,得到4组与原始点云序列长度一致的特征 $F_1 \in \mathbb{R}^{N \times C}$, $F_2 \in \mathbb{R}^{N \times C}$, $F_3 \in \mathbb{R}^{N \times C}$, $F_4 \in \mathbb{R}^{N \times C}$ 。

为了弥补投影和反投影带来的三维空间信息损失,本文使用逐点细化的解码方式。对原始非结构化点云使用一个多层感知机提取逐点特征 F_{point} ,然后将逐点特征和前面获得的反投影后的4个层次特征拼接,最后再经过一个MLP得到逐点语义分割结果 $\hat{y}$,计算过程如下式:

$$\hat{y} = \text{MLP} \left(\text{Concat} \left(\text{MLP} \left(F_{\text{point}} \right), F_1, F_2, F_3, F_4 \right) \right) \quad (4)$$

基于逐点细化的解码器融合了原始点云特征和多尺度特征,因此具有强大的表征能力,同时由于只使用了两个多层感知机,计算效率高。

1.5 网络训练与推理

在网络训练阶段,通过最小化交叉熵(cross entropy, CE)损失函数^[22]和Lovász损失函数^[23]来完成网络的训练。数据类别不平衡是限制语义分割性能的重要因素,因此本文引入了在线难

样本采样^[24](online hard example mining, OHEM)策略。为了充分利用不同尺度的特征,本文还增加了3个辅助损失来帮助网络模型训练。

CE损失函数是语义分割任务中最常用的损失函数,每个像素的CE损失计算如下:

$$L_{CE} = - \sum_{c=1}^C y(c) \log(p(c)) \quad (5)$$

其中, C 表示类别数, $y(c)$ 和 $p(c)$ 表示该像素第 c 类的标签值和预测概率。为了提升模型对困难样本的学习能力,本文使用了OHEM。预设一个损失阈值 t 和最小像素数 n , 当损失大于 t 的像素数量大于 n 时, 取所有大于 t 的损失计算总损失; 否则, 取最大的 n 个损失计算总损失。

Lovász 损失函数是针对交并比(intersection over union, IoU)设计的损失函数, 其计算公式如下:

$$L_{ls} = \frac{1}{C} \sum_{c \in C} \bar{\Delta}_{jc}(m(c), m_i(c)) = \begin{cases} 1 - p_i(c), & c = y_i(c) \\ p_i(c), & \text{otherwise} \end{cases} \quad (6)$$

其中, $\bar{\Delta}_{jc}$ 表示对 IoU 的 Lovász 拓展, C 表示类别数, $y(c)$ 和 $p(c)$ 表示该像素第 c 类的标签值和预测概率。主干部分总损失函数为

$$L_{main} = OHEM(L_{CE}) + L_{ls} \quad (7)$$

受文献[17]启发, 本文对后3个阶段得到的特征用一个辅助分割头得到粗分割结果, 并根据上述损失函数计算损失值 L_{aux}^i 。一方面, 对反投影后的结果直接监督, 可以减弱上采样带来的精度损失; 另一方面, 对不同层次的结果进行监督, 可以提升模型对多尺度特征的学习能力。最终的损失函数为

$$L = L_{main} + \sum_{i=1}^3 L_{aux}^i \quad (8)$$

在推理阶段, 直接加载已训练好模型, 对输入三维点云进行逐点正向传播概率预测, 得到语义分割结果。

2 实验与分析

2.1 实验数据集

本文在 SemanticKITTI 数据集上训练和评估网络模型。SemanticKITTI^[25]是一个用于自动驾驶场景点云分割任务的大规模数据集, 它包含22个序列共计43551帧激光雷达扫描。其中序列00至10共计19130帧都包含密集的逐点标签,

用于网络训练。序列11至21共计20351帧用于在线测试。和通常的设置一样, 使用序列08共计4071帧用于验证。

2.2 评价指标

为了公平比较, 使用语义分割任务中通用的平均交并比(mean intersection over union, mIoU)对性能进行评估:

$$mIoU = \frac{1}{C} \sum_{c=1}^C \frac{TP_c}{TP_c + FP_c + FN_c} \quad (9)$$

其中, TP_c 、 FP_c 、 FN_c 分别表示第 c 类的真正值(true positive)、假正值(false positive)、假负值(false negative)。

2.3 实验环境和参数设置

本实验在 Ubuntu 22.04 系统下进行, 代码基于 Python 3.8.16 编程语言和 PyTorch 1.13.1 深度学习框架搭建。硬件平台 CPU 为 Intel(R) Xeon (R) E-2224 CPU@3.40GHz, GPU 为 NVIDIA GeForce RTX 3090。代码可在 <https://github.com/ch2718n/RPNet> 上获取。

模型训练时采用 AdamW 优化器和 Cosine-Warmup 优化策略。初始学习率为 0.00001, 权重衰减系数为 0.01, 总训练轮次为 100, Warmup 轮次为 10, 最大学习率为 0.001, batch-size 为 8。

球面投影尺寸为 64×2048。对于基于距离视图的编码器, 4个阶段的下采样因子为 [4, 2, 2, 2], 特征通道数为 [64, 128, 320, 512], 注意力头数为 [1, 2, 5, 8], MLP 通道扩展比为 [4, 4, 4, 4], 模块数为 [2, 2, 2, 2], 缩减比例为 [8, 4, 2, 1], DropKey 系数为 [0.3, 0.2, 0.1, 0]。对于基于逐点细化的解码器, 通道数设置为 128。

2.4 实验结果与分析

2.4.1 定量数值对比

为了探究各类算法的特点和验证提出算法的性能, 本文将 RPNet 在 SemanticKITTI 测试集上的表现与其他算法相比较, 比较的算法类别包括基于点的算法、基于体素的算法和基于距离视图的算法, 结果见表1, 其中 mIoU 为主要性能评价指标, FPS 表示每秒处理的帧数, 后续各列表示各个单独类别的 IoU 指标。

表 1 SemanticKITTI测试集上的定量数值性能对比

Category	Methods	mIoU	FPS(Hz)	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign
Point-based	PointNet ^[1]	14.6	5.2 (2080Ti)	46.3	1.3	0.3	0.1	0.8	0.2	0.2	0.0	61.6	15.8	35.7	1.4	41.4	12.9	31.0	4.6	17.6	2.4	3.7
	PointNet++ ^[2]	20.1	0.1 (2080Ti)	53.7	1.9	0.2	0.9	0.2	0.9	1.0	0.0	72.0	18.7	41.8	5.6	62.3	16.9	46.5	13.8	30.0	6.0	8.9
	KPCConv ^[3]	58.8	1.4 (2080Ti)	96.0	30.2	42.5	33.4	44.3	61.5	61.6	11.8	88.8	61.3	72.7	31.6	90.5	64.2	84.8	69.2	69.1	56.4	47.4
	RandLA ^[4]	53.9	5.4 (2080Ti)	94.2	26.9	25.8	40.1	38.9	49.2	48.2	7.2	90.7	60.3	73.7	20.4	86.9	56.3	81.4	61.3	61.8	49.2	47.7
Voxel-based	MinkowskiNet ^[6]	63.1	7.2 (V100)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	SPVCNN ^[7]	63.8	6.7 (V100)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
	Cylinder3D ^[8]	67.8	5.9 (V100)	97.1	67.6	64.0	59.0	58.6	73.9	67.9	36.0	91.4	65.1	75.5	32.3	91.0	66.5	85.4	71.8	68.5	62.6	65.6
Range-based	SqueezeSegV3 ^[10]	55.9	8.8 (V100)	92.5	38.7	36.5	29.6	33.0	45.6	46.2	20.1	91.7	63.4	74.8	26.4	89.0	59.4	82.0	58.7	65.4	49.6	58.9
	RangeNet++ ^[11]	52.2	15(6000)	91.4	25.7	34.4	25.7	23.0	38.3	38.8	4.8	91.8	65.0	75.2	27.8	87.4	58.6	80.5	55.1	64.6	47.9	55.9
	KPRNet ^[12]	63.1	0.3 (2080Ti)	95.5	54.1	47.9	23.6	42.6	65.9	65.0	16.5	93.2	73.9	80.6	30.2	91.7	68.4	85.7	69.8	71.2	58.7	64.1
	SalsaNext ^[13]	59.5	24(6000)	91.9	48.3	38.6	38.9	31.9	60.2	59.0	19.4	91.7	63.7	75.8	29.1	90.2	64.2	81.8	63.6	66.5	54.3	62.1
	Lite-HDseg ^[14]	63.8	20 (2080Ti)	92.3	40.0	55.4	37.7	39.6	59.2	71.6	54.1	93.0	68.2	78.3	29.3	91.5	65.0	78.2	65.8	65.1	59.5	67.7
	FIDNet ^[15]	59.5	31 (2080Ti)	93.9	54.7	48.9	27.6	23.9	62.3	59.8	23.7	90.6	59.1	75.8	26.7	88.9	60.5	84.5	64.4	69.0	53.3	62.8
	RangeViT ^[16]	64.0	40(2080)	95.4	55.8	43.5	29.8	42.1	63.9	58.2	38.1	93.1	70.2	80.0	32.5	92.0	69.0	85.3	70.6	71.2	60.8	64.7
	CENet ^[17]	64.7	38(3090)	91.9	58.6	50.3	40.6	42.3	68.9	65.9	43.5	90.3	60.9	75.1	31.5	91.0	66.2	84.5	69.7	70.0	61.5	67.6
	RPNet(ours)	64.2	58(3090)	94.2	54.5	52.9	34.9	39.6	68.9	65.9	37.2	93.2	68.5	79.2	29.8	92.2	67.2	85.7	65.8	71.0	57.3	61.8

根据表1的结果,文本提出的RPNet结合了基于距离视图和基于点的方法,使用成熟的二维编码器提取深层特征,使用轻量的逐点解码器进行细化,在分割精度上 mIoU 指标达到 64.2%,超过了大部分现有方法,同时在推理速度上能够达到58帧/秒,比其他方法更快。

2.4.2 定性可视化分析

图5展示了在SemanticKITTI数据集上RPNet与经典的基于距离视图的算法SalsaNext^[13]进行比较的可视化效果。其中,第一列为语义分割标签,第二列和第三列分别为SalsaNext和RPNet的误差图。可以看出,本文提出的方法预测错误的点更少,准确率更高。

2.4.3 消融实验

为了验证本文提出的各个单独模块对算法性能的影响,本文在SemanticKITTI的验证集上进行消融实验。具体而言,基本框架保持不变,初始模型的编码器部分为原始Segformer的编码

器,不使用额外的调优方法。然后在此基础上,逐步加入距离视图补全算法、辅助损失函数 L_{aux} 、在线难样本采样OHEM、非对称上下文模块ACM、可变形嵌入DPE和DropKey,对比各个方法对网络模型分割精度和推理速度的影响,实验结果见表2。

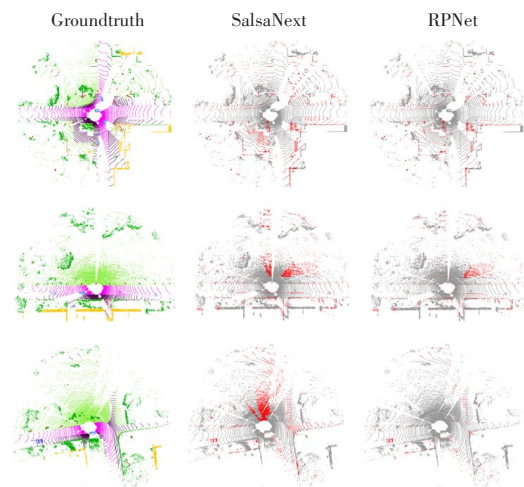


图5 可视化分割精度

表2 SemanticKITTI验证集上的消融实验

序号	补全	L_{aux}	OHEM	ACM	DPE	DropKey	mIoU/%	Latency/ms
1	—	—	—	—	—	—	55.8	13
2	√	—	—	—	—	—	57.4	14
3	√	√	—	—	—	—	57.8	14
4	√	√	√	—	—	—	58.7	14
5	√	√	√	√	—	—	63.1	16
6	√	√	√	√	√	—	64.2	17
7	√	√	√	√	√	√	64.6	17

由表2可知,距离视图补全算法、辅助损失函数、在线难样本采样、非对称上下文模块、可变形嵌入和DropKey在mIoU性能指标上分别带来了1.6%、0.4%、0.9%、4.4%、1.1%、0.4%的增益,证明了本文方法的有效性。其中,距离视图补全算法、非对称上下文模块和可变形嵌入仅仅分别带来了1 ms、2 ms和1 ms的延迟,对推理速度影响微小。

3 结语

本文提出了一种面向大规模点云场景的高效三维点云语义分割算法,名为RPNet。该模型架构结合了基于距离视图的编码器和基于逐点细化的解码器。基于距离视图的编码器以Seg-former作为骨干网络,然后提出了距离视图补全算法、非对称上下文模块和可变形嵌入来改进模型性能,引入DropKey减少模型的过拟合。基于逐点细化的解码器使用MLP融合逐点特征和多尺度深层特征,得到最终的逐点细化分割结果。本文还使用多个辅助头和在线难样本采样来训练网络,在不增加额外推理成本的情况下进一步提升网络的学习能力。在SemanticKITTI数据集上充分的实验结果表明,所提出的方法可以实现高准确度和高推理速度。

参考文献:

- [1] QI C R, SU H, MO K, et al. PointNet: deep learning on point sets for 3D classification and segmentation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017: 652-660.
- [2] QI C R, YI L, SU H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space [C] // Proceedings of the Conference on Advances in

Neural Information Processing Systems, 2017, 30: 5105-5114.

- [3] THOMAS H, QI C R, DESCHAUD J E, et al. Kp-conv: flexible and deformable convolution for point clouds [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, South Korea, 2019: 6411-6420.
- [4] HU Q, YANG B, XIE L, et al. RandLA-Net: efficient semantic segmentation of large-scale point clouds [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 2020: 11108-11117.
- [5] 刘苏毅,迟剑宁,吴成东,等. 基于递归切片网络的三维点云语义分割与实例分割[J]. 中国图象图形学报, 2023, 28(7): 2135-2150.
- [6] CHOY C, GWAK J Y, SAVARESE S. 4D spatio-temporal convnets: minkowski convolutional neural networks [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019: 3075-3084.
- [7] TANG H, LIU Z, ZHAO S, et al. Searching efficient 3d architectures with sparse point-voxel convolution [C] // Proceedings of the European Conference on Computer Vision, Virtual, 2020: 685-702.
- [8] ZHU X, ZHOU H, WANG T, et al. Cylindrical and asymmetrical 3D convolution networks for LiDAR segmentation [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 2021: 9939-9948.
- [9] 夏祥腾,王大方,曹江,等. 基于稀疏卷积神经网络的车载激光雷达点云语义分割方法[J]. 汽车工程, 2022, 44(1): 26-35.
- [10] XU C, WU B, WANG Z, et al. SqueezeSegv3: spatially-adaptive convolution for efficient point-cloud segmentation [C] // Proceedings of the European Conference on Computer Vision, Virtual, 2020: 1-19.
- [11] MILIOTO A, VIZZO I, BEHLEY J, et al. Rangenet++: fast and accurate LiDAR semantic segmentation [C] // Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 2019: 4213-4220.
- [12] KOCHANOV D, NEJADASL F K, BOOIJ O. KPRNet: improving projection-based LiDAR semantic segmentation [EB/OL]. arXiv.2007.12668, 2020.
- [13] CORTINHAL T, TZELEPIS G, ERDAL AKSOY E. Salsanext: fast, uncertainty-aware semantic segmentation of LiDAR point clouds [C] // Proceedings of the International Symposium on Visual Computing

- (ISVC), San Diego, California, USA, 2020:207-222.
- [14] RAZANI R, CHENG R, TAGHAVI E, et al. Lite-HDSeg: LiDAR semantic segmentation using lite harmonic dense convolutions [C] //Proceedings of the 2021 IEEE International Conference on Robotics and Automation(ICRA), Shanxi, China, 2021: 9550-9556.
- [15] ZHAO Y, BAI L, HUANG X. FidNet: LiDAR point cloud semantic segmentation with fully interpolation decoding [C] //Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 2021: 4453-4458.
- [16] ANDO A, GIDARIS S, BURSUC A, et al. Range-ViT: towards Vision Transformers for 3D Semantic Segmentation in Autonomous Driving [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, 2023:5240-5250.
- [17] CHENG H X, HAN X F, XIAO G Q. CeNet: toward concise and efficient LiDAR semantic segmentation for autonomous driving [C] //Proceedings of the 2022 IEEE International Conference on Multimedia and Expo(ICME), Taipei, Taiwan, China, 2022:1-6.
- [18] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: transformers for image recognition at scale [EB/OL]. arXiv. 2010. 11929, 2021.
- [19] XIE E, WANG W, YU Z, et al. SegFormer: simple and efficient design for semantic segmentation with transformers [C] //Proceedings of the Conference on Advances in Neural Information Processing Systems, 2021, 34: 12077-12090.
- [20] LI B, HU Y, NIE X, et al. DropKey for vision transformer [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, 2023:22700-22709.
- [21] ZHU X, HU H, LIN S, et al. Deformable convnets v2: more deformable, better results [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019:9308-9316.
- [22] ZHANG Z, SABUNCU M. Generalized cross entropy loss for training deep neural networks with noisy labels [C] //Proceedings of the Conference on Advances in Neural Information Processing Systems, 2018, 31:8792-8802.
- [23] BERMAN M, TRIKI A R, BLASCHKO M B. The Lovász-Softmax loss: a tractable surrogate for the optimization of the intersection-over-union measure in neural networks [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018:4413-4421.
- [24] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training region-based object detectors with online hard example mining [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016:761-769.
- [25] BEHLEY J, GARBADE M, MILIOTO A, et al. SemanticKITTI: a dataset for semantic scene understanding of LiDAR sequences [C] //Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, South Korea, 2019:9297-9307.

Point cloud semantic segmentation based on range view representation and point-wise refinement

Chen Han, He Hongtian, Lei Yinjie*

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

Abstract: 3D point cloud semantic segmentation is an important way for machines to achieve environment awareness. In existing research, voxel-based methods and point-based methods are computationally inefficient in the face of large-scale point cloud data. And range-based methods inevitably cause accuracy loss when projecting and back projecting point clouds. To solve the above problems, this paper proposed RPNNet, a novel 3D point cloud semantic segmentation framework based on range view representation and point-wise refinement. Sufficient experiments show that the proposed method achieves a mIoU of 64.2% and an inference speed of 58 frames per second on the 3D point cloud outdoor scene dataset SemanticKITTI, balancing high accuracy and speed.

Keywords: semantic segmentation; 3D point cloud; range view; point-wise refinement

文章编号: 1007-1423(2024)04-0023-06

DOI: 10.3969/j.issn.1007-1423.2024.04.004

基于改进点线特征的多源融合 SLAM 算法

林嘉洁¹, 徐 胜¹, 苏成悦^{1, 2*}, 陈元电¹, 刘力斌¹, 施振华¹

(1. 广东工业大学物理与光电工程学院, 广州 510000; 2. 广东工业大学先进制造学院, 揭阳 522000)

摘要: 传统视觉 SLAM 算法在缺乏明显特征的场景或移动平台快速运动时, 特征难以被检测和跟踪。提出了一种基于改进点线特征多源融合 SLAM 算法。前端通过相似短线段合并的策略改进 EDLines 线特征提取算法, 提取长线段特征, 后端融合了点、线特征和 IMU 数据, 并采用非线性优化方法进一步估计相机位姿。实验结果表明, 改进 EDLines 线特征提取算法相比传统 LSD 线特征提取算法, 其特征提取速度提高 4 倍。在 EuRoc 数据集的测试中, 本算法在不同场景下都具有很好的定位和建图效果, 且具有较高的鲁棒性和实时性, 对机器人实时的避障导航应用有较大的参考价值。

关键词: SLAM 算法; EDLines 算法; IMU; 点线特征; 非线性优化

0 引言

随着无人系统和自主导航技术的迅速发展, 多传感器融合的同时定位与地图构建(simultaneous localization and mapping, SLAM)技术在机器人、自动驾驶等领域得到广泛应用。SLAM 算法通过对环境感知和自主运动的联合估计, 实现对未知环境的实时定位和地图构建, 使得机器人能够在没有 GPS 信号时进行自主导航^[1]。视觉 SLAM 因位姿估计策略不同, 衍生出不同类型的算法。传统的基于特征点的 SLAM 算法通过提取图像中的特征点并通过特征点之间的匹配来估计相机的位姿, 这种方法在大规模场景中具有一定的鲁棒性和实时性, 但可能会受到环境纹理缺失、遮挡和环境光照变化等因素的影响, 导致特征跟踪失败^[2]。

Engel 等^[3]提出了基于稀疏直接法的视觉里程计 DSO, 其不含回环检测, 会出现累计误差。He 等^[4]改进了 VINS-Mono 算法, 结合线特征信

息提出 PL-VIO 系统, 计算复杂度高且实时性低。Xu 等^[5]设计一个基于点和直线的 VIO 系统, 提高了复杂环境下姿态估计和映射的准确性和鲁棒性。付煜等^[6]提出基于 LBD 描述子的 EDLine 线特征提取匹配的线特征融合 SLAM 算法, 相比传统的 LSD 算法有更快的线段提取速度。Chen 等^[7]提出了一种基于轻量级多状态约束卡尔曼滤波(MSCKF)的完全正则化平面特征单目视觉系统, 在结构化环境的性能优于基于点特征的 VIO。Hu 等^[8]提出基于点平面地图的大规模 VIO 系统, 可以提高在低纹理和以平面为主的场景的定位精度, 但是需要依靠先验地图。

本文在 VINS-Mono 的基础上提出一种基于改进点线特征多源融合 SLAM 算法。在前端视觉里程计部分加入改进自适应 EDLines 线特征提取算法, 线特征匹配采用基于梯度分布和几何特征的线特征匹配算法, 解决了传统基于 LSD 线特征提取 SLAM 算法存在的特征提取时线段不连续以及实时性差的问题。

收稿日期: 2023-09-18 修稿日期: 2023-10-30

作者简介: 林嘉洁(1995—), 男, 广东惠州人, 硕士研究生, 研究方向为多传感器融合 SLAM; 徐胜(1976—), 男, 广东梅州人, 博士, 讲师, 研究方向为运动控制; *通信作者: 苏成悦(1961—), 男, 湖南长沙人, 博士, 教授, 研究方向为应用物理、光电技术, E-mail: cysu@gdut.edu.cn; 陈元电(1977—), 男, 广东云浮人, 博士, 讲师, 研究方向为无人系统技术、无线通信; 刘力斌(1974—), 男, 湖南株洲人, 工程硕士, 讲师, 研究方向为运动控制; 施振华(1994—), 男, 广东江门人, 学士, 中级工程师, 研究方向为嵌入式、机器人

1 系统架构

算法采用了一种基于改进点线特征的实时单目视觉惯性SLAM算法,融合了点线特征与IMU信息,在低纹理环境下能实时稳定运行。该算法分为前端视觉里程计、后端非线性优化,其整体框架如图1所示。

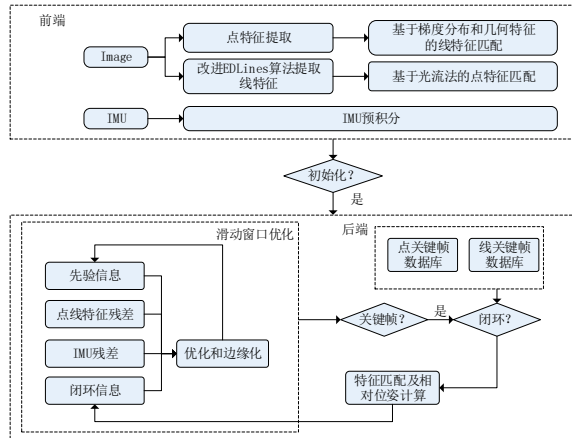


图1 本文算法系统框架

前端对输入图像进行特征提取。线特征提取使用改进自适应EDLines线特征提取算法,对给定的灰度图像进行平滑处理后,逐像素计算梯度幅值和梯度方向得到边缘像素链,依次遍历边缘像素链后采用最小二乘法对像素进行拟合得到线特征^[9]。在线特征匹配上,对待匹配的线段两端点进行对齐,然后计算对齐端点的平均梯度以表示预筛选的线特征梯度。本算法还使用待匹配线特征的角度差、长度比、点距以及垂直距离作为线特征匹配的几何特征因素。在连续的两帧之间将IMU的数据进行预积分处理,得到当前帧的位姿和速度。

数据预处理后,采用固定大小的滑动窗口,通过联合最小化多残差函数寻找包括位姿、速度、空间点线、加速度、陀螺仪偏差在内的最优状态向量,对视觉与IMU进行紧耦合优化。闭环检测阶段,采用DBoW词袋模型检测并构建关键帧数据库,计算图像间点特征相似度和线特征相似度检测闭环。

2 实现方法

2.1 线特征提取与匹配

传统点线特征融合SLAM系统普遍使用LSD

线特征提取算法与LBD线特征匹配算法^[10]。LSD算法可以检测亚像素级别的直线,适用于大多数类型图像,但是它需要消耗大量计算资源,实时性较差^[11]。EDLines算法具有较LSD算法更快的线特征提取速度,消耗的计算资源更少,实时性较高^[12]。本文提出改进自适应EDLines线特征提取算法,流程如图2所示。

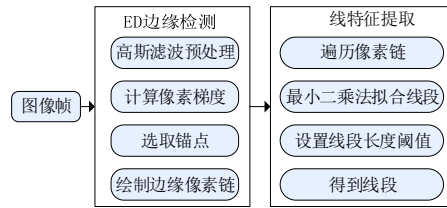


图2 自适应EDLines算法框架

2.1.1 改进EDLines线特征提取算法

EDLines使用ED(edge drawing)边缘检测器检测给定的灰度图像,生成一系列连续的边缘线段像素链,然后采用最小二乘直线拟合法从边缘线段中提取长直线,最后用Helmholtz原理进行直线验证,消除检测错误的线段。具体步骤如下:

(1) ED边缘检测:对给定的灰度图像进行噪声抑制和平滑处理,使用高斯滤波抑制噪声,逐一计算每一像素的梯度大小和方向。梯度最大的像素组成锚点,从任意锚点开始,计算邻近像素的梯度大小和方向,选取相邻梯度最大的锚点作为需要连接的下一个锚点,依次类推,最后连接所有选取锚点形成边缘线段。

(2) 直线段提取:得到相邻连续边缘像素链后,再把边缘像素链分割成一个或多个直线段。按顺序遍历像素,并使用最小二乘直线拟合法对像素进行直线拟合,直到某个像素点到直线段的距离超过某个阈值或是直线的方向发生改变,则生成一个新线段。通过计算NFA决定是否拒绝或接受已被检测的线段。

2.1.2 线特征的空间几何表示

提取线特征后,采用普吕克坐标法(Plücker)表示线特征^[13]。给定一条空间直线 L_w ,其普吕克坐标表示为 $L_w = (n^T, d^T)^T \in R^6$ 。 n 为坐标原点和直线 L_w 确定的平面法向量, d 为直线

L_w 的方向向量。对于直线的几何变换，给定一个从世界坐标系 w 到相机坐标系 c 的变换矩阵 $T_{cw} = \begin{bmatrix} R_{cw} & p_{cw} \\ 0 & 1 \end{bmatrix}$ ，得到直线在相机坐标系下的普吕克坐标表示：

$$L_c = \begin{bmatrix} n_c \\ d_c \end{bmatrix} = T_{cw} L_w = \begin{bmatrix} R_{cw} & [p_{cw}]_{\times} R_{cw} \\ 0 & R_{cw} \end{bmatrix} L_w \quad (1)$$

式中 $[\cdot]_{\times}$ 为斜对称矩阵，变换矩阵 T_{cw} 由旋转矩阵 R_{cw} 和平移向量 p_{cw} 组成。将直线 L_c 投影到图像平面上，其普吕克坐标表示为

$$l_c = K_l n_c = \begin{bmatrix} f_x & 0 & 0 \\ 0 & f_y & 0 \\ -f_x c_x & f_x c_y & f_x f_y \end{bmatrix} n_c = \begin{bmatrix} l_1 \\ l_2 \\ l_3 \end{bmatrix} \quad (2)$$

其中 K_l 为直线的投影矩阵。假设 L_w 表示第 i 帧相机 c_i 观察到第 j 条空间线 L_j ，则线特征重投影误差可以定义为

$$r_l(z_{L_j}^{c_i}, \chi) = d(m, l_c) = \frac{m^T l_c}{\sqrt{l^2 + l^2}} \in R^1 \quad (3)$$

其中 $d(m, l_c)$ 为点到线的距离函数； m 为线特征中点的齐次坐标。

2.1.3 基于梯度分布和几何特征的线特征匹配

本文采用基于梯度分布和几何特征的线特征匹配算法，对线段附近像素点的相似度进行评估，获取更准确、更快的线段匹配结果^[14]。线段匹配优先考虑长度超过35像素的线段，按照线段长度顺序选取 N 对线段。在被选取的线段中选择距离最近的两个端点进行对齐，然后计算直线水平方向和垂直方向的梯度。如果直线的方向不一致，则直线被拒绝。之后通过计算线段之间的角度差、长度比、端点距离或垂直距离，作为相似线段的判断指标。将角度差小于 $\pi/10$ 、长度比大于0.8、端点距离小于35像素或垂直距离小于25像素的线段视为相似线段。算法流程如图3所示。

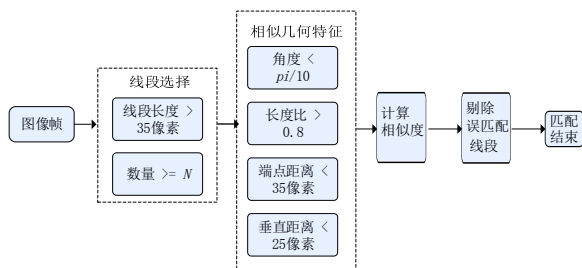


图3 线特征匹配算法框架

2.2 视觉点线信息与IMU滑动窗口优化

为了提高系统的实时性和鲁棒性，后端采用多源信息融合的滑动窗口优化法。首先定义点、线和IMU测量信息相关的滑动窗口全状态向量。

$$X = \begin{bmatrix} x_n, x_{n+1}, \dots, x_{n+N}, \\ \lambda_m, \lambda_{m+1}, \dots, \lambda_{m+M}, \\ O_o, O_{o+1}, \dots, O_{o+O} \end{bmatrix}^T \quad (4)$$

$$x_i = [p_{wb_i}, q_{wb_i}, v_i^w, b_a^{b_i}, b_g^{b_i}]^T, i \in [n, n+N] \quad (5)$$

其中 x_n 表示IMU在第 i 帧的状态，包括在世界坐标系下的IMU位置 p_{wb_i} 、姿态 q_{wb_i} 、速度 v_i^w 以及在body坐标系下的IMU加速度计偏置 $b_a^{b_i}$ 、陀螺仪偏置 $b_g^{b_i}$ 。 b 表示body坐标系，即IMU坐标系， w 表示世界坐标系。 λ_m 表示第 m 个空间特征点的逆深度， O_o 表示第 o 条空间线特征的正交表示。 N 是滑动窗口的关键帧数量。

通过最小化目标函数优化所有测量残差项总和来优化滑动窗口中的所有状态变量^[15]。目标函数包括点、线特征重投影残差、IMU测量残差以及边缘化的先验信息。四种残差都用马氏距离表示。

$$\min_x \rho \left(\left\| r_p - J_p \chi \right\|_{\sum_p}^2 \right) + \sum_{i \in B} \rho \left(\left\| r_b(z_{b_i, b_{i+1}}, \chi) \right\|_{\sum_{b_i, b_{i+1}}}^2 \right) + \sum_{(i,j) \in F} \rho \left(\left\| r_f(z_{f_j}^{c_i}, \chi) \right\|_{\sum_{f_j}}^2 \right) + \sum_{(i,l) \in L} \rho \left(\left\| r_l(z_{L_l}^{c_i}, \chi) \right\|_{\sum_{L_l}}^2 \right) \quad (6)$$

其中 $r_p - J_p \chi$ 为先验信息，在滑动窗口中边缘化一帧后计算得到， $r_b(z_{b_i, b_{i+1}}, \chi)$ 为 x_i 与 x_{i+1} 帧间的IMU预积分测量残差^[13]， B 是滑动窗口中所有预积分IMU测量值的集合。 $r_f(z_{f_j}^{c_i}, \chi)$ 为点特征的重投影残差。本文在文献^[16]基础上，提出了 $r_l(z_{L_l}^{c_i}, \chi)$ 作为线特征重投影残差的最后一项， F 和 L 分别为滑动窗口中相机观察到的点特征与线特征。 $\sum_p$ 、 $\sum_{b_i, b_{i+1}}$ 、 $\sum_{f_j}$ 和 $\sum_{L_l}$ 分别表示先验残差、IMU预积分测量残差、点特征重投影残差和线特征重投影残差的协方差矩阵， ρ 是柯西鲁棒函数。

3 实验结果及分析

实验采用苏黎世联邦理工学院 EuRoC 公开数据集^[17]评估本文算法和其他三个经典SLAM

算法的性能。实验平台 CPU 为 AMD Ryzen 7 4800H, 内存大小为 16 GB, 操作系统为 Ubuntu 18.04, ROS 版本为 Melodic。

3.1 线特征提取

在 EuRoC 数据集中选取两个关键帧, 分别使用 LSD 算法与本文改进 EDLines 算法进行线特征提取。提取结果如图 4 所示。

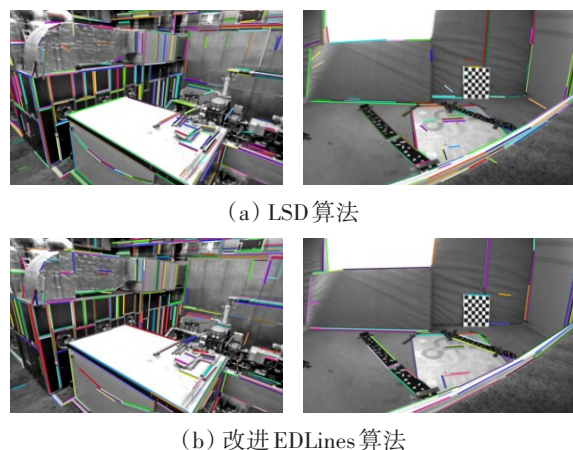


图 4 LSD 算法和改进 EDLines 算法线特征提取效果

从图 4 可得, 当场景纹理较多时, LSD 算法会提取较多短线特征, 改进 EDLines 算法会剔除短线段, 保留长线特征。当场景纹理较少时, LSD 算法提取的短线特征明显多于长线特征, 改进 EDLines 算法可以保留较多长线特征。对 LSD 算法和改进 EDLines 算法线特征提取数量及耗时进行比较分析, 其中提取耗时是对每张图片进行 500 次提取后的平均耗时, 结果见表 1。

表 1 LSD 算法和改进 EDLines 算法性能对比

图像	LSD 算法		改进 EDLines 算法	
	时间/ms	线数量/条	时间/ms	线数量/条
图 4(a)	17.6531	279	4.0630	264
图 4(b)	11.4372	97	3.7387	54

由表 1 可知, LSD 算法提取平均耗时较长, 本文算法在保留主要长特征的情况下对线特征的提取速度是 LSD 算法的 3 到 4 倍。

3.2 轨迹对比与误差分析

绝对轨迹误差 (absolute trajectory error, ATE) 是真实位姿与估计位姿的直接差值, 可以直观

反映 SLAM 算法精度和轨迹的全局一致性。本文使用 ATE 的均方根误差 (root mean square errors, RMSE) 评估算法性能。RMSE 值越小, 说明算法整体定位精度越高。在 EuRoC 数据集的 11 个序列中分别对 VINS-Mono、PL-VINS、PL-VIO 与本文算法进行测试, 结果见表 2, 粗体表示最优结果, — 表示跟踪失败。

表 2 多种算法的绝对轨迹误差 RMSE 值对比

Dataset	本文算法	VINS-Mono	PL-VINS	PL-VIO
MH_01	0.170	0.172	0.162	0.138
MH_02	0.185	0.187	0.176	0.148
MH_03	0.233	0.239	0.224	0.267
MH_04	0.273	0.383	0.299	0.364
MH_05	0.296	0.343	0.307	0.301
V1_01	0.156	0.179	0.159	0.165
V1_02	0.141	0.125	0.141	—
V1_03	0.214	0.202	0.158	0.218
V2_01	0.122	0.094	0.102	0.121
V2_02	0.142	0.246	0.156	0.177
V2_03	0.167	0.296	0.289	0.216

表 2 和图 5 表明, 在 11 种测试数据中, 本文算法在 5 种数据集上的 RMSE 值低于其他算法, 与 VINS-Mono 相比, 在 8 种数据集上的 RMSE 值较低; 与 PL-VINS 相比, 在 5 种数据集上的 RMSE 值较低; 与 PL-VIO 相比, 在 8 种数据集上的 RMSE 值较低。

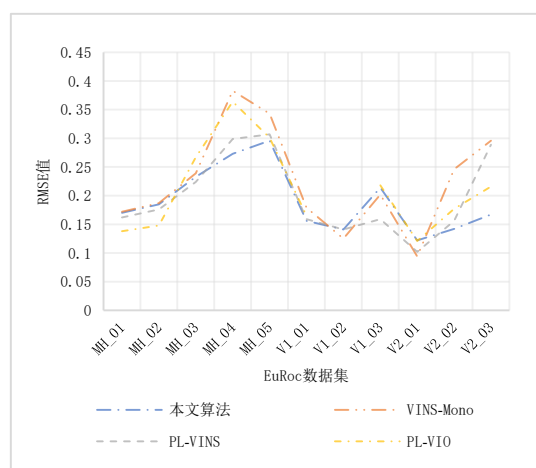
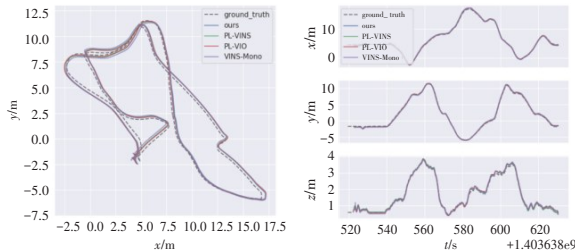


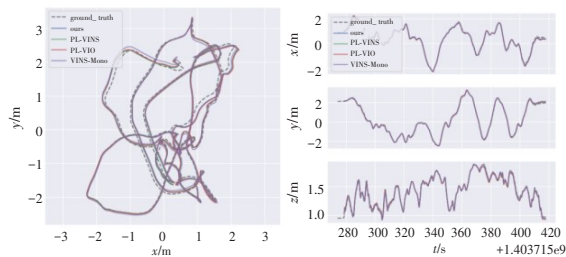
图 5 多种算法的绝对轨迹误差 RMSE 值对比折线图

对比分析 RMSE 值可以看出, 本文算法在不同数据集上取得较好效果, 对比 PL-VINS 算法, 优势不明显。但是 PL-VINS 算法的线特征提

取方法采用LSD，线特征匹配方法采用LBD，因此在线特征的提取和匹配上消耗的算力较高，耗时较长，不适合应用于实时嵌入式平台提供实时避障导航。图6为估计轨迹与真实轨迹对比。



(a) MH_05



(b) V1_01

4 结语

本文提出一种基于改进EDLines算法的点线特征多源融合视觉惯性SLAM系统，与传统的LSD算法相比，线特征的提取速度提高4倍。线特征匹配采用基于梯度分布和几何特征相似度的线特征匹配算法，提高了线段匹配准确度并且降低了匹配时间。后端采用滑动窗口优化方法，有效融合了点、线特征和IMU数据，提高了系统整体性能。在公开数据集EuRoC实验结果表明，提出的算法与传统的点线特征视觉惯性融合SLAM算法相比，具有更低的线特征误匹配率，更低的算力消耗，更高的定位准确度和鲁棒性。本研究提出的方法对机器人实时的避障导航应用有较大的参考价值。

参考文献：

[1] 王金科,左星星,赵祥瑞,等. 多源融合SLAM的现状与挑战[J]. 中国图象图形学报,2022,27(2): 368-389.

[2] 赵伟博,田军委,王沁,等. 基于点线特征融合的视觉惯性SLAM算法[J]. 计算机应用研究,2023,40

(2):445-449.

- [3] ENGEL J, KOLTUN V, CREMERS D. Direct sparse odometry [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40 (3) : 611-625.
- [4] HE Y, ZHAO J, GUO Y, et al. PL-VIO: tightly-coupled monocular visual-inertial odometry using point and line features [J]. Sensors, 2018, 18 (4):1159.
- [5] XU B, WANG P, HE Y, et al. Leveraging structural information to improve point line visual-inertial odometry [J]. IEEE Robotics and Automation Letters, 2022, 7(2):3483-3490.
- [6] 付煜,郑爽,别桐,等. 融合点线特征的视觉惯性SLAM算法[J]. 计算机应用研究, 2022, 39(2) : 349-355.
- [7] CHEN C, GENEVA P, PENG Y, et al. Monocular visual-inertial odometry with planar regularities [C] //Proceedings of the 2023 IEEE International Conference on Robotics and Automation(ICRA), London, United Kingdom, 2023:6224-6231.
- [8] HU J, REN K, XU X, et al. Efficient visual-inertial navigation with point-plane map [C]//Proceedings of the 2023 IEEE International Conference on Robotics and Automation(ICRA), London, United Kingdom, 2023:10659-10665.
- [9] AKINLAR C, TOPAL C. EDLines: a real-time line segment detector with a false detection control [J]. Pattern Recognition Letters, 2011, 32 (13) : 1633-1642.
- [10] ZHANG L, KOCH R. An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency [J]. Journal of Visual Communication and Image Representation, 2013, 24(7):794-805.
- [11] VON CIOI R G, JAKUBOWICZ J, MOREL J M, et al. LSD: a line segment detector [J]. Image Processing On Line, 2012, 2:35-55.
- [12] 张宁,王竞雪. EDLines和LSD直线提取算法性能探究[J]. 测绘科学, 2020, 45(12):116-125.
- [13] FU Q, WANG J, YU H, et al. PL-VINS: real-time monocular visual-inertial SLAM with point and line [EB/OL]. arXiv:2009.07462, 2020.
- [14] ZHAO S, ZHANG T, WEI H. Fast VIO with point-line features using adaptive EDLines algorithm [J]. Measurement Science and Technology, 2022, 33 (10):105401.

- [15] KAESSE M, RANGANATHAN A, DELLAERT F. iSAM: incremental smoothing and mapping[J]. IEEE Transactions on Robotics, 2008, 24(6): 1365-1378.
- [16] QIN T, LI P, SHEN S. VINS-Mono: a robust and versatile monocular visual-inertial state estimator [J]. IEEE Trans on Robotic, 2018, 34(4): 1004-1020.
- [17] BURRI M, NIKOLIC J, GOHL P, et al. The EuRoC micro aerial vehicle datasets [J]. The International Journal of Robotics Research, 2016, 35(10): 1157-1163.

Multi-source fusion SLAM algorithm based on improved point-Line features

Lin Jiajie¹, Xu Sheng¹, Su Chengyue^{1,2*}, Chen Yuandian¹, Liu Libin¹, Shi Zhenhua¹

(1.School of Physics and Optoelectronic Engineering, Guangdong University of Technology, Guangzhou 510000, China;

2.School of Advanced Manufacturing, Guangdong University of Technology, Jieyang 522000, China)

Abstract: Traditional visual SLAM algorithms are difficult to detect and track features in scenes lacking obvious features or when mobile platforms move rapidly. This article proposes a multi-source fusion SLAM algorithm based on improved point line features. The front-end improves the EDLines line feature extraction algorithm by combining similar short line segments, extracting long line segment features. The back-end integrates point, line features, and IMU data, and uses nonlinear optimization methods to further estimate the camera pose. The experimental results show that the improved EDLines line feature extraction algorithm has a feature extraction speed that is four times faster than the traditional LSD line feature extraction algorithm. In the testing of the EuRoc dataset, this algorithm has good positioning and image building effects in different scenarios, and has high robustness and real-time performance, which has great reference value for real-time obstacle avoidance navigation applications of robots.

Keywords: SLAM algorithm; EDLines algorithm; IMU; point line features; nonlinear optimization

(上接第 15 页)

Canine nose print recognition based on FaceNet

Li Menghan, Zhang Chenlei, Tian Cunwei*

(School of Physical Science and Information Engineering, Liaocheng University, Liaocheng 252000, China)

Abstract: With the increasing number of pet-owning households, canine identity recognition has become an important issue. Canine nose prints are unique, and can be used to accurately identify canine identities. In order to achieve canine nose print recognition, designed a method based on the FaceNet network. The Backbone of FaceNet utilizes ResNet18, and the loss function employs Triplet loss. Additionally, SRGAN network preprocessing is applied to canine nose print images to enhance image resolution, making the nose print images clearer. Based on the FaceNet network offers an effective and feasible solution for canine identity recognition and authentication.

Keywords: FaceNet; canine nose print recognition; ResNet18

文章编号: 1007-1423(2024)04-0029-05

DOI: 10.3969/j.issn.1007-1423.2024.04.005

多尺度注意力机制下的人脸表情识别算法设计

蒋文豪*

(重庆航天职业技术学院智能信息工程学院, 重庆 400021)

摘要: 为了优化在人脸表情较模糊情况下的识别效果, 并更好地获取表情的表征数据, 设计一种多尺度注意力机制下的人脸表情识别方法。对人脸表情图像进行缩放与扩充预处理操作, 从图像中提取人脸表情解耦表征皮沟数据, 通过卷积神经网络对提取到的解耦表征皮沟数据进行特征捕捉。引入了多尺度注意力机制, 有选择性地关注重要的表情特征。同时, 利用多通道的表情识别方法, 自适应地提取人脸组件区域内的表情信息, 从而识别人脸表情。实验分析结果表明, 所提方法在四类不同表情标签对应的人脸表情识别召回率始终高于对照组, 均达到了98%以上, 识别效果优势显著。

关键词: 多尺度注意力; 人脸特征; 表情识别; 皮沟特征; 神经网络

0 引言

人脸表情从广义角度上来说属于人类表情情绪、传递情感的重要途径之一^[1], 在人类交流过程中, 语言属于交流基础的载体, 负责承接各项数据信息, 面部表情能够强化及修饰语言基础信息, 通过面部表情丰富语义特征与形象特征, 侧面反映说话者的生理状况及心理体验^[2]。相关研究表明, 在交流中, 只有大约7%是通过语言进行的, 而大约55%是通过面部表情进行的, 由此可知人脸表情是最直接的非语言化情感表达方式^[3]。但由于目前模糊的人脸图像中存在细节不明显或信息模糊的问题, 导致在人工智能领域, 提取的表情特征质量下降, 对表情的智能识别方法存在一定的弊端。为此, 需要研究一种更为准确的脸部表情识别方法。

现阶段的人脸表情识别方法较多。例如文献[4]提出一种多注意力机制下自愈人脸表情识别方法, 该方法注意到了特征丢失的问题, 但是无法摆脱冗余特征干扰。文献[5]提出基于多方向特征融合的动态人脸微表情识别方法, 该

方法通过方向特征约束冗余特征的干扰, 取得一定进展, 但是在特征过多, 或者特征模糊的情况下, 算法鲁棒性下降。文献[6]提出人脸部分遮挡条件下的表情识别方法, 利用少量特征完成表情识别, 但是, 对模糊人脸表情的约束条件过多, 限制了其应用性。为解决以上问题, 本文开展了多尺度注意力融合下的人脸表情识别方法的研究。

1 人脸表情识别方法研究

1.1 人脸表情图像的预处理

在进行人脸表情识别时, 首先, 需要对人脸表情图像进行预处理, 提高表情识别的准确性。人脸表情图像预处理主要包括两个组成部分, 分别为图像缩放与图像扩充。因为针对人脸而言, 缩放人脸图像通常是为了统一样本的尺寸大小。由于人脸图像来自于不同的数据源或者采集设备, 其尺寸可能存在差异, 如图像的分辨率、图像所占的像素大小等。为了保持训练样本的一致性和可比性, 需要对人脸图像

收稿日期: 2023-09-22 修稿日期: 2023-11-10

基金项目: 重庆市教育委员会科学技术研究项目(KJZD-K202003001)

作者简介: *通信作者: 蒋文豪(1982—), 女, 四川大竹人, 硕士, 副教授, 研究方向为大数据和大数据安全, E-mail: a18618251406@126.com

进行缩放操作,使得它们具有相同的尺寸或相似的尺度^[7]。本文采用双线性差值方法,对图像进行缩放处理。双线性差值原理示意图,如图1所示。

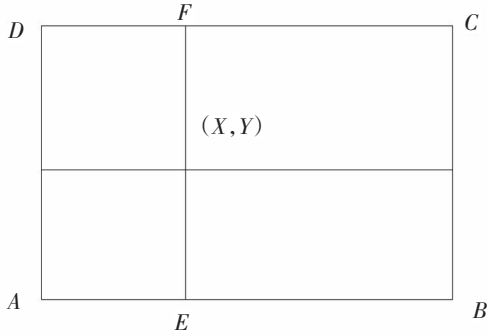


图1 双线性差值原理示意图

图1中, A 、 B 、 C 、 D 分别表示人脸图像中四个邻近的像素点, (i, j) 表示 A 点的坐标, 利用 (X, Y) 处邻近的四个灰度值, 计算 (X, Y) 处的灰度值, 公式如下所示:

$$g(E) = (X - j)[g(B) - g(A)] + g(A) \quad (1)$$

$$g(F) = (X - j)[g(D) - g(C)] + g(C) \quad (2)$$

$$g(X, Y) = (Y - j)[g(F) - g(E)] + g(E) \quad (3)$$

其中, $g(A)$ 、 $g(B)$ 、 $g(C)$ 、 $g(D)$ 、 $g(E)$ 、 $g(F)$ 分别表示 A 、 B 、 C 、 D 、 E 、 F 像素点对应的灰度值。通过计算, 得出 (X, Y) 处的灰度值 $g(X, Y)$, 即双线性差值, 根据双线性差值, 对输入的人脸图像进行缩放处理。经过缩放处理后的人脸图像, 其细节信息更丰富, 提取出的局部特征更具有代表性^[8]。

其次, 对图像进行扩充处理, 主要目的在于扩大数据集的规模和多样性。在人脸检测相关任务中, 如人脸识别、人脸表情识别等, 数据量的大小和质量对于模型的训练和性能具有重要影响。如果训练集的数据量有限, 或者数据集缺少某些类别的样本, 模型可能会面临性能下降或泛化能力受限的问题。通过扩充人脸图像, 可以生成更多的训练样本, 增加数据集的规模, 强化后续多尺度注意力特征融合的泛化性能^[9]。

1.2 引入多尺度注意力机制的人脸表情特征融合

在进行人脸表情识别时, 图像预处理主要

包括图像缩放和图像扩充两个步骤, 基于上述人脸表情图像预处理结果, 可以提高图像的质量, 保持图像细节, 增强图像的泛化性能。接下来, 提取人脸表情皮沟特征解耦表征数据, 人脸图像中的表情信息包含了丰富的情绪和认知状态的线索。通过提取人脸表情解耦表征数据, 可以将表情信息从面部尺度中减去, 使得后续任务关注于除表情外的其他特征, 减少特征之间的冗余和相互影响。这有助于提高对不同光照、姿态、噪声等条件下的人脸表情的识别和理解能力, 增强模型的鲁棒性和泛化性^[10]。为了提高人脸表情图像的特征提取效果, 提出了一种基于多尺度注意力机制的人脸表情特征融合方法。使用卷积神经网络对提取到的人脸表情解耦表征数据进行处理, 以有效地提取出人脸表情图像的全局特征。在建立的卷积神经网络结构中, 设置合适的卷积核尺寸大小, 以避免过多地提取局部特征而忽视全局特征, 从而造成特征信息冗余。

在此基础上, 引入了多尺度注意力机制来进一步提升特征融合的效果。在卷积神经网络中, 使用多个大小不同的卷积核来提取图像的多尺度特征, 并将它们进行分类。根据分类结果得到最终的人脸表情特征注意力参数:

$$F_m = g(X, Y) \left(\sum_{j=1}^4 f_j^2(x, y) \right)^{\frac{1}{2}} \quad (4)$$

其中, f_j 表示第 j 类人脸表情Gabor特征。

利用上述公式, 进行同一尺度的人脸表情特征融合, 在强注意力区域, 可进行深度融合, 弱区域舍弃融合。在此基础上, 根据表情特征融合结果, 得到特征图融合示意图, 如图2所示。

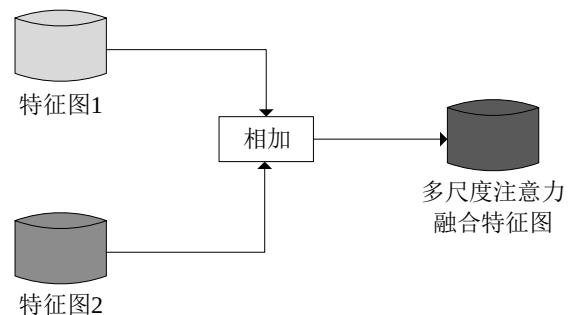


图2 人脸表情特征图融合示意图

如图2所示,在注意力参数下对两个特征图的所有元素逐一相加,生成多尺度融合特征图。无论上述多尺度图像还是多尺度特征,均需要对人脸表情特征图进行融合,提取多尺度融合特征,为提升后续表情识别效果奠定良好的基础。

1.3 多通道人脸表情识别算法设计

在1.2小节中通过多尺度注意力机制,将不同尺度提取的表情特征进行融合处理,通过融合不同尺度下的特征,可以综合利用各个特征的优势,提高整体的表情识别性能,获得更全面和详细的人脸表情信息。在此基础上,开展鼻子、嘴巴、眼睛、眉毛等多通道的人脸表情识别。

在人脸表情中,眉毛、眼睛、鼻子、嘴巴等部位,均蕴含了丰富的表情信息,通过观察这几个部位的动态变化,初步判断人脸表情类别。利用人脸组件,对人脸不同部位作出识别,获取包括中性表情在内的四种表情。在此基础上,利用多通道的表情识别方法,在神经网络中自适应提取人脸组件区域内的表情信息。首先,划分人脸表情区域,获取不同的人脸组件,包括鼻子、嘴巴、眼睛、眉毛等,控制每个人脸组件作为一个相对独立的通道,获取对应的表情信息^[11]。设定完整的人脸作为一个子通道,通过子通道提取全脸表情信息。其次,利用MTCNN对划分后的人脸表情区域进行检测。输入人脸图像大小,并根据人脸图像大小的实际情况,设置匹配度较高的边界框,以鼻子对应的关键点位置作为边界框的中心,截取边界框大小的区域,获取眼睛与鼻子图像,得到人脸组件关键点位置。以嘴角对应的关键点的连线中点作为中心,使每个关键点分别对应人脸各个组件,提取不同的人脸组件图像。基于尺度归一化操作原理,对提取到的图像处理,通过对人脸表情图像进行分割,得到不同表情组件,全面识别不同通道的人脸表情特征。多通道人脸表情识别表达式如下所示:

$$c = \arg \max_j \{b_j(x) | j = 1, 2, \dots, M\} F_m \quad (5)$$

其中, c 表示多通道人脸表情识别结果; $b_j(x)$ 表示待识别的人脸表情样本 x 中判别为第 j 类表情

目标类别的模糊值; M 表示表情目标类别模糊值数量。通过上述表达式,获取人脸表情识别结果,实现表情识别目标。

2 实验分析

在提出的识别方法投入实际应用前,需要对该方法的可行性及识别效果作出客观测试,确认人脸表情识别效果能够达到预期设计要求后,方可投入实际工作中使用。因此,开展了如下文所示的实验测试分析。

2.1 实验准备

按照上述提出的人脸表情识别方法的运行需求,搭建此次实验所需的测试环境。实验测试环境配置见表1。

表1 实验测试环境配置参数设置

编号	名称	配置参数
1	操作系统	Windows 10
2	开发环境	PyCharm
3	CPU	Intel(Core)i5-8400
4	GPU	NVIDIA GTX1050Ti
5	深度学习框架	PyTorch 1.8
6	开发语言	Python3.6

按照表1所示的硬件与软件配置,搭建实验测试环境。在此基础上,在自然环境下采集人脸表情图像,生成MMI人脸表情数据集,保证图像的规范性。该数据集由不同性别、年龄层次的人在自然环境条件下拍摄而成,总共408个图像序列。

此次实验测试中,随机选取其中四类人脸表情标签,如图3所示。



图3 实验选用人脸表情标签

选用图3所示的四类表情标签生成人脸表情数据集,数据集人脸表情分布见表2。

表2 数据集人脸表情分布情况

编号	表情	数量
01	生气	103
02	恐惧	85
03	高兴	126
04	伤心	94

如表2所示,MMI人脸表情数据集内总共包括四类表情标签,人物表情前后不一致,且不同人物的人脸表情状态存在一定的差异。此次实验所使用的人脸表情图像为正脸序列中的三帧图像,保证了实验的客观性。在此基础上,应用上述提出的人脸表情识别方法,开展实验测试分析。

2.2 结果分析

在本次实验中,将本文提出的多尺度注意力融合下的人脸表情识别方法设置为实验组,将文献[2]提出的基于局部脸形注意力的识别方法设置为对照组1、与文献[3]提出的基于NIFEIEN的识别方法设置为对照组2,分别获取三种方法的人脸表情识别结果,作出客观对比,判断本文提出的识别方法是否具有可行性。选取人脸表情识别召回率作为此次实验分析的性能评价指标,其计算表达式如下所示:

$$A = \frac{m}{\sum_{i=1}^N n_i} \quad (6)$$

其中, A 表示人脸表情识别召回率; m 表示数据集中第 i 类分类识别正确的人脸表情图像样本数; n_i 表示数据集中第 i 类的人脸表情图像样本总数。通过计算,得出实验的性能评价指标。 A 越高,表示提出的识别方法越有效,能够准确地识别出人脸表情,识别精度越高,反之同理。利用MATLAB模拟分析软件,模拟上述三种方法的人脸表情识别全过程,并利用SPSS统计分析软件,统计四类表情标签对应的人脸表情识别召回率,作出客观对比,结果如图4所示。

通过图4的性能评价指标对比结果可知,三种方法应用后,人脸表情识别结果存在较大差异。其中,本文提出的多尺度注意力融合下的

人脸表情识别方法应用后,四类不同表情标签对应的人脸表情识别召回率始终高于另外两种识别方法,均达到了98%以上,而对照组1提出的识别方法应用后,其人脸表情识别召回率最高不超过92%,对照组2提出的识别方法应用后,其人脸表情识别召回率最高不超过90%。由此对比结果不难看出,本文提出的人脸表情识别方法具有较高的可行性,能够准确地识别出人脸表情,识别精度较高,方法更加有效。

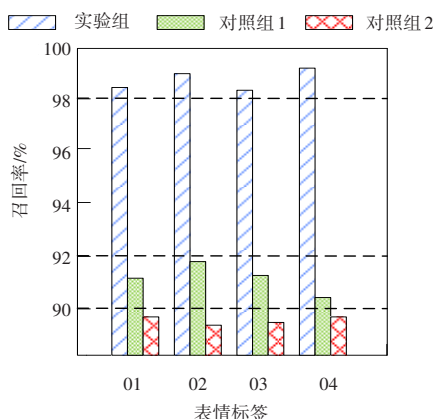


图4 人脸表情识别召回率对比结果

3 结语

综上所述,为了优化人脸表情识别方法的效果,提高识别结果的准确性,引入多尺度注意力融合机制,开展了基于多尺度注意力融合的人脸表情识别方法研究。通过图3的性能评价指标对比结果可知,提出的方法应用后,四类不同表情标签对应的人脸表情识别召回率较高,均达到了98%以上,能够准确地识别出人脸表情,识别精度较高,验证了该识别方法的有效性与鲁棒性。本文提出的研究,对推动人机交互智能化发展起到了至关重要的作用,在未来具有良好的应用前景。

参考文献:

- [1] 叶耀光,陈宗楠,陈丽群,等.基于通道注意的可变形金字塔表情识别网络[J].计算机技术与发展,2022,32(11):64-71.
- [2] 杜洋涛,杨鼎康,翟鹏,等.基于局部流形注意力的人脸表情识别[J].计算机系统应用,2022,31(10):15-24.

- [3] 张伟,伊力哈木·亚尔买买提. 基于NIFEIEN和GLBP特征的人脸表情识别[J]. 计算机仿真, 2022, 39(10):239-244,466.
- [4] 汪文翔,查铖,闵卫东,等. 多注意力机制下自愈人脸表情识别[J]. 微电子学与计算机, 2022, 39(9): 55-62.
- [5] 徐胜超,叶力洪. 基于多方向特征融合的动态人脸微表情识别方法[J]. 计算机与数字工程, 2022, 50(8):1818-1822.
- [6] 蒋斌,李南星,钟瑞,等. 人脸部分遮挡条件下表情识别研究的新进展[J]. 计算机工程与应用, 2022, 58(12):12-24.
- [7] 蓝峥杰,王烈,黄莹. 一种改进型Dense-HRNet和基尼指数动态加权决策的表情识别算法[J]. 电讯技术, 2022, 62(11):1683-1690.
- [8] 胡贵恒. 基于人工智能技术的人脸图像局部细节信息识别研究[J]. 沈阳工程学院学报(自然科学版), 2021, 17(3):66-70.
- [9] 谢惠华,黎明,王艳,等. 基于DE-Gabor特征的人脸表情识别[J]. 南昌航空大学学报(自然科学版), 2021, 35(2):82-91,124.
- [10] 钟马驰,张俊朗,蓝扬波,等. 基于人脸检测和模糊综合评判的在线教育专注度研究[J]. 计算机科学, 2020, 47(S2):196-203.
- [11] 孙冬梅,张飞飞,毛启容. 标签引导的生成对抗网络人脸表情识别域适应方法[J]. 计算机工程, 2020, 46(5):267-273,281.

Design of face expression recognition algorithm under the multi-scale attention mechanism

Jiang Wenhao^{*}

(School of Intelligent Information Engineering, Chongqing Aerospace Polytechnic, Chongqing 400021, China)

Abstract: In order to optimize the recognition effect of blurred face expression and better obtain the representation data of facial expression, a face expression recognition method under the multi-scale attention mechanism is designed. By zoom and expansion, the face expression decoupling representation feature data is extracted from the image, and the extracted decoupling representation feature data is captured through the convolutional neural network. Introduce multiscale attention mechanisms to selectively focus on important expression features. at the same time. Using the multi-channel expression recognition method, the expression information in the face component area is adaptively extracted to identify facial expressions. The experimental analysis results show that the recall rate of face expression recognition corresponding to the four different types of expression labels is always higher than that of the control group, which is more than 98%, and the recognition effect has significant advantages.

Keywords: multi-scale attention; face features; expression recognition; skin features; neural network

文章编号: 1007-1423(2024)04-0034-05

DOI: 10.3969/j.issn.1007-1423.2024.04.006

基于 LightGBM 和蚁群算法优化的特征选择方法

别春洋, 陶贻勇*

(安徽理工大学计算机科学与工程学院, 淮南 232001)

摘要: 在处理高维特征数据时通常会面临冗余和不相关的问题, Relief 作为一种传统的特征选择算法因其具有较高的稳定性和计算效率, 从而被广泛应用。但其特征选择结果具有随机性, 且对于特征之间存在较强依赖关系的数据集, 如共线性等, 可能会导致结果不准确。基于对特征选择方法的研究, 给出了基于 LightGBM 和蚁群算法的 L-ACO 方法, 使用 LightGBM 算法的特征重要性来表示 L-ACO 算法蚁群路径搜索过程的启发式信息。同时, 使用特征之间的皮尔森相关系数来调整信息素浓度, 以便更好地控制特征的相关性。实验证明, L-ACO 方法可以在保证分类准确率的前提下, 减少特征数量, 降低特征冗余, 并提高算法性能。

关键词: 特征选择; LightGBM; 蚁群算法; 皮尔森系数

0 引言

在机器学习处理实际问题中, 往往存在着大量冗余、无关和不必要的特征^[1], 这些特征会占用大量的存储空间、计算资源和时间成本, 同时还会导致模型过拟合和泛化能力下降。因此, 特征筛选^[2-3]已经成为了机器学习领域的重点研究问题之一。通过合理的特征选择, 可以减少数据的特征维度, 提高模型的泛化能力和可解释性, 从而更好地解决实际问题。

本文提出了一种特征选择方法(L-ACO), 将 LightGBM 和蚁群算法结合在一起, 根据 LightGBM 的内置特征重要性评分进行特征排序^[4], 同时利用特征间的 Pearson 相关系数表示特征间的距离, 根据特征相关性系数调节信息素浓度, 使用各特征重要性作为本文蚁群算法部分的启发式函数。实验表明, 这种方法可以确保选取的特征之间不会产生强烈的线性关联, 减少特征冗余, 提高数据分类的精确性, 同时还可以减少选择的特征数量。

1 研究背景

在数据预处理过程中, 特征选择非常重要, 通过特征选择可以去除数据中无关冗余的特征, 从而提高模型的泛化性能和解释性。特征选择方法的研究一直受到学术界和工业界的广泛关注。近年来, 随着机器学习算法和数据科学应用的广泛应用, 特征选择的重要性也越来越突出。然而, 由于现实数据集通常包含大量特征, 因此进行特征选择时往往会面临一些挑战。例如, 可能存在高度相关的特征, 这会导致模型过度拟合和性能下降。另外, 某些特征可能会包含噪声或无关信息, 这也会影响模型的精度和可解释性。

为了应对这些挑战, 学术界和工业界已经提出了许多特征选择方法和算法^[5-9], 这些方法可以根据不同的目标 and 需求, 选择最佳的特征子集, 以提高模型的性能和效率。一些常用的特征选择方法包括基于嵌入式的方法以及基于过滤器的方法。Relief 算法就是一种典型的特征选择算法, 具有较高的稳定性和计算效率, 但

收稿日期: 2023-10-16 修稿日期: 2023-11-01

作者简介: 别春洋(1997—), 男, 江苏南通人, 在读硕士研究生, 主要研究方向为金融科技与大数据、自然语言处理;

*通信作者: 陶贻勇(1997—), 男, 安徽阜阳人, 在读硕士研究生, 主要研究方向为自然语言处理, E-mail: yiyongtao1022@163.com

Relief 算法的特征选择结果具有随机性, 不同的初始采样会有不同的结果, 且对于特征之间存在较强依赖关系的数据集, 如共线性等, 可能会导致结果不准确。

为了解决上述问题, 本文提出了一种将 LightGBM 与蚁群算法相结合的特征选择方法, 并使用皮尔森(Pearson)系数表示特征之间的相关性, 经实验对比验证, 本文基于 LightGBM 和蚁群算法优化的特征选择方法具有更好的泛化能力。

2 相关理论

2.1 LightGBM 算法

梯度提升决策树(light gradient boosting machine), 即是一种高效的梯度提升决策树算法。它在原有的 GBDT 基础上进行了改进, 并且能够在不损害准确率的前提下提升模型的训练速度, 使得模型效果得到大大提升。作为一种前向加法模型, 他的核心是采用集成思想——Boosting 思想, 将多个弱学习器整合为一个强学习器^[10-11]。即用多棵树共同决策, 并且每棵树的结果都是目标值与之前所有树的预测结果之差, 并累加所有的结果, 得到最终的结果, 以此达到整个模型效果的提升^[12-13]。

LightGBM 是由多棵 CART(classification and regression tree), 即分类回归树组成, 因此它可以处理分类回归等问题。研究证明, 在处理高维数据特征时, LightGBM 算法可以依据特征重要性自动选择特征, 从而提高模型的泛化能力, 降低过拟合。

2.2 LightGBM 特征重要度

本文方法选择 LightGBM 算法的基尼系数来度量特征重要性, 具体原理公式如下:

假设有 m 个特征 $X_1, X_2, \dots, X_m$, 用 V 表示特征重要性, 计算出每个特征 X_j 的 Gini 指数评分 V_j^{Gini} 。Gini 指数的计算公式为

$$G_m = \sum_{k=1}^{|K|} \sum_{k' \neq k} p_{mk} p_{mk'} = 1 - \sum_{k=1}^{|K|} p_{mk}^2 \quad (1)$$

其中, V_j^{Gini} 表示特征 X_j 在节点 m 的重要性, 即节点 m 分枝前后的 Gini 指数变化量:

$$V_{jm}^{Gini} = G_m - G_l - G_r \quad (2)$$

对求得的各项特征的重要性进行归一化:

$$V_j^{Gini} = \frac{V_j^{Gini}}{\sum_{i=1}^c V_i^{Gini}} \quad (3)$$

其中, V_j^{Gini} 是特征 X_j 的基尼指数, $\sum_{i=1}^c V_i^{Gini}$ 是特征的增益之和。

2.3 皮尔森(Pearson)系数

皮尔森相关系数, 即皮尔森积矩相关系数(Pearson product-moment correlation coefficient), 是最常用的一种线性相关系数。用来反映两个属性 X 和 Y 的线性相关程度, 绝对值越大表明属性间相关性越强。两个属性间的皮尔森系数越高, 属性间的相关性越强, 在特征选择时, 尽量避免同时选取相关性较强的特征, 皮尔森系数的计算公式如下:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \quad (4)$$

其中, $cov(X,Y)$ 表示属性 X, Y 的协方差, $\sigma_X \sigma_Y$ 表示属性 X, Y 的标准差的乘积。

2.4 蚁群算法

蚁群算法(ant colony optimization, ACO)由意大利学者 Colormi 等^[14]于 1991 年提出, 是一种基于蚂蚁觅食行为的启发式优化算法。该算法模仿了蚂蚁觅食的全过程: 每只蚂蚁在觅食的过程中, 会释放出化学物质信息素, 且蚂蚁能够感知到其它伙伴释放的信息素。蚂蚁之间正是通过这种信息素来选择觅食路线, 信息素浓度越高, 对应路线被选择的概率便越高^[15]。蚁群算法通过信息素的挥发与更新来搜索最优解, 常用于解决优化问题, 如最小化函数值, 寻找最优路径等。

3 基于 LightGBM 和蚁群算法的特征选择方法

针对传统 Relief 选择算法对多维度数据特征选择有效性差、对存在较强依赖关系的特征选择结果不精确等问题, 本文提出了一种基于蚁群算法和 LightGBM 优化的特征选择方法, 简称 L-ACO。

3.1 路径转移概率计算

对于每一只蚂蚁来说,它经过的每一个特征,都有两种状态:0或1,0表示该节点未被选中状态;1表示被选中状态。每只蚂蚁从初始位置出发到终点的路径上,对于特征*i*和*j*,则存在4条可供选择的路径,分别是0→0、0→1、1→0、1→1。

对于每只蚂蚁所经过路径上的每一个节点,其被选择的概率如下:

$$p_{i,j}^k(t) = \begin{cases} \frac{\tau_{i,j}^\alpha \mu_{i,j}^\beta}{\sum_{k \in C} \tau_{i,j}^\alpha \mu_{i,j}^\beta} & (5) \\ 0, \text{ 否则} \end{cases}$$

其中,*i*表示当前的节点;*j*表示下一个待选的节点;*C*表示蚂蚁能够访问且尚未被访问到的节点集合; $\tau_{i,j}$ 表示节点间的信息素; α 是信息素因子; $\mu_{i,j}$ 表示节点的启发性信息; β 是启发式信息因子。

3.2 参数设置与信息素的更新

设初始时刻有*n*只蚂蚁,并随机初始化*n*只蚂蚁的初始位置,各条路径上的信息素 τ_0 初始值相等。 $\tau_{i,j} = V$,LightGBM特征重要性*V*的计算见公式(2),经过*t*时刻,当蚁群的所有蚂蚁全部完成一次遍历后,及时更新各条路径上的信息素^[13],首先是各路径上信息素的挥发,更新公式如下:

$$\tau_{i,j}(t) = (1 - \rho) * \tau_{i,j}(t - 1) \quad (6)$$

其中, ρ 为信息素挥发系数,通常设置 $0 < \rho \leq 1$ 。其次是蚂蚁在它们所经过边上释放信息素。

$$\tau_{i,j}(t) = \tau_{i,j}(t - 1) + \sum_{k=1}^n \Delta \tau_{i,j}^k \quad (7)$$

其中, $\Delta \tau_{i,j}^k$ 是第*k*只蚂蚁向它经过的边释放的信息素,定义为

$$\Delta \tau_{i,j}^k = \begin{cases} \frac{Q}{\rho_{i,j}} \\ 0, \text{ 否则} \end{cases} \quad (8)$$

其中*Q*表示信息素常数,表示蚂蚁循环一次所释放的信息素的总量, $\rho_{i,j}$ 为路径长度,即皮尔森相关系数,见公式(4)。结合公式(5)和公式(8)可知,皮尔森线性相关系数越小,该路径上的信息素浓度越高,且特征重要度越高,该路

径被其他蚂蚁选择的概率就高。

3.3 算法描述

本文结合LightGBM算法和蚁群算法,并引入皮尔森系数作为节点间的路径长度,提出一种特征选择方法,L-ACO算法。算法描述如下:首先运用LightGBM算法来计算每个特征的特征重要度*V*;然后利用皮尔森系数计算函数(见公式(4))算出特征间的相关性系数 ρ ;接下来构建L-ACO算法,计算各个特征的选择概率*P*,并进行参数的初始化和信息素 τ 的更新;最后运用L-ACO方法输出选择的最优特征子集*S_x*。

为了对本文算法进行更为准确直观的描述,现给出L-ACO算法伪代码,如下。

输入:数据集*S*

输出:最优特征子集*S_x*

1. 计算数据集各特征重要性*V*。

2. 构建L-ACO算法:

a. 初始化信息素 τ_0 ,蚂蚁个数*n*,启发性函数*V*,蚂蚁随机初始位置,最大迭代次数max。

b. 运用式(4)计算特征间的相关性系数。

c. for *i*=1 to max do

for *i*=1 to *m* do

运用式(5)计算每个蚂蚁在每条路径上的选择概率*P*。

End for

运用式(6)、式(7)、式(8)更新信息素 τ 。

End for

d. 输出最优特征子集*S_x*。

4 实验结果与分析

本文从UCI数据库中选择了一些经典的数据集对算法进行验证。经典的数据集见表1。

表1 UCI数据集描述

数据集	样本数	特征数	分类数
wdbc	198	33	2
wdbc	569	30	2
Sonar	208	60	2
Spambase	4601	57	2
Wine	178	13	3

4.1 实验环境与方法

本文实验环境为ThinkPad E550c笔记本,

Intel (R) Core (TM) i3-4005U CPU @ 1.70 GHz, 4.0 GB 内存, Windows10 64 位操作系统, 软件环境为 Jupyter 6.3.0、Python 3.8.3。

在分类器选择上, 选取逻辑回归作为分类器, 将本文方法选取的特征与原始数据特征进行了分类, 预测准确率比较; 首先对实验数据进行预处理, 即进行归一化处理, 将每个特征属性列的数据都归一到 $[0, 1]$ 之间, 然后对处理后的数据进行特征选择实验。本文方法与数据集原始特征数据在逻辑回归分类器上的分类准确率和特征个数的对比结果见表2。

表2 分类准确率和特征个数对比

数据集	本文方法	原始数据	本文选取特征个数	原始特征数
wpbc	0.935	0.912	4	33
wdbc	0.949	0.931	7	30
Sonar	0.813	0.725	5	60
Spambase	0.918	0.889	6	57
Wine	0.978	0.947	6	13

4.2 实验分析

对比分类精度, 从表2不难看出, 在UCI数据集上, 本文方法分类准确率相较于原始数据的分类准确率均取得较大提升。本文方法选取特征更少, 拟合效果较优。通过对比表3中的精确率、F1分数以及AUC三项评价指标, 发现L-ACO算法大多数情况下三项评价指标都略高于Relief算法。

表3 L-ACO算法与传统Relief算法实验性能比较

算法	指标	数据集				
		wpbc	wdbc	Sonar	Spambase	Wine
Relief	ACC	0.917	0.942	0.812	0.882	0.961
	F1	0.892	0.912	0.824	0.856	0.956
	AUC	0.904	0.933	0.765	0.874	0.952
L-ACO	ACC	0.935	0.949	0.813	0.918	0.978
	F1	0.905	0.931	0.858	0.921	0.981
	AUC	0.924	0.945	0.781	0.929	0.987

表3为本文方法L-ACO与传统Relief特征选择算法在UCI数据集上各项评价指标的对比

结果。与Relief算法相比, 本文方法在分类准确率、F1分数、AUC等指标性能上都有明显的优势。在Sonar数据集上, 本文方法分类准确率略低于Relief算法; 但本文方法的F1分数为0.858, 高于Relief算法的F1分数0.824; 在wpbc、wdbc和Spambase数据集上, 本文方法分类准确率、F1分数、AUC均优于传统Relief; 在多分类数据集Wine上, 本文方法在分类准确率、F1分数、AUC三项评价指标上均优于Relief算法。

综上所述, 本文L-ACO特征选择方法在分类准确率、F1分数、AUC上均优于传统的特征选择算法, 不仅达到了数据降维的目标, 同时提高了算法的泛化能力。

5 结语

本文提出的基于LightGBM算法和蚁群算法的特征选择方法(L-ACO), 运用各个特征的LightGBM特征重要度作为蚁群算法部分的启发式函数, 同时利用特征间的皮尔森(Pearson)相关系数表示各个特征之间的距离长短, 进而选择最优特征子集, 从而达到数据降维的效果。实验表明, 本文方法能在降低特征维度的同时, 确保数据集分类的准确率, 减少特征冗余, 提高算法的各项性能, 为实际生产中的特征选择问题提供了重要的方法依据。

参考文献:

- [1] 孙洁丽, 刘沛, 翟浩文. 基于高维数据的聚类研究综述[J]. 河北省科学院报, 2022, 39(5): 1-6.
- [2] 邹丽英, 刘祎. 超高维缺失响应数据的特征筛选[J]. 中国海洋大学学报(自然科学版), 2023, 53(1): 147-156.
- [3] 孟军, 尉双云. 基于近邻传播聚类的集成特征选择方法[J]. 计算机科学, 2015, 42(3): 241-244.
- [4] 罗妍, 王杲, 叶文玲. 基于LightGBM和SHAP的急性肾损伤可解释预测模型[J]. 电子与信息学报, 2022, 44(1): 27-38.
- [5] 崔树银, 汪昕杰. 基于最大信息系数和多目标Stacking集成学习的综合能源系统多元负荷预测[J]. 电力自动化设备, 2022, 42(5): 32-39.
- [6] 徐久成, 孟祥茹, 瞿康林, 等. 基于模糊邻域相对依赖互信息的特征选择方法[J]. 模糊系统与数学, 2023, 37(1): 121-135.

- [7] 王生武,陈红梅. 基于粗糙集和改进鲸鱼优化算法的特征选择方法[J]. 计算机科学, 2020, 47(2): 44-50.
- [8] 孙林,施恩惠,司珊珊,等. 基于AP聚类 and 互信息的弱标记特征选择方法[J]. 南京师大学报(自然科学版), 2022, 45(3): 108-115.
- [9] ABDULHUSSIEN A A, NASRUDIN M F, DARWISH S M, et al. Feature selection method based on quantum inspired genetic algorithm for Arabic signature verification [J]. Journal of King Saud University-Computer and Information Sciences, 2023, 35(3): 141-156.
- [10] 刘江,许康智,蔡伯根,等. 基于 LightGBM 的列车车载设备故障预测方法[J]. 北京交通大学学报, 2021, 45(4): 95-106.
- [11] SUENAGA D, TAKASE Y, ABE T, et al. Prediction accuracy of Random Forest, XGBoost, LightGBM, and artificial neural network for shear resistance of post-installed anchors [J]. Structures, 2023, 50: 1252-1263.
- [12] ALI S, KHORRAMI B, JEHANZAIB M, et al. Spatial downscaling of GRACE data based on XGBoost model for improved understanding of hydrological droughts in the Indus Basin Irrigation System (IBIS) [J]. Remote Sensing, 2023, 15(4): 873.
- [13] REN Q, WANG J. Research on enterprise digital-level classification based on XGBoost model [J]. Sustainability, 2023, 15(3): 2699.
- [14] COLORNI A, DORIGO M, MANIEZZO V. Distributed optimization by ant colonies [C] // Proc of the First European Conf on Artificial Life, Paris, France: Elsevier Publishing, 1991, 142: 134-142.
- [15] 郭城成, 田立勤, 武文星. 蚁群算法在求解旅行商问题中的应用综述[J]. 计算机系统应用, 2023, 32(3): 1-14.

Feature selection method based on LightGBM and ant colony optimization

Bie Chunyang, Tao Yiyong*

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

Abstract: When processing high-dimensional feature data, there are usually issues of redundancy and irrelevance. As a traditional feature selection algorithm, Relief is widely used due to its high stability and computational efficiency. However, the feature selection results are random, and for datasets with strong dependencies between features, such as collinearity, it may lead to inaccurate results. Based on the research on feature selection methods, an L-ACO method based on LightGBM and ant colony algorithm was proposed. The heuristic information of the L-ACO algorithm ant colony path search process was represented by the feature importance of LightGBM algorithm. At the same time, the Pearson correlation coefficient between features is used to adjust the concentration of pheromones in order to better control the correlation of features. Experiments have shown that the L-ACO method can reduce the number of features, reduce feature redundancy, and improve algorithm performance while ensuring classification accuracy.

Keywords: feature selection; LightGBM; ant colony optimization (ACO); Pearson coefficient

文章编号: 1007-1423(2024)04-0039-05

DOI: 10.3969/j.issn.1007-1423.2024.04.007

混叠场景下的轻量级口罩佩戴检测算法

安祯阳*

(重庆交通大学机电与车辆工程学院, 重庆 400074)

摘要: 针对混叠场景下口罩佩戴检测识别率低, 而现有检测模型结构复杂难以部署的难题, 提出了一种轻量级口罩佩戴检测算法。首先, 轻量化网络 MobileNetv3 作为混叠场景图像的特征提取网络; 其次提出通道混洗, 空间上利用不同感受野的卷积核进行特征提取的注意力机制, 实现特征信息的强化; 最后设计了损失函数解决了数据类不平衡问题, 提高了模型检测精度。在公开数据集测试表明, 模型平均检测精度为 78.1%, FPS 达到 65.53 Hz, 满足在小型设备部署的要求。

关键词: 轻量级; 目标检测; 部分卷积; 特征融合

0 引言

混叠场景下的检测任务是当前计算机视觉检测的热点问题^[1], 而利用深度卷积神经网络 (deep convolutional neural network, DCNN) 强大的特征提取能力构建目标检测模型愈发流行^[2-3]。基于 DCNN 的目标检测算法可分为两类: 一类是先提取目标框再进行预测的双阶段目标检测算法, 例如 R-CNN^[4]、Faster R-CNN^[5]等算法, 此类算法模型规模大, 检测速度慢; 另一类算法是基于回归的单阶段目标检测算法, 例如 YOLOv3^[6]、YOLOv4、SSD^[7]等算法, 此类算法实时性强, 但检测结果准确性有待提高。

本文为实现口罩佩戴检测模型的轻量化检测, 同时提高模型检测性能, 利用了图像增强算法扩充原始数据集, 而为充分挖掘图像特征信息, 减小模型尺寸, 提出了基于通道-空间注意力机制的特征融合方法, 最后再重新设计了损失函数, 解决了原始数据类不平衡的问题。

1 所提算法

1.1 整体网络结构

本文提出混叠场景下的轻量级口罩佩戴检测模型结构如图 1 所示。整个网络主要分为三部

分: 骨干特征提取网络 (Backbone)、颈部加强特征提取网络 (Neck) 以及头部检测网络 (Head)。对网络结构的改进部分主要集中在 Backbone 以及 Neck 两部分, 首先利用轻量化网络 MobileNetv3 作为特征提取网络, 然后为加强网络对目标的关注度, 需要融合不同尺度特征信息, 因此提出了基于通道-空间注意力机制的特征融合方法, 通道混洗注意力增强模块 (channel shuffle attention enhancement module, CSAEM)。

1.2 通道混洗注意力增强模块

模型为降低计算量, 在 3×3 卷积前使用 1×1 卷积实现降维, 使得通道充满约束, 降低检测精度, 而通道混洗^[8]是将原始特征的通道顺序打乱, 实现特征通道信息的变形重组。因此本文模型在嵌入注意力机制时引入通道混洗, 实现检测性能的提升, 结构如图 2 所示。

首先将不同尺度融合后的特征图记为 $F \in \mathbf{R}^{C \times H \times W}$, 将特征图分别经过全局平均池化和最大池化获取全局信息, 池化计算公式如下:

全局平均池化:

$$Z_k = \frac{1}{W \times H} \sum_i \sum_j u_{(i,j)} \quad (1)$$

$$(1 \leq i \leq w, 1 \leq j \leq H, 1 \leq k \leq C)$$

收稿日期: 2023-11-28 修稿日期: 2023-12-27

作者简介: *通信作者: 安祯阳 (1997—), 男, 重庆渝北人, 在读硕士研究生, 研究方向为机器人, E-mail: 1273880551@qq.com

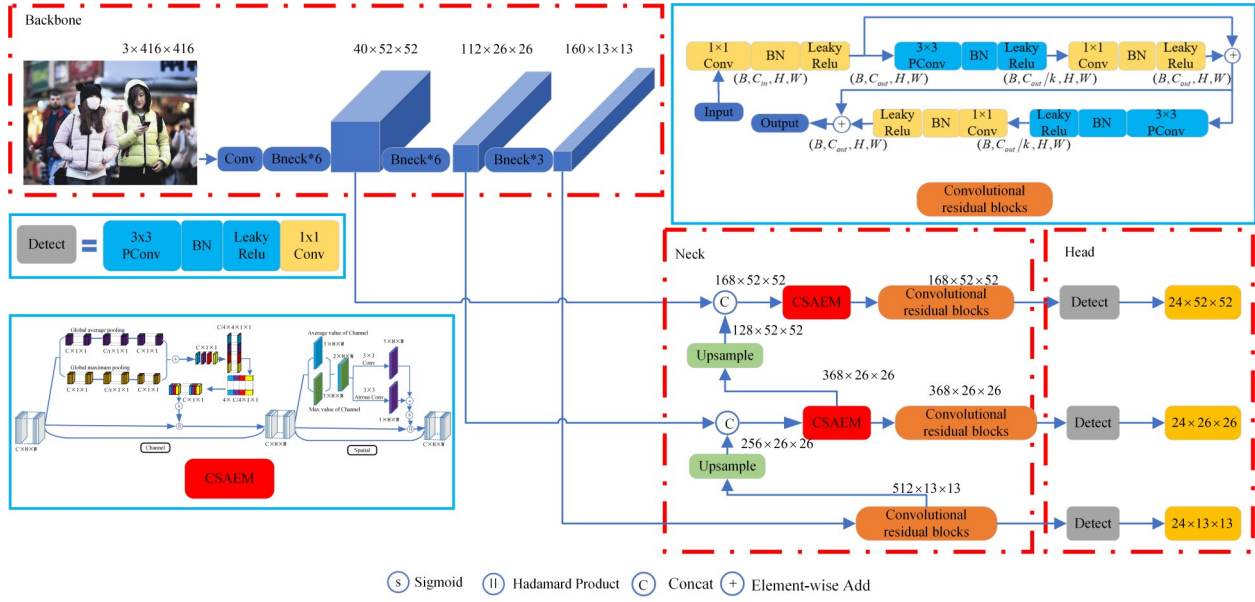


图1 网络结构

全局最大池化:

$$Z_k = \max[u_{(i,j)}(k)] \quad (2)$$

$$(1 \leq i \leq w, 1 \leq j \leq H, 1 \leq k \leq C)$$

其中: Z_k 代表 F 在第 k 个特征图上的池化输出值, $u_{(i,j)}$ 表示第 k 个特征图在 (i,j) 上的值, 最终得到两个 $C \times 1 \times 1$ 的图像特征。

池化图像特征后分别将其进行两次卷积(输出特征分别为 $C/r \times 1 \times 1$ 、 $C \times 1 \times 1$), 将其记为 φ , 参数量为 $C/r \times C$, 相较于全连接层构成的瓶颈结构, 减少了参数量。将最后输出的特征进行基于 element-wise add 操作, 得到通道权重, 然后通道权重执行通道混洗以及 Sigmoid 函数激活操作, 输出通道权重 W_c , 并将 W_c 与输入特征 F 逐元素相乘, 得到重标定的特征 F_c 。

$$F_c = F \otimes \left\{ \text{shuffle} \left\{ \mu \varphi [\text{AvgPool}(F)] + \varphi [\text{MaxPool}(F)] \right\} \right\} \quad (3)$$

其中: $\text{AvgPool}(F)$ 和 $\text{MaxPool}(F)$ 为输入特征经平均池化和最大池化后的输出, φ 为两次卷积操作, shuffle 为通道混洗, σ 为 Sigmoid 函数激活操作。

空间特征融合部分, 将重标定特征 F_c 使用平均池化和最大池化聚集空间信息, 拼接得到空间特征 F_{mid} , 将其分别经过标准 3×3 卷积和

3×3 空洞卷积(膨胀系数为 3, 感受野与 7×7 标准卷积相同)后, 然后将不同感受野获得的特征完成融合后再通过 Sigmoid 函数激活, 得到空间特征权重 W_s , 并将权重 W_s 与重标定的特征 F_c 逐元素相乘, 最终输出特征 F_s 。

$$\begin{cases} F_{\text{mid}} = [\text{AvgPool}(F_c); \text{MaxPool}(F_c)] \\ F_s = F_c \otimes \sigma [K^{3 \times 3}(F_{\text{mid}}) + K_D^{3 \times 3}(F_{\text{mid}})] \end{cases} \quad (4)$$

其中: $K^{3 \times 3}$ 代表普通 3×3 卷积, $K_D^{3 \times 3}$ 代表 3×3 空洞卷积。

通过通道混洗注意力增强模块可抑制特征融合时产生的冗余信息, 改善通道约束问题。

1.3 损失函数优化

为解决原始数据集类不平衡的问题, 对训练时的损失函数做了优化。损失函数表达式为

$$\text{loss}_{\text{all}} = \text{loss}_{\text{obj}} + \text{loss}_{\text{cls}} + \text{loss}_{\text{loc}} \quad (5)$$

其中: loss_{obj} 为置信度误差损失, loss_{cls} 为分类误差损失, loss_{loc} 为回归误差损失。

对于实验数据集类别失衡, 提出在损失函数的分类误差当中设计类权重系数。

$$\begin{cases} \text{loss} = [\hat{p}_{i,c}^j \log(p_{i,c}^j) + (1 - \hat{p}_{i,c}^j) \log(1 - p_{i,c}^j)] \\ \text{loss}_{\text{cls}} = - \sum_{i=0}^{s^2} \sum_{j=0}^B I_{ij}^{\text{obj}} \sum_{c \in \text{classes}} (\text{loss} * \text{weight}_c) \end{cases} \quad (6)$$

其中: s^2 代表不同尺度的输出, $p_{i,c}^j$ 代表坐标位

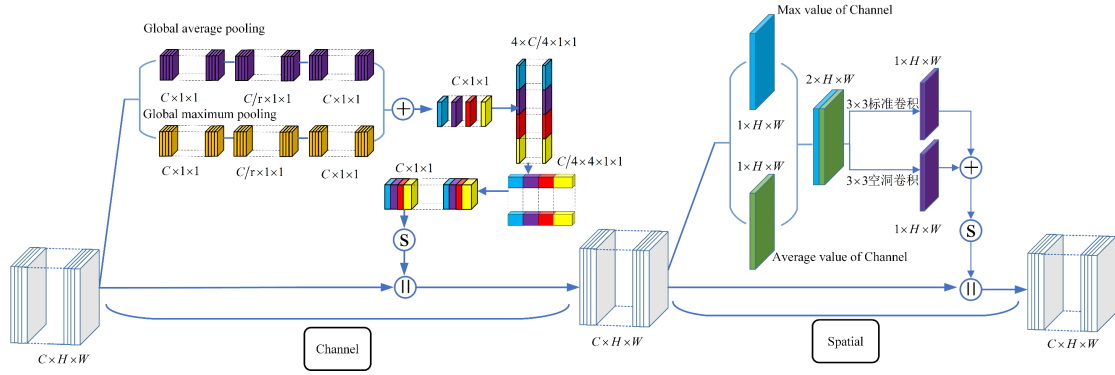


图2 通道混洗注意力增强模块

置在 (i, j) 预测框属于类别 c 的概率，而 $\hat{p}_{i,c}$ 代表标记框所属类别的真实值。类权重系数则满足：

$$weight_{c \in classes} = \frac{N}{n_{c \in classes} * num_class} \quad (7)$$

其中： num_class 为数据集类别数目， $n_{c \in classes}$ 为类别 c 的标记框个数，而 N 为数据集标记框总个数。

当预测框与真实框不相交时， IoU loss 为0，不能反映预测框与真实框的距离。而 $GIoU$ 作为距离度量指标，能够同时关注预测框与真实框的重叠区域和非重叠区域。因此，本文将引入 $GIoU$ 来设计回归误差损失函数。

回归误差损失：

$$loss_{loc} = 1 - GIoU = 1 - \left(\frac{A \cap B}{A \cup B} - \frac{|C - A \cup b|}{|C|} \right) \quad (8)$$

2 实验与分析

2.1 评价指标

本文采用类别平均精度 mAP 、类别平均召回率 mAR 、 $F1$ 以及模型尺寸Size来对模型进行评价。通过设计混淆矩阵对精确率 P 和召回率 R 进行计算，混淆矩阵包含 P_T （真正例）、 P_F （假正例）、 N_T （真负例）、 N_F （假负例）：

$$\begin{cases} P = \frac{P_T}{P_T + P_F} \\ R = \frac{P_T}{P_T + N_F} \end{cases} \quad (9)$$

则平均精度 AP 和平均召回率 AR ：

$$\begin{cases} AP = \int PdR \\ AR = 2 \int_{0.5}^1 R(o)d(o) \end{cases} \quad (10)$$

类别平均精度 mAP 和类别平均召回率 mAR 表达式如下：

$$\begin{cases} mAP = \frac{\sum_{i=1}^k AP_i}{k} \\ mAR = \frac{\sum_{i=1}^k AR_i}{k} \end{cases} \quad (11)$$

为综合考虑类别平均精度 mAP 与类别平均召回率 mAR ，以 $F1$ 作为评价指标之一，表达式如下：

$$F1 = 2 \frac{mAP * mAR}{mAP + mAR} \quad (12)$$

2.2 模型轻量化测试

模型参数数量和计算量对比见表1。在保证模型性能的基础上，通过将模型轻量化处理，减少模型的参数数量与计算量，可有效避免模型在训练时出现过拟合，还可以降低模型的大小，提高模型的检测速度，方便在小型设备当中的部署。

表1 参数数量和计算量对比

模型	参数数量/M	计算量/G
YOLOv4	63.95	29.89
SSD	23.88	30.54
GhostNet-YOLOv4	11.01	3.27
YOLOv4-tiny	5.89	3.41
YOLOv5s	7.09	8.20
本文模型	4.16	1.33

2.3 实验结果与分析

本文对改进的模型进行了六组消融实验，见表2。实验1为以MobileNetv3作为模型的特征

提取网络的实验结果；实验2在实验1的基础上对数据进行了重采样扩充预处理；实验3、实验4、实验5分别在实验2基础上结合分类误差损失优化、回归误差损失优化以及通道混洗注意力增强模块。

由表2可以看出，因原始数据较少且类别不平衡，使得检测效果较差，在实验2增加数据重采样扩充处理后 mAP 与 mAR 分别提升 18.9 和 19.7 个百分点；实验3对分类误差损失函数设计

类权重比例，缓解数据类不平衡带来的影响， mAP 下降 0.3 个百分点，但 mAR 提升 1.1 个百分点，检测综合效果得到改善；实验4采用 $GloU$ loss 作为回归误差损失函数， mAP 与 mAR 分别提升 15.4 和 11.9 个百分点；实验5在损失函数优化基础上，嵌入了设计的 CSAEM，实现了对图像特征信息的强化，抑制了在特征融合时产生的冗余信息， mAP 与 mAR 分别提升 2.1 和 3.4 个百分点。

表2 消融实验得到的实验结果(%)

	with_mask		without_mask		mask_wearied_incorrect		$mAP@0.5$	mAR	F1
	AP	AR	AP	AR	AP	AR			
实验1	68.7	73.5	42.6	52.2	10.0	12.5	40.5	46.1	43.1
实验2	72.6	77.3	48.5	57.8	57.1	62.3	59.4	65.8	62.4
实验3	71.4	77.4	49.6	59.6	56.2	63.6	59.1	66.9	62.7
实验4	83.0	85.0	63.4	67.7	77.9	80.3	74.8	77.7	76.2
实验5	82.3	84.6	68.4	73.8	80.2	84.8	76.9	81.1	78.3

3 结语

针对混叠场景下口罩佩戴检测识别率低，现有模型尺寸较大，结构复杂难以实现小型设备部署等难题，提出了一种混叠场景下的轻量级口罩佩戴检测算法，模型以轻量化网络 MobileNetv3 作为特征提取网络，然后提出了一种注意力机制结合通道混洗特征融合方法，降低了模型尺寸，实现了图像特征信息的强化；最后重新设计了损失函数，缓解了数据类不平衡的问题。在公开数据集 Face Mask Detection 与同类模型对比，本文模型平均检测精度为 78.1%，平均召回率达到 82.5%，满足在小型设备部署的要求。

参考文献

- [1] 张艳,张明路,吕晓玲,等.深度学习小目标检测算法研究综述[J].计算机工程与应用,2022,58(15):1-17.
- [2] 吴敌,李明辉,马文凯,等.基于改进YOLOv5的钢材表面缺陷检测[J].陕西科技大学学报,2023,41(2):162-169.
- [3] LI S, WU L, MENG W Z, et al. DVPPIR:

privacy-preserving image retrieval based on DCNN and VHE [J]. Neural Computing & Applications, 2022,34(17):14355-14371.

- [4] LEE H, EUM S, KWON H. ME R-CNN: multi-expert R-CNN for object detection [J]. IEEE Transactions on Image Processing, 2020, 29: 1030-1044.
- [5] SHARMA V, MIR R N. Saliency guided faster-RCNN (SGFr-RCNN) model for object detection and recognition [J]. Journal of King Saud University-Computer and Information Sciences, 2022,34(5):1687-1699.
- [6] TIAN Y N, YANG G D, WANG Z, et al. Apple detection during different growth stages in orchards using the improved YOLO-V3 model [J]. Computers and Electronics in Agriculture, 2019, 157: 417-426.
- [7] ZHAI S P, SHANG D R, WANG S H, et al. DF-SSD: an improved SSD object detection algorithm based on DenseNet and feature fusion [J]. IEEE Access, 2020, 8: 24344-24357.
- [8] Larxel. Mask dataset [EB/OL]. [2023-07-23]. <https://makeml.app/datasets/mask>.

(下转第47页)

文章编号: 1007-1423(2024)04-0043-05

DOI: 10.3969/j.issn.1007-1423.2024.04.008

基于自适应阈值的 ORB 特征点提取算法研究

谭帅奇, 帅鹏举*

(重庆开放大学重庆工商职业学院电子信息工程学院, 重庆 401520)

摘要: 在弱纹理场景下, 针对 ORB 算法提取特征点过程中, 固定阈值检测 FAST 角点可能会导致提取效果不佳进而影响匹配精度, 本文提出了自适应阈值的 ORB 特征点提取算法, 通过图像灰度差值和像素分布概率来计算图像对比度, 根据对比度动态计算角点检测阈值。然后根据动态阈值算法实现特征点的提取, 采用暴力匹配算法和快速最近邻匹配(FLANN)两种匹配算法, 在 EuRoc 数据集上分别对比了 ORB 算法、SIFT 算法与本文算法的特征点匹配精度和耗时。结果表明, 在匹配精度上比 ORB 算法提升了 26.6%, 比 SIFT 算法提升了 32.7%。

关键词: ORB 算法; 自适应阈值; 特征点; 视觉 SLAM

0 引言

随着科技的进步, 机器人技术飞速发展, 智能机器人也随之应运而生。智能机器人是一种基于传感器、数据管理和信息决策等多种技术的集合体, 它们可以在各种环境下自主完成任务, 应用领域也越来越广泛^[1]。

要实现机器人的智能化, 就要解决机器人的定位与导航的问题, 即 SLAM (simultaneous localization and mapping)^[2]问题。在实际应用中, SLAM 一般基于不同类型的传感器来获取环境信息。根据传感器的不同, SLAM 主要分为惯性 SLAM、激光雷达 SLAM 和视觉 SLAM。

现阶段视觉 SLAM 可分为: 基于滤波器的方法, 基于环境外观信息的直接法, 以及基于环境特征信息的特征法。不同种类的算法都有相应的优缺点, 详见表 1。

基于环境特征的特征点法, 是一种常用的视觉 SLAM 方法。通过使用特征检测算法, 提取图像中的角点, 并基于这些角点的周围环境信息来推断出相应的特征。即通过特征匹配和数据关联来优化问题, 从而求解状态变量。

表 1 视觉 SLAM 技术优缺点

算法	优势	不足
MonoSLAM	首个单目视觉 SLAM 系统	应用范围有限
DTAM	直接法稠密建图	计算量大
OKVIS	基于单目惯性导航的 SLAM 系统	难回环
ORB-SLAM	效果最佳的基于特征点的视觉 SLAM 系统	弱纹理场景建图效果不佳

目前, ORB-SLAM 是较为成熟的特征点算法, 它使用 ORB 算法来提取图像中的特征点。但是在弱纹理场景下: 实验室的楼道和楼梯、科技园区的写字楼与道路、室内的墙壁和其他装饰物品等, 这些场景大多都存在颜色和亮度变化不明显, 环境的特征信息不足, 基于固定阈值的 ORB 特征点提取算法难以提取有效特征, 则会大大降低定位与建图的实际效果。

为了解决环境特征不足的问题, 本文提出自适应阈值的 ORB 特征点提取算法, 通过图像灰度差值和像素分布概率来计算图像对比度, 根据对比度动态计算角点检测阈值。

收稿日期: 2023-12-22 修稿日期: 2024-01-10

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)

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: 帅鹏举 (1995—), 男, 重庆人, 硕士研究生, 研究方向为人工智能, E-mail: maimaimai1298@gmail.com

1 自适应阈值的特征点提取算法

特征点是指在图像中具有鲜明、稳定和可重复性的像素点，如角点、边缘等。通过特征点提取，可以将图像中的信息从高维度到低维度进行降维，从而使得机器人在处理图像信息时更加高效。同时，特征点提取还可以实现对图像中具有代表性的区域进行抽象和描述，从而实现机器人在未知环境中的定位和建图。

为了确保图像特征点提取的鲁棒性，需要考虑相机的运动、环境的变化等多种因素。为此，图像视觉领域的学者提出了多种经典的特征提取算法，如 SIFT^[3]、SURF^[4]和 ORB^[5]等。ORB 特征点考虑了图像像素的灰度差，能够实现快速的特征点检测和提取，具有较好的实时性和稳定性，但图像灰度差值的计算阈值为固定常数，对于环境图像灰度值变化过大，则固定阈值的图像灰度差计算方式难以适用于不同场景。

1.1 自适应阈值的 ORB 点特征提取

ORB 算法依靠固定阈值计算中心像素与周围点像素的差值，从而判断该中心点是否为特征点。但是在弱纹理场景下时，图像的像素变化不明显，但环境中的确存在特征点，若还采用 ORB 特征点提取算法来计算像素差，则可能导致特征点提取的数量和质量都不满足要求，难以实现特征跟踪，最终导致建图效果不佳。如何对像素阈值进行动态化调整，是解决该问题的关键所在。本文提出基于图像对比度的方法，计算特征区域的阈值。

1.1.1 自适应阈值

ORB 特征的提取包括：提取 FAST 关键点和构建 BRIEF 描述子，FAST 关键点如图 1 所示。

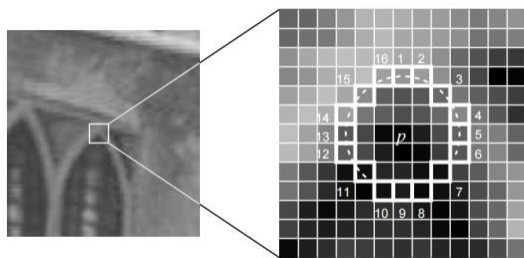


图 1 FAST 关键点

FAST 关键点又叫做角点，将角点中心 p 与周围所有像素值的灰度值进行比较，判断该点是否为像素变换显著的角点。在灰度值比较过程中，若采用一个固定的阈值作为比较标准，则对于图像对比度变化不明显的区域，可能导致角点提取失败。例如：设定阈值为 10，若 p 点灰度值-周围某一点灰度值 > 10 的点超过 n 个，则认为该点为一个角点。本文的自适应阈值计算步骤如下：

Step1: 计算点 p 周围 16 个像素点的灰度均值 $\mu_{I_{w \times h}}$:

$$\mu_{I_{w \times h}} = \frac{1}{w \times h} \sum_{I_{w \times h}} I(x, y) \quad (1)$$

其中： $I(x, y)$ 为每一个像素点的灰度值， w 为区域像素宽度， h 为区域像素高度。

Step2: 16 个像素点的灰度值 $I(x, y)$ 与灰度均值 $\mu_{I_{w \times h}}$ 的差值 ε 与差值平方和 ξ_i 分别为

$$\varepsilon = I(x, y) - \mu_{I_{w \times h}} \quad (2)$$

$$\xi_i = \left(I(x, y) - \mu_{I_{w \times h}} \right)^2 \quad (3)$$

Step3: 计算 16 个像素点灰度与灰度均值差的像素分布概率 P :

$$P = P_\varepsilon(i, j) \quad (4)$$

Step4: 待提取图像区域的灰度对比度 G 为

$$G = \sum_{\varepsilon} P \xi_i \quad (5)$$

Step5: 阈值 T 为

$$T = kG \quad (6)$$

其中 k 为 0~1 之间的常数，根据不同场景环境调整参数。

1.1.2 FAST 角点提取

得到自适应阈值 T 后，中心点 p 周围的 16 个像素点编号为 $p(1) \sim p(16)$ ，计算 4 个像素点 $p(1)$ 、 $p(5)$ 、 $p(9)$ 、 $p(13)$ 与中心点 p 的像素差。若有 3 个像素点差值都大于或者小于阈值 T ，则认为该点为一个角点。

为了保持尺度不变性，ORB 特征使用图像金字塔从原始图像中提取角点。在每个金字塔层中，使用 Harris 响应值来选择特征点，以避免提取过多的特征。此外，ORB 特征使用灰度质

心法来确定特征点的方向。对于每个特征点，定义一个小的图像块，并计算其矩 m_{pq} 为

$$m_{pq} = \sum_{x,y \in B} x^p y^q I(x,y), p, q = \{0, 1\} \quad (7)$$

其中 $I(x,y)$ 表示图像上某个像素点的灰度值，则该特征点的领域质心 C 表示为

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \quad (8)$$

特征点与该图像块的质心夹角 θ 表示特征点的方向：

$$\theta = \arctan 2(m_{01}, m_{10}) \quad (9)$$

利用图像金字塔方法提取的FAST特征点具有旋转不变性和尺度不变性，随着相机位姿的变换，所提取的角点在不同图像帧之间的匹配成功率较高，提高了系统的鲁棒性。特别是，针对本文所研究的弱纹理场景下的特征点提取，一样能提取数量较多、质量较好的角点。

1.2 特征点匹配

BRIEF 是一种衡量两个特征之间的相似度的二进制描述符，BRIEF 描述子的计算过程为：以 FAST 角点 p 为中心，构建半径为邻域半径 3 的圆 O ，并在圆内随机选取 n 个点，将这 n 个点记作：

$$p_1(A,B), p_2(A,B), \dots, p_n(A,B) \quad (10)$$

则可计算描述子 $F(p(A,B))$ 为

$$F(p(A,B)) = \begin{cases} 1, & I_A > I_B \\ 0, & I_A \leq I_B \end{cases} \quad (11)$$

计算得到描述子之间的相似度之后，则对相邻帧进行特征匹配。ORB-SLAM 主流匹配方法主要有暴力匹配和快速最近邻接匹配。暴力匹配的基本思想是对于每个特征点，在另一张图像中找到与之最相似的特征点。暴力匹配实现简单，易于理解和调试，也可以找到所有可能的匹配点。但它计算量较大，时间复杂度 $O(N^2)$ ，其中 N 为特征点的数量，对于大规模的图像，计算时间会非常长，也可能出现误匹配的情况。快速最近邻接匹配是一种更高效的特征点匹配算法，它的基本思想是利用数据结构（如 $kd-tree$ ）来加速匹配过程，可以通过设置阈值来排除一些不合适的匹配点，减少误匹

配的情况。具体实现时，需要将每个特征点的描述子插入到 $kd-tree$ 中，然后对于每个特征点，在 $kd-tree$ 中查找与之最相似的点作为匹配点。优点是计算速度快，时间复杂度为 $O(\log N)$ ，其中 N 为特征点的数量。

综上所述，本文采用快速最近邻接匹配 (FLANN) 算法进行特征点匹配。其步骤如下：

(1) 对参考图像和待匹配图像中的 ORB 特征点进行描述符计算。

(2) 使用 FLANN 算法对参考图像中的 ORB 特征点描述符进行索引，以加速匹配过程。

(3) 对待匹配图像中的 ORB 特征点描述符进行匹配。具体地，对于待匹配图像中的每个 ORB 特征点描述符，通过 FLANN 算法在参考图像中找到最近的 k 个 ORB 特征点描述符，并计算它们之间的距离。

(4) 根据距离比值进行剔除。具体地，对于待匹配图像中的每个 ORB 特征点描述符，将其与参考图像中的最近 ORB 特征点和次近 ORB 特征点之间的距离比值进行计算，如果比值小于一个预设的阈值，则将其匹配上。

(5) 对匹配结果进行筛选和优化。具体地，对于匹配上的 ORB 特征点对，可以使用 RANSAC 算法对其进行筛选和优化，以进一步提高匹配结果的精度和鲁棒性。

2 实验结果与分析

通过以上步骤，快速最近邻接匹配算法可以实现高效、准确的特征点匹配。本文分别采用 SIFT、ORB 与本文算法提取图像特征，并使用暴力匹配和快速最近邻接匹配来对相邻帧进行特征点匹配，在公开数据集上特征点匹配实验结果如图 2 所示。

从图 2 可以看出，ORB 特征点提取数量最多，但正确匹配的点对数明显低于本文算法。SIFT 特征点提取数量相较于 ORB 少一些，但所耗费时间比 ORB 和本文算法都长。针对弱纹理环境场景，特征点数量原本就不丰富，又考虑到系统的实时性，因此本文算法更优。三种算法具体匹配精度与耗时情况数值数据见表 2。



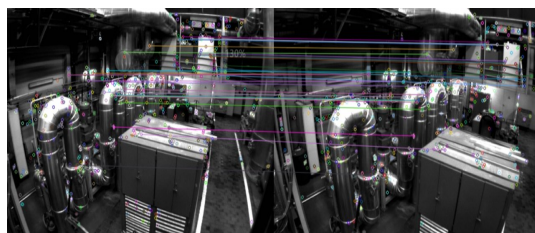
(a) SIFT特征点提取与暴力匹配效果



(b) SIFT特征点提取与快速最近邻接匹配效果



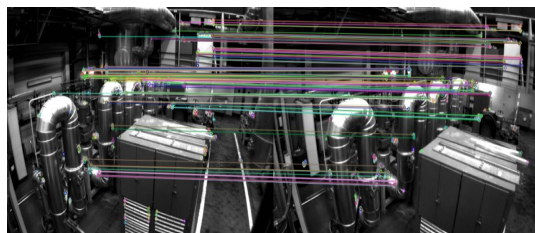
(c) ORB特征点提取与暴力匹配效果



(d) ORB特征点提取与快速最近邻接匹配效果



(e) 本文算法特征点提取与暴力匹配效果



(f) 本文算法特征点提取与快速最近邻接匹配效果

图2 SIFT、ORB与本文算法特征点提取和匹配效果

表2 SIFT、ORB与本文算法匹配精度与耗时

算法类型	SIFT	ORB	本文算法
所有匹配点对	500	500	500
正确匹配点对	122	128	162
匹配精度	0.244	0.256	0.324
耗时(s)	0.182	0.091	0.112

从表2可以看出,在500对匹配点中,本文算法匹配精度比SIFT算法高8个百分点,比ORB算法高6.8个百分点;耗时也低于SIFT,比ORB仅高21个百分点,并不影响算法的实时性。由此,本文算法综合效果显著。

3 结语

本文阐述了改进的ORB特征点提取算法原理。考虑到ORB算法固定阈值提取特征点的缺陷,将原本的固定阈值根据对比度改进为自适应阈值,计算特征区域的图像对比度,根据对比度来动态计算阈值,用自适应阈值的方法提取图像特征点。通过对比实验,对比各主流算法提取特征点的效果与本文自适应阈值算法提取特征点的效果,从匹配精度和耗时两方面总结实验结果。

参考文献:

- [1] 康荣杰,杨铨浩,杨名远.会思考的机器:机械智能[J].机械工程学报,2018,54(13):29-38.
- [2] DURRANT-WHYTE H, BAILEY T. Simultaneous localization and mapping: part I[J]. IEEE Robotics & Automation Magazine, 2006, 29(6): 99-110.
- [3] NG P, HENIKOFF S. SIFT: predicting amino acid changes that affect protein function[J]. Nucleic Acids Research, 2003, 31(13): 3812-3814.
- [4] BAY H. SURF: speeded up robust features[J]. Computer Vision & Image Understanding, 2006, 110(3): 404-417.
- [5] RUBLEE E, RABAU D V, KONOLIGE K. ORB: an efficient alternative to SIFT or SURF[C]//Proceedings of the 2011 International Conference on Computer Vision, Barcelona. Spain: IEEE Press, 2011: 2564-2571.

Research on ORB feature point extraction algorithm based on adaptive threshold

Tan Shuaiqi, Shuai Pengju*

(School of Electronic Information Engineering, Chongqing Open University Chongqing Technology and Business Institute, Chongqing 401520, China)

Abstract: In weak texture scenes, in the process of extracting feature points with the ORB algorithm, fixed threshold detection of FAST corner points may lead to poor extraction results and affect the matching accuracy. This paper proposes an adaptive threshold ORB feature point extraction algorithm, which uses image grayscale The difference value and pixel distribution probability are used to calculate the image contrast, and the corner detection threshold is dynamically calculated based on the contrast. Then the feature points were extracted based on the dynamic threshold algorithm, using two matching algorithms: brute force matching algorithm and fast nearest neighbor matching (FLANN). The feature point matching accuracy and accuracy of the ORB algorithm, SIFT algorithm and the algorithm in this paper were compared on the EuRoc data set. time consuming. The results show that the matching accuracy is 26.6% higher than the ORB algorithm and 32.7% higher than the SIFT algorithm.

Keywords: ORB algorithm; adaptive threshold; feature points; visual SLAM

(上接第42页)

Lightweight mask wearing detection algorithm in aliasing scenario

An Zhenyang*

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

Abstract: A lightweight mask wearing detection algorithm is proposed to address the low recognition rate of mask wearing detection in mixed scenes and the difficulty of deploying existing detection models with complex structures. Firstly, the lightweight network MobileNetv3 serves as a feature extraction network for aliasing scene images; Secondly, channel mixing is proposed, which utilizes convolutional kernels from different receptive fields in space for feature extraction and enhances feature information; Finally, a loss function was designed to solve the problem of imbalanced data classes and improve the accuracy of model detection. In public dataset testing, the average detection accuracy of the model is 78.1%, and the FPS reaches 65.53 Hz, meeting the requirements for deployment on small devices.

Keywords: lightweight; target detection; partial convolution; feature fusion

实践与经验

文章编号: 1007-1423(2024)04-0048-05

DOI: 10.3969/j.issn.1007-1423.2024.04.009

基于注意力机制的三维模型特征提取

王欢欢*, 李舒晴

(华北水利水电大学信息工程学院, 郑州 450018)

摘要: 目前在基于视图的三维模型检索技术中, 对多视图特征提取的大多方法, 关注于视图的全局特征信息而忽略了对视图局部特征信息和多视图之间的相关性的探究。针对此问题提出一种新的特征提取方法, 利用深度学习中的卷积神经网络, 并结合注意力机制提取特征, 以提升其判别性。方法在 ModelNet40 上进行实验分析, 将三维模型的多个视图作为输入, 在网络层中加入注意力模块进行特征提取分类, 结果表明, 该方法在分类准确度方面优于已有的典型算法。

关键词: 三维模型检索; 卷积神经网络; 注意力机制; CBAM

0 引言

近年来, 随着三维建模技术不断发展, 三维模型已经在多个领域得到了广泛应用, 例如建筑设计、机械设计、动画制作等。其中, 模型设计是应用中的关键环节, 完成一个模型作品, 是非常费时又费力的。但有关研究表明, 多数设计师可在原有模型基础上设计开发新模型, 从而缩短设计所用的时间, 降低开发过程中的各项成本, 并保证其可靠性和质量。如今, 随着三维模型的日益增多, 如何在海量的三维模型中快速、准确地找到与用户需求相符合的模型成为一个亟待解决的问题。三维模型检索成为众多研究者为解决这一问题的研究热点, 通过三维模型检索技术, 用户可以快速找到所需的模型, 并进行修改或组合, 以满足实际需求。在实际应用中, 三维模型检索技术可以用于许多领域, 例如工业产品设计、虚拟现实、虚拟人、三维游戏、教育、影视动画等。

研究者们提出了多种三维模型检索方法, 其一般流程如图 1 所示。早期采用文本标注的方法, 虽然简单直观, 但是存在很多局限性, 文本信息大多时候很难直观地描述三维模型的检

索需求, 且对这些模型进行标注需要消耗大量的人力资源, 无法满足实际应用的需求^[1]。三维模型的结构和形状信息更能够反映模型的本质特征, 准确描述出其特征, 成为提高检索分类准确性的关键。描述符是用来描述三维模型的特征和属性的, 按照不同的三维模型描述子表示形式, 又可将三维模型检索方法分为三类: 基于体素、基于点云和基于视图的方法^[2]。其中基于视图的方法将三维模型分类转化为二维视图的分类问题, 更符合人类对三维模型的视觉认知, 因此得到了广泛的应用^[3]。

近年来, 随着各深度学习模型在图形图像和计算机视觉等领域相继取得的显著成果, 越来越多的研究人员将基于深度学习的技术引入三维模型检索领域。基于多视图方法通过从不同角度或视角观察三维模型, 并将这些视图的信息融合起来, 以获得更丰富的特征表示。这种方法的优点在于它可以更好地捕捉三维模型的局部特征、形状细节和多尺度信息, 从而提高了检索性能。目前, 基于多视图的三维模型检索已经取得了一定的研究成果。其中, 研究者们主要关注于如何利用多视图信息提高三维模型检索的准确率和效率。在这些方法中, 深

收稿日期: 2023-10-09 修稿日期: 2023-11-04

作者简介: *通信作者: 王欢欢(1995—), 女, 河南驻马店人, 硕士, 研究方向为三维模型检索、深度学习等, E-mail: wanghuanhuan@stu.ncwu.edu.cn; 李舒晴(2000—), 女, 河南商丘人, 硕士, 研究方向为目标识别、深度学习

度学习模型被训练为从多个二维视图中提取特征后聚合, 并进行分类。其中多视图方法的最早研究之一是 Su 等^[4]提出的 MVCNN, 该方法具有里程碑的意义, 它将每个视图看作一个图像, 然后将这些视图输入到特征提取网络(CNN)中进行处理, 引入了最大池化特征融合策略, 将每个视图的特征进行融合, 以获得全局的三维形状特征。这有助于捕捉形状的多尺度信息和局部细节。然而, 只保留每个特征图的最大值, 而忽略了其他值, 无法充分利用多视图特性。Yu 等^[5]提出了 MHBNet, 在获取各视图的特征后, 方法引入了双线性池化技术, 有助于捕捉特征之间的关系, 从而提高了特征融合的效果。Wang 等^[6]提出了一种基于视图集聚类和池化的多视图三维物体识别算法。该算法的主要思想是利用视图集聚类和池化操作对多视图信息进行融合和特征提取, 从而实现更准确的三维物体识别。Feng 等^[7]提出了组视图 CNN (group view CNN, GVCNN), 以利用视图之间固有的层次相关性和区别。该模型由层次化的视图组形状体系结构组成, 该体系结构分为视图层、组层和形状层, 并采用分组策略进行重新排列。此外, 还有研究者利用循环神经网络(RNN)和长短期记忆网络(LSTM)等对三维模型的序列特征进行分析和建模。比如 Ma 等^[8]使用高速公路连接的长短期记忆单元(LSTM); Dai 等^[9]采用了 Siamese 结构的 LSTM; Han 等^[10]采用了双向 LSTM; Jiang 等^[11]在多个循环视图序列上应用了多个 LSTM; 此外, Chen 等^[12]提出通过结合三维模型的视图信息和循环注意力机制来提高分类准确率。

然而, 现有的方法存在一些缺点, 首先, 典型的方法通常提取多个视图的全局特征, 忽略了对视图局部特征的探索。其次, 在对多视图特征的提取中没有考虑不同图像区域之间的通道之间的关系, 影响了对描述符判别的准确性。为解决这些问题, 本文提出一种方法, 在特征提取过程中, 将注意力机制引入卷积神经网络, 在特征提取骨干网络的全连接层前面加入 CBAM 模块, 利用其通道注意力和空间注意力模块, 关注视图中局部特征和特征图中的区域和通道相关性, 提高分类准确性。

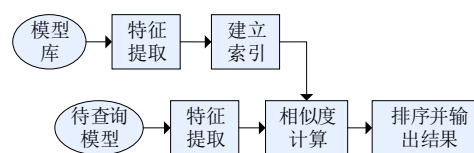


图1 三维模型检索流程

1 方法

本文是基于注意力机制的特征提取方法：
①采用卷积神经网络, 来提取视图的特征。将原始的 3D 物体模型从多角度投影到二维平面上, 并渲染成 n 个视图作为输入视图, 每个视图都被输入到具有共享权重的卷积神经网络中, 进行特征提取。
②在全连接层前面, 利用 CBAM 模块中的通道注意力和空间注意力分别探索视图内部和视图之间的关系来生成全局形状描述符。
③由全连接层组成, 将池化特征的形状描述符输入到全连接网络中, 完成目标识别, 框架如图 2 所示。

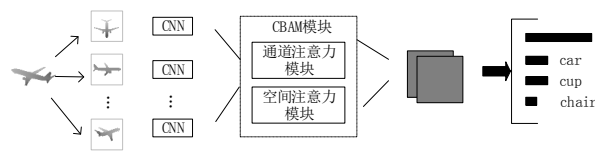


图2 结合 CBAM 模块的特征提取框架

1.1 特征提取网络

在三维模型检索中, 特征提取网络用于从三维模型中提取有效的特征表示, 以便进行相似度匹配和检索。特征提取网络可以采用多种方法, 比如可以采用卷积神经网络(CNN)对三维模型进行特征提取。CNN 可以对三维模型进行多尺度、多视角的特征提取, 从而得到更全面的特征表示。此外, 还可以采用循环神经网络(RNN)对三维模型的序列信息进行建模, 从而得到更准确的特征表示。特征提取常用的卷积神经网络包括 VGG-M、GoogleNet、ResNet、DenseNet 等, 本文采用了在二维图像特征提取方面表现突出的 ResNet 网络。ResNet 网络是微软实验室在 2015 年提出的深度神经网络模型, 该模型在当年的 ImageNet 竞赛中荣获分类任务第一名, 目标检测第一名。ResNet 的亮点在于

其超深的网络结构(突破 1000 层)和提出的残差(residual)模块。ResNet 通过在深度神经网络中加入残差单元(Residual Unit)使得训练深度比以前更加高效。ResNet 网络又有 ResNet18、ResNet50、ResNet101 等不同规模。本文采用 ResNet50 作为特征提取网络。结构如图 3 所示。

1.2 注意力模块 CBAM

注意力机制是一种在视觉和认知科学中广泛使用的技术,它模拟了人类的视觉或思考过程中的“注意力”。该机制允许模型有选择性地关注输入数据中的特定部分,而忽略其他部分,从而使模型能够在处理信息时更加聚焦和有效。在多视图特征提取任务中,注意力机制可以帮助模型更好地关注视图局部特征信息,从而提升分类准确性。CBAM 模块是 Woo 等^[13]提出的一种注意力模块,可以与任何类型的卷积神经网络(CNN)一起使用。CBAM 模块的作用是增强模型对于输入数据的关注能力,使其能够更加有效地学习和利用重要的特征信息。利用通道注意力机制和空间注意力机制,该模块能够自适应地对不同通道和空间位置的特征进行加权,提升了模型的表达能力和性能。

CBAM 由一维通道注意力模块 $F_c \in \mathbf{R}^{C \times 1 \times 1}$ 和二维空间注意力模块 $F_s \in \mathbf{R}^{1 \times H \times W}$ 组成,它们按顺序排列,其公式如下:

$$F' = M_c(F) \otimes F \quad (1)$$

$$F'' = M_s(F') \otimes F' \quad (2)$$

1.2.1 通道注意力

通道注意力模块关注于特征图的通道维度,通过对每个通道的重要性进行评分,得到每个通道的权重。采用全局平均池化(Global Average Pooling)和全局最大池化(Global Max Pooling)两种方式,分别得到每个通道的平均值和最大值。然后将这两个值分别送入一个共享的多层感知机(MLP)网络,得到每个通道的权重。最后,将这些权重与原始特征图相乘,得到经过

通道注意力加权后的特征图。

通过全局平均池化和全连接操作,生成一个通道维度上的注意力权重,用于调整不同通道特征的重要性。这有助于模型自适应地选择对于分类或检测任务最有用的通道特征。通道注意力模块可以表示为

$$F_c(f) = \sigma \left(\text{MLP} \left(\text{AvgPool}(f) \right) + \text{MLP} \left(\text{MaxPool}(f) \right) \right) \quad (3)$$

首先,将特征图输入通道注意力中进行池化操作,然后,通过共享网络添加 MLP 输出的特征,再利用 ReLU 功能激活,得到最终的信道注意力特征图。

1.2.2 空间注意力

空间注意力模块则关注于特征图的空间维度,通过对每个位置的重要性进行评分,得到每个位置的权重。采用卷积神经网络的方式,对通道注意力加权后的特征图进行卷积操作,得到每个位置的重要性评分。然后将这些评分进行归一化,得到每个位置的权重。最后,将这些权重与通道注意力加权后的特征图相乘,得到经过空间注意力加权后的特征图。

通过在空间维度上进行特征的最大池化和最小池化操作,捕捉特征图中的全局上下文信息,并生成一个空间维度上的注意力权重。这有助于模型自适应地选择对于分类或检测任务最相关的空间位置。空间注意力模块可以表示为

$$F_s(f') = \sigma \left(\text{Conv} \left[\text{mean}(f'); \text{max}(f') \right] \right) \quad (4)$$

该模块的输入是特征图 f' ,由通道注意力模块输出,我们将特征图沿通道维度进行最大池化,并对降维进行均值池化。然后,将通道数为 1 的两个特征图合成为一个,并在 Conv 操作后得到卷积的注意力特征图。最后,最终的空间注意力特征图 $F_s(f')$ 是 ReLU 功能激活后得到的。

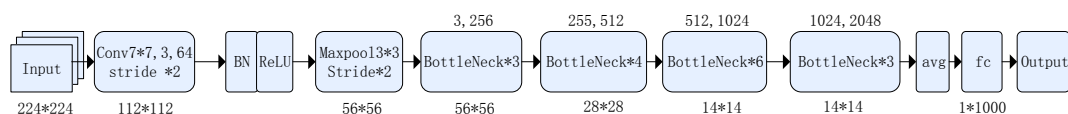


图 3 ResNet50 结构

2 实验结果与分析

2.1 实验环境

该实验在 Linux 系统上进行, 使用 AMD EPYC 7551P 256 GB 64 核 CPU 和 NVIDIA RTX A4000 GPU 并使用 PyTorch1.12 和 Cuda11.3 进行深度模型训练。

2.2 数据集

该实验采用的数据集为 ModelNet, 是普林斯顿 ModelNet 项目的 3D 对象计算机辅助设计 (CAD) 模型集合。该数据集包含 660 个对象类别

和 151128 个模型。从数据集中选择的 40 和 10 个公共类别构成核心数据子集, 分别表示为 ModelNet40 和 ModelNet10, 两个数据集都有方向对齐的版本。ModelNet40 包含飞机、椅子、杯子、吉他等 40 个类别, 不包含颜色信息, 共计 12311 个型号。训练数据与测试数据大约分别占 80% 和 20%。ModelNet 数据集示例如图 4 所示。

实验准确度指数为每个样本的总体准确率 (OA) 和每个类别的总体准确率 (AA), OA 为所有样本中预测正确的百分比, AA 为每个类别的平均准确率。

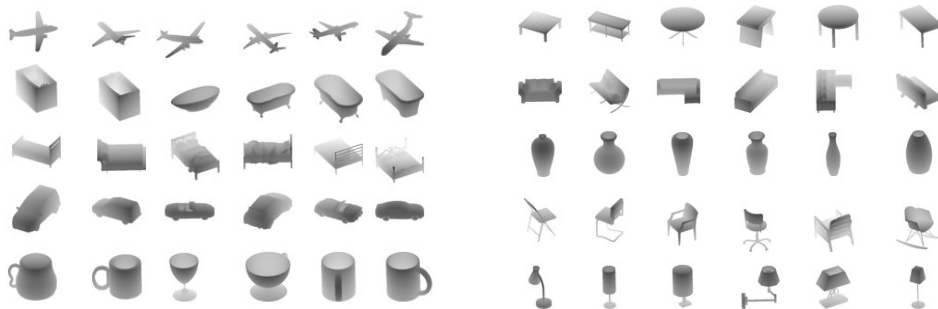


图 4 ModelNet 数据集

2.3 实验分析

在三维模型检索的分类任务中, 特征提取采用了多种特征提取网络 ResNet-50、DenseNet 和 VGG-M 进行实验分析。本文在数据集 ModelNet40 上进行实验研究, 将 CBAM 模块分别连接到上述不同的 CNN 模型中, 对总体准确率 (OA) 和平均准确率 (AA) 进行对比。结果见表 1。ResNet-50 在 OA 和 AA 性能均优于其他网络, 因此, 我们选择 ResNet-50 作为特征提取网络。从表 2 中可以看出本文方法性能有进一步提高。

表 1 不同卷积神经网络分类准确度对比

CNN	OA/%	AA/%
VGG-11+CBAM	93.67	91.92
DenseNet+CBAM	94.82	92.83
ResNet-50+CBAM	95.61	93.24

表 2 不同方法总体准确率 (OA) 对比

方法	视图	ModelNet40/%
MVCNN	12	89.90
GVCNN	12	93.10
SeqViews2SeqLabels	12	93.30
3D2SeqViews	12	93.40
MHBH	6	94.70
Ma et al.	12	91.05
本方法	12	95.61

3 结语

本文提出了一种基于 CNN 和注意力机制的特征提取分类方法, 该方法通过卷积神经网络提取特征, 将通道和空间注意力模块嵌入网络中, 使注意力模块能够有效识别视图中不同区域特征的重要性, 不再同等对待每个视图, 充

分考虑了视图局部特征和每个视图通道和区域之间的相关性。从实验结果可知,在加入注意力机制模块后,与最新方法对比,所用的方法在实验集 ModelNet40 上证明了其在分类准确率上的优越性。并且,通过将注意力模块与不同的特征提取卷积神经网络结合,使用 ResNet50 基本网络模型进行实验的性能更佳。由于本方法在对视图特征聚合时仅采用池化操作,会导致描述性信息的丢失,从而限制了三维模型描述的有效性。在未来的工作中,将重点关注特征聚合方法的研究,以提升三维模型分类及检索的准确性。

参考文献:

- [1] 裴焱栋,顾克江. 基于内容和语义的三维模型检索综述[J]. 计算机应用, 2020, 40(7): 1863-1872.
- [2] 刘安安,李天宝,王晓雯,等. 基于深度学习的三维模型检索算法综述[J]. 数据采集与处理, 2021, 36(1): 1-21.
- [3] 丁博,汤磊,何勇军,等. 基于代表性视图的三维模型检索[J]. 哈尔滨理工大学学报, 2021, 26(6): 18-23.
- [4] SU H, MAJI S, KALOGERAKIS E, et al. Multi-view convolutional neural networks for 3D shape recognition[C]//Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 945-953.
- [5] YU T, MENG J, YUAN J. Multi-view harmonized bilinear network for 3d object recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, 2018: 186-194.
- [6] WANG C, PELILLO M, SIDDIQI K. Dominant set clustering and pooling for multi-view 3D object recognition[EB/OL]. arXiv: 1906.01592, 2019.
- [7] FENG Y, ZHANG Z, ZHAO X, et al. GVCNN: group-view convolutional neural networks for 3D shape recognition [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, 2018: 264-272.
- [8] MA C, GUO Y, YANG J, et al. Learning multi-view representation with LSTM for 3D shape recognition and retrieval [J]. IEEE Transactions on Multimedia, 2018, 21(5): 1169-1182.
- [9] DAI G, XIE J, FANG Y. Siamese CNN-BiLSTM architecture for 3D shape representation learning [C] //Proceedings of the 27th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 2018: 670-676.
- [10] HAN Z, SHANG M, LIU Z, et al. SeqViews2SeqLabels: learning 3D global features via aggregating sequential views by RNN with attention [J]. IEEE Transactions on Image Processing, 2018, 28(2): 658-672.
- [11] JIANG J, BAO D, CHEN Z, et al. MLVCNN: multi-loop-view convolutional neural network for 3D shape retrieval [C] //Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, Hawaii, 2019, 33(1): 8513-8520.
- [12] CHEN S, ZHENG L, ZHANG Y, et al. VERAM: view-enhanced recurrent attention model for 3D shape classification [J]. IEEE Transactions on Visualization and Computer Graphics, 2018, 25(12): 3244-3257.
- [13] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C] //Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 2018: 3-19.

Three-dimensional model feature extraction based on attention mechanism

Wang Huanhuan*, Li Shuqing

(School of Information Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450018, China)

Abstract: Currently, in view-based 3D model retrieval technology, most methods for multi-view feature extraction focus on the global feature information of the view and ignore the exploration of the correlation between the local feature information of the view and the multi-view. To solve this problem, a new feature extraction method is proposed, which uses the convolutional neural network in deep learning and combines the attention mechanism to extract features to improve its discriminability. The method conducts experimental analysis on ModelNet40, takes multiple views of the three-dimensional model as input, and adds an attention module to the network layer for feature extraction and classification. The results show that this method is superior to existing typical algorithms in terms of classification accuracy.

Keywords: 3D model retrieval; CNN; attention mechanism; CBAM

文章编号: 1007-1423(2024)04-0053-05

DOI: 10.3969/j.issn.1007-1423.2024.04.010

LSTM-Prophet 混合模型在物料储备需求预测中的应用

辛 唯*

(青岛科技大学信息科学技术学院, 青岛 266000)

摘要: 建立 LSTM-Prophet 混合模型, 预测 A 公司未来每日物料消耗量, 为工段每日物料储备提供参考, 数据集来源于 A 公司 2015 年来每日物料实际消耗量。实验结果显示, LSTM-Prophet 混合模型预测时序数据的 MAE 值为 40.905, MAPE 值为 0.044, R^2 值为 0.863, 三个评价指标都优于 LSTM 模型、Prophet 模型、ARIMA 模型三个单一模型的评价指标值。验证了 LSTM-Prophet 混合模型有着更好的预测精度, 可以更好地应用于工业场景。

关键词: LSTM 模型; Prophet 模型; 时序数据预测; 混合模型预测

0 引言

随着工业 4.0 与工业互联网的发展, A 公司在数字化转型过程中, 工业生产过程日益复杂, 工业生产规模不断膨胀, 每日都会产生大量工业时序数据。现场物料消耗量是该生产过程中一项重要的时序数据, 对该时序数据的精确预测可以避免出现现场物料储备过多或临时紧缺的问题, 物料储备过多一方面会增加生产现场安全隐患, 另一方面物料长时间暴露在空气中增加了被氧化的风险; 物料临时紧缺则会耽误生产时间, 降低效率, 不利于生产实际。对未来每日物料消耗量这一时序数据进行高精度的预测成为 A 公司迫切的需求。

工业时序数据预测是通过采集工业生产过程历史数据来预测未来生产数据, 它一般用于评估和预测未来生产状态, 辅助调度管理部门制定高效经济的生产计划, 对于管理者合理安排机组出力、保证工段的安全性和稳定性、提高经济效益以及减少不必要的能源消耗有着重要的意义。与此同时, 随着工业生产的智能化发展, 对高精度的原料需求预测也越来越迫切, 高精度的时序数据预测不仅能带来生产线产业化升级, 减少劳动力, 还能提升生产效率。传统的时序数据预测通常是使用基于统计模型的

方法, 例如, 1926 年 Yule 对自回归过程进行了初步研究, 提出自回归模型 AR (auto regression model)、Walker 提出自回归滑动平均模型 ARMA (auto-regressive moving average model)^[1-2]、70 年代初 Box 等^[3]提出了著名预测方法——差分自回归滑动平均模型 ARIMA (autoregressive integrated moving average model), 这三者都是基于回归的单元时序数据预测方法。以上三种基于统计的模型要求时序数列是平稳、线性相关的, 对于工业时序数据中出现的非线性部分往往不能起到很好的数据挖掘作用, 进而影响预测效果。于是研究人员提出了基于机器学习的预测算法, 例如基于人工神经网络 ANN (artificial neural network)、基于支持向量机 SVM (support vector machine)、基于遗传算法 GA (genetic algorithm) 的时序数据预测方法。实践表明, 基于机器学习及其衍生算法能够较好地处理复杂时序数据序列, 相对于传统的预测方法, 预测结果更准确。但是随着数据量的扩张, 许多采集到的时序数据具有海量、时空依赖性的特点, 经典的机器学习和统计方法难以处理大量时间序列数据预测^[4]。因此许多研究人员尝试提出基于数据驱动的深度学习方法来预测时序数据, 1982 年 Hopfield^[5]提出一种单层反馈神经网络

收稿日期: 2023-10-03 修稿日期: 2023-10-31

作者简介: *通信作者: 辛唯 (1999—), 男, 山东威海人, 硕士, 研究方向为时序数据预测, E-mail: xw.199996@163.com

Hopfield network, 这是最早的循环神经网络 RNN (recurrent neural network) 的雏形, 1986 年循环神经网络在时间序列预测方面取得了很大进展, 它能够从时间序列数据中的不规则趋势中提取时间信息^[6]。1997 年 Hochreiter 等^[7]首次提出长短期记忆神经网络 LSTM (long short term memory), LSTM 作为 RNN 的变体, 通过引入门控机制, 自主地选择保留或是遗忘哪些信息, 还可以提高模型捕捉时序数据非线性特征的能力, 同时解决了 RNN 训练梯度爆炸和梯度消失问题^[8]。这些方法基于循环神经网络建立了时间序列之间时间依赖关系, 对于处理时序数据非线性部分有着较好的效果。

在工业领域, 时序数据受多项外界因素的影响, 如假日、温度、季节性需求等, 工业时序数据具有短周期性、波动性的特点, 此前专门针对此情况的研究探索还很少, 加之作者对 A 公司工业时序数据预测初步探索时, LSTM 等模型对预测需求的拟合预测结果不够理想。

于是, 本文针对此生产实际提出搭建基于 LSTM 模型和 Prophet 模型的混合模型, 该混合模型结合了两个模型的优点, LSTM 模型学习时序数据长期历史特征, 针对不同时间尺度以及数据非线性关系, Prophet 模型则针对数据线性部分以及节假日等变化波动点, 最终通过方差倒数法赋予两个模型预测权重, 将 LSTM 模型预测值和 Prophet 预测值通过权值相加得到最终预测结果。结果证明, 该混合模型提高了预测结果的准确性和鲁棒性, 更加适用于复杂的工业生产实际。

1 模型介绍

1.1 Prophet 模型

Prophet 模型是 Facebook 团队开发的一款基于 Python 和 R 语言的时序数据模型, 此模型根据数据集的特点, 会自动选择运用加法或乘法模型得到最终预测值, 可以实现时序数据曲线的快速拟合^[9]。相较于传统的时序数据预测模型 ARIMA, Prophet 集建模和评估两大板块, 使得模型在对时序数据预测上有更好的表现和更高的灵活性, 不仅能引入不同持续时间长短的周期性趋势, 还能自动处理时序数据集中的异

常值。Prophet 模型流程图如图 1 所示。

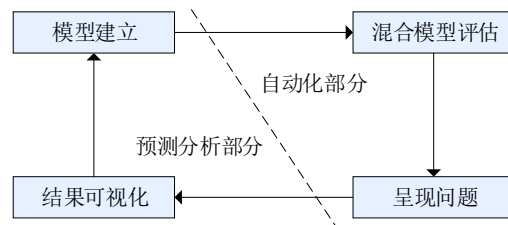


图 1 Prophet 模型流程图

1.2 LSTM 神经网络模型

LSTM 是循环神经网络 (RNN) 的一种特殊形式, 可以解决 RNN 模型存在的梯度消失问题。同时 LSTM 具有增强了循环神经网络的记忆能力, 可以更有效地处理长距离时序性数据。LSTM 网络结构如图 2 所示。

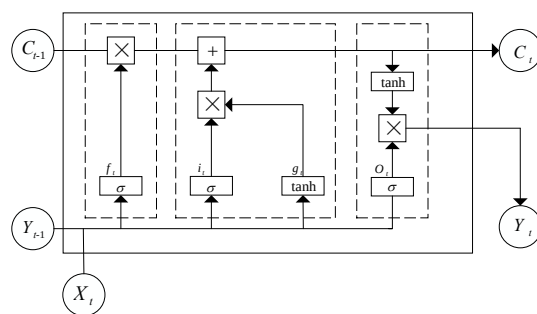


图 2 LSTM 网络结构

在 LSTM 神经网络中, t 时刻的输入由当前时刻输入 X_t 、上一个 LSTM 单元的输出 Y_{t-1} 、上一时刻 LSTM 单元状态 C_{t-1} 构成。 t 时刻的输出由当前时刻输出 Y_t 、LSTM 单元状态 C_t 构成。

2 LSTM-Prophet 混合模型建模

LSTM-Prophet 混合模型建模过程如下: 首先将 A 公司物料消耗量数据进行预处理工作, 确保排除明显传感器误差以及数据时序数据格式呈现, 将处理后的时序数据集分为训练集和测试集。其次根据训练需要设置 LSTM 模型参数后, 使用训练集训练模型, 得到预测结果 Y_L , 再使用训练集训练 Prophet 模型, 得到预测结果 Y_P 。最后对预测结果 Y_L 、 Y_P 赋予权值 α_1 、 α_2 , α_1 、 α_2 由方差倒数法确定^[10], 则混合模型最终

预测结果为 $Y = \alpha_1 Y_L + \alpha_2 Y_P$ 。混合模型构建流程如图3所示。

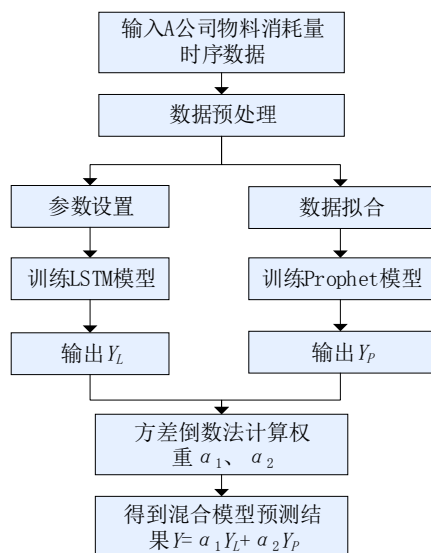


图3 LSTM-Prophet混合模型构建流程图

采用平均绝对误差 MAE (mean absolute error)、平均绝对误差百分比 MAPE (mean absolute percentage error)、决定系数 R^2 对预测结果进行评价，三个指标的数学表达式如下：

$$y_{MAE} = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| \quad (1)$$

$$y_{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|f_i - y_i|}{y_i} \quad (2)$$

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (3)$$

式中， y_i 是实际值， f_i 是预测值， $\bar{y}$ 是实际值的平均值， n 为数据个数。

3 模型应用与分析

3.1 LSTM模型预测效果

根据 A 公司日物料消耗量时序数据特征，通过网格寻优的方式，确定 LSTM 模型中最优超参数，部分参数如下：time-step=72、hidden-size=64、num-layer=1、features=hour_day_week_holiday_load。使用构建的 LSTM 神经网络对数据集进行预测，预测结果见图4。因数据量太大导致数据标注重叠、图像模糊，特放大部分预测结果，

放大部分为数据编号 2000~3000，LSTM 模型预测结果截选放大，如图5所示。

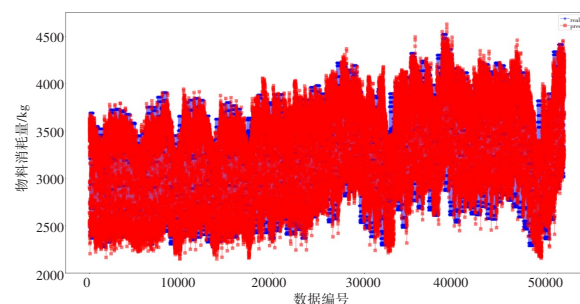


图4 LSTM模型预测效果

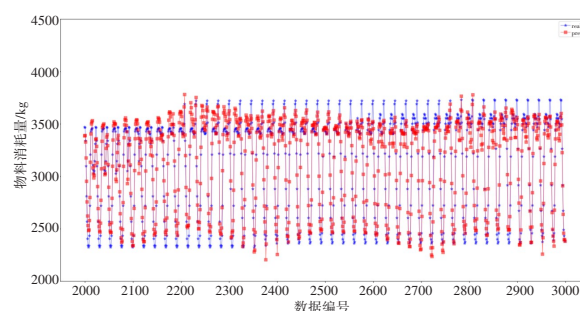


图5 LSTM模型预测结果截选放大

3.2 Prophet模型预测效果

根据 A 公司日物料消耗量时序数据特征，确定部分 Prophet 模型参数如下：growth=linear、weekly_seasonality=True、daily_seasonality=True，参数分别表示时序数据增长趋势为线性趋势、季节趋势以周和天为周期，且设定季节趋势的模式为乘法模式。参数设置后，使用 Prophet 模型对训练集数据进行拟合，拟合结果如图6所示。

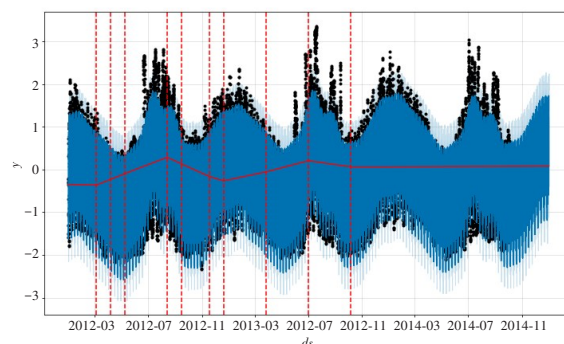


图6 Prophet模型拟合结果

黑色表示原始的时间序列离散点；深蓝色

的线表示使用时间序列来拟合所得到的取值；浅蓝色的线表示时间序列的置信区间，即合理的上下界；红色的折线表示训练集线性趋势曲线图。

使用训练好的 Prophet 模型对测试集进行数据预测，得到 Prophet 模型下的预测值和真实值，结果见图 7。

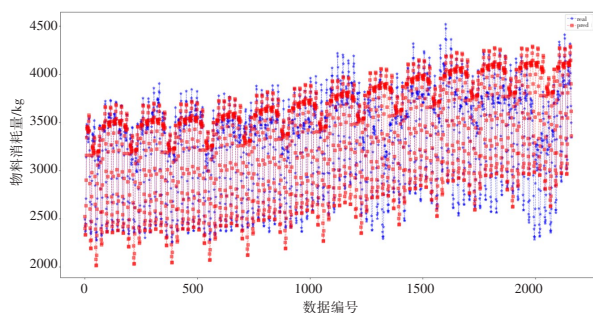


图 7 Prophet模型预测结果

从图 7 可以看出，Prophet 模型对测试集的预测效果较 LSTM 模型稍好，但仍存在个别数据预测值偏小的情况。

3.3 LSTM-Prophet 混合模型权重

根据两种单一模型得到的测试集的预测值的分析，利用方差倒数法计算得到 LSTM 模型和 Prophet 模型在 LSTM-Prophet 混合模型中权重分别为 $\alpha_1=0.11$ ， $\alpha_2=0.89$ 。

3.4 LSTM-Prophet 混合模型结果分析

通过赋予两个单一模型权重得到的 LSTM-Prophet 混合模型并进行训练，使用混合模型对测试集进行时序数据预测，预测图像结果如图 8 所示。采用 MAE、MAPE、 R^2 作为评价指标，并与其他模型对此测试集预测后的指标进行对比，不同模型预测效果比较见表 1。

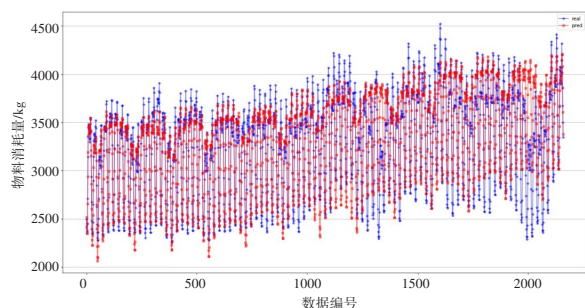


图 8 LSTM-Prophet混合模型预测效果

表 1 不同模型预测效果比较

模型	MAE	MAPE	R^2
LSTM	68.511	0.053	0.784
Prophet	51.932	0.050	0.828
ARIMA	85.433	0.062	0.719
LSTM-Prophet	40.905	0.044	0.863

由图 8 可以看出，LSTM-Prophet 模型对测试集的预测结果较单一模型更加收敛，预测效果明显优于单一模型。由表 1 可以看出，LSTM-Prophet 混合模型的 MAE、MAPE、 R^2 三项评价指标均优于 LSTM 模型、Prophet 模型和 ARIMA 模型。综合可知，LSTM-Prophet 混合模型对 A 公司日物料消耗量的预测精度优于其他单一模型。

4 结语

对比分析 LSTM-Prophet 混合模型、LSTM 模型、Prophet 模型、ARIMA 模型对 A 公司 2020 年以来日物料消耗量时序数据的预测结果，得出以下结论：

(1) 通过赋予权值的方式，结合 Prophet 模型处理趋势变化点和周期性的优势和 LSTM 模型发掘数据非线性特征的优势，实验证明混合模型的搭建可以得到更为精准的预测结果。

(2) 该方法所用数据集来源于实际工业场景，预测结果越精准越可以更好地指导实际生产，为管理者未来储备物料决策提供技术支持，该方法具有广阔的工业场合应用前景。

(3) 该方法能否推广到工业领域中更多的场景还需要进一步的研究。

参考文献：

- [1] BROERSEN P M. Finite-sample bias propagation in the yule-walker method of autoregressive estimation [J]. IFAC Proceedings Volumes, 2008, 41 (2) : 2744-2749.
- [2] ESWARAPPA C, HOEFER W. Autoregressive (AR) and autoregressive moving average (ARMA) spectral estimation techniques for faster TLM analysis of microwave structures [J]. IEEE Transactions on Microwave Theory and Techniques, 1994, 42 (12) : 2407-2411.

- [3] BOX G E, PIERCE D A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models[J]. Journal of the American Statistical Association, 1970, 65(332): 1509-1526.
- [4] BOX G E, JENKINS G M, REINSEL G C, et al. Time series analysis: forecasting and control [M]. NEW YORK: John Wiley and Sons, 2015.
- [5] HOPFIELD J J. Neural networks and physical systems with emergent collective computational abilities [J]. Proceedings of the National Academy of Sciences, 1982, 79(8): 2554-2558.
- [6] CHENG Y, ZHU H, WU J, et al. Machine health monitoring using adaptive kernel spectral clustering and deep long short-term memory recurrent neural networks [J]. IEEE Transactions on Industrial Informatics, 2018, 15(2): 987-997.
- [7] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural computation, 1997, 9(8): 1735-1780.
- [8] CHREITER S. The vanishing gradient problem during learning recurrent neural nets and problem solutions [J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 1998, 6(2): 107-116.
- [9] 黄心裕, 唐军, 王晓宇. 基于 Prophet 算法的海南近海波浪长时段时序分析与预测 [J]. 海洋学报, 2022, 44(4): 114-121.
- [10] 詹英. 组合预测方法在我国人均 GDP 预测中的应用 [D]. 武汉: 华中师范大学, 2014.

Application of LSTM-PROPHET mixed model in material reserve demand forecasting

Xin Wei*

(College of Information Science and Technology, Qingdao University of Science and Technology, Qingdao 266000, China)

Abstract: The LSTM-Prophet mixed model was established to predict the daily material consumption of Company A and provide reference for the daily material reserve of the workshop. The data set was derived from the actual daily material consumption of Company A since 2015. The experimental results showed that the MAE value, MAPE value and R^2 value of the LSTM-Prophet mixed model were 40.905, 0.044 and 0.863, all of which were better than the LSTM model, Prophet model and ARIMA model. It is verified that the LSTM-Prophet hybrid model has better prediction accuracy and can be better applied to industrial scenarios.

Keywords: LSTM model; Prophet model; time series data prediction; mixed model prediction

文章编号: 1007-1423(2024)04-0058-05

DOI: 10.3969/j.issn.1007-1423.2024.04.011

机械完工尾项管理系统在 FPSO 建造项目的应用

邢同超*, 石振华, 赵西磊, 王 亮, 伍江勇

(海洋石油工程(青岛)有限公司, 青岛 266520)

摘要: 在国际海洋工程建造项目中, 机械完工被广泛采用。作为从施工到调试前的重要过渡阶段, 机械完工是一种工程管理理念, 也是项目精细化管理的体现。施工和检验人员在根据 ITR 执行机械完工检查的过程中会记录大量尾项, 并形成尾项清单。如何快速高效地对完工尾项进行实时跟踪和管理是项目管理人员和现场检验人员关注的重点。基于以上需求, 提出了一种解决方案, 基于 Django 建立了一套机械完工尾项管理系统。在某 FPSO 项目的实际运行效果良好, 其中的应用经验可供其他类似项目参考。

关键词: 机械完工; 尾项管理; Django; Python; FPSO

0 引言

近年来, 在涉外海洋工程项目建造过程中机械完工(mechanical completion)被广泛应用, 并形成了一套完善的管理机制^[1]。机械完工的目的是对海洋工程和船舶的某个系统进行详细的完整性检查, 从而确认该系统符合项目和图纸文件的设计要求^[2], 并为下一步预调试、调试工作的开展打好基础。

机械完工贯穿于项目的设计、建造、检验、调试等各个阶段, 是项目精细化管理的有效体现^[3]。机械完工检查工作开始之前, 需完成系统/子系统的划分, 明确子系统的构成以及每一个位号对应的 ITR (inspection & test reports) 类型。施工和检验人员根据 ITR 执行机械完工检查的过程中会记录大量尾项, 并形成尾项清单, 尾项清单是机械完工检查的重要成果。尾项清单里的每一条尾项会根据优先级划分为不同的类别, 如 A 项、B 项、C 项, 尾项类别是后续安排尾项销项工作的重要依据。如何快速高效地对数量众多的完工尾项进行实时跟踪和管理是一

项十分必要而且有难度的工作, 也是项目管理人员和现场检验人员关注的重点。

1 机械完工尾项管理系统

作为大型复杂的海洋工程项目, 浮式生产储卸装置(FPSO)项目的系统/子系统数量多, 相应的 ITR 数量也非常多。随着机械完工检查工作的开展, 尾项数量可达数万条, 后续的统计跟踪工作量非常大。如果使用 Excel 来进行管理, 数据只能离散地保存在个人电脑上, 无法做到多人同时录入数据, 也不便于数据的查询和共享。为解决以上问题, 本文基于 Django 框架建立了一套机械完工尾项管理系统, 该系统实现了权限管理、机械完工基础数据的上传下载、尾项数据录入、尾项状态批量更新、尾项状态实时统计等功能。

2 系统主要功能设计

通过对实际业务需求分析可知, 尾项管理工作涉及基础数据录入、尾项信息录入、尾项状态更新、尾项统计等。相应地, 系统按功能

收稿日期: 2023-11-14 修稿日期: 2024-01-18

作者简介: *通信作者: 邢同超(1983—), 男, 山东冠县人, 硕士, 从事海洋石油工程建造信息系统开发与运维相关工作, E-mail: xingtch@cooec.com.cn; 石振华(1984—), 男, 山东淄博人, 本科, 从事海洋石油工程质量管理相关工作; 赵西磊(1984—), 男, 山东枣庄人, 本科, 从事海洋石油工程电气仪表安装检验工作; 王亮(1979—), 男, 山西桓曲人, 本科, 从事海洋石油工程机械设备安装检验工作; 伍江勇(1983—), 男, 湖北松滋人, 本科, 从事海洋石油工程电气仪表安装检验工作

可以划分为以下几个模块：权限管理模块、基础数据模块、尾项管理模块、尾项统计模块。

2.1 权限管理模块设计

权限管理是保证数据信息安全的重要手段。在项目实施过程中，需要保证不同部门、不同岗位的人员看到和能够操作的数据是不一样的。例如只有具备权限的人员才能关闭尾项。

2.2 基础数据模块设计

基础数据模块用来实现基础数据的编辑、导入、导出等功能。录入尾项时需要详细描述该尾项是跟哪个对象关联的，如某模块某个子系统下的某个位号的设备或者电缆。基础数据主要包含设计图纸中的各类信息，如：模块号、图纸号、专业、系统号、子系统号、位号/Tag No.、ITR类型等。

2.3 尾项管理模块设计

该模块用来管理机械完工尾项的新增、批量修改、删除、报告导出等。

现场检查完成后，尾项录入人员每天把尾项录入系统。系统需要自动为每一个尾项生成唯一的尾项号，在录入尾项时，系统需要支持模糊匹配功能以减少录入工作量，加快录入速度，同时保证信息的准确性。

在尾项更新时，用户可以根据唯一的尾项号去更新该尾项的状态。实际工作中，每天都有大量尾项需要更新状态，其中多个尾项的关闭人、关闭日期等字段信息可能是相同的，这时候就需要有批量更新的功能让用户可以同时更新多条尾项的状态。这种向量式操作模式，比起一条条的更新，可以大幅提升效率。

2.4 查询统计模块设计

查询统计模块主要用于对已录入的尾项信息进行统计汇总。除了要展示尾项总数、尾项关闭数量、尾项剩余数量、每日新增数量等统计信息，还要能展示每份ITR对应尾项的关闭情况。根据这些统计信息，管理人员可以对后续施工安排做出针对性调整。

3 系统功能实现与展示

基于Python语言开发的Web框架有很多，

其中重量级的代表便是Django，许多网站和应用都是利用Django开发的。Django框架是基于模型、视图、控制器(models views controller, MVC)模式设计的，这种模式非常便于程序后续的扩展和维护。由于Django开发应用时主要是在models.py、templates文件夹里的各类模板和views.py里进行，因此Django的开发模式也被称为MTV(models templates views)模式。Django的MVC流程架构示意图如图1所示。

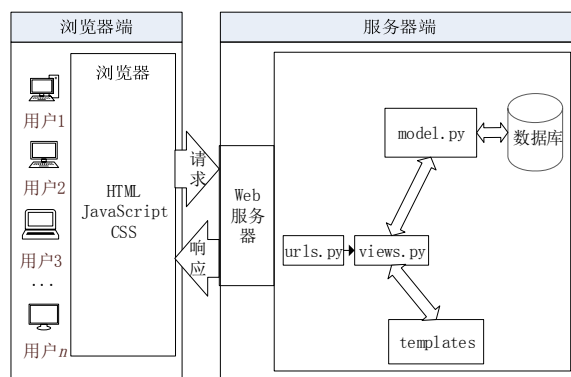


图1 Django的MVC流程架构示意图

本系统前端使用了AdminLTE框架。AdminLTE是一款基于Bootstrap的开源免费的模板主题工具，它提供了一系列响应的、可重复使用的组件，并内置了多个模板页面，同时自适应多种屏幕分辨率，兼容PC和移动端。

系统后端数据库采用开源的MySQL数据库，Django通过内置的MySQL适配器访问底层数据。Django的ORM模型本质上会根据对接的数据库引擎，翻译成对应的SQL语句，所以使用Django开发的项目无需关心程序底层使用的是MySQL、Oracle、SQLite，如果数据库迁移，只需要更换Django的数据库引擎即可。

3.1 权限管理模块实现

Django框架本身带有权限管理，但是功能相对简单，而且不具备较好的通用性和扩展性。因此，本系统专门开发了基于RBAC(role-based access control)模型的权限管理功能。RBAC模型将用户分配到不同的角色，再通过定义角色与权限之间的关系实现用户对系统资源的访问控制和权限管理。RBAC模型具有灵活性、可扩展性和易管理性等优势，被广泛应用于企业级应

用和系统中。

本系统中权限的验证通过 Django 中间件 Middleware 来实现。Django 中间件 Middleware 是介于 request 请求与 response 响应之间的一道处理过程，可以在全局上改变 Django 的输入与输出。

3.2 基础数据模块实现

基础数据管理人员需要先在 Excel 格式的模板中整理好基础数据，然后再上传系统。系统后台会解析 Excel 文件提取数据然后再存到 MySQL 数据库。本系统在解析 Excel 时用到了 Pandas，Pandas 是基于 Python 开发的专门进行数据操作和分析的开源软件库。Pandas 纳入了大量库和一些标准的数据模型，提供了高效地操作大型数据集所需的工具。

3.3 尾项管理模块实现

尾项管理模块是本系统的核心。在 Django 框架中，每个 ORM 模型对应着一个数据库表格。在定义模型时，可以指定每个字段的类型、长度、是否为空等属性，以及设置约束条件。机械完工尾项信息表对应的 ORM 模型中包含的部分核心字段见表 1。

一个尾项单中会包含多个尾项，尾项单号可以根据项目要求设定编号前缀，在第一份尾项单创建完成后，后续的编号采用流水号的方式自增即可。用户可以根据关键字，如尾项单号、模块号等进行筛选，迅速找到所需的尾项单。机械完工尾项单列表如图 2 所示。

表 1 机械完工尾项信息表对应的 ORM 模型中包含的部分核心字段

字段名	字段类型	字段别名 (verbose_name)
project_name	models.CharField	项目名称
location	models.CharField	位置
module_no	models.CharField	模块号
punch_note_no	models.CharField	尾项单号
punch_no	models.CharField	尾项号
tag_no	models.CharField	位号
punch_Chinese	models.TextField	尾项中文描述
punch_English	models.TextField	尾项英文描述
punch_status	models.CharField	尾项状态
system	models.CharField	系统
sub_system	models.CharField	子系统
construction_company	models.CharField	施工单位
punch_type	models.CharField	尾项类别
proposed_by	models.CharField	尾项提出人
closed_by	models.CharField	关闭人
closed_date	models.DateField	关闭日期

3.4 查询统计模块实现

尾项的关闭与否直接影响 ITR 的签署，也会影响子系统机械完工证书的签署。利用本系统，项目管理人员可以轻松获取尾项统计信息，如尾项的关闭数量，前一天的新增尾项数量等。还可以知道每个 ITR 的尾项状态，比如哪些 ITR 的尾项已经全部关闭具备签署条件。系统还支持数据导出功能，用户可以将结果导出为 Excel 表格或 CSV 格式的文件进行后续处理或分享。系统尾项状态汇总数据见图 3，某 ITR 尾项统计数据如图 4 所示。

C清除

+新增报告

删除

导出报告

导入报告

尾项单号:

按检单号:

设备位号:

模块号:

☐	尾项单号	AFI NO按检单编号	模块号	位号/TAG No.	修改人	最后修改时间	检验日期	操作
1	☐ punch-2502	10472	LQ	A-93018	岳璐	2022/11/28 09:52	2022/10/19	详情-编辑 删除
2	☐ punch-2563	无	L20	86NWR-001	岳璐	2022/11/28 09:51	2022/11/28	详情-编辑 删除
3	☐ punch-2562	无	L10	70DGP-07C218	岳璐	2022/11/26 10:56	2022/11/26	详情-编辑 删除
4	☐ punch-2561	无		94JK-601	岳璐	2022/11/25 14:23	2022/11/25	详情-编辑 删除
5	☐ punch-2550	无	L10	99-L10-LVL02	岳璐	2022/11/24 14:16	2022/11/12	详情-编辑 删除
6	☐ punch-2560	无	L20	70CDHM-07A01	岳璐	2022/11/24 14:03	2022/11/24	详情-编辑 删除
7	☐ punch-2559	无	L20	86NAX-001	岳璐	2022/11/23 14:21	2022/11/23	详情-编辑 删除
8	☐ punch-2558	无	P20	70DGP-06C25	岳璐	2022/11/23 14:04	2022/11/22	详情-编辑 删除
9	☐ punch-2404	20	L10	86NPD-001B	岳璐	2022/11/21 15:45	2022/09/26	详情-编辑 删除
10	☐ punch-2556	无	ITG	79CJJ-004AE	岳璐	2022/11/21 14:03	2022/11/19	详情-编辑 删除

图 2 机械完工尾项单列表

xxxx项目-尾项状态汇总				
尾项总数	尾项关闭	尾项剩余	昨日(2022-11-12)新增	昨日(2022-11-12)关闭
25143	24196	947	0	0
A项总数	A项关闭	A项剩余	A项昨日新增	A项昨日关闭
4630	4323	307	0	0
B项总数	B项关闭	B项剩余	B项昨日新增	B项昨日关闭
20513	19873	640	0	0

图 3 尾项状态汇总数据

下载					
	ITR Type	位号/TAG No.	尾项总数量	尾项关闭数量	尾项剩余数量
1	E51A	70JTA-952	10	10	0
2	E51A	87JBF-646	9	9	0
3	E51A	70JTB-834	9	9	0
4	E51A	70JTB-915	9	9	0
5	E51A	70JTB-833	8	8	0
6	E51A	70JTB-914	8	8	0
7	E51A	70JTA-990	8	8	0
8	E51A	70JTB-987	8	8	0
9	E51A	70JTB-934	8	7	1
10	E51A	87JBF-647	7	7	0
11	E51A	70JTA-927	7	7	0
12	E51A	61UY-002	7	7	0
13	E51A	77JTB-976	7	7	0
14	FF5A	76JTC-691	7	6	1

图 4 某 ITR 尾项统计数据表

4 系统应用和效果

机械完工尾项管理系统应用前后对比见表 2。

表 2 机械完工尾项管理系统应用前后对比

工作项	应用前	应用后
权限管理	没有权限管理，能打开 Excel 就能修改数据	完善的基于 RBAC 的权限管理，不同角色具有不同权限
基础数据	离线数据，无法实时共享和更新	在线数据，实时共享和更新
尾项管理	利用 Excel 进行管理，周期性进行汇总，无法实现实时多人录入和更新尾项	尾项数据实时更新，可多人同时操作，大幅提升效率，也解决了以往数据混乱的问题
尾项状态统计	在 Excel 表格中进行手动统计，数据量大，统计工作重复进行，统计结果难以实时共享	状态统计由系统自动完成，多用户可在各自终端实时获取数据

5 结语

机械完工检查过程中每天都会形成大量尾项，这些尾项信息的录入、更新、统计汇总工作需要大量的人力，信息化系统的使用可以让原本复杂纷乱的数据变得清晰而明确，系统提供的统计数据可以让项目管理人员更合理地分配施工人力，更有针对性地安排每天的工作。本文介绍了一套机械完工尾项管理系统，通过在某 FPSO 项目的实际运行，提高了工作效率，提升了项目管理工作的信息化水平，为后续类似项目提供了一定的借鉴和参考。

参考文献：

[1] 王金源,朱传超,高伟,等. 海洋工程建造项目机械完工阶段管理[J]. 项目管理技术, 2020, 18(7): 122-125.

[2] 颜培雷. 机械完工在 CR 公司海洋钻井平台项目管理中的应用研究[D]. 大连:大连海事大学, 2014.

[3] 石振华,朱开斌,杨星宇. 海洋工程项目机械完工的质量控制与管理[J]. 中国石油和化工标准与质量, 2021, 41(23):13-14.

(下转第 82 页)

文章编号: 1007-1423(2024)04-0062-04

DOI: 10.3969/j.issn.1007-1423.2024.04.012

基于数据交换格式标准的通用一致性测试框架

雷海申^{1,2}, 徐 巍^{1,2}, 裴 安^{1,2*}, 董 楠^{1,2}(1. 工业和信息化部电子第五研究所, 广州 511370;
2. 工业软件工程化与应用技术工业和信息化部重点实验室, 广州 511370)

摘要: 基于数据交换格式标准的通用一致性测试框架首先需要研制一致性测试标准, 再由测试开发人员依据一致性测试标准, 完成数据格式标准文本形式到程序可读文件形式的转化及一致性测试系统的设计和开发, 最终以程序可读文件作为基准测试用例集, 以导出软件导出的被测文件为被测件, 通过对结构、语法的分析完成一致性测试。通用测试框架通常不关注具体的标准, 只需依据一致性测试标准提炼通用结构分析和语法分析规则, 维护基准测试用例集, 代码开发量有效降低, 测试人员可以快速完成对测试需求的支撑。

关键词: 数据交换格式标准; 程序可读; 一致性测试

0 引言

数据格式交换标准主要用于软件之间的数据统一和共享, 提升数据模型完整性, 提高信息传递一致性, 对实现产品数据信息安全、可靠、保真、高效共享应用具有重要意义。当前国际应用最广的 STEP(Standard for the Exchange of Product Model Data), 即产品模型数据交换标准, 提供了一种不依赖具体系统的中性机制, 旨在实现产品数据交换和共享, 显著降低了产品生命周期内的信息交换成本, 提高了产品研发效率。国外发达国家已经把 STEP 标准推向了众多工业应用领域。我国工业门类广, 工业场景复杂, 在不久的将来, 也会有适合我国工业现状的标准出现。

数据格式交换标准研究工作重点有两个, 一是标准研制, 依赖众多行业专家群策群力; 二是标准落地实施, 需要企业深度参与开发应用和有效的测试手段保证, 测试手段主要是标

准一致性测试。一致性测试提供了一种用来决定被测实现(implementation under test, IUT)是否忠实地满足标准规定的要求和规范的方法^[1]。数据格式交换标准的数据载体是数据交换文件。生产数据交换文件的系统是导出处理器, 使用数据交换文件的系统是导入处理器, 导出导入既可以是同业务类型软件, 也可以上下游关系, 一致性测试验证二者之间数据交换是否符合一致性要求, 主要分为三个方面, 语法分析、结构分析、语义分析。导出处理器验证语法、结构、语义, 导入处理器验证语义, 而语义分析是验证信息是否被完整保留, 可以通过人工直观验证数据模型即可, 不建议花费大量时间在自动化中验证。邓镐胜等^[2]依据 STEP 定义的一致性测试标准, 提出了一种基于 STEP 标准的一致性测试方法, 在语法、结构、语义方面实现测试需求, 但是这种方法的核心技术严重依赖 STEP 应用开发框架 ROSE 和限制于 EXPRESS 语

收稿日期: 2023-12-27 修稿日期: 2024-01-12

基金项目: 广州市基础与应用基础研究项目基金(202201010776)

作者简介: 雷海申(1990—), 男, 河南周口人, 硕士, 助理工程师, 研究方向为工业软件测试技术; 徐巍(1991—), 男, 江西九江人, 硕士, 工程师, 研究方向为工业软件测试技术; *通信作者: 裴安(1986—), 男, 辽宁丹东人, 博士, 工程师, 研究方向为工业软件可靠性技术、适配验证, E-mail: xian@ceprei.com; 董楠(1994—), 男, 河南新乡人, 硕士, 助理工程师, 研究方向为软件可靠性

言,无法快速复用到其他数据交换格式标准中。

本文提出一种不局限于数据交换格式标准及编程语言的通用一致性测试框架,测试开发人员只需专注于研究一致性测试标准、基准测试用例设计、结构分析和语法分析的代码逻辑实现和替换,可以快速实现对测试需求的支撑。

1 通用一致性测试框架

基于数据交换格式标准的通用一致性测试框架如图1所示。首先需要研制一致性测试标准,再由测试人员依据一致性测试标准,实现数据格式标准文本形式到程序可读文件形式的转化,再完成一致性测试系统的设计和开发,最终以程序可读文件作为基准测试用例集,以导出软件导出的被测文件为被测件,通过对结构、语法的分析完成一致性测试。

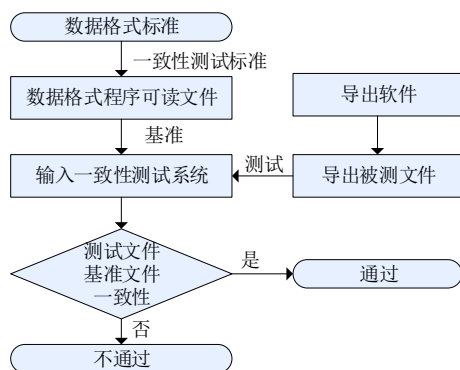


图1 测试框架

数据格式标准的通用一致性测试方法有三个关键点:①一致性测试标准的研制;②如何将文本形式标准转换为程序可读形式标准;③一致性测试系统的设计和开发。

1.1 一致性测试标准研制

一致性测试标准作为数据格式标准的一部分,适用于软件厂商、测评机构、用户及有关单位开展数据格式的一致性测试。

一致性测试标准描述数据格式一致性的内涵、测试对象、测试方法、测试内容、判定准则、测试流程等一般要求,适用于对格式标准的程度开展一致性测试,确保测试的客观性、可重现性和可追溯性。

一致性含义明确一致性范围、一致性要求、

一致性类等重要事项。一致性要求分为必选要求、条件必选要求和可选要求。不同的一致性要求组合成一致性类,一致性实现必须支持所声明一致性类中的全部内容,若一个实现只支持了所声明内容的子集,则不能被称为一致性实现。

测试对象包含导出处理器与导入处理器两个软件单元,测试方法定义IUT作为导出处理器或导入处理器被测试时,IUT具体化为导出处理器或导入处理器分别独立进行测试,及对应的测试流程。测试内容定义对IUT执行的测试分析有:语法分析、结构分析与语义分析,并分别规定导出处理器测试内容和导入处理器测试内容。

一致性测试流程分为五个阶段,包括:①测试需求分析;②测试方案编制;③测试设计;④测试执行和结果分析;⑤测试报告编制。

1.2 数据格式标准程序可读化

依据一致性测试标准,将数据格式标准的语法分析要求、结构分析要求可视化,程序可读化。文本形式的数据格式标准不可被程序解析,程序可读化的数据格式标准可以方便地被程序解析,作为测试分析预期输出。

当前数据文件的主流格式是XML、JSON等。XML是一种结构化的标记语言,结构性强,数据简单而精确,它可以提供一种易被程序读取的与软硬件平台无关的基于文本格式的数据表示方法,它的这些优点表明XML很适合用来描述测试需求数据^[3]。同时XML对于复杂软件结构的数据更容易表示。也可以使用XSD,即XML Schema Definition来表示数据格式标准。JSON(JavaScript object notation)是一种轻量级的数据交换格式。它由Douglas Crockford在2001年首次提出,并逐渐成为一种广泛应用的数据格式。JSON采用简洁的文本形式表示结构化数据,可读性较高,易于理解和生成。程序可读化需要和导出软件及导入软件使用的格式一致,随着大数据分析等技术的发展,许多通用数据分析开源组件开放使用,计算机可以很便捷地识别XML、JSON等格式的结构、语法。

依据一致性测试标准,通过对文本形式标准的分析,提炼出一致性测试标准中规定的关

键要素,形成程序可读化的文件,即基准测试用例集。实现方式通过在广泛使用的生成工具中填入关键信息,生成程序可读化文件。对于庞大的数据格式标准,需要根据业务领域的不同,完成对业务数据分模块分析,数据依赖通过引用方式实现。

1.3 一致性测试系统设计

一致性测试系统的架构如图2所示。系统可以采用BS架构,view层用于接收用户的请求和返回用户数据,model层完成一致性测试数据的分析、过滤、计算等逻辑,最终返回view层分析数据,完成HTML格式的转换后通过view层返回浏览器解释执行,用户可以看到一致性测试结果展示。

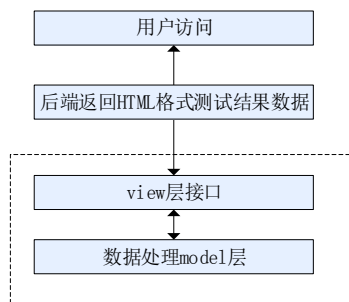


图2 系统架构

一致性测试分析在model层完成,流程如图3所示。有以下几个关键点:

(1) 读取导出软件生成的数据文件和基准程序可读化标准文件,分别生成树状数据结构并存入数据对象;

(2) 定义通用化的结构分析和语法分析判断逻辑,结构分析主要包含以下几类,属性的数据类型、数据长度、是否必需、属性约束是否符合一致性要求;

(3) 基准程序可读化标准的数据对象为基准,遍历测试数据对象,判断一致性要求是否符合,符合继续遍历,不符合则将不符合位置、原因存入resultData对象,直到全部遍历通过为止;

(4) 对resultData对象的数据进行重复数据的过滤,数据统计分析、记录日志等;

(5) resultData对象的数据转换为HTML格

式,返回view层。

上述是一致性测试分析技术要点,基准测试用例集需要根据不同标准的一致性要求进行设计,通用化的结构分析和语法分析判断逻辑需要适配不同的文件格式,框架的其他部分则可以复用。

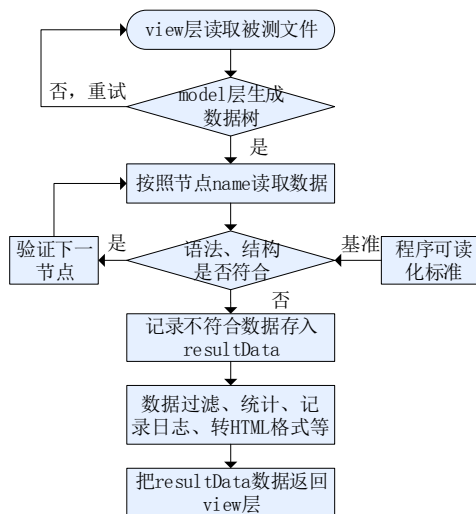


图3 一致性测试分析

2 实例

实例采用BS架构开发一致性测试系统,导出处理器和导入处理器使用的数据文件格式XML,因此本实例使用XSD格式承载数据格式标准程序化后的文件。

依据一致性测试的要求,将数据格式标准转化为XSD结构,片段如图4所示。

```
<xs:element name="entProduct" type="pasm:entProduct" substitutionGroup="pasm:entRoot" nillable="true"/>
<xs:complexType name="entProduct">
  <xs:complexContent>
    <xs:extension base="pasm:entRoot">
      <xs:sequence>
        <xs:element name="ProductType">
          <xs:simpleType>
            <xs:restriction base="xs:int">
              <xs:enumeration value="0"/>
              <xs:enumeration value="1"/>
            </xs:restriction>
          </xs:simpleType>
        </xs:element>
        <xs:element name="ProductObject" type="xs:IDREF"/>
      </xs:sequence>
    </xs:extension>
  </xs:complexContent>
</xs:complexType>
```

图4 XSD片段

导出处理器导出的XML测试文件如图5所示。

```
<entProduct ID="i1" Name="Product0">
  <ProductType>10</ProductType>
  <ProductObject>i2</ProductObject>
</entProduct>
```

图5 XML片段

根据XSD文件表示的一致性要求，entProduct属性的ProductType的value只能是0或1。当前ProductType的value是10，显然不符合一致性要求，运行软件执行测试后得到图6所示结果。

不符合规范的结果如下：共有 1 条原因
1. 路径:/CAD_3D_XML/entProduct/ProductType, 不通过原因(value must be one of {0, 1}), 检测位置ID="i1" Name="Product0"

图6 测试结果

3 结语

本文提出一种基于数据交换格式标准的通

用一致性测试框架，相比于将数据格式标准当作需求逐条检测的方式，本框架的优点是，测试框架通用，不用关心具体的标准，只需依据一致性测试要求提炼通用结构分析和语法分析规则，维护基准测试用例集，代码开发量有效降低，测试人员可以快速完成对测试需求的支撑。

参考文献

- [1] ISO 10303-31: conformance testing methodology and framework: general concepts[S]. 1994.
- [2] 邓镐胜, 杨占勇, 邓家祺, 等. STEP/AP203 一致性测试软件的实现[J]. 计算机工程与应用, 2002, 38(6): 123-126.
- [3] 王学奇, 肖明清, 陈希林, 等. 基于XML的测试需求描述及其实现[J]. 计算机工程与应用, 2005, 41(23): 112-114.

Universal conformance testing framework based on data exchange format standards

Lei Haishen^{1,2}, Xu Wei^{1,2}, Xi An^{1,2*}, Dong Nan^{1,2}

(1. The Fifth Electronic Research Institute of Ministry of Industry and Information Technology, Guangzhou 511370, China;

2. Ministry of Industry and Information Technology Key Laboratory of Industrial Software Engineering Application Technology, Guangzhou 511370, China)

Abstract: A universal conformance testing framework based on data exchange format standards. First, a conformance testing standard needs to be developed. Then, test developers can convert the data format standard from text form to program-readable file form and design and develop the conformance testing system according to the conformance testing standard. Finally, the program-readable file is used as the benchmark test case set, and the tested file exported by the software under test is used as the tested object. The conformance testing is completed through structural and syntactic analysis. The testing framework proposed in this paper is universal and does not care about specific standards. It only needs to extract universal structural analysis and syntactic analysis rules based on the conformance testing standard, maintain the benchmark test case set, and effectively reduce the amount of code development. Testers can quickly support test requirements.

Keywords: data exchange format standards; program readable; conformance testing

开发案例

文章编号: 1007-1423(2024)04-0066-09

DOI: 10.3969/j.issn.1007-1423.2024.04.013

基于学习体验的中医药文化数字展示系统开发研究

祝敏娇*

(苏州信息职业技术学院大数据与互联网系, 苏州 215200)

摘要: 中医药文化历史悠久、内容丰富、内涵深刻, 为促进中医药文化在数字时代的传承创新发展, 从学习体验角度探索中医药文化数字展示系统的设计开发。分析数字展示的多种形式, 解读学习体验的内涵, 搭建数字展示系统设计框架, 运用多种平台、工具, 如 3Ds Max 软件、720 云平台、万维引擎和 Unity 3D 软件等, 完成中医药文化数字展示系统的数字资源建设和交互设置。系统具备 AR 导览、VR 漫游和交互媒体三大模块, 集三维模型、二维图像、文本、音视频等数字资源于一体, 交互方式丰富。基于学习体验开发系统, 系统用户即中医药文化学习者, 中医药文化自然而然地植入学习者的心中, 从而得到广泛的传播和发扬。

关键词: 学习体验; 中医药文化; 数字展示; 平台; 工具

0 引言

习近平总书记在二十大报告中提出: 促进中医药传承创新发展。中医药文化发展历史悠久, 是中华优秀传统文化的重要组成部分, 包含了丰富的哲学思想和文化内涵, 而其深刻内涵对传播形成了一定程度的阻力^[1]。随着数字化技术的不断发展, 文化的数字化传承保护受到重视, 中医药文化的数字化发展势在必行, 对突破中医药文化传播阻力起到了积极的作用。在分析成熟数字展示案例的基础之上, 以学习体验为视角, 设计并开发中医药文化数字展示系统, 推动中医药文化的传承与发展。

1 数字展示

2021 年 3 月《中华人民共和国国民经济和社会发展的第十四个五年规划和 2035 年远景目标纲要》提出以数字化转型整体驱动生活方式变革, 催生新产业、新业态、新模式, 壮大经济发展新引擎。2022 年 11 月工业和信息化部、教育部等五部联合发布《虚拟现实与行业应用融合发展行动计划(2022—2026 年)》, 提到要推动

发展数字化体验, 让优秀文化借助虚拟现实技术“活起来”。

数字化将深刻改变人类生产生活方式。数字展示是采集展品相应实物资料, 以数字形式处理并再现实物的相关信息^[2]。在文化传播领域, 通过增强现实应用、虚拟漫游、交互媒体展示、体感交互、全息投影等技术途径将优秀文化数字化、虚拟化、智慧化, 选择合适的数字技术对文化进行数字展示, 达到高效传播的效果。

1.1 增强现实展示

增强现实(augmented reality), 简称 AR 技术, 是将虚拟图像元素投射到现实情境, 实现虚拟与现实同步和融合的一类创新交互技术^[3]。AR 技术在文化、教育等多领域应用广泛, AR 导览巧妙地把虚拟讲解和真实环境融合在一起, 呈现出生动有趣的视觉导览效果。如苏州博物馆西馆的 AR 导览服务(如图 1 所示), 游客戴上 AR 眼镜, 对准文物即可智能识别、复原远古的文物, 讲述展品背后的故事、“修复”破损文物等。

收稿日期: 2023-10-12 修稿日期: 2023-11-10

作者简介: *通信作者: 祝敏娇(1982—), 女, 江苏宜兴人, 硕士, 副教授, 研究方向为视觉传达、三维建模、虚拟现实, E-mail: 83754585@qq.com



图1 苏州博物馆AR导览

1.2 虚拟全景展示

全景漫游技术是以2D图像为基础的虚拟现实技术,利用全景图片,使用漫游系统制作工具构造出全景空间,可用鼠标或者移动键在网页端拖动观看,也可用相关设备进行环视观看,具有十分强烈的现实感和交互感^[4]。当下,各大展会或展馆均已提供线上办展、观展服务,一种情况是在现有实际场馆的基础上,通过全景摄影方式制作数字资源,利用漫游全景工具设置图片、音频等交互热点,最终发布虚拟展示作品,如秦始皇兵马俑博物馆“全景兵马俑”展示(如图2所示);另一种是没有实际场馆的情况下,在数字资源环节,通过三维软件制作虚拟场景,导出模型或全景图,如北京故宫博物院推出的“V故宫”,将未对外开放的“养心殿”“倦勤斋”“灵沼轩”三座宫殿设计虚拟展示系统,高效而直观地展示宫殿修缮复原或预期设计效果,传播文化和理念。图3为灵沼轩的虚拟修复场景。本研究的VR模块主要讨论第二种展示设计,用户在虚拟场景漫游的过程中通过热点触发丰富的数字资源。

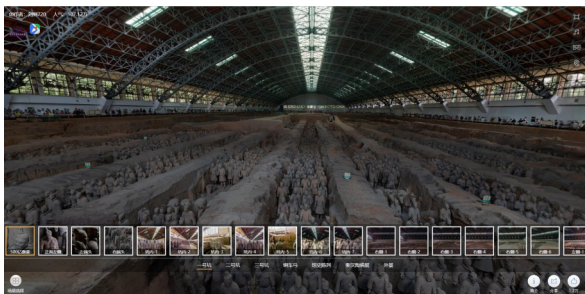


图2 全景兵马俑



图3 故宫“灵沼轩”虚拟修复场景

1.3 交互媒体展示

随着网络高速发展、智能手机的兴起,传统的纸质、电视、广播等媒体完成了传播信息的任务,但受众只能单向、被动地接受信息,无法进行双向性的交流沟通,在此背景下交互式媒体应运而生,其应用场景有交互式教学平台、学习视频和课件书籍等。互动媒体是有交互功能的多媒体形式,人们自主性操作多媒体内的元件或元素时,就实现了媒体互动^[5]。如苏州博物馆西馆设置的交互媒体展示台(如图4所示),突破了传统展示的封闭性,具有开放性、发散性,用户与之互动体验,提高了参与积极性,加深了对相关文化知识的理解。



图4 苏州博物馆交互媒体展示

2 学习体验

2.1 学习体验的内涵

荷兰学者Floor提出学习体验是创造体验的科学和艺术,通过以用户为中心和目标导向的方法,帮助学习者实现期望学习成果^[6]。近年国内学者也在不同方面对学习体验进行了研究,如谢翌等^[7]研究了基于学习体验的过程性课程评价,将学习体验的内在结构细化为“一个载体:事件,两个维度:认知与情感,四种类型:

学习任务体验、学习活动体验、学习情境体验、学习结果体验”。张露等^[8]基于学习体验视角研究游戏化学习理论，归纳总结学习体验的三种类型：个人基于情境的认知体验、个人基于动机的主体性体验和群体基于协作的社会性体验。余明华等^[9]分析总结学习体验首先是一种学习过程，包括学习者在学习环境中外在亲历学习活动及其内在所形成的心理活动；其次是一种学习结果，包括认知和情感反应。

本研究概括出注重学习体验要做到：以学习者为中心，分析其特点，了解其需求，结合学习主题确定目标，进行认知和情感方面的学习内容架构，设计科学而丰富的交互学习过程，学习有成效。学习体验的关注要素概括为：学习者、目标、内容、过程和成效。

2.2 基于学习体验的数字展示系统设计框架

以 Jesse James Garrett 的用户体验模型为参考，融入学习体验的关键要素，提出基于学习体验的数字展示系统的设计框架，如图 5 所示^[10]。其中，学习体验模型从抽象到具体分为策略层、模块层、交互层和表现层：策略层根据对学习者的特征和需求分析，研究学习主题，确定设计目标；模块层把学习者需求和设计目标转化为系统内容，考虑学习者的认知和情感体验，对内容进行模块化架构；交互层响应学习者的需求，使系统与学习者进行多维度互动，如立体空间维度、二维界面维度和交互操作维

度，体现了学习体验的过程元素；表现层将前期的数据信息、交互方式汇聚由感知设计产生最终系统产品，是学习者直接关注到并产生学习成效的层级。从目标确定到内容架构、交互设计和感知设计是数字展示系统从概念生成到完成实现的主要流程，即基于学习体验的系统设计框架。

3 中医药文化数字展示系统设计

基于学习体验开发中医药文化数字展示系统，搭建设计框架如图 6 所示。

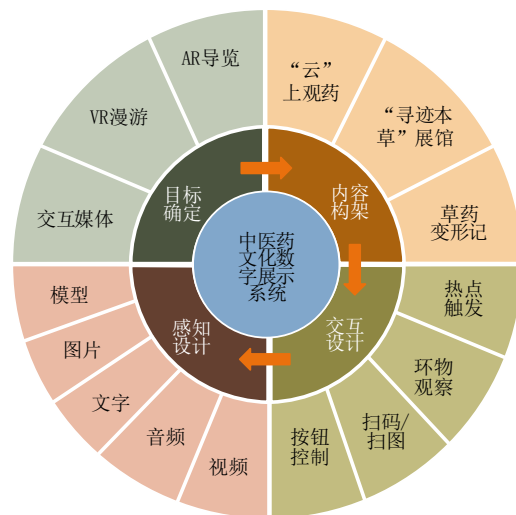


图 6 中医药文化数字展示系统的设计框架

确定用 VR/AR 形式载体展示中医药文化的系统目标；从用户的认知和情感角度出发将系

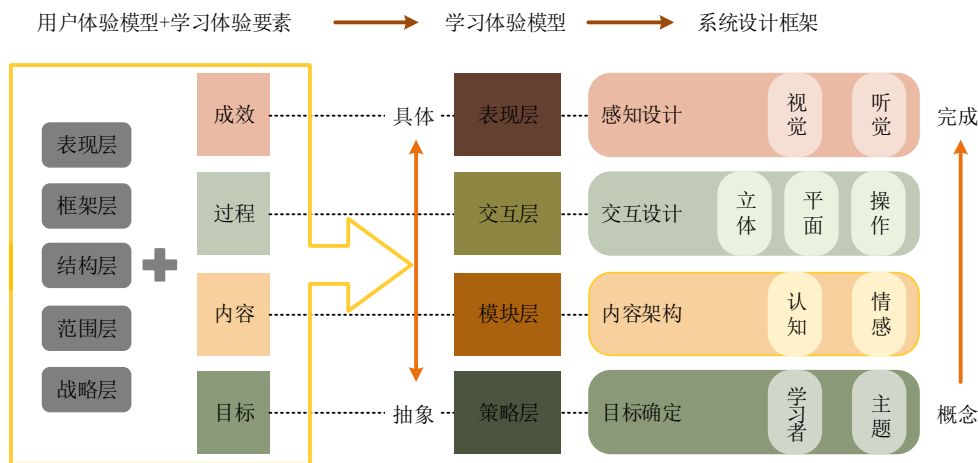


图 5 基于学习体验的数字展示系统的设计框架生成

统内容架构为三大模块；对立体、平面等数据执行多种不同交互操作方法；整个系统的数字资源最终以模型、图片、文字、音视频等形式被用户感知。系统设计旨在将我国优秀的中医药传统文化自然而然、潜移默化地植入学习者的心中，使之得到广泛的传播和发扬。

3.1 目标确定

中医药文化植根于中华传统文化之中，其历史悠久、内容丰富。近年来随着中华传统文化普及活动的开展，大众对中医药文化也有了粗略的了解且兴趣逐渐浓厚；在直接受众群体中，中医药得到了充分的肯定，但因内涵十分深刻，其广泛、积极的传播受到了一定阻碍。伴随数字媒体的蓬勃发展，人们需要有更加生动和实践性更强的信息接收途径。

针对以上需求，查询搜集多方面权威历史资料，采用AR导览、VR漫游和交互媒体的形式传承中医药相关的文化、精神和理念，如图7所示。学习者可体验AR交互或置身于“寻迹本草”展馆中身临其境地参观学习，也可采用交互媒体形式，收获全面而丰富的信息。

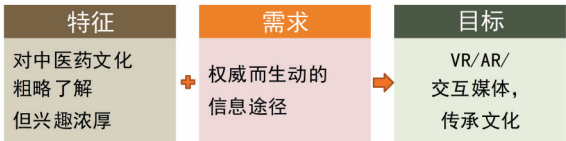


图 7 系统设计目标

3.2 内容架构

本系统内容由三个模块架构而成，分别是“云上观药”、“寻迹本草”展馆、“草药变形记”，如图8所示。在认知上对中草药药材及对应植物有基本的识别，了解中医药学多方面的知识和药材的加工制作过程；情感上，用户经历虚实结合或身临其境的体验，对中医药文化加深了理解，从而对相关事物产生主观上的情感体会并生成有独特意义的心理体验。

“云上观药”模块用AR导览折页的形式展示中草药，介绍每一味中草药的简介、用法、功效和搭配等。

“寻迹本草”展馆模块用VR场景漫游的形式，介绍内容涉及中医名著、古代中医和中药材的鉴别、生长环境等相关知识等。

“草药变形记”模块将中医药材的制作过程用富有趣味的对话小场景表达，设置简单交互，生成数字绘本软件。

3.3 交互设计

AR导览形式下用户通过手机微信扫描二维码再扫描锚点图，锚点图相当于超级链接，扫描锚点图可迅速链接更多丰富的资源，如图9所示。对于三维模型，在手机屏幕上用手指触摸拖动可以环绕观察；选择不同选项卡则可滑动浏览多幅图片、查看文本内容、观赏动画视频和聆听介绍音频或背景音乐。

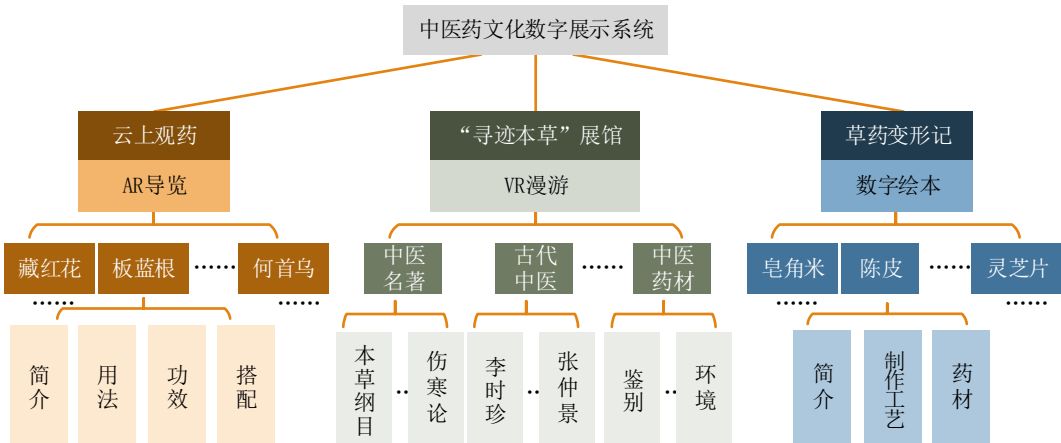


图 8 系统内容架构



图9 二维码和锚点图

VR漫游形式的交互,可佩戴VR设备进行体验,如性价比较高的头戴式VR纸盒眼镜,用户可以身临其境地游览于不同的虚拟场景,通过热点激活新功能从而更全面地获取数字信息。

数字绘本软件主要通过简洁易操作的按钮控制来实现翻页交互效果,形式简约,重点体现绘本内容内涵的表达,较传统的图片播放更具主动性,又避免过于繁杂的交互体验干扰用户的注意力而喧宾夺主。

3.4 感知设计

AR和VR效果最终集成于二维码或网页链接,用户可由此查看丰富的视听数字资源,如三维模型、介绍图片、文本、音频视频等。为了进一步体现系统的完整性、增加可推广性,将二维码及锚点图设计排版于其他媒介上进行

展示、传播,如折页、卡片、海报等。数字绘本软件呈现了美观的画面和动听的背景音乐,让用户拥有美好的使用感受。

4 中医药文化数字展示系统开发实践

中医药文化数字展示系统的不同模块使用不同的开发工具,ARVR基于在线开发平台,交互媒体基于UNITY 3D引擎软件。首先通过3Ds Max、PhotoShop、Animate等软件制作各类数字资源,再将这些资源导入相应的平台,设置交互进行管理。

4.1 三维建模

本系统主要有三维人、物和场景模型资源,建模流程如图10所示。其中人、物模型包含静态的书籍、植物、人物模型以及带有骨骼动画的人物模型,采用低精度建模加手绘材质贴图方式确保交互实时渲染的流畅性。

使用3Ds Max软件用多边形建模方法完成低模的制作,将多边形处理为三角形或四边形,若模型面数多、存在多边形,很可能会导致模型导入平台失败、平台运行卡顿等情况;若人物模型需要继续制作骨骼绑定动作,最好均为四边形结构且关节处的布线需适当增加。然后对模型进行UV展开,生成合理布局的UV线框

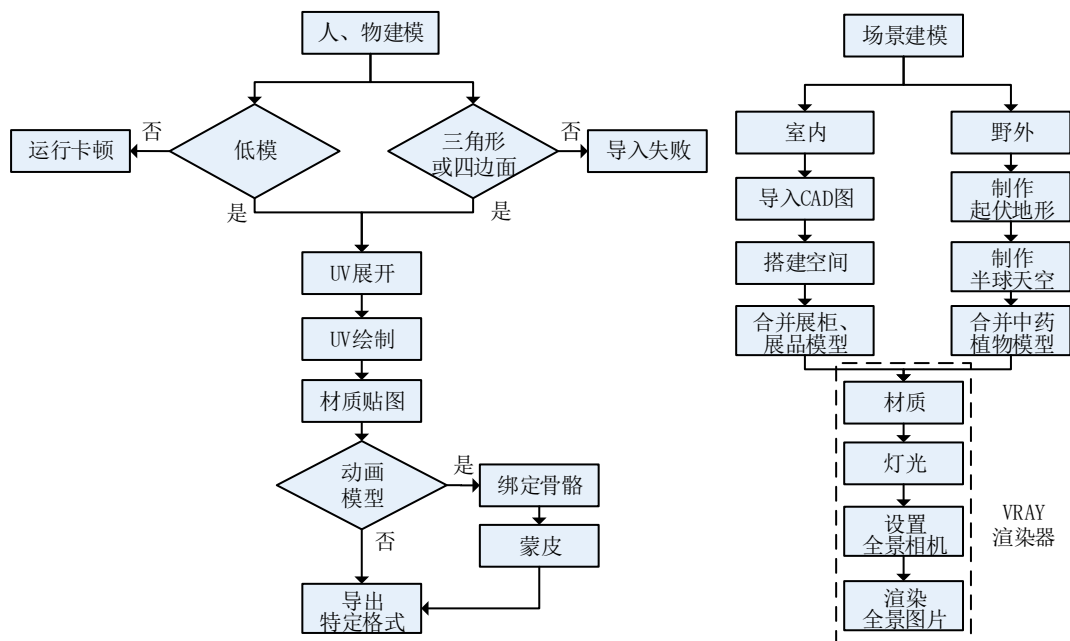


图10 三维建模流程

图，在线框图的定位辅助下，用Photoshop(简称PS)软件处理素材图片，或与BodyPaint软件联合，绘制UV贴图，再回到3Ds Max软件中作为材质贴图赋予对象，此时可对人物模型绑定骨骼、蒙皮设置动作，最后导出FBX格式的文件，并生成交互设置平台可识别的glb格式模型。

VR漫游的场景建模时，室内场景可借助AutoCAD软件绘制展厅的平面图，导入3Ds Max软件中，搭建好空间框架，合并入展柜、展品等模型；野外场景由起伏地形和半球天空构成，合并入中草药植物模型。两种场景的材质和灯光

均为VRAY类型，渲染相机设置为球形360°覆盖视野效果，最终输出长宽比为2:1的全景效果图。

4.2 图片绘制及排版

本系统的图片资源包含锚点图、介绍图片和绘本图。首先确定表现风格，如何首乌手绘介绍图，采用饱和度低的中国传统配色，给人以淡雅、舒适的视觉感受；根据中药植物的真实照片用PS软件绘制线稿、上色，完成图文排版，制作过程如图11所示。



图 11 介绍图片制作过程

绘本内容主要由对话和流程讲解两个部分组成，通过按钮进行串联。按钮设计参考印章造型，体现水墨晕染的效果，与内容页风格呼应，和谐统一，如图12所示。



图 12 介绍图片绘制过程

4.3 其他资源制作

其他资源包含动画视频的制作、文本内容的选择和介绍音频录制，多种数字资源给用户带来更多视听感受，丰富体验、增强学习效果。视频制作使用Animate绘制图形元件输出基本影片后导入剪映软件剪辑、增加配音；音频包含背景音乐和介绍音频，选取舒缓悦耳的国风纯音乐作为背景音乐更贴合系统主题，介绍音频

采用自然亲切的真人配音，配合文本内容，用户可听可阅，提升体验感。

4.4 交互设置

“云上观药”AR折页可在AR在线平台Kivicube或万维引擎编辑制作，选择扫描触发锚点图方式链接前期准备好的图集、模型、音频、视频等数字资源；“寻迹本草”展馆漫游在VR在线平台720云编辑制作，设置环物观察、场景跳转、音视频或文本内容介绍等不同功能的热点，强调漫游过程的交互性。

“草药变形记”数字绘本使用Unity 3D软件实现显示、翻阅页面的交互。现在最常用Unity自带的可视化UGUI创建界面UI，较少用OnGUI()创建UI，因为每一帧画面都会调用方法，开销较大。但本工程体量较轻，且在Editor脚本中可扩展Unity编辑器的功能，用OnGUI()在屏幕上直接绘制UI较为方便。如启动绘本软件时，全屏加载“Start”封面场景，

按下按钮则加载显示内容页的“Main”场景，具体实现如下所示。

```
void OnGUI()
{
    if(myskin)
    {
        GUI.skin= myskin;
    }
    GUI.DrawTexture(new Rect(0,0,Screen.width,
        Screen.height),Resources.Load("Start") as
        Texture2D);
    a+=Time.deltaTime;
    if(a>=1)
    {
        a=1;
        GUI.color = new Color(1,1,1,a);
        if(GUI.Button(new Rect(Screen.width-160,
            Screen.height-140,160,140),"", myskin.
            customStyles[0]))
        {
            Application.LoadLevel("Main");
        }
    }
}
```

药对应的数字资源。一般情况下，三折页和卡片也可以表达为数字资源形式，在网络上展示推广。



图 13 AR 导览折页图

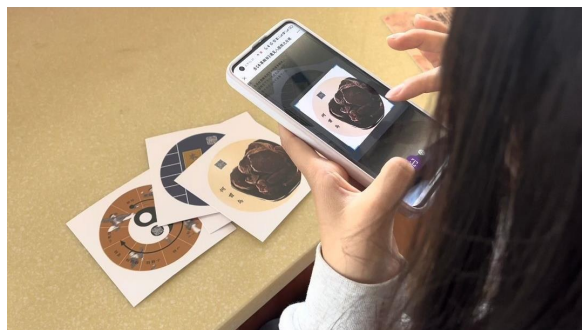


图 14 扫描 AR 卡片

5 结果与讨论

5.1 AR 导览展示

本系统的 AR 导览进一步将二维码及锚点图设计以三折页或卡片形式进行展示、传播，如图 13、图 14 所示。用户先用手机微信扫描封面上的二维码，出现锚点图扫描框，打开折页，扫描想要了解的中草药锚点图，即可激活中草

丰富的数字资源可在手机或电脑屏幕上清晰显示、播放，如图 15 所示，由万维引擎实现的 AR 导览显示界面，用户在屏幕上划动手指对三维模型缩放或旋转观察，左右划动翻看图片。三维模型在 Kivicube 平台的显示效果如图 16 所示，旋转设备摄像头可环绕观察模型。

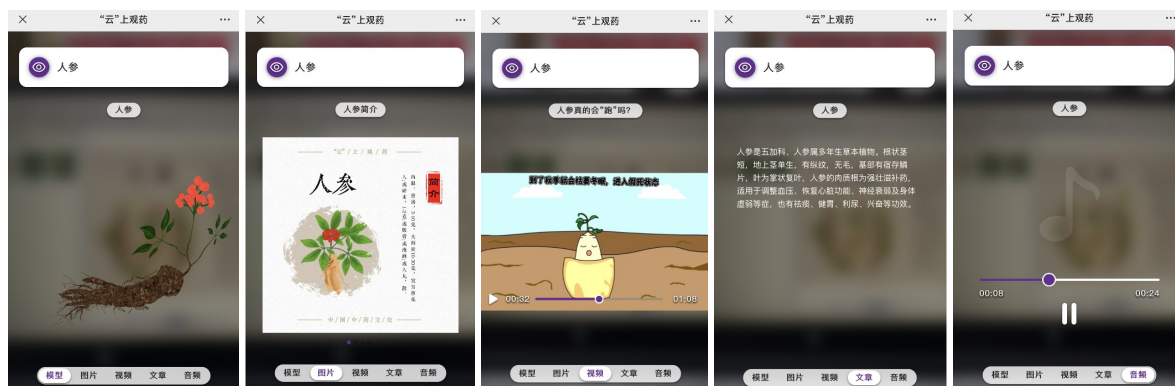


图 15 AR 导览激活的数字资源

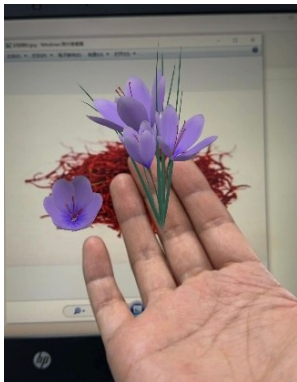


图 16 人参模型展示效果

5.2 VR场景漫游

VR 场景漫游有两类场景：室内展馆和野外识别。首先来到室内场景，展柜中陈列中医古籍，热点链接了中医古籍和名医的介绍图片、模型、音频等资源；药材热点链接到野外场景，可选择正常、小行星、水晶球和鱼眼四种视角。VR 模式较普通模式具备更强烈的身临其境感，通过准心瞄准热点即可触发新的功能；普通模式即通过鼠标或手指点击热点。效果如图 17、图 18 所示。

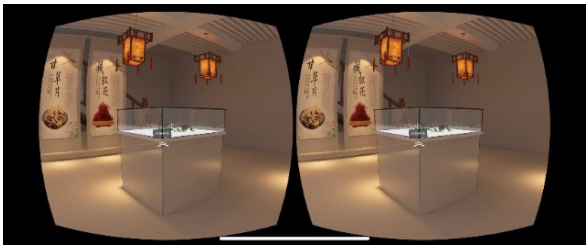


图 17 VR模式室内VR漫游效果



图 18 普通模式野外VR漫游效果

5.3 数字绘本运行

数字绘本可在不同移动终端上安装运行，用按钮切换封面启动页、内容页和结束页，整体画面国风浓郁，如图 19 所示，生动美观而富有趣味，打破了传统纸质书籍在时空方面的局限。



图 19 数字绘本软件

由上述结果得出，基于学习体验开发的中医药文化数字展示系统具有很好的数字资源集成展示功能，学习者可以体验到 AR、VR 和交互媒体软件等丰富的交互形式，在中医药文化不同的领域汲取知识认知和情感感知的养分，整体实现效果好。

6 结语

中医药文化是中华优秀传统文化的重要组成部分，体现了中华优秀传统文化的优秀内涵。中医药文化发展历史悠久、内容丰富、内涵深刻，得到了大部分直接受诊群体的肯定，但其广泛、积极

的传播因深刻的内涵和传统的宣传方式而存在一定阻碍。借助多平台软件工具，基于学习体验设计、开发中医药文化数字展示系统，系统用户即学习者，通过 AR 导览、VR 漫游和交互媒体的形式调动学习者的积极性，启发观察与思考，浸润于中医药文化、精神和理念，实现更深层的体验，从而使中医药文化发扬光大。在将来的研究中，要将最新的数字技术与文化传承结合，针对不同受众进行更加个性化的数字展示，为广大群众更好地了解中医药文化提供更多便捷有效的渠道。

参考文献:

- [1] 陈潇. “一带一路”倡议下河北省中医药对外文化传播交流策略探析[J]. 太原城市职业技术学院学报, 2023(4):194-196.
- [2] 王伟华. 博物馆文化遗产的数字展示与实体展示[J]. 东南文化, 2011(5):91-95.
- [3] 崔楠, 陈全, 徐岚, 等. 当历史文创产品遇上AR: 增强现实技术产品展示对消费者历史文创产品评价的影响[J]. 南开管理评论, 2021, 24(6):50-63.
- [4] 刘崧印, 朱学芳, 李川. 基于VR技术的虚拟图书馆全景漫游系统的设计与实现[J]. 图书馆学研究, 2022(11):47-56.
- [5] 杨青青. 中国戏曲服饰纹样元素码应用的“新媒体”交互视觉设计[J]. 美术研究, 2020(3):67-70.
- [6] 王楠. 学习体验设计的缘起、概念及框架: 用户时代教学设计的实践取向[J]. 现代教育技术, 2018, 28(12):28-33.
- [7] 谢翌, 程雯, 李亚培, 等. 基于学习体验的过程性课程评价[J]. 课程·教材·教法, 2021, 41(5):18-25.
- [8] 张露, 尚俊杰. 基于学习体验视角的游戏化学习理论研究[J]. 电化教育研究, 2018, 39(6):11-20, 26.
- [9] 余明华, 彭红超, 祝智庭. “互联网+”视域下的无缝学习体验设计[J]. 电化教育研究, 2017, 38(11):19-25.
- [10] 钱静, 胡永斌. 人工智能时代的资源新形态: AR电子教材及其学习体验设计框架[J]. 数字教育, 2022, 8(2):77-84.

Research on the development of a digital display system for traditional Chinese medicine culture based on learning experience

Zhu Minjiao*

(Department of Big Data and Internet, Suzhou College of Information Technology, Suzhou 215200, China)

Abstract: Traditional Chinese medicine culture has a long history, rich content, and profound connotations, in order to promote the inheritance, innovation, and development of traditional Chinese medicine culture in the digital era, we explore the design and development of a digital display system for traditional Chinese medicine culture from the perspective of learning experience. Analyze the various forms of digital display, interpret the connotation of learning experience, build a digital display system design framework, and use various platforms and tools such as 3Ds Max software, 720 cloud platform, World Wide Engine, and Unity 3D software to complete the digital resource construction and interactive settings of the traditional Chinese medicine culture digital display system. The system has 3 major modules: AR navigation, VR roaming, and interactive media, it integrates digital resources such as 3D models, 2D images, text, audio and video, and it has rich interactive methods. Based on the learning experience, the system users are learners of traditional Chinese medicine culture. Traditional Chinese medicine culture is naturally implanted in the minds of learners, as a result, it has been widely disseminated and promoted.

Keywords: learning experience; traditional Chinese medicine culture; digital display; platforms; tools

文章编号: 1007-1423(2024)04-0075-08

DOI: 10.3969/j.issn.1007-1423.2024.04.014

基于微信小程序的鸟鸣声识别系统的研究与实现

陈凌芳¹, 周 雁^{1,2*}, 王庆娟^{1,2}, 林佳皓¹, 谌业恒¹

(1. 北京理工大学珠海学院, 珠海 519088; 2. 北京理工大学, 北京 100081)

摘要: 鸟类是生态系统中的重要组成部分, 传统的鸟类调查与监测需耗费大量的人力物力, 效率十分低下。对此, 基于微信小程序, 利用 Flask 框架和深度神经网络, 研究设计了一个鸟鸣声识别系统, 通过声音识别技术实现了鸟鸣声的自动识别和分类。用户可通过微信小程序录制或上传鸟鸣声音频文件, 系统将自动对音频进行预处理和特征提取, 然后使用训练好的 CNN 模型对鸟鸣声进行识别。同时系统还提供了鸟类字典、数据监测可视化等相关信息的查询和展示。

关键词: Flask 框架; 微信小程序; 鸟鸣声识别; CNN 模型

0 引言

鸟类是生态环境的重要组成部分, 其数量和种类的减少引起了广泛的关注。鸟类有着独特的发声器官, 可通过间断的鸣声传达信息^[1], 如攻击、警告以及其他的一些情绪等^[2]。鸟鸣是鸟类之间交流的重要手段, 也是研究鸟类行为和生态学的重要工具。通过分析鸟鸣声中蕴含的特征来识别鸟类, 可以为野生鸟类的监测和保护、生态环境的保护、评价和管理等提供依据, 在自然教育、生态旅游等领域也大有可为。

目前, 国内对语音处理的研究大都集中于语音和音乐的处理, 对鸟类鸣声的识别研究相对较少^[3]。已有的一些研究成果也大都停留在实验和发表论文上, 实际应用的并不多见。

近年来, 微信在国内普及度异常高, 根据 2022 年相关统计数据, 微信月活跃用户已经达到了十亿的量级水平, 且增长量超过百分之三^[4]。微信小程序作为依附在微信上的小程序, 具备无需下载、可便捷地获取和传播等特色, 建立了用户与服务连接的全新模式, 备受用户青睐。

基于此, 本文研究与实现了一个基于微信

小程序的鸟鸣声识别系统, 以期实现随时可用移动设备快速通过声音识别鸟的种类, 节省鸟类观测的人力物力, 同时利用微信小程序平台的优势, 使系统得到较好的普及使用。

1 鸟鸣声识别系统设计

1.1 系统总体设计

系统分为四大模块, 分别为数据收集模块、用户模块、管理员模块和鸟鸣声识别模块。系统的结构如图 1 所示。

各模块功能设计如下。

1.1.1 数据收集模块

从网络上获取和提取有效鸟鸣声数据, 完成前期数据的收集。

1.1.2 用户模块

即用户的微信小程序前端页面, 为用户提供基本的使用功能, 包括:

(1) 授权登录: 用户初次使用识别小程序时, 会弹出弹窗询问用户是否授权; 授权后, 可以访问到用户的昵称、头像等基本信息, 若是新

收稿日期: 2024-01-24 修稿日期: 2024-01-31

基金项目: 广东省普通高校特色创新项目(2022KTSCX202); 广东省科技创新专项资金(“攀登计划”专项资金, pdjh2022b0711); 北京理工大学珠海学院大学生创新创业训练项目(2023062DCXM)

作者简介: 陈凌芳(2001—), 女, 江西吉安人, 本科生, 研究方向为软件工程; *通信作者: 周雁(1981—), 女, 广东高州人, 副教授, 研究方向为语音信号处理、嵌入式系统、自然语言处理等, E-mail: yanzhou0910@qq.com; 王庆娟(1987—), 女, 浙江义乌人, 博士, 讲师, 研究方向为人工智能、嵌入式系统设计等; 林佳皓(2001—), 男, 广东揭阳人, 本科生, 研究方向为软件工程; 谌业恒(2002—), 男, 广东广州人, 本科生, 研究方向为计算机科学与技术

用户,将在数据表中新建该用户的信息记录。

(2) 录音识别鸟类: 用户点击录音按钮, 将询问用户的录音设备授权, 完成录音后, 自动调用鸟鸣声识别模块对录音进行识别。

(3) 上传音频识别鸟类: 用户可选择上传本地鸟鸣音频文件, 上传成功后自动进行识别。

(4) 鸟声字典: 用户可打开鸟声字典认识鸟声库中的鸟类。

(5) 鸟类搜索: 用户想要快速查询鸟类的信息时, 可在搜索栏中进行查找。

(6) 监测鸟类概况: 保护动物不仅仅是个人的责任, 更是全人类的责任, 因此对所有已经识别的鸟类数据进行可视化, 所有用户可查看。

(7) 意见建议: 用户在遇到系统上的问题或者发现伤害鸟类的行为时, 可提交建议或举报。

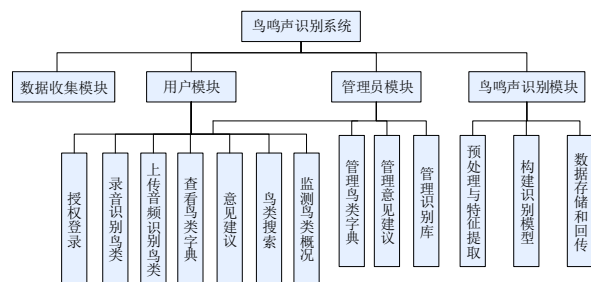


图 1 系统功能结构图

1.1.3 管理员模块

管理员角色包括系统开发人员、鸟类相关领域专家、提供最新鸟类数据的机构等。他们除了可以使用普通用户的功能外, 还可以使用以下权限:

(1) 管理鸟类字典: 查看和管理鸟类数据库。

(2) 管理意见建议: 查看和管理用户提交的意见建议数据。

(3) 管理鸟类识别库: 查看和管理所有用户的识别数据。

1.1.4 鸟鸣声识别模块

此模块为系统的核心模块, 用户提交数据后, 连接到后台进行声音的预处理, 模型的加载、识别, 并将处理结果进行传输和存储。包括:

(1) 预处理与特征提取: 用户的环境复杂多样, 接受到复杂音频时需对数据进行预处理, 再提取蕴含鸟鸣信息的特征, 供识别模型所用。

(2) 构建识别模型: 识别模型作为系统的核心部分, 用于根据鸟鸣声特征识别出鸟类。

(3) 识别结果存储和回传: 识别鸟类后将识别结果存储到数据库及回传至前端页面。

1.2 技术路线设计

1.2.1 数据收集技术

系统的数据获取主要采用爬虫技术, 以自动化模式爬取目标网页的数据。

(1) 发起请求: 使用 Requests 技术实现。Requests 技术基于爬虫的基础工具包, 实现自动爬取页面信息的功能。Requests 配备对应的 GET、POST、HEAD、PUT、PATCH、DELETE 六个基础方法和一个 Request 通用方法, 具有 HTTP 连接池自动化、持久 Cookie 会话、SSL 认证等基本功能^[5]。HTTP 库发送 Request 请求到鸟声库网页一目标站点, 类似于在浏览器中输入目标网址并回车, 等待服务器响应。

(2) 响应内容获取: 若服务器有正常响应, 其中的响应内容有可能是整个网页的代码, 有可能是页面中的字符串, 也有可能是二进制的

数据。

(3) 数据解析: 对获取的内容, 使用 bs4 第三方库的 BeautifulSoup 解析文档, 或使用 Json 解析数据。BeautifulSoup 可以从 HTML 文档中提取数据, 解析出特定的内容, 避免了许多繁琐过程^[6]。而 Json 作为一种较为轻便的数据交换格式, 常被应用于数据交互场景。爬取网页时, 若服务器以 Json 格式的形式返回数据, 可使用 Json 方法转换数据为字典类型的格式。

(4) 保存数据: 解析后的数据保存到数据库里面, 声音文件保存为特定的 wav 格式。

1.2.2 前端技术

系统的微信小程序前端采用微信原生框架 MINA, MINA 框架由微信小程序提供, 此框架不仅封装了微信客户端里面的文件系统、网络通信和任务管理等基础功能, 还可以使用一整套 JavaScript API, 方便开发者使用微信的各种功能。

框架系统分为两部分, 即逻辑层和视图层。在视图层, 使用了部署在小程序上的云服务, 即云开发小程序, 结合小程序自带的云数据库, 利用云数据库来储存微信小程序上较大的图片, 并在数据

展示页面结合Echarts技术进行数据可视化。

1.2.3 后端技术

系统后端主要采用 Flask 作为框架，Pycharm 作为开发平台，数据库采用 MySQL Community Server，数据库管理工具使用 Navicat Premium，Flask 和数据库之间采用 ORM 插件 Flask-SQLAlchemy 进行连接。

其中，Flask 作为一个依赖于 Python 的轻量级 WEB 应用开发框架，其中核心是 Jinja2 和 Werkzeug，具有轻便、简单、强拓展性等特点^[7]，在微信小程序的开发上有着加强兼容性^[8]。系统使用 Flask 框架提供的功能处理 request 请求，并返回值给客户端。

1.2.4 鸟鸣声识别技术

鸟鸣声比较复杂多变，但其信号与语音信号相似，都具有短时平稳的特性，因此可借鉴语音处理的方法来进行分析和识别。在鸟鸣声数据预处理中，需要对鸟鸣声进行分帧加窗、预加重、去噪、标准化等操作，以提高模型的鲁棒性^[9]。其中降噪叠加使用谱减法与带通滤波器。

谱减法原理如以下公式所示，假设 y 为噪声污染的语音信号， d 为信号中的加性噪声，而 x 为 y 中的纯净语音信号，利用信号前后变化平稳的静音段来估计噪声谱，那么可求得带噪语音信号中纯净信号 x 的功率谱：

$$|X(\omega)|^p = |Y(\omega)|^p - |D(\omega)|^p$$

带通滤波器用以筛选特定频率信号，将频率信号控制在一定范围内，可将其视为高、低

通滤波器协同作用。根据鸟鸣声的频率分布，设置带通滤波器，可排除其它频段噪声干扰。

在鸟声识别神经网络模型中，将 MFCC 图像特征作为模型输入，使用 CNN 模型对其进行训练和识别分类^[9]。

鸟鸣声数据的预处理和神经网络的搭建均使用 PyTorch 框架来完成。

模型搭建完成后，使用 Flask 框架集成 API。将预测模型打包好，使用 HTTP API 接口进行参数传递，结合 Flask 框架将对其进行部署。在 Flask 中，全局的 request 对象提供了这些需要传输的信息。使用 Flask 框架封装 API 可以提高开发效率、降低开发成本、提高可维护性和可扩展性，是一种非常优秀的 API 开发方式。

1.3 数据库设计

系统数据库设计的 E-R 图如图 2 所示。

对数据表简介如下。

(1) human_data 表：用户信息表，存储系统的用户信息，主要字段包括用户 ID、用户昵称、系统角色、反馈信息 ID 号等。

关联关系：与 result_data 表是一对一关系。

(2) main_bir 表：主要鸟类信息表，存储系统可识别的鸟类信息，用于识别鸟鸣声识别模块，主要字段包括鸟类 ID、鸟类中文名、学名、音频 ID 等。

(3) result_data 表：识别结果表，存储用户识别的鸟类信息，主要字段包括结果 ID、识别数据结果、鸟类名称、结果存储地址等。

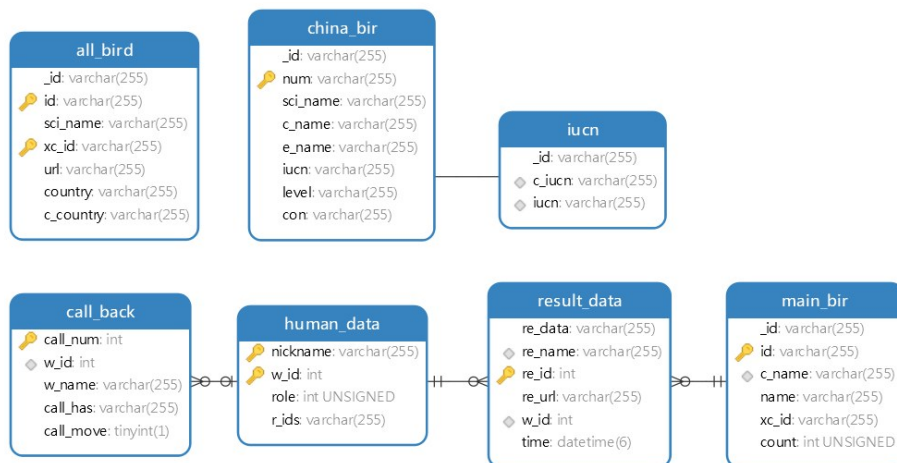


图 2 数据库 E-R 图

(4) all_bir表：国外鸟类信息表，存储国外鸟类的信息，主要字段包括鸟的ID、学名、数据链接地址、国家中英文名字、鸟鸣声文件存储位置等。

(5) china_bir表：国内鸟类信息表，存储国内鸟类的信息，主要字段包括鸟的ID、学名、中文名、濒危程度、保护级别等。

关联关系：与iucn表是一对一关系。

(6) iucn表：鸟类濒危程度表，存储国内的存储濒危鸟类信息，主要字段包括濒危程度及其缩写。

(7) call_back表：用户反馈表，存储用户反馈的意见建议、投诉等，主要字段包括反馈ID、用户ID、用户昵称和反馈内容。关联关系：与human_data表是一对多关系。

2 鸟鸣声识别系统实现

2.1 数据收集

配置好爬虫软件和工具，在网络上查找具有鸟鸣声数据资源的网站并进行操作。以 xeno-canto 官网为例，如图 3 所示，查看网页上鸟类数据的格式特点，找到页面制定的 a 标签，目标是爬取声音数据较完整且样本较多的鸟类数据。

爬取相关数据后，还需要下载对应的音频文件。xeno-canto 官网上的音频都是被压缩过的，不可用于模型训练，因此将文件以 wav 的格式进行保存。

2.2 前端功能实现

系统前端界面首页如图 4 中的左图所示。由于篇幅所限，仅对前端主要功能的实现进行简单阐述。

2.2.1 普通用户功能

(1) 鸟类识别：用户点击主页的“识别鸟类”按钮，可选择在线录音或上传录好的音频文件。若选择录音，将确认录音设备的授权并进行录音，完成录音后，自动将录音传至后端进行识别。用户也可点击录音上传，上传成功后将进行识别。识别后，将返回识别结果和置信度，如图 4 所示。

(2) 鸟类搜索：用户若想快速查找鸟类信息，可在首页搜索栏中输入关键字进行查找。搜索功能是通过数据库进行模糊搜索实现的，模型文件里利用“类模型.query.filter(类模型.字段.like(“%查询条件%”))”语句可实现模糊查找的功能。

(3) 查看鸟声字典：用户想要认识鸟类时，可点击主页的“所有鸟类”按钮。系统鸟类库分为三个：国外鸟类库、国内鸟类库及本系统可识别的鸟类库。用户可点击感兴趣的库查看对应的鸟类信息和聆听鸟鸣声，如图 5 所示。

(4) 监测鸟类概况：在小程序主页面的下部，会显示若干条最近的用户识别鸟类记录，点击此位置即可进入鸟类识别数据可视化界面，如图 6 所示。

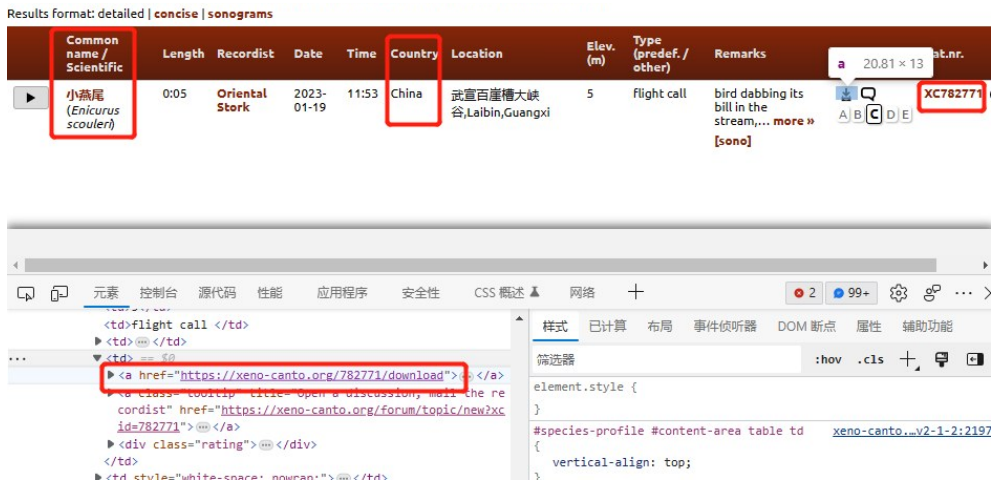


图 3 数据爬取实例

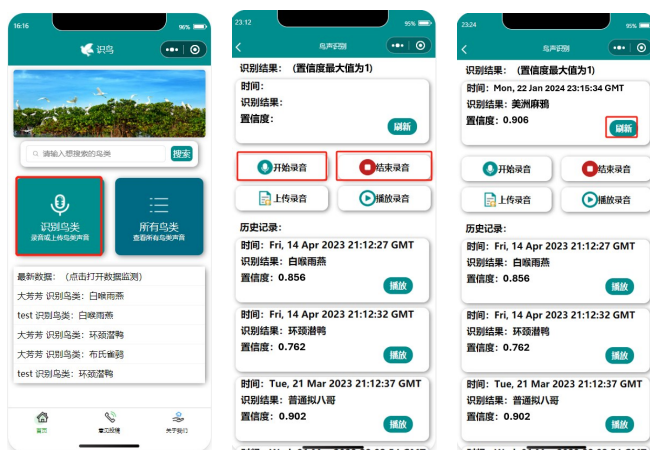


图4 鸟鸣声识别界面

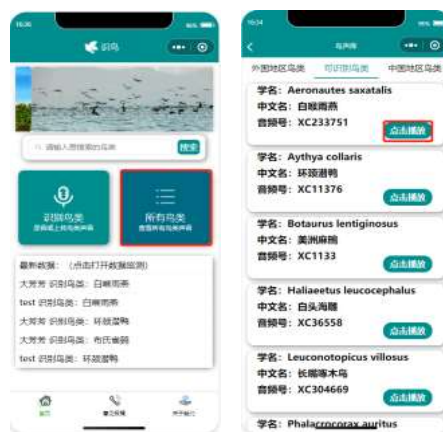


图5 查看鸟声字典界面

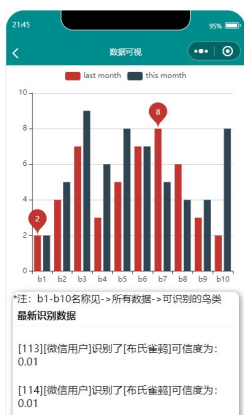


图6 数据监测可视化界面

2.2.2 管理员功能

该模块主要实现管理员特有功能，包括数据库的管理、意见建议的处理等，大都涉及到前端小程序与后端服务器之间的连接。本系统采用小程序内测版开发，利用本地的域名作为服务器地址。前后端数据传输基于Flask框架，数据采用Json格式。为了避免中文乱码的问题，需在header中对字符集进行设置，以确保数据的正确性。以获取国内鸟类数据库数据为例，实现前后端数据传输的关键代码如下：

前端关键代码：

```
wx.request({
  url: "http://169.254.241.85:5000/birddata"
  methods: "POST",
  header: {
    contest-type: 'application/json'
    chartset: 'utf-8',
```

```
},
data: {
  china bir: this.data.china bir,
},
success: function (res) {
  console.log(res.data);
  that.setData({
    birddata: res.data,
  });
  wx.showToast({
    title: "ok",
    icon: 'success',
    duration: 1000,
  });
}
```

后端关键代码：

```
@app.route('/birddata', method=[ 'POST' ])
def getchina():
  print("get data")
  data_json = request.get_json()
  print(data_json)
  bir_info = data_json('bir_info')
  all_bir = get_boxes(bir_info['china_bir'])
  res = all_bir
  return jsonify(res)
```

2.3 鸟鸣声识别模块

2.3.1 预处理与特征提取

本文从网络上爬取鸟鸣声数据并筛选后，得到10种鸟类的1823条样本，作为搭建鸟鸣声识别模型的数据集。导入Python的librosa音频

处理库,可方便地对音频进行各种处理。

(1) 预处理

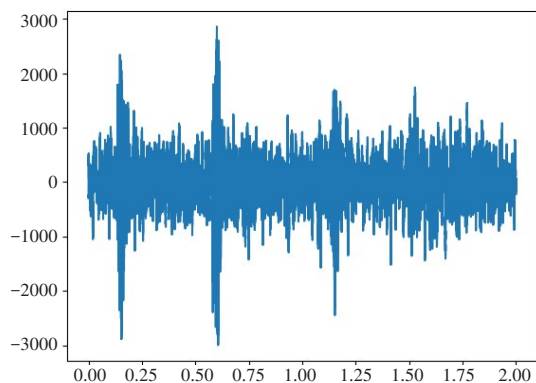
在进行鸟鸣声信号的分析 and 处理之前,需要进行一系列预处理操作,包括数据归一化、预加重、分帧、加窗、去噪等,尽可能消除环境、鸟类发声器官和录音设备等带来的影响,提高特征提取和分类识别的质量。

鸟鸣声的降噪先后使用了谱减法与带通滤波器。

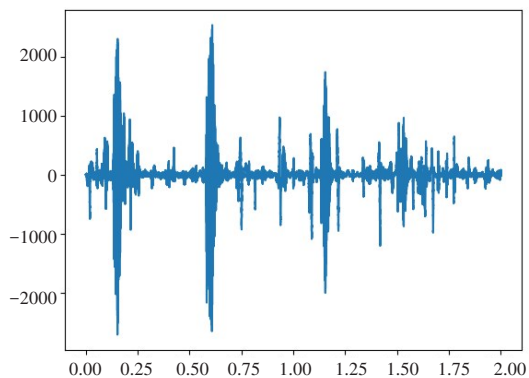
具体实现时,使用的是改进的谱减法:Berouti谱减法。其特点是调整了谱减法噪声处理规则,为噪声设置下限而不置零,以此解决谱减法使用过程中出现的音乐噪声问题。

鸟类鸣声频率主要分布在 $2\sim 9\text{ kHz}$ ^[10],而环境噪声频率大致处在 2 kHz 以下,因此对音频添加 $2\sim 9\text{ kHz}$ 阈值区间的带通滤波器,以去除多余频带的干扰。

图7所示为一个鸟鸣声示例音频降噪处理前、后的对比波形图。



(a) 降噪前音频波形图



(b) 降噪后音频波形图

图7 降噪前、后音频波形对比图

(2) 特征提取

梅尔频率倒谱系数(MFCC)由戴维斯和梅尔斯坦在1980年代提出,在语音处理中得到了广泛应用^[11]。本系统从预处理后的音频中提取MFCC系数,并转化为图像,用于后续的模式训练和识别。

2.3.2 识别模型的建立

本文使用速度较快的深度学习模型——CNN网络,对提取出来的特征值进行训练。

(1) 数据准备

将MFCC特征提取后得到的图像数据集,按比例 $7:2:1$ 划分为训练集、测试集和验证集。

(2) 模型设计

输入层接收MFCC图像数据,采用最常用的3通道的图像数据,即每个MFCC图像的通道数为3。

用多个卷积层提取MFCC图像的特征,卷积层每一层包括多个卷积核,每个卷积核的大小为 3×3 ,步长为1,激活函数使用ReLU函数。这样的卷积层输出为64个特征图。卷积层的输出通过多个池化层进行下采样。使用最大池化,池化核的大小选择为 2×2 ,步长选择为2。将池化层的输出展开成一维向量,然后使用多个全连接层进行分类。使用Softmax函数将全连接层的所有输出转换为概率分布,这样做有助于进行分类任务^[12]。

(3) 模型训练

使用训练集对CNN模型进行训练,采用随机梯度下降优化算法。以准确率为指标,使用验证集对模型进行评估。

(4) 模型测试与部署

使用预先划分好的测试集对训练好的模型进行测试分析,识别准确率为93.04%,召回率为90.83%。

最后,将训练好的模型部署到路由器中与微信小程序相连,使用Python的Flask框架进行快速部署。

3 系统测试

为了解系统的功能和性能,本文对系统进行了多层次、全功能模块的测试,测试的过程贯穿了系统开发的全过程。

3.1 单元测试

在系统开发的过程中,设计了单元测试,对系统的关键函数、微信小程序按钮等进行检测。

(1) 使用微信小程序开发工具测试。由于许多图片都存放在云数据库里,因此在开发完每个页面后,在微信开发者工具中进行真机测试,以检测页面显示是否有延迟。

(2) 使用 miniprogram-simulate 微信小程序测试框架。其主要功能包括模拟触碰事件、自定义时间的触发、更新自定义组件数据以及触发生命周期等。

对于单元测试的所有测试用例,实际结果均与预期一致。

3.2 功能测试

先进行黑盒测试。针对系统各个模块的主要功能,设计样例进行测试,在涉及输入数据的用例中,尽量输入多样的数据,完成完整的业务逻辑以发现问题。

同时,对系统执行了白盒测试。白盒测试基于系统的代码,用于测试鸟鸣声识别系统的内部代码逻辑和系统结构是否正确。其目的是发现代码中的缺陷和错误,以便在开发和测试过程中进行修复和改进。

本文设计了详尽的用例对系统的各个功能进行黑盒、白盒测试,所有样例实际结果与预期一致。

3.3 接口测试

后端开发时,利用接口测试工具 Postman 进行后端的接口测试,模拟前端传递信息至后端,检验后端功能的完整性。经测试,实际结果均与预期返回结果一致。

4 结语

本文利用 Flask 框架和深度神经网络,研究设计了一个基于微信小程序的鸟鸣声识别系统,实现了鸟鸣声系统的数据收集、预处理、识别模型搭建、系统测试等,可利用鸟类鸣声方便快捷地识别鸟类,以期为鸟类保护和生态环境监测等领域提供支持。

然而,系统还存在着一些不足之处,本团队下一步的工作包括:

(1) 爬取更多鸟鸣音频数据,丰富鸟鸣声数据库,增加可识别的鸟类数量。

(2) 进一步研究鸟鸣音频的预处理技术,提高系统在不同环境下的鸟声识别精度。

(3) 保护物种、濒危物种鸟类样本往往较少,但对其监测却尤为重要。因此,下一步将考虑数据集样本不平衡情况,研究针对长尾、开放数据集的识别算法。

(4) 进一步完善系统后端,拓展微信小程序的功能。

(5) 探索更轻量的深度网络模型。

参考文献:

- [1] 雷富民,王钢,尹祚华. 鸟类鸣唱的复杂性和多样性[J]. 动物分类学报,2003,28(1):163-171.
- [2] 肖华,张雁云. 鸟类鸣声研究[J]. 生物学通报,2009,44(3):11-13.
- [3] 乔玉,钱昆,赵子平. 基于机器听觉的鸟声识别的中文研究综述[J]. 复旦学报(自然科学版),2020,59(3):375-380.
- [4] 新华社. “数字”点亮美好生活:透视第50次《中国互联网络发展状况统计报告》[J]. 信息系统工程,2022(10):4-5.
- [5] 杨健,陈伟. 基于Python的三种网络爬虫技术研究[J]. 软件工程,2023,26(2):24-27,19.
- [6] 李晨昊. 基于BeautifulSoup+requests和selenium爬虫网页自动化处理的实现和性能对比[J]. 现代信息科技,2021,5(16):10-12,18.
- [7] 叶锋. Python最新Web编程框架Flask研究[J]. 电脑编程技巧与维护,2015(15):27-28.
- [8] 张勤. 基于flask的IP管理微信小程序设计与应用[J]. 电脑知识与技术,2020,16(5):92-93.
- [9] 欧昀,周晓彦,李大鹏. 基于卷积神经网络的鸟声识别实验设计[J]. 实验室研究与探索,2022,41(4):99-102,112.
- [10] 季婷,张雁云. 环境噪音对鸟类鸣声的影响及鸟类的适应对策[J]. 生态学杂志,2011,30(4):831-836.
- [11] 赵扬青,彭智才,蒋雨涵,等. 音频的梅尔频率倒谱系数特征抽取过程[J]. 信息技术与信息化,2023(1):104-111.
- [12] YILDIZ I B, VON KRIEGSTEIN K, KIEBEL S J. From birdsong to human speech recognition: Bayesian inference on a hierarchy of nonlinear dynamical systems[J]. PLOS Computational Biology, 2013, 9(9):e1003219.

Research and implementation of birdsong recognition system based on WeChat mini program

Chen Lingfang¹, Zhou Yan^{1,2*}, Wang Qingjuan^{1,2}, Lin Jiahao¹, Chen Yeheng¹

(1. Beijing Institute of Technology Zhuhai, Zhuhai 519088, China; 2. Beijing Institute of Technology, Beijing 100081)

Abstract: Birds play a crucial role in ecosystems, but the traditional ways of studying and monitoring birds are inefficient. In order to solve this problem, this paper develops a birdsong recognition system integrated into a WeChat mini-program using the Flask framework and deep neural network. The system employs sound recognition technology to identify birdsongs automatically. Users can record or upload the bird song audio files through the mini-program. The system then automatically preprocesses and extracts features from the audio, and use a trained CNN model for the recognition of bird songs. In addition, the system provides functionalities like a bird dictionary and visualization tools for data monitoring for the users to look up or display.

Keywords: flask framework; WeChat mini program; birdsong recognition; CNN model

(上接第 61 页)

Application of mechanical completion punch management system in FPSO construction project

Xing Tongchao*, Shi Zhenhua, Zhao Xilei, Wang Liang, Wu Jiangyong

(Offshore Oil Engineering Co., Ltd., Qingdao 266520, China)

Abstract: Mechanical completion is widely used in international marine engineering construction projects. As an important transition stage, Mechanical completion is a project managing theory and fine management of project. Construction and inspection personnel will log a large number of punch in the process of performing mechanical completion according to ITR. A punch list will be generated. How to track and manage the punchlist quickly and efficiently in real time is the focus of project management personnel and field inspectors. To solve above problems, a mechanical completion punch management system is built based on Django in this paper. Through the actual operation of a FPSO project, the application effect is good. The application experience can be used as a reference for other similar projects.

Keywords: mechanical completion; punch management; Django; Python; FPSO

文章编号: 1007-1423(2024)04-0083-06

DOI: 10.3969/j.issn.1007-1423.2024.04.015

基于 Spring Boot 的蒿草智能识别小程序的研究与实现

赵文欣^{1*}, 苏 峰²

(1. 内蒙古电子信息职业技术学院计算机与网络安全学院, 呼和浩特 010010;

2. 内蒙古自治区气象服务中心, 呼和浩特 010051)

摘要: 蒿草智能识别小程序是基于大数据分析 with 智能图像识别技术的融合应用, 贯穿蒿草智能识别、用户信息上报、数据可视化、知识科普、在线社区和防治鼻炎资讯推送六大功能为一体的蒿属植物识别小程序, 同时开发蒿属植物时空分布情况的大数据可视化平台, 从而帮助政府及有关部门, 更好地进行蒿草治理, 解决生态民生问题, 并且可以帮助深受蒿草困扰的过敏患者在蒿草分布扩散的密集期提前做好预防。

关键词: 图像识别; 蒿草智能识别; 数据可视化; 鼻炎防治

0 引言

近年来, 我国各地城市民众在每年的某些时间段都会有过敏症状, 如过敏性鼻炎、结膜炎、耳膜炎等五官炎症, 更甚者可能会患皮肤疾病, 这种现象尤以北方干旱、半干旱地区为多。大量生物和生物医学研究表明, 这些过敏症状多源于空气污染物, 当空气中漂浮的各类微小颗粒物达到一定浓度时, 就会引起某类易感人群出现过敏病症。从各种研究结果可以看出, 大多数过敏现象主要是因为空气中漂浮的花粉粒所造成的, 尤其是蒿属植物、豚草和葎草类植物的花粉, 是主要的季节性致敏因素。这些植物品种在华北、西北、东北和西南贫瘠寒冷干旱地区相对较多, 其中蒿属植物分布广且品种多, 导致过敏反应也最为明显。

本文设计使用 Spring Boot 架构^[1], 运用软件工程的开发方法, 整合 VUE 技术^[1]、智能图像识别技术^[2]、数据可视化分析技术等, 提升系统的智能性、交互性和可操作性。

1 需求分析

近年来, 蒿属植物泛滥成灾, 极大地影响

了人们的日常生活, 通过市场调研分析, 本文致力于开发一款集蒿草智能识别、用户信息上报、数据可视化、知识科普、在线社区和防治鼻炎资讯推送六大功能为一体的微信小程序, 从而帮助政府及有关部门更好地进行蒿草治理, 解决生态民生问题, 并且可以帮助深受蒿草困扰的过敏患者(如鼻炎、过敏和哮喘患者)尽可能避免接触自己过敏的蒿属植物。

1.1 背景分析

1.1.1 社会背景分析

近年来, 呼和浩特不少市民出现鼻痒、鼻塞、眼睛红肿、发痒等过敏性鼻炎症状, 生活备受困扰。许多市民热盼政府能对即将到来的杨柳飞絮、蒿属植物花粉进行有效治理。为应对市民各类过敏症状, 呼和浩特相关部门进行了诸多举措。

1.1.2 经济背景分析

市场上已有的识花“神器”和识别植物的软件, 约 3~5 s 就能显示结果, 开发成本低, 用户量较大, 目前, 互联网已深入到各地区各行业, 依靠微信小程序、手机 APP 进行技术研

收稿日期: 2023-11-21 修稿日期: 2024-01-15

作者简介: *通信作者: 赵文欣(1993—), 女, 内蒙古通辽人, 硕士研究生, 助教, 从事计算机专业类教学工作, 研究方向为云计算及大数据评测、工业互联网, E-mail: 957701860@qq.com; 苏峰(1993—), 男, 内蒙古通辽人, 硕士研究生, 研究方向为气象服务

发,打造全新的商业模式,对传统的企业进行数字化转型具有很好的前景。

1.1.3 产品背景分析

2020年,针对蒿草致敏问题,呼和浩特市园林建设服务中心制作了《呼和浩特地区较强致敏植物蒿属植物识别手册》,并为每种较强致敏蒿属植物建立了识别二维码。用户通过扫描二维码,就可以看到详细的预防措施,包括如何识别和回避易致敏的蒿草物种等。

但是,这种方式对用户来说存在很多弊端,总结为以下四点:

(1) 治理方式单一:虽然蒿草的致敏性非常强,但作为自然界的一种植物,它的存在也有其价值,比如,青蒿素就是从蒿属植物中提取的。

(2) 治理困难:通过播撒药水抑制蒿草开花,难度太大,且意义不大,作为草本植物,蒿草是靠种子繁殖,且繁殖速度很快。

(3) 分布范围广,致病率高:蒿草分布较广,包括沙漠地区和城市,沙漠地区的蒿草是用于防风固沙的,不能直接处理,而城市中的蒿草会导致过敏性疾病、哮喘、皮炎等,极大影响人们的生活和工作。

(4) 已有软件功能单一:目前已存在的小程序是呼和浩特市园林建设服务中心组织所属科研部门制作的“蒿属植物识别手册”二维码,必须通过扫描二维码,才能查看常见的16种蒿属植物,功能比较单一。

1.2 总体业务需求分析

根据背景分析描述及痛点分析可知,为解决上述问题,本文的蒿草智能识别小程序的业务流程如图1所示。

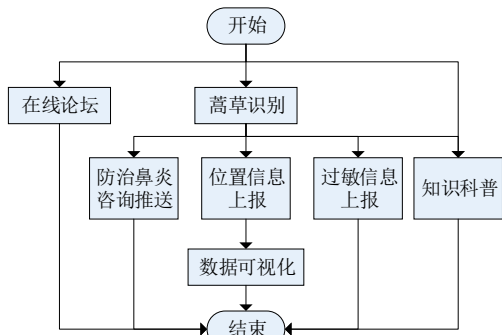


图1 业务流程图

1.3 系统功能性需求

根据市场定位,将系统用户划分为小程序用户、相关政府部门、相关医疗部门及合作商家四类,基于四种用户类型的业务交互如图2所示。

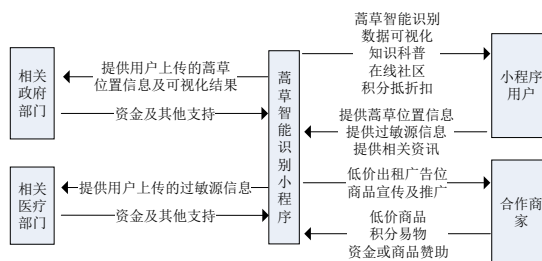


图2 业务交互图

另外,结合需求分析和用户分析,将小程序划分为蒿草智能识别、用户信息上报、数据可视化、知识科普、在线社区和防治鼻炎资讯推送六大功能模块,功能模块图如图3所示。

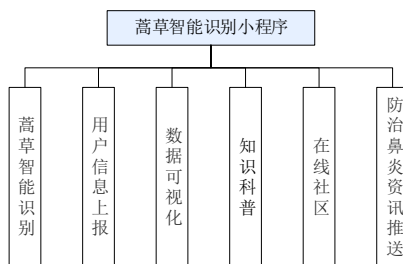


图3 功能模块图

1.3.1 蒿草智能识别模块

采用智能图像识别技术,使用微信小程序进行蒿草识别,通过扫描蒿草图片或直接扫描蒿属植物,显示蒿属植物的名称,根据名称在知识科普模块查看相关蒿草的致敏信息、分布范围、致敏时间及预防方式等各种信息。

1.3.2 用户信息上报模块

采用大数据分析技术与智能图像识别技术进行信息上报,包括蒿草位置上报、过敏源上报,位置上报是上传用户识别的蒿属植物所处位置,通过上传位置及名称获得积分奖励;过敏源上报是用户在个人信息模块填写自己的过敏源信息。

1.3.3 数据可视化模块

采用数据可视化技术、大数据分析技术与智能图像识别技术^[3]融合应用等,通过前端传回来的蒿属植物的位置及名称,做一个基于大数据分析的可视化大屏,更直观地显示各类蒿属植物分布的区域,将结果提供给有关政府部门,帮助他们更好地进行蒿草治理,最后形成蒿草植物花期时空分布可视化大屏。

1.3.4 知识科普模块

与气象部门合作,进行蒿草功效科普,定期播报呼和浩特市花粉浓度并提供鼻过敏相关的科普知识。

1.3.5 在线社区模块

为鼻炎患者专门设立的在线社区,在这个社区中每位用户都可以分享防治鼻炎的经验,以论坛的形式进行,加强平台的活跃度。

1.3.6 防治鼻炎资讯推送模块

蒿草智能识别小程序每天都会在出租的广告位主动为用户推送较新的防治鼻炎资讯,并且以图片的形式向广大用户展示,同时可以通过视频详细了解。最终,通过建立蒿草分布数据智能平台,建立内蒙古地区蒿属植物大数据全景数据库,围绕蒿草信息上报、在线社区数据共享、数据可视化、资讯及积分四大体系,结合人工智能技术处理蒿草数据,帮助鼻炎患者有效预防蒿草过敏。

1.4 系统非功能性需求

从响应速度、可用性和可扩展性三方面进行分析:

1.4.1 响应速度

使用C/S模式,通过HTTP进行数据交互,如果图像识别速度过慢,会影响用户体验。所以,要求在开发过程中提升接口调用效率,最终将响应时间控制在5 s之内。

1.4.2 可用性

应提供美观大方的小程序界面,用户可以在短时间内快速且灵活地使用本系统。另外,考虑到蒿草知识科普模块内容及图片,采用图片轮播的技术对图片进行处理,既美观又实用。

1.4.3 可扩展性

应具有良好的可扩展性,能够为相应合作部门提供指定扩展接口,并且可以灵活对接相关政府及气象部门提供的数据库扩展接口。

2 系统的设计

2.1 框架设计

蒿草智能识别小程序采用C/S模式下的Spring Boot架构,将整体划分为客户层、服务层和数据层三层结构。

蒿草智能识别小程序三层结构的系统架构如图4所示。

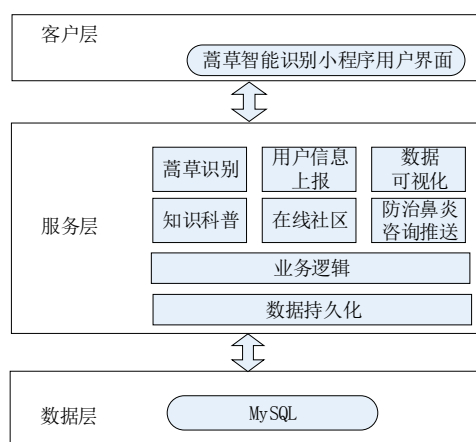


图4 系统架构图

2.2 数据库详细设计

蒿草智能识别小程序划分为六个主要模块,使用MySQL数据库,对蒿草智能识别的逻辑模型进行设计。针对本文研究的重点内容,设计了小程序用户、蒿草位置、过敏信息、蒿草、评论、话题、广告等共七个主要的数据库实体,各数据库实体的详细设计如下。

2.2.1 小程序用户

小程序用户拥有openid、ID、用户名、密码、头像、积分六个属性,用户实体图如图5所示。

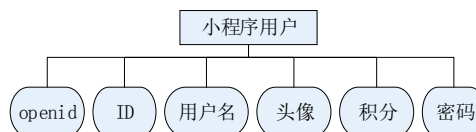


图5 用户实体图

2.2.2 蒿草

蒿草拥有ID、蒿草名称、蒿草图片、科属、拉丁学名、形态特征、致敏信息及预防方式等八个属性，蒿草实体图如图6所示。

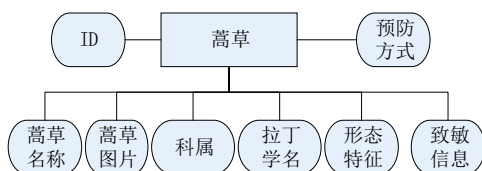


图6 蒿草实体图

2.2.3 蒿草位置

蒿草位置拥有ID、蒿草图片、位置信息共三个属性，蒿草位置实体图如图7所示。

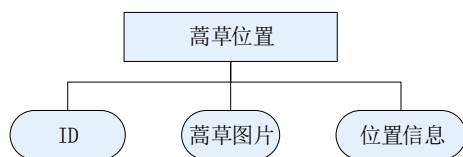


图7 蒿草位置实体图

2.2.4 过敏信息

过敏信息拥有ID、蒿草图片、用户名、过敏源、过敏反应共五个属性，过敏信息实体图如图8所示。

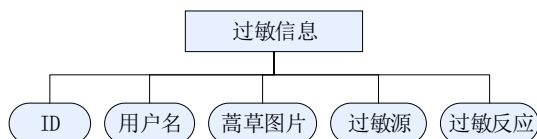


图8 过敏信息实体图

2.2.5 评论

评论拥有ID、评论员、内容、创建日期、回复ID共五个属性，评论实体图如图9所示。

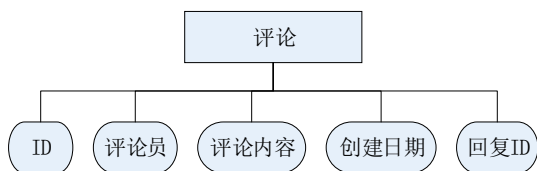


图9 评论实体图

2.2.6 话题

话题拥有ID、作者ID、作者名、话题内容、回复次数、浏览次数、回复时间、标题、主题、点赞数共十个属性，过敏信息实体图如图10所示。

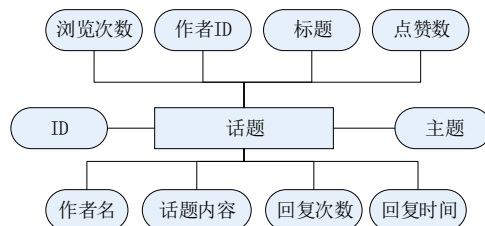


图10 话题实体图

3 系统实现

3.1 蒿草智能识别的实现

该模块主要是蒿草智能识别，试验后识别准确率达到了90%以上，首先通过拍照或者图片上传的方式选择一张蒿草图片，然后点击蒿草识别，通过与数据库中的信息对比，按照图片识别相似率，从高到低给出三种可能的结果。

3.2 知识科普的实现

该模块主要是知识科普，包括十六种常见蒿草基本知识科普，致敏信息及预防方式等，点击不同的蒿草名称，可以进入蒿草详细信息介绍模块查看各类信息，例如科属、拉丁学名、形态特征等。

3.3 数据可视化的实现

该模块主要是数据可视化，能够在地图上清晰地显示蒿草的分布情况，点击后还可查看蒿草详情信息。

3.4 用户信息上报的实现

该模块主要包括过敏源信息上报和蒿草位置信息上报两个主要功能，用户在蒿草智能识别功能使用完毕后，可以点击填报信息按钮进行蒿草位置信息上报，用户通过及时上报蒿草位置来获取相应的积分奖励，最后可以用所得积分易物或者抵折扣。

3.5 在线社区的实现

该模块主要是在线交流与咨询，可以在线发布防治鼻炎资讯及治理蒿草的新闻及资讯等，用户可以在线交流，发帖子，查看帖子的评论数、点赞数及回帖数，同时可以查看帖子的最后评论时间等。

3.6 个人中心的实现

该模块主要是个人中心，主要包括天气信息、用户积分信息、广告位等。

4 系统测试

遵循软件测试规范，制定出符合蒿草智能识别小程序的测试计划，将从单元测试和性能测试两个方面进行相关的阐述。

4.1 单元测试

开发过程是通过搭建Spring Boot+Vue^[4]的体系架构进行的，所以在进行单元测试时，优先选用单元测试工具Junit进行自动化测试^[5]，然后采用Postman工具分别对Controller层、Service层和Dao层中的所有接口进行测试。同时，基于Spring Boot的三层架构，制定了详细的单元测试方案，最终得出结论，各个单元模块的接口均通过测试。

4.2 性能测试

根据实际场景，使用LoadRunner对蒿草智能识别小程序进行性能测试。以蒿草智能识别功能为例，测试并发用户数对系统性能的影响，打开小程序首页后，进行50、100、300个用户并发操作，结果见表1。

表1 性能测试结果

并发用户数/个	50	100	300
通过的虚拟用户数/个	50	100	300
平均响应时间/s	1.543	2.253	3.835
每秒HTTP响应数/次	36.396	35.399	34.256

5 结语

在“互联网+”时代，人们深受蒿草困扰的环境下，本文基于“互联网+”的背景^[6]，设计并实现了基于Spring Boot框架的蒿草智能识别小程序，通过资源整合与共享的方式来获取相关盈利，一方面针对需要蒿草位置信息的相关政府部门及需要过敏源信息的医疗部门来说，通过将信息提供给他们，从而获得他们的投资与支持；另一方面针对使用小程序的用户及销售防治鼻炎商品及药品的商家来说，通过提供“积分抵折扣”的形式向用户售卖产品，通过出租广告位及招商合作的形式，与相关商家达成商品销售的内部合作价，从而来满足用户“积分抵折扣”购买商品的需求，还可以商定一个能使双方互利互赢的广告位出租价，为想要承租广告位进行防治鼻炎商品宣传推广的商家提供广告宣传，利用最便捷的微信小程序实现了六大核心功能模块。最后，通过使用Postman和LoadRunner，验证了系统的可用性和实用性。

参考文献：

- [1] 贾志杰. Vue+Spring Boot 前后端分离开发实战[M]. 北京:清华大学出版社,2021.
- [2] 陈银萍. 基于人工智能中的图像识别技术的分析[J]. 信息与电脑(理论版),2019(1):165-166.
- [3] 李赫. 人工智能中的图像识别技术分析[J]. 无线互联科技,2021,18(17):93-94.
- [4] 王伟,常庆丽,吴朝霞. 基于Springboot+Vue的精准帮扶系统研究与实现[J]. 河南科技,2021,40(27):12-14.
- [5] 赵文欣,李华,云晖,等. 基于Spring MVC架构的档案管理系统设计与实现[J]. 内蒙古工业大学学报(自然科学版),2018,37(6):448-459.
- [6] 沈皎佳. “互联网+”背景下微信小程序的运用研究[J]. 信息记录材料,2021,22(6):119-120.

(下转第93页)

文章编号: 1007-1423(2024)04-0088-06

DOI: 10.3969/j.issn.1007-1423.2024.04.016

基于 Web 的异常数据检测系统设计与实现

常 鑫, 田立勤*

(华北科技学院计算机学院, 廊坊 065201)

摘要: 针对大数据处理中数据来源复杂, 且在清洗过程中存在较多相似的问题, 开发了一款异常数据检测系统。系统使用流行的 Python 编程语言和轻量级的 Flask Web 框架, 能够简化数据处理工作中的重复流程, 极大减轻数据分析的工作量, 提高工作效率, 利用系统内置的异常检测算法和用户上传的自定义算法完成对数据中异常的检测工作。

关键词: Web; Python; Flask; 异常检测

0 引言

在对物联网采集的水质和空气等类型数据进行分析的过程中, 需要对相同数据进行清洗工作, 发现存在的异常数据^[1]。异常数据包括空白数据、由传感器故障导致值连续不变的数据、不符合现实规律的数据(如河水水温达到 200 摄氏度在正常自然环境下是不现实的)等构成简单的异常类型, 也有由数据多个维度属性共同影响的异常数据, 不符合时间规律的异常数据等构成的复杂异常项。前者的检测相对简单, 只需对数据进行数值大小等简单对比就能发现, 后者影响因素较多, 隐藏的规律通过人工很难发现, 需要运用统计分析、机器学习等多个学科知识和技术对数据进行系统分析才能判断。

除了物联网数据外, 网络安全、物资管理、财务管理等多个领域均需对数据进行异常检测^[2]。由于领域不同, 需要用到的算法和标准也不同, 但大部分基础操作是相同的。开发一款通用的异常软件进行数据处理能大大减少异常检测的工作量, 方便各行业人员进行异常数据检测工作。

随着信息技术的不断进步, Web 开发技术也不断更新, Python 作为近几年流行的程序开发语言, 虽然运行速度相对 Java 等传统编程语言略慢, 但凭借强大的软件生态环境和语言本身的特点, 能够提供远超 Java 等编程语言的软件开发速度。基于 Web 开发的应用程序能够为多人团队提供更好的协作配合功能, 且 HTML 和 JavaScript 等前端技术的不断发展, Web 程序的 UI 设计和开发简单便捷, 功能强大, 基于 Web 平台是进行综合系统开发的优秀载体^[3]。

本文使用 Python 和 Flask Web 框架进行 Web 开发, 设计并实现能够对数据进行清洗和异常检测的通用系统。

1 系统功能需求分析

1.1 数据上传和存储

1.1.1 多种载体数据上传

能够存储数据的文件格式和数据表现形式不尽相同, 本系统需要对数据的表现形式做出规定, 接受以 Excel 表格、CSV、JSON 等多种常见文件格式存储的数据并进行处理。

收稿日期: 2023-11-27 修稿日期: 2023-12-18

基金项目: 河北省重点研发计划项目(19270318D); 河北省物联网监控技术创新中心项目(2156793H); 中央高校基本科研业务费资助项目(2142021009); 青海省重点实验室(2017-ZJ-Y21)

作者简介: 常鑫(1994—), 男, 河北廊坊人, 硕士研究生, 研究方向为数据异常检测; *通信作者: 田立勤(1970—), 男, 陕西定边人, 教授, 研究方向为数据异常检测, E-mail: 1561753947@qq.com

1.1.2 数据存储

不同行业、不同领域需要处理的数据所具有的维度不尽相同，常见的SQL数据库中单一表结构无法直接存储形式多变的数据，需要在数据存储时能够动态确定数据表结构，实现对不定维数、不定属性数据的存储和提取。

1.2 简单数据异常检测

在数据上传后，需要对数据进行简单数据异常检测，包括空值检测、连续异常检测、超界异常检测等多种形式，并能在检测后对数据和数据参数列进行标记。

其中连续异常和超界异常需提供能让用户自主设定数据上下界、连续次数阈值等参数的功能，以应对不同领域、不同学科的要求。

1.3 算法上传和调用

数据可能出现的异常多种多样，能够进行异常检测的算法层出不穷，提供算法上传和调用的功能，用户就可以编写新算法对系统中已上传的数据进行检测，也可以重新实现系统中已有算法，重新设定算法参数，对用户数据做定制，实现更高性能或准确性。

1.4 快速部署

Web程序虽开发便捷、可拓展性良好、便于团队协作，但相对于桌面程序，Web程序部署却较为复杂，需在物理机上安装众多依赖软件，且大多依赖软件只提供特定操作系统的发行版本^[4]。另Python语言本身有丰富的库能够提供多种功能，但软件社区较活跃，Python类库和软件包更新较频繁，有时新版本的包对旧版本功能支持不好，易造成包调用错误^[5]。上述原因导致没有计算机相关背景的用户使用起来较困难，因此需实现一种机制，使得软件能在不同平台的操作系统上快速进行部署。

1.5 数据导出

在数据分析工作中，数据清洗和数据异常检测往往并不是工作的最终目的，是为接下来工作做准备，在之后工作中可能还需对数据进行更深入分析，寻找数据中存在的规律来指导生产生活，或使用数据生成图标图像等作数据可视化展示。本系统提供数据导出功能，用户可以根据自身需要选择数据并以指定文件格式导出，进行后续数据分析工作。

2 系统功能实现

2.1 数据上传与存储

2.1.1 数据表现形式

在充分考虑系统编码复杂度和系统易用性后，对最终接受的数据表现形式做出规定，具体见表1，第一行为数据各列参数名，之后每一行为一条数据的形式。

Excel和CSV数据本身符合数据表样式要求，只需稍作调整即可，而JSON格式数据要先将数据转换为二维数组，且第一行数组各元素为每列列名，即可符合系统要求的数据格式。

2.1.2 多文件格式上传

系统在上传阶段要处理的数据文件格式形式多样，在之后升级时还有可能增加新文件格式，如果为每种文件格式都编写对应的上传代码会产生大量重复代码，使系统代码难以阅读与维护，也浪费大量开发时间^[6]。

为了解决这一问题，使用流行的Python数据分析包Pandas中提供的DataFrame数据类型作为数据中间形式，如图1所示，在用户上传数据后，系统根据用户上传的文件格式将数据转换为DataFrame类型，再使用统一的持久化方法将数据存入数据库中，完成数据上传功能。

表1 数据表现形式

水温(°C)	pH(无量纲)	溶解氧(mg/L)	电导率(μS/cm)	浊度(NTU)	高锰酸盐指数(mg/L)	氨氮(mg/L)
6.0	8.25	15.19	1643.4	3.8	4.40	0.002
5.8	8.26	15.39	1631.4	3.7		

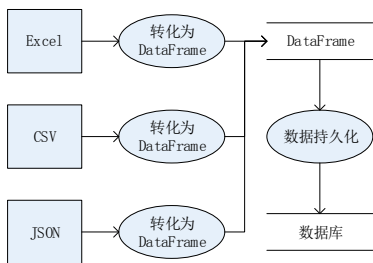


图 1 数据上传数据流程图

2.1.3 数据库表设计

为实现数据表结构的通用性，系统需要根据用户上传的数据动态生成表结构，SQL 数据库的单张表无法满足需求^[7]。目前可行方法有使用 Redis 这一类非关系数据库，或对数据库结构进行特殊设计，使数据表能够满足要求^[8]。前一种方法使用较简单，但 Redis 等非关系数据库并不支持 SQL 语句对数据进行结构查询，且往往将数据存储在内存中，若发生断电等意外情况将会丢失数据，还需在物理机上安装专门的服务，增大了系统耦合性；后一种方法虽在设计阶段比较复杂，且产生部分冗余数据，消耗一定存储空间，但在后续数据处理中能够方便使用 SQL 语句对数据进行查询，数据安全性较高，因此选择后一种解决方式。

设计后的数据库结构如图 2 所示，通过多对多关系来关联数据及其对应参数，使参数数量能够实现动态变化，数据和异常类型也建立多对多的关系，主要考虑到一条数据中可能出现多种异常，建立多对多关系方便进行查询和操作，也能更方便查询异常类型表，对出现某类异常数据进行查询。

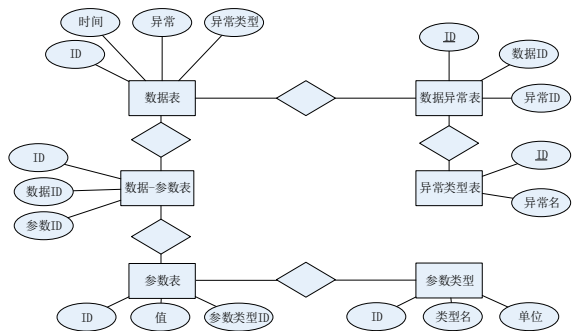


图 2 数据库 ER 图

2.1.4 上传流程

系统界面如图 3 所示，当用户点击原始数据导入后进入项目选择页面，该页面如图 4 所示，在该页面中用户选择要将数据上传到具体项目，点击上传进入后续页面，或可点击创建项目按钮跳转到项目创建页面创建新项目后再进行上传。之后用户进入文件格式选择页面，如图 5 所示，在该页面用户选择要上传的数据类型，目前提供 Excel、CSV、JSON 三种文件格式。

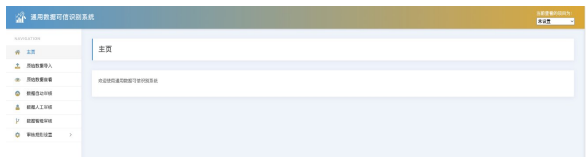


图 3 数据库 ER 图

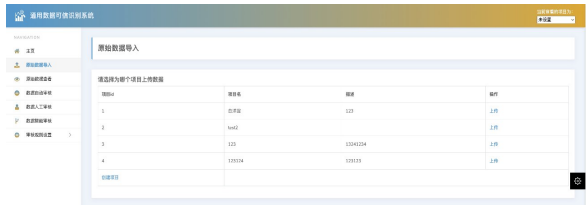


图 4 项目选择页面

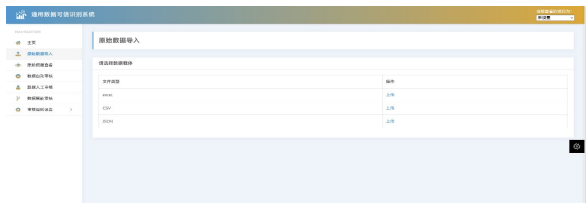


图 5 文件格式选择页面

用户完成文件格式选择后进入正式上传页面，如图 6 所示，用户选择数据文件后点击上传即开始上传数据。

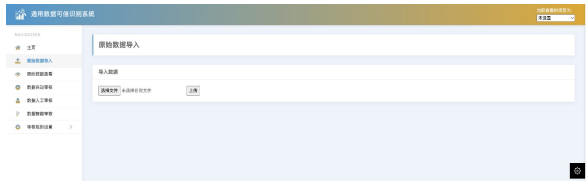


图 6 文件格式选择页面

如果该项目是首次上传数据，将进入如图 7 所示的参数设定界面，可以设定数据上下限和连续异常检测的连续次数阈值。结束用户设置完成后，即结束上传流程，系统将会按照用户

设定的参数对数据进行简单异常检测，最后返回如图8所示的页面，提示用户上传完成。

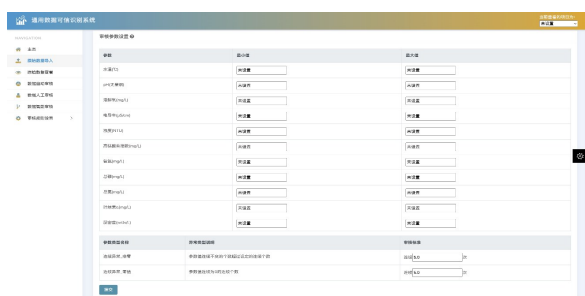


图7 初始参数设定页面

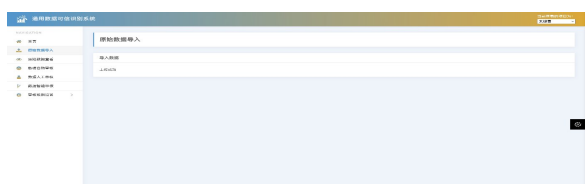


图8 数据上传成功页面

至此数据上传流程完成，用户可以点击系统导航条中原始数据查看、数据自动审核、数据人工审核三个按钮进入相关页面查看上传的数据。其中原始数据查看页面查看的是未经标注的数据，数据以简单的表单形式进行展示。数据自动审核页面中，如果某条数据的某一元素被系统检测为异常，该条数据的该元素在表中底色将变为红色，如图9所示。可通过点击数据后的查看按钮查看该数据所有出现异常的列和出现的异常。数据人工审核页面和数据自动审核页面基本相似，只是添加了筛选功能供用户筛选数据。查看按钮更改为修改按钮，允许用户对数据进行修改。

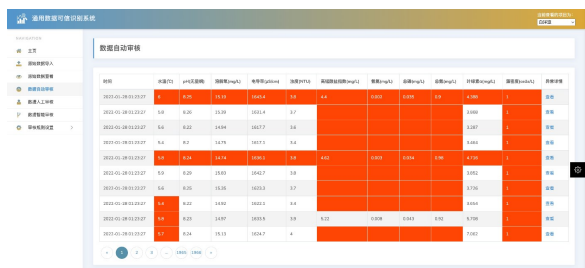


图9 数据自动审核

2.2 自定义算法的上传与调用

2.2.1 算法上传与调用流程

算法上传与调用流程如图10所示，首先用

户在算法上传页面填写算法信息，页面如图11所示，之后进入算法文件上传页面，用户选择要上传的算法文件，算法信息存入数据库的算法表，算法文件存入用来存放算法文件的文件夹，算法上传流程结束。

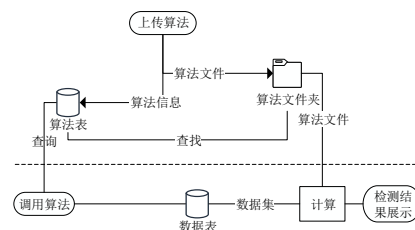


图10 自定义算法上传和调用流程图

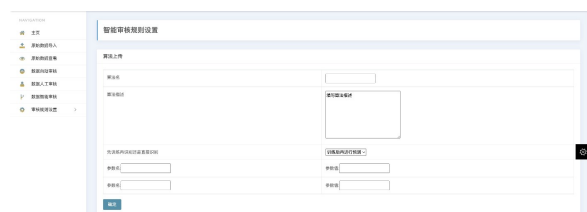


图11 算法信息填写

用户完成算法上传流程后，便可通过点击页面导航栏的数据智能审核按钮进入智能审核页面，页面如图12所示，在页面中用户可以选择已经上传的算法，然后再选择要应用的算法进行项目的分析，项目中的哪几列需要进行分析，然后点击开始就会开始分析。



图12 智能审核结果页

这时系统将会在数据库算法表中查找相应算法信息，然后根据算法文件名在算法文件中提取相应算法文件，在数据表中提取出用户选定项目的所有数据，最后应用选定的算法对数据进行分析。分析结束后，将会自动进入到审

核结果页,如图12所示。

审核结果页显示检测的数据量、发现的异常数据量、异常数据所占比例等数据,并显示一个饼图,让用户更加直观地看到该项目中异常数据所占比例。

2.2.2 算法模板

智能算法能够调用,主要依赖Python作为解释型编程语言的特性。这使得程序能够在运行过程中使用Python内置的import函数动态地加载Python包,以此实现调用用户自定义算法的目的^[9]。但在算法调用时,依然对算法文件有一些要求。要求用户在算法文件中按照提供的模板类,实现一个Algorithm类,在类中设定几个必要参数,并运用用户自己编写的实现remove_empty, full_result, fit_predict, fit, predict等方法,这样系统就能够顺利执行用户上传的算法,且不需要对用户算法本身做出改变,降低系统耦合性。

2.3 快速部署

为了简化系统部署流程,降低使用系统的门槛,使用容器技术对底层系统进行封装,选用容器软件为Docker, Docker容器能够让软件不受操作系统和硬件平台限制,运行在所有安装了Docker软件的计算机上,由Docker根据硬件系统和操作系统自动按照用户提前编写的Docker配置文件在容器中安装、运行软件,且和物理机本身操作系统做软隔离,保证安全性^[10]。使用Docker后,系统架构如图13所示。

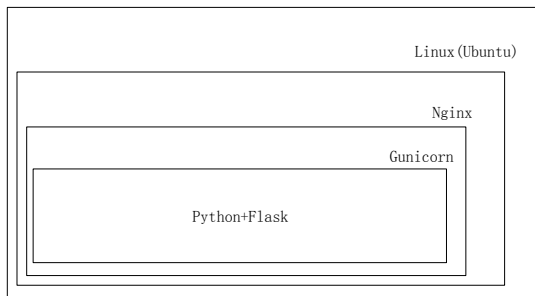


图 13 软件架构图

编写的Docker文件如图14所示,详细描述了运行系统所要安装的所有依赖软件,并在运行后自动将代码复制到容器中运行。在配置好

Docker文件后,用户部署系统只需使用Docker调用编写好的Docker文件就可快速将系统部署于物理机上。

```
FROM ubuntu:20.04
ENV TZ=Asia/Shanghai
RUN apt-get update
RUN apt-get install -y python3 python3-pip python3-dev
RUN pip3 install --no-cache-dir flask gunicorn
COPY . /app
WORKDIR /app
EXPOSE 80
CMD ["gunicorn", "-w", "flask", "-b", "0.0.0.0:80"]
```

图 14 Docker 文件

3 结语

本系统使用Python和Flask语言编写,对以多种文件格式上传的多维数据初步处理并进行异常检测,减轻异常处理过程中的重复工作,并与Docker软件配合,应用容器技术实现了系统的快速部署功能。

参考文献:

- [1] 余晓帆,朱丽青.基于Flask框架的社交网站数据爬取及分析[J].微型电脑应用,2022,38(3):9-12.
- [2] 杨威超.数据驱动的物联网安全威胁检测与建模[D].郑州:战略支援部队信息工程大学,2019.
- [3] 吕丰秀. HTML5技术在移动Web前端设计中的应用[J].电子技术,2022,51(12):222-223.
- [4] 刘进学,徐诗理,曾凡平.基于Web的实验室信息管理平台设计与实现[J].电脑知识与技术,2022,18(16):57-59.
- [5] 徐园,梅勇,龚俊,等.基于改进的OCSVM算法的工控网络异常检测算法[J].兵工自动化,2022,41(4):49-52.
- [6] 杜雨荃,王晓菊,田立勤.基于微信小程序的网上购物系统的设计与实现[J].网络安全技术与应用,2022(4):60-62.
- [7] 杜继明. Python语言的Web开发应用分析[J]. 数字通信世界,2022(1):76-78.
- [8] 宋永生,黄蓉美.基于Python的学习成绩分析及可视化分析[J].信息技术与信息化,2021(12):100-102.
- [9] 王锐.一种基于Grubbs和孤立森林的多维序列异常检测方法[J].移动通信,2022,46(4):55-59.
- [10] 谢兆贤,曹香美,王超.基于Docker容器的快速开发网页服务器[J].计算机系统应用,2022,31(4):99-109.

Design and implementation of abnormal data detection system based on Web

Chang Xin, Tian Liqin^{*}

(School of Computing, North China Institute of Science and Technology, Langfang 065201, China)

Abstract: Aiming at the complexity of data sources in big data processing and the existence of more similar problems in the cleaning process, an anomaly data detection system is developed. The system uses the popular Python programming language and lightweight Flask Web framework, which can simplify the repetitive processes in the data processing work, greatly reduce the workload of data analysis, improve work efficiency, and complete the detection of anomalies in the data using the built-in anomaly detection algorithms and custom algorithms uploaded by the user.

Keywords: Web; Python; Flask; outlier detection

(上接第87页)

Research and implementation of wormwood intelligent recognition applet based on Spring Boot

Zhao Wenxin^{1*}, Su Feng²

(1. School of Computer and Network Security, Inner Mongolia Electronic Information Vocational and Technical College, Hohhot 010010, China; 2. Inner Mongolia Autonomous Region Meteorological Service Center, Hohhot 010051, China)

Abstract: Wormwood intelligent recognition applet is based on the fusion application of big data analysis and intelligent image recognition technology. It integrates six functions of wormwood intelligent recognition, user information reporting, data visualization, knowledge popularization, online community and rhinitis prevention and control information pushing. At the same time, it develops a big data visualization platform for the spatiotemporal distribution of wormwood plants. Thus, it can help the government and relevant departments to better manage wormwood, solve ecological and livelihood problems, and help allergy patients who are troubled by wormwood to prevent in advance the intensive period of wormwood distribution and diffusion.

Keywords: image recognition; wormwood intelligent recognition; data visualization; rhinitis prevention and treatment

文章编号: 1007-1423(2024)04-0094-05

DOI: 10.3969/j.issn.1007-1423.2024.04.017

基于工业互联网的能碳数字化和智能化建设

张 伟, 安振文, 刘子强*

(中海油信息科技有限公司研发中心, 天津 300450)

摘要: 通过对接边缘侧数据采集服务, 实时获取设备参数信息, 实现能源管理的采集、监控、统计和分析, 并能结合各能源使用情况, 实时计算汇总企业碳排放量, 帮助企业摸清碳家底。分析能源和碳排放的历史数据, 利用变量灰色模型建立自变量的预测值, 建立能源灰色预测 GM(1,1) 回归模型, 预测数据为决策者提供决策支持, 从而实现企业能源和碳排放管理的数字化和智能化。

关键词: 灰色预测; 能源管理; 碳排核算

0 引言

我国 2020 年提出碳达峰和碳中和的“双碳”战略, 为实现 2030 年“碳达峰”目标, 未来低碳经济市场下, 产品的碳排放量将是企业未来缴纳碳税的重要依据, 因此, 企业建立能源和碳排放管理体系, 监控统计产品使用的能源使用量, 根据碳排放核算指南^[1]来核算产品的碳排放量, 掌握企业生产产品的碳排放量, 为企业未来参与国际市场竞争提供数据支持。

目前, 企业已建设了相对完善的工业互联网, 能碳系统通过工业互联网获取能源数据, 实现能源的自动采集, 并结合符合化工行业碳核算标准, 汇算能源对应的碳排放量, 遵循科学规范、技术先进、易扩展、易维护、通用性等原则, 采用在线方式广泛接入生产过程中能源相关的指标数据, 并利用预测算法, 实现能源和碳排放监控预警。企业通过能碳数据的协同融合, 完善企业能源管理体系框架, 建立碳排放监控核算预测流程, 实时监测和强化跟踪能源消耗和碳排放状态, 协助企业辨识节能降碳机会, 协助管理者合理制定节能降碳目标, 进一步压降企业成本, 促进企业可持续发展。

1 关键技术

1.1 工业互联网

工业互联网把工业和互联网紧密联合在一起, 实现了人、机、环境的互联。学者和专家也对工业互联网有着不同的理解, 能碳系统基于现有的工业互联网架构, 从信息安全角度将网络分为生产网和办公网, 并结合云计算、大数据和 SaaS 服务理念, 实现数字和设备一体化, 融合 IT 和 OT, 动态反馈工业环境, 构建“工业+互联网+安全生产”的新型安全生产模式, 连接人、机、物、法、环五个生产的关键要素。在集团、企业、工厂不同层面, “工业+互联网+安全生产”有不同侧重点的融合应用, 工厂层属于生产现场网络, 采用有线网络与无线网络结合使用, 并使其隔离于其他网络, 使用交换机与办公网络相连, 并采用信息安全技术对数据进行隔离。企业层面侧重数字化管理设备、工艺等要素。根据“工业+互联网+安全生产”的内涵以及安全生产实际, 采用专线和私有云的方式保证企业数据安全, 使企业工业互联网具有快速感知、实时监测、超前预警、动态计

收稿日期: 2023-12-21 修稿日期: 2024-01-10

作者简介: 张伟(1978—), 男, 河南焦作人, 硕士, 高级工程师, 研究方向为工业互联网; 安振文(1979—), 男, 河北沧州人, 硕士, 高级工程师, 研究方向为工业互联网; *通信作者: 刘子强(1992—), 男, 河北石家庄人, 本科, 工程师, 研究方向为工业互联网, E-mail: ex_liuzq3@cnooc.com.cn

算等能力，其中快速感知是利用智能传感、边缘计算等技术全面采集现场信息，实时监测是私有云共享数据，超前预警可以根据预测模型进行预测和预判。工业互联网架构如图1所示。

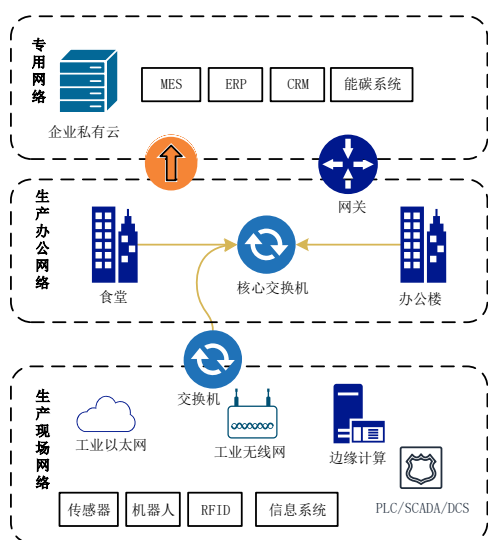


图1 工业互联网架构

1.2 微服务设计

为保证整个数据体系的生命力，采用DDD (Domain-driven design) 领域设计^[2]，通过“领域分析”“领域建模”统一建模语言，降低业务人员和开发人员沟通难度，提升敏捷开发效率。根据行业特性、经验、工艺等信息划分业务边界，聚焦业务范围，明确模糊边界，依据单一职责和开闭设计原则，将能碳系统划分为能源中心主题域、碳核算中心主题域、碳资产中心主题域、主数据中心主题域、碳基础中心主题域，如图2所示。各个主题域间采用轻量级基于HTTP的RESTful API通信机制进行数据交互，

每个主题域之间数据尽量互不干扰，用户中心统一权限管理和身份角色管理，每个服务和主题域都可单独选择技术架构，降低服务之间的耦合度，方便服务维护和扩展。

1.3 容器化部署

随着分布式系统和云计算的飞速发展，微服务和容器的应用越来越广泛，通过将微服务容器化实现自动化部署和持续集成。容器与底层共享一个操作系统，共享网络等，省去了管理程序的额外开销，降低了系统负载^[3]，在同等条件下可以支持更多的应用运行，从而可以充分利用系统资源。容器内可以部署多应用，并且多应用之间不会相互影响，高效利用硬件资源。使用k8s容器编排实现应用的自动扩容和减缩。基于容器技术的交付流程，可以加快微服务的落地，更好地以应用为中心实现应用的快速部署，无缝切换。

微服务容器化部署主要分为微服务开发、微服务容器镜像构建、微服务容器镜像管理、微服务容器编排管理四个步骤。通过使用容器技术，部署多个微服务的工作将变得更为简便，实现微服务一键化部署。部署流程如图3所示。

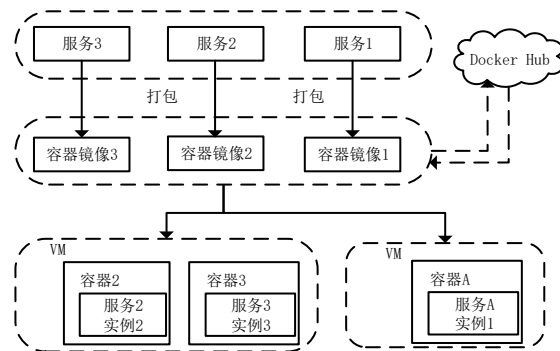


图3 微服务容器化部署流程



图2 业务领域设计

1.4 灰色预测模型

灰色系统理论是一门应用数学学科，它主要研究的是信息部分清楚、部分不清楚并带有不确定性现象。传统的系统预测理论大都针对信息较为清楚的领域，但是在现实世界中存在部分已知、部分未知的系统。目前为止，对部分信息清楚，部分信息不清楚的系统研究较少。可能正是由于这种原因，灰色系统理论的发展获得了足够的重视和关注。

通过对灰色预测模型GM(1,1)模型进行分析，找到了GM(1,1)模型的不足之处，并对其进行了改进。改进之处具体包括引入了新型矫正算子和对背景值的改进。引入新的矫正算子后，实际上它也是一种数据处理的方式，使得数据具有了很好的光滑性，同时体现了新信息优先的原理。对背景值进行改进后使得模型的预测精度有了很好的提高。

2 能碳数字化和智能化建设方案

能源和碳排放融合目前还属于创新融合阶段，能源管理平台分为能源数据采集、能源监控、能源管理、能源模型等模块，形成能源采集、统计、监控和分析的能源管理体系，在此基础上，依据能源碳排放标准计算方法，建立碳核算中心、碳资产管理中心，自动核算碳排放量，管理企业碳资产，实现能源和企业碳排放的数字化和智能化管理。

2.1 功能设计

能碳系统分为能源管理和碳排放管理两个部分，并结合其他现有系统的现有数据，能源管理将工厂层的生产、办公所产生的能源信息

记录并进行统计分析，功能包含仪器仪表数据管理、测点管理、能源统计等；双碳管理部分功能包含排放源和能源管理、排放因子、活动水平、核算公式等，其中，基础数据的部分数据如设备管理、生产计划、用户管理由其他三方服务或者用户中心进行获取。最后，整合历史数据，对未来的能源和碳排放进行预测，实现能源和碳排放的数字化和智能化管理，整体功能如图4所示。

2.2 能碳融合

在能源管理中，能源种类根据国标《能源分类与代码》(GB/T 29870—2013)进行初始化设置，以标准化的能源种类进行管理，同时，排放源的品种即能源种类，通过能源管理中能源种类统计数据，可以直接计算出对应能源种类的碳排放量，进而得出企业总体的能源使用量和碳排放量，能碳数据融合管理后，为企业直观地展示出企业使用的能源量和碳排放量，并结合碳市场管理企业碳资产，为企业决策者提供数据支持，助力实现企业双碳目标。

2.3 能碳“一张图”

能源使用量与企业碳排放量是呈正比的关系，规划“一张图”进行“能碳”数据展示，如图5所示。通过可视化技术、图表展示方式，辅助管理者在“一张图”上直观、清晰地了解到能源消耗情况、企业碳排放量、能耗消耗趋势和碳排放量趋势，让管理者心中有帐，为实现“碳达峰”“碳中和”双碳目标打好坚实的基础。

2.4 能碳预测

为了协助企业管理者更好地掌握未来能源

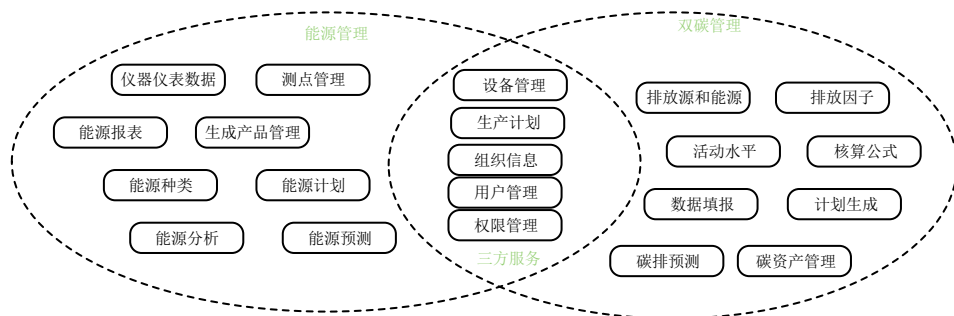


图4 功能设计



图5 能碳“一张图”

消耗量和碳排放量，合理安排生产和更高效地经营，需要预测未来一定时期内的能源消耗量和碳排放量，通过对企业生产经营的总结和归纳，采用灰色预测模型GM(1,1)进行预测，具体建模步骤^[4]如下：

(1) 确定预测系统中的原始序列

$$X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), \dots, X^{(0)}(n)\} \quad (1)$$

(2) 进行一次累加得到累加生成数列：

$$X^{(1)}(k) = \sum_{i=1}^k X^{(0)}(i), (k = 1, 2, \dots, n) \quad (2)$$

(3) 均值生成序列 $Z^{(1)}(k)$ 为

$$Z^{(1)}(k) = \frac{X^{(1)}(k) + X^{(1)}(k-1)}{2} \quad (3)$$

(4) 利用均值生成序列 $Z^{(1)}(k)$ 来构造矩阵：

$$\begin{bmatrix} -Z^{(1)}(2) & 1 \\ -Z^{(1)}(3) & 1 \\ -Z^{(1)}(4) & 1 \\ \dots & \dots \\ -Z^{(1)}(k) & 1 \end{bmatrix} \begin{bmatrix} a \\ u \end{bmatrix} = \begin{bmatrix} X^{(0)}(2) \\ X^{(0)}(3) \\ X^{(0)}(4) \\ \dots \\ X^{(0)}(k) \end{bmatrix} \quad (4)$$

(5) 采用最小二乘法求解参数向量：

$$\begin{bmatrix} a \\ u \end{bmatrix} = (X^T X)^{-1} X^T Y \quad (5)$$

(6) 将第(5)步中参数向量值代入下面公式：

$$X^{(1)}(k+1) = \left(X^{(0)}(1) - \frac{B}{A} \right) e^{-Ak} + \frac{B}{A}, k = 0, 1, 2, \dots \quad (6)$$

(7) 对一次累加生成数列的预测值进行一次累减生成，得到累减序列：

$$X_{\min us}^{(1)} = [X_{\min us}^{(1)}(1), X_{\min us}^{(1)}(2), \dots, X_{\min us}^{(1)}(n)] \quad (7)$$

原始数列的最终预测值：

$$X_{pred}^{(0)}(k) = X_{\min us}^{(0)}(k) - X_{\min us}^{(0)}(k-1) \quad (8)$$

(8) 相对误差检验：

$$\lambda(k) = \frac{|X^{(0)}(k) - X_{pred}^{(0)}(k)|}{X^{(0)}(k)}, k = 1, 2, 3, \dots, n \quad (9)$$

当 $\lambda(k) < 0.2$ 时，可认为模型达到了一般要求；

当 $\lambda(k) < 0.1$ 时，可认为模型达到了较高要求。

(9) 第(8)步通过后，通过模型对未来数据进行预测。

为验证 GM(1,1) 模型的准确性和可行性，现以国家统计局发布的数据进行预测。

为了数据的准确性和可靠性，拟研究的数据来自于国家统计局 2015—2021 年中国天然气(万吨标准煤)。使用 2015—2019 年天然气的使用量，预测 2000 年和 2021 年，见表 1。

表1 天然气使用量统计数据

年份	使用量(万吨标准煤)
2015	25178.554
2016	26931.012
2017	31452.063
2018	35866.3
2019	38999.04
2020	41858.376
2021	46636

根据上述步骤(1)，形成 2015—2019 年天然

气的原始队列:

[25178.554, 26931.012, 31452.063, 35866.3, 38999.04]

通过级比公式校验数据是否符合指数规律, 公式如下:

$$\sigma(k) = \frac{X^{(0)}(k-1)}{X^{(0)}(k)} k = (2, 3, 4, 5, \dots, n) \quad (10)$$

比值结果为:

[0.93492788, 0.85625582, 0.87692522, 0.91967136]

校验在合理区间(0.36787944117144233, 1.0), 比值结果均在区间内, 用此原始序列可以预测。

根据步骤(2)和(3)对原始序列进行累加生成序列, 结果为

[25178.554, 52109.566, 83561.629, 119427.929, 158426.969]

并产生均值序列

[38644.06, 67835.5975, 101494.779, 138927.449]

再对上述值进行最小二乘法计算, 得出矩阵参数[A, B]值, 结果为

[-1.20556228e-01, 2.28568080e+04]

将 k 代入白化方程的时间响应式, 得到一次累减序列为

[25178.554, 52696.18448675, 83739.48858328, 118760.19030007, 158267.89041226, 202837.48187208]

根据公式计算预测序列为

[27517.63048675, 31043.30409653, 35020.7017168, 39507.70011218, 44569.59145983]

最后, 使用相对误差校验公式校验预测结

果与实际结果的误差值为

[0.02178227, 0.01299625, 0.0235764, 0.01304289]

可以直观地看出, 数据集合中的数据均小于0.1, 达到了较高的水平, 可以对天然气使用量进行预测。那么, 预测2020和2021这两年的数据为[44569.59145983, 50280.03344298], 与实际的误差分别是0.065和0.078, 在可接受范围内, 是可信的预测信息。

3 结语

节能减排、绿色环保是可持续发展战略的重要组成部分, 结合能耗和碳排放两个概念, 利用成熟稳定的工业互联网、云计算等技术, 通过对能耗数据的收集、检测、统计、分析和评估, 再利用灰色预测算法进而实现对碳排放的检测分析, 最终完成能碳融合的数字化和智能化平台。该平台能够协助企业实现能源管理, 摸清碳家底, 制定决策, 实现了能源节约、成本降低, 增加了企业的经济及环境效益。

参考文献:

- [1] 国家发展和改革委员会. 中国化工生产企业温室气体排放核算方法与报告指南(试行)[S/OL]. <https://www.gov.cn/gzdt/att/att/site1/20131104/001e3741a2cc13e13f1c04.pdf>.
- [2] 董向阳. 领域驱动设计DDD在IT企业内部网站开发中的运用[J]. 电脑知识与技术, 2022(10): 44-47.
- [3] 周跃勇, 周施文, 李水龙, 等. 地震预警系统的容器化部署[J]. 现代信息科技, 2023(1): 87-89.
- [4] 马萍. 灰色系统GM(1,1)模型的改进及其灰色统计模型研究[D]. 长春: 吉林大学, 2007.

Digital and intelligent construction of energy and carbon based on industrial internet

Zhang Wei, An Zhenwen, Liu Ziqiang*

(CNOOC Information Technology Co., Ltd., Research and Development Center, TianJin 300450, China)

Abstract: Through the data collection service at the docking edge, obtained equipment parameter information on real time, collection, monitoring, statistics, and analysis. Additionally, calculation and summarization of carbon emissions according to collected data. Helping businesses understand their carbon footprint. Applying grey theory GM (1,1) model, analyzing historical data and forecast energy data. Policy maker can use the forecast data. Improve the enterprise energy and carbon emission management level of digitalization and intelligence.

Keywords: gray forecast; energy manager; carbon accounting

文章编号: 1007-1423(2024)04-0099-06

DOI: 10.3969/j.issn.1007-1423.2024.04.018

嵌入式实验室门禁人脸识别系统设计

翟 丽^{1*}, 邹琦璋²

(1. 南京审计大学实验中心, 南京 211815; 2. 南京审计大学经济学院, 南京 211815)

摘要: 为加强对实验室门禁系统的控制, 降低门禁程序的误检率, 对嵌入式实验室门禁人脸识别系统进行设计与验证。简述系统的构建需求, 有针对性地进行嵌入式系统总体结构设计。基础电路中接入 0K6410 核心板, 关联前后识别端口的同时, 植入图像采集装置, 完成系统硬件的设计; 建立嵌入式交叉识别编译环境, 对所识别的图像预处理并设计定位指令, 关联 ORL 嵌入式人脸数据库, 完成系统软件的设计。系统测试结果表明: 针对于静态检测识别环境和动态检测识别环境下的测试, 在 50 名用户的识别测验后, 得出的误检率均控制在 15% 以下, 表明在实际嵌入式人脸识别过程中, 识别误差小、可控、识别速度快、更易调整, 具有实际的应用价值。

关键词: 人脸识别; 嵌入式; 系统设计; 实验室门禁; 识别程序; 远程监测

0 引言

实验室门禁系统多被应用安装在大型的实验室外部, 主要用来保护重要实验仪器及实验数据, 同时也可以对实验室人员的日常工作情况进行限制, 在进出实验室的过程中, 需要以指纹识别、密码输入、证件识别的方式进行身份的验证, 避免非法人员潜入给实验室造成不可估量的损失^[1-2]。传统的实验室门禁系统对于人们的识别形式过于单一, 多为指纹识别, 虽然可以达到身份验证的目的, 但是存在一定的漏洞, 极可能出现识别误差, 或者身份误验, 甚至存在一些不法分子, 利用特殊的方式获取指纹, 替代验证, 以达到自身的目的^[3-4]。嵌入式人脸识别系统的覆盖范围相对更广泛, 控制方式也十分简洁, 具有一定的应用优势, 为此, 本文设计嵌入式实验室门禁人脸识别系统。

1 需求分析

传统的钥匙开门、密码开门、数字化验证开门虽然具有一定的特点, 但是在现代化的背

景之下, 缺乏一定的灵活性和稳定性, 安全程度也并不是十分高, 因此, 人们对于门禁系统的需求及标准逐渐提升^[5]。生物识别是目前最为安全且多元的一种高新技术, 自身的针对性较强, 基于生物特征对人脸做出动态化扫描, 对特征位置进行标记, 同时也具有非接触性、可靠性、准确率高等优点, 与传统的加密系统相对比, 很难伪造, 能够进一步确保实验室的安全^[6]。针对不同识别模块, 进行需求分析: ①视频监控需求: 对于来访者积极进行人脸扫描, 对捕获的人脸数据、信息集中存储, 并要求对实时画面及时刷新、更新。同时还需要增加图像采集功能, 对标定区域进行图像采集, 完成图像信息的存储。②电控锁调整需求: 当合法人员进入实验室时, 通过打开电控锁, 实现门禁控制功能, 并增加动态化识别的效果, 与报警程序相关联, 形成实验室门禁人脸识别结构, 具体如图 1 所示。根据图 1, 可以完成对实验室门禁人脸识别需求结构的设计及需求分析。以此为基础, 有针对性地设定系统的侵入需求设定。一旦系统识别到非法入侵行为, 会

收稿日期: 2023-09-19 修稿日期: 2023-11-03

基金项目: 南京审计大学教改课题(2022JG056)

作者简介: *通信作者: 翟丽(1984—), 女, 江苏大丰人, 硕士, 工程师, 研究方向为教育管理、实验室建设与管理, E-mail: wangwenjing8964@163.com; 邹琦璋(2001—), 女, 江西抚州人, 硕士研究生, 研究方向为数字经济

立即触发警报,以音频播放程序传输音频信号,并对异常的人脸进行存储,传输到管理人员的控制系统之中,同时与嵌入式的功能模块相搭接,实现需求目标的执行。

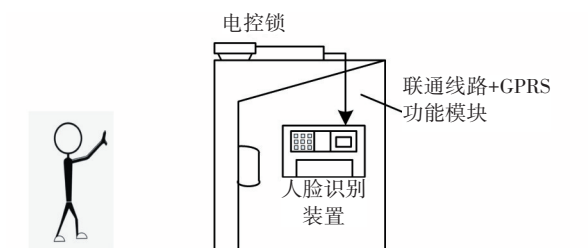


图 1 实验室门禁人脸识别需求结构示意图

2 嵌入式系统总体结构设计

结合实际的系统构建需求和标准,进行总体结构的设计。通常情况下,实验室的门禁系统由硬件和软件两部分组成。硬件一般包括摄像头、电控锁、远程控制装置、PC 装置、外部识别装置等;而软件一般包括识别程序、嵌入式存储数据库、人脸采集功能模块、匹配指令等。在实验室人员进行门禁识别过程中,系统会先采集人脸的图像,进行一系列处理后,将其与数据库预留的人脸资源匹配,根据所匹配的结果为实验室人员打开嵌入式电控锁,形成具有总体性的循环系统结构,具体如图 2 所示。

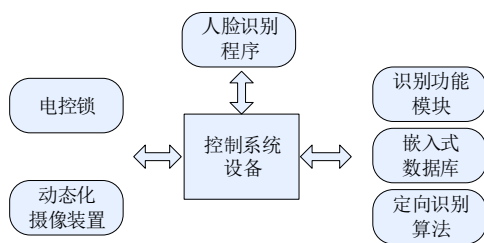


图 2 嵌入式系统总体结构示意图

根据图 2,可以完成对实验室门禁人脸识别系统总体结构的设计与调整。接下来,根据实验室的工作人数,调整图像采集摄像头的覆盖位置,将嵌入式识别程序与功能模块相连接,进行模数的动态化 A/D 转换,逐步转变为可识别的图像、数字形式,传输至对应的管理程序之中,完成对实验室门禁人脸识别系统总体结构的设计与调整。但是这部分需要注意的是,

针对于复杂的嵌入式识别程序,单一的图形采集、数据处理框架可能无法识别,可以在硬件中同时增加 CMOS 传感器及 CCD 传感器,扩大实际的识别覆盖范围,增加单元可识别的人像的数量和区域。在软件中,增加整体的嵌入式识别处理端口,设定多个电控锁开门指令通道,关联基础性的识别控制模块,增加对访问人员的排查和身份核定,完善嵌入式实验室门禁人脸识别系统总体结构的完整性和可使用性。

3 系统硬件设计

3.1 0K6410 核心板设计

根据实际的嵌入式人脸识别系统构建需求及标准,进行硬件设计,以 0K6410 核心板的构建为引导。首先,将 S3C6410 处理器接入系统的可识别电路之中,以 ARM11 作为主控制内核,联通 5G 网络,在主控电源的附近安装一个小型的加速器和监控装置,实现多元控制。构建核心板+底板的双层硬件控制结构,设定核心板的尺寸为 5 cm×6.5 cm,底板尺寸设定为 8.5 cm×12 cm,确保两板之间存在一定的空间,可以接入 3 个连接器,引脚需要布设 120 个,便于二次调整。以此为基础,在硬件系统中构建强电磁环境,采用 S3C6410 处理器将 0K6410 核心板和连接器关联在一起,设定系统的主频运行频率为 667~820 MHz。后续为增加系统在运行过程中的灵活性,还需要在 0K6410 核心板与主控电源之间,设定一个 4 层的复位键,主要针对于芯片与核心板的复位处理。在初始的接口处连入一个串口,采用 DM9000AE 网络芯片对异常信号进行捕捉,一旦出现问题,立即切断相关的硬件装置,形成一个稳定、安全的硬件控制电路,再加上 0K6410 核心板的双向控制,可以进一步增加硬电路的安全程度,逐步完善优化 0K6410 核心板的实际应用效果。

3.2 前后识别端口设计

在 0K6410 核心板的内侧位置建立一个小型的定向控制电路,接入 TVP5158 芯片,电压调整为 3.5 V,在 DM9000AE 网络芯片的接入端口处增设一个视频端口,将 TVP5158 引脚通过电压转换芯片进行控制,形成动态化的逻辑电压,

形成稳定、灵活的硬件执行环境。随后，遵循嵌入式原则，在电路中添加多个前后端同步识别端口，结合定位装置及 MDIO 模块作出控制，但是需要注意的是，针对于数据的处理及人脸图像的识别，一般不支持 DSP 端口的控制，所以需要选择嵌入式 MDIO 控制模块执行。

将物理层的芯片与单通道相互联通，接入一个自适应均衡器，时刻对芯片及前后识别端口的运行状态进行平衡控制，以此来维持硬件的日常运行。

3.3 图像采集装置设计

针对于嵌入式人脸识别系统，图像采集装置是最为重要且核心的硬件。使用 T9V136 芯片，将所设定的小型控制电路进行分接，布置红外光感光环境，对芯片的模拟电压进行调整，内核核片均由 I/O 电源进行双向控制。在电路中安装一个图像识别板，作为串行通信命令的过渡与承接。

在标定的范围之内，联通 SDATA 和 SCLK 装置，形成动态化的图像识别区域，将响应串行通信命令均衡输出，组成可见光感光 CMOS 图像识别程序，利用传感器对 SPI 端口和 FLASH 进行连接，形成调平的可识别范围，可以更为快速地采集高质量的近红外图像与显示图像，通过传感器进行过滤和传输，利用系统对图像作出分析、匹配，完成图像采集装置的设计和安装。

4 系统软件设计

4.1 嵌入式交叉识别编译环境建立

根据系统的构建需求及标准，进行系统软件的设计。为增强系统实际的应用情况，需要构建具有针对性的嵌入式交叉识别编译环境。这一类针对性较强的系统一般需要采用双向内核程序进行同步控制，设定软件的执行模式为开源。可以先构建一个等效的交叉识别协议，将源代码、信息资源以及基础性的指令进行交接和融合，实现系统控制最大的自由度。随后，遵循嵌入式的原则，测定系统的交叉识别覆盖范围，并进行基础识别指标参数的调整与设置。

在总控制程序中增加一个微处理协议，完

善系统的实际执行网络功能，以嵌入式控制结构作为引导，构建系统的交叉识别范围，并计算出识别编译移植距离，具体如公式(1)所示：

$$H = g + \sum_{i=1} 0.2k \times \frac{iw_2}{w_1 - \alpha g^2} \quad (1)$$

公式(1)中： H 表示识别编译移植距离， g 表示交叉编译机制运行差， k 表示嵌入范围， w_1 表示预设识别速度， w_2 表示实测识别速度， i 表示单向识别次数， α 表示标定移植距离。通过上述计算，最终可以得出实际的识别编译移植距离。结合得出数值，在实验室的门禁系统中划定对应的侵入式识别范围，并构建嵌入式的交叉识别引导程序，将其写到控制板卡上，以系统实际的需求对其进行裁剪、定制，营造交叉识别编译控制环境，以此来进一步增加系统识别的灵活性和可靠性，为后续的软件设计奠定基础。

4.2 图像预处理及定位指令设计

结合嵌入式人脸识别技术，对人脸图像进行预处理，并设计定位指令。首先，需要在可识别程序中建立一个工具链，按照嵌入式的处理步骤，将预处理的程序和协议逐步嵌入系统之中，与此同时，还需要将相应的数据包解压，全部拷贝到系统的识别程序之中。利用识别程序，对实验室人员的脸部特征进行识别，作出嵌入式标记之后，以待后续使用。但是这部分需要注意的是，人脸图像采集，一般是对人脸的特征进行采集，具体如图3所示。

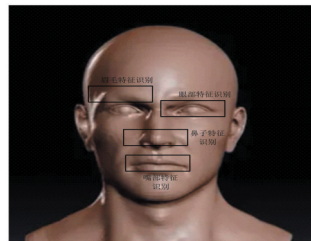


图3 人脸特征采集图示

根据图3，可以完成对人脸特征采集的设计与研究。以此为基础，结合嵌入式的处理技术，对所采集的人脸特征作出实际的标记，启动嵌入式操作系统内核的文件协议，对传输的图像进行分解预处理，设定可识别的帧数标准。根据识别特征，构建定位指令，由于人脸的特征

具有多样性和唯一性，所以在完成特征的提取之后，还需要对图像进行降维处理，并计算出特征空间的基值，具体如公式(2)所示：

$$N = \int o^2 do - \sqrt{\varepsilon \theta_1} \times \theta_1 \theta_2 \quad (2)$$

公式(2)中： N 表示人脸识别特征空间基值， o 表示降维差， ε 表示实际定位覆盖区域， θ_1 表示检测单向距离， θ_2 表示检测双向距离。通过上述计算，最终可以得出实际的人脸识别特征空间基值。结合嵌入式的识别程序，针对人的眼睛、瞳孔、鼻孔和嘴、人脸等作出同步定位，对重点区域进行定向识别匹配。将得出的人脸识别特征空间基值作为定位的标准，设定在指令之中，再搭配各个识别位置的特殊识别协议，形成同步多元化的嵌入式人脸识别定位指令群，设定在主控程序之中，完成对图像预处理及定位指令设计。

4.3 ORL嵌入式人脸数据库设计

为扩大系统实际的应用效果，需要进行ORL嵌入式人脸数据库的设计。通过接收指令进行模块串口的关联，将指令集群设定在嵌入式控制程序之中，增设查找和分析功能。同时，在标定的系统覆盖范围之内，构建ORL嵌入式动态化识别程序，以单向的字符串作为识别引导目标，转换所识别的图像数据形式，依据实际的分辨率，设定人脸数据库的范围，将所识别的图像转换成单幅图像的形式，原始人脸图像均为灰度级，且拍摄的视角、定位、尺寸均是不同的，举例来说：对于所有识别者表情的解析，形成集成人脸库，具体如图4所示。

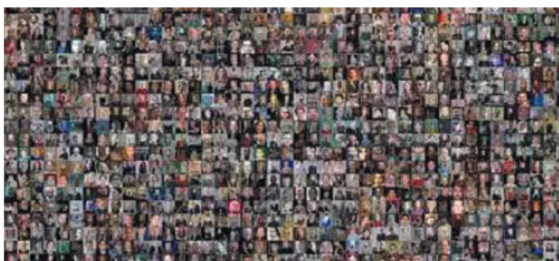


图4 ORL嵌入式人脸数据库呈现图示

根据图4，可以完成ORL嵌入式人脸数据库的设计。随后，在基础的识别程序之中，增加嵌入式的识别目标，与ORL人脸数据相关联，

形成动态化的识别采集执行程序，最终完成对ORL嵌入式人脸数据库的设计与关联。

5 系统测试

本次主要是对嵌入式实验室门禁人脸识别系统的实际应用效果进行分析和验证。考虑到最终测试结果的真实可靠，选择A实验室作为测试的主要目标对象，确保测试环境的单一稳定，在基础的门禁程序中进行多方向调整，实现嵌入式人脸识别程序的融合与设计，获取最终的测试结果。接下来，根据实际的测试需求，为系统搭建测试环境。

5.1 测试准备

在对嵌入式实验室门禁人脸识别系统的实际应用效果进行分析和验证之前，需要先搭建相应的测试环境。选择50名用户作为测试目标，将其中40人的人脸图像提前录入系统之中，并针对每一个人的人脸特征，进行识别标记。其他10人的人脸图像暂未录入系统之中。获取基础的可识别数据，根据嵌入式技术和AdaBoost算法，可以先测算出实际的识别均值，具体如公式(3)所示：

$$S = \sqrt{0.3\delta x_1} + x_2 \left(\mu x_1 - \frac{x_1 x_2}{t} \right) \quad (3)$$

公式(3)中： S 表示识别均值， x_1 表示识别基础范围， x_2 表示识别扩展范围， δ 表示嵌入转换差， μ 表示动态化识别误差， t 表示单向识别时间。通过上述计算，最终可以得出实际的识别均值，根据其制定系统的可识别标准，接入嵌入式结构的同时，营造一个可控、可调整的嵌入式实验室门禁人脸识别环境。核定测试的装置及设备是否处于稳定的运行状态，并确保不存在影响最终测试结果的外部因素，核定无误后，开始具体的系统测试。

5.2 测试过程及结果分析

结合实际的实验室门禁识别标准及需求，进行具体的测定分析。测试共分为两个阶段，分别是动态识别和静态识别，具体如下测定：

(1) 测试用户处于静态环境下的识别测试。将测试挑选的50名用户划分为5个测试小组，每组10人，顺序打乱，按照预设的识别流程进

行测验。完成之后,获取相对应的测试数据、信息。对于静态人脸图像的检测,一般是先识别眼睛、刘海、胡子等具有特征性的标记点,捕捉特征图像的同时,进行图像数据归类。

(2) 同样是测试50名用户,处于动态环境下的识别测试。本次不对用户划分小组,50名用户在嵌入式人脸识别系统的可覆盖区域之内进行合理位置移动,测定出该系统能够识别出的用户数量。对于动态人脸图像的检测,一般是先识别人脸的数量,标记之后,再对所识别人的无关特征进行匹配验证,形成特征标记点,捕捉特征图像,完成人脸的同步识别。

根据上述测定,结合获取的数据、信息,在标定的区域之内,结合AdaBoost算法,最终测定出系统在处于静态环境和动态环境下的误检率,具体如公式(4)所示:

$$I = \beta^2 \times \sum_T (0.5n + zT) \quad (4)$$

公式(4)中: I 表示误检率, β 表示嵌入转换差值, n 表示可识别范围, z 表示嵌入波动比率, T 表示识别次数。通过上述计算,最终可以得出实际的误检率。针对测试结果,进行图像分析,如图5所示。

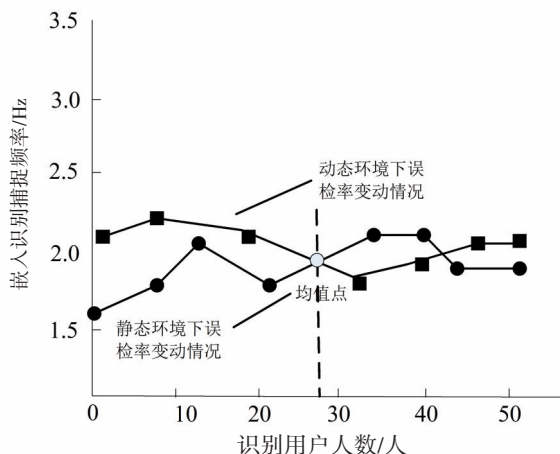


图5 测试结果图像分析

根据图5,可以完成对测试结果的分析与研究,具体见表1。

根据表1,可以完成对测试结果的分析 and 研究:针对于静态检测识别环境和动态检测识别环境下的测试,在50名用户的识别测验后,最终得出的误检率均控制在15%以下,表明在实

际嵌入式人脸识别过程中,识别误差小、可控、识别速度快、更易调整,具有实际的应用价值。

表1 测试结果分析验证表

测试指标	静态检测识别	动态检测识别
识别耗时/s	0.48	0.62
嵌入识别偏差	0.92	1.03
识别率/%	96.55	90.25
误检率	10.31	14.66

6 结语

综上所述,便是对嵌入式实验室门禁人脸识别系统的设计与验证。与传统的指纹识别、密码识别以及证件识别系统相对比,嵌入式人脸识别系统具有更大的优势,自身的识别范围更大,对于识别人员的存储量增大,识别误差小、速度快、精度高,促使其应用十分广泛。另外,针对于部分机密性较高的实验室,嵌入式人脸识别系统可以针对不同的实验室属性,设定对应的人脸识别级别和程序,从多方面增加门禁的安全识别程度,进一步提升系统的交互性和灵活性,另外添加非接触式数据采集、身份单层存储等功能,一定程度上也确保系统的隐蔽性和机密性,扩展相关技术的应用前景,为实验室门禁识别系统未来的创新和发展奠定坚实的基础。

参考文献:

- [1] 白燕燕,陈浩浩,范亚州,等.基于多模态融合的人脸识别门禁系统设计[J].电子制作,2022,30(12):28-30,34.
- [2] 陈保国.基于无线网络的人脸识别门禁系统设计与实现[J].无线互联科技,2021,18(24):48-49,53.
- [3] 廖文杰.基于树莓派的门禁系统设计与实现[J].信息与电脑(理论版),2021,33(21):144-146.
- [4] 吴立军,宁睿,侯德华,等.基于人脸识别的智能门禁系统设计[J].河南科技,2021,40(26):19-21.
- [5] 余秀普,高明,陈嘉铭,等.基于人脸识别和热成像门禁系统的研究与设计[J].职业技术,2021,20(9):98-103.
- [6] 周飞,李峥,高万明.嵌入式防疫识别系统设计[J].牡丹江师范学院学报(自然科学版),2021(2):30-33.

(下转第108页)

文章编号: 1007-1423(2024)04-0104-05

DOI: 10.3969/j.issn.1007-1423.2024.04.019

基于三维点云建模的操作驾驶舱设计

宋亚楠*, 姜 晗, 吕厚稳, 葛楚琳, 曹纪慧

(海洋石油工程(青岛)有限公司数字化中心, 青岛 266000)

摘要: 针对工程企业施工管理的特点并结合最新的元宇宙概念, 提出了一种新型数字化驾驶舱管理的模式。以往的数字化管理模式侧重于采集基础底层信息并统计成看板, 对于可视化也仅仅停留在管理系统以外, 忽视了管理者与执行者的交互和工程场地数据的集成。通过所设计的新型操作驾驶舱系统可以实现场地模型、基础数据和管理交互的实时有效协同, 真正实现工程场地的元宇宙模式, 大幅提高数字化管理效率、管理水平以及管理体验。

关键词: 工程管理; 信息化; 三维重建; 元宇宙; 人工智能

0 引言

随着工业 4.0 时代的来临, 各领域制造业均在尝试利用信息技术来促进产业变革, 改善工作流程、提高工作效率、节省工作成本、增强工作体验。这促使了智能化的工具在当下愈加丰富, 恰又适逢我国迎来了首轮人口负增长, 人口红利逐渐减弱。在此情况下可以明确, 未来的制造业面临着不得不进行的重大转型, 而管理的转型则是首当其冲和重中之重。目前, 各大行业领先的公司和组织都在进行生产管理的数字化转型实践, 取得了瞩目的成果, 也遇到了许多实际落地的问题。如数据无法智能化及时采集, 使数据录入工作较多需要人来参与; 数字化管理无法深入替代传统管理, 致使线上的工作与线下工作重复; 现有 MES 制造执行系统与对应行业业务的对接不完善, 导致无法完全提取数据, 管理深度浮于表面; 管理工具着重于替代传统管理业务, 而未从全面的角度创新管理思维等。这些问题是当前许多系统的特点, 也是数字化转型的瓶颈, 极大影响了用户在工作中的使用体验^[1]。

常规的制造业生产管理系统主要包括制造执行 (MES) 系统、监督控制和数据采集 (SCADA) 系统、数字孪生可视化平台和数据综合管理驾驶舱。MES 系统实用的基础取决于 SCADA 系统的自动化程度或人为参与的管理水平。数字孪生可视化平台需要与 MES 系统进行深度融合, 才能使两者发挥出作用, 提高生产管理的可行性。而数据综合管理驾驶舱目前基本上都是以综合看板为主, 偏重于其他生产系统环节的数据呈现, 忽视了数据交互和在线管理。基于目前的整体行业状况和海洋石油工程的行业特点, 本文在综合梳理了 MES 系统和可视化平台流程的前提下, 提出了一种智能化可交互的新型管理操作驾驶舱概念, 并给出了具体设计。

1 海洋工程行业生产管理架构及业务需求分析

海洋工程行业生产管理体系主要涉及以下几个方面: ①生产计划的排程; ②材料、零件以及消耗品的跟踪; ③实时生产进度明细; ④质量、安全、技术文档更新、流转及查阅; ⑤人力及

收稿日期: 2023-08-03 修稿日期: 2024-01-15

作者简介: *通信作者: 宋亚楠(1994—), 女, 内蒙古乌兰察布人, 工学学士, 计算机软件初级工程师, 从事海洋石油工程建造信息系统开发与运维研究, E-mail: songyn4@cooec.com.cn; 姜晗(1992—), 男, 山东青岛人, 硕士研究生, 系统运维工程师, 助理工程师, 研究方向为石油工程计算机视觉算法、石油工程数字化管理及技术应用、工程管理优化等; 吕厚稳(1997—), 男, 山东菏泽人, 工学学士, 机械工程师, 研究方向为机械设计、石油工程数字化管理及技术应用、工程管理优化等; 葛楚琳(1990—), 女, 天津人, 本科, 工程师, 主要从事海洋工程信息化系统应用与管理方面的研究; 曹纪慧(1994—), 女, 山东日照人, 本科, 工程师, 主要从事海洋工程项目管理和信息化系统应用与管理方面的研究

工机具的管理。

(1) 生产计划排程：输入信息包括材料进度、工机具安排、人力情况。输出施工计划和反馈调节，自适应要求高。得到的动态排程计划用以指导后续生产，对生产实际的反馈进行响应。可嵌入人工智能算法，得出最佳人力安排和资源配置，为人工决策给出专业指导。

(2) 材料、零件以及消耗品的跟踪：输入信息为材料的采购信息、材料的上游生产及其运输情况以及生产过程中的材料零件的全流程跟踪。这些信息来源于厂家信息、施工信息和监控信息，由相应负责人录入，并在数字孪生场地地图中标识。

(3) 实时生产进度明细：生产管理系统(MES)需建立完整并可伸缩配置的流程引擎，以期能够适应各种类型的车间场地。针对MES的各流程工序使用智能化的数据收集方法，使工厂管理更加具象化。

(4) 质量、安全、技术文档更新、流转及查阅：建立文档管理，便于施工人员查阅、线上审批报检和数字化交付相关事宜。

(5) 人力及工机具的管理：人力资源情况和工机具的明细需实时更新，所有相关数据都在系统中体现，便于其管理和建模。

针对海洋石油工程行业业务需求进行分析，具有数据量大、数据变动频率高以及非流程性数据多的特点，人工统计和实时更新不仅费时费力，而且错误率高、时效性差。另外，工程数据报表的生成也是工程管理过程中的一个重要环节，通过对数据的分析进行下一步生产的决策，这一过程目前是通过线下会议来实现上传下达，时效性差、管理不够灵活。

业务需求除了生产数据的管理和流通外，对场地真正实现数字孪生有着迫切需求。数字孪生元宇宙指对生产场地的全部生产数据在服务器上数字投影，现实中的生产活动与元宇宙中的数据实时联动，元宇宙为数据的可视化提供了一个中间平台。

2 模块与算法设计

2.1 三维点云场地建模

为了更加真实地反映工程现场情况，达成数字孪生，利用三维重建的方法效率和准确率

要远胜人工建模，也可满足管理系统海量数据处理的时效性要求，对场地整体和车间内外事物利用点云三维重建技术可以实现建模。

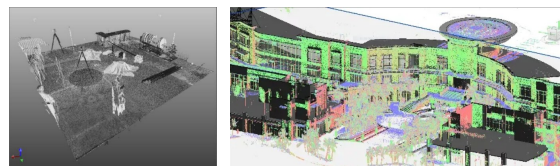


图1 点云建模效果

基于点云的三维建模需利用摄像头、激光仪等对厂区进行扫描，方法有单目视觉、双目视觉、深度视觉、激光扫描等。在室外的场景下利用单目视觉进行重建计算量大^[2]、存储要求高，且需要进行运算存在时差；而双目视觉利用“视差法”通过一帧多个图片同时匹配能更加快速地计算相机位姿和对象位置^[3]；深度视觉是使用一种单目摄像头结合红外传感器的相机，利用飞行时间法(TOF)能够直接获取位置坐标的一个维度数据，再通过计算其他维度数据得到准确的目标定位和位置信息。然而，红外传感器在室外复杂光线的影响下容易造成黑圈和空洞问题，造成数据丢失不利于三维重建。因此，选用双目或多目相机对室内外三维重建比较适合。

根据在海洋石油工程场地的实际情况，可以利用场地遍布的摄像头加自动无人机来作为视觉设备，同步更新可视化模型。多目视觉进行稠密点云重建，第一步是双目图像特征匹配，计算形成深度图，再结合平面图计算出每一个点的三维坐标，在地图中绘制。并利用深度学习进行边缘检测、对比点云数据通过聚类算法与学习算法结合形成具有语义信息的场景三维地图。再结合人工添加的组块三维地图可以形成整个厂区场景语义。得到的点云场地数据聚类可形成更加平滑的表面样式，进一步染色和填充表面则可形成更友好的视觉画面。这一模块是数字孪生智慧厂区的基础。

2.2 场地对象识别算法设计

通过双目视觉算法可以扫描获取整个厂区的点云数据，而点云数据并不具有语义信息，无法满足场地管理的需求。目前基于神经网络

深度学习的识别算法得到了广泛的研究和思考,如SSP-Net、Fast R-CNN、Faster R-CNN、MS-RCNN、Mask R-CNN、SSD等。首先进行网格化处理和边缘检测,将点云进行分类,这里涉及到聚类算法,通过边缘检测和聚类算法将在场景点云中提取的各种物体的点云进行集合,形成对象。对象的点云集合需要进行大量数据训练,得到符合海洋石油工程场景的深度学习数据集,对各种点云集进行识别分类,赋予语义,形成厂区管理数据模型的数据单元^[4]。厂区对象标记和使用,可以实现厂区的动态管理和实时监控,实现生产细节的可视化以及管理驾驶舱的交互性。

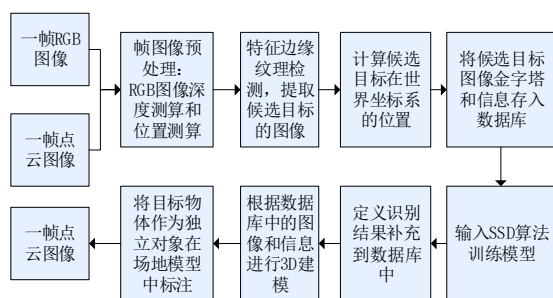


图2 场地目标识别建模流程

2.3 生产数据智能采集方法

基于3D可视化厂区语义模型可以实现车间内外的整体建模。除此之外,还需要智能地获取生产数据与厂区语义模型结合起来,才能实现3D可视化管理。生产业务数据和场地模型数据进行映射和关联是很重要的一步。采集信息的方法可以人工录入,然而实效性差、错误多,且程序不够严谨,容易人为影响程序进程,造成死锁。因此,针对于智能化管理工厂需要更加智能的数据采集手段,才可能避免人工管理的问题,真正解放人力。

场地生产数据的采集需安装多种传感器设备,以期实现厂区业务数据的全面监控与收集。厂区业务数据经过分析可得出,主要包括:①风水电表显数据的整合;②人力情况的分布与分配;③工机具及大型设备的定位和使用情况;④施工材料跟踪和仓库的存储。针对以上问题,可利用集成的策略对整个场地业务信息进行整合。集成风水气电等相关仪表数据到系统上,

与生产管理系统MES集成并通过手持终端设备填报各类施工数据,利用机械臂和自动导航车AGV来实现自动仓储管理和运输任务。零件管理是一个难题,对于小型大批量的零件很难进行单个数字化管理,因此批量管理托盘是解决这个问题重中之重,小型零件以托盘为单位进入MES系统工序流,以此来实现数字化管理。

2.4 辅助决策系统的算法设计

辅助决策系统是人工智能领域比较高阶的应用,涉及到大数据分析、云知识库、机器学习以及自适应算法等。目前海洋石油工程企业在海外项目投入力度愈发增强,潜在风险把控和节本增效的需求也日益增强。与此同时,面对复杂的工程情况依旧停留在以人工经验为基础的判断决策层面上,并未建立自适应的、科学标准的风险管理和项目排程体系。对于目前存在的问题,亟待建立一种全面、系统、自适应、自动化的计算机辅助决策系统,以便工程项目管理更加先进有效。

对于海洋石油工程企业而言,辅助决策系统体现在以下三个方面:

(1)项目商务的合同风险管理决策、物料运输风险评估、施工安全风险;

(2)工程项目施工计划排程和资源分配动态决策;

(3)人工智能应对随机突发事件,给出解决方案建议。

辅助决策依赖于大数据技术,数据具有价值型、多样性、高速性和海量性等特点,大数据需要海量存储的服务器作为基础,由于数据海量且多样,造成了大数据检索过程中高并发、高噪声的特点^[5]。分离关键数据,用以训练上述问题的最优解是启发式辅助决策需要完成的基础。

步骤1:辅助算法以施工计划排程为主要研究对象。首先需要从生产管理系统中提取各类生产对象的详细信息,并对其做数据预处理,随后进入到决策系统数据库中已备进一步计算,形成大量学习数据表喂给辅助决策学习系统。

步骤2:目标函数的导入和更改。辅助决策最终目的是实现决策,机器学习应用在这个问题中的目的是数据挖掘,即通过大量计算提取

出人力难以迅速实现的结果,为人工决策提供基础信息。因此,在本系统中采取两种辅助决策目标方案。第一种为人工输入最优解函数,计算机利用输入的数据按照人工设定的函数计算最优解;第二种为概念目标的输入,即不指定具体函数公式,以机器学习为基础进行参数的调整,最终得到最优解。实现第二种方案可以利用 chatgpt 进行实现。

步骤3:训练得到的函数可用于新输入数据的处理,计算得出成本、工作量以及最优方案参考,供给管理驾驶舱模块用以人工决策辅助。

2.5 管理驾驶舱模块设计

结合上面的设置,即具备实现高互动性的自动驾驶舱程序的可能,管理驾驶舱并非简单的数字看板、统计数据,而是具有实时性、交互性和可视化的中控平台^[6]。针对海洋石油工程企业的特点,梳理数字化管理流程。首先是场地组块铺排,当一个新的项目来临,能得到工程的设计数据,包括体量大小、工艺用料等信息,可以据此确定该项目的场地铺排,驾驶舱根据场地情况进行操作,简单地设置项目开始地点,再通过生产管理系统导入项目详细信息并逐步导入项目设计信息。生产管理系统按照原有逻辑进行数据流动,驾驶舱模块根据需要抽取数据。驾驶舱可以观察整个场地情况,包括场地的动态3维模型,人员、设备的分布。对整个场地情况一目了然。除此之外,驾驶舱可以输入生产指令等,将数据传达到目标客体,输入之前,驾驶舱将通过辅助决策模块给出所有工作的预计成本、投入物料和大约人工等,为人工提出生产任务提供科学的数据支撑。

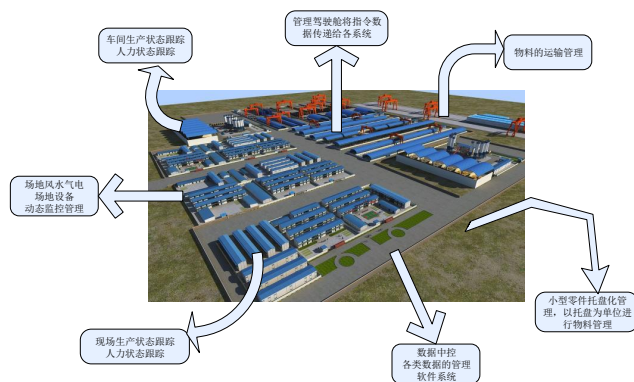


图3 场地管理驾驶舱整体模块架构

3 效益分析

在信息化转型中,实际效益在许多应用中无法清楚,这也是制约众多企业发展的一大因素,科学的投入产出比计算来估计信息化系统的实际效益是项目最重要的环节。

通过信息化的实时监测计算,可以让机器设备在单位时间内的利用率最大化。并且可以通过信息化系统对设备进行全生命周期的监测,通过对设备的各个部件工作参数监控,可以准确地预测设备可能出现的故障,及时排查检修,既降低因为设备不可控的故障影响生产效率的可能性,又可以减少维修产生的费用。

信息化平台可以对人员从多个方面进行降本增效管理,通过算法计算人员的工作能力因子(包括工作能力、工作效率、创新能力、合作意识、出勤情况等),对人员进行合理的分配,提高人员单位时间内的工作效率,提高生产进度,从而降低生产作业成本。

4 结语

本文主要设计了一套海洋石油工程企业生产信息一体化平台,其中包括了物联网硬件部分的搭建使用。通过信息化平台,可以有力提高管理效率,降低管理成本,使工作更具流程化和科学性。利用了大数据智能分析技术来改善人力资源管理水平和计划管理的水平。经过论证,本文方法能够有效提高生产制造企业的管理水平。

参考文献:

- [1] 刘鑫,王李管,彭朝晖,等.地下矿卡智能调度与生产管理系统设计及应用[J].黄金科学技术,2023,31(4):680-688.
- [2] 康景尧.面向室外场景的三维重建研究[D].西安:陕西科技大学,2023.
- [3] 郝晓东.建造半潜式平台三维快速测量技术研究[D].青岛:中国石油大学(华东),2012.
- [4] 姜晗,贺付亮,王世元.基于生长四叉树结构的二维建图算法[J].西南大学学报(自然科学版),2020,42(6):128-139.
- [5] 李小冬,杨帆,李吉勤,等.国际工程项目风险管理辅助决策支持体系及要素研究[J].建筑经济,2014(5):34-39.
- [6] 祁家祯,梁田雨,杨驰,等.军工制造企业数据管理驾驶舱的设计与应用[J].智能制造,2022(3):121-124,128.

Design of operating cockpit based on 3D point cloud modeling

Song Ya'nan*, Jiang Han, Lyu Houwen, Ge Chulin, Cao Jihui

(Digital Center of Ocean Oil Engineering Co., Ltd., Qingdao 266000, China)

Abstract: According to the characteristics of the construction management of engineering enterprises and the latest concept of the meta-universe, a new digital cockpit management mode is proposed. The previous digital management mode focuses on the collection of basic underlying information and statistics into spectacles. The visualization only stays outside the management system, ignoring the interaction between managers and executors and the integration of engineering site data. Through the new operating cockpit system designed, it can realize the real-time and effective coordination of the site model, basic data and management interaction, truly realize the meta-universe mode of the engineering site, and greatly improve the digital management efficiency, management level and management experience.

Keywords: project management; informatization; three-dimensional reconstruction; meta-universe; artificial intelligence

(上接第 103 页)

Embedded laboratory access control face recognition system design

Zhai Li^{1*}, Zou Qizhang²

(1. Lab Center of Nanjing Audit University, Nanjing 211815, China;
2. School of Economics, Nanjing Audit University, Nanjing 211815, China)

Abstract: In order to strengthen the control of the laboratory access control system and reduce the false detection rate of the access control procedures, the embedded laboratory face recognition system of the laboratory access control system is designed and verified. Simply describe the construction requirements of the system, and design the overall structure of the embedded system. The OK6410 core board is connected to the basic circuit to complete the system hardware design, establish an embedded cross-recognition compilation environment, pre-process the identified image and design positioning instructions, and associate the ORL embedded face database to complete the design of the system software. System test results show that: for static detection recognition environment and dynamic detection recognition environment test, after 50 users identification test, the error rate control below 15%, shows that in the process of actual embedded face recognition, identification error is small, controllable, identify fast, easier to adjust, has the practical application value.

Keywords: face recognition; embedded; system design; laboratory access control; recognition program; remote monitoring

文章编号: 1007-1423(2024)04-0109-04

DOI: 10.3969/j.issn.1007-1423.2024.04.020

基于 CNN 的人脸年龄与性别检测系统设计

孙小广^{1*}, 万若楠¹, 余光正²

(1. 广州城市理工学院电子信息工程学院, 广州 510800; 2. 华南理工大学物理与光电学院声学研究所, 广州 510640)

摘要: 基于 CNN 的人脸年龄与性别检测系统以 PyTorch 中 ResNet-50 模型为主干网络, 利用 Wiki 平台进行人脸原始图像采集预处理, 构建年龄预测与性别判断分类器作为模型网络输出层, 使用均方误差与交叉熵损失函数对年龄与性别结果进行预测输出。系统前端使用 PyQt5 构建图像识别、视频识别和实时摄像识别的交互界面, 采用 OpenCV 对图像、视频或摄像头读取的每一帧图像进行人脸标注, 将系统模型识别的人脸性别与年龄预测检测结果绘制成图像并展示。

关键词: CNN; ResNet-50 模型; OpenCV; 人脸识别

0 引言

随着人工智能技术的迅速发展, 尤其是基于深度学习的人脸识别研究的深入^[1], 基于计算机视觉的人脸年龄与性别检测系统研究已经成为一个热门领域^[2]。基于 CNN 的人脸年龄与性别检测系统^[3], 以 PyTorch 中 ResNet-50 作为后端模型主干网络^[4], 将年龄预测分类器^[5]与性别预测分类器整合为一个模型进行预测输出^[6]。系统将前端界面、摄像头模块等交互功能接入后端完成训练模型^[7], 可以快速准确地判断消费者的年龄和性别, 识别不同年龄和性别的人员。系统能实时监控公共场所, 提高社会安全性, 在医疗、教育和娱乐等领域也得到广泛应用, 能满足不同用户的需求。该系统可以更好地应用计算机视觉技术, 提高人工智能化水平, 为社会和经济发展提供更好的支持和保障。

1 数据预处理

Wiki 是一个基于 Nvidia Tegra3 处理器的开放实验人脸数据集平台, 在此平台上进行人脸检

测和跟踪, 使用多个不同摄像头采集数据。数据集包含 500 张来自世界各地的人脸原始图像, 涵盖照片、3D 模型、视频等。在这个数据集上, 研究者可以分析人脸特征值、人脸比例、脸部形状等信息。Wiki 原始数据集包含的原始图片数量很多, 需要从中筛选大约 10000 张图片(男女各 5000 张)进行预处理, 从 10000 张图片中抽取约 1000 张(男女比例相当)作为测试集, 剩余 9000 张作为训练集。OpenCV 识别出一张人脸图像^[8], 将图片由 RGB 图转换为灰度图, 将图片尺寸比例统一为 100 mm*100 mm。文件名格式统一命名为: 编号-年龄-性别.png, 其中性别 1 代表男性, 0 代表女性, 部分人脸图像如图 1 所示。

对数据进行预处理后, 需要使用特定的模型对训练集进行训练, 以使模型能够正确预测任务, 当模型构建完成进行测试时, 需要从训练集中随机选取一定数量的样本用于模型性能评估, 以确定模型的准确性和稳定性。

收稿日期: 2023-09-22 修稿日期: 2023-10-15

基金项目: 国家自然科学基金项目(12074129)

作者简介: *通信作者: 孙小广(1983—), 男, 湖南武冈人, 硕士, 讲师, 研究方向为信息处理技术、超声技术及应用, E-mail: sunxg@gu.edu.cn; 万若楠(1983—), 女, 湖北潜江人, 硕士, 讲师, 研究方向为信息处理技术、声子晶体技术; 余光正(1978—), 男, 四川泸州人, 博士, 教授, 研究方向为声信号处理、心理声学、双耳听觉等领域



图1 人脸数据集

2 模型设计

系统使用 ResNet-50 作为模型的主干网络来提取人脸图像的特征，模型由多个组件构成，通过导入预训练模型及其权重，构造出对人脸图像的年龄和性别分类结构，相当于构建一个二分类组网。模型通过图像预处理、人脸检测和性别识别完成多个任务^[9]，实现对面脸图像中的年龄和性别进行快速、准确的识别，模型结构如图2所示。

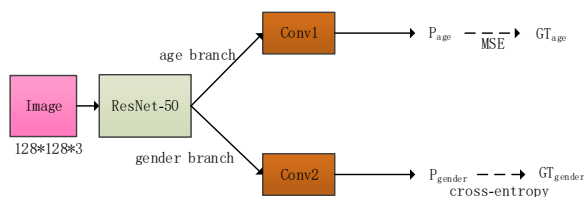


图2 模型结构

ResNet-50 的输出特征图传输到年龄与性别分类组网，年龄与性别分类组网分别输出一个向量。性别识别模型使用了三层卷积层，第一层是池化层，第二层是全连接层，第三层是输出层。卷积层1：在 ResNet-50 的输出特性映射上，以 3×3 的卷积核尺寸和 1 步的方式使用，为了加快模型的收敛性，在卷积层之后增加了一个归一化层，在归一化层之后又添加一个 ReLU 激活函数，用来提高模型的非线性能力。卷积层2：在第一个卷积层之后，增加的另一个卷积层，也包含一个卷积层、一个归一化层和一个激活函数，卷积核数量分别为 64、128 和 256。卷积层3：包含一个自适应平均池化层和一个全

连接层，池化层主要体现在自适应上，对于任何大小的输入，输出均自适应调整为指定大小。

整个模型结构考虑了输入数据的尺寸和存储格式以及与 ResNet-50 模型的兼容性，同时对于分类器 loss function 和 optimizer 的设置也进行了选择，它们对模型的性能和训练效率有着重要影响。

loss function 的选择：交叉熵损失函数是最常用的计算分类损失的函数，可用于多分类问题，也可用于二分类问题。交叉熵损失是一个针对概率测量的评估指标，它量化了实际输出与期望输出之间的差异，将预测概率最小化。且在实际问题中，目标是使模型预测正确的分类概率最大，因此使用交叉熵损失函数可以有效地达到这个目标。而且，交叉熵损失函数通用地适用于多分类问题时不必指定每个类别的权重或阈值等额外参数，在训练过程中只需要让模型最小化交叉熵损失，它就能够自动地调整权重参数以达到一个比较好的分类结果。因此，它很适用于本系统模型中的人脸图像性别分类器。

optimizer 的选择：在深度学习模型时，Adam 优化器通常是最好的选择之一。Adam 具有自适应超参数功能，能在动态中自行更新参数。这种性质使得 Adam 优化器通常比一些传统的优化器(如随机梯度下降)更容易收敛，并且更可靠地处理高维数据。因此，在本文识别人脸图像判断性别系统中，选择 Adam 优化器对模型进行训练。

年龄分类模型与性别分类模型相似，同样使用三层卷积层。为优化神经网络回归，选择均方误差 MSE 损失函数，MSE 损失函数广泛应用于线性回归和多元线性回归等任务中。当均方误差为代价函数时，为了使均方误差最小，一般都会通过梯度下降法来实现。在 PyTorch 中，可以使用内置的 MSE 损失函数来计算均方误差，同时使用 Adam 优化器对模型进行优化，使得模型能够在训练数据上达到较好的拟合效果。

3 交互界面

系统的前端基于 PyQt5 库进行开发，使用 PyQt5 自带的 QT Designer 对交互界面进行编辑。

根据系统上传图片、视频，开启摄像头等需求自定义和设计交互界面，且留出一定的宽体界面用于识别预测结果展示。基于该系统需求，设计添加了拍摄视频，抓拍人脸等功能控件，同时定义控件点击跳转函数，并使用OpenCV工具对图像中人脸进行标注，并展示识别结果。交互主界面如图3所示。



图3 系统交互界面

OpenCV工具是一款面向Linux、Windows、Android、Mac等多个操作系统的跨平台视频数据库，其由一系列的C函数以及少数的C++类函数组成，并为Python、Ruby、MATLAB等语言提供接口，可以实现许多用在图像处理和计算机视觉领域的通用算法。系统定义控件点击使用OpenCV函数开启电脑摄像头，将图像上传于模型进行人脸识别标注，并实时将每帧识别图像结果展示于界面框中。本系统中，OpenCV不仅在数据的预处理过程使用，而且在最后的识别预测环节也起到不可或缺的作用。

4 功能测试

实际应用中，一个任务可以选择多种学习算法。对同一个算法，当使用不同的参数配置时，也会产生不同的模型。本系统设计，我们通过衡量模型的泛化误差，通过试验评估获取了一定的性能度量，对学习器的性能也进行比较，得到了较理想的模型。在本模型的性别分类中，使用准确率与召回率作为评价模型表现的重要指标。准确率是指在模型预测为正样本

的数据中真正为正样本的比例($P = TP / (TP + FP)$)，召回率是指在真实为正样本的数据中被模型预测为正样本的比例($R = TP / (TP + FN)$)。模型训练测试准确率如图4所示。

```
root INFO Training epoch:43/100 batch:1/586 lr:0.000755 age_loss:0.015751
gender_loss:0.535456 loss:0.551207 acc:0.750000;
root INFO Training epoch:43/100 batch:51/586 lr:0.000755 age_loss:0.016912
gender_loss:0.202182 loss:0.219094 acc:0.924042;
root INFO Training epoch:43/100 batch:101/586 lr:0.000755 age_loss:0.017323
gender_loss:0.194603 loss:0.211925 acc:0.931312;
root INFO Training epoch:43/100 batch:151/586 lr:0.000755 age_loss:0.017437
gender_loss:0.201491 loss:0.218928 acc:0.922185;
root INFO Training epoch:43/100 batch:201/586 lr:0.000755 age_loss:0.019190
gender_loss:0.200332 loss:0.219523 acc:0.920398;
root INFO Training epoch:43/100 batch:251/586 lr:0.000755 age_loss:0.018755
gender_loss:0.197581 loss:0.216336 acc:0.919323;
root INFO Training epoch:43/100 batch:301/586 lr:0.000755 age_loss:0.018556
gender_loss:0.196523 loss:0.215079 acc:0.920889;
root INFO Training epoch:43/100 batch:351/586 lr:0.000755 age_loss:0.018543
gender_loss:0.192140 loss:0.210683 acc:0.922543;
root INFO Training epoch:43/100 batch:401/586 lr:0.000755 age_loss:0.018375
gender_loss:0.191714 loss:0.210089 acc:0.920979;
root INFO Training epoch:43/100 batch:451/586 lr:0.000755 age_loss:0.018260
gender_loss:0.190643 loss:0.208902 acc:0.922256.
```

图4 模型测试输出日志

模型运行的学习曲线表示在训练过程中，不同训练轮次的训练误差与验证误差的变化趋势，验证曲线则反映在不同的模型配置下，验证误差与模型复杂度之间的关系。利用学习曲线和验证曲线，我们很清晰地看到模型性能的提升度，以及模型中是否存在过拟合或欠拟合。本模型学习曲线与验证曲线如图5所示。

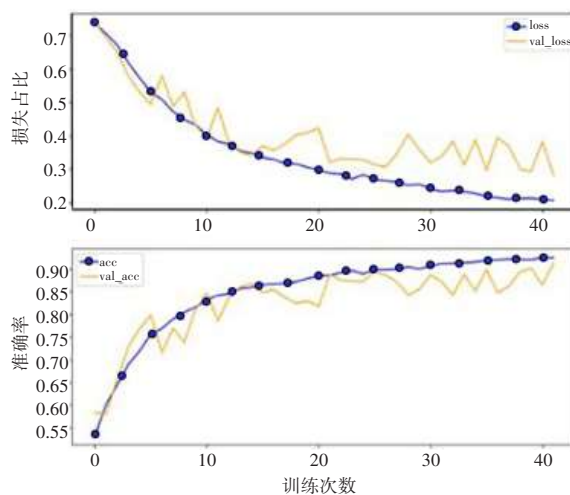


图5 模型性能评估

如图5所示，经过多轮训练，训练集与测试集的loss不断下降，与此同时，训练集和测试集的ACC在持续上升，在训练轮数达到40后，模

型的准确率达到90%。

5 结语

基于CNN的人脸性别与年龄预测系统，是机器学习、深度学习以及计算机视觉等技术的结合。系统构造了基于ResNet-50的年龄预测与性别判断的主干网络模型，构建了可进行图像视频识别预测，亦可进行摄像头实时辨别预测的交互界面，前段和后端设计合理衔接，识别预测人脸图像稳定性好，准确率达到90%。

模型的目标函数、学习率、优化器等参数的调整，会影响系统的稳定性和识别预测的准确率。通过多种人脸识别方案的对比，可对该人脸识别性别年龄系统进一步优化，进而减少误差，使人脸图像的判断预测更加准确。

参考文献：

- [1] 黄煜真,元泽怀,陈嘉瑞,等. PyTorch框架下基于CNN的人脸识别方法研究[J]. 信息与电脑(理论版),2022,34(10):197-199.
- [2] EIDINGER E, ENBAR R, HASSNER T. Age and

gender estimation of unfiltered faces[J]. IEEE Trans Inf Forensics Secur, 2014, 9(12):2170-2179.

- [3] 何永强,秦勤,王俊鹏. 改进深度学习块卷积神经网络的人脸表情识别[J]. 计算机工程与设计, 2019, 40(3):850-855.
- [4] 邵定琴,张乾,岳诗琴,等. 一种融合Resnet-50与SVR的年龄估计[J]. 计算机仿真, 2023, 40(1): 280-285.
- [5] 缪辉昊. 基于CNN的跨年龄人脸识别在线训练算法设计与实现[D]. 南京:东南大学,2021.
- [6] 闫鹏飞,张忠民. 基于多任务学习的跨年龄人脸识别[J]. 哈尔滨商业大学学报(自然科学版), 2023, 39(1):55-61.
- [7] 谷广兵,沈晓萍,王玉婷. 基于CNN的人脸识别研究与实现[J]. 现代信息科技, 2021, 5(13): 22-25, 30.
- [8] 查宏超,王琪,高进可,等. 基于树莓派和Open CV的深度学习人脸检测和性别年龄预测方法的研究[J]. 科技技术创新, 2020, 9:77-78.
- [9] 宋梦媛. 一种基于多任务CNN的多模态人脸识别模型[J]. 云南师范大学学报(自然科学版), 2022, 42(5):51-56.

Design of facial age and gender detection system based on CNN

Sun Xiaoguang^{1*}, Wan Ruonan¹, Yu Guangzheng²

(1. College of Electronic Information Engineering, Guangzhou City University of Technology, Guangzhou 510800, China;

2. Institute of Acoustics, School of Physics and Optoelectronics, South China University of Technology, Guangzhou 510640, China)

Abstract: The ResNet-50 model in Python is used as the backbone network for the facial age and gender detection system based on CNN, the original facial image collection is preprocessed on Wiki platform. An age prediction and gender judgment classifier is constructed as the output layer of the model network, and the mean square error and cross entropy loss function are used to predict and output the results of age and gender. An interactive interface of image, video and real-time camera recognition is built on front-end system used PyQt5, and OpenCV is used to annotate the face of each frame of images, videos, or camera readings. The predicted detection results of facial gender and age recognized by the model system are plotted and displayed as the image.

Keywords: CNN; ResNet-50 model; OpenCV; face recognition

文章编号: 1007-1423(2024)04-0113-04

DOI: 10.3969/j.issn.1007-1423.2024.04.021

三维虚拟直播间的设计与制作

陈 军*, 陈俊汶

(广东技术师范大学教育科学学院, 广州 510665)

摘要: 基于三维虚拟现实技术, 分析了虚拟演播室技术的特征, 从开发工具、设计思路和设计实现的角度, 详细阐述了三维虚拟直播间的构建过程, 构建了一个具有高度真实感的三维虚拟直播间。探讨了三维虚拟直播间的设计与实现, 进一步优化虚拟场景在网络直播中的应用和发展。

关键词: 三维虚拟直播间; 虚幻引擎; 三维建模; 开发工具

0 引言

随着科技的不断发展, 网络直播已经成为人们生活中的一部分。在虚拟现实技术的推动下, 虚拟直播间的发展逐渐引起了人们的关注, 传统的直播形式已经无法满足人们对于多元化、立体化的需求。因此, 三维虚拟直播间的出现, 为直播行业注入了新的活力。选择专业的游戏开发引擎 Unreal Engine 和三维建模软件 Maya 等开发工具, 构建虚拟直播间的各项元素。这些工具具有强大的功能和广泛的应用领域, 可以大大提高三维虚拟直播间开发效率和效果。通过场景建模、纹理贴图、烘焙渲染、绿幕搭建和虚实画面合成构建三维虚拟直播间。为观众呈现出一个极具沉浸感和真实感的虚拟世界, 使得观众能够身临其境地参与到直播中来。这种新颖的直播形式可以吸引更多观众的关注和参与, 推动直播行业的发展。

1 虚拟演播室技术及三维虚拟直播间开发工具

1.1 虚拟演播室技术

虚拟演播室技术是一种利用计算机技术创

造出的虚拟环境, 将演员或主持人放置在其中进行表演和互动。这种技术可以创造出逼真的虚拟背景、道具和场景, 以及虚拟角色和特效等元素, 为观众呈现出一个极具沉浸感和真实感的虚拟世界。虚拟演播室技术的工作原理是将计算机制作的虚拟三维场景与电视摄像机现场拍摄的人物活动图像进行数字化的实时合成, 使人物与虚拟背景能够同步变化, 从而实现两者的融合, 以获得完美的合成画面^[1]。这种技术可以应用于新闻、体育、娱乐、电视剧、直播等各种类型的电视节目制作中。虚拟演播室技术的优势在于其可以创造出极具真实感和沉浸感的虚拟环境, 同时也可以提高制作效率和降低制作成本。此外, 虚拟演播室技术还可以实现多种形式的互动和特效, 为观众带来更加丰富、生动的观看体验。

1.2 三维虚拟直播间开发工具

构建三维虚拟直播间需采用到的开发软件主要有:

(1) 3D 建模软件: 虚拟场景的创建需要使用 3D 建模软件, 例如 Blender、Maya、3DsMax 等, 这些软件能够创建复杂的 3D 模型, 包括场

收稿日期: 2023-11-17 修稿日期: 2023-12-11

基金项目: 广东省哲学社会科学“十三五”规划 2020 年度一般项目(GD20CJY21); 广东省 2018 年高职教育教学质量与教学改革工程项目(GDJG2019080)

作者简介: *通信作者: 陈军(1983—), 男, 广东雷州人, 硕士, 高级实验师, 研究方向为虚拟现实, E-mail: chenjun@gpnu.edu.cn; 陈俊汶(2001—), 男, 广东广州人, 学士, 研究方向为数字媒体技术

景布局、角色、道具等。

(2) 游戏引擎：游戏引擎是虚拟场景的实现基础，常用的游戏引擎有 Unity3D、Unreal Engine 等。这些引擎能够实现虚拟场景的渲染和交互，包括场景的光影效果、角色的移动和动作等^[2]。

(3) 视频直播软件：视频直播软件用于将虚拟场景实时编码成视频流，并将其上传到直播平台。常用的视频直播软件有 OBS、XSplit 等。

(4) 直播平台 SDK：直播平台 SDK 可将编码后的视频流传输到直播平台，例如抖音、哔哩哔哩 bilibili、YouTube、Twitch 等。

2 设计思路

电子竞技比赛解说直播间由外景和主题建筑两部分组成。考虑到受众人群为成年人，直播间并不需要儿童化或更加靓丽的色彩。镜头景别设计有全景、中景、特写，以及开场动画与退场动画。在场景中添加烟雾、火花、激光等特殊的视觉元素，增加比赛的视觉效果和娱乐性。

开发流程：三维虚拟直播间的开发一般分为四部分。虚拟部分，使用三维建模软件进行场景建模；现实部分，架设绿幕与摄像机，通过采集卡或数据线传输现实画面；合成部分，将三维场景和现实拍摄的主播导入虚幻引擎中进行虚实融合；输出部分，采用 OBS 软件获取窗口画面后输出。

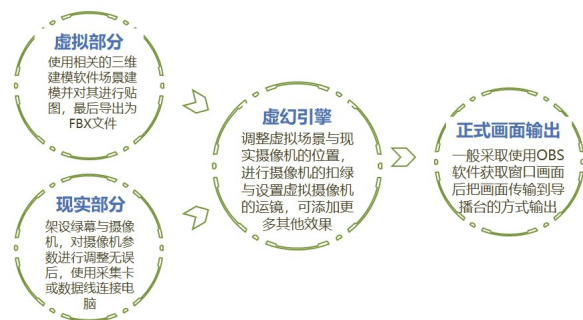


图1 三维虚拟直播间设计流程

3 设计与实现

3.1 三维模型构建

3.1.1 模型建立

模型是整个虚拟直播间设计的基础，它的

逼真程度直接关系到虚拟直播间的沉浸性和真实性，目前主要采用的建模方法有两种：①标准几何体建模法，利用已有的几何体模型搭建，再后期贴图；②组合标准几何体模型建模法，纹理贴图增加模型的逼真度，烘焙后压缩文件大小，提高直播的实时渲染速度^[3]。由于直播受到网络带宽的限制，在模型的构建中应特别注意，在提高模型逼真度的同时，尽可能简化模型和压缩文件大小。

依据以上建模原则，电子竞技比赛解说直播间的场景古建筑模型建模设计，利用三维建模软件 Maya2022 完成虚拟场景的建模。

(1) 添加分段：在 Maya 软件添加分段需要切换为对象模式后选中物体，在“编辑网格”选项，找到添加分段选项，调整 UV 两项段数，调整好分段数后，切换到选择点或线，把多余的点和线删除。

(2) 挤出工具的应用：选择立方体后切换为选择面模式，选择所需要挤出的面，按住“shift”键和鼠标右键，选择“挤出面”选项，按“W”设置为移动工具，即可把面挤出。

(3) 收拢工具：将多个点收拢为一个点，选中所要收拢的模型，长按鼠标右键，切换为选择边模式，点击收拢快捷键，即可把边的两个顶点收拢成一个点。对于一些需要复制的模型，使用快捷键“Ctrl+D”，复制出一个一模一样的模型。

利用添加分段、挤出工具、收拢工具完成虚拟场景的建模，如图2所示。

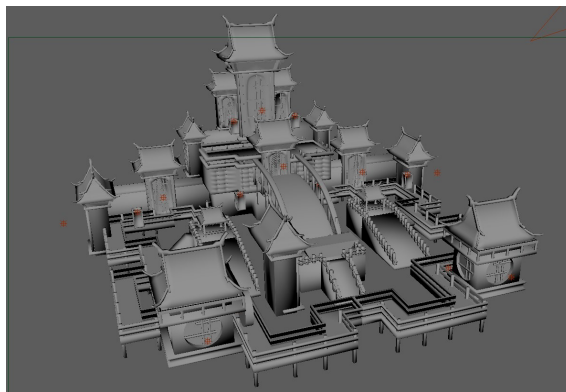


图2 虚拟场景建筑三维模型

3.1.2 场景贴图和烘焙

选择贴图的模型，在“多边形建模”的选

项里点击“按类型删除历史”。选择对象模式，在UV编辑器窗口上方的导航栏中选择“图像”“UV快照”。使用Photoshop软件打开UV快照图片，在UV快照图片下方新建图层，把所需要的贴图导入Photoshop图层内，根据黑色辅助线来选择每个面所要的贴图，完毕后隐藏UV快照图片，导出图片。回到Maya软件，打开“公用材质属性”“创建渲染节点”窗口，选择Maya分类下的“2D纹理”，导入Photoshop导出的图片，即可完成贴图。模型贴图完成后，进一步对场景添加灯光效果，使用烘焙技术对整个场景模型进行烘焙处理，将光照信息渲染成贴图，并将这个烘焙后的贴图再贴回到场景中去。这个过程将光照计算的结果提前写入到贴图中，从而在实时渲染中不需要进行耗时的光照计算，大大提高直播实时渲染的效率，最终电子竞技比赛解说三维虚拟直播间模型如图3所示。

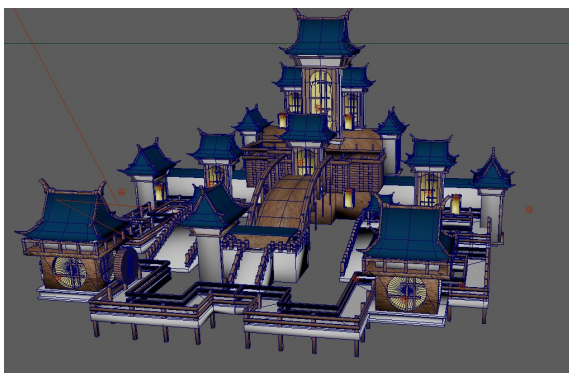


图3 三维虚拟直播间效果图

3.2 绿幕直播间的搭建

绿幕直播是一种创新的直播形式，通过绿幕背景和相关技术手段，无需花费大量资金去布置实景或者租用场地，主播可以在直播中呈现出各种不同的场景和背景，给观众带来更加丰富、多样化的视觉体验。

绿幕直播间的搭建方法如下：

(1) 准备绿幕：选择一块绿色幕布，布质或塑料的均可。尺寸需要根据直播间的实际情况进行选择，一般建议选择较大尺寸的幕布，可以模拟出更真实的场景效果。

(2) 安装绿幕：将绿幕悬挂在直播间的一侧，确保其平整无褶皱。如果是使用支架式绿幕，需要确保支架稳固，避免出现晃动或倾斜

等问题。

(3) 准备灯光：灯光是绿幕直播中非常重要的设备之一。需要选择合适的灯光设备，如LED灯、闪光灯等，根据直播需要进行布置。

(4) 安装摄像设备：选择合适的摄像设备，如相机、手机等，并确保其电量充足。同时，需要调整好摄像设备的设置，确保画面清晰、稳定。

(5) 连接绿幕和摄像设备：将摄像设备与绿幕进行连接，确保画面能够正确地映射到绿幕上。

(6) 调整灯光和摄像设备：根据需要调整灯光和摄像设备的角度和位置，确保画面效果达到最佳。光线控制是绿幕技术的关键之一。需要选择合适的灯光设备，并确保其光线均匀、柔和，以避免产生阴影或反光等问题。

3.3 直播画面合成

实时抠像技术是实现虚拟背景与主播画面合成的重要手段。通过使用抠像软件(如OBS、XSplit等)，可以将主播的画面与虚拟背景进行合成。抠像软件会根据图像边缘检测算法，将主播的实时画面与虚拟背景进行合成^[4]。通常，抠像软件会提供一系列工具和功能。连接设备和调试，将主播的画面传输到抠像软件中，并对其进行调整和优化。在这个过程中，需要不断调整抠像软件的参数和设置，如边缘检测、色彩校正、透明度调整等，以帮助将主播的画面与虚拟背景进行完美合成。在合成画面时，需要使用视频编辑软件(如Adobe Premiere、Final Cut Pro等)或虚拟现实软件(如Unity、Unreal Engine等)对画面进行调整和压缩编码。在这个过程中，需要注意画面的分辨率、颜色深度、压缩格式等因素，以确保画面质量和流畅度。

电子竞技比赛解说直播间通过Unreal Engine虚幻引擎来合成，在主窗口“电影摄像机actor”把三维模型拖入画面中，调整好三维模型的位置和旋转，使其对着场景当中的解说台。在“当前摄像机设置”里调整镜头设置、当前焦距和当前光圈。新建虚拟摄影机，在下方摄影机的选项中，调整好虚拟摄影机的位置，在“Sequencer”窗口中“Transform”选项打下关键帧，把时间轴拖动到最后一帧，调整好摄影机

的位置,再打下一个关键帧,完成关键帧动画镜头制作。将“数字键1”节点的“Pressed”与“创建关卡序列播放器”进行链接,将“创建关卡序列播放器”节点的“Return Value”与“播放”节点的“目标”进行链接,将“创建关卡序列播放器”与“播放”链接,完成后点击“编译”并保存。最终合成效果如图4所示。



图4 直播画面合成效果

3.4 网络传输和直播

将合成的虚拟直播画面通过网络传输到直播平台进行直播。在这个过程中,需要对网络带宽和延迟进行监控和管理,以确保画面的质量和流畅度。同时,要确保直播内容符合相关法规和规定。直播过程中,网络稳定性至关重要,如果网络出现波动或中断,会影响直播的质量和流畅度,需要选择稳定的网络环境,并采取措施,如使用网络加速器等来确保网络稳定性。直播过程中,延时问题是一个常见的问

题,延时过长会影响观众的观看体验,因此需要采取措施减少延时。可以通过压缩视频数据、使用更快的编码器等方式来减少延时^[5]。

4 结语

三维虚拟直播间是一种基于虚拟现实技术的创新直播形式,具有广泛的应用前景和发展潜力。通过对三维虚拟直播间的设计与制作的实践,可以更好地了解和应用虚拟直播技术,为未来的直播产业发展和创新打下坚实的基础。虚拟直播间的搭建和应用是一项前沿性和创新性的技术,在实践过程中,需确保虚拟直播系统的稳定性和安全性,同时,也需要对直播内容进行审核和管理,避免出现不良信息或侵权行为等问题,为用户提供更加优质的直播体验。

参考文献:

- [1] 周梅. 融媒体时代虚拟演播技术在电视节目制作中的应用[J]. 科技传播, 2020, 12(3): 80-82.
- [2] 涂欢. 网络游戏直播中的互动行为研究[D]. 贵阳: 贵州民族大学, 2022.
- [3] 崔新友, 王培培. 贴图烘焙技术及其在城市三维景观系统中的应用初探[J]. 科技信息, 2010(15): 473-474.
- [4] 陈秋凤, 申群太, 张可为, 等. 边缘保持KNN抠图[J]. 小型微型计算机系统, 2015, 36(7): 1591-1596.
- [5] 吕晓静. 自适应码率户外视频直播系统的设计与实现[D]. 哈尔滨: 哈尔滨工业大学, 2020.

Design and production of 3D virtual broadcasting room

Chen Jun*, Chen Junwen

(College of Educational Science, GuangDong Polytechnic Normal University, Guangzhou 510665, China)

Abstract: Based on the 3D virtual reality technology, the characteristics of the virtual studio technology are analyzed. From the perspectives of development tools, design ideas and design and implementation, the construction process of the 3D virtual studio is described in detail, and a 3D virtual studio with high sense of reality is constructed. This paper discusses the design and implementation of 3D virtual broadcast room, and further optimizes the application and development of virtual scene in network broadcast.

Keywords: 3D virtual broadcasting room; unreal engine; three-dimensional modeling; development tool

文章编号: 1007-1423(2024)04-0117-04

DOI: 10.3969/j.issn.1007-1423.2024.04.022

基于 OpenStack 云架构下 Hadoop 分布式集群部署 实训教学过程研究

周少珂^{1*}, 王 雷¹, 石磊磊², 李 凯²

(1. 河南应用技术职业学院信息工程学院, 郑州 450042; 2. 河南数安教育科技研究院, 郑州 450042)

摘要: 高职院校纷纷开设云计算技术和大数据技术专业, 专业间既紧密联系各有侧重, 专业课程也有交叉, 使学生在专业学习时存在一定误区和困惑。为解决教师教学实训过程中的瓶颈难点, 深化学生对大数据和云计算技术的掌握程度, 针对云计算技术中 OpenStack 项目和大数据技术中 Hadoop 项目进行融合部署, 在 OpenStack 私有云成功配置的基础上, 申请云中的资源创建云主机, 进行 Hadoop 完全分布式集群部署和配置, 使师生在该教学实训环节成功实现配置, 进一步加深对大数据和云计算技术的理解, 促进专业课程知识的深入学习。

关键词: 大数据; Hadoop; OpenStack; 云计算; 教学

0 引言

随着大数据和云计算技术发展, 高职院校纷纷开设大数据技术和云计算技术专业, 二者之间既紧密联系又各有侧重, 其专业课程也有部分交叉, 使学生在专业学习时存在一定误区和困惑^[1]。学生在进行专业学习的过程中对这两门专业的认知程度较浅, 大部分停留在平日的高新技术流行热词方面, 没有能够着手演练的实操项目, 以便让学生从实操中直观地认知和掌握学习^[2]。

鉴于此, 为解决教师教学过程中的瓶颈难点, 深化学生对于大数据和云计算技术的掌握程度, 本文以云计算技术中目前流程的 OpenStack 开源项目为研究对象进行部署和配置, 以大数据技术中目前流程的 Hadoop-2.7.6 版本开源项目为研究对象, 将两者项目进行融合部署配置, 在 OpenStack 私有云成功配置的基础上, 申

请云中资源池中的资源进行 Hadoop 完全分布式集群部署和配置, 最终使师生在该教学实训环节中能够成功实现技术配置, 进而对大数据和云计算技术有进一步的直观理解^[3]。

1 教学实训研究

基于 OpenStack 云架构的 Hadoop 分布式集群部署教学实训, 是两个开源合二为一的项目, 该实训项目可以设计在大数据技术专业中 Hadoop 课程章节学习过程中的一个项目实训, 也可以设计在云计算技术专业中 OpenStack 课程章节学习过程中的一个项目实训^[4]。该项目实训需学生提前具备或掌握《Linux 操作系统基础》《MySQL 数据库》和《计算机网络基础》等相关前置基础知识, 因其涉及知识点较为宽泛、部署步骤较多, 建议教学实训学时数 12~16 学时^[5]。

针对专业的课程侧重不同, 在该教学实训过程中也可具体侧重: ①针对计算机网络技术

收稿日期: 2023-09-27 修稿日期: 2023-10-28

基金项目: 河南省科技厅科技攻关项目(222102240010); 河南省教育厅教育科学、就业创业类研究项目(2020YB0499); 河南省大中专院校就业创业课题(JYB2023223); 河南应用技术职业学院青年骨干教师(2023-GGJS-X002); 河南省高等职业学校青年骨干教师培养计划(2020GZGG013)

作者简介: *通信作者: 周少珂(1987—), 男, 河南郑州人, 硕士, 讲师, 研究方向为计算机网络技术、云计算技术, E-mail: zhoushaoke@126.com; 王雷(1983—), 男, 河南郑州人, 硕士, 副教授, 研究方向为计算机应用技术; 石磊磊(1987—), 男, 河南郑州人, 本科, 工程师, 研究方向为云计算技术; 李凯(1995—), 男, 河南郑州人, 本科, 工程师, 研究方向为云计算技术

专业和云计算技术专业,仅重点学习掌握 OpenStack 私有云部署和配置即可;②针对大数据技术专业需要学习 OpenStack 私有云部署同时,进一步掌握 Hadoop 分布式集群部署和配置。

为使师生在实训项目部署过程中,更加深入地理解原理和掌握技能,整个项目使用网络中易于下载安装的 CentOS-7-x86_64-DVD-1511.iso 镜像、VMware16 虚拟机软件、SecureCRT 远程登录软件和开发的相关脚本工具。在具体部署时先基于底层的 VMware16 等虚拟机进行 OpenStack 私有云的双节点主机项目部署,而非使用网络中公有云进行云主机资源的申请,以便在部署过程中,让师生更加深入理解私有云项目的配置和最终的效果。在此成功部署基础上对 OpenStack 私有云资源池进行主机资源申请,进而对 Hadoop 大数据项目的配置部署,具体见表 1。最终完成全部实训部署,通过两个实训项目的对比操作,更加深入理解云计算和大数据两者技术的不同。

表 1 OpenStack 和 Hadoop 项目配置需求

配置	OpenStack 项目	Hadoop 项目
内存	≥32 GB	≥16 GB
CPU	≥16 核	≥8 核
部署顺序	先	后
前置知识	Linux、MySQL	Linux、MySQL
是否开源	开源	开源
Linux 操作系统	CentOS-7-x86_64-DVD-1511.iso	
VMware 虚拟机	VMware-workstation-full-16.0.0-16894299.exe	

2 OpenStack 私有云部署

云计算按照服务模式主要可分为基础设施即服务(infrastructure as a service, IaaS)、平台即服务(PaaS)和软件即服务(SaaS)三种;按照运营模式则主要是公有云、私有云和混合云^[3]。IaaS,提供的服务是对所有计算基础设施的利用,包括处理 CPU、内存、存储、网络和其它基本的计算资源,用户能够部署和运行任意软件,包括操作系统和应用程序,OpenStack 就是其私有云 IaaS 层面的开源项目。本项目部署采用双节点主机部署,即控制 Controller 节点和计算 Compute 节点,如图 1 所示。

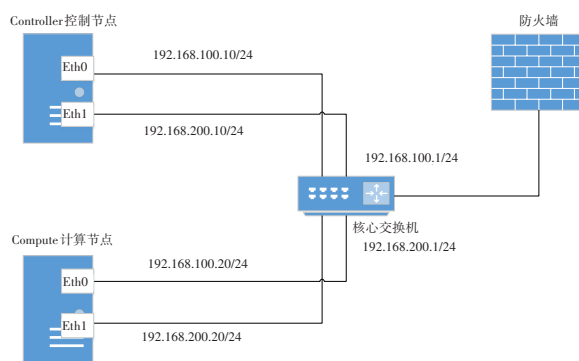


图 1 OpenStack 双节点拓扑图

在 VMware 虚拟机基础配置环节,需要开启虚拟化功能。在安装 CentOS7.2 操作系统时两台节点再添加一个网卡,VMnet1 使用“仅主机”模式、VMnet8 使用“NAT”模式,保证两个节点共计四张网卡按照规划的拓扑图中的 IP 地址配置能够成功 Ping 通。然后进行双节点网络 IP 地址和主机名的配置、CentOS7.2 镜像和 IaaS 镜像 YUM 源的配置、防火墙和 Selinux 等基础环境配置。使用提前编辑的基础环境脚本 iaas-xiandian 和 iaas-pre-host.sh 进行双节点安装和环境参数配置。

(1) 控制节点 Controller: 主要进行服务调度和控制工作。内存和 CPU 无需较多(8 GB 内存、4 核),能够保证节点中的服务正常稳定运行即可。该节点主要的服务功能为:MySQL 数据库服务、Keystone 认证服务、Glance 镜像服务、Nova 计算服务、Neutron 网络服务、Cinder 磁盘块服务、Swift 对象存储服务、Heat 资源编排服务、Cilometer 监控服务和 Dashboard 可视化仪表盘服务。

(2) 计算节点 Compute: 主要进行具体服务工作的执行。内存和 CPU(≥16 GB 内存、≥16 核)较多,最终云平台中的资源池可申请资源较多。该阶段主要的服务功能为: Nova 计算服务、Neutron 网络服务、Cinder 磁盘块服务、Swift 对象存储服务、Cilometer 监控服务,具体见表 2。

Cinder 磁盘块服务和 Swift 对象存储服务需要部署在计算节点 Compute 的空白分区中,应该在系统安装过程中留下足够的磁盘空间,系统安装完成后,使用 Fdisk 或 Parted 命令进行分区划分,然后使用 mkfs.xfs 进行文件系统格式化,

表 2 双节点主机所需安装的服务脚本

服务组件	Controller控制节点	Compute 计算节点
MySQL数据库	iaas-install-mysql.sh	无需安装
Keystone 认证	iaas-install-keystone.sh	无需安装
Glance 镜像	iaas-install-glance.sh	无需安装
Nova 计算	iaas-install-nova-controller.sh	iaas-install-nova-compute.sh
Neutron 网络	iaas-install-neutron-controller.sh	iaas-install-neutron-compute.sh
Cinder 磁盘块	iaas-install-cinder-controller.sh	iaas-install-cinder-compute.sh
Swift对象存储	iaas-install-swift-controller.sh	iaas-install-swift-compute.sh
Dashboard仪表盘	iaas-install-dashboard.sh	无需安装

完成空白分区的划分。

双节点中部署完毕所有组件功能脚本后，在电脑浏览器中输入 http://Controller 节点 IP 地址/dashboard，完成 OpenStack 私有云系统平台登录，在该云资源平台中可以查询出计算、存储和 CPU 等资源池中的所有资源。在该平台中进行网络(子网)、路由和三台云主机(使用 CentOS7.2 系统镜像、2 核、4 GB 内存)的创建，并启动该三台云主机，使用 SecureCRT 远程登录软件进行成功连接，进一步为在该 OpenStack 私有云平台上部署 Hadoop 完全分布式大数据集群(3 台主机)做好准备。

3 Hadoop 分布式集群部署

2002 年，Hadoop 起始于 Apache 项目的 Nutch 项目。开发者通过学习和借鉴 Google 的开源的 GFS、MapReduce 和 BigTable 三篇论文，开发创建出 Hadoop 项目。随着技术的进步，Hadoop 版本已经发展到 Apache Hadoop 3.3.6，其生态圈发展日益完善，包括 HDFS 分布式文件系统、MapReduce 计算模型、Zookeeper 协调一致性服务、Hbase 列式数据库、Hive 数据仓库、Flume 日志采集、Sqoop 数据泵等众多生态圈中的组件。

Hadoop 项目整体包括底层的大数据分布式集群部署和上层的大数据应用开发，本文主要针对底层大数据分布式集群部署进行研究，在 Hadoop 分布式集群部署中可以分为三类：①单机版 Hadoop 集群；②伪分布式 Hadoop 集群；③完全分布式 Hadoop 集群。其中完全分布式集群部署是真实集群环境，前两类集群仅供初学者学习参考。本文主要针对完全分布式 Hadoop

集群环境进行部署配置。此种方式能够直观看到分布式效果，并进一步对比与 OpenStack 私有云项目的不同。

在 OpenStack 私有云项目上运行的 3 台云主机进行集群环境部署，该 3 台云主机使用 Linux 系统的 CentOS7.5 发行版，由于 Hadoop 集群系统需要使用 Java 环境，因此首先对这 3 台主机分别进行 JDK1.8 环境变量的配置。整个分布式集群系统共有 3 台主机，其中 1 台主机同时承担 NameNode 节点和 DataNode 节点两种工作任务。剩余 2 台主机仅仅承担 DataNode 节点工作任务，整个集群配置过程中需对此 3 台机器进行同样配置，以保证数据的一致性，在配置部署前需将各主机防火墙进行关闭，SeLinux 功能禁用，并配置使用本地 Yum 软件源，分别修改主机名为 Hadoop01、Hadoop02 和 Hadoop03，修改主机名和 IP 地址的映射。

在进行文档配置部署时需对以下六个核心文件进行配置：①Hadoop-env.sh 文件；②Core-site.xml 文件；③Yarn-site.xml 文件；④Mapred-site.xml 文件；⑤Hdfs-site.xml 文件，该文件主要配置 HDFS 文件系统的一些目录、访问地址和块大小；⑥Master-Slave 文件，该文件主要保存和记录分布式集群数量和名称，使集群成为统一运行整体。该配置文件中需把所有 DataNode 数据节点主机名添加至该文件中。

通过上述六个核心文件的配置，整个 Hadoop 集群系统对 NameNode 主机的控制节点进行 Hadoop namenode format 命令格式化，在格式化节点过程中无需进行多次，以免使 NameNode 节点和 DataNode 节点不同步，影响系统的稳定性，

最后通过使用Start-all.sh命令启动集群。

分布式集群系统成功运行后,就可以在任
何一台主机上进行测试,有两种方式可以进行
测试:①通过浏览器Web登录界面方式,使用
谷歌浏览器(兼容性较强)输入http://hadoop01:
50070进入HDFS分布式存储组件服务;②通过
系统调用后台命令的方式,使用“jps”等命令。

4 结语

基于OpenStack云架构的Hadoop分布式集
群部署在教学实训过程中是两个有着先后顺序、
循序渐进,互为技能提升的实训项目。在前者
成功配置OpenStack私有云的基础上,能够促进
Hadoop大数据项目快速掌握,从而能够更加直
观地感知和理解云计算技术和大数据技术底层
的不同。该实训项目的学习在进一步促进巩固
Linux操作系统知识的同时,也为计算机网络技

术专业和云计算技术专业的虚拟化技术向深度
学习方向的转变,起到了促进的作用。

参考文献:

- [1] 周少珂,王雷,刘振民. 高职大数据技术与应用专业
Hadoop课程教学改革探究:以河南应用技术职业
学院为例[J]. 河南教育(高等教育),2021(7):
72-73.
- [2] 夏靖波,韦泽鲲,付凯,等. 云计算中Hadoop技术研
究与应用综述[J]. 计算机科学,2016,43(11):6-11,48.
- [3] 宋伟. 基于OpenStack私有云平台上的Hadoop完
全分布式环境应用研究[J]. 电脑知识与技术,
2020,16(11):234-236.
- [4] 谢显杰. 基于OpenStack的私有云平台构建研究
[J]. 信息与电脑(理论版),2022,34(5):88-91.
- [5] 尹乔,魏占辰,黄秋兰,等. Hadoop海量数据迁移系
统开发及应用[J]. 计算机工程与应用,2019(13):
66-71.

Research on the training and teaching process of Hadoop distributed cluster deployment based on OpenStack cloud architecture

Zhou Shaoke^{1*}, Wang Lei¹, Shi Leilei², Li Kai²

(1. Henan Vocational College of Applied Technology, Zhengzhou 450042, China;

2. Henan Shu'an Education and Technology Research Institute, Zhengzhou 450042, China)

Abstract: Vocational colleges are offering majors in cloud computing technology and big data technology, which are closely related and have different focuses. Professional courses also intersect, causing students to have certain misunderstandings and confusion when learning professional knowledge. Given that this is to solve the bottleneck and difficulty in the teaching and training process of teachers, and deepen students' mastery of big data and cloud computing technology. We will integrate and deploy the OpenStack project in cloud computing technology and the Hadoop project in big data technology. Based on the successful configuration of the OpenStack private cloud, we will apply for cloud resources to create a cloud host for Hadoop fully distributed cluster deployment and configuration. This will enable teachers and students to successfully achieve configuration in this teaching and training phase, further providing a more intuitive understanding of big data and cloud computing technology, and promoting in-depth learning of professional course knowledge.

Keywords: big data; Hadoop; OpenStack; cloud computing; teaching