现代计算机

XIANDAI JISUANJI

第29卷第1期（总第769期）

半月刊（1984年创刊）

2023年1月10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 广州一龙印刷有限公司

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

ISSN 1007-1423
01>

现代计算机 第29卷 第1期（总第769期） 2023年1月

2023年1月 第29卷

第1期 （总第769期）

ISSN 1007-1423
CN 44-1415/TP

MODERN COMPUTER

现代计算机

中山大学出版社 主办

MODERN COMPUTER

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

现代计算机
XIANDAI JISUANJI
第29卷 第1期（总第769期）2023年1月10日出版

目 次

研究与开发

图形图像

实践与经验

开发案例

期刊基本参数： CN44-1415/TP*1984*S*16*120*zh*P*¥30.00*1200*22*2023-01

Modern Computer
(Vol. 29, No. 1; Jan. 10, 2023)

CONTENTS

Research and Development

Graphic and Image

Practice and Experience

Development Solution

研究与开发

文章编号: 1007-1423（2023）01-0001-09 DOI: 10.3969/j.issn.1007-1423.2023.01.001

基于分时电价和多目标优化的家庭电能管理系统的设计与实现

者佳男*，吴 莹，李 皎

（西安石油大学计算机学院，西安 710000）

摘要：现如今，分时电价的盛行改变了原来的用电模式，不仅要保证用电费用降低，更要保证用户用电的舒适度。因此提出了基于分时电价和多目标优化的家庭电能管理系统。首先，将常用的家用设备按照家用设备的用电特性进行了分类。然后，家庭电能管理系统优化调度模型的建立以家庭用电成本的优化、用户舒适度的提升和碳排放量的追踪为目标。最后，对家庭电能管理系统模型使用改进的粒子群算法进行优化验证。通过使用Matlab进行仿真实验，验证该方法在降低用电成本、提高用户舒适度和追踪碳排放量方面有显著的优化效果。

关键词：家庭电能管理；分时电价；多目标优化；改进的粒子群算法；优化调度

0 引言

从目前国内外的研究来看，对于家庭微电网优化调度问题有大量的研究，文献[1-2]提出如何将居民用电设备进行分类；文献[3-6]提出基于居民用电设备的优化调度模型；文献[7-8]针对多目标的优化模型提出求解算法，然而在实现家庭微电网优化调度中仍存在许多问题。文献[9-10]分别提出了短期家庭负荷预测模型和日前日内需求响应优化调度两部分策略，但并未考虑到用户舒适度。文献[11]提出了在分时电价的大环境下，对家电负荷的运行进行优化，然而存在决策变量过多、人工求解困难等问题。文献[12-13]提到将电费最省和碳排放量最少作为目标，但目前国内的小区并没有普及清洁能源来代替煤炭发电的装置，因此碳排放量的减少需要更加深入的研究。文献[14]基于经济性和舒适度的多目标优化模型，但文中对于基本设备的优先级没有讨论，使得用户在设备优化上还有待提高。

综上，本文在没有光伏发电与风力发电等能源装置，结合了已有的家庭电能管理系统研究基础，考虑多维优化目标——用户舒适度、电价成本和碳排放量的追踪——建立在分时电价下的多目标优化模型，以保证不同家电用电行为下的最优调度。因此提出了基于分时电价和多目标优化的家庭电能管理系统的优化模型，该模型对家用设备进行了分类，并对设备进行监控，管理使其合理的运行达到最佳的运行状态。最后通过Matlab仿真算例验证该模型对家庭用电设备有较好的适应性，也保证用户舒适度。

1 家庭电能管理系统架构

家庭电能管理系统是电网在用户消费侧的重要组成部分，它利用一定的技术手段来实现

收稿日期：2022-05-26 **修稿日期：**2022-07-09

基金项目：国家自然科学基金资助项目（51707158，61802301）

作者简介：*通信作者：者佳男（1997—），男，硕士研究生，主要研究方向为家庭能量管理，E-mail：zjn806170452@163.com）；吴莹（1981—），女，副教授，硕导，主要研究方向为微电网能源系统；李皎（1982—），女，副教授，硕导，主要研究方向为区块链

用户对家庭用电设备的检测、管理和减少电费开支。在分时电价的大环境下合理地使用电器来达到减少电费，有效控制以家庭为主的碳排放，起到保护环境的作用。利用多目标优化算法使用户舒适度、电能消耗和碳排放达到最优。其架构如图1所示。

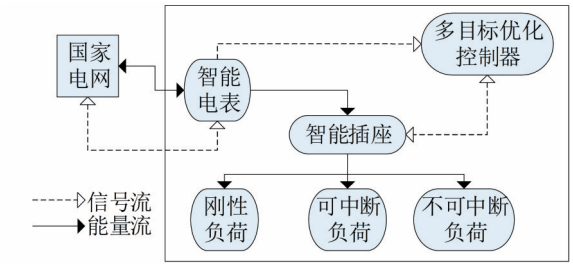


图1 家庭电能管理系统架构

系统主要包括智能插座，智能电表，多目标优化控制器。智能电表可以把家庭的用电情况上传给国家电网，然后电网以此来发布第二天分时电价的价格(引自国家能源网)，电网通过智能电表给用户家庭提供电能和提供分时电价的价格。智能插座连接用户的用电设备，获取用电设备的基本数据，例如功率，电压和用电时长等，并且通过电闸进一步来控制设备。收集到的数据通过智能插座上传给多目标优化器来进行分析。多目标优化控制器将得到的数据结合三个优化条件(用户舒适度、碳排放量和电能的消耗)和电网公布的分时电价及用户对设备的使用意向来进行优化，并把优化的结果传送给智能插座来控制设备的使用。

2 家庭电能管理系统优化模型

家庭电能管理系统是针对拥有多个家居用电设备的单个用户，这些设备通过智能插座全部与管理系统相连。当HEMS开始运行后，这些设备就会立即作出反应。在整个优化过程里，系统会将优化的时间划分为长度相同的子时间段。

2.1 用电负荷模型

在日常生活里，每个用电设备都有不同的用电方式，本文依据设备间的用电特性将所有家用设备分为刚性设备、柔性设备和HVAC设备。刚性设备表示用电设备的使用时间不会根据优化系统而改变，例如冰箱，电灯等。柔性设备又根据设备使用时间在平移过程中可否中断分为两类：①可平移不可中断设备，如洗碗机、洗衣机，其运行时段可以进行移动、调整，但工作期间设备不可中断；②可平移可中断设备，表示此类家用设备的使用时间根据优化系统而发生变化，并且在运行期间可中断使用，如电动车，热水器。HVAC设备表示此类设备的使用情况只与温度有关，如空调。

2.1.1 刚性设备

该类设备直接影响着用户的正常生活，不会随着优化发生改变，例如电冰箱，电灯等。

刚性设备模型：

$$\begin{cases} P^{c} = S_{c}(t) * P^{c,\,\mathrm{rated}} \\ E^{c} = P^{c,\,\mathrm{rated}} * d_{c} \\ S_{c}(t) = 1, t \in [T_{c}^{\mathrm{start}}, T_{c}^{\mathrm{end}}] \\ S_{c}(t) = 0, t \notin [T_{c}^{\mathrm{start}}, T_{c}^{\mathrm{end}}] \end{cases} \tag{1}$$

其中：$c \in C$为刚性设备；P^{c}为设备c在t时刻的瞬时功率；$S_{c}(t)$为设备c在t时刻的工作状态，$S_{c}(t)=1$表示设备处于开启状态，$S_{c}(t)=0$表示设备处于关闭状态；$P^{c,\,\mathrm{rated}}$为设备c的额定功率；T_{c}^{start}为设备开启时间，T_{c}^{end}为设备结束时间；E^{c}为设备C的总用电量；d_{c}为刚性设备C的总工作时间。

2.1.2 柔性设备

该类设备的用电时间较为灵活，依据用电特性和用电过程中可否中断分为可平移不可中断设备和可平移可中断设备。

可平移不可中断设备模型

$$\begin{cases} T_{i}^{\mathrm{start},\,s} \leqslant T_{i}^{\mathrm{start}} \leqslant T_{i}^{\mathrm{start},\,e} \\ T_{i}^{\mathrm{end}} = T_{i}^{\mathrm{start}} + d_{i} \\ E^{i} = P^{i,\,\mathrm{rated}} + d_{i} \end{cases} \tag{2}$$

式中：$i \in I$为不可中断设备；$T_{i}^{\mathrm{start},\,s}$、$T_{i}^{\mathrm{start},\,e}$为启动时间范围的上下限；$T_{i}^{\mathrm{start}}$为设备$i$的启动时间；$T_{i}^{\mathrm{end}}$为设备$i$的结束时间；$d_{i}$为设备$i$的工作时长；$E^{i}$是不可中断设备$i$的总用电量。

可平移可中断设备模型：

$$\begin{cases} T_k^{\text{start},s} \leqslant T_{k,1}^{\text{start}} \leqslant T_k^{\text{start},e} \\ d_k = \sum_{j=1}^{c}(t_{k,j}^{\text{end}} - t_{k,j}^{\text{start}}) \\ t_{k,j}^{\text{end}} \geqslant t_{k,j}^{\text{start}}, j = 1,2,\cdots,M \\ t_{k,j+1}^{\text{start}} \geqslant t_{k,j}^{\text{start}}, j = 1,2,\cdots,M-1 \\ E^k = P^{k,\text{rated}} * d_k \end{cases} \tag{3}$$

式中：$k \in K$为可中断设备；$t_{k,j}^{\text{start}}$为设备k的第j个工作时段的起始时间；$t_{k,j}^{\text{end}}$为设备k的第j个工作时段的终止时间；M为总工作时段数量。E^k是可中断设备的总用电量。

2.1.3　HAVC设备

暖通空调系统(heating，ventilation and air conditioning，HVAC)是包含温度、湿度、空气清净度以及空气循环的控制系统，因此室内外温度，以及用户自行设置的最佳温度值和房屋的热参数决定了该类设备的运行时间。所以设备功耗P_{AC}与室温T_{in}之间的关系模型为

$$T_{\text{in}}(t+1) = \mathrm{e}^{-\triangle t/RD} T_{\text{in}}(t) + R(\mathrm{e}^{-\triangle t/RD} - 1)P^{AD}(t) + (1 - \mathrm{e}^{-\triangle t/RD})T_{\text{out}}(t) \tag{4}$$

式中：T_{in}、T_{out}分别为室内温度，室外温度，单位℃；R为房间热阻，单位℃/kW；D为房间热容量，单位kWh/℃；P^{AD}为空调功率，单位kW；$\triangle t$为温度差。

2.2　优化调度目标函数

2.2.1　经济目标函数

经济目标函数是把家庭用电费用作为优化目标。优化后的值越小表示优化效果越好。

$$F_1 = \sum_{a=1}^{n} P_a(t) * i(t) \tag{5}$$

其中：F_1为到t时刻为止所用的电费；$P_a(t)$为t时段设备的用电量；$i(t)$为设备在t时段电量的分时电价。

2.2.2　用户舒适度函数

用户舒适度函数是将用户对用电设备优化后的使用时间段的满意度作为优化目标。优化后的数值越大，表示用户对于优化后的设备使用时间满意度越高。

$$f_x = \begin{cases} \beta_x * \dfrac{t_x^{\text{end}} - T_x^{\text{end}}}{T_x^{\text{start}} + T_x^L - T_x^{\text{end}}}, T_x^{\text{start}} + T_x^L \leqslant t_x^{\text{end}} \leqslant T_x^{\text{end}} \\ 0, t_x^{\text{end}} < T_x^{\text{start}} + T_x^L \text{或} t_x^{\text{end}} > T_x^{\text{end}} \end{cases} \tag{6}$$

$$F_2 = \sum_{x=1}^{k} f_x \tag{7}$$

其中：β_x为用户对设备x的满意度影响因子，β_x取值为非负整数，影响因子越大，表示用户希望在优化时间内此设备能优先完成工作；f_x为用户对用电设备x的使用舒适度；T_x^L为用电设备x需要工作的时长；T_x^{start}、T_x^{end}为用户希望用电设备x的启/停时间；t_x^{end}为用电设备x实际的停止时间；F_2为总的用电设备满意度；K为可调度设备总数。

2.2.3　碳排放量函数

家庭侧碳排放有很多种，例如汽车尾气排放，煤气灶生活做饭等，本文选择煤炭发电导致的碳排放，根据市场调查每消耗一度电的碳排放量是0.785 kg。就目前国内的情况，小区没有安装光伏发电装置，碳排放量与用电量消耗的线性关系如下：

$$F_3 = \left(\sum_{a=1}^{n} P_a(t)\right) * 0.785 \tag{8}$$

其中：F_3是碳排放量；$P_a(t)$是a设备所消耗的电量。

在以后的发展中，小区里有了光伏发电装置后，碳排放量与光伏发电的使用效率有关，当光伏发电的用电效率达到最大时，碳排放量达到最小；反之当光伏发电的用电效率达到最小时，碳排放量达到最大。在光伏发电装置中，需要有发电、储电的过程，本文不再赘述。

2.3　约束条件

2.3.1　功率对等约束条件

在家庭系统中电能需要守恒，即生产的电能与消耗的电能应该保持平衡。功率平衡约束如下：

$$\sum_i W_i * X_{ij} + M_j = P_j \tag{9}$$

式中：W_i表示设备i的平均功率；X_{ij}表示第i个电器在j时间段的工作状态，工作时值为1，不工作时值为0；M_j表示各时间片家庭中不可调度

或必须运行的电器的总功率；P_j表示在j时间段内电网提供的功率。

2.3.2 温度设备约束条件

用户对于温控设备的运作时间与工作时段并不关心，关心的是在特定的时间内由温控设备控制的温度有没有达到要求。所以对温控设备来说不存在时间约束，只存在温度约束。其温度约束如下：

$$T_K(t) = T_S(t) + \triangle T \tag{10}$$

式中：$T_K(t)$为t时刻房间的实际温度；$T_S(t)$为t时刻房间的理想温度；$\triangle T$为用户允许的温差。

2.3.3 工作时间约束条件

对某一电器设备来说，需要工作的时间片数为N_i，HEMS需要保证每一个电器设备都能够完成工作。所以每个电器设备工作的时间片数应该与N_i相等，即：

$$\sum_{j=1}^{24} X_{ij} = N_i, \forall i \in A \tag{11}$$

其中：X_{ij}表示第i电器在第j时间段的工作状态，工作时值为1，不工作时值为0；A为可中断设备与不可中断设备的集合。

设备最短连续工作时长约束中，可中断设备的最短连续工作时长一般小于需要工作的总时长，而不可中断设备的这两个值相等。最短连续工作时长约束表示如下：

$$X_{i(j+x)} = 1, \forall x \in [0, L_i] \text{且} B_{ij} = 1 \tag{12}$$

式中：L_i表示最短连续工作时间片个数，即最短连续工作时长；B_{ij}表示电器i在时间片j执行了启动操作。

2.4 多目标优化调度策略

2.4.1 多目标优化算法

在2.2节介绍的3个目标(用电费用、用户舒适度和碳排放量)中，如果仅将用电费用优化到最低，在电价较低时才使用电，不可避免地会引起用户的不满。同样，当一个目标达到最佳时，会对其他两个优化目标产生一定的影响。本文的多目标优化调度策略综合考虑用电费用、用户舒适度以及碳排放量的因素，使得达到最优。

传统多目标优化方法都是根据对不同目标的不同要求，将目标通过某种方式转变为单目标优化可以求解的问题。缺点是仅能求出优化问题的局部最优解，求解的结果强烈依赖于初始值。智能优化算法的提出改善了这方面的问题。目前最流行的智能算法就是粒子群算法。

2.4.2 粒子群算法

粒子群(particle swarm optimization，PSO)算法从随机触发，通过迭代寻找到最优解。粒子群的标准算法容易陷入局部最优，导致结果不准确，且不能有效解决离散及组合优化问题。因此本文使用改进的多目标粒子群算法，在该算法中每个粒子的速度和位置按照如下方程来更新：

$$v_{s,d}^{(k+1)} = wv_{s,d}^{(k)} + c_1 r_1 (p_{s,d}^{(k)} - x_{s,d}^{(k)}) + c_2 r_2 (p_{g,d}^{(k)} - x_{s,d}^{(k)}) \tag{13}$$

$$x_{s,d}^{(k+1)} = x_{s,d}^{(k)} + v_{s,d}^{(k+1)} \tag{14}$$

$$\begin{cases} c_1 = c_{\max} - (c_{\max} - c_{\min}) * \dfrac{k}{k_{\max}} \\ c_2 = c_{\min} + (c_{\max} - c_{\min}) * \dfrac{k}{k_{\max}} \end{cases} \tag{15}$$

式中：$x_{s,d}$为第s粒子在d维搜索空间上的位置；$v_{s,d}$为粒子的速度；$p_{s,d}$为个体极值；$p_{g,d}$为种群的群体极值；w为惯性因子；c_1、c_2是非负的常数，称为学习因子；r_1、r_2是分布在0和1之间的随机数；k为当前迭代次数；$k_{\max}$为最大迭代次数。

$$\begin{cases} w = w_{\min} - (w_{\max} - w_{\min}) \dfrac{f - f_{\min}}{\overline{f_h} - f_{\min}}, & f \leqslant \overline{f_h} \\ w = w_{\max}, & f > \overline{f_h} \end{cases} \tag{16}$$

式中：f是当前适应度值；$\overline{f_h}$是平均适应度值；$f_{\min}$是最小适应度值；$w_{\max}$、$w_{\min}$分别为惯性因子最大值和最小值。

2.4.3 算法比较

图2是标准的粒子群算法流程图，图3是改进的粒子群算法流程图。

比对图2与图3能看出，改进的粒子群算法在标准的粒子群算法基础上，改变了权重因子w和学习因子c的取值方法，有效地提高了粒子群算法的全局搜索能力，避免陷入局部最优。

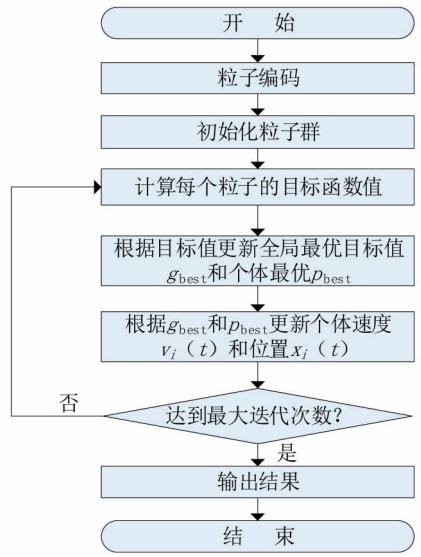


图 2　标准的粒子群算法流程

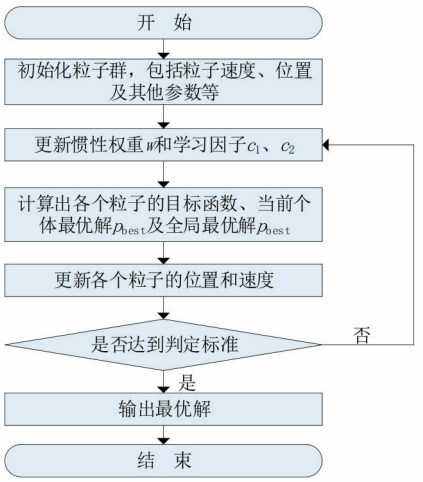


图 3　改进PSO算法基本流程

3　家庭电能管理系统

家庭电能管理系统运作流程如下：

（1）在优化前，多目标优化控制器获取到设备的所有参数，其包括用电设备的额定功率和工作时长，以及用户希望设备工作的时间段。

（2）智能电表将获取到的分时电价信息上传给多目标优化控制器。

（3）多目标优化器通过平移柔性设备的运行时间来节省电费，通过每一天的电量使用检测追踪用户侧的碳排放量，通过使设备在用户理想的时间段运行来保证用户满意度，最后结合约束条件，制订各柔性负荷的调度计划。

（4）多目标优化控制器把调度计划上传给智能插座，使其在相应的时间段运行各用电设备。

调度流程图如图4所示。

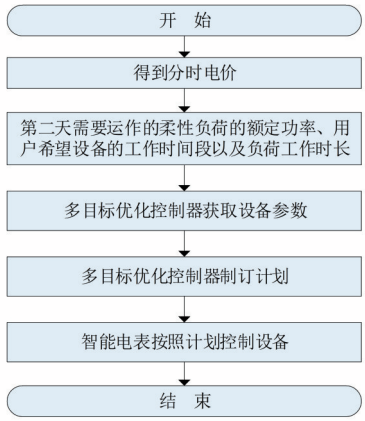


图 4　家庭电能管理系统运作流程

4　实例分析

4.1　算例参数

本文以某一家庭数据为例进行分析。该用户家庭安装有智能电表和多目标优化控制器，根据设备的特性不同，其家庭用电负荷情况不同，刚性负荷如表1所示，柔性负荷如表2所示。

表 1　刚性设备参数

设备名称	功率/kW	工作时间范围	数量
电冰箱	0.60	0：00—24：00	1
电视机	0.08	0：00—24：00	2
客厅荧光灯	0.05	0：00—24：00	1
卧室白炽灯	0.03	0：00—24：00	3
计算机	0.30	0：00—24：00	1

家居设备优化的目的在于保证各设备完成既定工作的前提下，尽可能压缩用户的电能成本，同时保证负荷曲线峰谷差不致过大；用户可以根据各自的习惯在HEMS中设置设备优化参数，根据这些参数以及电网提供的分时电价信息，HEMS可对家居用电设备进行运行优化。例如，用户可设置在晚上回家后的18：00至次日7：00之间对电动车进行充电操作，为了不影响用户在第二天对电动自行车的正常使用，

表 2 柔性设备参数

设备名称	功率/kW	工作区间	工作时长/h	偏好时间范围	满意度影响因子
洗衣机	1.0	08：30—21：00	1.0	15：30—16：30	16
洗碗机	0.4	08：00—12：00 13：00—17：00 19：00—24：00	2.0	08：00—10：30 13：00—15：30 19：00—21：30	14
热水器	1.5	08：00—21：00	0.5	19：00—21：00	15
电水壶	1.5	05：00—07：00 08：00—12：00 13：00—17：00	0.5	05：30—06：30 11：00—12：00 15：30—16：30	20
消毒柜	0.5	08：00—12：00 14：00—17：00 20：00—24：00	0.5	11：00—12：00 16：00—17：00 21：00—22：00	14
电动车	2.0	20：00—08：00	2.0	00：00—08：00	13
电饭煲	0.6	05：00—07：00 10：00—12：00 15：00—18：00	0.8	05：30—06：30 11：00—12：00 16：00—17：00	20

须保证充电结束时充电量达到其总容量的90%以上，HEMS会根据电网提供的分时电价信息选择合适的充电时段，达到压缩电能成本的目标。其次对于空调的使用，全时段待机等候命令，当室内温度不在规定的温度范围时，空调开始运作，直到温度达到用户设定的温度则停止运作。空调的参数设置如表3所示。

表3 空调相关参数

功率/kW	室温温度设定/℃	温度范围/℃	R房间热阻/℃/kW	C房间热容量/kWh/℃
2	24	23～26	40	0.025

表 4 某地区分时电价定义

时段	时间跨度	电价/(元/(kW·h))
峰时段	10：00—15：00 18：00—21：00	0.9864
平时段	07：00—10：00 15：00—18：00 21：00—23：00	0.6770
谷时段	23：00—07：00	0.3766

4.2 仿真结果

在对整个数据进行分析的过程中，使用了Matlab软件，并将最终得到的结果以图的方式表现出来。

从图5可以看出，当室外温度大于室内温度时，空调开始运作，直到室温保持在温度设定范围内则停止工作。当空调处于待机状态时，消耗的电量可忽略不计。

图6是标准的粒子群算法下的优化调度结果，图7是改进的粒子群算法优化后调度结果。从两幅图可知，将负荷从峰时段平移到谷时段或者平时段，来减少电费。

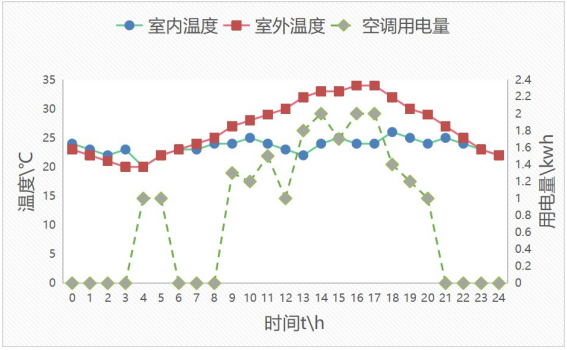


图 5 HVAC设备的优化结果

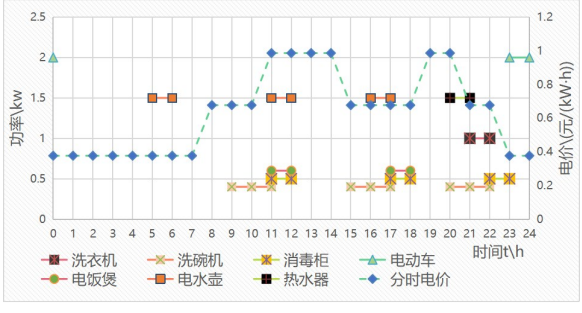


图 6 标准粒子群算法优化结果

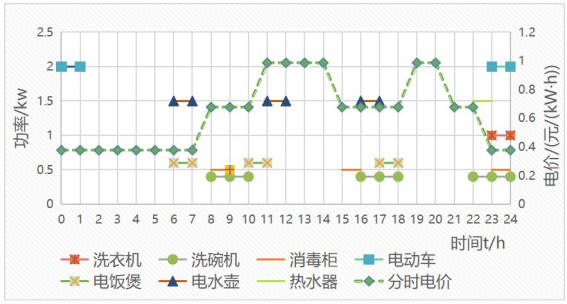


图 7　改进粒子群算法优化结果

图8为使用多目标优化系统前后各个时段的电量费用对比，家庭电能管理系统负荷优化调度结果充分说明了系统策略调度的经济性电费大部分都产生在峰时段，而优化后HEMS系统将一部分负荷转移到平时段和谷时段，从而减少了峰时段电量的高运作。

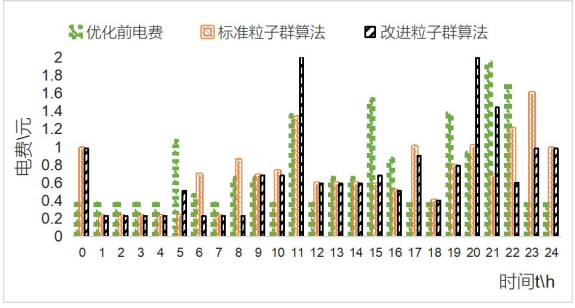


图 8　每个时段电费对比

图9多目标优化系统前后各个时段的电量总费用对比，可以看出优化后的费用增加平缓，并低于优化前电费。从两幅图中能够明显地对比出，峰时段和平时段购电支出都有所降低，谷时段略有升高。从图8和图9能看出，使用标准的粒子群算法和改进的粒子群算法都能够减少电费，标准的粒子群算法比优化前减少了7.2%，改进的粒子群算法减少了4.3%。

图10为使用多目标优化系统前后各个时段的功率对比图，使用优化系统之前，大部分的柔性设备都集中在了峰时段，也就是11：00—13：00，19：00—21：00；在优化后，将一部分设备从峰时段移动到谷时段，有效避开了峰值电价时间段，达到了设备平移的目的，起到了有效的削峰填谷作用。

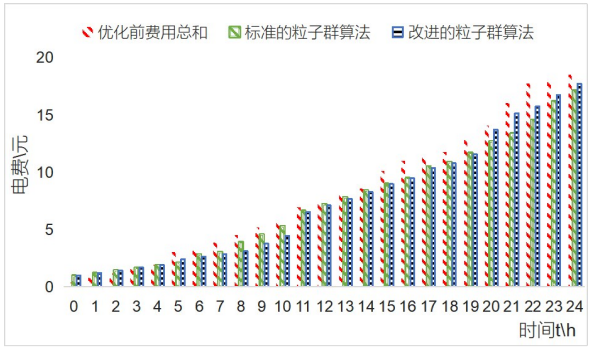


图 9　优化前后电费总和对比

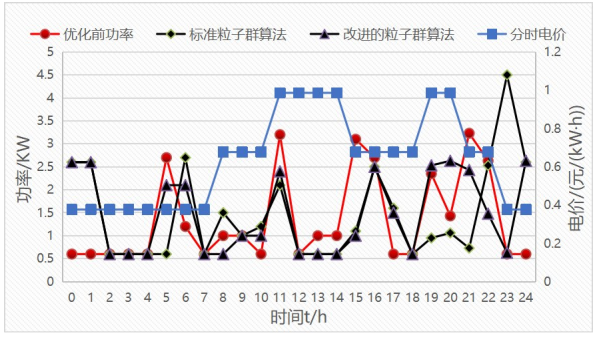


图 10　优化前后总功率对比

表5为优化前后关于电费、满意度和碳排放量的数据对比。从表中可以清晰地看出使用了优化算法后电费都减少了，但用户对设备的使用时间满意度，改进的粒子群算法比标准的粒子群算法要高。标准的粒子群算法的电费降低了7.2%，满意度降低了23.9%，改进的粒子群算法电费降低了4.3%，满意度降低了11.8%，从数据能够看出，标准的粒子群算法容易陷入局部最优，最后的优化结果不是很理想。

表 5　优化前后的对比

	电费/元	舒适度	碳排放量/kg
优化前	18.50	113.70	18.38
标准的粒子群算法	17.16	86.50	18.38
改进的粒子群算法	17.71	100.20	18.38

图11为两种算法的迭代次数收敛曲线。从图中可以明显看出，改进的粒子群算法在迭代150次左右趋于平稳。而标准的粒子群算法在

250次左右才趋于平稳。这也证明了无论是收敛效果还是优化精度，改进的粒子群算法比标准的粒子群算法都更好一些。

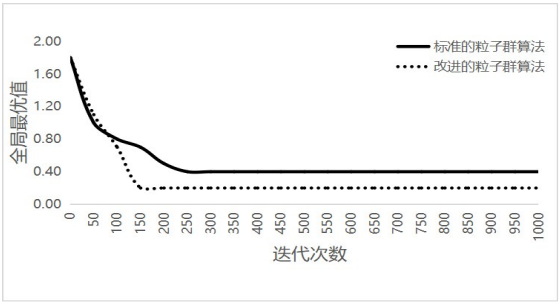


图 11　两种算法的收敛曲线

5　结语

本文在构建家庭电能管理系统架构，家庭用电负荷模型的基础上，提出了一种基于实时电价和多目标优化的家庭电能管理系统的调度与实现。以某一个家庭用户为例，在Matlab平台下进行了数据的仿真模拟。通过数据的仿真结果，我们可以得出：本文提出的家庭电能管理系统能够更好地兼顾用户侧的用电成本和用户对用电设备的舒适度，也从家庭方面针对碳排放做了检测与追踪，对环境保护起到一定的作用，对以后减少碳排放做出贡献。

在未来，以小区为单位有了光伏发电装置之后，可进一步优化电费，同时对碳排放量的减少也可进行优化，来增加HEMS的经济效益。并且小区的所有用户都使用了能量管理系统后，与区块链技术相结合，以此来共享发电成本。每个家庭都采用其最佳策略来最小化能源消耗成本，用户可以维护自己的能源消耗隐私。此外，能量管理系统分布在区块链上，为参与者提供一个可信的通信媒介。稳固智能合约旨在交易的执行，无需智能社区第三方的参与。

参考文献：

[1] 武东升. 家庭能量管理系统用电与电能调度优化策略研究[D]. 哈尔滨:哈尔滨工业大学,2017.

[2] ROH H T, LEE J W. Residential demand response scheduling with multiclass appliances in the smart grid [J]. IEEE Transactions on Smart Grid, 2016, 7(1): 94-104.

[3] 于军琪,侯雪妍,赵安军,等. 基于改进天牛须算法的家庭能量管理系统多目标优化调度[J]. 电力系统及其自动化学报,2022,34(5):79-86.

[4] WANG X, MAO X, KHODAEI H. A multi-objective home energy management system based on internet of things and optimization algorithms[J]. Journal of Building Engineering, 2021, 33: 101603.

[5] ALIÇ O, FILIK Ü B. A multi-objective home energy management system for explicit cost-comfort analysis considering appliance category-based discomfort models and demand response programs[J]. Energy and Buildings, 2021, 240: 110868.

[6] KAKRAN S, CHANANA S. Optimal energy scheduling method under load shaping demand response program in a home energy management system[J]. International Journal of Emerging Electric Power Systems, 2019, 20(2).

[7] WANG S C, YAN J, XU X, et al. An optimization strategy of smart home energy management system [C]//IOP Conference Series: Earth and Environmental Science. IOP Publishing, 2019, 237(4): 042026.

[8] 张延宇,曾鹏,臧传治. 智能电网环境下家庭能源管理系统优化调度算法[J]. 电力系统保护与控制,2016,44(2):18-26.

[9] ZHAO X, GAO W, QIAN F, et al. Electricity cost comparison of dynamic pricing model based on load forecasting in home energy management system[J]. Energy, 2021, 229: 120538.

[10] 仲志强,成海生,蔡华,等. 基于分时电价和需求响应的家庭微电网系统协同控制策略及其实现[J]. 现代电力,2021,38(1):69-78.

[11] 何可人,滕欢,郭宁. 考虑峰谷差因素的用户侧负荷优化运行[J]. 科学技术与工程,2016,16(27):177-182.

[12] 官拓颖. 家庭能量管理系统中能量优化调度模型与算法的研究[D]. 长沙:湖南大学,2016.

[13] 陈娅兰. 基于多目标优化的家庭能量管理系统调度模型及算法研究[D]. 北京:北方工业大学,2017.

[14] 李菁,王志新,严胜,等. 计及舒适度的家庭能量管理系统优化控制策略[J]. 太阳能学报,2020,41(10):51-58.

（下转第15页）

文章编号: 1007-1423(2023)01-0009-07 DOI: 10.3969/j.issn.1007-1423.2023.01.002

图书馆智能定位盘点机器人的研究——以上海图书馆为例

吴佳圆

（上海图书馆上海科学技术情报研究所，上海 200031）

摘要：介绍了上海图书馆近几年对智能盘点机器人的研究与应用，回顾了国内图书盘点的发展历史，汇总了我国优秀自主研发案例，并在理论基础之上对智能定位盘点机器人进行现场测试分析，探讨新技术、新模式对图书盘点业务的提升与改善，并在最后给出进一步的思考与建议。

关键词：智慧图书馆；RFID；盘点机器人

0 引言

传统图书盘点需要耗费大量的人力与时间，而 RFID（radio frequency identification）技术在图书馆行业的普及拓展了图书智能定位盘点业务的发展空间。现今，在机器人技术日益成熟的大环境下，RFID 技术与智能机器人的有机结合将有力推进智慧图书馆的发展。为了早日实现这一目标，上海图书馆以东馆开放为契机，着手引入智能定位图书盘点机器人。这一智能设备将大大提升图书盘点的效率，不仅减轻了馆员的体力工作，还能为读者提供更好的借阅体验。

1 智能定位盘点的介绍

图书盘点是图书馆文献管理的一项重要的基础工作，其目的是发现馆藏文献资源的排放问题、书目数据等问题。同时，通过图书盘点也可以查找遗失文献，从而确保馆藏文献的完整性。图书盘点作为图书馆日常工作之一，是保障图书馆开展业务的前提。

早期的图书盘点方式主要有人工清点、条码扫描盘点，如今随着 RFID 技术在国内图书馆的普及，基于此技术的智能定位盘点系统开始崭露头角。它是指图书馆工作人员利用智能设备对书架上的书进行核对和清点，检查有无错架、漏架。这些智能设备的最大特点就是全都运用了 RFID 技术，通过该技术设备可自动扫描多本图书的电子标签信息，并给出结果提示，从而引导工作人员完成图书盘点工作。该盘点模式根据盘点设备的不同，衍生出静态人工盘点也称手持盘点、盘点车盘点、动态 RFID 智能书架、智能盘点机器人这四种形式。

智能定位盘点的应用简化了图书馆的日常工作，将一线工作人员从繁杂的图书盘点中解放出来。除此之外，与传统图书盘点方式相比较，智能定位盘点在盘点成本、盘点效率和准确率、盘点工作的常态化、创新服务等方面都具有一定的优势[1]。

2 智能盘点机器人的研究现状

智能盘点机器人是智能定位盘点的一种应用形式，它将 RFID 技术与机器人技术结合起来，能为图书馆的图书动态盘点工作提供支撑[2]。相较于其他三种智能定位盘点技术，盘点机器人在精确度、高效性、智能化方面都领先一步。未来，图书馆运用盘点机器人进行图书盘点将是大势所趋，而本章节内容也将就智能盘点机器人进行更深入的研究。

2.1 产品主要设计理念

目前，基于对图书智能盘点的需求，盘点

收稿日期：2022-08-25 **修稿日期：**2022-10-26

作者简介：吴佳圆（1992—），女，江苏海门人，本科，助理馆员，从事领域为图情管理

机器人需要拥有数据采集系统和导航定位系统。

2.1.1 超高频RFID与机器人的结合

数据采集系统主要依赖于RFID技术，又称无线射频识别。该技术可通过无线电讯号识别特定目标并读写相关数据，而无需建立机械或光学接触。在图书盘点方面，图书馆一般采用芯片内置方式把电子标签放置在书籍当中，使每一本书籍的馆藏条码与电子标签相关联。届时，只需通过智能设备终端对图书进行扫描，就可以非接触式地快速识别粘贴在图书上的RFID标签，从而实现对图书的盘点管理。

RFID技术按照工作频率，可分为低频、高频、超高频和微波系统。相较于其他频段，超高频系统具有传输距离远、穿透性强、使用寿命长、稳定性好、存储容量大等特点。在图书盘点领域，香港城市大学的研究表明，超高频RFID在标签读取速率、标签读取范围、标签检测方向、防冲突等方面都优于高频RFID[3]。

2.1.2 SLAM自适应导航功能

盘点机器人如果要实现全方位替代人工进行顺架、排架工作，必不可少的一项功能就是灵活自主地在书架间进行移动。它要能在移动的过程中实时检测环境，自主建造增量式地图，感知到整层楼中书架排放、充电设备等环境，然后利用构建的地图进行自主定位和导航[4]，而不会因为环境的变化，比如书架、桌椅的挪动或者读者、工作人员的阻挡等造成机器人不能正常行走的情况。目前大部分智能移动机器人是通过SLAM(simultaneous localization and mapping)，也就是同步定位与地图构建算法来实现定位导航。SLAM是一种基于概率计算的增量式定位建图算法，它利用各类传感器，如激光测距器、深度相机、超声波等，搜集场景数据信息并对当前位置进行后验概率估计，而后利用运动模型加以校正，从而得到机器人当前位置信息，构建高精度周围场景地图，并进行导航工作。SLAM算法是目前最成熟的机器人定位导航算法，已在商用机器人上广泛应用[5]。

运用在图书馆内的智能盘点机器人同样配备SLAM自适应导航功能。设备在正式投入使用时，需要先进行图书馆现场环境建图。机器人自身搭载的激光雷达发射出激光，通过计算反射回的激光的角度、时间等确定障碍物的位置，达到构建2D地图的作用。一般来说，为15000 m^2的环境建图大约需要10天至15天。

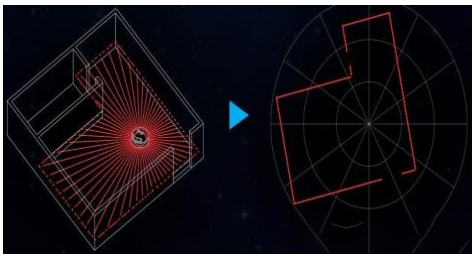


图1 SLAM地图构建示意图(一)

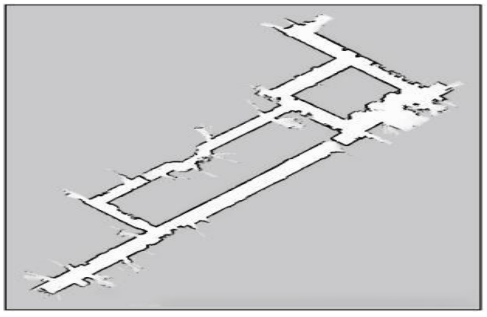


图2 SLAM地图构建示意图(二)

2.2 我国自主研发案例

以上从理论角度探讨了RFID技术、SLAM导航与机器人结合的可能性。技术层面，在2015年，南京大学计算机软件新技术国家重点实验室陈力军教授团队和南京大学图书馆联合研发了全球首台智能图书盘点机器人——图客，开创了理论实践的先河。在此之后，上海纵智、杭州横竖、北京金至等科技公司也轮番入场了图书盘点机器人的领域。

2.2.1 江苏图客

2019年7月，江苏图客机器人有限公司注册成立，其研发的“图客”机器人已更新到第四代产品：移动硬件搭载差速轮，并配备无刷伺服电机，实现了机器人的静音行走；导航模块由SLAM双激光360°导航，使机器人可以在书架间双向行走，路径更为灵活、准确；此外，机器人有两套独立升降模块，可以更好地适应不同书架的层高；在安全性能上添加了阻燃层模块，保障机器人的运行安全。“图客”在结合

图书识别定位技术、多模态自主导航技术和OCR视觉技术(智能避障和书脊信息识别)的基础上，实现了盘点机器人每小时两万册的盘点效率，98%的定位精确度，1%的漏读率。如今，该款机器人已经在南京大学、同济大学、宁波大学等图书馆投入使用。“图客”实物图如图3所示。

图3　图客机器人

2.2.2　上海飒智

2017年底，上海高新技术企业飒智智能科技有限公司成立，该公司致力于智能移动共融机器人产品的研发、集成与应用，并于2018年推出了其自主研发的融合人工智能、计算机视觉、大数据处理、物联网等先进技术的图书盘点机器人。

在机械方面，该款机器人配有四个高频RFID天线，可同时对四层书架进行盘点；增加了可调天线间距的设计，天线升降范围可达90～2400 mm，适用于不同书架，提高盘点的读到率和层定位精度。机器人的机械结构经过专门的防共振设计，工作最大噪声不超过50分贝，可在安静的图书馆环境中使用。标签读取方面，除了为机器人配备高端千标签/秒级吞吐量RFID读写器外，同时辅以OCR视觉识别技术，采用人工智能算法对索书号标签进行识别，精准定位标签损坏或RFID信号弱的书籍。RFID和OCR两种读取方法相互补充，可实现高达99.7%的盘点效率。导航与行动方面，在设备部署阶段，机器人根据自主研发的建图算法和其自带的高精度激光雷达进行场馆地图建立工作。除了地面附近的激光雷达之外，机器人四周一定高度都装有超声探测装置，用于规避桌椅、行人和透明物体。当检测到机器人周围有障碍物时，机器人会首先进行减速，直到停止。2021年，飒智机器人在不断更新换代后已经实现了100架(7层书架)/小时的盘点效率，高于99.7%的图书读取率，8小时续航时长以及厘米级的定位粒度。是人工盘点效率的60倍。“飒智”实物图如图4所示。

2.2.3　北京金至

北京金至科技股份有限公司是国家高新技术企业，始立于1998年中关村。公司业务方向涉及云计算、大数据、智能维保、信息安全等领域。“金至国胜壹号”是金至科技凭借多年的RFID技术实践经验和深入调研，在分析图书馆业务的基础上，结合智能机器人技术、激光扫描技术、RFID射频识别技术、云计算与大数据分析，自主研发的一款图书馆智能盘点机器人，实现了图书馆无人值守的全自动盘点、顺架、扫描等工作。通过WIFI网络将数据上传至云平台，进行大数据分析、运算，生成在架、错架、遗失、外借等数据报表。设备设计新颖，运行稳定，安全可靠，是一款高性能的盘点助手。“金至”实物图如图5所示。

图4　飒智机器人

图5　金至机器人

3　智能盘点机器人在上海图书馆(东馆)的应用展望

上海图书馆东馆开放在即，新馆更大的阅

读空间与更多的在架图书都对未来上图东馆的综合服务能力提出了更高的要求。尤其是三楼开放式的阅读广场，对读者来说是一种全新的阅读体验，38万册图书也将对书架准确率提出新的挑战。因此，在上图东馆运行智能盘点机器人进行图书盘点势在必行，为了检验其可操作性，上海图书馆挑选了其中一款盘点机器人在中文书刊外借室进行先行测试。希望通过此次现场测试，再结合东馆的实际情况，加强与厂商的交流沟通，从而挑选、打磨出一款适合上海图书馆东馆的盘点机器人。

本次测试分为两个阶段进行，分别为常规场景和非常规场景的适应性测试。

第一阶段常规场景测试选取了两排四列书架共965本图书，主要通过RFID+OCR两种方式进行相互补充盘点，对标签弱的图书采用OCR对索书号进行识别。测试结果证明OCR对RFID有明显的补充作用，两者结合后的盘点率可达98%以上，具体结果如表1所示。

表1 常规书架的盘点结果汇总 单位:%

	单RFID	单OCR	RFID+OCR	OCR补充率
第5面第一列	93.644	82.203	99.576	93.333
第6面第一列	90.625	86.806	99.306	92.592
第7面第一列	84.414	77.973	95.154	71.428
第8面第一列	97.196	76.636	99.065	66.667
总和	91.295	81.347	98.342	81.928

第二阶段测试主要针对特定书架和书籍区域进行分段测试。上海图书馆的工作人员挑选了三类具有明显特征的书架，分别为：被大柱子隔断的书架(如图6所示)、有凸出书籍的书架以及有密集图书和横放图书的书架。

3.1 非常规场景的测试——被大柱子隔断的书架

针对该类场景，盘点机器人仍采用RFID+OCR的盘点方式，但在导航方式上运用了SLAM自适应算法，遇到柱子这类障碍，盘点机器人会自动后退，然后向外移动一定距离，并不断进行尝试，直到机器人能通过该障碍物。在通过障碍物之后，盘点机器人会以书架另一端为起点，继续进行剩余部分的盘点，以此保证最高的导航精度和盘点率，盘点机器人运动轨迹如图7所示。

图6 被柱子隔断的书架

图7 盘点机器人运动轨迹

机器人通过对导航算法的优化，调整自身和书架的间距，从而获得最优盘点效果。图8盘点的数据结果表明，盘点机器人不仅可以完全适应书架被柱子等凸出障碍物隔断的结构环境，采用RFID和OCR进行盘点还可以实现高达98%以上的盘点率。

图8 盘点结果汇总

3.2 非常规场景的测试——有凸出书籍的书架

在盘点进程中遇到大型的凸出书籍(如图9

所示），盘点机器人会采用和上述大柱子场景类似的导航逻辑。机器人侧面是视野 360° 的激光雷达，可以探测平行于天线平面的障碍物。如果探测到书籍进入危险距离，机器人会停下来，然后沿书架后退 30 cm 再前进 30 cm，期间会让机器人与书架间的距离增加 3 cm（单次移动间距可根据实际情况设置），如果不再遇到障碍，保持该距离直到本架结束，到下一架会再调整回原本距离，继续盘点。如果不能过去，再以此法增加 3 cm 距离。如果反复 6 次，或进入对侧书架的危险范围也无法成功，则判定失败，停止盘点后面的书架，升高天线，执行前面书架的上半部分盘点。其运动原理如图 10 所示。

图 9　有凸出书籍的书架

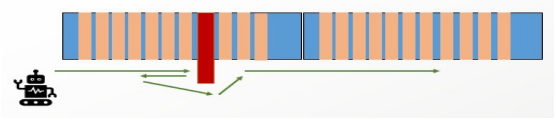


图 10　机器人运动原理

根据图 11 数据结果显示。

图 11　针对凸出书籍的智能盘点结果汇总

针对凸出书籍的盘点，盘点机器人经过导航算法优化，可以尽可能地将该因素对盘点效果的影响控制在极小范围内，盘点率可达 97%～98%，和正常书架的盘点率相对接近。

3.3　非常规场景的测试——有密集、横放图书的书架

图 12　有密集、横放图书的书架

由于该部分书籍厚度≤5 mm，且未按常规方向堆叠放置，无法使用 OCR 识别到索书号标签，故盘点机器人仅采用 RFID 进行盘点。针对该场景除了需要解决超薄书籍间的信号干扰问题，还需要克服横放图书凸出书架的问题，以及适应书架层高变化、不在同一直线等难点。在导航算法优化下，盘点机器人对该列密集图书进行了多次测试，测试结果如图 13 所示。

图 13　针对有密集、横放图书书架的智能盘点结果汇总

可以看到，密集图书由于书籍太薄且摆放不规则，图书间信号干扰较强，无法使用 OCR 等其它辅助盘点方式，故盘点率略低于正常书架的盘点，但仍可达到 87% 以上。

4 盘点机器人在未来使用中可能会遇到的问题

通过以上现场测试，我们对此类智能定位盘点设备的效能有了直观的了解。然而，新型设备或者说创新模式在应用过程中必然伴随着有待磨合之处，结合国内其他图书馆较早使用盘点机器人的经验，和在上海图书馆的现场测试，我们发现该智能设备在未来长期使用中可能会遇到以下这些问题。

4.1 SLAM算法精度降低

盘点机器人的定位与导航依赖于SLAM算法，该算法在计算过程中需要将观测数据、预测数据和历史数据合成一个高维向量并进行概率估计。随着时间的推移，数据结果会产生累积误差，长时间运行后定位精度可能会有偏差。而图书盘点和查找最重要的就是定位精度，精度偏差容易导致图书位置、排序和数量识别错误，从而降低了盘点的准确率[5]。更有甚者，定位精度降低严重的话可能会造成机器人和狭窄的书架之间发生碰擦。

4.2 少儿绘本盘点难度大

一般来说，少儿绘本书脊薄、规格多，这些图书特征给智能盘点机器人的盘点工作带来了不小的困难。首先，书脊薄会造成书架上少儿图书的密集堆放，从而干扰到盘点机器人对每本图书RFID信号的读取，而辅助盘点的OCR视觉识别功能也同样受限于较薄的书脊，无法完成有效的识别。其次，规格多表现在少儿绘本往往大小不一，开本种类从8开到64开都很常见，甚至还有异型开本。不规律的书籍大小同样给盘点机器人的RFID信号识别造成困难，因为在盘点时，机器人与书架之间的距离是固定的，遇到较小开面的图书会拉长芯片读取的距离，可能导致信号较弱从而识别不到；较大开面的图书则会对机器人的行径路线产生阻碍。前期测试的数据证实，盘点机器人对少儿图书的盘点准确率为80%，对于少儿绘本的盘点依然处于技术瓶颈期。

4.3 对网络连接要求较高

盘点机器人对网络的依赖度较高，在盘点完成后，机器人需将数据上传至后台的服务器输出定制化报表，同时要将相关结果同步到folio等索书平台。万一在数据传送阶段遇到无网络连接或网络较慢等意外状况，我们要求盘点机器人能做到把数据先行缓存，待网络良好再进行上传。此外，将来在上图东馆阅读广场的书架上会采用墨水屏层架标，如果能实现智能盘点机器人和与其系统对接，那么当层架标变化时，盘点数据库后台也会实时更新，保证盘点数据的准确性。这也对未来场馆的网络稳定性提出了更高的要求。

5 总结与展望

近年来，随着智慧图书馆理念的提出，代表着智能化、自动化、高效化的智能定位盘点模式将被更广泛地运用到图书馆的日常运营中去。盘点机器人是当下这一创新模式最先进的技术之一，其历经数次迭代升级，在系统功能、运行性能、结构设计等方面都日臻完善。我们希望，通过上海图书馆在这一领域的实践与测试，能够带动更多新时代下的传统图书馆在常态化图书盘点业务上的改进与突破。

参考文献：

[1] 樊慧丽，邵波. 智能机器人图书盘点创新实践与思考：以南京大学图书馆为例[J]. 图书馆，2018(9)：96-100.

[2] 樊慧丽，邵波. 国内外图书馆机器人的研究应用现状与思考[J]. 图书馆杂志，2017(6)：88-93.

[3] CHING S H，TAI A. HF RFID versus UHF RFID：technology for library service transformation at City University of Hong Kong[J]. The Journal of Academic librarianship，2009，35(4)：347-359.

[4] 沈奎林，邵波，陈力军. 情景共融环境下图书馆盘点机器人的发展研究[J]. 图书馆研究，2020，50(5)：98-102.

[5] 沈奎林，邵波，陈力军. 图书馆智能盘点机器人定位研究[J]. 图书情报导刊，2020，5(7)：34-40.

Research on intelligent positioning inventory robot for libraries —take Shanghai library as an example

Wu Jiayuan

（Institute of Scientific and Technical Information of Shanghai, Shanghai Library, Shanghai 200031）

Abstract: The essay introduces the research and application of intelligent inventory robots in recent years by Shanghai Library. It reviews the development of book inventory and summaries the excellent independent inventions in China. Through the field test on one type of the intelligent positioning inventory robots in Shanghai Library, the essay discusses the enhancement and improvement of book inventory business by new technologies and modes, and gives further thoughts and suggestions at the end.

Keywords: smart library; RFID; inventory robot

（上接第8页）

Design and implementation of household electric energy management system based on time-of-use power price and multi-objective optimization

Zhe Jianan*, Wu Ying, Li Jiao

（School of Computing, Xi'an Shiyou University, Xi'an 710000）

Abstract: Nowadays, the prevalence of TOU electricity price has changed the original mode of electricity consumption, not only to ensure the reduction of electricity costs, but also to ensure the comfort of users. Therefore, the home energy management system based on TOU price and multi-objective optimization is proposed. First, the commonly used household equipment is classified according to the electricity characteristics of the household equipment. Then, the establishment of the optimal scheduling model of the home electric power management system aims at the optimization of the household electricity cost, the improvement of user comfort and the tracking of carbon emissions. Finally, the model is optimized using a modified particle swarm algorithm. Through simulation experiments using matlab, the proposed method has significant optimization effects in reducing electricity costs, improving user comfort, and tracking carbon emissions.

Keywords: home energy management system; time-of-use price; multi-objective optimization; improved particle swarm optimization algorithm; optimal operation

文章编号: 1007-1423(2023)01-0016-06 DOI: 10.3969/j.issn.1007-1423.2023.01.003

机器学习预测化合物的雌雄激素受体活性

胡 帅，孔 韧*，谢良旭

（江苏理工学院电气信息工程学院生物信息与医药研究工程研究所，常州 213001）

摘要：预测一种化合物的雌雄激素受体的活性，对于避免暴露于环境中类似雌雄激素的化学物质是非常重要的。采用包括支持向量机、随机森林等多种机器学习方法，利用Binding Database数据库建立了预测活性的定量结构-活性关系(quantitative structure-activity relationship，QSAR)模型，并对模型进行验证与评估。评估结果发现，随机森林结合扩展连通性分子指纹对数据集预测准确率为0.83，其受试者工作特征曲线(receiver operating characteristic curve, ROC)得到的数据集曲线下面积(area under curve, AUC)为0.892，表明该模型具有广泛且良好的预测能力。该研究建立的活性预测模型可用于化合物的活性预测，为内分泌化合物的活性评估和风险管理提供参考。

关键词：雌雄激素受体；QSAR；活性预测；机器学习

0 引言

许多天然或者非天然的化合物被广泛应用于人类的日常生活，如防腐剂、紫外线过滤器、增塑剂、香料、抗菌剂、杀虫剂和阻燃剂等个人护理和消费品，食品都有化合物的存在[1-2]。值得注意的是，许多化合物作为内分泌干扰物(endocrine disrupting chemicals，EDC)，可过干扰人类激素的合成和作用，导致多种疾病，如生殖能力下降，癌症，甚至幼体死亡[3]。根据内分泌学会的定义，内分泌干扰化学物质是“一种外源性的非自然的化学物质，或化学物质的混合物，它能干扰激素作用的任何方面”[4]。目前国内外采取了许多方法来确定一种化合物是否具有内分泌活性。内分泌干扰者筛查计划和21世纪毒理学测试项目建立了各种体外或体内检测方法，以测量化学物质对人类或野生动物内分泌系统的潜在影响[5]。尽管细胞培养，特别是干细胞培养可以替代动物个体进行实验，从而缩短评估过程、提高实验的灵敏度，化合物毒性的实验评估仍然耗时费力[6]。使用数学模型来预测化合物的活性已经成为计算机辅助药物设计的研究热点。数据建模工具在一定的框架下对现有的实验数据进行扩展，可以减少人力物力消耗。

定量构效关系(quantitative structure-activity relationship，QSAR)被广泛应用于化合物对内分泌受体如雌雄激素受体的活性预测，如协同雌激素受体活性预测项目和雄激素受体活性协同建模项目构建就是通过不同QSAR方法训练的雌激素或雄激素受体活性预测[7]。该方法经常结合机器学习算法一起使用[8]。在综合国内外利用机器学习预测化合物的雌雄激素受体活性的基础上，本文采用支持向量机，随机森林等方法，采集来自Binding Database数据库的已知活

收稿日期：2022-06-09 **修稿日期：**2022-12-31

基金项目：国家自然科学基金面上项目(22003020,81603152,81903661)

作者简介：胡帅(1995—)，女，安徽宿州人，硕士，主要研究方向为机器学习、计算机辅助化合物筛选；*通信作者：孔韧(1981—)，女，河南岳阳人，博士，研究员，主要研究方向机器学习、计算机辅助化合物筛选；谢良旭(1986—)，男，山东济南人，博士，副研究员，主要研究方向生物信息学、生物材料的计算机仿真、基于机器/深度学习的智能药物设计

性数据集，建立了定量结构–活性关系模型用于雌雄激素受体的活性预测，有利于实现化合物的高通量筛选。

1　材料来源

Binding Database（BindingDB）（https://www.bindingdb.org/bind/index.jsp）是一个公开访问的数据库[9]，目前包含8185种蛋白质和超过920703个类药分子的2096653个活性数据。这些数据从科学文献中提取，数据收集的重点是作为药物靶点或候选药物靶点的蛋白质，其结构数据存在于蛋白质数据库中，数据的类型包括Ki、IC_{50}、Kd、EC_{50}等[10]。由BindingDB查询生成的数据集可以以带注释的mysql数据库的形式下载，以便进行进一步分析。研究发现，内分泌干扰物中的雌雄激素受体类似物可以跟雌雄激素受体结合，激活或抑制这些激素受体，从而干扰内分泌系统，影响人体健康[11]。BindingDB全面记录了雌雄激素受体类似物的数据情况[12]。本文所用的数据为从该数据库下载的有实验数据记录影响雌雄激素类似化合物，共计13190条数据，其中毒性评估准则以50%生长抑制浓度的对数log（IC_{50}）单位nmol/L表示，经过去除重复数据清除IC_{50}值相差大的错误数据后得到了8357条数据。为了进行量化分析数据，利用pandas画出了8357条数据的IC_{50}值的大致分布，如图1所示。

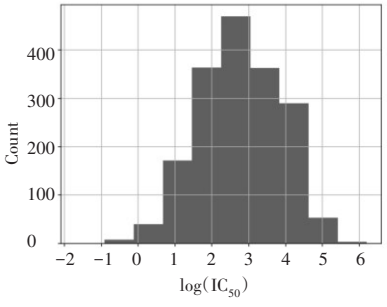


图1　数据分布

从图1可以看出数据集里的数据基本符合正态分布，所以将数据集里的数据以中位数log2.28为分界点，分为两部分。小于中位数为雌雄激素活性强的数据，或称活性数据，在模型算法中用class1表示；大于中位数为雌雄激素活性弱的数据，或称非活性数据，在模型算法中用class0表示。

2　数据处理

2.1　分子指纹

在比较两个化合物之间的相似性时遇到的最重要问题之一是任务的复杂性，这取决于分子表征的复杂性。在一定程度上的简化或抽象可以使比较分子更加容易。分子指纹就是一种抽象化表征，它的流程一般为归纳提取分子的结构特征，再用哈希(Hashing)生成比特向量。本文采用的基于子结构的分子指纹为MACCS和ECFP。

MACCS密钥通过使用RDKit进行计算[13]。ECFP指纹由Pande开发的DeepChem开源软件包进行转换。然后利用交叉验证方法[14]将数据集分为训练集和测试集，二者的比例为8∶2。

2.2　方法

2.2.1　随机森林

随机森林(random forests，RF)是一种比较典型的机器学习模型。经典的机器学习模型神经网络虽预测准确，但是计算量很大[15]。为了解决这个问题，上个世纪分类树的算法被Breiman等科学家提出。2001年Breiman把分类树组合成随机森林[16]，即通过对数据集的采样生成多个不同的数据集，并在每一个数据集上训练出一颗分类树，最终结合每一颗分类树的预测结果作为随机森林的预测结果。随机森林，被誉为当前最好的分类算法之一[17]。因为本文数据比较离散，所以选择了ID3(iterative dichotomiser 3)算法。下图为使用ID3算法构建随机森林的流程。

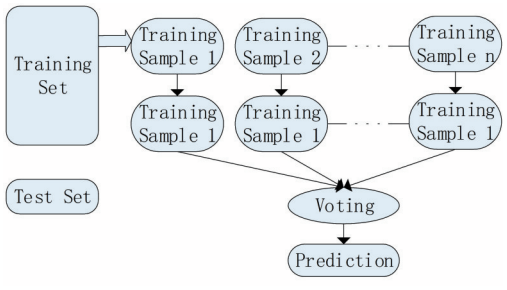


图2　RF模型的流程图

2.2.2 支持向量机

支持向量机(support vector machines，SVM)是一种构建分类器的强大方法[18]。它的目的是在两个类之间创建一个决策边界，这个边界叫做超平面[19]。这个超平面是从每个类的最近的数据点获取。SVM算法最初是由Vapnik在1963年提出的线性分类器，针对非线性数据，它还有一个用途就是核函数。在非线性问题中，可以使用核函数向原始数据添加额外的维数，使用核函数求解的过程如下：

第一步：将问题转化为原始问题及变形，用到的公式为

$$\min_{w,b,\xi} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\xi_i$$

$$\text{s.t. } y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i + b) \geqslant 1 - \xi_i, i = 1, 2, \cdots, N$$

$$\xi_i \geqslant 0, i = 1, 2, \cdots, N$$

第二步：原始问题对偶化，首先构建拉格朗日函数：

$$L(\boldsymbol{w}, b, \xi, \sigma, \mu) = \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}a_i\left(y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i + b) - 1 + \xi_i\right) - \sum_{i=1}^{N}\mu_i\xi_i$$

其中，$a_i \geqslant 0$, $\mu_i \geqslant 0$。

然后原始问题对偶化并选择核函数：

$$\min_{a} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}a_i a_j y_i y_j \mathrm{K}(\boldsymbol{x}_i\cdot\boldsymbol{x}_j) - \sum_{i=1}^{N}a_i$$

$$\text{s.t. } \sum_{i=1}^{N}a_i y_i = 0;\ 0 \leqslant a_i \leqslant C,\ i = 1, 2, \cdots, N$$

第三步：利用KKT条件求解，首先利用SMO算法求解$\boldsymbol{a}^*$，然后在向量$\boldsymbol{a}^*$中选一个$0<a_j<C$，求解$\boldsymbol{b}^*$。

$$\boldsymbol{b}^* = y_j - \sum_{i=1}^{N}\boldsymbol{a}_i^* y_i \mathrm{K}(\boldsymbol{x}_i\cdot\boldsymbol{x}_j)$$

最后得到分类决策算法：

$$f(x) = \mathrm{sign}\left(\sum_{i=1}^{N}\boldsymbol{a}_i^* y_i \mathrm{K}(x_i\cdot x_j) + \boldsymbol{b}^*\right)$$

3 结果与讨论

3.1 评估标准

在本研究中，采用准确性(Accuracy)、精准率(Precision)、召回率(Recall)、精度和召回率的调和平均(*F*1_score)、对每个类别的精准、召回和*F*1_score加和求平均，即宏平均(Macro_{avg})，对宏平均的一种改进，考虑了每个类别样本数量在总样本中占比，即加权平均($\mathrm{Weighted}_{avg}$)这六种统计指标来评估所提模型的性能，计算公式如下：

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (1)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$F1\_score = 2*Precision*\frac{Recall}{Precision + Recall} \quad (4)$$

$$Macro_{avg} = \frac{P_{no} + P_{yes}}{2} \quad (5)$$

$$Weighted_{avg} = P_{no}*\left(\frac{Support_{no}}{Support_{all}}\right) + P_{yes}*\left(\frac{Support_{yes}}{Support_{all}}\right) \quad (6)$$

对公式中出现的英文简写解释如下：

TP(True Positive)：做出Positive的判定，而且判定是正确的，即检测为雌雄激素活性强且雌雄激素活性强的物质；*FP*(False Positive)：做出Positive的判定，但判定是错误的，即检测为雌雄激素活性强但实际雌雄激素活性弱的物质；*TN*(True Negative)：正确的Negative判定，而且判定是正确的，即检测为雌雄激素活性弱且雌雄激素活性弱的物质；*FN*(False Negative)：错误的Negative判定，而且判定是错误的，即检测为雌雄激素活性弱但雌雄激素活性强的物质；雌雄激素活性弱的精准率用P_{no}来表示，雌雄激素活性强的精准率用P_{yes}表示；*Support*代表支持样本数，雌雄激素活性弱样本数用$Support_{no}$表示，雌雄激素活性强的样本数用$Support_{yes}$表示。

此外还应用了受试者工作特征曲线(receiver operating characteristic，ROC)和曲线下面积(area under curve，AUC)来评价模型的分类性能：ROC曲线的*Y*轴为真阳性率(true positive rate)，*X*轴为假阳性率(false positive rate)；*AUC*的值越大表明模型的预测结果越好[20]。

3.2 预测结果和分析

本文采用了二组分类模型RF和SVM，分别结合两种分子指纹做预测，在使用Sklearn.metric.classification_report工具对模型的测试结果进行

评价时，整理输出结果，如表1～表4(图3)所示。

表 1　随机森林(RF)结合ECFP指纹

类型	精准率	召回率	精度和召回率的调和平均	样本数
Class0	0.83	0.79	0.83	859
Class1	0.79	0.82	0.83	809
Accuracy	—	—	0.83	1668
$Macro_{avg}$	0.83	0.83	0.83	1668
$Weighted_{avg}$	0.83	0.83	0.83	1668

表 2　随机森林(RF)结合MACCS指纹

类型	精准率	召回率	精度和召回率的调和平均	样本数
Class0	0.81	0.81	0.81	859
Class1	0.80	0.80	0.80	809
Accuracy	—	—	0.80	1668
$Macro_{avg}$	0.80	0.80	0.80	1668
$Weighted_{avg}$	0.80	0.80	0.80	1668

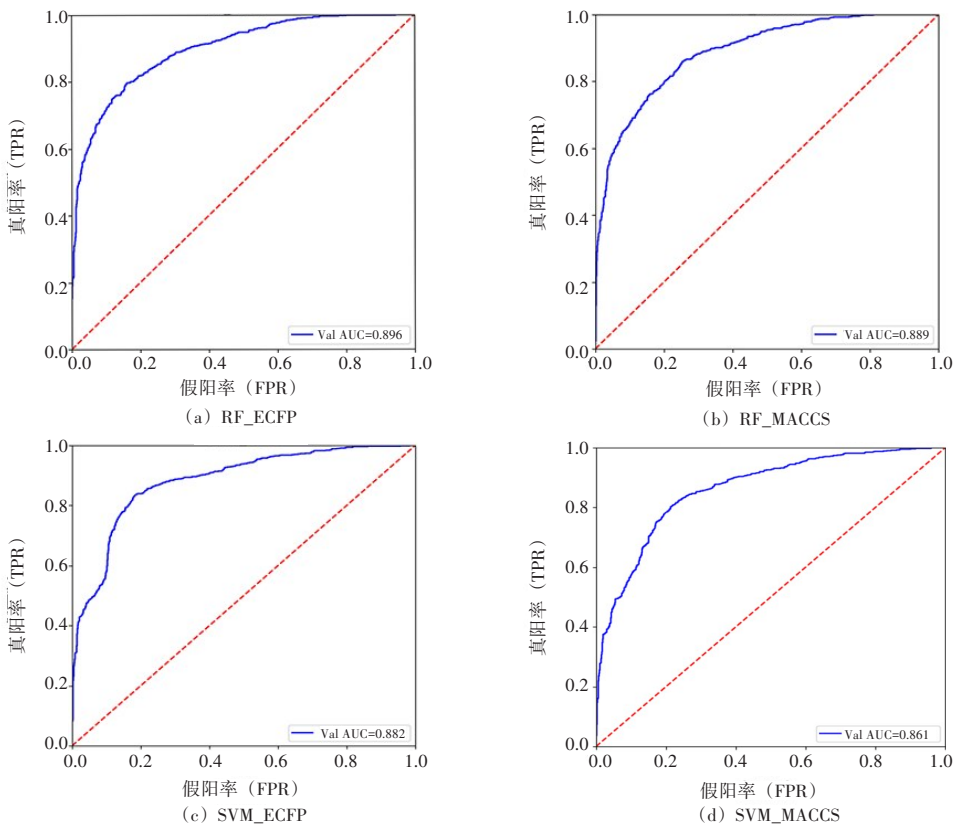
表 3　支持向量机(SVM)结合ECFP指纹

类型	精准率	召回率	精度和召回率的调和平均	样本数
Class0	0.82	0.83	0.83	859
Class1	0.82	0.81	0.81	809
Accuracy	—	—	0.82	1668
$Macro_{avg}$	0.82	0.82	0.82	1668
$Weighted_{avg}$	0.82	0.82	0.82	1668

表 4　支持向量机(SVM)结合MACCS指纹

类型	精准率	召回率	精度和召回率的调和平均	样本数
Class0	0.79	0.83	0.81	859
Class1	0.81	0.77	0.79	809
Accuracy	—	—	0.80	1668
$Macro_{avg}$	0.80	0.80	0.80	1668
$Weighted_{avg}$	0.80	0.80	0.80	1668

从表1～表4可以看出，RF，SVM算法结合ECFP指纹模型准确率和AUC值要分别高于RF，SVM结合MACCS，主要是因为ECFP可用来表



（注：AUC表示曲线下面积）

图 3　集受试者工作特征(receiver operating characteristic, ROC)曲线

示功能基团是否存在，对于分析分子活性至关重要。RF的结果要好于其他结果，从直观现象来解释，主要是因为RF的每棵决策树都相当于一个分类器，对于输入的每一个样本来说，*M*个分类器就有*M*个分类结果。RF集成了所有结果，最终把类别中的结果做为输出成果，RF是一种最容易达到好的分类结果的算法。RF结合ECFP指纹在对化合物的雌雄激素活性的预测上可以达到0.83的准确率，*AUC*值可以达到0.896，且ROC曲线光滑，说明数据样本空间分布合理，模型预测结果优良。其次，SVM结合ECFP结果稍逊于RF，从侧面证明了SVM适用于样本空间小的数据。SVM结合ECFP，由于样本小，ROC曲线略有凹凸。在对化合物的雌雄激素活性的预测上可以达到0.83的准确率，*AUC*值可以达到0.882，证明模型的预测能力良好且广泛。

4 结语

本文对8357条雌雄激素受体毒性值有影响的小分子化合物进行分子指纹的转换，选择学习方法建模并进行了毒性的预测评估，RF模型结合ECFP指纹明显优于参比模型。本文构建的雌雄激素受体预测模型，因纯粹选用理论计算参数而不依赖实验测定，故较方便、省时且节约费用，可对环境分布的化合物进行快速的激素受体干扰能力判断，为环境内分泌干扰分子的风险管理提供基础。

本文所使用的数据和源码可访问https://gitee.com/code-mk/classification获取。

参考文献：

[1] TIJANI J O, FATOBA O O, BABAJIDE O O, et al. Pharmaceuticals, endocrine disruptors, personal care products, nanomaterials and perfluorinated pollutants: a review[J]. Environmental Chemistry Letters, 2016, 14(1): 27-49.

[2] PARK S, CHUNG C. Effects of an action-guided intervention on optimistic bias and protective behaviors regarding endocrine disruptors in young women [J]. Women & health, 2022, 62(3): 234-244.

[3] 王浩安. 药物临床前安全性研究：组织交叉反应[C]//中国毒理学会第九次全国毒理学大会论文集, 2019: 281.

[4] 李金荣, 郭瑞昕, 刘艳华, 等. 五种典型环境内分泌干扰物赋存及风险评估的研究进展[J]. 环境化学, 2020, 39(10): 2637-2653.

[5] JUDSON R S, KAVLOCK R J, MARTIN M T, et al. Estimating toxicity-related biological pathway altering doses for high-throughput chemical risk assessment [J]. Chemical Research in Toxicology, 2011, 24(4): 451-162.

[6] BURELLO E, WORTH A. QSAR modeling of nanomaterials [J]. Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology, 2011, 3(3): 298-306.

[7] MANSOURI K, KLEINSTREUER N, ABDELAZIZ A M, et al. CoMPARA: collaborative modeling project for androgen receptor activity [J]. Environmental Health Perspectives, 2020, 128(2): 027002.

[8] DUCHOWICZ P R, BENNARDI D O, ORTIZ E V, et al. QSAR Models for the Fumigant Activity Prediction of Essential Oils [J]. Journal of Molecular Graphics and Modelling, 2020, 101: 107751.

[9] SUN L, MENG S. Structure-based model profiles affinity constant of drugs with hPEPT1 for rapid virtual screening of hPEPT1′s substrate[J]. SAR and QSAR in Environmental Research, 2016, 27(8): 1588.

[10] CHEN X, LIU M, GILSON M K. BindingDB: a web-accessible molecular recognition database [J]. Combinatorial Chemistry & High Throughput Screening, 2001, 4(8): 29-36.

[11] CHIGRINETS S V, BRYUHIN G V. Environmental exposure to endocrine disruptors on semen quality of men[J]. Andrology and Genital Surgery, 2018, 19 (3): 198.

[12] JUDSON R S, HOUCK K A, KAVLOCK R J, et al. In vitro screening of environmental chemicals for targeted testing prioritization: the ToxCast project[J]. Environmental health perspectives, 2010, 118 (4): 277-289.

[13] GAGKAYEVA Z V, GORSHUNOV B P, KACHESOV A Y, et al. Infrared fingerprints of water collective dynamics indicate proton transport in biological systems[J]. Physical Review, 2022, 105(4): 1-9.

[14] MAYNE G C, WOODS C M, HUSSEY D J, et al.

Cross validated serum small extracellular vesicle microRNAs for the detection of oropharyngeal squamous cell carcinoma [J]. Journal of Translational Medicine, 2020, 18(1): 1-6.

[15] MUEGGE I, MUKHERJEE P. An overview of molecular fingerprint similarity search in virtual screening [J]. Expert Opinion on Drug Discovery, 2016, 11(2): 148.

[16] ZOELLER R T, BROWN T R, DOAN L L, et al. Endocrine-disrupting chemicals and public health protection: a statement of principles from the endocrine society[J]. Endocrinology, 2012, 153(9): 4097-4110.

[17] ZHANG H, ZHANG P. Cross-voting SVM method for multiple vehicle classification in wireless sensor networks[J]. Sensors, 2018, 18(9): 3108.

[18] WU G, ZHENG R B, TIAN Y J, et al. Joint Ranking SVM and Binary Relevance with robust Low-rank learning for multi-label classification[J]. Neural Networks, 2020, 122(1): 24-39.

[19] XU S, CHEN C-C, WU Y, et al. Adaptive residual weighted K-nearest neighbor fingerprint positioning algorithm based on visible light communication [J]. Sensors, 2020, 20(16): 4432.

[20] 王艺霖, 范俊韬, 王书平, 等. 机器学习预测内分泌干扰物水生生物毒性效应[J]. 生态毒理学报, 2022, 17(2): 148-163.

Machine learning predicts androgynous receptor activity in molecule compounds

Hu Shuai, Kong Ren*, Xie Liangxu

(Institute of Bioinformatics and Medical Engineering(School of Electrical and Information Engineering, Jiangsu University of Technology), Changzhou 213001)

Abstract: Predicting estrogen and androgen receptor activity of a small molecule compound is important to avoid exposure to environmental androgynous chemicals. In this paper, a Quantitative Structure-Activity Relationship (QSAR) model for predicting Activity was established by using Binding Database, including support vector machine, random forest and other machine learning methods. The model is verified and evaluated. The evaluation results show that the prediction accuracy of random forest combined with extended connectivity molecular fingerprint on the dataset is 0.83. The Area Under curve (AUC) of the dataset obtained from the Receiver Operating Characteristic curve (ROC) was 0.892, indicating that the model had good and extensive predictive ability. The activity prediction model of androgynous receptor established in this study can be used to predict the activity of compounds, in order to provide a reference for the activity assessment and risk management of endocrine compounds.

Keywords: estrogen and androgen receptors; QSAR; activity prediction; machine learning

文章编号: 1007-1423(2023)01-0022-06 DOI: 10.3969/j.issn.1007-1423.2023.01.004

基于Transformer的船舶故障预测研究

黄佳汲

（上海船舶运输科学研究所有限公司，上海 200135）

摘要：为了解决循环神经网络不能并行训练导致的在时间性能上的缺陷，对Transformer架构应用于船舶运行的趋势预测进行研究，并围绕Transformer设计一套更高效的故障预测方法。在数据预处理上，设计了一种滑动窗口的方法筛选平稳数据，取得了良好的结果；在故障诊断上，使用多项式回归，通过对历史数据的训练，在测试集上获得了较高评分；在趋势预测上，使用Transformer架构，经过训练可以做出较为精准的预测，并在与GRU网络的比较中体现出了显著的性能优势。综合来看，所设计的方法有一定实践意义。

关键词：故障诊断；故障预测；机器学习；Transformer

0 引言

随着大数据、物联网、云计算等科学技术能力的迅猛提升，智能船舶已成为海事界的重点研究对象，是全球航运界的未来发展趋势。作为世界上的航运大国，我国对于船舶智能化的发展有着极大的重视。2015年，中国船级社发布了《智能船舶规范2015》；2020年，中国船级社在更新的《智能船舶规范2020》中增加了远程控制船舶和自主操作船舶在设备配备、状态监测与健康管理、设计原则与性能等方面的要求；工业与信息化部也于2017年成立了智能船舶专项，并联合多部委编制相应的智能船舶发展的行动计划[1]。

智能机舱是智能船舶的功能之一，依据《智能船舶规范2020》，智能机舱应具备基于状态监测数据对设备与系统的运行状态、健康状况进行分析和评估的能力[2]。智能设备故障诊断与健康管理是基于状态维修的升级发展，聚焦于复杂工程设备的健康状态监测和管理，可通过经验知识完成对监测设备的运行状态(包括设备结构、故障失效机理、运动学原理、设计制造、安装运行和后期维修等)识别预测，从而判断运行机制，分析故障原因，确定故障程度，达到降低成本、提高效率、使设备安全可靠运行的目的。智能船舶作为航运新领域，目前正处在发展初期，还存在设备故障预测能力较弱、数字化管理水平欠缺等不足[3]。

在现有的研究中，文献[4]提出了一种以麻雀搜索算法优化SVM作为故障分类器，以LSTM作为趋势预测器的故障预测方法；文献[5]提出了一种基于CNN-GRU组合神经网络的变压器短期故障预测方法，并在与SVM、LSTM、GRU、CNN-LSTM的对比中获得了更好的预测精度。目前在时间序列的问题上，普遍使用循环神经网络作为预测模型。

循环神经网络固有的顺序属性，使得训练样本间无法进行并行训练，因此这一类模型在时间性能上受到了限制。而Transformer架构的提出，很好地解决了这一问题，通过多头注意力机制的使用，突破了循环神经网络不能并行计算的限制。本文的目标是尝试将主要用于自然语言处理的Transformer架构应用在趋势预测上，并结合基于机器学习的故障诊断，提出一套比循环神经网络更高效的故障预测方法。

收稿日期：2022-08-26 **修稿日期：**2022-09-06

作者简介：黄佳汲(1994—)，男，上海人，助理工程师，管理学学士，研究方向为智能船舶前后端开发及数据分析

1　整体框架

本文的数据来自某液体散货轮最近一年的实船数据，在建立故障诊断基准值时所使用的为秒级数据，在使用Transformer进行趋势预测时所使用的为分钟级数据。本文数据分析所用的语言为Python，主要的第三方库为numpy、pandas、matplotlib、sklearn以及pytorch。

本文方法主要分为三步：首先是数据预处理，对历史数据进行筛选，选出平稳运行时间段的数据，减少异常值的影响；然后是通过历史数据对模型进行训练，使用多项式回归建立平稳运行状态下设备运行的基准值模型，使用Transformer建立设备运行趋势预测模型；最后通过预测数据与基准值的比较判断设备运行是否处于正常状态，总体框架如图1所示。

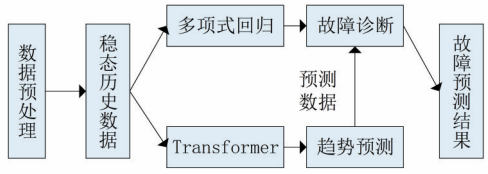


图1　总体框架

2　相关技术介介绍

2.1　基于滑动窗口的平稳数据筛选

由于本文的数据集体量较大，因此对于常见的缺失值问题可以直接采用删除该条记录的方法进行处理。数据预处理的重点工作是筛选运行状态平稳的数据区间，实船在航行的过程中，运行状态并非总是保持较为平稳的状态(如图2所示)，而在非平稳状态下记录的数据会存在大量异常点，影响模型的训练，因此需要对数据进行筛选。

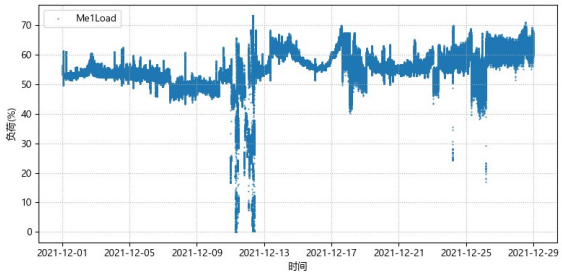


图2　负荷概览

本研究通过实践探索了一种滑动窗口的方法，通过对按时间排列的历史数据分段拟合并求导，对导函数依据其输入范围按一定的间隔求解，并设定一个上界，若所有结果均小于上界，则认为该段数据为稳定工况；若任意一个求解结果超过了上界，则认为该段数据不是稳定工况，并将之前所有相邻的稳定工况分段连接，获得一个稳定工况时间段的起止时间，完成判断后将窗口前推，继续判断下一段数据(如图3所示)。

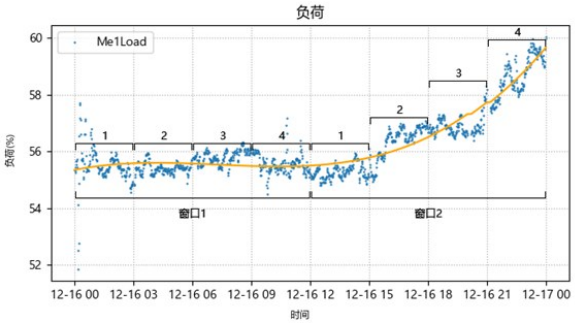


图3　滑动窗口演示

关键代码示例：

```
count = len(data)                                    # 数据长度
split = 100                                          # 窗口大小
index = 0                                            # 计数器
while index < count - split:
    tempX = np.arange(split)
    tempY = data[mpoint].iloc[index:index + split]
    tempP = np.polyfit(tempX, tempY, 3)
    tempDer = np.polyder(tempP)
    isStable = True
    for i in range(1, 20):
        tempVal = np.polyval(tempDer, split // 20 * i)
        if np.abs(tempVal) > 0.2:
            isStable = False
            breakPoint = index + split // 20 * i
            break
    if isStable:
        if lastIsStable == False or lastIsStable == None:
            tempStartTime = data["Time_1"].iloc[index]
            tempStartIndex = index
            tempEndTime = data["Time_1"].iloc[index +
              split]
        else:
            tempEndTime = data["Time_1"].iloc[index +
              split]
        lastIsStable = True
```

```
        index += split // 4          # 前推1/4个窗口的距离
    else:
        tempEndTime = data["Time_1"].iloc[breakPoint-
          split // 4]
        if tempStartTime != None and (breakPoint-split
            //4-tempStartIndex) > 1000:
            tempTimeTups.append((tempStartTime, temp
              EndTime))
            tempStartTime = None
            lastIsStable = False
            index = breakPoint + split // 4  # 从中断点前推
                                        #1/4个窗口的距离
```

2.2 基于皮尔逊相关系数的相关性分析

故障诊断基准值方法建立在船舶设备运行时的关联性上，如柴油机在运行时，负荷越高，往往转速越快，排气温度越高，使用时一般使用负荷作为输入。建立故障诊断基准值模型前首先要对测点进行相关性分析，相关性分析采用皮尔逊相关系数，该系数广泛用于度量两个变量之间的相关程度，其值介于-1与1之间。两个变量之间的皮尔逊相关系数定义为两个变量之间的协方差和标准差的商，如式(1)所示。

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \tag{1}$$

2.3 基于多项式回归的故障诊断

在确定了两个测点之间的相关性之后，需要通过机器学习的方法建立测点之间的回归模型，本文选择多项式回归作为故障诊断的基准值模型。

多项式回归的公式如式(2)所示，回归的目的是找到一组$w_0, w_1, \cdots, w_d$，使得拟合方程尽可能与实际样本数据相符合，一般使用最小二乘法实现梯度下降，本文选择最高次为3次的多项式回归。

$$f(x) = w_0 + w_1 x + \cdots + w_d x^d \tag{2}$$

2.4 基于Transformer的趋势预测

本文采用Transformer实现设备运行状态趋势预测，Transformer[6]是一种模型架构，它的提出改变了以往序列问题建模和循环神经网络之间划等号的思路。目前，Transformer已被广泛应用于自然语言处理、计算机视觉、音频处理等许多人工智能领域，并取得了巨大的成功。

文献[7]中BERT模型取得了NLP领域11项任务的最佳成绩，而在BERT中发挥重要作用的结构就是Transformer，之后相继出现的XLNET、roBERT等模型击败了BERT，但是这些模型的核心没有变，仍然是Transformer。

相比之前占领市场的LSTM和GRU模型，Transformer有两个显著的优势：

(1) Transformer能够利用分布式GPU进行并行训练，提升模型训练效率。

(2) 在分析预测更长的文本时，捕捉间隔较长的语义关联效果更好。

Transformer模型采用的是编码器-解码器架构，在该模型中，编码器和解码器不再使用循环神经网络，而是使用连续N个具有相同结构的编码器和解码器所连接的网络结构。Transformer总体架构图如图4所示。

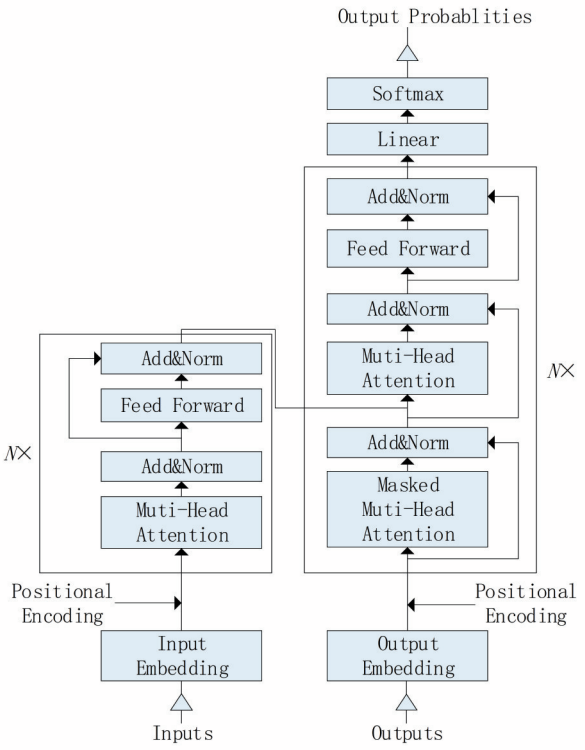


图4 Transformer总体架构图

每一个编码器中又包含两个串联的子层连接结构，其中一个为多头注意力层，另一个为前馈全连接层。多头注意力层的作用是分割输出的张量，分别进行注意力计算，前馈全连接层则是考虑注意力机制可能对复杂过程的拟合程度不够，通过增加两层网络来增强模型的能力，每个

子网络都具有残差连接。在每个解码器中，除了包含与解码器类似的多头注意力层和前馈全连接层外，还在两个子网络之间添加了另外一个注意力模块。与编码器类似，解码器中的三个子网络也均具有残差连接，并且在每个残差合成后都进行归一化操作[8]。

3　故障预测方法验证

3.1　数据预处理

首先使用前文提到的滑动窗口方法对时间段进行筛选，经过检验，这一方法可以很好地筛选出平稳时间段的数据，见表1。

表1　平稳运行时间段

	起始		结束	
0	2021-12-23	04：45：19	2021-12-23	08：13：24
1	2021-12-23	08：43：49	2021-12-23	09：08：03
2	2021-12-23	09：38：08	2021-12-23	10：19：05
3	2021-12-23	10：32：25	2021-12-23	13：48：01
4	2021-12-23	13：56：21	2021-12-23	16：50：36
5	2021-12-23	17：03：57	2021-12-23	18：52：17
6	2021-12-23	19：00：49	2021-12-23	19：43：42
7	2021-12-23	19：52：06	2021-12-23	20：20：50
8	2021-12-23	20：29：11	2021-12-23	22：41：04

3.2　相关性分析

本文以主机负荷、主机转速、扫气压力、增压器转速四个测点为例，选取历史数据，使用pandas自带的corr函数计算相关性矩阵，并使用seaborn库提供的heatmap展示测点之间的相关性，结果如图5所示。

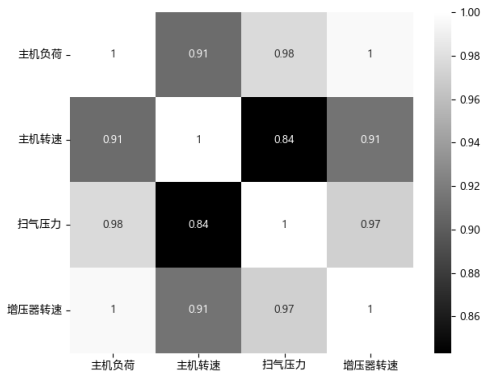


图5　相关性分析

3.3　故障诊断基准值建立

选取主机负荷与增压器转速这两个测点，将平稳运行时间段数据合并，在其中随机选取10000个作为测试集，其余作为训练集，使用多项式回归的方式建立以主机负荷作为输入，增压器转速作为输出的回归模型。通过对筛选出的数据进行训练，模型在测试集上的结果如图6所示。

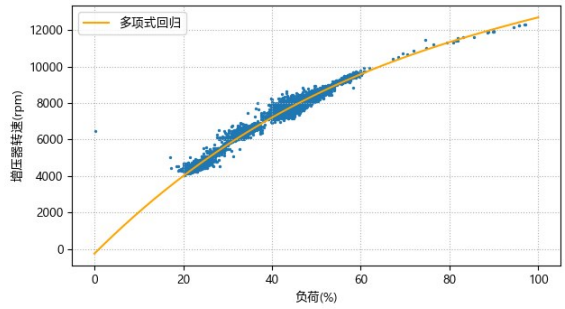


图6　主机负荷与增压器转速拟合

可以看出，主机负荷与增压器转速之间的关系较好地符合连续函数的特点，多项式回归的r^2得分在0.9以上，拟合程度较高。在模型的保存和使用上，多项式回归模型只需要存储一组系数即可做到随时取用，即使跨语言使用也非常便捷。

3.4　趋势预测

完成故障诊断基准值的建立后，需要再建立测点的趋势预测模型，预测模型采用Transformer架构来建立。本文使用主机转速作为样例，通过pytorch搭建Transformer模型，尝试通过前8个历史数据预测后1个数据。

首先创建一个Transformer架构的模型，考虑到本文研究的并不是机器翻译问题，而是一个短序列预测的问题，因此在构建模型时舍弃了文本嵌入层和位置编码器。同时为了验证此模型对于循环神经网络的性能优势，额外引入一个GRU模型作为对照。

代码示例：

```
class MyTF(nn.Module):
    def_init_(self, channels, nhead, dropout=0, num_
        layers=1):
        super(MyTF, self)._init_()
```

```
        encoder_layer = nn.TransformerEncoderLayer(
            d_model=channels,
            nhead=nhead,
            dropout=dropout,
            dim_feedforward=2 * channels,
        )
        decoder_layer = nn.TransformerDecoderLayer(
            d_model=channels,
            nhead=nhead,
            dropout=dropout,
            dim_feedforward=2 * channels,
        )
        self.encoder = torch.nn.TransformerEncoder(en-
          coder_layer, num_layers=num_layers)
        self.decoder = torch.nn.TransformerDecoder(de-
          coder_layer, num_layers=num_layers)
        self.fc = torch.nn.Linear(8, 1)
    def forward(self, x):
        return self.fc(self.decoder[x, self.encoder(x)])
class GRU(nn.Module):
    def__init__(self, input_size, hidden_size, output_size,
        num_layers=1):
        super(GRU, self).__init__()
        self.hidden_size = hidden_size
        self.num_layers = num_layers
                # 实例化预定义的nn.GRU, 它的三个参数
                #分别是input_size, hidden_size, num_layers
        self.gru = nn.GRU(input_size, hidden_size)
        self.linear = nn.Linear(hidden_size, output_size)
        self.tanh = nn.Tanh()
    def forward(self, input, hidden):
        input = torch.reshape(input, (1, 1, self.gru.input_
          size))
        rr, hn = self.gru(input, hidden)
        return self.linear(rr), hn
    def initHidden(self):
        return torch.zeros(self.num_layers, 1, self.hidden_
          size)
```

在参数选择上，设置Transformer模型的词嵌入维度为8，多头注意力层的头数为4，置零比率为0，编码器和解码器的层数都为1，并通过一个全连接层将输出张量的特征维度转换为1。设置GRU模型输入张量的特征维度为2，隐层张量特征维度为4，隐含层的数量为1，同样通过一个全连接层将输出张量的特征维度转换为1。

完成模型建立后，在实船数据上开始测试，随机取某一平稳运行时间段的31个连续数据，选择前30个作为训练集，从头开始每次取9个数据，前8个作为输入，最后一个为真值，反向传播后将起始位置前推一位，重复训练过程，当起始位置大于21后将起始位置重新设为0。选择均方误差作为损失函数，选择Adam优化器并设置学习率为0.01，设置训练退出条件为最近22次损失函数结果均值小于0.5。

代码示例：

```
data_all = pd.read_excel("data/data.xlsx")
rd_index = np.random.randint(0, data_all.shape[0]
  -31)
data = data_all.iloc[rd_index:rd_index + 31]
mytf = MyTF(8, 4, 0)
# mygru = GRU(2, 4, 1, 1)
criterion = nn.MSELoss()
optimizer_tf = optim.Adam(mytf.parameters(),
  lr=0.01)
# optimizer_gru = optim.Adam(mygru.parameters(),
  lr=0.01)
start = time.time()
split = 30000
optimizer = optimizer_tf
loss_arr = []
for i in range(10000):
    optimizer.zero_grad()
    index = i % 22
    input_values = torch.Tensor(data["Me1Nms"].
iloc[index:index + 8].values).view(1, 1, -1)
    output_true = torch.Tensor(data["Me1Nms"].iloc
[index + 8:index + 9].values).view(1, 1, -1)
    outputs = mytf(input_values)
    # hidden = mygru.initHidden().to(device)
    # for j in range(4):
    #   outputs, hidden = mygru(input_values[0, 0, j * 2:
        j * 2 + 2], hidden)
    loss = criterion(outputs[0][0], output_true)
    loss_arr.append(loss)
    if i >= 21:
        loss_arr.pop(0)
        temp_avg = torch.mean(torch.Tensor(loss_arr))
        print(temp_avg)
        if temp_avg <= 0.5:
```

```
            break
        loss.backward()
        optimizer.step()
    end = time.time()
```

根据测试结果，GRU模型和Transformer模型均可以使损失函数收敛在0.5以内，在时间消耗上，Transformer模型平均在1秒左右，而实现同样的均方误差GRU模型所需的时间平均在6秒以上，Transformer的性能优势非常显著。

完成训练后，以第23到第30个数据为输入，第31个数据作为真值，测试结果如图7所示。

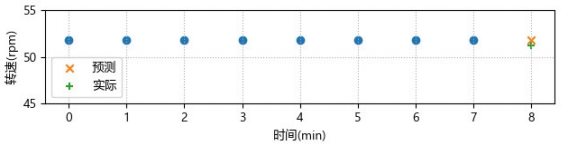


图7　预测效果

可以看到，数据在测试集上的预测结果和真值基本重合，预测效果较好。至此，故障预测的全部步骤验证完成。

4　结语

本文围绕Transformer建立一套相比循环神经网络更高效的故障预测方法。在数据预处理上，使用一种滑动窗口的方法筛选平稳数据，取得了良好的结果；在故障诊断上，使用多项式回归，通过对历史数据的训练，模型的r^2得分达到0.9以上，证明模型的准确度较高；在趋势预测上，将Transformer架构应用于船舶运行状态的趋势预测，并与GRU网络进行了比较，证明Transformer的时间性能相对循环神经网络有显著优势，并可以做出较为精准的预测。综上所述，本文提出的方法有一定的实践意义。本方法的主要缺陷在于缺少故障环境，由于故障的发生属于小概率事件，随机性很强，因此故障数据的收集较为困难，本文的测试数据为正常运行状态下的数据，缺少真实故障数据的验证，在未来的研究中，如何有效收集真实故障数据是需要关注的重点。

参考文献：

[1] 严新平,褚端峰,刘佳仑,等. 智能交通发展的现状、挑战与展望[J]. 交通运输研究. 2021,7(6):2-10,22.

[2] 中国船级社. 智能船舶规范[S]. 2020-03-01.

[3] 李震邦. 数字孪生技术与智能船舶发展[J]. 船舶工程,2022,44(S1):543-547.

[4] 陈万志,李昊哲,刘恒嘉,等. 融合LSTM和优化SVM的风力发电机组故障预测方法[J]. 辽宁工程技术大学学报(自然科学版),2022,41(4):379-384.

[5] 杨威,蒲彩霞,杨坤,等. 基于CNN-GRU组合神经网络的变压器短期故障预测方法[J]. 电力系统保护与控制,2022,50(6):107-116.

[6] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [EB/OL]. arXiv, 2017. https://doi.org/10.48550/arXiv.1706.03762.

[7] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[EB/OL]. arXiv,2018. https://doi.org/10.48550/arXiv.1810.04805.

[8] 洪季芳. Transformer研究现状综述[J]. 信息系统工程,2022(2):125-128.

A research about ship fault prediction based on transformer

Huang Jiaji

(Shanghai Ship and Shipping Research Institute Co., Ltd., Shanghai　200135)

Abstract: In order to solve the performance defects of recurrent neural networks caused by its being not able to be trained parallelly, trying to use Transformer in the trend prediction of ship, and designing a more efficient fault prediction method based on Transformer. In data preprocessing, a sliding window method is designed to screen stationary data, and good results are obtained. In fault diagnosis, polynomial regression is used, and through the training of historical data, high scores are obtained in the test set. In trend prediction, Transformer can make more accurate prediction after training, and it has significant performance advantages compared with GRU network. In summary, the method designed has certain practical significance.

Keywords: fault diagnosis; fault prediction; machine learning; transformer

文章编号: 1007-1423(2023)01-0028-06 DOI: 10.3969/j.issn.1007-1423.2023.01.005

一种结构光组合编码模式和解码算法

汪维刚*，王泽民，赵 选

（西安工业大学电子信息工程学院，西安 710021）

摘 要：为了满足高分辨率的动态三维测量，设计了一种结构光组合编解码模式。将伪随机阵列编码和单幅正弦条纹图案组合形成编码图案。单幅正弦条纹编码图案能够满足高分辨率的三维测量，通过伪随机阵列编码对单幅正弦条纹进行再编码，使得单幅正弦条纹图案能够适用于动态场景下的三维测量。为了确保相位值的准确性，设定编码参数，使得伪随机阵列子窗口中只包含一个完整周期的正弦条纹。组合图案解码算法通过伪随机阵列的窗口识别匹配，再进行窗口内的正弦条纹解码。实验结果表明了上述编解码模式的有效性。

关键词：结构光编码；伪随机阵列；窗口匹配；正弦条纹

0 引言

结构光法作为非接触式三维测量技术的一个重要分支，具有测量速度快、成本低、鲁棒性好等优势。按编码方式的不同，编码结构光法可分为直接编码、时间编码和空间编码，文献[1]中概括了时间编码和空间编码主要算法。

随着动态场景下三维测量需求不断增多，有关结构光空间编码的研究也越来越热门。文献[2]设计了一种自均衡伪随机序列，当被测量物体带有多种颜色特征时，无法避免物体自身的颜色对编码图案造成干扰而导致的识别错误。文献[3]提出一种伪随机序列结合相位信息的编码方法，但解码精度会受到编码宽度的影响。文献[4]利用主成分分析和K均值聚类方法设计了一种自适应颜色聚类方法，完成了对颜色强度高低的识别。文献[5]提出利用两种颜色编码的网格条纹模式，但该方法下的网格条纹检测和识别算法均较复杂。

伪随机阵列编码是由伪随机序列通过对角线折叠法或者拼片算法生成，文献[6]中产生大尺寸伪随机阵列的拼片算法较复杂，且码字利用率也较低，随着伪随机阵列尺寸的增大，编码符号也不断增加，增加了后续解码的难度。文献[7]将伪随机序列通过对角线折叠法生成伪随机阵列，但即使在三元域中使用九阶的本原多项式生成伪随机序列，采用对角线折叠法得到的伪随机阵列仍然无法满足高分辨率测量的要求。文献[8]给出了一种三级棋盘的非正式编码，但并没有生成全局唯一的编码图案。

本文设计了一种结构光组合编码模式，将单色正弦条纹和小尺寸伪随机阵列通过不同颜色通道相叠加，得到组合编码图案。单色正弦条纹能够满足高分辨率的测量，且保证了受到被测物体表面颜色干扰的影响更小。相较于大尺寸伪随机阵列，小尺寸的伪随机阵列更容易实现，并且小尺寸伪随机阵列所用码字较少，编码时所需的二值符号种类也更少，降低了后续伪随机阵列的解码难度。所设计的编码图案能够满足动态场景下的高分辨率测量需求。

收稿日期：2022-10-14 **修稿日期：**2022-12-02

作者简介：*通信作者：汪维刚(1996—)，男，甘肃天水人，硕士研究生，研究方向为非接触测量技术，E-mail: 18709207273@163.com；王泽民(1978—)，男，河北人，博士，教授，硕士生导师，研究方向为机器视觉检测技术、非接触测量技术；赵选(1997—)，男，安徽宿州人，硕士研究生，研究方向为非接触测量技术

1　结构光组合编码模式

1.1　基于伪随机序列测量方法分析

伪随机序列具有良好的窗口特性，窗口内的子序列在整个序列中具有全局唯一性。一维伪随机序列可由本原多项式生成，用移位寄存器作为伪随机码产生器，本原多项式及其所指定的反馈移位寄存器共同决定了伪随机序列输出结果。由二元五阶本原多项式所生成的伪随机序列，在大小为5的窗口内的子序列全局唯一。如图1所示。

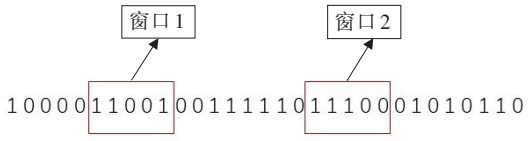


图1　一维伪随机序列示意图

伪随机阵列通常由一维伪随机序列通过折叠算法或者拼片算法生成，再选取基元符号进行编码，其生成的二维伪随机阵列同样具有窗口特性，即子窗口在整个二维阵列中的全局唯一性。伪随机阵列解码时，通过窗口识别进行特征点检测匹配。由于编码时基元符号需要以领域像素点来表示，多个像素点标识一个基元符号，导致测量分辨率较低。伪随机阵列基元窗口的大小直接影响到编码图案所包含的特征点数，窗口过小会导致解码算法检测基元符号特征点困难，窗口过大会导致空间分辨率低。以一幅大小包含50×50像素点的图像为例，假设编码基元符号大小为10×10像素，如图2所示。

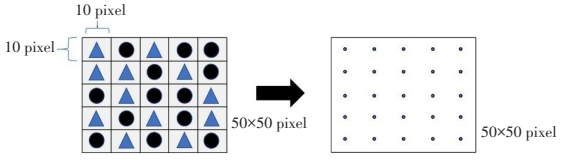


图2　伪随机阵列编码原理图

可以看出，包含50×50像素点的图像，采用以上伪随机阵列编码所得到的特征点仅为25个像素点，严重降低了测量分辨率。为了尽可能地提高测量分辨率，通过减少基元符号的编码像素个数或者采用更大的伪随机阵列模板来获得更多的特征点数。例如，使用更高阶的本原多项式生成长度更长的一维序列来获得更大尺寸的伪随机阵列，再通过多元基元符号进行编码，但即便如此，伪随机阵列编码仍然无法达到高分辨率的测量要求。

1.2　基于相位的测量方法分析

基于相位测量的主要原理是将生成的一系列相移条纹依次按序投射到被测物表面，由于物体高度的调制相位发生变化，通过相位的变化值可以得到被测物体高度信息。N步相移法是指正弦条纹图案的相位值在一个周期内等间距移动N次，即每次移动$\frac{2\pi}{N}$步长。其中第i幅条纹图像可表示为

$$I_i(x,y)=I_a(x,y)+I_b(x,y)\cos\left(\phi(x,y)+\Delta\varphi_i\right)\quad i=1,2,3,\cdots,N \tag{1}$$

式中：$I_a(x,y)$为环境背景光强，$I_b(x,y)$为条纹对比度，$\Delta\varphi_i$为相移步长。通过N幅条纹图案之间的相位差可计算得到包含被测物体高度信息的相位值

$$\phi_l(x,y)=-\arctan\left[\frac{\sum_{i=1}^{N}I_i(x,y)\sin\frac{2\pi i}{N}}{\sum_{i=1}^{N}I_i(x,y)\cos\frac{2\pi i}{N}}\right] \tag{2}$$

由于通过反正切函数得到的相位值是被包裹在$[-\pi,+\pi]$区间中，需要进行相位展开获取绝对相位$\phi_r(x,y)$，绝对相位和包裹相位的关系可表示为

$$\phi_r(x,y)=\phi_l(x,y)+2k\pi \tag{3}$$

由于相移法需要投射一系列的相移条纹，因而不适用于动态场景三维测量。其次，由于相移法无法准确获得被测物体的绝对相位值，因此为了保证相位值的唯一性，通常将格雷码与相移法相结合来确保相位的准确性。另外，在空间相位展开过程中，当某一点相位值由于噪声或遮挡发生错误时，该错误会随相位展开路径向后传播，影响相位展开的准确性。

1.3　本文组合编码模式

通过上述分析可知，伪随机阵列编码测量方法能够满足动态三维测量，但空间分辨率较低。基于相位编码的相移法能够提供像素级的高分辨率测量，但由于需要投射多幅条纹图案，

因此不适用于动态三维测量。

为了满足动态场景下的高分辨率的三维测量，本文提出了一种组合编码方法。首先，为了满足高分辨率的三维测量，选择基于相位测量的正弦条纹编码；其次，为了能够适用于动态场景三维测量以及保证测量过程中相位值的准确性，选择伪随机阵列对正弦条纹进行再编码。为了减少被测物体表面颜色带来的干扰，使用单色正弦条纹编码。下面将详细论述本文的组合编码原理。

正弦条纹编码中，条纹的周期性无法确保相位值的唯一。为了保证编码图案相位值的唯一性，通常采用格雷码对相位值进行分区定位，但这种方法并不适用于动态场景三维测量。本文选择伪随机阵列对正弦条纹图案再编码，伪随机阵列的窗口特性确保了条纹图案中相位值的全局唯一性，使得正弦条纹编码能够满足高分辨率的动态三维测量。

伪随机阵列的窗口和正弦条纹周期对齐是组合编码方法的核心。本原多项式的阶数决定了伪随机阵列的窗口大小，且窗口大小是固定的，因此从大小来看窗口也具有周期性。由于正弦条纹也具有周期性，因此当窗口大小和正弦条纹编码的周期大小相同时，即可保证在一个窗口中只包含一个完整周期的正弦条纹，单个周期内条纹的相位值唯一，保证了编码图案中相位值的准确。假设正弦条纹的周期为T，窗口宽度大小W，则窗口中是否唯一包含单个完整周期的正弦条纹可表示为

$$F(T,W)=\begin{cases}\text{true}, & T=W\\ \text{false}, & T\neq W\end{cases} \tag{4}$$

被测物体表面高度变化过快或者表面面型具有复杂曲面会导致局部条纹缺失，如图3和图4所示。

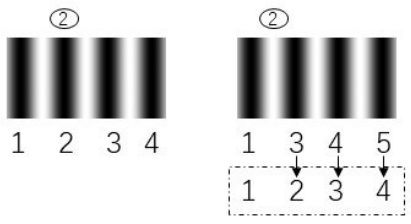


图3 正弦条纹编码图案局部条纹缺失

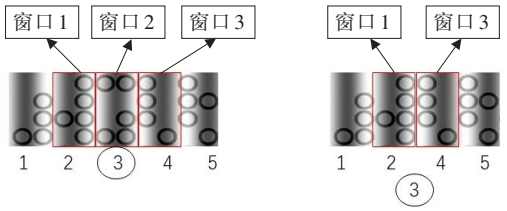


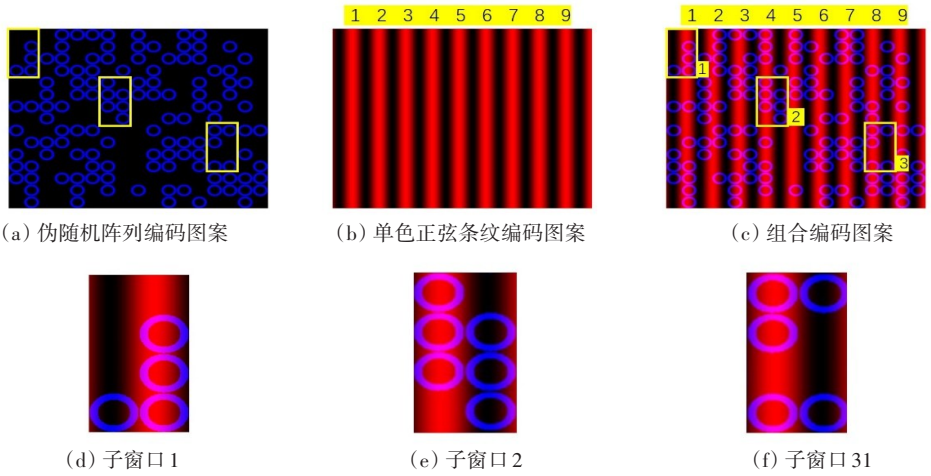
图4 组合编码图案局部条纹缺失

在传统的相移法中，当2号条纹发生缺失，单一相位编码方法无法确定缺失的条纹编号，导致产生的误差会从缺失的2号条纹处开始向后传递直到结束，造成相位值紊乱，致使解码出错。对于本文所设计的组合编码，通过伪随机阵列对正弦条纹进行了再编码，当发生条纹缺失时，通过伪随机阵列窗口特性依然可以保证后续的解码正确性，不会发生误差传递的现象。当被测物表面高度变化过快或者存在断层时发生条纹缺失的情况，当第3号条纹发生缺失，条纹处的伪随机阵列编码也发生缺失。通过伪随机阵列编码的窗口特性，通过窗口1和窗口3可以得出窗口2缺失的信息，从而判断出3号条纹缺失的信息，确保局部条纹的缺失不会影响到后续条纹的解码。

伪随机阵列编码中基元符号的选择至关重要，符号提取和识别与解码效率直接相关。因此，选取的符号应有明显的形状特征，且具备良好的平移、旋转和尺度不变特征。因此，选取圆环作为本文伪随机阵列编码的基元符号。通过Hu矩可以看出，圆环发生变形时Hu矩的七个特征值并没有很大变化，依然能够很好地反映符号特征。

表1 圆环及其变形图案的Hu矩

序号	图案	H[1]	H[2]	H[3]	H[4]	H[5]	H[6]	H[7]
1		0.3117	1.8798	5.2082	5.6888	11.4857	6.7103	10.9994
2		0.3472	2.3159	6.0722	6.0531	12.1180	8.5128	13.1779
3		0.3132	1.9120	5.1372	5.8916	11.5070	6.8575	11.3313
4		0.3140	1.9202	5.7560	5.8732	11.6879	8.9836	13.4816



（a）伪随机阵列编码图案　（b）单色正弦条纹编码图案　（c）组合编码图案

（d）子窗口1　（e）子窗口2　（f）子窗口31

图5　空间组合编码图案

本文选用二元八阶本原多项式生成长度为255的伪随机序列，按对角线折叠法形成大小为15×17的伪随机阵列，窗口大小为4×2。根据以上算法对大小为1024×768的图像编码，生成的组合编码图案以及截取的部分子窗口图案如图5所示。

表2　编码方法比较

编码方法	投射图案数量	点云密度
伪随机阵列	单幅	稀疏
相移法	N幅（N≥3）	稠密
本文组合编码图案	单幅	稠密

2　组合图案解码算法

组合图案解码算法分为伪随机阵列解码和正弦条纹解码两部分，首先对伪随机阵列图案进行窗口识别匹配，通过窗口定位找到正弦条纹图案中相应窗口位置，对正弦条纹进行傅里叶变换，得到包含物体高度信息的相位值。

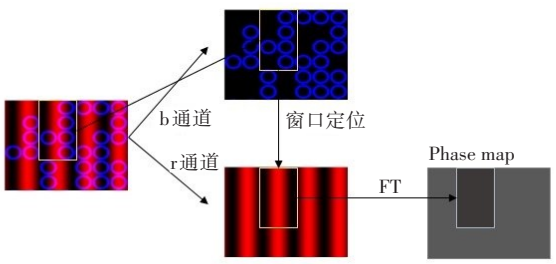


图6　组合图案解码示意图

伪随机阵列解码的本质是对符号特征的提取、识别以及匹配，编码符号良好的平移、旋转和尺度不变特性有利于提高解码效率，图像去噪、滤波等预处理能够使识别匹配更加准确。文献[9]中详细描述了伪随机阵列符号识别匹配过程，其原理如图7所示。

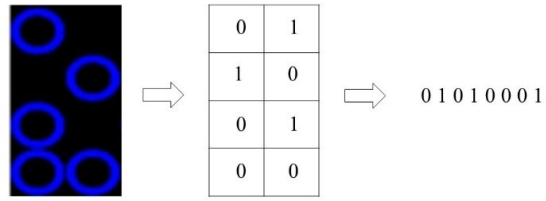


图7　窗口匹配原理

窗口内受到物体高度调制的正弦条纹图像可表示为

$$g(x,y)=a(x,y)+b(x,y)\cos\left(2\pi f_0 x+\phi(x,y)\right) \tag{5}$$

其中，$a(x,y)$为环境背景光强，$b(x,y)$为条纹对比度，f_0为正弦条纹基频，$\phi(x,y)$为被测物体高度信息的调制相位。对其进行傅里叶变换后通过滤波得到基频分量，再进行傅里叶逆变换得到物体调制相位值如下：

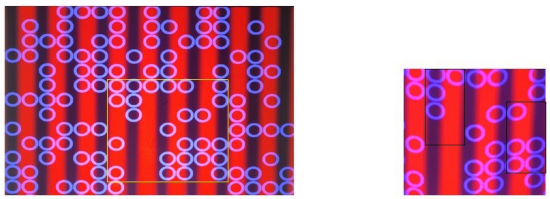
$$\phi(x,y)=\arctan\frac{\mathrm{Im}[q(x,y)]}{\mathrm{Re}[q(x,y)]} \tag{6}$$

通过伪随机阵列图案对正弦条纹图案进行了再编码，伪随机阵列的窗口唯一标识了相位图中的相位值，因此得到的相位无需进行相位展开。

3　实验结果及分析

实验采用的投影仪型号为爱普生（EPSON）

CB-2155W，分辨率为1280×800 pixel，工业相机型号为CK U300系列，分辨率为1282×768 pixel。将本文组合图案投射于待测物体表面，通过相机获取受到物体高度调制的变形图案以及局部图案，如图8所示。



（a）变形图案　（b）局部图案

图 8　采集的变形图案及其局部图案

从图8可以看出，变形图案的伪随机阵列窗口中依然唯一包含单周期正弦条纹，确保了后续解码过程中相位值的准确性。为了方便，截取大小为512×512的局部图案进行计算。将采集的原图案通过颜色通道分离得到伪随机阵列图案和正弦条纹图案，首先对伪随机阵列图案进行解码，通过图案预处理以及符号识别提取，其结果如图9所示。

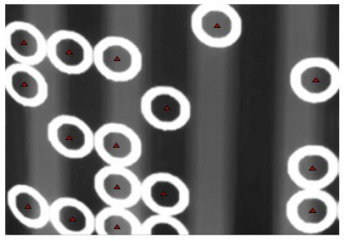
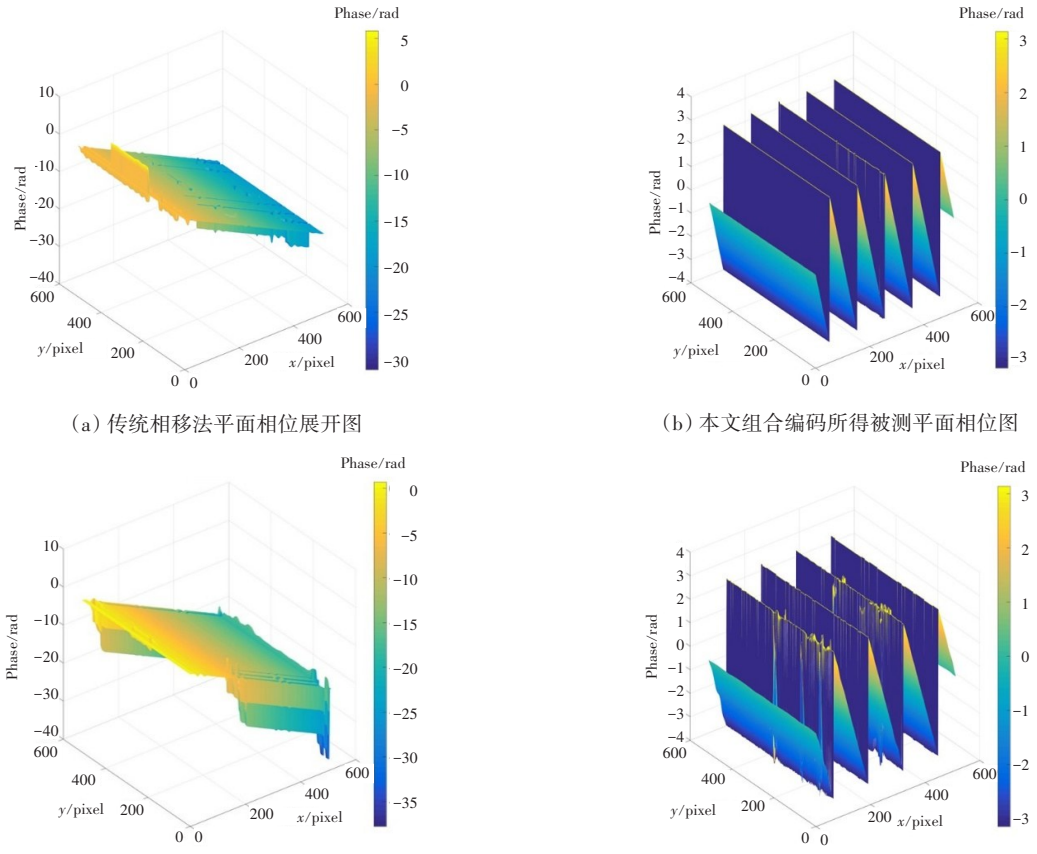


图 9　符号识别结果局部

伪随机阵列完成窗口匹配，通过窗口定位到正弦条纹图案对应窗口中的正弦条纹。前文编码算法确保窗口中只包含一个周期正弦条纹，确保相位值唯一。

通过傅里叶变化求得相位值，如图10所示。



（a）传统相移法平面相位展开图　（b）本文组合编码所得被测平面相位图

（c）传统相移法被测曲面相位展开图　（d）本文组合编码所得被测曲面相位图

图 10　相移法和本文算法所得相位图

其中，图(a)、(c)为传统相移法通过相位展开得到被测平板物体和柱体表面的相位图，图(b)、(d)为本文组合解码算法所得被测平板物体和柱体表面相位图。从图10可以看出，传统相移法中对于噪声等影响产生的噪点，由于相位展开过程中误差累积传播，所得相位图边缘处误差较大。在组合解码算法中，通过伪随机阵列对正弦条纹再编码确保相位值的唯一性，无需进行相位展开，因此相位图中的误差不会影响领域像素点。根据本文算法所得相位图，结合现有标定技术即可完成稠密三维测量。实验证明本文所提组合编码模式和解码算法的有效性，在快速稠密三维测量中具有很大应用前景。

4　结语

针对高分辨率的三维测量，提出了一种组合编码模式和解码算法，将伪随机阵列编码和单幅正弦条纹图案组合形成组合编码图案。基于相位测量的正弦条纹编码能够满足高分辨率的三维测量，伪随机阵列编码对正弦条纹进行再编码，使得单幅正弦条纹图案能够适用于动态场景下的三维测量。由于窗口特性确保了测量过程相位值的准确，因此得到的相位值无需再进行相位展开，结合现有的标定技术即可完成高分辨率的动态三维测量。实验结果说明了本文组合编码模式和解码算法的有效性。

参考文献：

[1] VAN DER JEUGHT S, DIRCKX J J J. Real-time structured light profilometry: a review[J]. Optics and Lasers in Engineering, 2016, 87: 18-31.

[2] PETKOVIĆ T, PRIBANIĆ T, ĐONLIĆ M. Single-shot dense 3D reconstruction using self-equalizing De Bruijn sequence[J]. IEEE Transactions on Image Processing, 2016, 25(11): 5131-5144.

[3] FU B, LI F, ZHANG T, et al. Single-shot colored speckle pattern for high accuracy depth sensing[J]. IEEE Sensors Journal, 2019, 19(17): 7591-7597.

[4] LU J, ZHANG X, GAO L, et al. Research on color structured light encoding and decoding method based on De Bruijn sequence[J]. Journal of Optoelectronics·Laser, 2014, 25(1): 147-155.

[5] SAGAWA R, FURUKAWA R, KAWASAKI H. Dense 3D reconstruction from high frame-rate video using a static grid pattern[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(9): 1733-1747.

[6] JIA X, WEI Y W, ZENG D. Binary encoding based on multi-element M array for structured light[J]. Journal of Optoelectronics·Laser, 2017, 28(9): 1025-1035.

[7] GREEN D H, TAYLOR I S. Irreducible polynomials over composite Galois fields and their applications in coding techniques[C]//Proceedings of the Institution of Electrical Engineers. IET Digital Library, 1974: 935-939.

[8] ITO M, ISHII A. A three-level checkerboard pattern (TCP) projection method for curved surface measurement[J]. Pattern Recognition, 1995, 28(1): 27-40

[9] LU J, LI J J, HUANG C M. Decoding of structured light based on symbol M-array[J]. Optics and Precision Engineering, 2013, 21(4): 1069-1078.

A combined encoding mode and decoding algorithm for structured light

Wang Weigang*, Wang Zemin, Zhao Xuan

(College of Electronic Information Engineering, Xi'an Technological University, Xi'an 710021)

Abstract: In order to meet the high-resolution dynamic 3D measurement, a structured light combined codec mode is designed. A coded pattern is formed by combining a pseudorandom array code and a single sinusoidal fringe pattern. The single sinusoidal fringe coding pattern can meet high-resolution three-dimensional measurement, and the single sinusoidal fringe pattern is recoded by pseudorandom array coding, so that the single sinusoidal fringe pattern can be applied to three-dimensional measurement in dynamic scenes. In order to ensure the accuracy of the phase value, the encoding parameters are set so that only one complete cycle of sinusoidal fringes is included in the sub-window of the pseudorandom array. The combination pattern decoding algorithm recognizes and matches through the window of the pseudorandom array, and then decodes the sinusoidal stripes in the window. Experimental results show the effectiveness of the above encoding and decoding modes.

Keywords: structured light encoding; pseudorandom array; window matching; sinusoidal stripes

文章编号: 1007-1423(2023)01-0034-06 DOI: 10.3969/j.issn.1007-1423.2023.01.006

基于自适应特征融合的场景文本检测

马艺舒，余艳梅*，陶青川

（四川大学电子信息学院，成都 610065）

摘要：基于深度学习的自然场景文本检测发展快速，其中基于分割的文本检测算法因其对多方向和弯曲文本检测效果好而备受关注。目前大多数基于分割的文本检测方法为了更加充分利用高层语义特征和底层细粒度特征，特征提取部分通常采用ResNet+特征金字塔(FPN)结构，特征融合部分多用concat或者add进行融合，但FPN存在的不同特征尺度不一致问题可能导致融合结果冲突，进而影响后续分割效果。因此，基于目前快速高效的DBnet网络，对其特征融合方式进行改进，提出了一种基于自适应特征融合的场景文本检测网络。在公开数据集Icdar2015和ICDAR 2017-MLT上的实验结果均表明：文本改进网络与经典的DBnet相比，准确率、召回率、F分数均有所提升，仅FPS稍有降低。

关键词：深度学习；特征金字塔；文本检测；自适应特征融合

0 引言

由于文字可能携带着重要且精准的信息，因此从文本图像中自动识别出文字逐渐成为重要的研究领域，光学字符识别(OCR)技术正是研究这一问题的有效手段[1]。目前OCR任务大多分两个阶段完成：文本检测和文本识别，文本检测旨在定位出图像中文字区域的位置，文本识别则把检测到的文本区域里的文字识别出来。文本检测作为OCR任务的第一步，是识别任务的前提，检测效果将会直接影响文本识别的效果[2]。而自然场景中的文本具有多样性，其受光照、噪声、遮挡等影响从而背景十分复杂，这给文本检测带来了极大的困难。传统的场景文本检测算法主要依赖人工设计特征，然后再用机器学习的方法学习文本信息，这样的方法虽然取得一定的成效，但具有很大的弊端，因为它很大程度依赖人的先验知识[3]。

近年来随着深度学习的发展，场景文本检测发展迅速，目前基于深度学习的文本检测大致分为基于回归和基于分割的方法：基于回归的文本目标通常用具有一定旋转角度的四边形去表示，但通常不能处理具有任意方向且形状不规则的文本；而基于分割的方法其基本思路为通过分割网络结构进行像素级别的语义分割，再基于分割的结果构建文本边界框，该方法目前因其灵活性备受关注[4]。Dan等[5]提出基于分割的PixelLink，其将同一实例内的像素链接在一起分割出文本实例，然后直接从分割结果中提取文本边界框而不进行位置回归，但后处理复杂，且模型推理速度非常慢；Zhou等[6]提出端到端的文本检测算法EAST，其增加了方向信息，可以检测任意形状和方向上的文本，但由于感受野不够，对长文本的检测效果不好；Lyu等[7]提出一种结合目标检测和语义分割的文本检测算法Corner，其可以同时处理多方向文本和极端高宽比的文本，虽然速度上有所突破，但依然达不到理想效果；Liao等[8]提出可微分二

收稿日期：2022-06-11 **修稿日期：**2022-08-06

作者简介：马艺舒(1997—)，女，陕西汉中人，硕士研究生，主要研究方向为计算机应用与图像处理；*通信作者：余艳梅(1975—)，女，四川广安人，博士，副教授，硕导，研究方向为图像处理，E-mail:yuyanmei@scu.edu.cn；陶青川(1972—)，男，四川南充人，博士，副教授，硕导，研究方向为计算机视觉

值化网络DBnet，其将二值化操作插入到分割网络中进行联合优化，并且在训练时是完全可微的，该方法凭借高效的推理速度被广泛应用在各大头部互联网的OCR任务中，但是，由于其特征提取部分使用的FPN会导致不同特征尺度不一致问题，直接concat进行特征融合可能会导致后续检测结果不准确。为了进一步提升检测准确率，同时不显著降低速度，本文对DBnet[8]进行改进，将FPN输出的多层特征采用自适应特征融合的策略进行融合。

本文所做的工作如下：

（1）改进了DBnet[8]中的特征融合结构，将concat结构改为本文提出的自适应特征融合结构。

（2）通过消融实验评估了参与自适应特征融合层级的数量。

（3）通过实验证明了在常用公开数据集上均得到了很好的检测效果。

1　本文的方法

1.1　总体网络架构

DBnet网络结构主要分为三个部分：由Resnet+FPN组成的特征提取部分；特征融合concat以及可微分二值化模块，本文在其基础上进行改进。图1所示为本文的整体网络架构，分为特征提取、自适应特征融合、可微分二值化模块三部分，其中特征提取部分和可微分二值化模块与DBnet保持一致。由于FPN存在不同特征尺度不一致的问题，直接将FPN每一输出特征层进行concat的融合结果可能会产生冲突，导致后续无法区分文字和背景[9]。为了解决上述问题，受文献[10]的启发，将DBnet中的特征融合部分改为本文提出的自适应特征融合结构（如图1中间所示），该结构可以让网络在FPN不同输出层级上直接学习如何在其他层上对特征进行筛选，从而只保留有效信息进行自适应的融合，最终将各个层级融合后的结果再相加，有效解决了FPN导致的融合结果冲突问题。

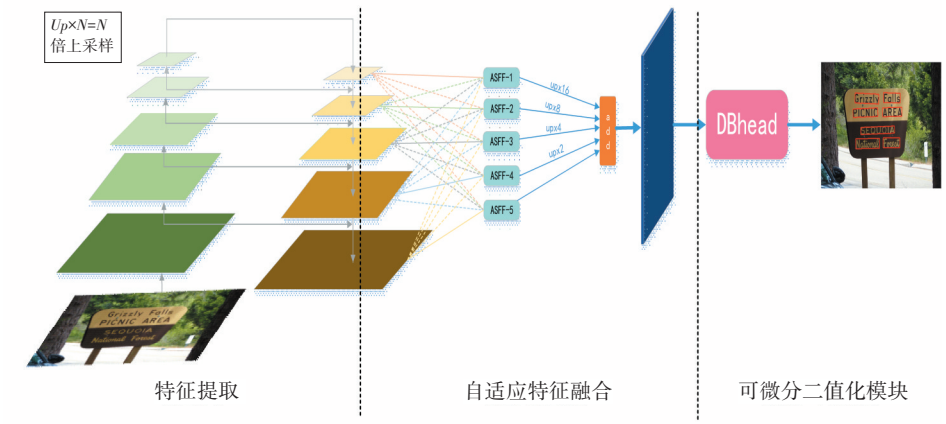


图1　本文的整体网络架构

1.1.1　特征提取结构

本文的特征提取部分采用与DBnet一致的策略，即ResNet+FPN结构，FPN包括一个自下而上的路径，一个自上而下的路径和一个横向连接路径，为了平衡模型检测准确率和模型的推理时间，减少网络需要学习的参数量，降低模型的复杂度[10]，其中FPN自下向上的部分，即特征提取主干本文选用ResNet-18进行构建。

1.1.2　自适应特征融合

由于FPN存在不同特征尺度不一致的问题，当图像同时含有大文本和小文本时，可能会导致后续融合结果发生冲突，无法将文本和背景区分开来。对此，本文改进了DBnet[8]中的特征融合部分，将其concat结构替换为本文提出的

自适应特征融合结构。如图2所示，对于FPN自顶向下每一层级的输出特征图采用类似文献[11]中的策略，将其他层级特征图通过卷积和上/下采样操作调整通道数和尺寸，从而与该层级保持一致，然后进行自适应融合，具体公式为

$$y_{ij}^{l}=\alpha_{ij}^{l}\cdot x_{ij}^{1\to l}+\beta_{ij}^{l}\cdot x_{ij}^{2\to l}+\gamma_{ij}^{l}\cdot x_{ij}^{3\to l}+\delta_{ij}^{l}\cdot x_{ij}^{4\to l}+\varepsilon_{ij}^{l}\cdot x_{ij}^{5\to l} \tag{1}$$

其中：$x_{ij}^{n\to l}$表示(i,j)处从层级n到层级l的特征向量；y_{ij}^{l}表示(i,j)位置融合后的结果；α_{ij}^{l}、β_{ij}^{l}、γ_{ij}^{l}、δ_{ij}^{l}、ε_{ij}^{l}为(i,j)位置的空间权重，其满足$\alpha_{ij}^{l}+\beta_{ij}^{l}+\gamma_{ij}^{l}+\delta_{ij}^{l}+\varepsilon_{ij}^{l}=1$，且$\alpha_{ij}^{l},\beta_{ij}^{l},\gamma_{ij}^{l},\varepsilon_{ij}^{l}\in[0,1]$，该权重可在反向传播中自动生成，通过训练自适应找到最优融合。本文最终从五个层级进行自适应融合，分别对应图2中ASFF-1、ASFF-2、ASFF-3、ASFF-4和ASFF-5，并将每个层级融合后的输出y^1、y^2、y^3、y^4、y^5分别调整至y^5大小然后进行逐像素相加，这样使得矛盾信息被过滤，进而后续检测结果更加准确。

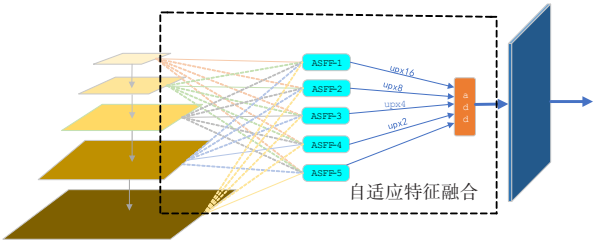


图2 自适应特征融合

1.1.3 可微分二值化模块

由于传统的二值化方法不可微，不能在训练时进行学习，所以本文使用了DBnet[8]网络中提出的可微分二值化模块，其用近似阶跃函数来近似二值化过程，具体公式为

$$B_{ij}=\frac{1}{1+e^{-k(P_{i,j}-T_{i,j})}} \tag{2}$$

其中：B为最终得到的近似二值图；T为学习得到的自适应阈值图；k则根据经验设置为50。

标签的生成采用PSENet[12]中的方法，由原始多边形以及收缩和扩张多边形得到，其中原始多边形为

$$G=\{SK\}_{k=1}^{n} \tag{3}$$

式中n代表文本框多边形的顶点个数，然后采用Vatti算法将G向内收缩为G_s，收缩的尺寸D为

$$D=\frac{A^2(1-r^2)}{L^2} \tag{4}$$

其中：A和L为原始多边形的面积和周长；r为收缩率，r根据经验设置为0.4。再将G以D扩张得到G_d，G_s和G_d之间的部分则为文本区域的边界，最后阈值图也采用类似的标注方式，可通过计算到达G中最近段的距离来生成阈值图的标签。

1.2 损失函数

为了得到更加准确的二值图，从而将文字区域和背景部分区分开来，本文采用了文献[8]中的损失函数，整个损失函数分为三个部分，即概率图P的损失函数L_s、二值图的损失函数L_b和自适应阈值图的损失函数L_t：

$$L=L_s+\alpha\times L_b+\beta\times L_t \tag{5}$$

其中：α和β分别设置为1.0和10，且L_s和L_b为二元交叉熵损失函数(BCE)：

$$L_s=L_b=\sum_{i\in S_l}y_i\log x_i+(1-y_i)\log(1-x_i) \tag{6}$$

此外，本文使用hard negative mining的策略来解决文本分类中正负样本不均衡问题，其中S_l是正负比为1∶3的采样集。L_t为生成的扩张多边形G_d里的预测结果和实际标签之间的L_1距离之和：

$$L_t=\sum_{i\in R_d}\left|y_i^{*}-x_i^{*}\right| \tag{7}$$

其中：R_d是扩张多边形G_d内像素的索引；y^{*}是自适应阈值图的真实标签。

2 实验及结果分析

2.1 实验设计

由于真实场景文本数据集的数量有限，为了达到更好的检测效果，本文首先对所有的模型使用合成数据集SynthText进行预训练，然后再分别在公开数据集Icdar2015和Icdar2017-MLT上进行1000次的微调，训练的批样本量设置为8，学习率设置采用Poly策略。

2.2 评价指标

本文使用与DBnet[8]一致的评价指标来评估改进后模型的文本检测效果，即准确率、召回率和F分数。

准确率计算公式为

$$P = \frac{TP}{TP + FP} \tag{8}$$

其中：TP代表实际为正例，预测也为正例的样本；FP代表实际为负例但预测为正例的样本，此数值越大越好。

召回率计算公式为

$$R = \frac{TP}{TP + FN} \tag{9}$$

其中：TP代表实际为正例，预测也为正例的样本；FN代表实际为正例但预测为负例的样本，此数值越大越好。

F分数是召回率和准确率的加权调和平均，计算公式为

$$F = \frac{2*PR}{P + R} \tag{10}$$

其中：P和R分别表示精准率、召回率，此数值越大越好。

2.3　实验环境

本文所有的实验均在Ubuntu16.04操作系统上进行，以Pytorch1.4.0作为深度学习框架，CUDA版本为10.1，实验所用编程语言均为Python3.6，CPU为Intel（R）Xeon（R）Platinum 8255C，GPU为NVIDIA RTX 2080 Ti。

2.4　实验结果与分析

在Icdar2015上的消融实验结果如表1所示。

表1　Icdar2015消融实验对比结果

序号	方法	P/%	R/%	F/%	FPS/(frame/s)
0	ResNet-18+FPN+DB	88.16	75.06	81.08	**29.40**
1	ResNet-18+FPN+ASFF1+DB	88.31	75.21	81.24	26.02
2	ResNet-18+FPN+ASFF2+DB	88.41	75.33	81.35	25.94
3	ResNet-18+FPN+ASFF3+DB	88.68	75.54	81.58	25.09
4	ResNet-18+FPN+ASFF4+DB	89.23	75.71	81.92	24.11
5	ResNet-18+FPN+ASFF5+DB	**89.48**	**75.79**	**82.07**	23.44
6	ResNet-18+FPN+ASFF6+DB	89.01	75.26	81.56	21.01

设未改进前的FPN+DB为方法0；在一个层级上进行自适应特征融合，即FPN+ASFF1+DB为方法1；在两个层级上进行自适应特征融合，即FPN+ASFF2+DB为方法2；在三个层级上进行自适应特征融合，即FPN+ASFF3+DB为方法3；在四个层级上进行自适应特征融合，即FPN+ASFF4+DB为方法4；在五个层级上进行自适应特征融合，即FPN+ASFF5+DB为方法5；在六个层级上（加上原图一起）进行自适应特征融合，即FPN+ASFF6+DB为方法6，可见方法5准确率、召回率和F分数均为最高。

在ICDAR 2017-MLT上的消融实验结果如表2所示，可见方法5在该数据集上虽然准确率比方法6略有降低（0.03个百分点），但方法6的FPS下降程度偏高，为了在准确率和速度两个方面进行权衡，最终本文选择在五个层级上进行特征融合。

表2　ICDAR 2017-MLT消融实验对比结果

序号	方法	P/%	R/%	F/%	FPS/(frame/s)
0	ResNet-18+FPN+DB	83.64	73.88	78.46	**21.41**
1	ResNet-18+FPN+ASFF1+DB	83.81	73.91	78.55	20.01
2	ResNet-18+FPN+ASFF2+DB	83.93	73.96	78.63	19.53
3	ResNet-18+FPN+ASFF3+DB	83.97	74.15	78.76	18.98
4	ResNet-18+FPN+ASFF4+DB	84.31	74.22	78.94	18.02
5	ResNet-18+FPN+ASFF5+DB	84.93	**74.39**	79.31	17.07
6	ResNet-18+FPN+ASFF6+DB	**84.96**	74.38	**79.33**	14.16

为了突出本文优势，我们还与其他方法进行了对比，表3展示了在Icdar2015数据集上进行对比实验的结果，实验结果直观地显示了本文方法具有不错的表现（准确率、召回率、F分数均为最高，速度比最快的DBnet略有降低）。

表 3 在Icdar2015上的对比结果

方法	*P*/%	*R*/%	*F*/%	*FPS*/(frame/s)
PixelLink[5]	81.06	69.01	74.55	1.96
EAST[6]	85.21	66.04	71.41	—
Corner[7]	85.62	71.23	77.76	3.91
TextSnake[13]	81.55	69.44	75.01	1.00
SBD[14]	87.43	75.21	80.86	2.01
CRAFT[15]	86.07	73.95	79.55	6.50
PAN[16]	83.20	73.58	78.09	8.53
DBnet[8]	88.16	75.06	81.08	**29.40**
Our approach	**89.48**	**75.79**	**82.07**	23.44

3 结语

为改善自然场景文本检测的效果，本文提出了一种基于DBnet改进的自适应特征融合的文本检测方法，该方法可以克服FPN不同特征尺度不一致性引起的特征融合干扰，在公开数据集Icdar2015和ICDAR 2017-MLT上进行的对比实验表明：相比较于其他方法，本文的方法在场景文本检测任务上有更好的表现，提升文本检测准确率的同时检测速度仅有略微下降。且根据消融实验结果，五个层级同时参与融合时效果最佳，基于此结果，本文最终选择将五个层级的自适应特征融合输出结果进行最终的融合。在未来的研究中，考虑加入其他预处理手段，使模型在速度不显著下降的前提下进一步提升准确率。

参考文献：

[1] 于文清. 基于深度学习的自然场景文本检测算法研究与应用[D]. 石家庄:河北地质大学,2022.

[2] 江杨心. 基于深度学习的OCR票据识别系统的研究与实现[D]. 上海:华东师范大学,2022.

[3] 王建新,王子亚,田萱. 基于深度学习的自然场景文本检测与识别综述[J]. 软件学报,2020,31(5):1465-1496.

[4] 刘艳菊,伊鑫海,李炎阁,等. 深度学习在场景文字识别技术中的应用综述[J]. 计算机工程与应用,2022,58(4):52-63.

[5] DAN D, LIU H, LI X, et al. PixelLink: detecting scene text via instance segmentation [C] //32nd AAAI Conference on Artificial Intelligence,2018.

[6] ZHOU X, YAO C, WEN H, et al. EAST: an efficient and accurate scene text detector[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE,2017.

[7] LYU P, YAO C, WU W H, et al. Multi-oriented scene text detection via corner localization and region segmentation [C] //Proceedings of 2018IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City,USA:IEEE,2018:7553-7563.

[8] LIAO M H, WAN Z Y, YAO C, et al. Real-time scene text detection with differentiable Binarization [C]//Proceedings of the 34th AAAI Conference on Artficial Intelligence. New York, USA: AAAI, 2020: 11474-11481.

[9] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C] //2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2017.

[10] LIU S, HUANG D, WANG Y. Learning spatial fusion for single-shot object detection[EB/OL]. arXiv: 1911.09516,2019.

[11] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas,NV,USA,2016:16541111.

[12] LI X, WANG W, HOU W, et al. Shape Robust Text Detection with Progressive Scale Expansion Network [C] //2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Long Beach,CA,USA,2020:19263662.

[13] LONG S, RUAN J, ZHANG W, et al. (2018). TextSnake:a flexible representation for detecting text of arbitrary shapes[M]//FERRARI V,HEBERT M, SMINCHISESCU C, et al. Computer Vision-ECCV 2018. ECCV 2018. Lecture Notes in Computer Science(),vol 11206. Springer,Cham,2018:19-35.

[14] LIU Y, HE T, CHEN H, et al. Exploring the capacity of sequential-free box discretization network for omnidirectional scene text detection[EB/OL]. arXiv: 1912.09629,2019.

[15] BAEK Y, LEE B, HAN D, et al. Character region awareness for text detection [C] //2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). IEEE,2019.

[16] WANG W, XIE E, SONG X, et al. Efficient and accurate arbitrary-shaped text detection with pixel aggregation network[J]. 2019 IEEE/CVF International Conference on Computer Vision(ICCV). Seoul,Korea(South),2019:19398281.

(下转第44页)

文章编号: 1007-1423(2023)01-0039-06 DOI: 10.3969/j.issn.1007-1423.2023.01.007

基于关键点的未爆弹图像目标检测算法

单成之[1]，张 健[2*]

（1.武警警官学院训练基地，广州 510440；2.武警海南总队，海口 570203）

摘要：针对排爆机器人智能识别未爆弹的问题，提出一种基于关键点的未爆弹图像目标检测算法。首先以ResNet-50为主干网络，对未爆弹图像进行初步特征提取，然后引入分离注意力模块的特征增强网络，增强了对未爆弹目标图像背景复杂和存在遮挡现象的适应能力，最后输入头部预测模块得到未爆弹类别和位置信息。实验结果表明，本算法能够取得较高的检测精度和实时性，为各类未爆弹的鉴定和排爆机器人向自主排爆提供了便利。

关键词：未爆弹；排爆机器人；关键点；目标检测

0 引言

近些年来，随着我国的交通运输、水利工程、房地产等基础设施建设的不断发展，在各类建设施工开挖过程中，常常挖掘出一些战争时代残留的各种不同型号、种类的废旧手榴弹、炮弹、迫击炮弹等未爆弹。为消除其对人民生命财产安全带来的严重隐患，在对未爆弹销毁前，需要经专业人员对其品种和危险程度逐个进行鉴别[1-3]。此外，依靠人工排除未爆弹效率较低，且危险性极大。利用排爆机器人代替人工完成这一危险性任务，成为当下的研究重点。排爆机器人通过摄像头采集回来图像数据进行处理，实时分析未爆弹类别及其所在位置。因此，基于人工智能算法实现对未爆弹的图像检测，使排爆机器人向自主排爆迈出关键一步。

未爆弹图像目标识别主要任务是识别未爆弹的种类，同时在图像中定位未爆弹的位置。由于应用场所和待识别目标的特殊性，在识别过程中待检测物品外观锈蚀严重和环境差异大，导致未爆弹图像成像背景复杂，检测物存在遮挡现象等问题，给识别带来了诸多挑战。文献[4]利用哈特利变换对炮弹图像进行特征提取后输入到SVM分类器，实现炮弹图像特征的快速分类。文献[5]通过Xception深度网络提取废旧手榴弹图像特征，使用L2-SVM的合页损失对归一化特征进行分类，取得较好的分类效果。文献[6]通过Faster R-CNN目标检测算法实现了对未爆弹目标识别与位置信息获取，算法具有较好的鲁棒性。文献[7]将深度图像量化为点云数据，图像坐标数据转化为世界坐标系数据，实现对现实物品的检测和抓取。由此可见，基于深度学习的未爆弹图像检测算法能够实现对目标物的自主识别和定位，提高了检测效率和自主性。

本文受CenterNet算法启发[8]，提出了基于关键点的未爆弹图像目标检测算法。采用ResNet-50进行初步特征提取，利用特征增强网络，提升对存在背景复杂和遮挡未爆弹目标的适应能力，最后输入头部预测模块得到未爆弹类别和位置信息。在预训练阶段不需要提供anchor的大小，在检测阶段不需要非极大值抑制的后处理，算法流程清晰简单。实验结果表明，本文算法具有更高的检测精度和较快的检测速度。

1 理论与方法

1.1 基于预选框的目标检测算法

目前，基于预选框(anchor-based)的目标检

收稿日期：2022-07-10 **修稿日期：**2022-08-05

作者简介：单成之(1969—)，男，江苏沭阳人，本科，教授，主要研究方向为作战指挥、炮兵射击指挥；*通信作者：张健(1980—)，女，安徽芜湖人，硕士，高级工程师，主要研究方向为信号与信息处理、数据分析，E-mail：zdzjlwyx@163.com

测算法分为单阶段(one-stage)和二阶段(two-stage)。二阶段即为首先需要利用区域候选网络(RPN)生成感兴趣区域(ROI)，然后再对该区域进行目标分类和位置回归。常见的二阶段目标检测算法有R-CNN、SPP-Net、Fast R-CNN、Faster R-CNN和R-FCN等。单阶段即为直接利用回归的思想，使用单一网络完成目标类别和位置的检测，检测速度大幅提升。常见的单阶段目标检测算法有SSD、YOLO系列、RetinaNet等。基于预选框(anchor-based)的目标检测算法在检测速度和精度上都取得了较大进步，但仍有以下缺陷：使用anchor时，需要在每个特征尺度上密集平铺，导致仅有很少一部分是正样本，造成正负样本不均衡问题；anchor设计需要很细致，一般利用统计学从训练集或者测试集中获取，难以推广到其他的数据集；每个像素网格点有多个锚点框，引入更多的超参数，加大训练难度；二阶段和单阶段都存在非极大值抑制(non maximum suppression，NMS)后处理的问题。

1.2 基于无预选框的目标检测算法

从CornerNet开始[9]，便开启了基于无预选框(anchor-free)的目标检测算法新篇章，其主要是抛弃了目前主流单阶段和二阶段目标检测算法都用到的anchor boxes。CenterNet在CornerNet基础上进一步优化，得到了更高检测精度和更快的检测速度。CenterNet的检测器将检测目标看作一个点，即目标框的中心点，从而将目标检测问题转化为对关键点的估计问题。即为将图像传入到全卷积网络中得到热力图，热力图峰值点(keypoint heatmap)是中心点，边界框的尺寸大小和其他对象属性根据中心点推断出来。

2 网络方案设计

2.1 网络结构

针对未爆弹图像特点，本文网络模型包括主干网络、特征增强网络和头部预测网络，如图1所示。主干网络负责获取输入图像的特征信息；特征加强模块实现对主干网络输出的初步特征信息进行特征强化，增强有益特征信息并抑制无用特征信息；头部预测模块主要包含了三个分支，对特征加强的信息进行类别中心点热力图、中心点的偏置和目标bbox的高宽计算。

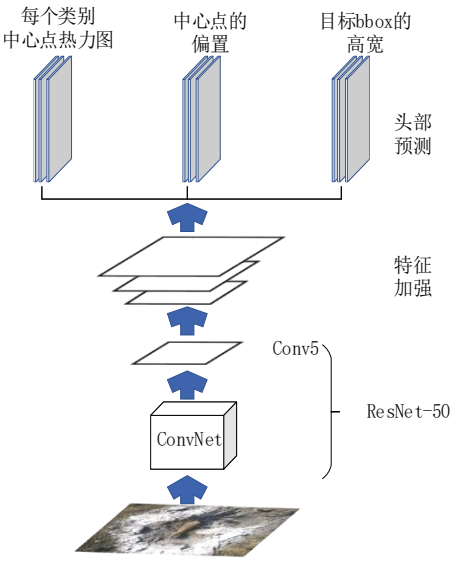


图1 未爆弹图像检测网络模型

将未爆弹图像$I \in R^{W \times H \times 3}$输入到ResNet-50主干网络中进行特征提取，数据流至最后一个子网络Conv5得到通道数为2048，特征图尺寸为16 × 16的初步特征层；接着初步特征依次完成三次深度卷积、反卷积上采样、深度卷积特征增强处理，从而获得高分辨率特征图输出，生成语义信息和空间信息更丰富的特征预测图；最后将该特征预测图作为头部预测网络的输入，完成目标热力图、中心点和宽高预测，输出得到违禁品的类别和位置。

2.2 特征增强模块

为使主干网络输出的初步特征层进一步适用于未爆弹图像检测，通过引入ResNeSt模块[10]，进一步加强图像特征粒度信息，如图2所示。

图2展示了ResNeSt模块，其中包含特征图组(feature-map group)和拆分注意力(split attention)操作。输入特征图可以沿通道维度分为几组，特征图组中特征图被分为多个组，每个组内又进行分组。超参数K和R分别表示特征图组的数量和基数组内的分组数，总的特征图分组数可以表示为$G = K \times R$。对每个单独的组应用一系列变换$\{F_1, F_2, \cdots, F_G\}$，每个组的中间表示为$U_i = F_1(X)$，其中$i \in \{1, 2, \cdots, G\}$。在基数组内的每个分组中进行1 × 1和3 × 3的卷积，得到R个特征图后进行拆分注意力操作。

2.3　头部预测模块

主干网络只进行了4倍下采样，特征图比一般网络的分辨率高很多，所以不需要进行多尺度预测和特征金字塔也可以较好地检测目标大小。通过特征增强网络后得到一个更有效的图像语义信息和位置信息，将这些信息输入到头部预测网络，对热力图、中心点和宽高进行预测，如图3所示。

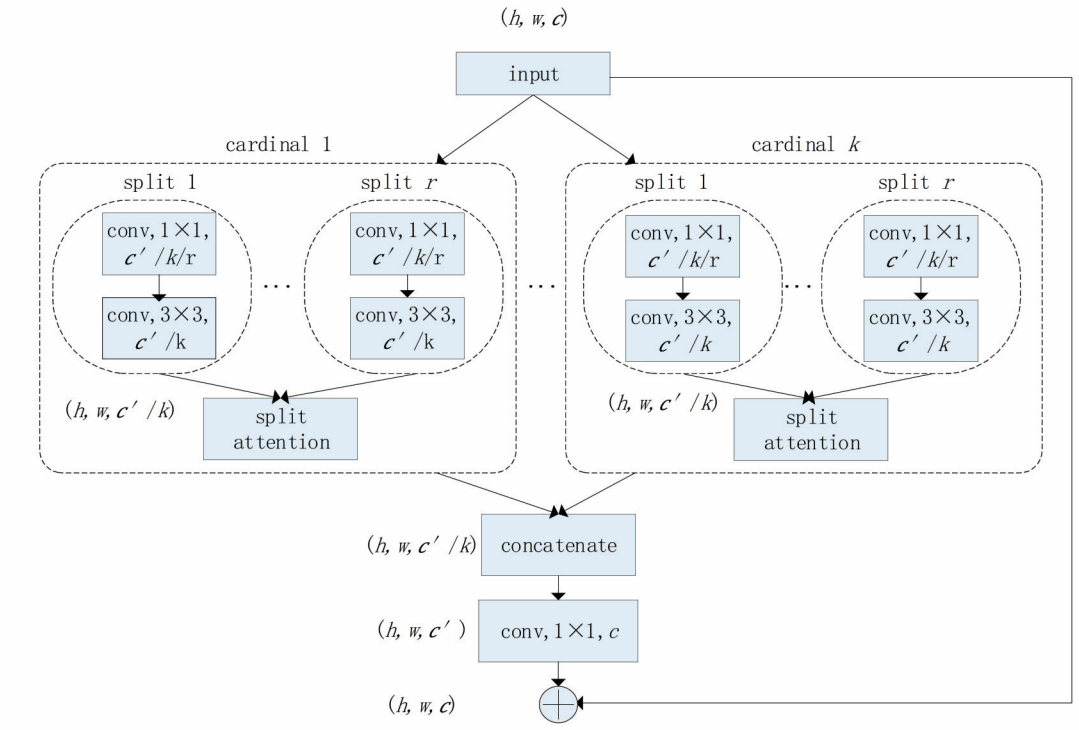


图2　ResNeSt模块

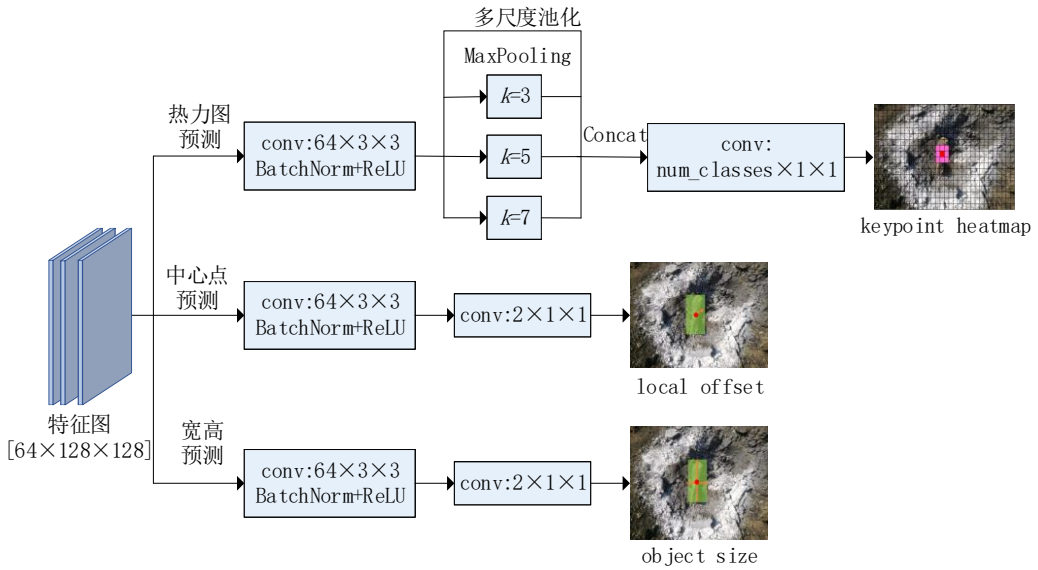


图3　头部预测模块

热力图预测分支中，对高分辨率特征图先进行常规的卷积操作，归一化处理和使用RuLU作为激活函数，然后进行多尺度池化(multi scale pooling，MSP)操作，通过融合不同大小的最大池化层来获得鲁棒的特征表示，最大池化的池化核 $k=\{1\times1,3\times3,5\times5,7\times7\}$，将来自不同内核大小的特征图连接(concat)一起作为输出，接着进行1×1卷积降维操作，通道数为num_classes，最终结果为(128,128,num_classes)，代表每一个热力点是否有物体存在，以及物体的种类。

中心点预测分支中，对热力图的输出进行精炼，提高定位准确度。对高分辨率特征图先进行常规的卷积操作，归一化处理和使用RuLU作为激活函数，再进行1×1卷积降维操作，此时卷积的通道数为2，最终结果为(128,128,2)，代表每一个物体中心距离热力点偏移的情况。

宽高预测分支中，预测以关键点为中心的检测框的宽高。对高分辨率特征图先进行常规的卷积操作，归一化处理和使用RuLU作为激活函数，再进行1×1卷积降维操作，此时卷积的通道数为2，最终结果为(128,128,2)，代表每一个物体宽高的预测情况。

3 实验结果与分析

3.1 实验环境与参数设置

本实验使用的硬件环境为：Intel(R) Core(TM) i7-8750H CPU@2.20GHz处理器，Nvidia GeForce GTX 1070 8 GB显卡，16 GB内存。软件环境为：深度学习框架Pytorch，加速环境CUDA10.1，编程语言Python3。

在模型训练过程中，输入图像尺寸：512 pixel×512 pixel，输出分辨率：128 pixel×128 pixel(即4倍stride)，针对每个样本采用简单的数据增强方式：随机翻转；随机放缩，其中比例在0.6到1.3；裁剪；色域变换，采用小批量样本集的梯度下降法，Adam优化器，初始学习率设定为 1.0×10^{-4}。

3.2 实验数据与评价指标

本文实验图像数据全部采自公开网络。由于未爆弹图像数据库较匮乏，采用预设的数据变换规则，在已有数据的基础上进行有监督的数据增强，对原始图像进行数据扩充。数据集含有5类共1200张未爆弹图像，利用LabelImg开源图像标注工具对数据进行标定[11]，最终形成本次实验的未爆弹数据集。

为了更好验证检测算法的有效性，本文采用平均精度(average precision，AP)、均值平均精度(mean average precision，mAP)和每秒帧率(frames per second，FPS)三项指标对算法模型评估。AP由精确率(precision，P)和召回率(recall，R)曲线和坐标围起来的面积组成，用于计算平均的检测精度和衡量检测器在每个类别上的性能好坏。mAP用于评价多目标的检测器性能，衡量检测器在所有类别上的性能好坏，即得到每个类别的AP值后再取所有类别的平均值。精确率(P)、召回率(R)、平均精度(AP)和均值平均精度(mAP)的数学定义如下所示[12]。各标准的具体公式如下：

$$R=\frac{TP}{TP+FN} \tag{1}$$

$$P=\frac{TP}{TP+FP} \tag{2}$$

$$AP=\int_0^1 P_{\text{smooth}}(r)\mathrm{d}r \tag{3}$$

$$mAP=\frac{\int_{i=1}^{K} AP_i}{K} \tag{4}$$

式中：*TP*表示被预测为正样本、实际为正样本的数量；*FP*表示被预测为负样本、实际为负样本的数量；*FN*表示被预测为正样本、实际为负样本的数量；$P_{\text{smooth}}(r)$表示平滑的P-R曲线，*K*表示类别数。

此外，*FPS*也是未爆弹图像目标检测的一个重要评价指标，指一秒内识别的图像数。*FPS*值越大，检测速度越快，实时性越高。

3.3 实验结果与实验分析

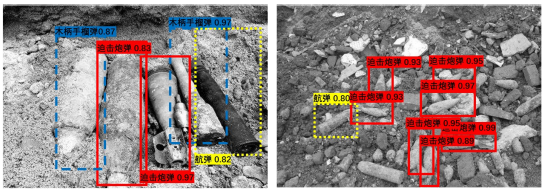
为了验证本文算法的性能，在未爆弹图像测试集上进行了实验，结果如表1所示。

表1 检测算法的检测精度

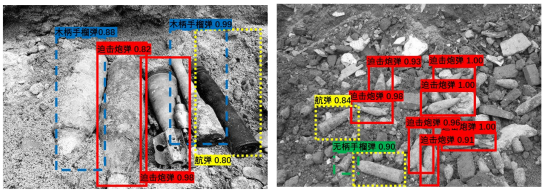
算法	Backbone	*mAP*/%	*FPS*/(frame/s)
本文算法	Resnet-50	94.16	34

由表1可见，本文算法在*mAP*和*FPS*性能

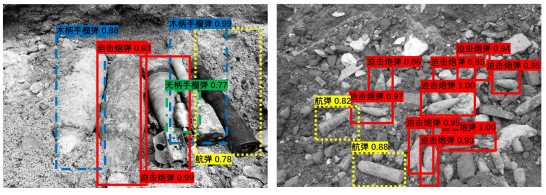
结果分别为94.16%和34 frame/s，验证了本文算法的有效性。为了进一步验证本文算法的有效性，分别将SSD[13]、YOLOv3[14]、CenterNet等经典模型与本文算法在未爆弹图像测试集上进行检测可视化，部分检测可视化结果如图4所示。



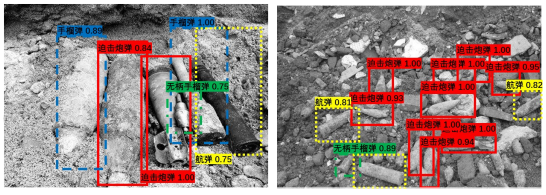
（a）SSD



（b）YOLOv3



（c）CenterNet



（d）ours

图4　未爆弹图像检测算法结果对比

图4可视化了不同算法未爆弹图像检测结果，其中数字为目标的置信度。由图4可以看出，相比SSD、YOLOv3、CenterNet等经典模型，本文算法在识别准确率和定位精度上取得了更好的检测效果。在图4中多个未爆弹目标相互遮挡，背景复杂混乱且与未爆弹目标颜色相似，SSD算法对违禁品目标检测的漏检率比较高，YOLOv3和CenterNet算法的检测效果有较好的提升，但也仍然存在漏检问题，本文算法对多个遮挡未爆弹检测具有较好的效果和较大的提升。

4　结语

为进一步提高未爆弹的检测速度和精度，本文提出了一种基于关键点的未爆弹图像检测算法。以目标中心为关键点的检测算法，保证在较快的检测速度前提下，能够准确地检测出背景复杂、遮挡的未爆弹目标。实验结果表明，在未爆弹测试数据集上与三种先进的目标检测方法进行实验验证，本文算法能够得到更高的准确率和速度。下一阶段研究中，将进一步扩展未爆弹图像目标数据，同时在确保检测速度不降低的前提下，提升算法在密集复杂物品背景下的未爆弹检测能力，为实现智能化排爆机器人精准抓取未爆弹打下基础。

参考文献：

[1] 张博文，肖明辉. 废旧民用爆炸物品销毁方法研究[J]. 大众标准化，2021(23)：38-40.

[2] 张炜. 爆炸法销毁废旧炮弹的实施过程及效果[J]. 工程爆破，2019，25(5)：87-90.

[3] 周晓光，董恒超，何泽，等. 销毁废旧炮弹及危险爆炸物品的设计与实践[J]. 爆破，2020，37(4)：132-137.

[4] 钱玮，张旭东，丁俊香，等. 基于SVM思想的炮弹图像识别方法[C]//第十四届中国神经网络学术会议论文集，2004：522-525.

[5] 林进健，陈世娥. 基于CNN-SVM的废旧手榴弹图像特征分类[J]. 武警工程大学学报，2020(4)：32-35.

[6] 胡聪，何晓晖，邵发明，等. 基于Faster R-CNN的未爆弹检测[J]. 机电产品开发与创新，2021，34(5)：105-107.

[7] 王浩，何昌见，倪俊杰. 基于深度图像的排爆机器人目标物检测抓取方法[J]. 中国安防，2019(12)：85-90.

[8] ZHOU X Y，WANG D Q，KRÄHENBÜHL P，et al. Objects as points[EB/OL].(2019-05-25)[2020-08-05]. https://arxiv.org/abs/1904.07850.

[9] LAW H，DENG J. Cornernet：detecting objects as paired key points[C]//Proceedings of the European Conference on Computer Vision (ECCV)，2018：734-750.

[10] 许骞艺. 基于改进的ResNeSt驾驶员头部状态分类算法[J]. 吉林大学学报(工学版)，2021(2)：704-711.

[11] 刘攀. 基于FasterRCNN的口罩佩戴检测研究[J]. 电脑知识与技术，2021(21)：14-16.

[12] 穆思奇,林进健,汪海泉,等. 基于改进YOLOv4的X射线图像违禁品检测算法[J]. 兵工学报,2021,42(12):2675-2683.
[13] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[EB/OL]. (2015-12-08)[2020-07-12]. https://arxiv.org/abs/1512.02325.
[14] REDMON J, FARHADI A. YOLOv3: an incremental improvement [EB/OL]. (2018-04-08) [2020-01-12]. https://arxiv.org/abs/1804.02767.

An algorithm for object detection in unexploded bombs images based on key points

Shan Chengzhi[1], Zhang Jian[2*]

(1. Officers College of PAP, Guangzhou 510440; 2. Hainan Corps of CAPF, Haikou 570203)

Abstract: Aiming at the problem of intelligent identification of unexploded bombs by explosive ordnance disposal robots, the paper proposes an image target detection algorithm of unexploded bombs based on key points. The paper firstly uses ResNet-50 as the backbone network to perform preliminary feature extraction on the unexploded bomb image, and then introduces the feature enhancement network of the separation attention module to enhance the adaptability to the complex background and occlusion phenomenon of the unexploded bomb target image, and finally input The head prediction module obtains the category and location information of unexploded bombs. The experimental results show that the algorithm in the paper can achieve high detection accuracy and real-time performance, which provides convenience for the identification of various types of unexploded bombs and the autonomous explosive discharge of robots.

Keywords: unexploded bombs; explosive ordnance disposal robots; key points; object detection

(上接第38页)

Scene text detection based on adaptive feature fusion

Ma Yishu, Yu Yanmei*, Tao Qingchuan

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

Abstract: Deep learning based natural scene text detection develops rapidly, among which segmentation based text detection algorithm has attracted much attention because of its good effect on multi-direction and curved text detection. At present, in order to make full use of high-level semantic features and low-level fine-grained features, most text detection methods based on segmentation usually adopt ResNet+ Feature Pyramid (FPN) structure for feature extraction, and CONCAT or ADD for feature fusion. However, the inconsistency of different feature scales in FPN may lead to the conflict of fusion results, which will affect the subsequent segmentation effect. Therefore, based on the current fast and efficient DBnet network, this paper improves its feature fusion method, and proposes a scene text detection network based on adaptive feature fusion. The experimental results on the public datasets Icdar2015 and ICDAR 2017-MLT show that compared with the classical DBnet, the accuracy, recall and F-score of the text improved network are improved, but the FPS is slightly reduced.

Keywords: deep learning; pyramid of features; text detection; adaptive feature fusion

文章编号: 1007-1423(2023)01-0045-06 DOI: 10.3969/j.issn.1007-1423.2023.01.008

基于灰狼优化算法的灌区流量预测模型研究

单无牵[1]，宁 芊[1,2*]，周新志[1]，张人元[1]，罗 强[3]

（1.四川大学电子信息学院，成都 610065；2.新疆师范大学计算机科学技术学院，乌鲁木齐 830054；
3.成都万江港利科技股份有限公司，成都 610064）

摘 要：快速准确的灌区流量预测对于水资源分配有着重要依据。现有的流量预测模型一般采用传统方法或是简单的神经网络模型，采用上述方式存在操作复杂或预测精度低等缺陷。因此，本文提出使用灰狼优化算法优化BP模型，设计出一种新的GWO-BP模型。为了验证模型的优越性，将该模型应用于甘肃省花海灌区的实测流量上，并与传统的经验公式法和BP神经网络模型进行对比。结果显示GWO-BP预测结果更加稳定可靠，满足灌区水资源计量的工程精度要求。

关键词：水资源；流量；灰狼优化算法；神经网络

0 引言

虽然我国土地辽阔，淡水资源总量为28000亿立方米，占全球水资源的6%，名列世界第四，但是我国人口众多，人均水资源量只有2300立方米，仅为世界平均水平的25%，所以我国是一个严重缺水的国家[1]。同时，不同地域不同时期水资源的分布不均匀，更加加剧了问题的严重性。在这样的一个背景下，很多地区都希望能够准确地测量出灌区的流量值，以此为依据，为计划用水以及水资源调配等工作进行量化参考，能够对水资源进行更合理的管理与分配[2]。

针对灌区流量测算的研究大致分为两种：第一种是在传统的流量测算方法的基础上，对测量设备进行改进，例如改进流速仪、水位测量装置等；第二种方法是研究与流量有关的中间变量，例如水位、流速、几何形状等，研究中间变量与流量之间的内在联系，通过相应的算法来实现。例如曹登峰等[3]、周冬生等[4]改进了水位计、流速仪等测量设备或使用雷达等新型测量工具提高流量计算的精度。Lin等[5]使用软测量技术，以闸门开度、水位等条件为输入进行模糊边界流场计算。Aggarwal等[6]以实测河流数据为基础，通过三种时序模型来进行流量评估。张振等[7]使用图像算法替代了流速测量工具，建立了RBFNN软测量模型。文献[8-10]采用自适应神经模糊推理系统(ANFIS)研究了90°急弯的流量，对其进行数值模拟以及水文信息预测。

为了提高灌区流量预测精度，本文提出使用灰狼优化算法(WGO)优化BP神经网络模型的权值和阈值。首先使用经验公式法和BP神经网络创建相应的预测模型，并指出相应缺陷；其次在花海灌区的实测数据的支撑下对三种模型进行比较。

收稿日期：2022-09-24 **修稿日期：**2022-10-17

基金项目：新疆维吾尔自治区区域协同创新专项(科技援疆计划)(2020E0247，2019E0214)

作者简介：单无牵(1998—)，男，四川广元人，硕士研究生，研究方向为水利信息化；*通信作者：宁芊(1969—)，女，四川成都人，博士，教授，研究方向为水利信息化、物联网等，E-mail:ningq@scu.edu.cn；周新志(1966—)，男，四川德阳人，教授，博士生导师，研究方向为智能计算与控制，水利信息化；张人元(1998—)，男，四川巴中人，硕士研究生，研究方向为水利信息化；罗强(1977—)，男，四川成都人，硕士，工程师，从事领域为水利信息化

1 研究区域概况

1.1 基本情况

花海灌区属于石油河下游的花海盆地，具体位置位于玉门市老市区以北76公里左右，灌区中心位于东经97°44′，北纬41°18′，南靠宽滩山北麓戈壁，北靠马鬃山前戈壁，西接昌马灌区青山农场，东与金塔县接壤，面积3878平方公里。灌区深居亚欧大陆腹地，为典型的大陆荒漠气候，终年干旱少雨或雪，风沙大，光照强，昼夜温差大，年降水量70.1毫米，年蒸发量2982.5毫米，区内海拔介于1215～1325米之间，南北自然坡度1/500，东西自然坡度1/1000。

1.2 数据来源

花海灌区内包含各种干渠以及众多支渠，无论干渠或是支渠，流量测算时均采用流速-面积法。测量时根据渠道的物理参数以及水位信息等合理布置流速-面积法所需要的测线数目、测点个数及位置。在对花海灌区的数据资料进行筛选后，选择了北直四农、北直七农、东二支三斗马家峪、北直十二农水桥、东干渠、西干渠等22条渠道、共计147条流量数据。

2 研究方法

2.1 经验公式法

法国工程师谢才在1769年提出了一种计算明渠和管道均匀流平均流速、流量沿程水头损失的经验公式[11]：

$$V = C\sqrt{RJ} \tag{1}$$

该公式的求解核心是求解谢才系数C，针对这一情况，不同学者提出了不同的经验公式，然而在实际运用工程中最为简便而运用广泛的是曼宁公式：

$$C = \frac{1}{n}\sqrt[6]{R} \tag{2}$$

利用谢才公式可以算出明渠断面的平均速度，曼宁公式则用来确定谢才公式中的谢才系数C。引入流量Q，谢才-曼宁公式[11]又可以写成：

$$Q = AV = AC\sqrt{RJ} = \frac{1}{n}AR^{\frac{2}{3}}J^{\frac{1}{2}} \tag{3}$$

式中：n为糙率；C为谢才系数；R为水力半径；A为过水断面面积；J为坡度。

2.2 BP神经网络

神经网络是由多个神经元以及层级相互连接组成的拓扑结构，BP神经网络[10]由于结构简单、使用方便常常用于解决各种工程问题。BP神经网络包括向前传播和误差反向传播两部分。向前传播时，网络中各神经元接受前一级的输入，并输出到下一级。误差反向传播部分是根据真实输出与期望输出之间的误差来不断修正神经网络的权值和阈值，从而使得神经网络的预测误差越来越小。训练的流程图如图1所示。

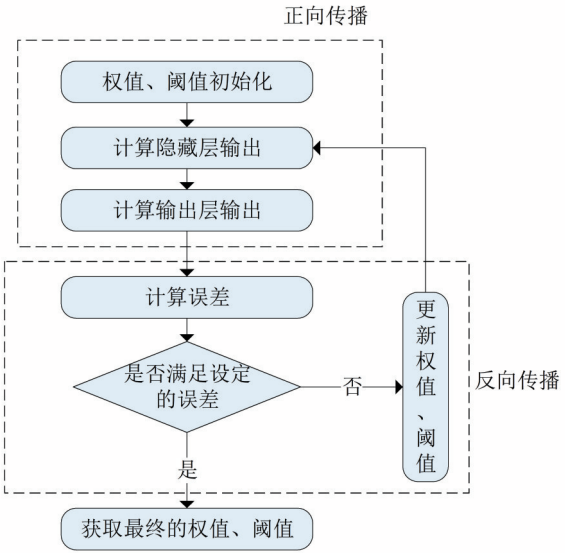


图1 神经网络训练流程图

由文献[12]可知，影响灌区流量测量的主要因素为测点参数以及渠道参数，故本文以渠道糙率n、渠底宽b、水深h、测点流速u、测点位置(x,y)为BP的六个输入因子，以流量Q为输出因子，建立的神经网络模型如图2所示。

2.3 GWO-BP模型

GWO算法是Mirjalil等[13]于2014年提出的一种智能优化算法，该算法模仿灰狼的领导层级和狩猎机制进行数学建模，从而完成优化工作。整个狼群按照自适应度由高到低，分为α、β、δ、ω四个等级。捕食过程中α、β、δ负责追捕猎物，ω跟随其他三者搜索。

灰狼与猎物之间的距离为

$$D = \left| C \cdot X_p(t) - X(t) \right| \tag{4}$$

式中：t表示迭代次数、$X_p(t)$表示猎物的位置、C为[0, 2]的随机数。灰狼位置更新为

$$X(t+1) = X_p(t) - A*D \tag{5}$$

$$A = 2\alpha \cdot r_1 - \alpha \tag{6}$$

式中：α是收敛因子，r_1是[0, 1]的随机数。

α、β、δ三头狼与猎物之间的数学描述如下：

$$\begin{cases} D_\alpha = \left| C_1 \cdot X_\alpha - X \right| \\ D_\beta = \left| C_2 \cdot X_\beta - X \right| \\ D_\delta = \left| C_3 \cdot X_\delta - X \right| \end{cases} \tag{7}$$

ω朝向α、β、δ的位置为

$$\begin{cases} X_1 = \left| X_\alpha - A_1 \cdot D_\alpha \right| \\ X_2 = \left| X_\beta - A_2 \cdot D_\beta \right| \\ X_3 = \left| X_\delta - A_3 \cdot D_\delta \right| \end{cases} \tag{8}$$

ω的最终位置为

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3} \tag{9}$$

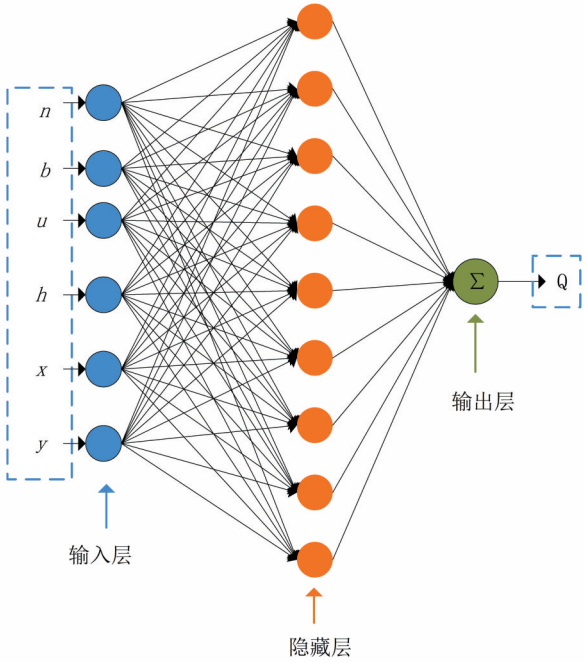


图 2　BP神经网络结构

由于BP神经网络通过误差反向传播，在训练过程中朝着误差梯度下降的方向收敛，因此在实际训练过程中存在收敛速度慢、陷入局部最优解、过拟合、参数选取敏感等问题。为了改善BP神经网络的缺陷，本文使用GWO方法对BP神经网络中较为敏感的权值和阈值进行优化。整个过程中将灰狼的位置作为BP的权重和阈值，通过不断更新灰狼的位置来找到最优的权值和阈值。训练的流程图如图3所示。

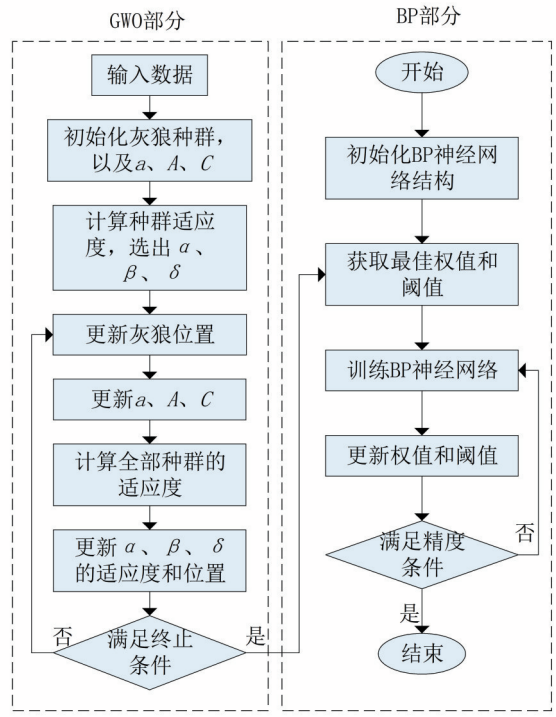


图 3　GWO优化BP神经网络流程

3　结果和讨论

3.1　评价指标

本文对三种流量预测模型进行研究，并得到了测流结果。单一的结果无法说明模型的好坏，故本文采用了决定系数R^2、均方根误差*RMSE*、平均绝对误差*MAE*、平均绝对百分比误差*MAPE*四种评价指标进行判断。R^2越接近1模型性能越好，*RMSE*、*MAE*、*MAPE*越接近于0，模型性能越好。这些指标的计算方式如下：

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(Q_{Pi} - Q_{Mi} \right)^2}{\sum_{i=1}^{n} \left(\overline{Q}_{Pi} - Q_{Mi} \right)^2} \tag{10}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Q_{Pi} - Q_{Mi}\right)^2}{n}} \tag{11}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Q_{Pi} - Q_{Mi}\right| \tag{12}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{Q_{Pi} - Q_{Mi}}{Q_{Mi}}\right| \tag{13}$$

式中：Q_{Pi}是预测的流量；Q_{Mi}是实际测量的流量；n为参与计算的流量个数。

3.2 灌区流量预测结果

本文共收集到147条花海灌区流量实测数据，并将数据集划分为两部分，训练集97条(70%)测试集50条(30%)，并对数据进行归一化处理，消除奇异样本导致的不良影响。相应模型完成训练后，采用三种模型对测试集的50条数据进行预测。

模型一：选取灌区的糙率n、断面面积A、坡度J、水力半径R为输入，利用经验公式模型对花海灌区流量进行预测，预测结果如图4所示。

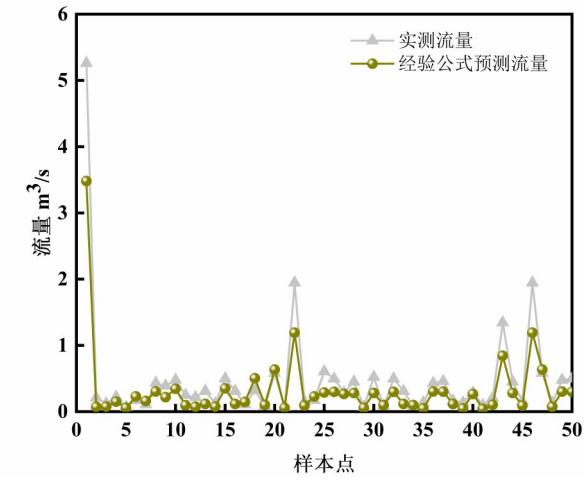


图4 经验公式预测流量结果

模型二：选取灌区的糙率n、渠底宽b、水深h、测点流速v、测点位置(x,y)为输入，利用BP神经网络模型对花海灌区流量进行预测，预测结果如图5所示。

模型三：选取灌区的糙率n、渠底宽b、水深h、测点流速v、测点位置(x,y)为输入，利用GWO-BP神经网络模型对花海灌区流量进行预测，预测结果如图6所示。

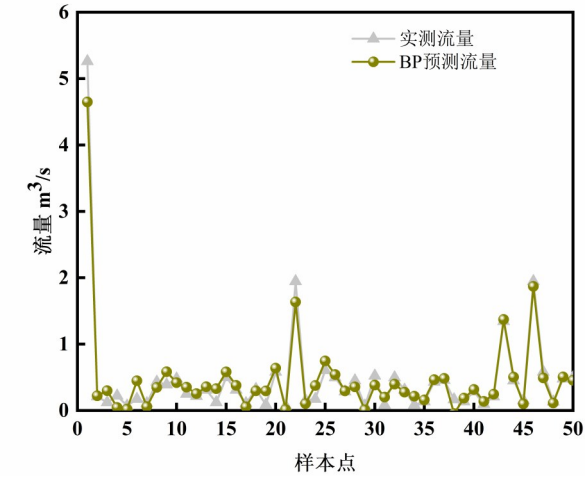


图5 BP神经网络模型预测流量结果

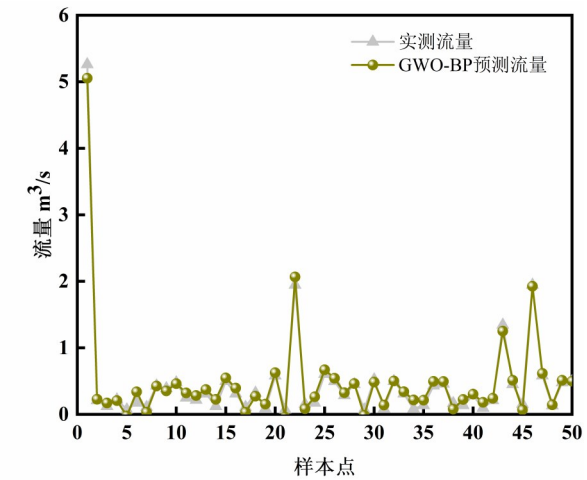


图6 GWO-BP模型预测流量结果

从三种模型的流量预测结果可以看到，经验公式模型的预测效果最差，尤其体现在峰值上，这是因为渠道的糙率以及坡度是影响预测结果的重要参数，然而工程实际中这两个参数往往不容易精确测量，故导致通过该方法得到的测流结果不够理想。BP神经网络模型和GWO-BP模型均能对灌区流量进行有效预测。而GWO-BP模型对峰值的预测效果更加接近真实值，误差分布也更加均匀，详细的评价指标汇总见表1。

表 1　评价指标

评价指标	R^2	*RMSE*	*MAE*	*MAPE*
经验公式	0.843	0.258	0.169	0.0093
BP	0.918	0.113	0.088	0.0008
GWO-BP	0.972	0.069	0.059	0.0005

灌区流量是水资源合理分配的重要依据，因此快速准确地流量预测具有重要意义。从表1可以看出，GWO-BP模型的精度是最高的，使用GWO优化BP后R^2提高0.054，*RMSE*减少0.044，*MAE*减少0.029，*MAPE*减少0.003。最终可以得出结论，使用灰狼优化算法能够有效提高BP神经网络模型的准确性。

4　结语

本文采用灰狼优化算法与单一的BP神经网络模型相结合，构建了GWO-BP模型。将该模型应用在花海灌区流量预测中，并与传统的经验公式模型以及BP神经网络模型进行了比较。就结果而言，在进行灌区流量预测时，本文实验充分证明了使用GWO优化BP模型是有效的，其决定系数R^2高至0.972，均方根误差*RMSE*低至0.059。综上，GWO-BP模型是对灌区流量预测问题研究体系的强力补充，在一定程度上补足了现有研究的不足，同时也能为花海灌区流量预测精度提升提供必要的参考依据。

参考文献：

[1] 李丽丽，陈雯雯，李东风. 经济增长与水资源消耗的脱钩实现路径研究[J]. 人民黄河，2022，44(4)：74-79.

[2] 热孜娅·阿曼，方创琳，赵瑞东. 干旱区城镇化发展与水资源利用耦合协调研究[J]. 人民黄河，2022，44(4)：67-73.

[3] 曹登峰，周叶，潘罗平，等. 基于流速仪法的灯泡贯流式水轮机绝对流量测试方法[J]. 水电能源科学，2021，39(10)：190-193，171.

[4] 周冬生，宗军，蒋东进，等. 雷达流速仪测量精度关键技术研究[J]. 水文，2018，38(5)：67-70.

[5] LIN Y C，HO H C，LEE T A，et al. Application of image technique to obtain surface velocity and bed elevation in open-channel flow [J]. Water，2022，14(12)：1895.

[6] Aggarwal S K，Saini L M，Kumar A. Electricity price forecasting in deregulated markets：A review and evaluation [J]. International Journal of Electrical Power & Energy Systems，2009，31(1)：13-22.

[7] 张振，徐立中，韩华，等. 基于径向基神经网络的明渠流量软测量方法[J]. 仪器仪学报，2011，32(12)：2648-2655.

[8] GHOLAMI A，BONAKDARI H，ZAJI A H，et al. Predicting the velocity field in a 90 open channel bend using a gene expression programming model[J]. Flow Measurement and Instrumentation，2015，46：189-192.

[9] LI J，SALIM R D，ALDLEMY M S，et al. Fiberglass-reinforced polyester composites fatigue prediction using novel data-intelligence model[J]. Arabian Journal for Science and Engineering，2019，44(4)：3343-3356.

[10] GHORBANI M A，KHATIBI R，KARIMI V，et al. Learning from multiple models using artificial intelligence to improve model prediction accuracies：application to river flows [J]. Water Resources Management，2018，32(13)：4201-4215.

[11] 裴斌，赵淑饶，郭银，等. 基于能量方程和曼宁公式的断面平均水深计算[J]. 人民黄河，2019，41(2)：10-13.

[12] 胡云进，郜会彩，耿洛桑，等. 梯形断面明渠流速分布的研究[J]. 浙江大学学报(工学版)，2009，43(6)：1102-1106.

[13] MIRJALILI S，MIRJALILI S M，LEWIs A. Grey wolf optimizer [J]. Advances in Engineering Software，2014，69(3)：46-61.

(下转第120页)

文章编号: 1007-1423(2023)01-0050-05 DOI: 10.3969/j.issn.1007-1423.2023.01.009

基于航班时刻优化的地面等待模型的建立与仿真

李科扬

（中国民用航空飞行学院空中交通管理学院，广汉 618307）

摘要： 地面等待是指机场或者管制区内容量受限时，飞机通过地面等待的策略来缓解机场或者管制区拥挤的方法。为了优化航班的单机场地面等待问题，考虑不同机型，针对不同航空公司航班之间的排序，建立了基于航班时刻优化的地面等待模型。结果表明，当机场发生拥挤时，对航班进行排序，使航班有序进行起降，同时使所有航班的地面等待总体损失最小。

关键词： 地面等待；计算机仿真；机场容量；航班排序

0 引言

近年来，随着民航业的快速发展，我国的民航运输业务量成倍增加，导致机场和航路的拥挤日益严重。为了减少拥挤，通常采用地面等待和空中等待两种方法。由于空中等待策略的安全性和经济性均不如地面等待策略，因此当管制员预测到延误会发生时，通常采用地面等待策略[1-2]。

针对地面等待模型的研究，2000 年胡明华等[3]提出了将网络流的方法应用于解决空中交通流量管理的问题中，建立了地面等待问题的网络流规划模型；2004 年罗喜伶等[4]提出了基于离散事件系统的起降容量受限的地面等待随机型数学模型；2009 年康瑞等[5]在建立地面等待模型时综合考虑了对续航的影响、延误时间和延误耗损，并在航班排序时引入随机型变量，提出了基于分组排序的随机型地面等待模型；2013 年王莉莉等[6]给出了考虑扇区容量限制的多机场地面等待问题的动态模型,并设计了一种以航班优先级别为核心的多机场地面等待启发式优化算法；2020 年张玉州等[7]针对多跑道机场的跑道分配问题，提出了一种多跑道进离港地面等待问题优化模型。

本文针对单个机场容量受限问题，综合考虑航班机型，航班所属航空公司等因素，建立了基于航班时刻优化的地面等待模型。

1 模型建立

1.1 确定型单机场地面等待问题归纳

在对单机场地面等待问题建立数学模型时，我们需要对实际问题做如下简化[8-12]：

（1）在单机场地面等待问题中，我们认为起飞机场 X 的容量，航班飞行穿越的管制区的容量是无限的，即在整个系统中，只有降落机场 Y 的容量是受到限制的；

（2）在确定型问题中，任一确定的时间段内降落机场 Y 的容量应当是确定的；

（3）在建立数学模型时，应保证存在某一时间段，时间段内机场的容量小于这个时间段内计划降落飞机的数量，即保证机场内的拥挤是真实存在的；

（4）在建立数学模型时，假定所有的航班只在计划到达时刻或者计划到达时刻之后到达；

（5）在建立模型时，保证同一航空公司内部的航班次序可以自由换序，以确保延迟损失最小；在对不同航空公司的航班进行排序时，考虑公平性原则，利用先到先服务准则对航班进行排序。

收稿日期： 2022-07-16 **修稿日期：** 2022-09-01

作者简介： 李科扬(1996—),男,河南洛阳人,硕士研究生,研究方向为空中交通流量管理

通过上述简化，可以将实际问题简化为数学问题，通过计算机仿真合理地调整航班计划，使得总的地面等待损失最小。

1.2　建立数学模型

给定时间间隔[0,T]，将时间间隔等分为m个时间段T_1、T_2、…、T_m，时间段T_j(j=1,2,…,m)对应的目的机场Y的容量为K_j(j=1,2,…,m),在[0,T]时间内共有n架航班降落，分别为F_1、F_2、…、F_n，航班F_i的计划降落时间为T_j，设航班F_i实际降落时间为T_p(p=1,2,…,m)，$p\geq i$(当p=i时，代表航班F_i准点到达)，设航班F_i单位时间地面等待损失系数为C_i^g，航班F_i的地面等待时间为(T_p−T_i)，航班F_i地面等待的损失为C_i^g*(T_p−T_i)[13]。

设决策变量为x_{ij}(i=1,2,…,n；j=1,2,…,m)，当x_{ij}=1时，表示航班F_i在时间T_j内降落；当x_{ij}=0时，表示航班F_i不在时间T_j内降落。设延迟成本为c_{ij}，c_{ij}=C_i^g*(T_p−T_i)[14]。

定义$Airlines_{a,b}$，其中a代表第a家航空公司，b代表该家航空公司的第b架航班；$Flight_c$代表对所有的航班进行排序，c代表航班的数量[15-16]。

目标函数：

$$w = \min \sum_{i=1}^{n} \sum_{j=T_i}^{m+1} c_{ij} * x_{ij} \tag{1}$$

约束条件：

$$\sum_{j=1}^{T_i-1} x_{ij} = 0, \forall i \in \{1,2,\cdots,n\} \tag{2}$$

$$\sum_{j=T_i}^{m+1} x_{ij} = 0, \forall i \in \{1,2,\cdots,n\} \tag{3}$$

$$\sum_{i=1}^{N} x_{ij} \leqslant K_j, \forall j \in \{1,2,\cdots,m\} \tag{4}$$

$$x_{ij} \in \{0,1\} \tag{5}$$

$$Flight_c = Airlines_{a,b}(Airlines_{ij}[w] < Airlines_{ij+1}[w]) \tag{6}$$

$$Flight_c = Airlines_{a,b}(Airlines_{a,b}[T] < Airlines_{b,b+1}[T]) \tag{7}$$

其中，式(1)为成本函数表达式；约束(2)，(3)代表航班只能在一个时间段内降落且实际降落时间不早于计划降落时间；约束(4)表示每个时间段内降落的航班数量应小于等于每个时间段对应的机场容量；式(5)代表整数约束，表明航班只有未降落和已降落两种状态；式(6)表示在依据损失最小原则对航班进行排序；式(7)表示在对不同航空公司的航班进行排序时，按照先到先服务原则进行排序。

2　仿真计算与分析

采用Python软件进行数值仿真。程序流程图如图1所示。

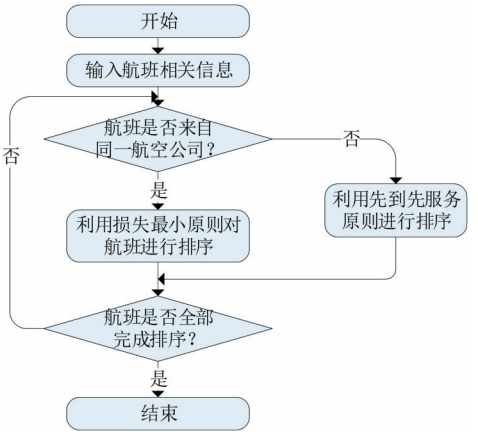


图1　仿真程序流程图

采用乌鲁木齐地窝堡机场某日13：00—14：00航班数据对模型的有效性进行验证，将时间区间T等分为12份，每5分钟为1个时间段，每个时间段的容量均设置为1。13：00—14：00时段内共有20架航班，航班相关情况如表1所示。

表1　航班相关信息

编号	航班	航空公司	机型	计划时刻
1	库尔勒	CCA	M	13：00
2	沈阳	CSN	H	13：05
3	北京	CES	H	13：10
4	天津	CCA	M	13：10
5	北京	CSN	H	13：15
6	广州	CES	H	13：15
7	南京	CSN	H	13：20
8	太原	CCA	H	13：25
9	石家庄	CES	H	13：25
10	榆林	CCA	M	13：30
11	兰州	CSN	H	13：30
12	洛阳	CSN	H	13：35
13	博乐	CCA	H	13：40
14	上海	CES	H	13：45
15	昆明	CSN	M	13：45
16	郑州	CCA	H	13：45
17	武汉	CCA	H	13：50
18	长沙	CES	M	13：50
19	厦门	CES	M	13：55
20	福州	CSN	H	13：55

在仿真时，将所有的航班划分为中国国际航空(CCA)、中国南方航空(CSN)和中国东方航空(CES)三家公司。飞机按照尾流大小可以分为重型机、中型机和轻型机三种，本文只涉及重型机和中型机。

各类机型的延误运营成本如表2所示。

表2 各类机型延误运营成本

机型	代表符号	尾流类型	每分钟延误运营成本/元
重型机	H	重型尾流	4167
中型机	M	中型尾流	2916
轻型机	L	轻型尾流	208

将仿真计算的20个航班按所属航空公司进行排序，分别得到三个航空公司各自的航班调整时刻、延迟时间和延迟损失等参数，如表3～表5所示。

表3 中国国际航空公司航班运行情况汇总

航班	机型	计划时刻	调整时刻	延迟时间/分钟	延迟损失/元	航班排序
库尔勒	M	13：00	13：00	0	0	1
天津	M	13：10	14：15	65	3120	6
太原	H	13：25	13：35	10	690	2
榆林	M	13：30	14：25	55	2640	7
博乐	H	13：40	14：10	30	2070	5
郑州	H	13：45	13：50	5	345	3
武汉	H	13：50	14：00	10	690	4
汇总	—	—	—	175	9555	—

表4 中国南方航空公司航班运行情况汇总

航班	机型	计划时刻	调整时刻	延迟时间/分钟	延迟损失/元	航班排序
沈阳	H	13：05	13：05	0	0	1
北京	H	13：15	13：20	5	345	2
南京	H	13：20	13：25	5	345	3
兰州	H	13：30	13：40	10	690	4
洛阳	H	13：35	14：05	30	2070	6
昆明	M	13：45	14：30	45	2160	7
福州	H	13：55	13：55	0	0	5
汇总	—	—	—	95	5610	—

表5 中国东方航空公司运行情况汇总

航班	机型	计划时刻	调整时刻	延迟时间/分钟	延迟损失/元	航班排序
北京	H	13：10	13：10	0	0	1
广州	H	13：15	13：15	0	0	2
石家庄	H	13：25	13：30	5	345	3
上海	H	13：45	13：45	0	0	4
长沙	M	13：50	14：35	45	2160	6
厦门	M	13：55	14：20	25	1200	5
汇总	—	—	—	75	3705	—

表3为隶属于国航的7架航班的运行情况，其中有3架中型机，4架重型机，总的延迟时间和延迟损失为175分钟和9555元，7架航班中共有5架航班调整了降落次序；表4为隶属于南航的7架航班的运行情况，其中有1架中型机，6架重型机，总的延迟时间和延迟损失为95分钟和5610元，7架航班中有3架航班调整了降落次序；表5为隶属于东航的6架航班的运行情况，其中有2架中型机，4架重型机，总的延迟时间和延迟损失为75分钟和3705元，6架航班中有2架航班调整了降落次序。

对比表3、表4和表5，三家航空公司之中，国航的延迟时间最长，延迟损失也最多；东航的延迟时间和延迟损失均为最小，一方面因为国航的航班数比东航的航班数多，另一方面国航的中型机数量要多于东航的中型机数量，由于中型机的延误运营成本低于重型机，因此在机场容量紧张时，会优先安排重型机进行降落，所以中型机较多的国航的延迟时间和延迟损失会比较大。各个航空公司运行情况对比如表6所示。

表6 各家航空公司运行情况对比

航空公司	航班架次	平均延迟时间/分钟	平均延迟损失/元	航班调整次数
CCA	7	25.0	1365	5
CSN	7	13.6	801	3
CES	6	12.5	607	1

在13：00—14：00时间段内，机场小时容量为12架，而在13：00—14：00内，共有20架

航班计划降落，所以20架航班中，至少有8架航班会发生延误，而仿真结果显示20架航班中有14架航班发生了延误，这是由于为了追求总体航班的最小延误损失，在进行航班排序时将某些原本不需要延误的航班进行了延误，此外，由于模型本身的约束条件，使得航班只能在计划到达时刻及以后的时刻到达，这就可能导致机场某些空闲时段的容量未被充分利用。航空公司的延迟损失与该航空公司的中型机所占比例有关，中型机占比越高，航空公司的延迟损失越大。

3　结语

由仿真结果可得：当某一时间段机场容量小于该时间段飞机计划到达数量时(拥挤发生时)，模型可以选择该时间段的某些航班延迟降落、调整航班降落次序，使得系统总延迟损失最小。

考虑到实际运行时，同一航空公司内部航班相较于不同航空公司的航班，航班的协调难度较低、效率较高的情况，模型实现了对航班按照航空公司进行分组排序，使得同一航空公司内部按照延误损失最小的原则进行航班排序；不同航空公司之间按照先到先服务的原则进行排序，提高了航班排序的公平性；同时按照航空公司进行分组排序，排列出的航班时刻表也更易被机组人员接受。

虽然该模型考虑的情况比较简单，但它是研究其它地面等待模型的基础。在此模型的基础上可以进一步研究机场容量不确定、多个机场及管制区容量同时受限的随机型多元受限地面等待模型。

参考文献：

[1] ZHONG H, GUAN W, ZHANG W, et al. A bi-objective integer programming model for partly-restricted flight departure scheduling[J]. PLoS ONE, 2018, 13(5): e0196146.

[2] 何爱平，张建伟，韩云祥. 基于期望Sarsa的进港航班排序模型研究[J]. 现代计算机，2021(7)：55-59，68.

[3] 胡明华，李丹阳，李顺才. 空中交通地面等待问题的网络流规划模型[J]. 东南大学学报(自然科学版)，2000(3)：104-109.

[4] 罗喜伶，张其善. 起降容量受限的地面等待模型研究[J]. 北京航空航天大学学报，2004(2)：127-130.

[5] 康瑞，彭莉娟，杨凯. 基于分组排序的随机性地面等待模型[J]. 四川大学学报，2009，41(6)：147-152.

[6] 王莉莉，顾秋丽，张兆宁. 多机场地面等待动态建模及其算法研究[J]. 科学技术与工程，2013，13(20)：5878-5882.

[7] 张玉州，张子为，江克勤. 多跑道进离港地面等待问题建模及协同优化[J]. 南京大学学报(自然科学版)，2020，56(1)：132-141.

[8] 朱晶波. 多元受限的航班时刻优化模型和算法研究[D]. 南京：南京航空航天大大学，2001.

[9] 徐肖豪，李雄. 航班地面等待模型中的延误成本分析与仿真[J]. 南京航空航天大学学报，2006(1)：115-120.

[10] 杨光. 航班地面等待模型研究[J]. 科技资讯，2012(23)：239，241.

[11] 高海超. 多元受限流量管理模型与仿真[J]. 指挥信息系统与技术，2014(05)：39-43.

[12] 凌枫. 空中交通流量管理中地面等待策略问题研究[J]. 科技展望，2016(3)：176.

[13] GUSTAF S, JOHN-PAUL C. Scheduling of airport runway operations using stochastic branch and bound methods[J]. Transportation Research Part C, 2014, 45：119-137.

[14] 潘贺. 单跑道机场非紧急状态下进离港航班调度问题研究[D]. 长春：吉林大学，2020.

[15] 马英钧，孙晓娜，赵东方. 基于改进粒子群算法的航班降落调度问题研究[J]. 计算机应用研究，2015，32(7)：2035-2038.

[16] 谢则辉，焦潇冰. 基于可变优先级的航班时隙排序[J]. 现代计算机，2013，30：16-20.

(下转第73页)

文章编号: 1007-1423(2023)01-0054-05 DOI: 10.3969/j.issn.1007-1423.2023.01.010

表格结构识别网络FcaTGRNet

赵玲俐，余艳梅*，陶青川

（四川大学电子信息学院，成都 610065）

摘要：如今重要的数据大都通过表格直接展示，在分析数据时，除数据和文字信息外，表格结构信息也很重要，要想更加准确、便捷地分析表格数据，能否自动准确地识别表格结构至关重要。现在有很多识别表格结构的方法，对表格单元格检测的准确率也都比较理想，但是很少有方法在特征提取方面做出改进。为了证明图像特征提取在表格结构识别过程中的重要性，针对这一问题在TGRNet的基础上引入了多频谱注意力机制，以便更好地提取图像多个频谱上的特征，使得到的特征更加全面。在公开数据集上的实验结果显示，所做改进较原方法在表格结构识别的单元格空间位置和逻辑位置检测的准确率有所提升。

关键词：表格；结构识别；多频谱；深度学习

0 引言

如今，社会迅速发展，是一个大数据的时代，很多文件中的重要数据通常以表格的形式出现，要更加准确分析这些数据，除了提取表格的数据还需要提取表格的结构信息，通过自动识别表格结构，不仅可以提高数据分析的准确率还可以提高效率。

现在有很多表格结构识别的方法，早期有使用传统图像处理的方法，通过对图像进行膨胀腐蚀、找连通区域和检测表格的横线竖线，得到表格的交点，从而提取表格的结构，但是这种传统的方法准确度不高，特别是不能适用于无线表格。由于深度学习的兴起，现在的表格结构识别算法大多基于深度学习，如DeepDeSRT[1]利用语义分割方法来识别表格单元格；文献[2]通过构建一个由450K的表图像和相应的LaTeX源组成的大数据集，进一步实现了端到端的表识别；文献[3]提出的模型先分别识别表格的横线和竖线，最后再经过合并单元模块将其整合到一起；文献[4]提出了一种端到端的模型，用于表格的检测和识别，还用到跨数据的迁移学习；文献[5]中的DeepTabStR利用可变形卷积网络来分析文档图像中的表格结构；文献[6]中的CascadeTabNet提出了一种端到端的级联结构来检测表格以及表格中的每个单元格；文献[7]提出了一种基于图像的表结构识别的深度语义分割模型，其中标签是行分隔符、列分隔符、单元格内容和背景；文献[8]的TabStruct-Net使用了两个阶段，一个是自上而下阶段(分解阶段)，这是基于Mask-RCNN(修改的FPN)的单元检测网络，另一个是自下而上阶段(合成阶段)，其从单元检测网络获取信息(自上而下阶段)，以及它们使用邻接矩阵的行-列关联，并重建整个表格；文献[9]提出了局部和全局金字塔掩模对齐的框架，该框架在局部和全局特征映射中都采用了软金字塔掩模学习机制；文献[10]主要针对HTML的表识别，包括表结构识别、文本行检测、文本行识别和方框分配；文献[11]提出的TabCellNet通过检测单元格而不是行和列来改进表结构检测，以更好

收稿日期：2022-08-13 **修稿日期：**2022-09-07

作者简介：赵玲俐(1997—)，女，四川南充人，硕士，硕士研究生，研究方向为计算机视觉；*通信作者：余艳梅(1975—)，女，四川广安人，博士，副教授，硕导，研究方向为图像处理，E-mail:yuyanmei@scu.edu.cn；陶青川(1972—)，男，四川南充人，博士，副教授，硕导，研究方向为计算机视觉

地提高异构表结构的泛化能力；文献[12]的TGRNet具有单元检测分支和单元逻辑位置分支这两个主要分支，可以共同预测不同单元的空间位置和逻辑位置；文献[13]提出了一种新的表结构分解方法，它直接检测表单元格，并且创建了一个带有单元格级注释的基于图像的大型表格数据集。上述方法大都没有在图像特征提取方面做出改进，我们知道图像的特征对最后表格结构识别准确率至关重要，所以本文在图像特征提取方面做出了改进，引入了多频谱注意力机制，并且进行了相应的实验。

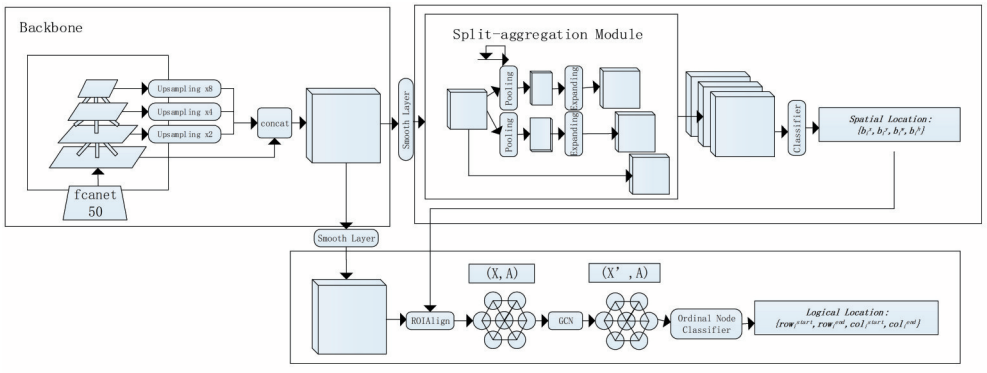
本文主要从以下几个方面展开：第一部分主要介绍本文需要用到的方法基础；第二部分介绍本文提出的新方法，FcaTGRNet的网络结构和评价指标，以及实验结果的对比；最后就是对本文的总结和展望。

1　论文涉及方法介绍

本部分主要对用到的一些方法进行介绍，主要包括TGRNet[12]方法和FcaNet(frequency channel attention networks)[14]。

1.1　TGRNet

TGRNet是由Xue等[12]于2021年发布的一种表格结构的提取方法，具体网络结构可以参考第二部分的图1，只是图中加粗黑色框内主干网络TGRNet[12]用的ResNet50。可以看到，此方法分为两部分：一个是通过目标检测网络提取表格图像的每个单元格的位置，这里称为表格的空间位置；另一个是根据图神经网络和得到的单元格空间位置获取单元格的相对位置，这里称作逻辑位置。根据逻辑位置，就可以将表格图像转为电子文档进行保存分析。



(与原网络TGRNet区别在于加粗黑色框内的主干网络，原文使用的是ResNet50)

图1　FcaTGRNet网络结构图

1.2　FcaNet

FcaNet[14]是浙江大学李玺团队在2021年提出的一种基于频谱的注意力机制，它弥补了现有注意力机制的一些不足，使模型能够获取图像不同频率分量的信息。全局平均池化(global average pooling，GAP)其实是二维离散傅里叶变换的低频分量，使用这种方法来提取图像信息，只使用了图像的低频分量，但是其他频率也包含了一些有用的信息，为了更好地利用图像信息，就需要使用文献[14]中提出的多频谱通道注意力网络FcaNet。

FcaNet[14]的基本原理是：首先，将输入的特征图沿着通道划分为多块，记为$[X^0, X^1, \cdots, X^{n-1}]$，其中每个$X^i \in \mathbf{R}^{C' \times W \times H}$，$i \in \{0, 1, 2, \cdots, n-1\}$，$C' = C/n$，每个块分配一个二维的DCT分量，那么每一块的输出结果如下：

$$Freq^i = 2DDCT^{u,v}(X^i) = \sum_{h=0}^{H-1}\sum_{w=0}^{W-1} X^i_{:,h,w} B^{u,v}_{h,w}$$
$$\text{s.t. } i \in \{0, 1, 2, \cdots, n-1\} \quad (1)$$

其中：$[u, v]$表示2DDCT的分量下标，表示对每一块采用不同的频率分量，因此式(2)得到最终的输出$Freq \in \mathbf{R}^C$，也就是得到多谱向量，然后再将这个向量送入通道注意力常用的全连接层中进行学习，得到注意力图

$$Freq = \text{cat}\left([Freq^0, Freq^1, Freq^2, \cdots, Freq^{n-1}]\right) \quad (2)$$

$$ms\_att = \text{sigmod}(\text{fc}(Freq)) \tag{3}$$

其中：cat(*)表示对输出的频谱进行通道拼接；fc(*)表示经过全连接网络。

2 FcaTGRNet表格结构识别算法

本节主要介绍FcaTGRNet的网络结构、评价指标以及实验结果。

2.1 本文方法介绍

本文是对文献[12]中提出的TGRNet进行改进，改进后的新网络命名为FcaTGRNet。主要改进是：在主干网络中添加了文献[14]所提出的多频谱通道注意力模块(FcaNet)，使得提取的特征图可以在多频通道下进行处理，网络结构图如图1所示，加粗黑色框内，原文使用的主干网络为ResNet50，而经过改进，本文使用了文献[14]提供的预训练模型Fcanet50，此模型是在ResNet50基础上进行修改，每个Bottleneck中均添加了fcanet模块成为FcaBottleneck，以便更好地提取图像多个频带的特征信息，添加的具体位置如图2所示。方法中用到的损失函数与原文保持一致，有回归损失、交叉熵损失以及focal损失[15]。

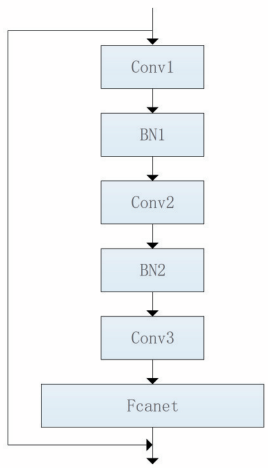


图2 Fcanet50的Bottleneck网络结构图

2.2 评价指标

对于本文的评价指标，也主要沿用了文献[12]中使用到的精确率(precision，简称为P)、召回率(recall，简称为R)、平均精确率(hmean，简称为H)以及预测单元格逻辑位置时用到的开始行A_{rowSt}和结束行A_{rowEd}，开始列A_{colSt}和结束列A_{colEd}，以及综合评价逻辑位置预测情况的指标：单元格四个点位置均正确的指标A_{all}。

2.2.1 预测空间位置使用的指标

空间位置在本文是指表格中每个单元格在图像中的位置，这里使用的是目标检测算法来检测表格单元格，所以指标有三个，分别是P、R和H。在目标检测结果中可能出现四种情况：样本为正样本，预测为正样本(TP)；样本为正样本，预测为负样本(FN)；样本为负样本，预测为负样本(TN)；样本为负样本，预测为正样本(FP)。假设这里总样本个数为N，可以得到P、R和H的公式如下：

$$P = \frac{TP + TN}{N} \tag{4}$$

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

$$H = \frac{2(P \times R)}{P + R} \tag{6}$$

其中：P表示最后生成的目标框中正确的个数与真实目标框数的比值；R表示最后生成的目标框中正确的个数与生成的目标框总数的比值；H则是结合P和R的一个综合性的指标。

2.2.2 预测逻辑位置用到的指标

逻辑位置评价指标包括A_{rowSt}、A_{rowEd}、A_{colSt}和A_{colEd}，以及综合评价这四个指标的A_{all}，前四个指标分别代表单元格的开始行、结束行、开始列以及结束列的预测正确率，最后一个指标代表单元格的位置的四个指标均预测正确的概率，由于最后的文本框位置需要四个指标均预测正确，所以A_{all}才是逻辑位置预测准确率的最终指标。这五个指标计算方法相似，均是将检测出来的值的正确个数除以前一步目标检测部分的单元格的正确个数，如公式(7)所示。

$$A = \frac{N_l}{N_s} \tag{7}$$

其中：A为相应的逻辑位置预测准确率；N_l为逻辑位置预测正确个数；N_s为单元格空间位置预测正确个数。

2.3 实验结果

2.3.1 实验环境及参数设置

本文代码使用的是pytorch框架，训练环境均为：一块Tesla P40 GPU，显存24 GB；12核CPU，型号为Intel（R）Xeon（R）CPU E-2650 v4 @2.20GHZ，内存60 GB。实验主要在ICDAR13-Table数据集和TableGraph-24K数据集上进行。在TableGraph-24K数据集训练时学习率设为0.001，在原文预训练模型的基础上训练了30个epoch；ICDAR13-Table数据集训练时学习率为0.0005，在预训练模型的基础上训练了700个epoch。这里，ICDAR13-Table数据集用的预训练模型为在TableGraph-24K数据集上训练好的模型。

2.3.2 实验结果

在TableGraph-24K数据集和ICDAR13-Table数据集的实验结果分别如表1和表2所示，表中各指标均是数值越大效果越好。

表1　TableGraph-24K数据集实验结果

评价指标	R	P	H	A_{rowSt}	A_{rowEd}	A_{colSt}	A_{colEd}	A_{all}
TGRNet	0.898	0.881	0.889	**0.517**	**0.527**	0.756	0.725	0.363
Ours	**0.935**	**0.887**	**0.910**	0.515	0.525	**0.761**	**0.736**	**0.374**

表2　ICDAR13-Table数据集实验结果

评价指标	R	P	H	A_{rowSt}	A_{rowEd}	A_{colSt}	A_{colEd}	A_{all}
TGRNet	0.916	0.895	0.906	0.917	0.916	**0.919**	**0.923**	0.832
Ours	**0.942**	**0.910**	**0.926**	**0.925**	**0.926**	0.911	0.916	**0.845**

表1、表2分别在TableGraph-24K和ICDAR13-Table两个数据集上进行了验证，从表中可以看到，在空间位置的预测上，指标P、R和H均有所提升，在逻辑位置的预测A_{all}也有超过百分之一的提升。所以本文改进的方法FcaTGRNet无论是在单元格空间位置检测部分还是逻辑位置检测部分均明显优于原始TGRNet方法，可见本文提出添加多频谱注意力机制来提取图像特征的改进确实有一定的效果。

3 结语

本文在TGRNet[12]的基础上添加了多频谱通道注意力机制Fcanet[14]，改进形成的新网络FcaTGRNet在检测精度上有一定的提升，FcaTGRNet主要是针对主干网络进行的改进，在ICDAR13-Table和TableGraph-24K两个数据集上进行了训练和测试，本文的方法在空间位置和逻辑位置上的准确率均有提升，取得了比原文更好的效果。在未来，可以继续在逻辑位置网络部分进行改进，使得表格间相对位置的预测准确率进一步提升。

参考文献：

[1] SCHREIBER S, AGNE S, WOLF I, et al. DeepDeSRT: deep learning for detection and structure recognition of tables in document images [C] //2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Kyoto, Japan, 2017: 1162-1167.

[2] DENG Y, ROSENBERG D, MANN G, Challenges in end-to-end neural scientific table recognition [C] // 2019 International Conference on Document Analysis and Recognition (ICDAR), 2019: 894-901.

[3] TENSMEYER C, MORARIU V I, PRICE B, et al. Deep splitting and merging for table structure decomposition [C] //2019 International Conference on Document Analysis and Recognition (ICDAR), 2019: 114-121.

[4] PALIWAL S, RAHUL R, SHARMA M, et al. TableNet: deep learning model for end-to-end table detection and tabular data extraction from scanned document images [C] //2019 International Conference on Document Analysis and Recognition (ICDAR), 2019: 128-133.

[5] SIDDIQUI S A, FATCH I A, RIZVI S T R, et al. DeepTabStR: deep learning based table structure recognition [C] //2019 International Conference on Document Analysis and Recognition (ICDAR), 2019: 1403-1409.

[6] PRASAD D, GADPAL A, KAPADNI K, et al. CascadeTabNet: an approach for end to end table detec-

tion and structure recognition from image-based documents[C]//Proceedings of the IEEE International Conference on Computer Vision (CVPR), 2020:2439-2447.

[7] MEUNIER J, DEIEAN H. Table rows segmentation [C]//2019 International Conference on Document Analysis and Recognition (ICDAR). Sydney, NSW, Australia, 2019.

[8] RAJA S, MONDAL A, JAWAHAR C V. Table structure recognition using top-down and bottom-up [C]//ECCV 2020. ECCV 2020. Lecture Notes in Computer Science(), 2020, 12373.

[9] QIAO L, LI, Z S, CHENG Z Z, ZHANG P, et al. LGPMA: complicated table structure recognition with local and global pyramid mask alignment[C]// SICDAR 2021. ICDAR 2021. Lecture Notes in Computer Science(), 2021, 12821.

[10] YE J Q, QI X B, HE Y L, et al. PingAn-VCGroup's solution for ICDAR 2021 competition on scientific literature parsing task B: table recognition to HTML [C]//2021 International Conference on Document Analysis and Recognition (ICDAR), 2021.

[11] JIANG J C, SIMSEK M, KANTARCI B, et al. (2021). Deep learning-based tabular cell structure detection[J]. Neurocomputing, 2021, 440:12-23.

[12] XUE W, YU B, WANG W, et al. TGRNet: a table graph reconstruction network for table structure recognition[C]//2021 IEEE/CVF International Conference on Computer Vision(ICCV), 2021:1275-1284.

[13] WEI D, LU H, ZHOU Y, et al. Image-based table cell detection: a novel table structure decomposition method with new dataset[C]//2020 25th International Conference on Pattern Recognition (ICPR), 2021:1-7.

[14] QIN Z, ZHANG P, WU F, et al. FcaNet: frequency channel attention networks[C]//2021 IEEE/CVF International Conference on Computer Vision(ICCV), 2021:763-772.

[15] YI L, GOYAL P, DOLL P, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision (CVPR), 2017:2980-2988.

Table structure recognition network FcaTGRNet

Zhao Lingli, Yu Yanmei*, Tao Qingchuan

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

Abstract: Nowadays, most important data are directly displayed through tables. In the analysis of data, in addition to data and text information, table structure information is also very important. In order to analyze table data more accurately and conveniently, whether the table structure can be automatically and accurately identified is very important. At present, there are many methods to recognize table structure, and the accuracy of table cell detection is ideal, but few methods make improvements in feature extraction. In order to prove the importance of image feature extraction in the process of table structure recognition, this paper introduces a Frequency channel attention Networks(FcaNet) based on TGRNet to solve this problem, so as to better extract the features in multiple spectrum of the image and make the features more comprehensive. Experimental results on open datasets show that the proposed method has improved the accuracy of spatial location and logical location detection in table structure recognition compared with the original method.

Keywords: table; structure recognition; multi-frequency; deep-learning

文章编号: 1007-1423(2023)01-0059-04 DOI: 10.3969/j.issn.1007-1423.2023.01.011

用于虚拟现实/增强现实人机交互的实时手势识别方法的研究

柯 健，张 量

（苏州市职业大学计算机工程学院，苏州 215104）

摘要： 虚拟现实/增强现实（简称VR/AR）中的人机交互方式是一个重要的研究方向，将手势与VR/AR中的操作控制相结合，提出一种利用手势实现VR/AR操作的人机交互方式。在实验中使用MediaPipe框架采集实时手势图像，提取手部关键点，设计手势关键点算法进行实时手势识别，结果显示该方法具有良好的性能，为VR/AR提供了一种更为自然的人机交互方式。

关键词： 手势识别；人机交互；手部关键点检测

0 引言

虚拟现实技术的三大基本特征是沉浸感、交互性和想象力，其中交互性是用户与虚拟世界通过特定的交流手段实现信息交互的方式。虚拟现实技术的实现离不开硬件设备的支持，虚拟现实设备主要有两类，第一类是移动头显设备，如Gear VR等，需要插入手机使用，第二类是外接式头显设备，如HTC Vive等，需要外接电脑使用，这两类虚拟现实设备的交互方式基本都是通过操作手柄实现。增强现实技术是一种将虚拟信息与真实世界巧妙融合的技术，目前大多数是通过手机摄像头拍摄现实世界，然后与虚拟信息相融合，从而实现对真实世界的增强，它的交互方式是通过操作手机屏幕实现的。

实现人与虚拟场景的交互方式主要有三种：一是虚拟交互，通过虚拟场景中的虚拟按钮进行控制；二是物理交互，利用控制手柄等物理设备，实现与虚拟场景的交互；三是直接交互，通过数据手套或者手势识别，做出各种不同的动作指令，从而实现对虚拟场景中物体的控制。在VR/AR中人机交互方式大多是通过虚拟交互或者物理交互实现的，随着VR/AR设备以及人机交互的不断发展，直接交互尤其是手势交互成为VR/AR中更为友好，更为自然的交互方式。

1 相关工作

手势识别是手势交互的前提条件，目前手势识别主要由两种实现方式，一是基于传感器的手势识别方法，文献[1]提出了一种基于加速度计和陀螺仪传感器数据的深度学习框架实现手势识别。文献[2]以手势识别和虚拟现实人机交互为研究对象，提出并设计了一种新型的，用于辅助VR设备进行人机交互的原型系统。文献[3]基于Leap Motion对模板匹配的手势识别加以改进。 文献[4]提出了基于Leap Motion手势识别和最邻近算法，建立了三维手势交互系统。文献[5]提出了基于Leap Motion的动态手势识别方法。文献[6]设计了一种VR游戏手势运动识别装置。文献[7]提出的基于Kinect视觉功能对体感手势目标识别。文献[8]提出了基于Kinect骨骼数据的手势识别。基于数据手套的手势识别方法，由于数据手套价格昂贵，容易损坏，不容易得到推广，基于Kinect骨骼数据的手势识别方法，由于Kinect只提供人体25个关节点，每只手的手部只提供4个关节点，手部关节点数量较少，影响了对手势的精确识别。

二是基于视觉的手势识别，使用普通摄像头采集数据，成本较低，利用机器学习或神经网络进行手势图像识别。文献[9]提出了基于关

收稿日期： 2022-07-10 **修稿日期：** 2022-09-04

作者简介： 柯健（1974—），男，江苏常熟人，硕士，研究方向为深度学习；张量（1982—），男，江苏苏州人，博士，研究方向为人工智能、进化计算

键点的残差全连接网络动态手势识别方法。文献[10]提出了基于改进的YOLOv3实现手势识别的人机交互方法。文献[11]提出了融合注意力机制的YOLOv4-Tiny算法，采用基础手势结合手势状态变化实现人机交互。文献[12]提出了基于深度学习的手势识别算法，利用卷积神经网络对关节特征图分类，达到手势识别的目的。基于视觉的深度学习手势识别方法是目前研究的热点，但是使用深度学习需要提供大量的手势样本数据用于训练，当只有少量样本时，深度学习容易出现过拟合。

在计算机科学和人机交互领域中，手势识别是通过数学算法来识别人类手势的一个议题。用户可以使用简单的手势来控制或与设备交互，让计算机理解人的意图。在AR中，通过感知手部形状和动作对于改善用户体验是至关重要的。在手势识别领域常用的一种方法是模板匹配，通过对两个手势样本进行欧式距离计算，得到两个手势样本之间的相似度，根据相似度加以判断手势。

2 手势识别方法

本文针对传统的VR/AR人机交互方式，提出了一种基于MediaPipe平台的手势识别[13]，用于VR/AR操作控制方案，计算手部关键点之间的位置、距离，从而进行手势识别决策，它是一种实时的不需要大量样本训练的手势识别方法。如图1所示。

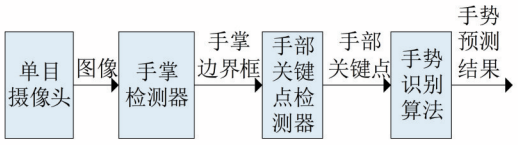


图1 手势识别流程

2.1 手势设计

手势识别的首要工作是先设计好能被识别的手势，手势的设计[14]要遵循几个原则：首先，在用户使用方面，手势的设计要方便简单，符合用户习惯，便于用户理解和掌握；其次，在交互过程方面，要求手势的设计保证流畅和降低操作疲劳；最后，在系统实现方面，由于手势识别的准确度与传统的交互方式存在一定的差距，识别过程容易出现误判，要求手势的设计功能交叉少，风格统一，使系统存在一定的容错性和可靠性。总之，利用手势交互的目的是使人机交互的过程更加自然方便。在单目摄像头下，得到的是二维图像信息，很难获得深度方向的信息，因此在设计手势的时候，不要使用手势离摄像头的远近距离作为判断手势的标准。可以根据手指关节的弯曲程度来决定每根手指是弯曲还是伸直的状态，每个手指的关节位置可以作为不同手势语义的设计标准。本文定义了4组基本的静态手势和标签，如表1所示。

表1 手势定义

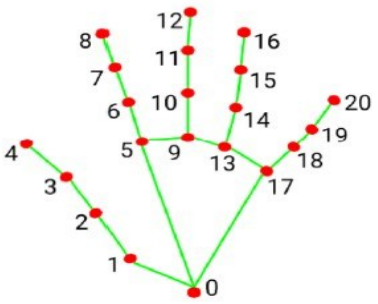
编号	标签	语义	描述
1	选择	选择物体	食指伸直，四指弯曲
2	确定	确定操作	食指、中指伸直，三指弯曲
3	抓取	抓取物体	五指弯曲
4	松开	放下物体	五指伸直

2.2 手掌检测器

检测人手是一个非常复杂的任务，模型必须适用于各种尺寸的手，而且需要检测被遮挡或者自遮挡的手，手的检测与人脸检测有所不同，人脸具有眼睛、嘴巴等高对比度的轮廓，而人手没有类似的特征，只通过视觉特征来检测很难实现可靠的手部检测。为了解决这个问题，使用手掌检测而不是手检测，这是因为估计手掌的边界框比估计具有关节的手指检测要简单得多，而且手掌目标更小，即使在握手等造成双手自遮挡的情况下，非极大值抑制算法也能良好地工作，另外，手掌是正方形的，只使用正方形的边界框来建模手掌，可以忽略其他长宽比，减少锚框的数量。使用类似于特征金字塔网络的编码解码器进行特征提取，这样即使在大场景中，大物体和小物体都能在不同尺度下感知到。在训练过程中使用FocalLoss损失函数，用于支持大尺度方差所产生的大量的锚框。

2.3 手部关键点检测器

输入图像经过手掌检测器后，检测手部关键点坐标的模型通过回归预测手部的关键点坐标，手部关键点检测器学到了内在的一致性手部姿势表示，即使在手部部分缺失或者自遮挡的情况下也能表现出非常好的鲁棒性。手部关



0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

图 2　手部关键点示意图

键点检测器将输入图像映射到21个关键点位置 $p_i \in \mathbf{R}^3$，关键点位置包括 X，Y 坐标和相对手掌的深度 Z 坐标，Z 值越小，表示离摄像头越近，输入图像中包含手的置信度，以及左右手二分类信息。如图2所示。

2.4　手势识别算法

手势识别算法将识别到的手部关键点结构分类成一系列离散手势，使用一系列算法来推导手势。手势识别经常需要判断某个手指是否弯曲，第一种方法是计算手指头关节的第一节和第二节的夹角，根据设定的相关阈值来判断手指是否弯曲，如果它们之间的夹角大于某个阈值则定义为弯曲，小于某个阈值则定义为伸直，但是这种方法有一定的局限性，比较适合手指侧面图像的场合。第二种方法是根据关键点的 y 坐标判断手指是否伸直，例如食指的关键点8和关键点6，如果关键点8的 y 坐标小于关键点6的 y 坐标，则定义食指为伸直。第三种方法是计算每个指尖关键点与手掌的距离，手指尖与手掌的距离小于阈值的时候，定义手指为弯曲，手指尖与手掌的距离大于阈值的时候，定义手指为伸直。在实验中要综合运用几种方法，根据手指是否弯曲，可以识别手势1和手势2。

手势3和手势4的检测，除拇指外，可以使用前面几方面确定其余四指是否伸直，再根据食指的状态确定拇指是否弯曲。如果食指是伸直状态，拇指的关键点4与食指的关键点14的距离小于拇指的关键点3与食指的关键点14的距离，则拇指为弯曲状态；拇指的关键点4与食指的关键点14的距离大于拇指的关键点3与食指的关键点14的距离，则拇指为伸直状态。如果食指是弯曲状态，拇指的关键点4与食指的关键点6的距离小于拇指的关键点3与食指的关键点6的距离，则拇指为弯曲状态；拇指的关键点4与食指的关键点6的距离大于拇指的关键点3与食指的关键点6的距离，则拇指为伸直状态。

3　实验及结果分析

本文的实验环境为Intel 9700K，内存32 GB，显卡Nvidia 2070，操作系统为Windows10，深度学习框架为Pytorch1.10。

本文为手势交互设计了4种手势，根据实验的需要从网络上收集或使用摄像头和计算机视觉技术捕获手势图像，每种手势180张，共采集720张图像。把数据集输入到手势识别系统，输出每个图像中的手部关键点和识别结果，最终得到检测置信度为70%时的各手势识别的混淆矩阵，如表2所示。

表 2　预测手势与真实手势的混淆矩阵

	手势1	手势2	手势3	手势4
手势1	177	1	1	0
手势2	0	175	2	0
手势3	0	1	174	0
手势4	0	1	1	180
其它	3	2	2	0

表2中 X 轴为手势的真实标签，Y 轴为手势的预测值，表中对角线显示了识别正确的手势数量，当预测的手势不在定义的4种手势内时，归到其它类中。

计算手势识别算法的精确率、召回率、$F1$ 综合评价指标，如表3所示。

表3 手势预测的精确率、召回率以及*F*1综合评价指标 单位:%

	精确率	召回率	*F*1综合评价指标
手势1	98.9	98.3	98.6
手势2	98.9	97.2	98.0
手势3	99.4	96.7	98.0
手势4	98.9	100.0	99.4

通过实验，手势识别平均精确率达到99.0%，平均召回率达到98.1%，平均*F*1综合评价指标达到98.5%，在实验过程中手势识别算法对拇指是否弯曲还存在一些误判，但整体识别精度能满足VR/AR中对手势交互的要求。

4 结语

本文主要研究了手势识别在VR/AR中的应用，在MediaPipe平台，通过单目摄像头采集手势图像，基于手部关键点特征实现实时手势识别算法，达到了通过手势控制VR/AR系统中的人机交互的目的，在实验中设计了4种不同的手势，平均*F*1综合指标达到98.5%，该方法的可行性和鲁棒性均得到了验证，可用于基于计算机视觉的手势交互场景的实时识别，为虚拟现实/增强现实的人机交互提供了一种可供选择的方案。

参考文献:

[1] 刘璇恒,邓宝松,裴育,等.穿戴式手势交互系统与识别算法研究[J].小型微型计算机系统,2020,41(11):2241-2248.

[2] 李兆基,王海鹏,阮伟华,等.辅助虚拟现实数据手套及电刺激力触觉增强反馈电路系统设计[J].电子科技,2021,34(3):28-35,42.

[3] 杜佳祥.基于Leap_Motion的手势识别改进算法[J].现代计算机,2021:123-125.

[4] 刘瑜兴,王淑侠,徐光耀,等.基于Leap_Motion的三维手势交互系统研究[J].图形学报,2019,40(3):556-564.

[5] 高宇,何小海,吴晓红,等.基于Leap Motion的动态手势识别方法[J].计算机系统应用,2019,28(11):208-212.

[6] 龙江腾,高永平.一种VR游戏手势运动识别装置[J].现代电子技术,2021,44(12):173-176.

[7] 李国城.基于Kinect视觉功能对体感手势目标识别[J].科技创新与应用,2021(14):38-41.

[8] 任重庚,沈捷,王莉,等.基于Kinect骨骼数据的手势识别[J].计算机工程与设计,2019,40(5):1440-1444,1450.

[9] 张云峰,张超,吕钊.基于关键点的残差全连接网络动态手势识别方法[J].安徽大学学报(自然科学版),2022,46(2):30-38.

[10] 苏静,刘兆峰,王媛,等.基于改进的YOLOv3实现手势识别的人机交互方法[J].天津科技大学学报,2021,36(6):49-54.

[11] 丛玉华,何啸,邢长达,等.基于计算机视觉手势识别的人机交互技术研究[J].兵器装备工程学报,2022,43(1):152-160.

[12] 宋一凡,张鹏,刘立波.基于视觉手势识别的人机交互系统[J].计算机科学,2019,46(11A):570-574.

[13] ZHANG F, BAZAREVSKY V, VAKUNOV A, et al. MediaPipe hands: on-device real-time hand tracking[EB/OL]. arXiv:2006.10214,2020.

[14] 娄泽华,殷继彬.人机交互中的手势设计原则分析[J].软件导刊,2018,17(4):19-24,35.

Research on real-time gesture recognition for human-computer interaction in virtual reality/augmented reality

Ke Jian, Zhang Liang

(School of Computer Engineering, Suzhou Vocational University, Suzhou 215104)

Abstract: Human-computer interaction in virtual reality/augmented reality is an important research direction. This paper combines gestures with operation control in VR/AR, and proposes a human-computer interaction method that uses gestures to realize VR/AR operation. In the experiment, the MediaPipe framework was used to collect real-time gesture images, extract hand key points, and design a gesture key point algorithm for real-time gesture recognition. The results show that the method has good performance and provides a more natural human-computer interaction mode for VR/AR.

Keywords: gesture recognition; human-computer interaction; hand key point detection

图形图像

文章编号: 1007-1423（2023）01-0063-07　　DOI: 10.3969/j.issn.1007-1423.2023.01.012

基于卷积神经网络的图像识别技术研究

弓佳明

（广州华商学院数据科学学院，广州 511300）

摘要： 深度数据学习追求的一个最终技术目标即是让智能机器也能够变得像普通人电脑一样的具有自主分析及学习能力，能够快速识别各种文字、图像信息和图像声音文本等各种数据。图像识别技术是指对图像内有用信息进行读取与识别，通过对提取信息进行判断与分类来降低人工劳动强度，提升效率。通过研究图像预处理技术与卷积神经网络技术，实现对图像中多个对象点的识别，实现对图像中重点对象的状态研判。

关键词： 深度学习；图像预处理；图像识别；卷积神经网络

0　引言

图像模糊识别理论问题研究中涉及的图像模式识别数学基础本质也就是从属于在一个统计模式空间中到每一个模糊类别空间之间的模糊映射的关系问题。目前，在计算机图像模糊识别理论技术的逐步深入及发展研究领域中，主要的研究领域有以下三种计算机图像空间识别数学建模与方法：统计模式空间的识别、结构模式空间的识别、模糊模式空间识别[1]。从20世纪中期至70年代，相关领域经历了漫长的发展历程，并受到国际学术界的广泛关注及重视，到21世纪，人们仍可以借助对图像算法的理论跟研究不断创新新的领域与研究，以此来推动社会的进步与发展，其中最值得关注的就是图像识别领域，有关这一领域的实验技术的相关实验理论研究工作至今仍在稳步进行[2]。

近年来，随着我国计算机科学技术水平进一步快速进步和全面发展，图像视觉生物识别及检测系统技术和科学计算技术的综合应用逐渐发展起来，并在实践中取得令人瞩目的、十分突出可喜的创新科技成果，成为目前国内又一个综合性重点学科，并逐渐被广泛应用于国内其他众多应用技术领域[3-7]。

1　图像识别技术概述

1.1　图像预处理技术

1.1.1　图像增强

图像的增强方法[8]是指计算机通过采取某种图像计算处理方式，通过数学公式将图片分解成像素点，然后将需要处理的重点区域提取或者强化重点区域的边界，以此来将图片中重要信息内容突出，并且将图像中不重要的或者不需要处理的信息部分进行减少像素点，即弱化图像重点区域中的边缘信息，来达到图像信息增强的目的，这样做能够为后续图像处理增强识别度，并且减少其他信息对图像识别的影响。

假设原始像素图$f(x,y)$的像素灰度级范围均为$[0,L-1]$，其原始像素灰度值可以用整数r来表示，假设整数r能够被归一化映射到像素区间$[0,1]$中。做到如下变换：$s=T(r)$，其满足以下几个条件：

（a）Tr在区间$0\leqslant r\leqslant 1$中，为最小单值且单调地递减；

（b）当$0\leqslant r\leqslant 1$时，$0\leqslant T(r)\leqslant 1$。

收稿日期： 2022-09-29　**修稿日期：** 2022-11-24

基金项目： 校级青年学术科研项目（2021HSQX48）：基于结构增强的极大团社区发现算法

作者简介： 弓佳明（1993—），男，陕西咸阳人，硕士研究生，助教，研究方向为图像识别、多目标优化算法

条件(a)中所要求函数的Tr值必须为单值的设置目的主要是为了能够自动保证在对其做变换计算的两个函数值之间存在着逆反式变换；单调值函数的设置目的之二则是为了保证其输出的灰度图像中的灰度值可以自动实现从灰度由偏黑开始逐渐过渡到偏全白，并逐步按顺序逐渐地自动增加。条件(b)保证了输出信号的灰度级之间与实际信号输入中的信号灰度级之间都可以一直大致地保持在一个相同的大小范围。由s到r之间的逆反变换可以分别来表示条件为$r = T - 1s$，其中$0 \leqslant s \leqslant 1$。

因此，对于某一幅灰度图像来说，它对应的高灰度级次函数可以很严格地被看作是表示一个灰度区间值为[0,1]的一个随机变量。设函数$P_r(r)$为概率密度函数，$P_s(s)$为表示灰度值为s的概率密度函数。设$P_r(r)$和Tr都是已知，且它们都能满足(a)条件时，则可得

$$P_s(s) = P_r(r)\left|\frac{\mathrm{d}r}{\mathrm{d}s}\right| \tag{1}$$

均衡化的变换函数可表示为

$$s = T(r) = \int_0^r P_r(r)\mathrm{d}r \tag{2}$$

上式最右半部分是表示变量r值的单调累积分布函数，表示该分布函数为单调递增，满足条件(a)。类似地，区间[0,1]域上变量的概率密度函数的积分因子也包括在区间[0,1]域变量中，因此它也同时满足条件(b)。离散化形式为

$$s_k = T(r_k) = \sum_{j=0}^{k} P_r(r_j) = \sum_{j=0}^{k} \frac{n_j}{n} \tag{3}$$

其中，$0 \leqslant r_k \leqslant 1, k = 0, 1, 2, \cdots, L - 1$。

1.1.2 图像复原

由于图像退化过程会造成不可逆转的像素丢帧，导致图像质量会有一定的下降，并且存储转存的过程也会对图片像素等造成影响，还有其他射线辐射、噪声干扰、非线性畸变、带宽限制等会造成图像的几何失真。可以用图1来表示图像的退化模型。

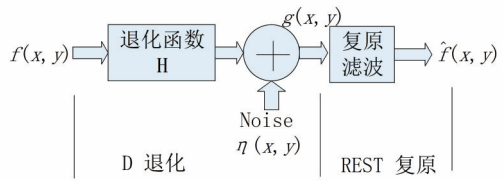


图1 图像的退化与复原过程

一般可以采用退化跟复原方法来实现计算机对退化图像信号的复原，利用对此类过程结果进行建模，寻找能去除影响或可削弱减少其影响效果的估计方法。另一种类型是指基于检测图像的估计方法，即如原始和未检退化的图像过程已知，我们同样可以选择通过对其原始退化图像结果进行建模，根据所建模型可以对原始退化后的图像结果进行复原。

1.1.3 形态学处理

最普遍的数学形态学中的处理与运算基本方法主要有以下7种，分别表示为腐蚀、膨胀、开运演算、闭式运算、击中、细化和粗化。这里重点讲一下腐蚀跟膨胀：提取的目标区域由0或1组成的二值区域图像，得到的图像边缘可能不是很理想，因此需要进行进一步的处理以此来消除边缘噪声点，类似于电路系统中的滤波操作。腐蚀“收缩”是将边缘像素点二值与邻近区域进行对比，消除邻近区域近似度低的样本像素点，以此来使边缘更加“圆滑”的操作，即减少像素边缘区域像素点，来消除噪声点的目的。膨胀“扩大”，通过自定义一个元素结构，在待处理的二值图像上进行“滑动”操作，将二值点与“滑动”到的区域进行近似度对比，通过二值运算增多此时像素点，达到“膨胀”效果，即增加选中区域的像素点来放大特征区域样本。

假设$f(x,y)$为输入图像，$g(i,j)$为结构元素，用$\triangleright$表示膨胀运算，$\triangleleft$表示腐蚀运算，这四种运算的表达式如下：

膨胀运算：

$$f \triangleright g = \max_{(i,j)} [f(x - i, y - j) + g(i,j)] \tag{4}$$

腐蚀运算：

$$f \triangleleft g = \min_{(i,j)} [f(x + i, y + j) + g(i,j)] \tag{5}$$

开运算：

$$f \circ g = (f \triangleleft g) \triangleright g \tag{6}$$

闭运算：

$$f \bullet g = (f \triangleright g) \triangleleft g \tag{7}$$

1.2 图像特征提取

图像信息的特征识别一般是建立以识别图像系统中的物体本身的主要特征为基础条件的，

在图像特征信息提取分析环节中最常被采用的主要识别特征通常是其颜色、形状轮廓。下面分别对这两种特征进行介绍。

1.2.1　颜色特征提取

在图像系统的各种视觉特征属性中，颜色特征属于最重要，但也是系统最基本、稳定可靠的重要特征属性之一，通常情况下将图像的颜色信息划分为三原色，通过对三原色所占比例、颜色种类、颜色配比、图像的高度、宽度等信息来标定一个图像的颜色信息；其次，要学会选择更合适的图像量化方法，即将图像以像素点的方式决定图像的颜色总数，将图像的颜色信息以数学公式将像素点转化为具体数值的方式实现对图像特征的准确量化判断；最后，定义另一个颜色相似性准则，以精确测量两个图像对象之间相同颜色级别的特征的相似性或差异性。

1.2.2　形状特征提取

形状特征通常被认为是一个只要有下列至少两种特征就可直接进行图形的形状表示的方法：轮廓跟区域特征。由于我们主要处理的图像大多是静态的照片，即目标图像几乎不会有变形存在，因此形状特征提取既能对目标中重点区域进行检索跟识别，又能减少识别时间，提高算法效率。现重点介绍基于轮廓特征方法中的Hough变换直线检测法。

Hough变换[9]的一项基本理论思想是空间图中的各点一线都存在着对偶性，即空间图像中共线上的各个点都在一个坐标系之内，空间中会组合成相交的线。假设存在一根直线方程l，原点到该直线方程的最小垂直线距离为r，垂线点与坐标系x轴方向的最大夹角为θ，则该直线方程可简单表示为

$$\rho = x\cos\theta + y\sin\theta \tag{8}$$

直线l点在直角坐标轴下显示出来的任意一个为极点。这种由时间线到点间距离的变换过程实际上本身就是Hough变换。

1.3　图像匹配技术

图像匹配算法主要解决多幅图像之间识别相同类型信息的相关系数，取相关系数较高的对应的图像位置，目前图像匹配方法主要是两大类：一类是基于图像灰度值的图像匹配技术，即将图像信息划分为0跟1两个不同的向量矩阵，通过矩阵的匹配来达到相似度较高的区域，一次完成信息的匹配；另一类是基于图像特征的匹配方法，即将识别图像的点、线、面、轮廓或者外观形状等鉴别度较高的位置区域当作特征点，转化为数学向量，通过向量的计算来得到相似度较高的区域，以此来完成图像的匹配。现主要介绍一种基于特征进行的匹配中使用的SIFT（scale-invariant feature transform）算法[10]。

正常情况下，目标检测算法对图片的大小、方向、角度，以及旋转非常敏感，而SIFT算法则是选择图像局部，因此在图像进行旋转、大小缩变的时候由于其通过计算高斯差分，以及后续的特征提取等信息表现优异，所以利用SIFT特征提取进行图像识别在处理具有噪声点的图像时具有很好地鲁棒性。SIFT算法具体流程如图2所示。

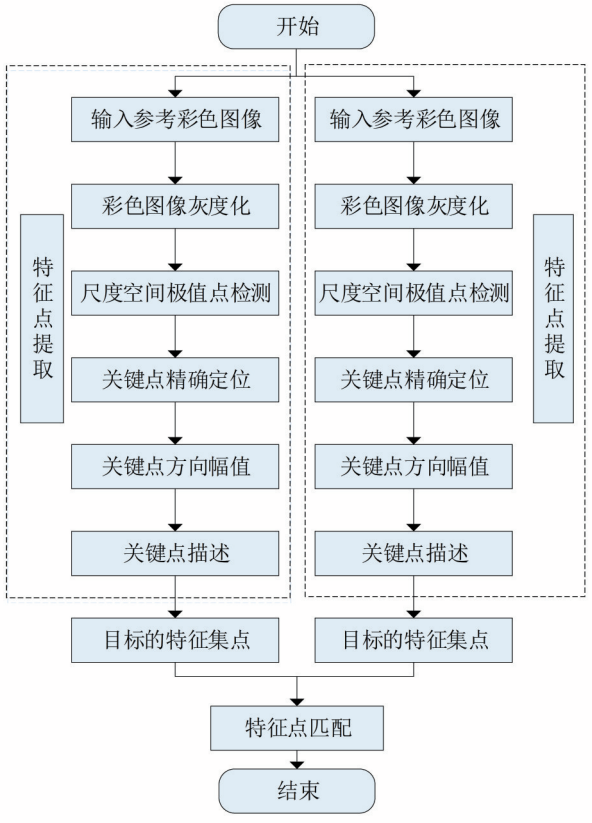


图2　SIFT算法流程图

传统计算机所用的SIFT算法目的主要在于实现模板图像特征与计算机实际采集及检测到的目标的模板图像特征信息间快速进行的精确匹配，采用的方式一般是渐进欧式距离判断法。如果该渐进最近距离的差值和除以该点次的渐近距离值均小于匹配算法中所能事先精确设定得出的距离匹配点阈值，则可以将该点确定为距离匹配点，阈值的范围控制在区间[0.6, 0.75]内。

为了进一步精确地优化和匹配计算的预期效果，采用匹配算法(random sample consensus, RANSAC)。此算法是一种通过反复循环、不断迭代而寻找最优模型的参数估计方法，它通过反复地随机选取样本点的方式凑成当前样本子集并迭代优化数字模型，使其跟整体样本更具有相关性。

2 卷积神经网络概述

2.1 卷积神经网络(CNN)

卷积神经网络系统(convolutional neural networks，CNN)[11]是指一类可包含卷积神经计算符且具有深度学习结构特点的前馈式神经网络技术(feedforward neural networks)[12]系统。卷积神经网络通过模型迭代，规则的优化，模型规则的权重的矫正等途径将模型不断地进行更新，使得训练的模型具有更强的鲁棒性，能更高效率地完成要执行的任务。其中，神经网络的数量在某一范围内与模型效果成正比，即某一范围内神经网络数量越多，效果就越明显。神经网络用于图像处理时首先会将图像进行特征提取，特征提取由卷积层完成，经过池化层将重复识别的特征剔除，最后通过全连接层特征计算将图像信息转化为矩阵或者向量，一次得到图像数字化特征提取跟图像计算的目标结果，以便后续模型进行迭代计算跟权重优化。

卷积层理论中最主要的另外两个基本特征之一就是局部的连接和权值的共享。由于在局部全区域网络连接的计算工作中实际参数过多将会直接造成无法正确计算实际参数的情况，可通过自动实现参数大小的共享操作来达到逐步地缩减计算使用实际参数时的数量，为今后进一步实现多层网络互连计算提供可能。

2.2 神经网络对图像识别的实现

卷积神经网络输出的表示方法如图3所示。卷积神经网络系统产生出来的输入图像信号一般是彩色的，也有全彩色灰度两种主要输入信息形式，可输出信号分别为纯黑白、彩色图像的输出信息和半纯黑色灰度图像。彩色图像矩阵是指显示介质在基于RGB灰度三种基本图像色彩模式条件下自动生成彩色的某一种基本图像。

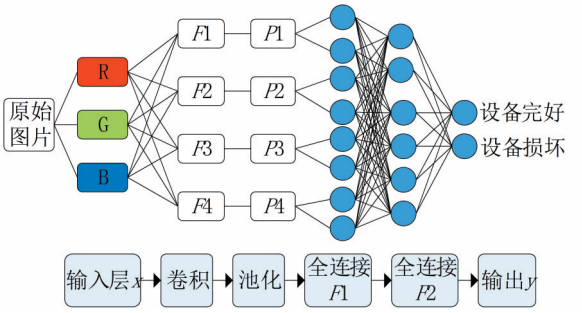


图3 卷积神经网络图

卷积层表示图如图4所示。由堆叠卷积核计算的交叉相关计算结果可知，随着堆叠卷积核层次图的连续堆叠，特征图表面上的尺寸收缩效应也会随之逐步地被放大减小，为此，填充技术作为一个可以在特征图尚未真正通过连续堆叠卷积核计算就通过人为的方法增大、减小其表面尺寸的收缩效应，来抵消连续堆叠对计算过程中产生的表面尺寸的收缩效应的影响及其作用效果的一种设计方法。常采用的两种进行填充算法的基本方法分别为按0进行的填充计算法和按重复计算填充边界值的填充算法(replication padding)等。

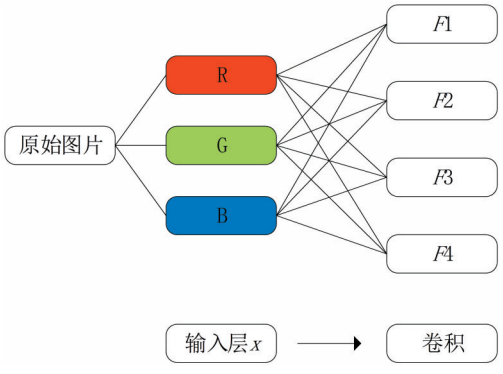


图4 卷积层部分图

卷积层输入公式为矩阵$x \in R^{W*H*C}$，这个输出公式也同样可以来表示矩阵特征图，输出矩阵为$y \in R^{w0*H0*C0}$。卷积层计算方法如下：

$$F1 = W11*R + W12*G + W13*B + b1 \quad (9)$$

$$F2 = W21*R + W22*G + W23*B + b2 \quad (10)$$

$$F3 = W31*R + W32*G + W33*B + b3 \quad (11)$$

$$F4 = W41*R + W42*G + W43*B + b4 \quad (12)$$

池化图层的具体实际作用范围包括降低特征图层中的图像平均分辨率，对于图像的平移过程会在一定程度上降低其鲁棒性，矩阵池化效果如图5所示。

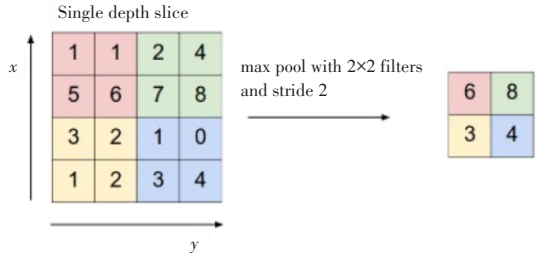


图5　池化效果图

ReLU层及相关层的一种变体，该层输出中包含的每个输入信号元素都是每一个输入信号元素的x。它是网络另一个非线性的输入操作，ReLU层的存在使得网络的表达方式变得更加复杂，ReLU函数为

$$R(X) = \max\{0, x\} = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{if } x \geqslant 0 \end{cases} \quad (13)$$

批归一化图层中的原始输入信号依然可以视为一个三维的输出信号，输出信号中的y值必须至少与其原始的输入信号值之间具有两个几乎相同的值。批归一化算法的基本设计方法之一是要通过对原始输入信号值上的每一个数据元素都能按照基本方式来进行批归一化，从而才能使得其原始的输出信号值的算术均值也可以被尽量地稳定保持为至少一个0，以及其算术方差至少也被尽量地固定保持为最少一个1。通过使用这种方式，每一层输出信号的原始数据的分布范围都将比较一致。该处理方法也将能够更大幅度地提升数据优化算法执行的速度。

从网络训练理论研究的一个数学角度，末端函数的主要数学功能其实是损失函数。通过随机梯度下降回归的一种方式，使得损失函数得到一种逐渐平滑的随机梯度的降低。目前使用比较广泛的一种方式应该是Softmax回归。

Softmax回归，其中：输入$x \in eL$表示图像在不同类别中的损失概率，用k来表示输入图像的标签，将输入归一化到[0, 1]之间，Softmax函数可表示如下：

$$f(x) = \frac{e^{x_j}}{\sum_j^n e^{x_j}} \quad (14)$$

通过交叉熵$y = -\log\left(f(x_k)\right)$来重新定义损失值。图6是一个线性Softmax函数的坐标轴图像，从实际数学应用和结果来看，其两侧区产生输出信号的输入增益变化甚小，在输出信号输入的特征空间中的映射应用理论上，有希望取得很好的数学应用效果。

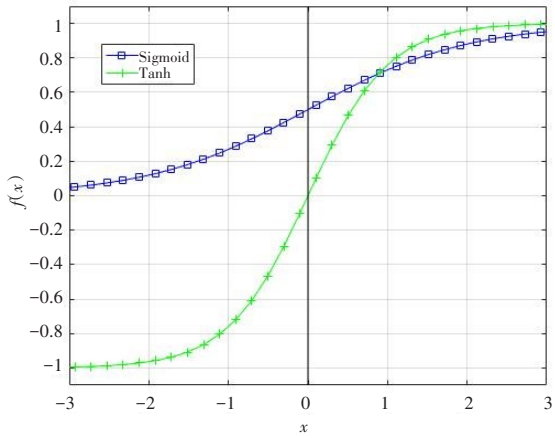


图6　Softmax函数图

3　基于卷积神经网络的图像识别技术

在一些实际案例的高清图像实时采集传输过程中，由于网络摄像机天线等一些电子设备本身的技术缺陷，或者由于外界环境以及各种环境不良干扰因素产生的影响，经常会导致实时采集检测到的视频图像可能存在图像噪声、色彩偏差、对比度变低以及画面失焦或者模糊等诸多问题。

在要识别的图像中，机器首先将图像识别为像素点，像素点之间的关联性与相似度成为图像识别的原理与基础，在一个图像中相邻像素点之间具有强关联性与强相似度，神经网络

通过识别像素点之间的相关联性来完成图像的识别与对比。在识别过程中我们先进行图像与处理，即提取了图像的特征信息，将这些特征信息交给神经网络来进行识别，神经网络通过对特征识别区域的每个像素点分配神经元，并进行邻近神经元之间的相关性与速度计算得到向量矩阵，这个向量矩阵就是图像的特征信息包，通过进行特征信息包的对比与鉴定，来完成对不同图像间对象的识别。算法流程如算法1所示：

算法1

Ⅰ：首先对图像进行预处理

输入：初始图像

Ⅱ：对图像进行特征提取

将图像中主要用到的部分提取出来；

弱化非重要因素对判断结果的影响；

较少后续图形中其他噪声点污染；

Ⅲ：对提取特征的图像进行卷积神经网络学习

（输入层，卷积层，池化层，全连接层）。 //更新迭代卷积神经网络学习基本流程：

Step 1：用随机数初始化所有的神经网络不同参数/权重

Step 2：输入要训练的图片，执行前向传播的步骤，计算出每个类别对应的输出概率。

Step 3：计算输出层的总误差

Step 4：反向传播算法计算输出的误差值相对于神经元所有神经元参数或权重值间的梯度，并用梯度下降迭代法来迭代或更新所有神经元的卷积核矩阵值和每个神经元参数值/所有神经元权重值之间的差值，以有效保证算法使算法输出值的误差范围达到或最小化卷积核矩阵的元个数、卷积核尺寸

Ⅳ：通过softmax回归函数将结果输出

end

4 结语

由于图像中对象目标众多，内容类型繁多，传统的图像模板匹配算法不能很好处理图像中存在多个对象点的图像识别，在平衡机器学习的高算力需求基础上，研究得出一套兼顾性能与准确度的图像识别算法。本文通过研究图像预处理技术和卷积神经网络技术，采用图像识别技术实现对图像中重点对象的状态研判。

参考文献：

[1] 李雪芳. 基于机器学习的计算机网络图像识别系统[J]. 信息技术与信息化，2022(8)：206-209.

[2] 李龙飞，李君军，赵杰，等. 基于图像识别技术和轨道机器人的大坝安全智能移动巡检系统研发[J]. 中国农村水利水电，2022(11)：215-221.

[3] 肖帆，杨超. 图像识别技术在钢板表面检查的应用[J]. 中国仪器仪表，2022(8)：44-46.

[4] 陈兵，蒋行国. 卷积神经网络用于人脸特征提取[J]. 现代电子技术，2022，45(18)：182-186.

[5] 丁建军，白飞燕，任学禹，等. 基于浮选泡沫图像识别的精煤灰分预测系统[J]. 选煤技术，2022，50(4)：89-93. DOI：10.16447/j.cnki.cpt.2022.04.016.

[6] 陈海燕，刘振亚，周逸，等. 基于对比学习的终端区相似气象场景识别(英文)[J]. Transactions of Nanjing University of Aeronautics and Astronautics，2022，39(4)：425-433.

[7] 马睿，王佳，赵威，等. 基于卷积神经网络与迁移学习的玉米籽粒图像分类识别[J/OL]. 中国粮油学报：1-10[2022-11-24]. http://kns.cnki.net/kcms/detail/11.2864.TS.20220803.1310.010.html.

[8] 闫根弟，戴丽莉，张广勇. 基于交叉滤波技术的可见光图像增强方法研究[J]. 激光杂志，2022，43(8)：146-150.

[9] 汪辉. 基于压缩感知和Hough变换的层叠纸张计数算法研究[D]. 武汉：武汉科技大学，2015.

[10] 陈宗桂，董晓军，曾令容，等. 改进SIFT算法在医学图像配准中的应用研究[J]. 计算机技术与发展，2022，32(8)：71-75.

[11] 董小虎. 基于卷积神经网络的图像目标检测技术研究[D]. 湖南：国防科技大学，2018.

[12] 赵勇，冯纯伯. 时变最优化信息处理技术及其应用(Ⅱ)：前馈神经网络学习算法[J]. 东南大学学报，1999(4)：100-104.

（下转第94页）

实践与经验

文章编号: 1007-1423（2023）01-0069-05 DOI: 10.3969/j.issn.1007-1423.2023.01.013

基于数字水印的即开型彩票防伪探究

樊东霞

（中国福利彩票发行管理中心数据部，北京 100101）

摘要：数字水印技术是一种有效的信息隐藏技术，通过探究数字水印技术能够实现对即开型彩票(以下简称即开票)的防伪验证，借助哈希算法单向、不可逆的特性结合随机数列生成水印信息，保证数字水印的随机性和安全性。采用DCT水印嵌入算法，通过对比原始水印和从待验证即开票中提取的水印信息的相似性，实现对即开票真伪的鉴定，同时不增加印刷成本。运用数字技术赋能，提高即开票的安全性，推进彩票行业高水平发展。

关键词：彩票；即开票；数字水印；DCT；防伪

0 引言

彩票是指国家为筹集社会公益资金，促进社会公益事业发展而特许发行、依法销售，自然人自愿购买，并按照特定规则获得中奖机会的凭证[1]。目前，市面上主流的彩票品种有即开型、数字型、乐透型、竞猜型、基诺型等多种类型，即开型彩票指由发行机构预先设定游戏名称、面值、奖组、玩法规则和奖级结构等内容，并组织印制发行的即买、即开、即兑型彩票品种[2]。

本文探讨的是即开型彩票的防伪印刷，目前印刷防伪主要通过专用纸张、专用油墨、激光防伪等技术，存在成本过高、工艺繁琐复杂、防伪有效期短等不足。如何在保证效果的前提下，降低防伪成本变成彩票机构需要研究解决的课题。本文提出通过数字水印技术达到防伪目的，水印检测率高、安全性高，能够验证即开票的真伪，防止不法分子伪造，同时水印技术成熟便捷，能够显著降低防伪成本。

1 数字水印技术

1.1 数字水印简介

数字水印技术是一种有效的信息隐藏与安全维护技术，最初是为了在开放的网络环境下表征数字产品产权而设计发明的。数字水印能够在不影响原始数据的价值和使用的情况下，将图像、文字、符号、数字等一切可以作为标记的数字信息与原始数据紧密结合并隐藏其中，原始数据又称为载体，可以是文本、音频、图像、视频等格式。

数字水印需具备以下几个特征：①抗篡改性，数字水印是难以伪造或改变的；②隐蔽性，数字水印的嵌入不影响载体的使用效果；③稳健性，数字水印难以被去除；④标识性，数字水印的容量要足以表达所有者的标识；⑤安全性，水印算法应该是公开的，不通过算法保证数字水印的安全。

数字水印常用的分类有以下几种：①根据数字水印的载体数据，分为图像水印、文本水印、音频水印、网格水印等；②根据数字水印的用途，分为证件防伪、版权标识、篡改提示等；③根据数字水印的应用要求，分为鲁棒性水印和脆弱性水印[3]。鲁棒性水印要求水印经得起变换处理或图像攻击，比如压缩、位移等；脆弱性水印对变更敏感，用于完整性保护。

1.2 数字水印算法

以图像为载体的数字水印算法，根据数字水

收稿日期：2022-07-09 **修稿日期：**2022-08-23

作者简介：樊东霞(1986—)，女，河北衡水人，硕士研究生，工程师，研究方向为信息安全、数据分析

印的嵌入位置，可分为空间域水印和变换域水印。

1.2.1 空间域算法

指对数字图像在空间域上进行处理的一类算法，这类算法一般处理速度较快，但是水印鲁棒性较差，容易遭受攻击被破坏。常见的空间域算法有最低有效位(LSB)法、拼凑(Patchwork)算法等。

1.2.2 变换域算法

指通过傅里叶变换(DFT)、离散余弦变换(DCT)、小波变换等将图像从空间域转换到相应的频率域，在频率域进行处理后，再反转变换到空间域的一类算法。相对空间域算法，变换域算法更复杂，计算量更大，但鲁棒性更强。

1.3 数字水印在印刷域应用

随着近年来信息处理技术和科技工业的发展，高精度的扫描仪、高质量的打印机等图像处理设备和功能强大的图像处理软件取得长足发展并在市场上广泛应用，给人们生活提供便利的同时，也使得证件、票据等重要文件的篡改伪造变得更加简单。

数字水印防伪技术在印刷业应用指的是将水印隐藏在印刷载体中，人眼不可见，但通过特定的设备或者软件可以识别检测。水印信息需要一直存在于印刷载体中，不能因为地点、时间的变换而产生变换，同时数字水印技术与印刷设备、材料无关，不会增加额外的印刷成本。由于水印是不可见的，嵌入的位置又是随机的，水印信息难以被轻易识别和伪造，提高了该技术的安全系数。

目前，印刷品防伪和版权保护方面的数字水印应用，在市场推广已有一定进展，例如：在传统媒体或票据中隐藏不可见的数字水印，通过网络摄像机扫描，可推送与传统媒体或票据相关的网址，并显示相关属性信息。通过数字水印阻止商标及重要文档的伪造、盗版和篡改。通过数据水印鉴定载体来源、证实载体完整性、判断是否被篡改以及定位篡改的位置。

2 即开型彩票防伪

彩票具备娱乐、公益的属性，尤其即开票主题丰富、画面精美、文化承载能力强，丰富多彩的票面设计能吸引彩民购彩、宣传公益理念，有利于文化的传播。2021年，高频快开游戏全面退市后，即开型彩票抓住时机，乘势而上，市场规模持续扩大，当年即开票销量达540多亿元，创下历年销量新高。

高额的奖等、大规模的市场份额、购彩者无须实名的特点，使得即开票很容易成为不法分子的目标，在此场景下，即开票如何高效并且低成本地防伪变得至关重要。本文设计用哈希算法结合随机数列生成的水印信息，并通过DCT算法嵌入到即开票的印刷图像标识中，通过扫描设备和专门的应用软件检测即开票印刷图像中的数字水印，水印防伪流程如图1所示。

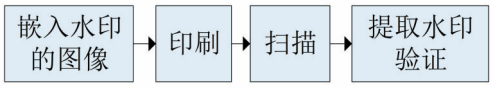


图1 水印防伪流程图

假如不法分子利用彩票扫描后图像再次印刷伪造，一般流程如图2所示。

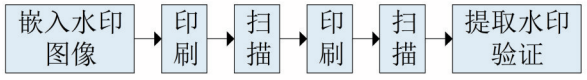


图2 伪造流程

由图2可以看出，伪造即开票多了一次扫描印刷的过程，由于电子电路的不稳定性会不可避免地产生噪声，所以图像在打印扫描的过程中，几乎都存在一定程度的失真问题，包括明暗差距、影像偏色、几何失真等。通过对比原始水印与扫描提取出的水印信息，根据水印信息的相似程度可以判断即开票真伪。

2.1 数字水印的生成

针对不同游戏不同批次的即开票产品，可由发行机构选择根据游戏编码、即开票票号、批次编号作为水印生成要素，通过哈希算法结合随机数列生成水印信息。

世界上广泛应用的哈希算法包括MD5与SHA系列，哈希算法能够对任意长度的输入信息进行压缩或者填充，输出固定长度的哈希值，

输出信息的长度只取决于所使用的哈希算法。随着计算机技术的发展，MD5和SHA1已相继被破解，本文采用更加安全的SHA256哈希算法，该算法的输出是256位的二进制串，可以表示为

$$W^1=\{w_i,1\leqslant i\leqslant 256\},w_i\in\{0,1\}$$

为了增强水印的鲁棒性，通过随机数发生器生成两个伪随机实数序列K^1，K^0：

$$K^1=\{K_i,1\leqslant i\leqslant 256\},K_i\in\{1,2,3,\cdots,16\}$$

$$K^0=\{K_i,1\leqslant i\leqslant 256\},K_i\in\{1,2,3,\cdots,16\}$$

重新组合生成伪随机数列W：

$$W=\{K_i,1\leqslant i\leqslant 256\},K_i=\begin{cases}K_i^0,\ w_i=0\\K_i^1,\ w_i=1\end{cases}$$

将W顺序按行转换为16 × 16的方阵，作为水印信息存储在系统后台。

不同的即开票对应了不同的游戏编码、票号，作为水印的生成要素，利用哈希算法单向、不可逆的特性，保证了数字水印的安全性、标识性和抗篡改性，结合随机数列能够保证数字水印具有随机性，增加了不法分子伪造的难度。

2.2　数字水印的嵌入

数字水印可以选择嵌入到载体的任意位置，比如选择嵌入在即开票面的中国福利彩票标识的徽标图像中。图像在计算机的存储可表示为数学中M行N列的二维矩阵，矩阵中的数值表示图像对应位置上像素的颜色信息，数字图像的处理实质是矩阵的各种数学运算和处理。在常用的图像格式文件中，由于BMP格式是不使用压缩技术处理的位图图像，虽然格式简单，但往往图片过大不利于存储，JPG格式采用有损压缩技术除去冗余信息，可能会导致数据丢失，所以即开票采用可支持无损压缩技术的PNG格式作为即开票画面图像的电子文件格式。PNG格式支持RGB真彩色图像，图像中每一个像素的长度是24位，由R(红)、G(绿)、B(蓝)三原色叠加组成，数据类型是8位无符号整型，所以每一种颜色的取值范围为0x00～0xFF，能够很好地展示彩票的票面设计。RGB真彩色图像用三维矩阵表示，即$M\times N\times 3$，三个矩阵分别表示对应行列值位置上的像素颜色分量。

基于人类视觉系统(HVS)的研究表明，人眼对蓝色的敏感性最弱，所以选择将水印嵌入到蓝色通道效果最好。作为载体图像的PNG格式大小一般为几十KB到几MB，远大于数字水印的字节长度，利用载体图像的冗余信息足以嵌入数字水印。

在印刷的过程中图像会不可避免地失真，因此需要选择具备强鲁棒性的水印。本文采用DCT算法作为数字水印的嵌入算法，DCT算法是DFT算法的特殊变种，具备计算复杂性相对简单，信息聚集能力和隐蔽性较好的特点，是目前应用最为广泛的变换域算法之一。通过提取载体图像的蓝色分量，将图像从色彩域转换到频率域，经过DCT变换得到的系数矩阵表示一系列频率中每一个频率所对应的变化程度。

对于$M\times N$大小的数字图像，DCT正变换公式为

$$F(\mu,\nu)=c(u)c(v)\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}f(x,y)\cos\frac{\pi(2x+1)}{2M}\cdot\cos\frac{\pi(2y+1)}{2N}$$

$$\mu=0,1,\cdots,M-1;\nu=0,1,\cdots,N-1$$

其中：

$$c(u)=\begin{cases}\sqrt{\dfrac{1}{M}},\ u=0\\\sqrt{\dfrac{2}{M}},\ u=1,2,\cdots,M-1\end{cases}$$

$$c(v)=\begin{cases}\sqrt{\dfrac{1}{N}},\ v=0\\\sqrt{\dfrac{2}{N}},\ v=1,2,\cdots,N-1\end{cases}$$

DCT逆变换公式为

$$f(x,y)=\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}c(u)c(v)F(u,v)\cos\frac{\pi(2x+1)u}{2M}\cdot\cos\frac{\pi(2y+1)v}{2N}$$

$$x=0,1,\cdots,M-1;y=0,1,\cdots,N-1$$

其中：x，y为空间采样值；u，v为频域采样值。

在图像取样中选择方阵，计算相对简单，即N=M，二维DCT正反变换可以简化为

$$F(\mu,\nu)=c(u)c(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f(x,y)\cos\frac{\pi(2x+1)}{2N}\cdot\cos\frac{\pi(2y+1)}{2N}$$

$$u=0,1,\cdots,N-1;v=0,1,\cdots,N-1$$

$$f(x,y)=\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}c(u)c(v)F(u,v)\cos\frac{\pi(2x+1)u}{2N}\cos\frac{\pi(2y+1)v}{2N}$$

$$x=0,1,\cdots,N-1;\ y=0,1,\cdots,N-1$$

对载体图像进行DCT变换，对变换后的某些DCT系数进行小幅度的变换，完成水印信息的嵌入。步骤如下：

（1）DCT变换先将图像分解为8×8大小的不重叠子像素块$T_i(i=1,2,\cdots,N/8>16)$；对每一个子块T_i分别进行DCT变换，对DCT分块变换后的图像频率系数进行Zig-Zag方式扫描排列，共64个DCT系数。

扫描顺序如下：

1	2	6	7	15	16	28	29
3	5	8	14	17	27	30	43
4	9	13	18	26	31	42	44
10	12	19	25	32	41	45	54
11	20	24	33	40	46	53	55
21	23	34	39	47	52	56	61
22	35	38	48	51	57	60	62
36	37	49	50	58	59	63	64

（2）将对T_i扫描后的DCT系数重新排列成一维向量$\boldsymbol{y}=\{y_1,y_2,\cdots,y_{64}\}$，选择其中的16个中频数据，取出序列中第20到35的中频系数部分，得到$\boldsymbol{Y}_i=\{Y_1,Y_2,\cdots,Y_{16}\}$。

在DCT系数矩阵中，左上角第一个值为直流系数，其余是交流系数。左上为直流和交流低频系数，中间为交流中频系数，右下为交流高频系数。DCT算法具备可分离性和能量集中的特点，图像的低频分量反映了基本图像，高频分量代表图像细节部分，DCT变换后图像能量集中在图像的低频部分，黄继武等[4]在对DCT直流系数（DC）和交流系数（AC）分量的定性和定量分析的基础上，指出DC分量比AC分量更适合嵌入水印，嵌入DC分量的水印具有更好的稳健性。因此，水印嵌入到低频分量，具备较好的鲁棒性，水印嵌入到高频分量，鲁棒性差但具备更高的不可见性，本文选择将水印嵌入到中频分量，实现鲁棒性与不可见性之间的平衡。

（3）将水印信息W依次顺序按行分解为16组，用K_i表示分解后其中的一组，排成一维向量。

$$W=\bigcup_{i=1}^{16}K_i$$

$$K_i=\{k_1,k_2,\cdots,k_{16}\}$$

（4）对T_i像素块的Y_i序列值进行修改，a表示水印强度系数，取a=0.5。

$$Y_i'=Y_i+a*K_i,\ 1\leqslant i\leqslant 16$$

（5）对修改后的中频系数Y_i'通过Zig-Zag逆变换形式重组，再进行DCT逆变换，得到嵌有数字水印的图像子块。

（6）将图像子块重新组合成嵌入数字水印的图像。

2.3 数字水印的提取

水印提取的过程是水印嵌入的逆操作，扫描需要验证的即开票，对扫描得到的图像进行水印提取，并比对后台存储的票号和水印信息，设原始水印信息为W，提取出的水印信息为W'，计算相似度NC（normalized correlation）验证该票真伪，通用的NC计算公式如下：

$$NC=\frac{\sum_{i=1}^{i=m}\sum_{j=1}^{j=n}W(i,j)W'(i,j)}{\sum_{i=1}^{i=m}\sum_{j=1}^{j=n}W(i,j)^2}$$

计算机系统和扫描仪采用RGB彩色空间描述颜色，但是目前印刷业普遍使用的是CMYK颜色空间，由于印刷和扫描过程中图像在不同彩色空间之间的转换可能会引起颜色丢失、亮点失真等，影响水印的质量，因此，可以根据水印信息的相似度判断即开票的真伪。根据经验正品水印的相似度NC都是大于0.7的，该技术简单易操作、速度快、效率高，同时因为数字水印的隐蔽性，肉眼无法察觉，不影响票面美观，适合应用在即开票领域。

3 结语

作为最具灵活性和最易获客的产品类型，即开票将进一步加大研发创新力度和资源投放力度，成为品牌宣传、营销获客、拓展市场的彩票拳头品类。2022年，各级彩票机构通过派奖促销、扩大销售渠道等途径，全年销售规模有望创造新高，如何高效保证即开票的安全是彩票发行机构的必修课。

本文探索即开票的防伪设计，运用数字技

术赋能，通过数字水印技术，利用哈希函数的单向性结合随机数列生成水印，选择鲁棒性强的DCT水印算法，增强破解难度，通过比较水印信息的相似度达到安全防伪的目的，相较于传统的防伪技术，不需增加印刷材料或设备，不会改变即开票原有的设计图案，成本低廉、效果更优，为未来的即开票市场发展做好技术保障，有助于即开票的安全健康发展。

参考文献：

[1] 国务院. 彩票发行管理条例[S].(2009-05-04).
[2] 中华人民共和国民政部. 中国福利彩票即开型彩票[S].(2017-01-06).
[3] 刘海燕. 数字音频水印技术及应用[M]. 北京:清华大学出版社,2016:46.
[4] 黄继武,SHI Y Q,程卫东. DCT域图像水印:嵌入对策和算法[J]. 电子学报,2000,28(4):386-389.

Research on anti-counterfeiting of instant lottery ticket based on digital watermarking

Fan Dongxia

(Department of Technology, China Welfare Lottery Issuance and Management Center, Beijing 100101)

Abstract: Digital watermarking technology is an effective information hiding technology, used to realize the anti-counterfeiting verification of instant lottery ticket. By the one-way and irreversible property of Hash algorithm combined with random sequence to generate watermark information, guarantee the randomness and security of digital watermarking. Using DCT watermarking algorithm, by comparing the similarity between the original watermark and the watermark information extracted from the instant ticket, achieve the authenticity of the identification, at the same time does not increase the cost of printing. Using digital technology to improve the security of instant tickets, promote the high-level development of the lottery industry.

Keywords: lottery; instant ticket; digital watermarking; DCT; anti-counterfeiting

(上接第53页)

Establishment and simulation of ground waiting model based on flight time optimization

Li Keyang

(College of Air Traffic Mangement, Civil Aviation Flight University of China, Guanghan 618307)

Abstract: Ground holding refers to a method for aircraft to relieve congestion in an airport or control area by ground holding strategy when the capacity of the airport or control area is limited. In order to optimize the single airport ground waiting problem of flights, a ground waiting model based on flight time optimization was established according to the order of flights of different airlines and considering different aircraft types. The results show that when the airport is congested, the flights should be ordered so that the flights can take off and land in an orderly manner, and the total ground waiting loss of all the flights can be minimized.

Keywords: ground waiting; computer simulation; airport capacity; flights sorte

开发案例

文章编号: 1007-1423(2023)01-0074-07 DOI: 10.3969/j.issn.1007-1423.2023.01.014

基于 Android 系统的大学生心理健康服务APP的设计与实现

刘忻瞳[1]，毛明志[1,2*]

（1.中山大学计算机学院，广州 510275；2.中山大学软件工程学院，珠海 519082）

摘要：大学生心理健康问题是世界各国高度关注与重视的问题，而网络的发展逐步改变了高校进行大学生心理健康测试的途径。为帮助大学生获得更便捷多样的心理健康服务，推动高校心理健康教育发展，对大学生心理健康进行了研究，在心理问卷的基础上，设计并实现了一款大学生心理健康测试APP系统，具有一定的可用性。

关键词：Android；心理健康；APP；设计与实现

0 引言

大学生作为一个特殊社会群体，正处于由校园走向社会的过渡时期，主要面临着包括环境变化引起的不良适应性、异性交往引起的情感困惑、人际关系的困扰、学习压力及求职择业引起的紧张焦虑问题以及经济负担等心理问题[1]。近年来，随着计算机技术在高校的不断成熟和发展，传统的大学生心理健康教育已经不能满足大学生的心理健康需求，利用网络开展心理健康教育成为一种新思路、新途径[2]。

国内外高校在心理健康相关组织机构和服务方面、教育模式方面也存在着明显差异[2-4]。与国外高校的心理健康教育相比，我国对高校学生的心理健康教育起步较晚，但呈现出迅速发展的趋势，目前已经形成了较为完整的理论体系和操作模式，取得了很多的成果[5-8]，但从实际的效果来看，我国大学生心理健康教育工作仍然存在一些不足，如教学模式不够灵活多样、服务实效不够理想、体系和平台建设不够健全的问题[9]。为此，本文在心理问卷的基础上，设计并实现了一款大学生心理健康服务APP系统，可为高校大学生提供形式更丰富的心理健康服务，方便大学生了解相关知识，对自身心理健康状态有更清晰的认知，尽早发现心理健康问题并寻求学校辅导员教师以及专业人员的帮助。

目前国内对心理健康的研究方法中，测验法运用最为广泛，即通过国内外测评量表对大学生心理健康进行评定分析，使用频率较高的量表有：90项症状自评量表(SCL-90)、自编量表、大学生人格问卷(UPI)、抑郁自评量表(SDS)、焦虑自评量表(SAS)、卡特尔16种人格因素测验(16PF)[6]。

1 相关基础理论与技术

本文采用症状自评量表(SCL-90)对学生进行心理健康测量评估，系统开发采用前后端分离的混合开发模式。数据库使用PostgreSQL实现，在Ubuntu系统下运行；后端接口使用Flask实现，在Ubuntu系统下运行；前端使用Ionic作为框架，Angular作为JS框架进行开发，教师端应用了Echarts插件展示分析数据。

1.1 症状自评量表(SCL-90)

症状自评量表(SCL-90)作为当前广泛应用于精神障碍和心理疾病门诊的检查量表，具有容量大、反映症状丰富、能准确刻画测试者的

收稿日期：2022-08-29 **修稿日期：**2022-12-31

基金项目：中山大学自选课题项目(35000-14070034)

作者简介：刘忻瞳(2000—)，女，本科，学士，专业为软件工程；*通信作者：毛明志(1966—)，男，博士，副教授，CCF杰出会员，研究方向为软件工程、人工智能、项目管理等，E-mail：mcsmmz@mail.sysu.edu.cn

自觉症状等特点，能有效对大学生群体展开心理健康状况测查。

该量表共包含90个自我评定项目，分为9个因子，分别为躯体化、强迫症状、人际关系敏感、抑郁、焦虑、敌对、恐怖、偏执、精神病性[8]，剩余未归入任何因子的7项作为第10个因子处理。症状自评量表(SCL-90)每个项目均采用五级评分(无、轻度、中度、偏重、严重)[10]。每个因子分计算方法为组成该因子的项目总分除以组成该因子的项目数。

因子分为1～2.5，说明被测者自觉无症状或症状较轻；因子分为2.5～3.5，说明被测者自觉有症状，严重程度为轻到中度；因子分为3.5～4.5，说明被测者自觉有症状，严重程度为中到严重；因子分为4.5～5，说明被测者自觉有症状，且症状非常严重。

1.2　混合开发模式

混合开发模式(hybrid)是指利用 Web 技术进行 Web APP 开发，再将程序打包封装成原生APP，同时也可以通过开发原生插件的方式调用移动设备上的硬件资源[11]。

混合开发模式集原生技术和 Web 技术于一体，既能实现设备功能调用，又能保证一定程度的跨平台特性[12]，具有开发效率高，更新和部署方便，代码维护方便，可离线运行等优点。

1.3　Flask

Python 框架 Flask 是一个只提供 Web 框架的核心功能的轻量级框架，具有自由、灵活、可扩展性强、第三方库的选择面广等优点，非常适用于开发 Web 服务的API接口。因此，本系统后端使用 Flask 框架实现数据接口，接收前端发送的HTTP 请求，数据接口响应请求后将数据以JSON 对象的形式返回给前端。由于Flask本身不具备ORM框架，因此使用falsk-sqlalchemy模块进行数据库的反射和执行ORM操作。

1.4　Ionic

Ionic 是基于 Web 技术应用 HTML5、CSS3 和 JavaScript 技术进行智能设备 APP 开发的轻量框架，具有良好的跨平台性能，被称为 Hybird APP 框架。Ionic 既是一个 CSS 框架，也是一个 JavaScript UI 库[13]，具有原生 APP 的卓越运行性能，提供简洁、美观的 UI 组件库和强大的命令行工具，具有可维护性高、运行速度快、开发效率高等特点，适合用于开发移动 APP。

1.5　Angular

Angular 框架由 Google 公司开发，主要特点包括与 MVC 设计思想一致的整体框架、模块化和依赖注入以及双向数据绑定。Angular 最初被设计成一个标准 MVC 模式，但是通过代码重构和 API 重写，呈现出越来越接近 MVVM 框架模式的趋势。Angular基于TypeScript，TypeScript是JavaScript的一组超集，最重要的区别在于增加了类型的概念，提高了代码的可读性和维护性。

2　系统总体设计

2.1　系统总体架构设计

系统的总体架构采用三层架构，将整个系统的业务划分为应用表示层、业务逻辑层和数据访问层，目的是实现“高内聚、低耦合”，有利于系统的开发、维护、部署和扩展[13]。系统总体架构设计如图1所示。

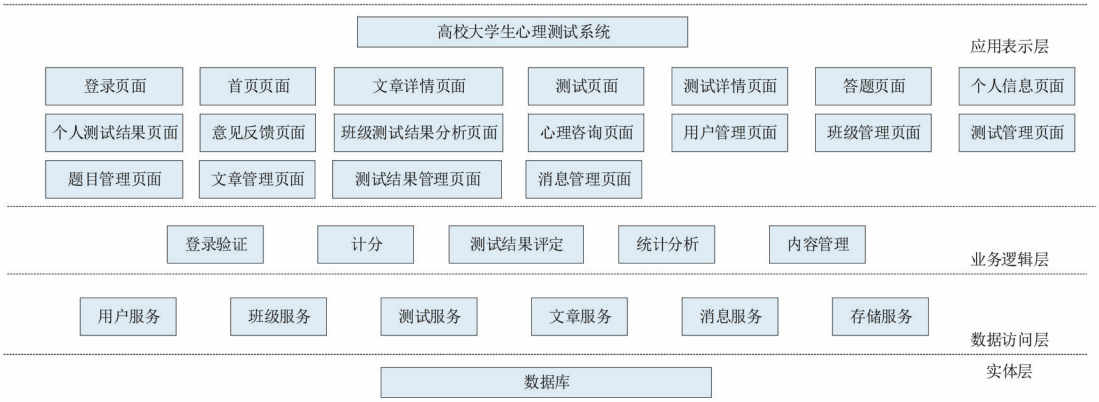


图1　系统总体架构

应用表示层主要负责直接跟用户进行交互，用于数据录入，数据显示等。逻辑业务层是系统架构中体现核心价值的部分，其主要任务是完成业务规则的制定、业务流程的实现以及其他与业务需求相关的系统设计，例如对具体问题进行逻辑判断与执行操作等。本系统中，逻辑业务层包括对用户输入内容的检验，计算测试分数，生成测试评定结果，统计分析测试成绩等功能。数据访问层是数据库的主要操控系统，主要负责为业务逻辑层或表示层提供数据服务，包括数据的增加、删除、修改、查询等基本操作，不包含逻辑处理。数据访问层是对数据的操作，而不是数据库本身。

2.2 系统功能模块设计

本系统使用角色分为学生、教师以及管理员三种，包括登录、心理健康测试、推荐内容浏览、成绩统计分析、心理咨询、用户管理、班级管理、测试管理、文章管理、测试成绩管理、消息管理等11个子功能模块。系统功能模块设计如图2所示。

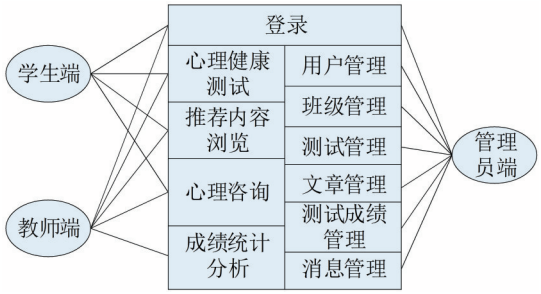


图2 系统功能模块设计

2.2.1 心理健康测试功能模块

心理健康测试功能模块主要由计分模块与测试结果评定模块组成。

测试开始后，测试题目的得分以数组形式存进存储服务中，切换下一题时，将本题得分推入数组；切换上一题时，将数组最后一个元素弹出。由于本系统测试题目均为单选题，选项采用1到5级评分，分别记1到5分，最后成绩为各题目得分之和除以测试题目总数。测试计分业务流程如图3所示。

测试评定结果文本以数组形式存储，系统根据不同测试成绩匹配相应的测试结果，测试评定结果分为以下四类：自觉无症状或症状较轻、自觉症状严重程度为轻到中度、自觉症状严重程度为中度到严重、自觉症状非常严重。测试结果评定业务流程如图4所示。

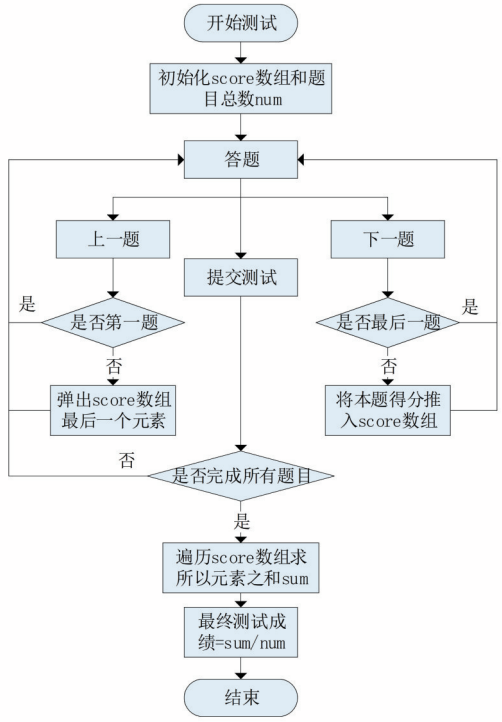


图3 测试计分业务流程

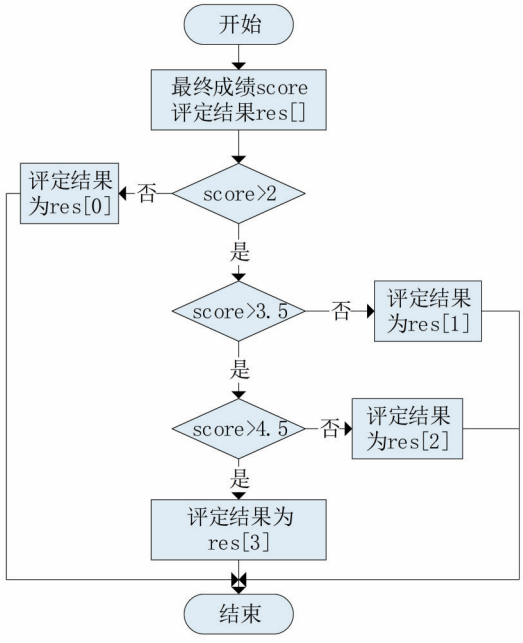


图4 测试结果评定业务流程

2.2.2　成绩统计分析功能模块

教师选择测试、班级后，系统将该班级学生在该测试下的成绩分为以下五类：未完成测试、自觉无症状或症状较轻、自觉症状严重程度为轻到中度、自觉症状严重程度为中度到严重、自觉症状非常严重。成绩统计分析业务流程如图5所示。

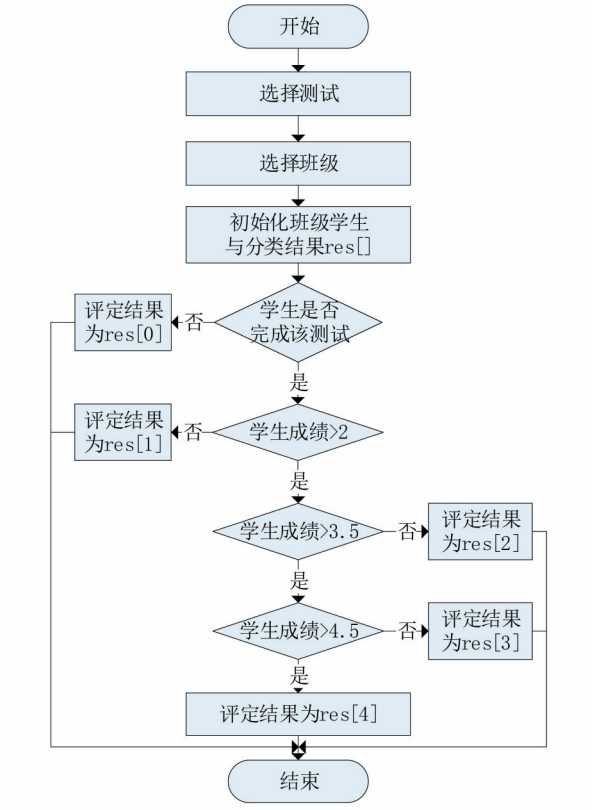


图5　成绩统计分析业务流程

2.3　数据库设计

本系统采用 PostgreSQL 数据库，并使用 pgAdmin 4 v5.0 版本作为开发工具。pgAdmin 4 是一款专门针对 PostgreSQL 数据库的管理与开发工具，它提供了友好的图形用户界面，大大提高了用户对数据库访问的便捷性。

本文主要运用对象角色建模（ORM），设计一个合理的数据库并建立相应的实体关系库。该数据库的结构具有设计合理的特点，易于创建和修改业务规则，可以减少数据库的存储量，增强数据一致性和完整性，提高系统响应速度，简化数据库应用数据[14]。

表1　数据库结构表

表名	备注
User_list	用户表
Class_teacher	班级-教师关系表
Test_list	测试表
Test_record	测试记录表
Message_list	反馈信息表
Class_list	班级表
Text-content	文本表
Question_list	题目表
Passage_list	文章表
Chat_list	消息表

3　系统实现

3.1　学生端

学生端可通过页面下方菜单栏切换“首页”“测试”与“我的”页面。

学生可通过下方菜单栏进入测试页面，页面分类展示已完成和未完成的测试，已完成的测试不可再次完成。选择要完成的测试后，可进入测试详情页面，查看测试简介与题目总数等信息。点击开始测试按钮后，可进入题目页面开始答题，用户可通过左、右箭头按钮切换上下题，测试结束后可在“我的”页面查看测试记录。学生端测试页面如图6所示。

图6　学生端测试页面

3.2　教师端

教师端应用的结构与学生端大体相似，不同的是“测试”页面不提供测试功能，而是为教师提供学生测试成绩统计分析功能。

教师可根据测试 ID、测试名称查找测试，

点击对应测试操作栏的“成绩分析”按钮进入成绩分析页面，查看所负责班级的测试统计分析结果。

成绩分析页面共分为三部分：最上方选择栏展示该教师负责的所有班级，可通过左右横向滑动浏览切换；中间饼图展示班级测试成绩分布情况，点击分块可查看对应成绩分段下的学生测试成绩；最下方展示该班级所有学生的该测试成绩，并提供按学生用户名查找测试记录的功能。未选择班级时，饼图和学生测试记录列表为空。选择班级后，对应班级的测试成绩分析页面如图 7 所示。

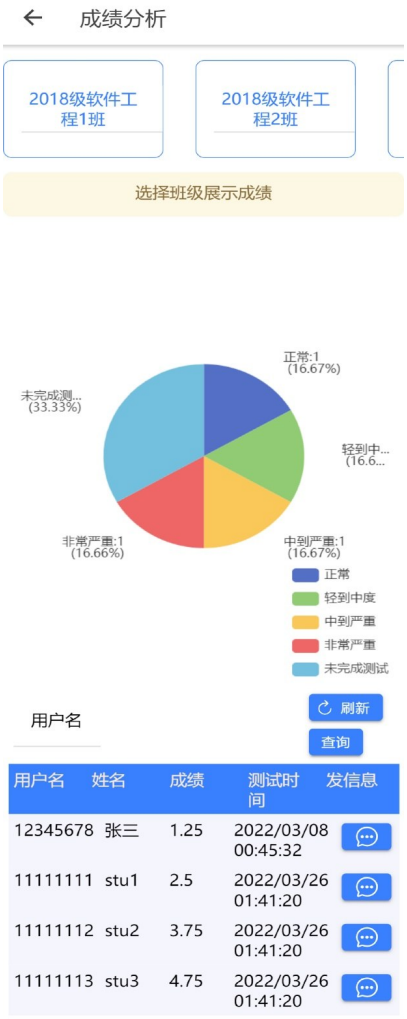


图 7　测试成绩分析页面

3.3　管理员端

管理员登录成功后，可以根据左侧的菜单栏进入对应功能的操作页面，对用户、班级、班级成员、测试、题目、测试记录、文章、用户反馈消息进行增、删、改、查等管理操作。管理员端页面如图 8 所示。

图 8　管理员端页面

3.4　心理咨询

学生点击“首页”页面右上方的信息按钮，可浏览到负责所在班级的教师用户，选择教师用户后可进入与该教师的聊天页面。

教师有两种方法选择学生，进行心理咨询服务：

（1）点击“首页”页面右上方的信息按钮，可浏览到发送过信息的用户，选择用户后可进入与该用户的聊天页面；

（2）进入成绩分析页面，点击对应学生“发信息”栏的按钮，可进入与该学生的聊天页面。

4　系统测试

4.1　测试目标

本文主要采用黑盒测试，根据程序功能的要求规范设计测试用例，并推导出测试结果的准确性。

4.2　测试实现与结果

本系统测试分为单元测试、集成测试、权限测试三部分。

4.2.1　单元测试

单元测试由一组独立的测试构成，每个测试针对软件中的一个独立的软件单元，用于检验被测代码所实现的一个独立功能的正确性。

本文对系统的 6 个管理员管理单元以及学生

答题单元进行了测试，测试结果正常，说明系统独立单元功能正常。

4.2.2　集成测试

集成测试是在单元测试基础上进行的一种有序测试，测试目的是验证软件单元之间、软件单元和集成的软件系统之间的接口关系，并验证已集成软件系统是否满足设计要求。

本文对学生端测试、测试结果生成与个人测试记录查看模块，意见反馈模块，心理咨询模块以及管理员模块功能进行测试，结果均为正常，说明系统各单元之间能顺利传输数据、协作运转，各功能模块运行正常，系统能完整实现需求。

4.2.3　权限测试

本系统用户划分为学生、教师以及管理员三种，进行权限测试的目的是检查是否出现用户越权行为。由于系统已划分为学生端、教师端以及管理员端三个不同的版本分别实现，只需在用户登录时检查用户角色是否具有登录当前版本的权限即可。

测试结果正常，说明用户可以正确登录对应版本完成相应操作。

5　系统部署与安装

5.1　搭建Android应用运行环境

Android 应用的运行主要在 Android Studio 上的模拟设备以及真实的物理设备上进行。Android Studio基于IntelliJ IDEA，为开发者提供集成的Android开发工具，便于应用程序的开发与调试。

搭建Android应用的运行环境包括模拟器运行与物理机运行。

5.1.1　模拟器运行

Android Studio中的设备模拟器提供多种虚拟设备供开发者调试运行项目，本文选用Pixel 2 API 30进行模拟运行。点击“Run”按钮后，Android Studio会在该虚拟设备上安装并运行应用包。经过模拟运行测试，APP能在虚拟设备上正常运行，运行页面如图9所示。

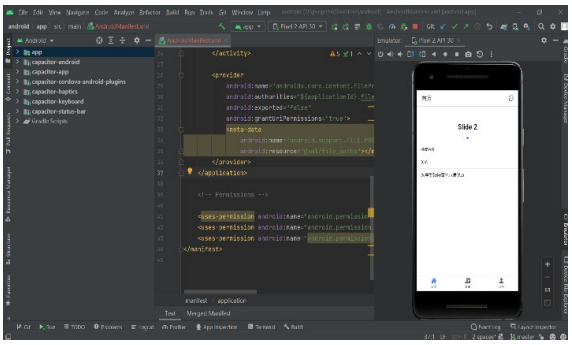


图9　虚拟设备运行页面

5.1.2　真机运行

真机运行使用USB数据线连接安卓手机与电脑，物理机匹配成功后点击“Run”按钮，Android Studio会在该物理设备上安装并运行应用包。经过模拟运行测试，APP能在真实物理设备上正常运行。

5.2　发布应用

测试通过后，需要将应用发布到应用商城中。Android系统要求签名机制，所有安装在Android系统上的软件都必须经过签名，目的是使用签名辨别软件的开发者。使用Android Studio在工程目录下成功创建APP-release.apk和testjsk.jsk文件后，即可将签名成功的apk文件发布到应用商城。

6　结语

本文在开发过程中结合心理学专业问卷，实现了高校大学生心理健康服务APP，实现了学生端测试、教师端成绩统计分析以及管理员端管理的基本流程，满足学生、教师以及管理员三种角色不同的使用需求。本文实现的移动端APP使用更为便捷，方便学生随时进行心理测试；功能也更为多样，例如增加了心理健康相关文章、内容的展示功能，有助于提高学生的使用频率，引起学生对心理健康的重视。

参考文献：

[1] 卡德热亚·买买提，安霞霞，阿拉发提·阿布拉江，等. 基于移动互联网环境下的大学生心理健康咨询平台的研究与设计[J]. 电脑知识与技术(学术版)，2020，16(35)：45-46，53.

[2] 陆文颖,冯学强. 国内外高校心理健康教育比较分析及对策研究[J]. 中国教育学刊,2015(S1):260-261.
[3] 许淑琴,陈丽华,哈斯其美格. 高校大学生心理健康教育的系统化探究[J]. 科教导刊,2022(5):130-133.
[4] 黄萍. 中美高校心理健康教育比较研究[J]. 长春教育学院学报,2016,2(32):33-35.
[5] 李力,王鑫. 基于KH-SVM的大学生心理健康评价[J]. 微型电脑应用,2021,37(5):45-47,51.
[6] 赖露珠,吕慧敏,窦颖,等. 2010—2020年国内大学生心理健康研究进展[J]. 校园心理,2022,20(2):152-157.
[7] 房美妍,曾辛茹,许世骅,等. 新媒体微信公众平台在大学生心理健康教育中的作用[J]. 课程教育研究,2019(13):176-177.
[8] 卿再花. 大学新生心理健康现状的调查分析及对策[J]. 世纪桥,2010(23):116-117.
[9] 郝颖. 新时代大学生心理健康教育创新的现实难题与对策[J]. 教育与职业,2020,5(9):107-111.
[10] 王征宇. 症状自评量表(SCL-90)[J]. 上海精神医学,1984(2):68-70.
[11] 时东晓,罗伟雄,刘岚,等. 基于Hybrid的高校移动门户设计与开发[J]. 软件,2019,40(5):140-146.
[12] 钟迅科. 基于HTML5的跨平台移动Web应用与混合型应用的研究[J]. 现代计算机(专业版),2014(7):32-36.
[13] 朱凯南,李艳平,申闫春,等. 基于Ionic和Cordova的跨平台移动APP的研究与应用[J]. 电脑知识与技术(学术版),2016,12(1):119-121.
[14] 张晶,沈东旭,白建,等. 基于混合开发模式的移动校园APP设计探讨[J]. 信息与电脑,2019(2):113-115.

Design and implementation of mental health service APP for college students based on Android system

Liu Xintong[1], Mao Mingzhi[1,2*]

(1. School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510275;
2. School of Software Engineering, Sun Yat-sen University, Zhuhai 519082)

Abstract: Universities all over the world attach great importance to college students' mental health problem, and the development of the Internet has gradually changed the way of college students' mental health testing. In order to help college students obtain more convenient and diverse mental health services and promote the development of mental health education in Colleges and universities, this paper studies the mental health of college students. Based on the psychological questionnaire, a college students' mental health service app system is designed and implemented, which has certain usability.

Keywords: Android; mental health; APP; design and implementation

文章编号: 1007-1423（2023）01-0081-07 DOI: 10.3969/j.issn.1007-1423.2023.01.015

基于Nvidia Jetson的自主环境探测机器人的设计与实现

闫存莹，王福康，田存伟*

（聊城大学物理科学与信息工程学院，聊城 252059）

摘要：移动探测机器人通常需要进入危险复杂的未知环境,代替人进行侦察、取样、救援救灾和应急处置等任务,可以大大降低甚至避免人员伤亡。基于任务需求，设计并实现了一款可用于复杂环境的移动探测机器人。该机器人采用履带作为移动机器人的移动结构以适应复杂的地形，硬件系统包括Nvidia Jetson NANO主控、STM32底盘控制主控、激光雷达传感器等。软件系统基于ROS操作系统和Ubuntu18操作系统开发，可以实现复杂环境下机器人自主扫描环境信息并建立精确地图。

关键词：Nvidia Jetson；SLAM；机器人；自主；环境探测

0 引言

近年来，国内外地震、煤气泄漏、火灾等灾害频发。这些灾害造成了严重的人员伤亡和财产损失。灾害发生后首要的任务是现场状况和被困人员的探测工作。然而这些探测行动都遇到了同一个难题，因灾害过后现场情况复杂（如：有害物质泄露、现场高温、灾后空间狭小），致使救援人员无法进入现场了解情况，救援工作难以展开[1]。目前已开发的探测机器人一种是多自由度、运动灵活，但是体积大、价格昂贵；另一种是体积小巧、运动灵活但功能单一，可靠性不高。为了有效回避以上探测机器人的缺点，本文设计了一种能在多种环境下进行自主探测的机器人，它具有体积小、运动灵活，能适应多种复杂情况、反馈多种信息的优点。

在环境探测机器人领域，为了满足低功耗的要求，大多采用低性能的处理器作为机器人中控。这就导致探测机器人仅能处理简单的遥控指令进行探测和搜索，但在复杂的环境中仅仅通过摄像头传输的图像画面，操控手并不能完全掌握环境的各种信息，导致传统探测机器人效率低下，无法完成较为复杂的探测任务，同时对环境的感知能力也较为薄弱。针对探测机器人工作环境的特点，在保证可靠性、低功耗的前提下选择高性能的处理器来实现复杂的环境感知和自主探测的任务，成为迫切的需求。

本文针对自主探测机器人的工作特点，提出了一种基于Nvidia Jetson主控和激光雷达的可自主建图、导航、探测的履带式机器人系统[2]。并通过gazebo软件仿真和实际测试，能够实现复杂环境下的探测任务。

1 机器人自主探测系统总体设计

1.1 机器人系统总体设计方案

机器人以STM32作为底层控制器，控制底盘的驱动电机、霍尔编码器、IMU惯性传感器模块、电量检测模块等设备运行。将机器人姿态传感器数据和里程计信息以串口通信的方式传送给主控制器进行分析。

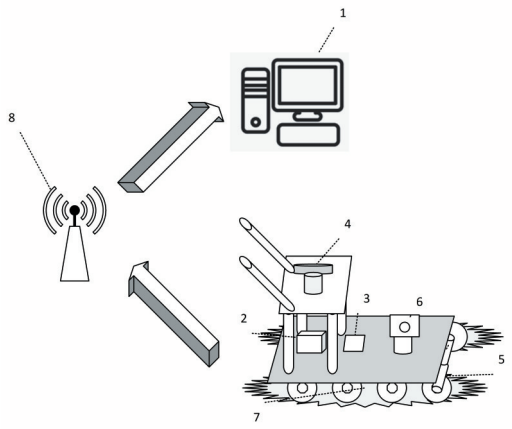
为了满足复杂环境下自主探测的功能，机器人以高性能的Nvidia Jetson NANO作为主控制器，通过对激光雷达的扫描数据进行分析，构

收稿日期：2022-08-01 **修稿日期：**2022-09-03

作者简介：闫存莹（1994—），男，山东聊城人，硕士研究生，研究方向为人工智能；王福康（1999—），男，山东聊城人，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

建未知探测环境的地图，并能通过带有补光灯的高清摄像头将环境图像信息传送回上位机。

机器人以PC为上位机，在Ubantu远程虚拟机下处理机器人传回的各种信息和状态数据、图像数据，并最终通过ROS可视化工具等展示出来。



1—上位机；2—Jetson NANO主控制器；
3—STM32底层控制器；4—激光雷达；
5—带有编码器的驱动电机；
6—带有360°云台的高清摄像头；
7—履带一体式底盘；8—远距离WiFi信号塔

图1 机器人总体设计方案图

机器人的控制方式分为自主搜索模式和手动控制机器人搜索模式。在手动模式下，机器人可根据上位机遥控指令进行运动，将图像信息和激光雷达信息传回上位机[3]，同时构建环境的2D地形地图并通过Rviz可视化软件展示。机器人在自主搜索模式下分为两种情况：①已获取环境的电子地图的情况；②探测环境的地图未知的情况。在①情况下，上位机可在电子地图上设置探索目标点，主控制器根据自身定位和目标地点，使用全局路径规划算法，规划出一条最优路径，底层控制器驱动电机按照规划路径进行运动，运动过程中底层控制器反馈的里程计和姿态信息将进行实时分析以判断运动状态。同时在运动过程中，激光雷达和摄像头实时探测周围环境，并运行gmapping建图算法将环境地图传回上位机，机器人同时根据探测到的周围环境进行避障，每一次避障完成后继续运行全局路径算法按照预设路径前进，直至达到预设目标点。在②情况下，因环境复杂未知且没有电子地图，可将机器人直接布置在需探测的环境中，通过上位机控制机器人运行自主建图程序，先通过对环境的探索构建出环境地图，再重复①情况下的任务模式。

1.2 Nvidia Jetson NANO主控系统设计

1.2.1 Nvidia Jetson NANO

复杂环境下机器人自主探测要求控制器体积小，但涉及到多种传感器的融合数据处理，同时需要运行多个建图和路径规划算法，本设计选取Nvidia Jetson NANO处理器（如图2所示）作为主控制器，可同时满足体积和算力需求[4]。

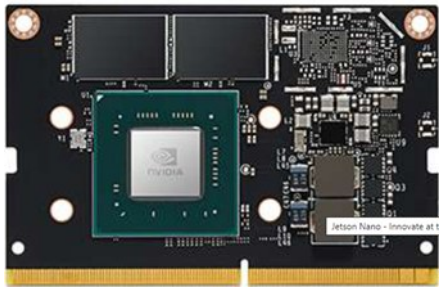


图2 Nvidia Jetson NANO主板

Jetson Nano模块是一款低成本的AI计算机，具备超高的性能和能效，可以运行现代AI工作负载，并行运行多个神经网络，以及同时处理来自多个高清传感器的数据。它专为支持入门级边缘AI应用程序和设备而设计，完善的NVIDIA JetPack SDK包含用于深度学习、计算机视觉、图形、多媒体等方面的加速库。

1.2.2 Ubantu系统和ROS机器人操作系统

Ubuntu是一个以桌面应用为主的Linux操作系统。Ubuntu基于Debian发行版和Gnome桌面环境，而从11.04版起，Ubuntu发行版放弃了Gnome桌面环境，改为Unity。从前人们认为Linux难以安装、难以使用，在Ubuntu出现后这些都成为了历史。Ubuntu也拥有庞大的社区力量，用户可以方便地从社区获得帮助。自Ubuntu 18.04 LTS起，Ubuntu发行版又重新开始使用GNOME3桌面环境。

机器人操作系统（robot operating system，ROS）是一个应用于机器人上的操作系统，它操作方便、功能强大，特别适用于机器人这种多

节点多任务的复杂场景。ROS是一个适用于机器人编程的框架，这个框架把原本松散的零部件耦合在一起，为它们提供了通信架构。ROS虽然叫做操作系统，但并非Windows、Mac那样通常意义的操作系统，它只是连接了操作系统和开发的ROS应用程序，所以它也算是一个中间件，在ROS的应用程序之间建立起了沟通的桥梁，所以也是运行在Linux上的运行环境，在这个环境上，机器人的感知、决策、控制算法可以更好地组织和运行。

1.2.3　ROS程2序开发流程

软件节点设计框图如图3所示。

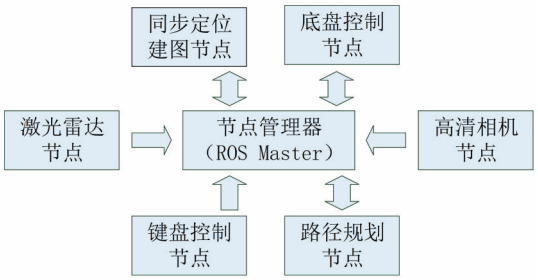


图3　系统软件节点设计框图

本设计根据ROS系统特点，对机器人的功能进行划分，设计了不同的功能节点，如表1所示。

表1　机器人各节点功能

节点名称	功能
节点管理器ROS Master	节点控制中心
底盘控制节点	向上发布速度、姿态信息，向下发布运动控制信息
激光雷达节点	初始化激光雷达设置，发布激光雷达数据
高清相机节点	初始化高清相机参数，发布高清相机数据
同步定位与建图节点	将激光雷达、底盘中IMU和编码器里程计等传感器的信息处理后，调用gmapping实时建立环境地图
路径规划节点	根据上位机发来的目标点坐标，将激光雷达、底盘中IMU和编码器、里程计等传感器的信息处理后，调用路径规划算法进行路径规划和避障
键盘控制节点	通过上位机PC远程控制机器人运动

机器人ROS层程序开发流程图如图4所示，分别为：

（1）启动ROS Master中央节点管理器。

（2）通过roslaunch命令启动编写好的.launch文件，启动多个ROS节点。

（3）与下层控制器进行串口通信，订阅/cmd_vel话题作为控制输入。

（4）判断是否有上位机指令，若收到开始接收数据，未收到则等待。

（5）将底层控制器的数据通过底盘控制节点读取。

（6）判断是上位机指令为SLAM或路径规划，对应执行以下步骤(7)或(8)。

（7）进行SLAM建图：调用激光雷达的雷达节点、同步定位与建图节点、底盘控制节点进行SLAM建图。

（8）进行路径规划：调用激光雷达的雷达节点、路径规划节点、底盘控制节点进行SLAM建图。

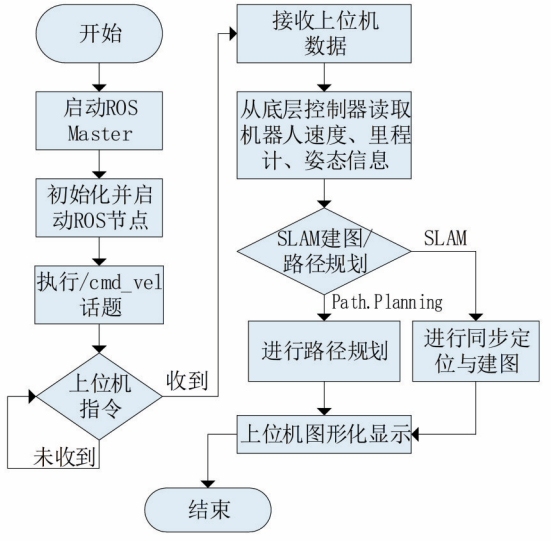


图4　机器人ROS层程序开发流程图

2　机器人的自主建图与路径规划

2.1　SLAM地图构建算法——gmapping

Simultaneous是实时同步，Localization是在给定地图的情况下，估计机器人的位姿，Mapping是在给定智能车位姿的情况下，建立环境的地图。SLAM算法就是指同时估计机器人的位

姿并且建立环境地图。机器人在移动过程中需要导航和路径规划，SLAM主要解决智能车在自主移动过程中的定位和建立地图的问题，获取地图信息从而进行路径规划避开障碍物[5]。

考虑到机器人探测环境大多为狭小空间，即需要构建小场景高精度的地图，同时还要尽量减小计算量。本设计采用的SLAM算法为gmapping，该算法是一个比较完善的地图构建开源包，使用激光和里程计的数据来生成二维地图。

SLAM程序设计流程图如图5所示。

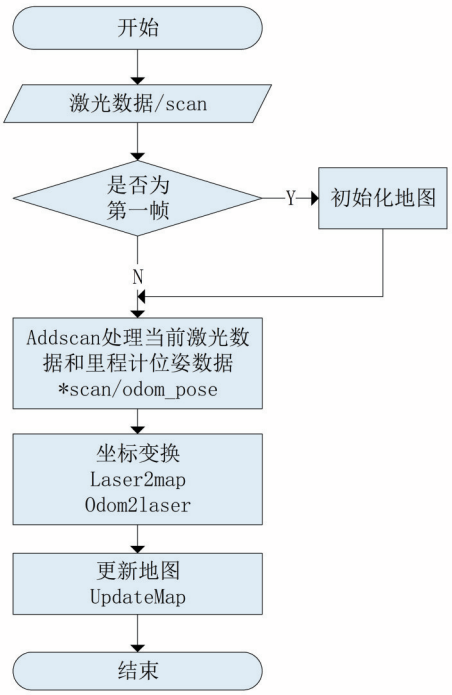


图5 SLAM程序设计流程图

2.2 路径规划算法——DWA

动态窗口法(DWA)主要是在速度(v,w)空间中采样多组速度，并模拟机器人在这些速度下一定时间(sim_period)内的轨迹。在得到多组轨迹以后，对这些轨迹进行评价，选取最优轨迹所对应的速度来驱动机器人运动。

DWA算法将避障问题转化为速度空间中的三个约束条件：①移动机器人的最大和最小速度约束;②移动机器人的动态性能约束；③移动机器人的安全约束[6]。

机器人的采样速度应该控制在移动机器人自身的最大速度与最小速度区间内，v表示线速度，w表示角速度，则约束公式：

$$V_t=\left\{(v,w)\,\middle|\,v_{\min}\leqslant v\leqslant v_{\max},w_{\min}\leqslant w\leqslant w_{\max}\right\} \tag{1}$$

由于机器人动力学的限制，线加速度$\dot{v}$和角加速度$\dot{w}$存在上限和下限，v_c,w_c指当前时刻的线速度、角速度，Δt为时间间隔，因此满足：

$$V_d=\left\{(v,w)\,\middle|\,\begin{matrix}v\in\left[v_c-v\Delta t,v_c+v\Delta t\right],\\ w\in\left[v_c-v\Delta t,v_c+v\Delta t\right]\end{matrix}\right\} \tag{2}$$

基于移动机器人的安全性考虑，防止机器人因速度因素而撞上障碍物，提出了满足机器人运动安全性考虑的速度空间：

$$V_a=\left\{(v,w)\,\middle|\,\begin{matrix}v\leqslant\sqrt{2\mathrm{dist}(v,w)\dot{v}},\\ w\leqslant\sqrt{2\mathrm{dist}(v,w)\dot{w}}\end{matrix}\right\} \tag{3}$$

式中，$\mathrm{dist}(v,w)$即为当前轨迹距最近障碍物的距离。该碰撞条件是在将机器人轨迹模拟出来以后，根据机器人和障碍物之间的距离，计算当前采样的速度是否可以在碰到障碍物之前停止，若可以停止，则该速度即为允许的速度[7]。

最终机器人动态窗口的速度范围应满足：

$$V_r=V_t\cap V_d\cap V_a \tag{4}$$

3 系统实验测试与分析

本次系统实验分为两部分，第一部分是检验软件系统的整体性和算法的可行性，在Gazebo仿真软件上所做的系统仿真实验；第二部分是实物实景测试。

3.1 仿真实验环境

整个仿真系统在Ubantu18.04系统上运行，ROS版本为Melodic版本，系统配置如表2所示。

表2 Ubantu系统配置信息

参数名称	参数
CPU	AMD Ryzen 5800X
CPU	NVIDA GeForce RTX 3060
内存	32 GB DDR4
硬盘	1 T SSD
操作系统类型	64 bit

基于ROS和Gazebo的无人船传感器融合仿真环境框架如图6所示。

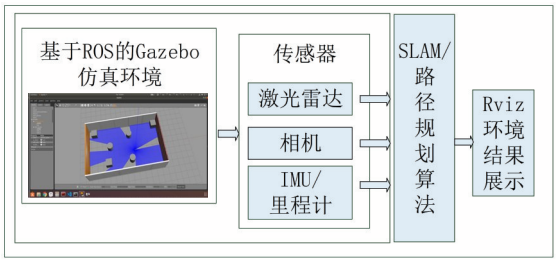


图6　基于ROS和Gazebo的无人船传感器融合仿真环境

3.2　实验结果

图7为Gazebo仿真环境下的模拟环境，设置圆柱体和正方体模仿障碍物。

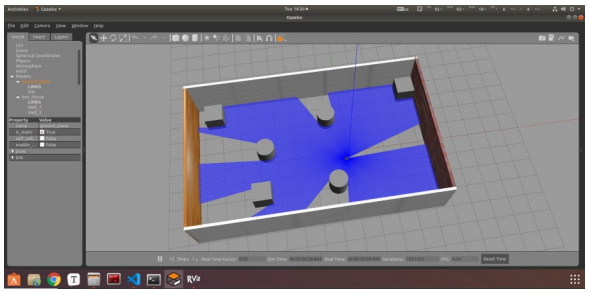


图7　Gazebo仿真环境下的模拟环境

运行SLAM算法进行建图，并在Rviz软件中显示建图结果，如图8所示。

在图8中，左下角显示为摄像头的实时图像数据，建图中边缘处红色点为激光雷达扫描点，根据实验结果可知本设计采用的SLAM算法达到预期目标。

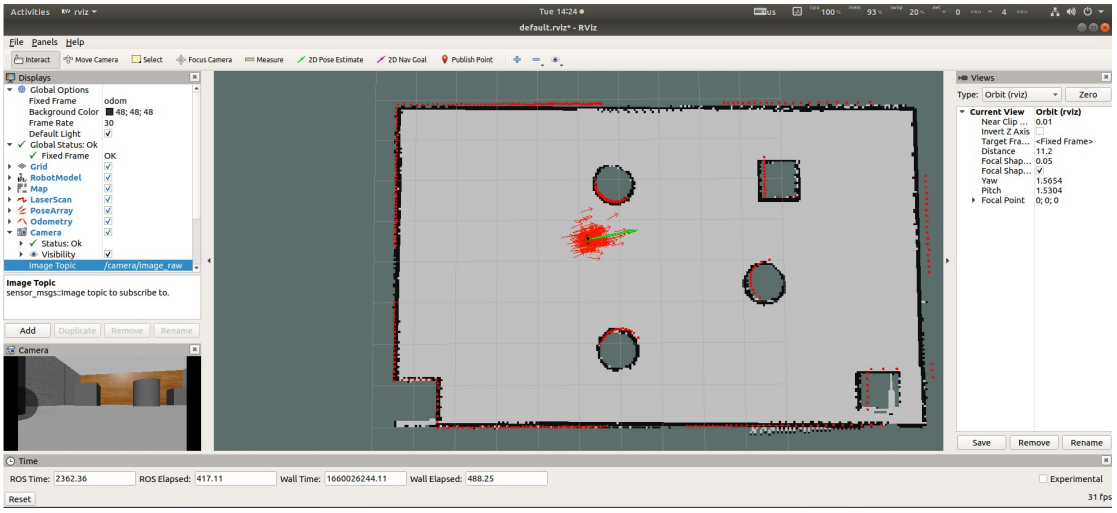


图8　Rviz软件中显示建图结果

在完成SLAM建图的基础上，我们对路径规划算法进行测试，在地图中设置目标点位，由机器人自主规划路径并前往。运行结果如图9所示。

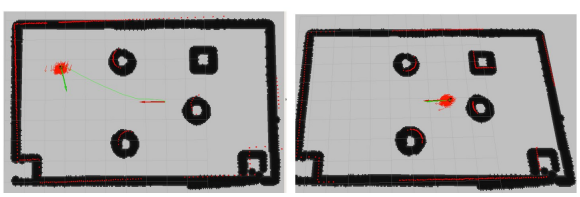


图9　路径规划运行结果

在图9可以看到，设置目标点后，路径规划算法规划出一条路径，如图中绿色轨迹所示，并最终到达设置目标点。根据实验结果可知路径规划算法达到预期目标。

3.3　实测环境

实测场景如图10所示。

机器人配置如表3所示，上位机配置如表4所示。

图 10 实测场景

表 3 机器人参数

参数名称	参数
主控制器	Jetson NANO
底层控制器	STM32 F103c8t6
激光雷达	思岚A1激光雷达
相机	高清摄像头(1080P)
相机云台	360°两轴云台
电机	520编码器电机*2
操作系统类型	Ubantu18.04 64bit

表 4 上位机配置

参数名称	参数
CPU	AMD Ryzen 5800X
CPU	NVIDA GeForce RTX 3060
内存	32GB DDR4
硬盘	1T SSD
操作系统类型	Ubantu18.04 64bit

3.4 实验结果

SLAM建图结果：运行SLAM建图程序，对实验环境进行扫描，结果如图11所示。

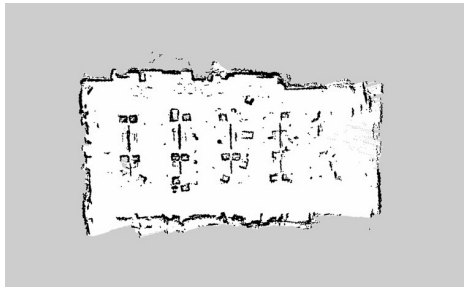


图 11 实验环境扫描图

导航及路径规划运行结果如图12所示，步

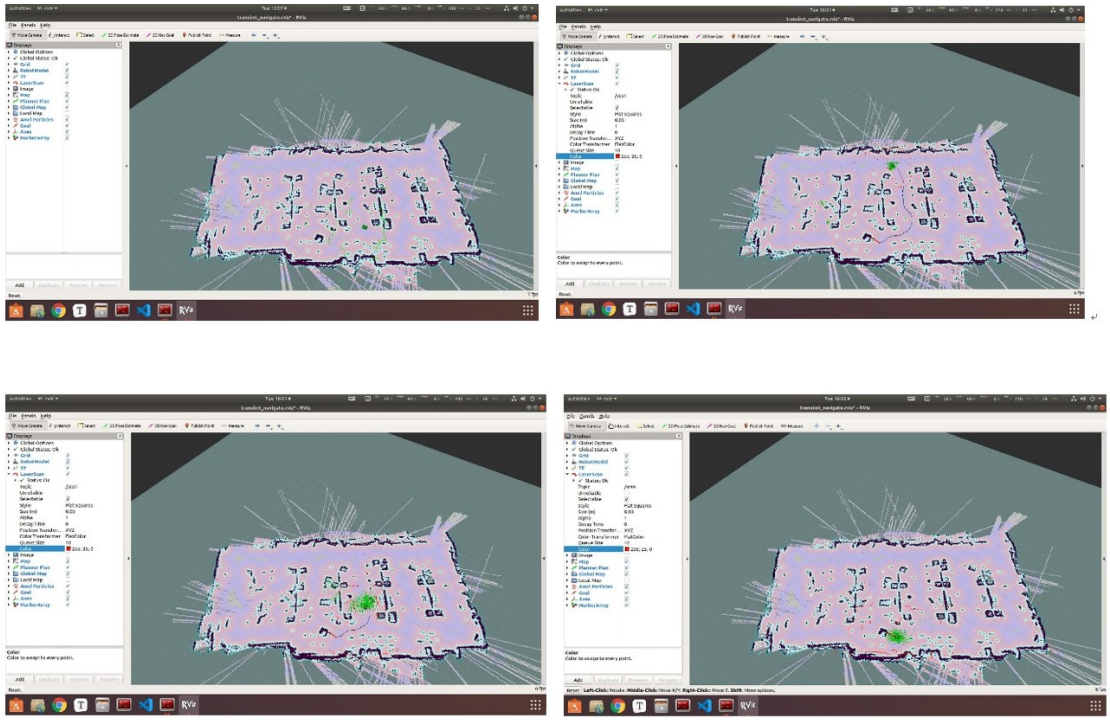


图 12 导航及路径规划结果

骤为导入地图，运行路径规划程序，设置目标点，展示规划路径，驱动机器人前往目标点。经过测试，机器人根据规划路径顺利到达目标点。

4　结语

本文提出了一种基于Nvidia Jetson的自主环境探测机器人，设计了机器人的软件系统和硬件系统，并通过SLAM算法和路径规划算法实现了机器人的自主环境探索功能，并通过Gazebo仿真软件对软件系统的可行性进行了测试。实验结果表明，本设计的软件系统设计达到应用目标。通过对实验室环境的实际测试，将机器人置于较为复杂的环境中进行实际测试，表明机器人的整体设计完备，具备实际应用价值。而对于地图构建的精度问题，未来还将通过继续优化建图算法和提高硬件性能来获取更好的建图效果。

参考文献：

[1] 李欢欢. 复杂环境移动探测机器人测控系统研究[D]. 南京：东南大学，2019.

[2] 丁奇安，刘龙申，陈佳，等. 基于Jetson Nano的哺乳期仔猪目标检测[J/OL]. 农业机械学报：1-12[2022-08-11]. http://kns. cnki. net/kcms/detail/11.1964.S.20220126.1932.034.html.

[3] 张堡瑞. 基于激光雷达扫描和视觉识别结合的水面环境探测[D]. 绵阳：西南科技大学，2022.

[4] 赵威，李振华. 基于ROS系统和SLAM技术的智能小车的研究[J]. 物联网技术，2022，12(7)：97-99.

[5] 梁双，倪晓昌，董娇玲，等. 基于激光雷达的SLAM算法综述[J]. 信息与电脑(理论版)，2022，34(3)：59-61.

[6] 王豪杰，马向华，代婉玉，等. 改进DWA算法的移动机器人避障研究[J/OL]. 计算机工程与应用：1-8[2022-08-11]. http://kns.cnki.net/kcms/detail/11.2127.TP.20220629.1043.005.html.

[7] 常新新，胡为，姬书得，等. 基于改进动态窗口法的移动机器人避障研究[J]. 组合机床与自动化加工技术，2021(7)：33-36，39.

Design and implementation of autonomous environment detection robot based on Nvidia Jetson

Yan Cunying, Wang Fukang, Tian Cunwei*

(School of Physical Science and Information Engineering, Liaocheng University, Liaocheng 252059)

Abstract: Mobile detection robots usually need to enter dangerous and complex unknown environment, instead of human reconnaissance, sampling, rescue and disaster relief and emergency disposal tasks, can greatly reduce or even avoid casualties. Based on the task requirements, a mobile detection robot which can be used in complex environment is designed and implemented. The robot uses track as the mobile structure of the mobile robot to adapt to complex terrain. The hardware system includes Nvidia Jetson NANO main control, STM32 chassis control main control, liDAR sensor and so on. The software system is based on ROS operating system and Ubuntu18 operating system. The robot can scan the environment information and build accurate map in complex environment.

Keywords: Nvidia Jetson; SLAM; robotic; autonomous; environment exploration

文章编号: 1007-1423(2023)01-0088-07 DOI: 10.3969/j.issn.1007-1423.2023.01.016

袖带式血压信号采集系统设计

向蓓蓓，郭　妍，谢光敏，陈吉鑫

（四川水利职业技术学院信息工程学院，成都 610039）

摘要：血压值是反映人体健康的重要参数，对于心脑血管病的治疗能够提供有力依据，因此血压测量值的精准度在医学临床上尤为重要，为了能够实时了解到血压值，医学领域关于血压的测量一直在不断研究与开创。在各类研究基础上，采用了示波法的血压测量方法，设计出基于此方法的袖带式血压信号采集系统，相当于一个模拟电子血压计。该血压计能够实现血压信号采集，并且能够在血压超出正常范围时发出报警信号。

关键词：血压测量；AT89C52；脉搏波；示波法

0　引言

随着社会经济突飞猛进的发展和人民生活水平的日益提高，人们患上血液疾病的概率越来越高，众所周知，血液的容积、血管硬化的程度、心脏功能、内分泌、运动、情绪等都是影响血压的因素。要想了解到底是什么因素，到底影响多大，还得依靠血压的精确测量加以判断。生理学家 Harrey 在 1628 年提出了血液循环理论，在那之后的几百年来，找到既方便有效又准确可靠的血压测量方法便成为研究者们一直追求的目标，时至今日，血压测量的方法已经有很多，但每种方法都存在一定的缺陷，始终差强人意。基于袖带充放气的示波法对比电子柯氏音法，具有很明显的优点，示波法的抗干扰性强，重复性好[1]，因此是电子血压计和血压监护仪应用最广泛的方法。基于此，本文旨在通过袖带式血压信号采集的设计，采用示波法原理进行血压测量，对比日常中各类血压测量技术，找到最准确、最便捷的检测方式，实现对血压精确的测量。

1　方案介绍

1.1　系统构成

设计的模拟电子血压计的主要组成部分包括单片机、压力传感器、初级次级运算放大电路、滤波电路、数模转换电路、显示电路。首先通过压力传感器采集血压信号，血压信号通过运算放大电路和滤波电路的预处理，从而分离出两路信号，模数转换电路再对两路信号进行模数转换，最后由单片机来完成信号的储存与计算[2]，计算所得到的数据控制血压计的外围电路，从而通过显示电路对收集到的血压数据的数值进行显示。实验所设计的整个血压采集系统主要由以下几部分组成。

1.1.1　信号采集

由压力传感器、前置放大器、带通与低通滤波器构成，主要功能是分离出袖带压力信号和脉搏波信号。

1.1.2　模数转换

由多路 D/A 转换器构成，主要是将袖带压力信号和脉搏波信号转换成数字信号。D/A 转换器为单片机的外设。

1.1.3　中央处理单元

系统的中央处理单元主要部分采用的是单片机 AT89C52，其主要的功能：微控制器接收特定的系统参数设置和来自微控制器的外部控

收稿日期：2022-08-13　**修稿日期：**2022-09-06

作者简介：向蓓蓓(1992—)，女，重庆潼南人，本科，助教，研究方向为电路与系统、信号与信息处理；郭妍(1994—)，女，吉林白山人，硕士，研究实习员，研究方向为图像处理、模式识别；谢光敏(1996—)，女，四川资阳人，硕士，助教，研究方向为图像篡改、密码学、Petri 网；陈吉鑫(1994—)，女，四川成都人，本科，助教，研究方向为财务管理、管理信息系统

制命令，并将测量数据传送回主微控制器。这些数据为血压计算提供了具体信息，包括收缩压、舒张压、平均血压数据、气泵模数转换控制以及测量过程中各种算法的实施。微控制器控制气泵的填充和排空过程，并实现对报警系统的控制。

1.1.4　显示模块

显示电路主要采用LCD1602，用于显示舒张压、收缩压和平均血压值。

1.2　系统构成框图

根据系统构成框图(如图1所示)，该系统的整个工作流程如下：手臂套上袖带，袖带充气放气，产生气压。压力下血管的血流产生振荡波，通过气管被压力传感器接收。压力传感器将接收到的振荡波信号传送到差分放大电路。差分放大器将振荡波信号分成两路，根据振荡波信号的不同频率，两路信号被一个用于分离袖带压力信号的低通滤波器和一个用于分离脉冲波信号的带通滤波器分开，并一起传送到D/A转换器，经由数字信号电路进行处理之后，由单片机计算出舒张压、收缩压和平均血压值[2]。计算数据输入至LCD显示。如果测量值超出正常范围，报警信号通过显示模块LCD显示为“!!!”。

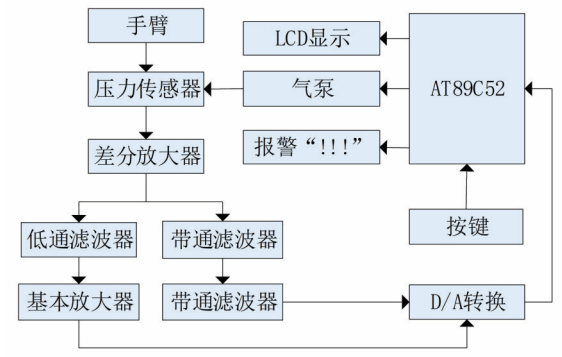


图1　系统构成框图

2　硬件设计

2.1　压力传感器BP01

采集血压信号，系统用到的是压力传感器BP01，压力传感器BP01具有高线性、低噪声的特性。通过温度补偿和内部标定的方式来提升测量精准度、系统稳定性。全范围内精度为±1%，零点失调不超过±300 μV。BP01的电路连接如图2所示。

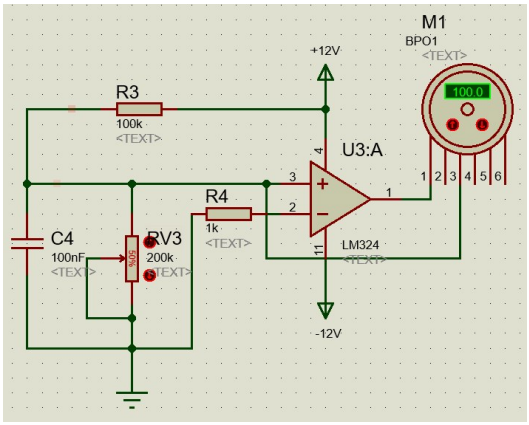


图2　BP01电路连接

2.2　信号预处理电路

2.2.1　前置放大电路

动脉血压的参数范围为10～400 mmHg，压力传感器BP01输出的差分信号为0～15 mV，无法对生理信号进行直接的记录与显示，因为生理信号幅度偏小、频率偏低、内阻高、噪声大、抗干扰性弱[3]。为了克服生理信号的这些特性，就要求初级放大电路有一个合适的通带宽度和动态范围，必须具备较高的共模抑制比、低噪声、高增益以及高输入阻抗，要达到这些高要求，前置放大器将是最好的选择。

放大电路中，U3:B和U3:C连接成同相输入的形式，能够对电压起到缓冲作用，同时增大输入阻抗。U3:D则构成差动放大器，提供差模增益，对从U3:B和U3:C传送来的共模电压起到抑制的作用，从而提高共摸抑制比。前置放大电路图如图3所示。

2.2.2　分离袖带压力信号和脉搏波信号

袖带压力信号变化非常缓慢，属于低频信号，因而需要采用低通滤波器来提取出袖带压力信号。本次实验设计采用的是二阶有源低通滤波器，如图4所示。

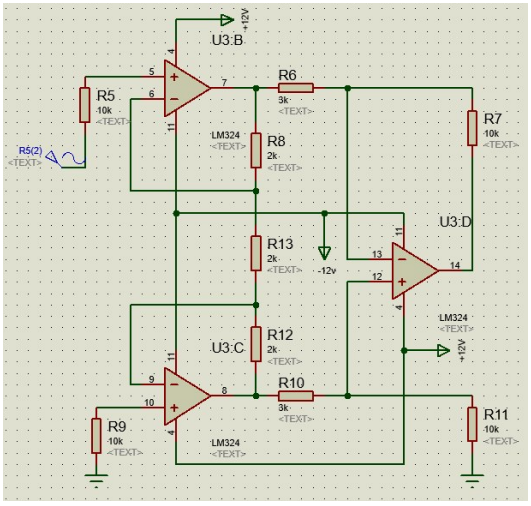


图 3　前置放大电路

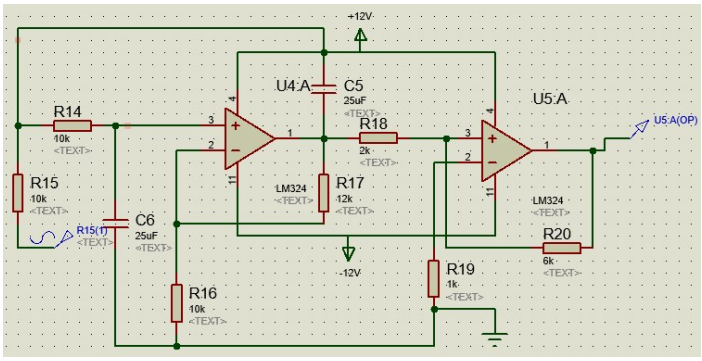


图 4　二阶有源低通滤波器及次级放大电路图

通常情况下，脉搏波信号的频率在0.6～6.4 Hz范围内，脉搏波信号还掺杂着一些高频干扰和低频分量。基于此，本实验设计采用带通滤波器进行信号除杂优化，采用两级带通滤波器的级联来提高增益。两级级联能使频率响应更加尖锐，从而提高输出的信噪比。获取脉搏波信号，将频率设定为0.4～6.6 Hz，电路图如图5所示。

一级带通滤波器：

$$\text{高截止频率 } f=\frac{1}{2\pi(0.2\ \mu\text{F}\times 120\ \text{K})}=6.63\ \text{Hz}$$

$$\text{低截止频率 } f=\frac{1}{2\pi(47\ \mu\text{F}\times 10\ \text{K})}=0.338\ \text{Hz}$$

$$\text{放大倍数 } A=-\frac{120\ \text{K}}{10\ \text{K}}=-12$$

二级带通滤波器：

$$\text{高截止频率 } f=\frac{1}{2\pi(0.24\ \mu\text{F}\times 33\ \text{K})}=19.91\ \text{Hz}$$

$$\text{低截止频率 } f=\frac{1}{2\pi(47\ \mu\text{F}\times 10\ \text{K})}=0.338\ \text{Hz}$$

$$\text{放大倍数 } A=-\frac{33\ \text{K}}{10\ \text{K}}=-3.3$$

综上可得总体放大倍数$A=15\times 12\times 3.3=594$

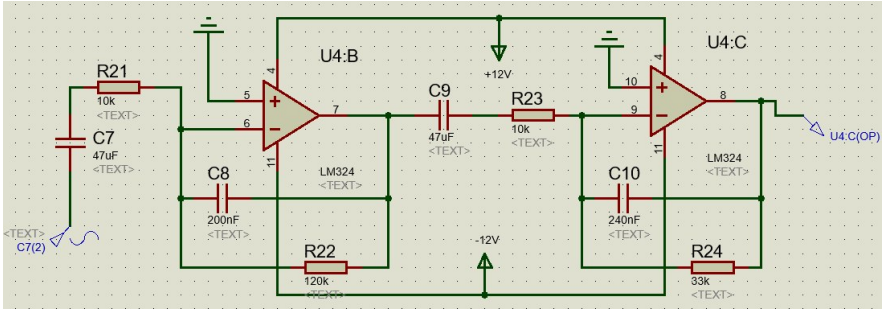


图 5　两带通滤波器的级联

2.3　模数转换电路

2.3.1　ADC0808模数转换芯片

本设计用到的模数转换芯片是ADC0808，如图6所示。它的构成包括一个地址锁存器、一个译码器、一个8路的模拟开关、一个D/A转换器和一个三态输出锁存器[4]。各引脚功能如下：

START：启动D/A转换的信号输入端，上跳沿时，所有内部寄存器处于清零状态；下跳沿时，开始进行D/A转换；在转换期间，START应保持低电平。

EOC：转换结束信号的输出引脚，低电平时开始转换，高电平时转换结束[5]。

OE：输出允许控制端，用于控制三条输出锁存器向单片机输出转换得到的数据。OE=1，输出转换得到的数据；OE=0，输出数据线呈

高阻状态。

ALE：地址锁存允许信号输入端，高电平时有效，此时地址锁存连同译码器将锁存3位地址线的地址信号，经译码的通道的模拟量被选中后被转换器转换。

IN0～IN7：8位模拟量输入引脚。

OUT1～OUT8：8位数字量输出引脚。

A、B、C：对IN0—IN7进行选通的一路模拟量输入，是3位地址输入线。

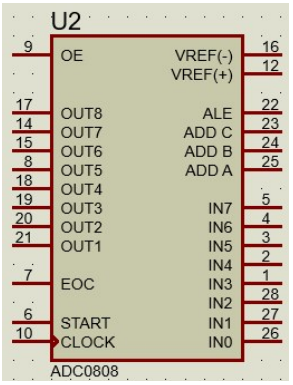


图6　ADC0808模数转换芯片

2.3.2　单片机AT89C52

单片机AT89C52是本设计的中央处理单元，如图7所示。内部结构与MCS-51系列单片机的构成基本相同，但数据存储器与51系列的其他单片机相比大很多，存储容量为256字节，单片机AT89C52的指令系数和引脚功能与MCS-51的完全兼容[6]。各引脚功能如下。

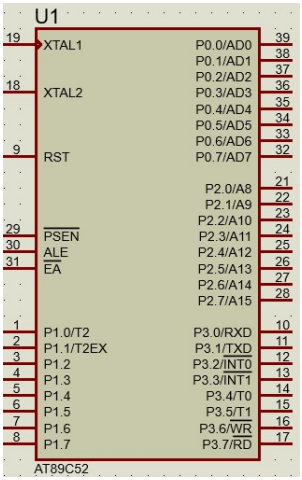


图7　AT89C52

XTAL1：振荡器反相放大器的输入端。

XTAL2：振荡器反相放大器的输出端。

RST：复位输入。

P0口：是一个8位漏级开路的双向I/O口。每位能驱动8个TTL逻辑电平。当输入高电平时，引脚作为高阻抗输入。当访问外部程序和数据存储器时，P0口也被作为低8位地址/数据复用[7]。

P1口、P2口、P3口：是具有内部上拉电阻的8位双向I/O口，P口输出缓冲级可驱动(吸收或输出电流)4个TTL逻辑电平[7]。当输入高电平时，内部上拉电阻将拉高端口，此时可作为输入口。作为输入使用时，被外部拉低的引脚由于内部电阻的原因，将输出电流ILL。

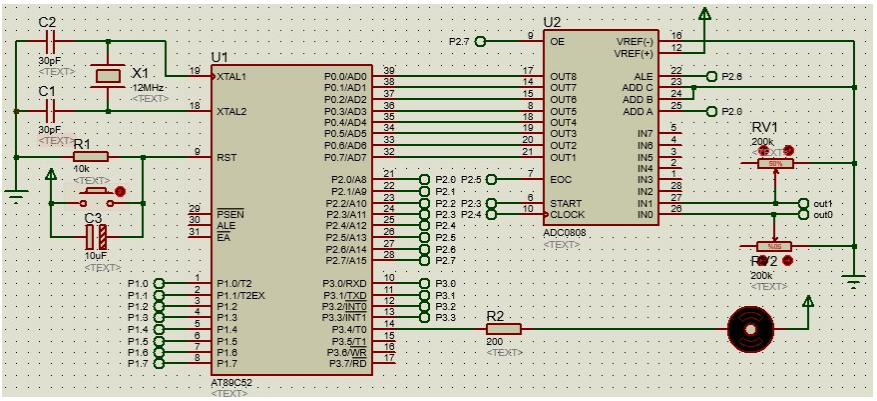


图8　D/A转换电路

2.3.3 ADC0808与单片机AT89C52的连接

模拟信号通过output0和output1进入ADC-0808，经模数转换之后，数字信号由OUT1-8输出到单片机P0口。如图8所示。

2.4 数据显示部分

在选择显示器上，考虑到仿真结果的血压计数据显示。数码管只能显示数字，一次只能显示一个内容，需要按键切换，操作相对麻烦。LCD显示的内容较多，一次能把全部需要的内容显示出来，操作更方便，看起来也更直观。另外LCD工作电流不大、体积轻巧、功耗低、寿命长，显示效果逼真[8]，因此本设计的数据显示采用的是LCD1602。

3 系统软件设计

3.1 系统软件设计流程

系统软件设计流程如图9所示。开始时，系统首先对按键进行判断。如果按键按下，单片机的控制口将启动气泵充气；若按键无效，则返回初始化重新判断。当ADC信道0所测的血压直流量大于4 V时，气压足量则充气结束，气泵开始放气，此时气压大致为180 mmHg或略微大于180 mmHg；若血压直流量不大于4 V，则返回初始化重新判断。通过ADC信道1采集血压交流分量测量出脉冲的峰峰值，并计算ADC通道0在该脉冲期间测量的袖带压力信号的平均值。该峰峰值和袖带压力信号平均值作为测量的一对数据，每个心跳脉冲对应这样的一对数据。

由ADC信道0测到的血压直流分量小于1 V就表示气压低于50 mmHg，单次测量结束。从记录下来的多对由峰峰值和袖带压力信号的平均值组成的数据中找到最大的峰峰值，再从这些数据中找出最接近最大峰值0.5倍的一对数据，对应的血压直流分量就为收缩压；最接近最大峰值0.7倍的一对数据，对应的血压直流分量就为舒张压。然后对测出的血压值是否在正常值范围内进行判断。本实验设计的合理的测量范围是收缩压为90～140 mmHg，舒张压为65～90 mmHg。如果测量数据不在合理范围内，将通过LCD显示出报警信号“!!!”。

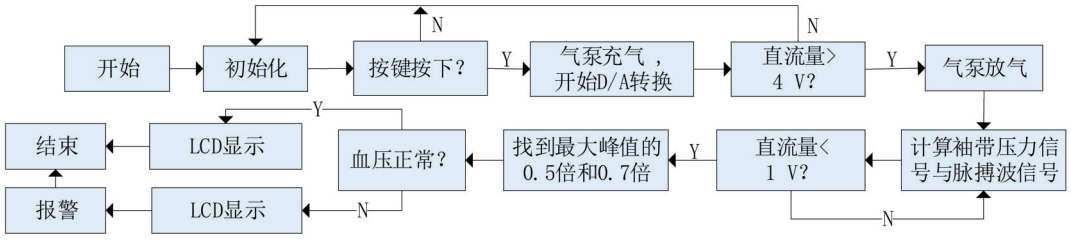


图9 系统软件设计流程图

3.2 模数转换的程序流程

系统在气泵开始充气时就开始了D/A转换，D/A转换与充气放气同时进行。单片机采用查询P2.5口来判断转换是否结束，一旦模数转换结束，P2.5接收到高电平，系统检测到高电平后开始读入数据。模数转换的程序流程如图10所示。

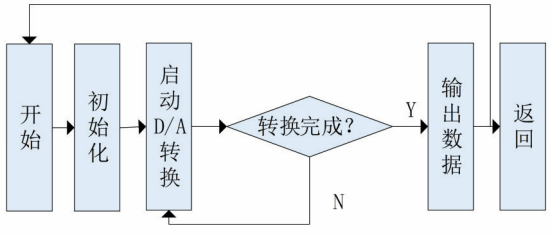


图10 模数转换程序流程图

4　调试与仿真

系统在调试仿真部分采用的是Keil uVision2和Proteus ISIS。按下按钮，气泵充气。血压直流量大于4 V时，气泵均匀放气，开始测量交流信号；当血压直流量小于1 V时，测量结束，显示模板显示测量值。

（1）血压测量值在系统设计的合理测量范围内，仿真结果如图11所示。

（2）血压测量值不在合理范围内，LCD显示屏上显示报警信号“!!!”。仿真结果如图12所示。

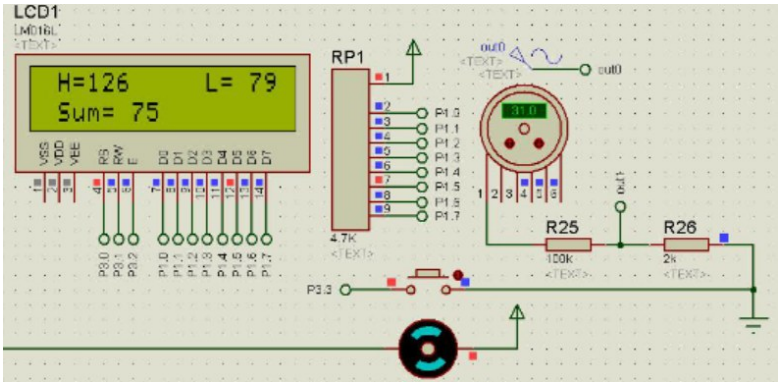


图11　血压正常显示图

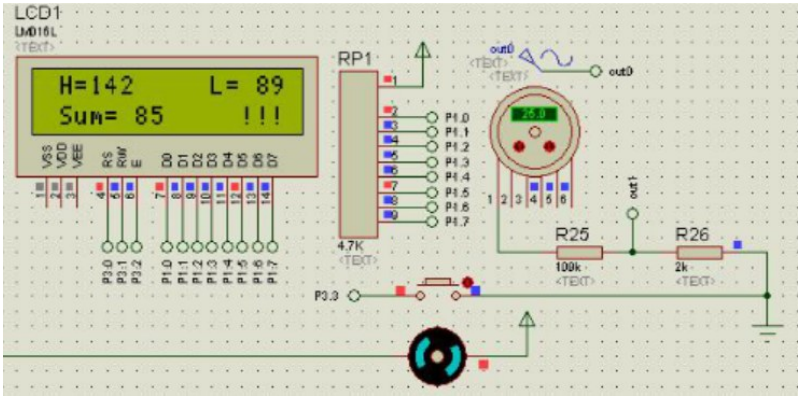


图12　血压偏高报警电路图

5　结语

本设计以单片机AT89C52为控制核心，利用单片机的高度集成化，大大简化了实验电路图，再辅以压力传感器BP01、气泵、报警、外围的模拟电路以及LCD驱动芯片，设计出简易的血压测量计，其基本工作过程为：压力传感器采集信号，通过差分放大器分离出袖带压力信号和脉搏波信号，再对这两种信号作滤波处理，之后所得的信号通过模数转换后，单片机AT89C52再计算分析数字信号，最后得出收缩压、舒张压，通过LCD显示。在软件设计中，采用数字信号处理技术对D/A采样的信号进行处理，并结合相应的算法及编程方法实现计算血压值的程序。

参考文献：

[1] 张畅，陈辉，郑秀娟. 基于优化脉搏波特征的无袖带血压检测方法[J]. 电子测量技术，2021，44(24)：1-7.

[2] 张铭. 基于脉搏波传导时间的无袖带血压测量仪设计[D]. 长春：吉林大学，2014.

[3] ALPERT B S，QUINN D，GALLICK D. Oscillometric blood pressure：a review for clinicians[J]. Journal of the American Society of Hypertension，2014，8

(12):930-938.
[4] 邓根强. 基于脉搏波的连续血压测量系统研究及设计[D]. 深圳:深圳大学,2017.
[5] 李玉俊,黄琪. 基于单片机的高精度AD转换器[J]. 电子设计工程,2011,19(17):21-23.
[6] 王德坤. 礼仪机器人控制系统的设计与实现[D]. 广州:广东工业大学,2015.
[7] 罗淳. 模糊自整定PID温室温度控制器的设计[D]. 武汉:武汉科技大学,2009.
[8] 沈倩倩. 基于51单片机LCD12864左右半屏交接处位置文字显示研究[J]. 滁州职业技术学院学报,2021,20(4):56-59,99.

Design of cuff blood pressure signal acquisition system

Xiang Beibei, Guo Yan, Xie Guangmin, Chen Jixin

(School of Information Engineering, Sichuan Water Conservancy Vocational College, Chengdu 610039)

Abstract: Blood pressure is an important parameter in response to human health, and can provide a strong basis for the treatment of heart blood vessel of brain vigorous sickness.The blood pressure measurement accuracy in clinical medicine is particularly important, In order to know the value of blood pressure in real time, the measurement of blood pressure in the field of medicine has been continuously studied and created. This design is in all kinds of research based on the wave method of blood pressure measurement methods, design a cuff blood pressure signal acquisition system based on this method, quite in a simulated electronic blood pressure meter. The blood pressure monitor can realize blood pressure signal acquisition, and can send out the alarm signal in the blood pressure outside the normal range.

Keywords: Blood pressure measurement; AT89C52; pulse wave; wave method

(上接第68页)

Research and analysis of image recognition technology based on convolutional neural network

Gong Jiaming

(Institute for Data Sciences, Guangzhou Huashang College, Guangzhou 511300)

Abstract: Deep learning ultimate goal is to enable the machine to have the ability to analyze and learn like people, and to recognize data such as words, images information and sounds. Image recognition technology refers to reading and identifying the useful information in the image, judging and classifying the extracted information to reduce the labor intensity and improve the efficiency. In this paper, the image preprocessing technology and convolution neural network technology are studied to realize the recognition of multiple object points in the image, and the state of key objects in the image is studied and judged.

Keywords: deep learning; image preprocessing; image recognition; convolution neural network

文章编号: 1007-1423(2023)01-0095-06 DOI: 10.3969/j.issn.1007-1423.2023.01.017

基于Rasa框架的智能对话系统研究与设计

李孟全，吴士泓，王志刚

（远光软件股份有限公司，珠海 519085）

摘要：随着预训练语言模型和知识图谱技术的快速发展，人机对话技术目前已成为人工智能领域新的研究热点，受到学术界和工业界的广泛关注。人机对话系统分为任务型和非任务型两类，其中任务型对话系统用于帮助用户完成某些特定任务，对其进行深入研究将具有很高的学术价值和应用价值。本文将利用Rasa开源框架研究和设计一套智能对话系统，帮助公司完成提质增效管理目标，提升公司整体智能化管理水平。

关键词：对话系统；自然语言处理；Rasa

0 引言

在企业推进以人机协同系统为基础的智能化进程中，任务型对话系统是其中最核心的领域，需要各方深入研究和推动。赵阳洋等[1]认为，尽管对话系统近年来有较大的发展，能够帮助用户完成简单的任务，但其在通用性、深度理解等方面还面临着许多挑战。笔者研究了Rasa开源框架，发现Rasa在通用性和深度理解方面具备一定的优势，该框架设计高效、灵活，提供了对话机器人开发时必要的基础架构和工具，尤其是Bunk等[2]基于Rasa框架提出的DIET模型，能够对用户意图和实体信息进行联合建模，且Rasa整体上与NLP技术的最新研究进展融合性很好。所以，本文将基于Rasa框架深入研究和设计一套智能对话系统，以便减轻相关人员的工作压力，满足公司“提质增效”的管理需求，为企业数字化转型奠定坚实基础。

1 研究现状和问题

1.1 对话技术研究现状

对话技术大体上经历了三个发展阶段，分别是基于符号规则和模板阶段、基于传统机器学习阶段、基于深度学习阶段。

基于符号规则和模板阶段。该阶段的技术严重依赖专家手工制定的语法规则和本体设计，优点是容易理解，缺点是人工成本非常高、扩展性差。例如，Goddeau等[3]提出的有限状态机方法需要人工预先定义好全部状态和状态转移条件，采用NLU输出中得分或概率最高的结果才能完成对话状态更新，该方法缺点是无法解决错误识别问题，且在复杂应用场景下灵活性不足。后来，Sun等[4]提出用于修正NLU中错误识别的N-Best列表分数方法，但缺点也同样明显：相关参数依赖人工编制，不能根据数据分布情况进行学习。

基于传统机器学习阶段。该阶段的技术方法包括K近邻、决策树、朴素贝叶斯、逻辑回归、集成学习、条件随机场、支持向量机(SVM)等。该阶段的技术虽然具备弱学习能力，但漏洞修补难度大，也不利于模型大规模应用。例如，Lewis等[5]、Zhang等[6]认为朴素贝叶斯模型的独立假设条件比较苛刻，在实际场景中应用价值不大；Yang等[7]、Weinberger等[8]认为K近邻算法在结果一致性方面简单、有效，且重

收稿日期：2022-08-13 **修稿日期：**2022-09-19

作者简介：李孟全(1980—)，男，山东定陶人，硕士，研究员，研究方向为人机对话、机器阅读；吴士泓(1978—)，男，江西景德镇人，博士，高级研究员，研究方向为人工智能和智能系统；王志刚(1986—)，男，湖北黄冈人，博士，研究员，研究方向为自然语言处理、知识图谱

新训练代价低，但其性能过于依赖模型查找到的最近邻距离度量，且面对高维数据时处理效果不佳。Joachims[9]指出SVM应对高维空间和过拟合时表现出来的稳健性不错，但其性能又高度依赖于相关参数和核函数的选择。Lubis等[10]通过变分自编码潜在动作空间对对话策略进行优化，主要探索了预训练、动作知情的先验和多任务学习三种借助辅助任务塑造潜在变量分布的方法，优点是与目前的Transformer不同，不需要额外的数据，且计算成本低，在端到端对话任务上效果显著，但在长距离上下文和对话轨迹准确性方面有待完善。

基于深度学习阶段。该阶段的技术促使了端到端的学习成为可能，是目前研究的主流方法，也涌现了大量的研究成果，当然也存在一定程度的不足。目前在处理序列标注任务时会在传统序列标注方法基础上加入深度学习方法，例如，Ma等[11]、Dilek等[12]提到的具有条件随机场层的双向LSTM，成为目前处理序列标注任务的主流模型，既可以兼顾上下文又可以兼顾依赖关系信息，满足过去和未来的信息均可被充分利用的要求，但该方法有时候还是需要人工构建特征；其他方面，冯强中[13]利用知识图谱和神经网络(主要是双向LSTM)进行知识增强，将复杂且非标准的自然语言查询语句转换为符合公司知识库标准的问题，再结合知识库中的问题、相似问题和答案等实体构建知识图谱，完成智能问答工作，但不足之处是模型仅限于其公司内部相关数据上的表现不错；Li等[14]提出了一种上下文感知的动态卷积网络，可以在生成卷积核的同时充分利用上下文信息，提出的新网络框架可被扩展至多任务学习下的双通道上，用于完成对话行为和情感分类的联合识别任务。

1.2 目前存在的问题

目前主流的深度学习方法在对话任务的表现不错，但有些问题需要进一步解决，比如：①模型中领域漂移(domain shift)、语义间隔等问题；②目前对于语言的理解多数停留在语义表层，复杂场景下融合知识后的逻辑推理能力偏弱；③域适应能力面临的问题亟需解决，以便应对自由度比较高的对话场景。

2 Rasa框架概述

Rasa是一个用于构建智能对话机器人的开源机器学习框架，拥有大量的可扩展特性组件，几乎覆盖对话系统的所有功能，是目前对话系统的主流开源框架[15]，主要包含以下四部分。

(1) Rasa NLU：提取用户想要做什么和上下文关键信息；

(2) Rasa Core：基于对话历史，选择最优的回复和动作；

(3) 通道(channel)和动作(action)：连接对话机器人与用户及后端服务系统；

(4) 辅助系统：Tracker Store、Lock Store和Event Broker等。

Rasa内部运行机制如图1所示。用户可以利用通道支持的客户端将查询信息输入对话系统，Rasa NLU负责将用户的输入转换成意图和实体信息，完成自然语言理解工作。然后，Rasa Core负责基于当前和历史的对话记录(其中NLU的输出是对话记录的一部分)采取下一个动作，下一个动作可以是回复用户信息、调用用户自定义的动作类。自定义动作运行在一个单独的服务进程中，被称为动作服务器，该服务器通过HTTP(hypertext transfer protocol)与Rasa Core进行相互通信。多数情况下，我们重点关注Rasa NLU和Rasa Core两个核心部分。

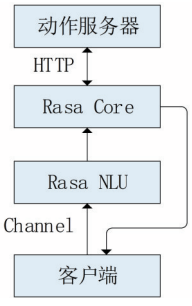


图1 Rasa内部运行机制示意图

3 系统设计与实现

3.1 自然语言理解模块

3.1.1 Rasa NLU模块中各类组件

Rasa NLU主要是通过监督学习算法完成意图提取和实体提取，所以需要自行准备一些包

含意图信息和实体信息的语料。这里将以某央企总部制定的“会计基础管理办法”及省公司本部制定的“会计科目责任制管理规范(业务规范)”等原始数据为基础，准备需要的相关语料，同时会根据最新的内外部政策要求更新语料库。

Rasa NLU在软件架构设计上比较灵活，允许开发人员自行使用各种算法完成相关工作，这些算法的具体实现被称为组件(component)。由于这些相互连接的各类组件及组件间的数据流共同组成了有向无环图，所以Rasa NLU也可以被看作一个基于有向无环图的通用框架。自然语言理解模块通常有实体识别和意图识别两个任务，为了完成这些任务，图2给出了一个典型的Rasa NLU配置通常包含的各类组件，下面针对这些组件进行逐一说明。

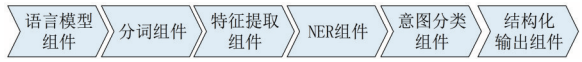


图2　典型Rasa NLU组件示意图

3.1.2　流水线配置

由上面分析可知，Rasa NLU是基于有向无环图进行组件配置，这种有向无环图在Rasa中被称为流水线(pipeline)，配置文件格式为YAML，配置样例如算法1所示。

算法1　流水线配置样例

```
recipe: default.v1
language: zh
pipeline:
  - name: JiebaTokenizer      #使用Jieba中文分词器
  - name: LanguageModelFeaturizer
                  #基于HuggingFace的Transformers库
    model_name: bert      #BERT的语言模型和架构
    model_weights: bert-base-chinese
                        #BERT的中文语言模型和架构
  - name: RegexFeaturizer
                  #读取训练数据中的正则表达式配置
  - name: DIETClassifier
                #Rasa提供的实体和意图联合提取组件
    epochs: 100                #模型运行100个epoch
    learning_rate: 0.001
    tensorboard_log_directory: ./log
  - name: ResponseSelector
    epochs: 100
    learning_rate: 0.001                #模型学习率
  - name: EntitySynonymMapper          #实体识别组件
```

3.1.3　Rasa NLU分析结果

如果输入“请问车辆燃油费报销时需要上传哪些附件”，NLU将会给出如算法2所示的解析结果，显然“attachment”和“attachment”之间的相关性明显高于与“goodbye”的相关性。

算法2　Rasa NLU分析结果

```
{
  “text”:“请问车辆燃油费报销时需要上传哪些
    附件”,
  “intent”:{
    “id”:5313274288154357548,
    “name”:“attachment”,
    “confidence”:0.999690031993
  },
  “entities”:[
    {
      “entity”:“reimbursement”
      “start”:2,
      “end”:9,
      “confidence”:0.999778123345692,
      “value”:“车辆燃油费报销”,
      “extractor”:“DIETClassifier”
    }
  ],
  “intent_ranking”:[
    “id”:5313274288154357548”,
    “name”:“attachment”,
    “confidence”:0.999688021375883,
  },
  {
    “id”:-3614266674174050593,
    “name”:“goodbye”,
    “confidence”:0.000122011470351,
  }
  ]
}
```

3.2　对话管理模块

Rasa Core是Rasa体系中负责对话管理(dialog management，DM)的部分，主要职责是根据NLU模块的输出实现会话行为状态的更新和跟踪，并结合一定的策略(policy)选择下一个动作输出。简言之，对话管理的核心在于基于意图

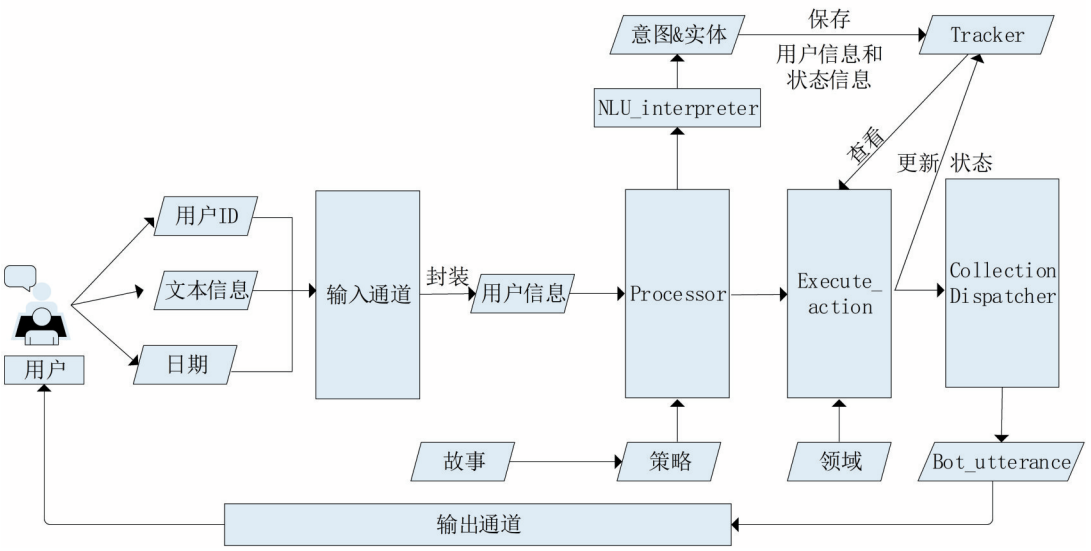


图 3 对话管理模块内部运行机制示意图

(intent)和状态(state)的动作预测。对话管理模块内部运行原理示意图如图3所示。

3.2.1 通道

通道的主要作用是对经过的数据进行统一封装，将符合规范的数据推送至下一个节点处理。这里的通道有两类：输入通道和输出通道，前者负责将用户的身份和输入信息封装成用户信息对象后推送给Processor处理，后者将对话系统需要返回给用户的信息进行统一封装。

3.2.2 领域

领域(domain)对象定义了对话系统需要掌握的所有信息，例如意图、实体(entity)、词槽(slot)、动作、表单(form)和回复(response)。这些数据明确界定了模型的输入和输出范围，一般由Rasa框架中的配置文件domain.yml统一提供。

本质上讲，领域有利于将对话系统需要使用的模板信息进行统一管理，包括定义的意图、实体、槽位、话术模板，以及自定义的动作和策略等。领域对象在动作执行时可以提供这些模板信息，如果需要添加或修改这些信息，只需要修改domain.yml文件里的数据就可以，不需要修改任何代码，实现了系统配置和代码解耦，便于整体上对对话系统进行管理。算法3给出了领域对象的样例。

算法3 领域对象样例

```
intents:
    - goodbye
    - greet
    - chitchat
    - deny
    - stop
    - affirm
    - info_date
    - info_ reimbursement
entities:
    - reimbursement
    - date-time
slots:
    address:
        type: text
        influence_conversation: false
        mappings:
            - entity: reimbursement
              type: from_entity
    date-time:
        type: text
        influence_conversation: false
        mappings:
```

```
        - entity: date-time
          type: from_entity
responses:
  utter_greet:
    - text: 您好，请问有什么事情可以帮到您?
  utter_goodbye:
    - text: 再见!
  utter_ask_ reimbursement:
    - text: 需要咨询哪类报销业务问题?
  utter_ask_date-time:
    - text: 业务什么时候发生的?
  utter_ask_continue:
    - text: 是否要继续?
  utter_default:
    - text: 小E没有理解您说的是什么意思。
actions:
  - utter_greet
  - utter_ goodbye
  - utter_ ask_ reimbursement
  - utter_ ask_date-time
  - utter_ask_continue
  - utter_default
  - respond_chitchat
  - action_ reimbursement _form_submit
forms:
  reimbursement_form:
    required_slots:
      - reimbursement
      - date-time
```

3.2.3　故事

故事(story)是一种较高语义层次上记录对话过程的方式，它不仅需要记录用户的语义表达，还需要记录系统内部正确的状态变化。故事本身是以字典的形式存在，包含两个键：story和steps。后面的策略就是通过学习故事内容才进行动作预测的。

3.2.4　策略

策略(policy)担负着根据当前用户反馈的信息告知Processor当前一轮对话中应该执行的后续动作以及如何更新对话状态信息等任务。Rasa框架中的策略管理模块支持人工规则的策略和依靠机器学习、深度学习完成的数据驱动策略，也支持组合策略。由于前期财务领域的样本数据偏少，所以采用以表单策略(form_policy)为代表的人工规则策略，该策略与动作管理中的表单动作(form_action)相对应。需要说明的是，Rasa框架提前预置了一个叫fall_back的策略，其目的是为对话系统执行一个兜底策略。

3.2.5　动作

动作接收用户输入和对话状态信息，按照业务逻辑进行处理，并输出改变对话状态的事件和回复用户的消息。

3.3　系统部署

在生产环境中部署对话系统时，使用单机部署模式，使用Redis作为Tracker store和Lock store背后的数据库，单机高并发设置为6，满足Rasa服务器使用6个worker，性能测试时使用REST接口连接Rasa进行对话处理，并使用JMeter等测试工具对Rasa的吞吐量和响应时间进行测量。

4　结语

本文首先从企业以人机协同系统建设推进其智能化建设的角度引出对话系统，并对选择Rasa框架搭建智能对话系统进行说明，提出构建智能对话系统的实现方式；其次，针对对话技术研究现状和目前的问题进行分析，给出相关研究进展和未来挑战；然后，针对Rasa框架内部运行机制进行解读；最后，针对智能对话系统进行详细设计，针对对话系统实现及系统部署进行说明。笔者认为，随着全国“全电发票”试点工作的深入推进，国内大型企业在以财务数智化为基础的智能系统建设步伐必将提速，以人机多元协同为主线的管理智能化创新应用将会越来越多。

参考文献：

[1] 赵阳洋,王振宇,王佩,等. 任务型对话系统研究综述[J]. 计算机学报,2020,43(10):1862-1896.

[2] BUNK T, VARSHNEYA D, VLASOV V, et al. DIET: lightweight language understanding for dialogue systems[EB/OL]. CoRR,abs/2004.09936.

[3] GODDEAU D, MENG H, POLIFRONI J, et al. A form-based dialogue manager for spoken language ap-

plications[C]//Proceedings of the 4th International Conference on Spoken Language Processing. Philadelphia,USA,1996:701-704.
[4] SUN K,CHEN L,ZHU S,et a1. A generalized rule based tracker for dialogue state tracking[C]//Proceedings of the Spoken Language Technology Workshop. South Lake Tahoe,USA,2014:330-335.
[5] LEWIS D D. Naive (Bayes) at forty: the independence assumption in information retrieval[C]//Proceedings of the European Conference on Machine Learning. Chemnitz,Germany,1998:4-15.
[6] ZHANG H. The optimality of Naive Bayes[J]. Miami Beach,USA,2004:562-567.
[7] YANG Y,LIU X. A re-examination of text categorization methods[C]//Proceedings of the 22nd Annual International ACM SIGlR Conferenee on Research and Development in Information Retrieval. New York,USA,1999:42-49.
[8] WEINBERGER K Q,BLITZER J. Distance metric learning for large margin nearest neighbor classification[J]. Journal of Machine Learning Research,2009,10:207-244.
[9] JOACHIMS T. Transductive inference for text classification using support vector machines[C]//Proceedings of the 16th International Conference on Machine Learning. B1ed,S10venia,1999:200-209.
[10] LUBIS N, GEISHAUSER C, HECK M, et al. LAVA:latent action spaces via variational auto-encoding for dialogue policy optimization[C]//Proceedings of the 28th International Conference on Computational Linguistics,2020:465-479.
[11] MA X Z, HOVY E. End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany, 2016:1064-1074.
[12] DILEK H-T, GOKHAN T, ASLI C, et a1. Multi-domain joint semantic frame parsing using bi-directional RNN-LSTM[C]//Proceedings of the 17th Annual Conference of the International Speech Communication Assoeiation. San Francisco, USA, 2016:715-719.
[13] 冯强中. 基于知识增强的企业技能智能问答应用研究[J]. 现代计算机,2022,28(9):8-14.
[14] LI J Y,JI D H. Modeling local contexts for joint dialogue act recognition and sentiment classification with bi-channel dynamic convolutions[C]//Proceedings of the 28th International Conference on Computational Linguistics,2020:616-626.
[15] 孔晓泉,王冠. Rasa实战:构建开源对话机器人[M]. 北京:电子工业出版社,2022.

Research and design of intelligent dialogue system based on Rasa framework

Li Mengquan, Wu Shihong, Wang Zhigang

(YG Soft Inc., Zhuhai 519085)

Abstract: With the rapid development of pre-trained language models and knowledge graph technology, human-computer dialogue technology has become a new research hotspot in the field of artificial intelligence, which has attracted extensive attention from academia and industry. Human-computer dialogue system can be divided into task-oriented dialogue systems and non-task-oriented dialogue systems, among which task-oriented dialogue systems provide a convenient interface to help users complete some specific tasks, and in-depth research on them will have high academic value and application value. This article will use the Rasa open source framework component to research and design an intelligent dialogue system to help the company achieve the management goals of improving quality and efficiency, and improve the company's overall intelligent management level.

Keywords: dialogue system; natural language processing; Rasa

文章编号: 1007-1423（2023）01-0101-05 DOI: 10.3969/j.issn.1007-1423.2023.01.018

基于Web的新型教育服务平台设计

于雯琦，于 杰，贾雪坷，赵 硕，樊文静

（河北环境工程学院信息工程系，秦皇岛 066100）

摘要：教育培训是近年来逐渐兴起的一种将知识教育资源信息化的机构或在线学习系统。对网课教育及教育培训信息获取的平台设计方案进行了详细分析论述。通过对教育资源的需求和市场前景的分析，阐述了建设平台的意义，明确了平台设计的目标、任务及主要功能，可以为学生提供网络学习环境，使学生便捷地筛选出当地优质教育资源，同时为教育培训行业提供二次发展机会。

关键词：网课教育；教培信息获取；服务平台

0 引言

教育培训是近年来逐渐兴起的一种将知识教育资源信息化的机构或在线学习系统。随着教培行业总量的快速膨胀，竞争也变得更加激烈。而2020年新冠疫情的到来更是对我国教培行业一次沉重的打击。相关数据显示，疫情期间中国89.4%教育机构营业收入减少[1]。另外随着信息技术的快速发展，借助网络进行教学的教育模式已经成为高等院校知识传播的重要途径，利用网络资源来满足自己的学习需求已经成为众多学生的选择。

此外，“双减”政策正式出台后，虽然减少了学生的压力，但是给家长带来了教育焦虑、给学科类培训机构带来了冲击。对于家长而言，他们更担心孩子输在起跑线上，教育资源处于不平衡态，家庭教育不够完善，升学焦虑依旧处于较高地位；对于学科类培训机构而言，大多数机构正在向非学科类和素质类教育转型，然而迅速转型存在一定的困难，需要一些软着陆的应对。夏越教育创始人谢业恩坦言：“希望相关部门能够为培训机构的转型提供更多指引和机会，让培训行业从业者有更多转型的选择”[2]。

1 市场前景

1.1 线上教育的发展趋势

截至2020年12月，中国在线教育用户规模达3.42亿，随着在线教育用户规模的快速增长，中国在线教育市场规模也明显上升。根据中国在线教育整体市场数据显示，2020年中国在线教育市场规模约达4230亿元，行业发展态势较好。并且国家针对在线教育行业出台了《关于大力加强中小学线上教育教学资源建设与应用的意见》等一系列积极政策，为中国在线教育行业发展推波助澜。

1.2 市场饱和度

目前，就教育市场的整体发展状况来看，在线教育渗透率不断提高，课程布局线上已成必选项，反观中国城乡区域，网络化教育相对匮乏，因而造成教育发展不平衡的现状。

收稿日期：2022-07-13 **修稿日期：**2022-09-06

基金项目：国家级大学生创新创业训练计划项目（202151721007X）

作者简介：于雯琦（2001—），女，河北邯郸人，本科，专业为物联网工程；于杰（2001—），女，河北承德人，本科，专业为软件工程；贾雪坷（2001—），女，河北邯郸人，本科，专业为物联网工程；赵硕（2001—），女，河北邢台人，本科，专业为软件工程；樊文静（2000—），女，河北沧州人，本科，专业为物联网工程

2 研发平台的意义

（1）线上教学突破了时间、地点和空间的限制，解决了因疫情防控而不能出行和聚集等问题，师生在家就可以完成教学任务。相比于传统教育，线上教育的成本更低，更具有灵活性和包容性，给教育资源相对匮乏地区的学生更多受教育的机会，丰富他们的学习方式。

（2）目前用户获取当地教育培训机构信息渠道较窄，具有一定的局限性，本平台将教育培训机构信息一体化，对优质教育资源进行整合和筛选，构筑“点—线—面—网”教育资源共享体系，聚焦师资、教研、管理等，成为点燃城乡区域教育振兴的“新引擎”[3]。

（3）本平台规范透明、全面和客观的评价机制，促使教育培训机构提高教学质量，完善教学方式，从而形成教育培训机构之间良性竞争，促进城乡区域教育的发展。

（4）“双减”政策下，为面临转型困难的学科类机构提供发展空间。

3 平台系统开发设计

3.1 需求分析

通过对教育机构分布与类型现状的分析以及对学生家长用户的调查发现，学生和家长用户所面临的问题是获取当地教育培训机构信息渠道较窄，而且得到的信息具有一定的局限性与片面性，大型机构垄断、信息的单向性等原因限制、阻碍了教育资源共享；其次，还面临着因疫情防控使得受教育的时间地点受到限制、城乡区域网络化教育较为匮乏、家庭教育不完善导致的教育焦虑等问题。

随着监管趋严、“双减”政策正式落地，教培行业迎来巨变，不少企业股价大跌，行业走到十字路口，该形势下教培行业面临如何转型自救的问题。

发现问题就是为了解决问题，特别是发现问题的根本所在，更利于问题的解决。通过探讨课程建设的发展趋势、优化系统内学习资源、完善教学评价反馈系统，打造出一款帮助用户选择优质网络课程、帮助教育培训机构转型和发展的服务平台。用户可以在使用过程中对相同类型的教育培训机构有全面的了解和对比，选择更适合自身的教育方式，同时优质师资稀缺引发的教育焦虑也能得到缓解。

本平台的应用范围广泛，适用于所有学习者。教育机构系统管理员应相对固定且要分二级或三级管理；教育系统信息的更新会记录在本平台中；平台界面风格统一、简洁大方，便于用户和管理员不同端口登录、操作和使用。

3.2 研究方法

本文研究主要采用调查研究、比较研究、评价研究和分析研究等方法。

（1）该系统的开发和研究采用比较、分析和评价研究；

（2）结合工作实际，理论研究和程序开发及系统模块的建立均在实践中得到考验和验证；

（3）通过对教培机构管理者和学习者实地调查和问卷分析，确切掌握相关需求，建立对应的系统功能模块；

具体的研究程序和步骤如下：

（1）首先搜集相关文献，仔细阅读并消化吸收，拓展知识面、奠定理论基础；

（2）带着问题设计调查问卷和走访教育培训机构，了解问题、困难、矛盾和实际需求；

（3）分析比较教培机构线上平台系统的优缺点，全面分析和推敲，对系统的体系结构和类型进行全面的了解[4]。

3.3 平台设计目标及原则

本平台的最终目标是全国的教育培训行业平台入驻率达到80%，建设丰富的网络课程；将学科机构信息一体化，增加机构的曝光率和咨询量；帮助部分机构应对转型困难，同时形成良性竞争；实现教学资源共享最大化，支持多人在线学习和学习资源之间的自由转换，并且实现多终端无缝切换，个性化教育等。最终搭建开发一个具有开放体系的、便捷的、易维护的、安全性高的、具有良好信息交互的网课及教培信息获取的网络平台。为用户提供准确且精细的线下教培行业信息以及优质的网课内容，让教学与学习更轻松、更高效。综上平台设计应遵循以下原则。

（1）稳定性，即网页的技术细节成熟稳定，能够满足用户多样性的需要。当资源使用率、每秒点击数以及并发用户数较多时，能保证服务器的稳定。

（2）便捷性，“分类 +推荐+ 筛选 + 搜索 +

预览 + 详情”是该平台的基本流程，该流程能让用户高效、简单地检索到自己想要的信息。

（3）安全性，即平台安全性的目标是满足相应的安全等级要求。

（4）维护性，即运营过程中，及时对近期维护工作分布情况进行统计，预防平台故障的发生。

3.4　平台系统总体设计

本系统是MVC架构的动态的教育培训机构选择服务网站，平台系统总体设计如图1所示。在设计过程中主要使用NetBeans IDE、后台数据库Navicat、服务器Tomcat，采用JSP和Web技术为开发手段。平台主要包含管理员模块、用户模块、教育培训模块三大模块。

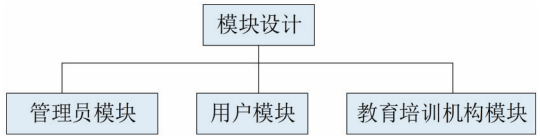


图1　模块设计

3.5　管理员模块

管理员模块(图2)主要包括AI审核，意见管理，教育培训机构信息管理等功能。

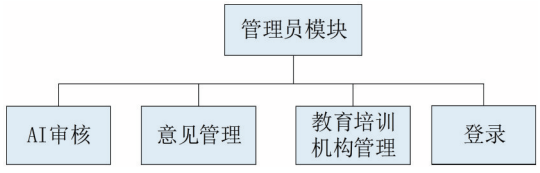


图2　管理员模块设计

3.5.1　AI审核功能

系统对所有发布的内容进行智能审核，并重点对其中出现的图像、音频和文本等进行全方位检测，一旦发现某一方面出现违规内容，将对整段视频进行过滤拦截，并打回，以防流入大众视野，并可以自动获取一定比例的内容转交后台管理，管理员检查是否有违规内容。

3.5.2　意见管理功能

用户对本平台反馈的信息都将在此列表中显示，本功能可以提取关键词并统计问题出现的频率。管理员可以查看反馈内容，并给予回复或删除。

3.5.3　登录注册功能

网站首页的标题栏提供了登录系统管理员入口，管理员通过输入正确的用户名和密码以及验证码即可登录到网站后台。

3.5.4　教育培训机构信息管理功能

教育培训机构入驻平台后，平台会根据教育机构提供的信息以及一定的社会调查基础对其信息进行审核。审核通过后，教育机构的简介、资质、教育资源以及合同时长等相关信息将被管理员正式录入到数据库中。若与平台解约，管理员可以一键删除其信息。

3.6　用户模块

用户模块(图3)主要分为首页、我的课程和个人中心三部分。首页界面主要包含搜索引擎、智能分类和智能推荐功能；我的课程界面包含已购买课程、我的收藏和课程评价功能；个人中心界面包含用户注册登录和意见箱功能。

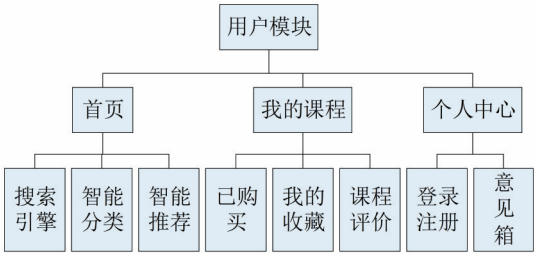


图3　用户模块设计

3.6.1　搜索引擎功能

可搜索区域内的教育培训机构及课程，查询结果包括该机构的师资力量、机构评分、价格、机构位置以及课程详情等，并且以图文或视频的形式对教育培训机构的线上线下课程进行系统展示。用户可选择适合的课程进行了解或咨询，网课被订购之后就会被系统自动保存到本系统数据库的网课信息表当中。

3.6.2　智能分类功能

本功能会根据用户年级、兴趣科目等将课程进行智能分类。方便用户按需查找课程，节省时间。

3.6.3　智能推荐功能

本功能会根据搜索记录、浏览顺序等用户

的历史行为，利用大数据进行兴趣建模，推荐适合用户的课程，给予用户智能化体验。

3.6.4 注册和登录功能

普通用户访问本系统主页面时，需要注册账号并登录。输入手机号和验证码。注册完成后，自动跳转到登录页面，用户需完善密码、身份类型、学历、所在地等信息。

3.6.5 意见箱功能

用户可以对系统以及平台的服务等问题发表意见和看法，方便平台进行完善与改进，更好地为用户服务，满足用户需求。

3.6.6 已购买功能

用户可以在已购买功能(图4)查看已报名的线上课程及相关信息和学习进度。

3.6.7 我的收藏功能

我的收藏功能(图5)分别显示收藏的课程和教育培训机构列表，“我的收藏”方便将课程集中保存，避免丢失或遗忘。

图4 已购买课程

图5 我的收藏

3.6.8 课程评价功能

凡是实名认证并观看过此类课程的时长达到50%以上的用户(防止恶意评论与刷评)可以对课程进行多方面评价，并且对教学方式、课程的内容以及平台使用感三方面打分。点击并查看评价信息，帮助用户从多角度了解要选择的课程，为用户提供参考。

3.7 教育培训机构模块

教育培训机构模块(图6)主要分为用户中心、信息发布和学生信息管理三部分。用户中心界面包括登录注册和反馈等功能；信息发布界面包括网课发布、招生信息发布和推广等功能。

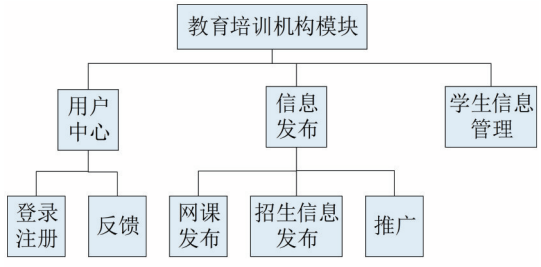


图6 教育培训机构模块设计

3.7.1 登录注册功能

教育培训机构进行注册时除填写常规信息外还需上传“营业执照”“办学许可证”“民办非企业”等材料，待管理员审核通过后即可登

录，教育培训机构需完善名师简介、教学方式、招生信息等宣传信息。

3.7.2 反馈功能

用户在平台上对教育培训机构进行的反馈、提出的意见和问题等会在该栏中显示，使本机构找到自身的不足，更加完善。

3.7.3 推广功能

教育培训机构通过发布课程和信息等方式进行推广宣传。用户可通过搜索引擎、智能分类、智能推荐等功能查阅教育培训机构的招生信息。此外，还设有特惠课程的定制推广功能，教育培训机构可提前一周向网站运营管理处申请，审核通过后将在首页滚动播放其课程的海报等相关内容。

3.7.4 发布招生信息功能

发布教育培训机构的招生信息及招生要求等。

3.7.5 发布网课功能

教育培训机构可以在此进行课程发布，已报名的用户可以及时收到发布信息。

3.7.6 学生信息管理功能

已报名线下课程的用户个人信息会存放在表单中(包括姓名手机号、任课老师、所在班级等)教育培训行业的负责人可以对表单内容进行录入、删除、更新、检索等操作。报名网络课程的用户信息以及学习进度等也将录入到后台数据库中。

4 结语

目前，本项目已纳入国家级大学生创新创业训练计划平台，教育面临着世界性新技术革命的挑战，在这种新形势下，本平台的搭建有利于发展我国的教育事业，使之更好地适应经济社会发展的需要，更好地为社会主义现代化建设服务。

参考文献：

[1] 张景璐. 我市校外培训机构转型开新局[N]. 珠海特区报,2022-01-14(6).

[2] 于君瑶,王莉红. 新冠疫情背景下国内教育培训行业发展策略研究[J]. 企业改革与管理,2021,(16):99-100.

[3] 潘旭阳. 构建区域教育“命运共同体”[N]. 中国教师报,2021-12-08(15).

[4] 刘云,吕富岩. 基于Web的高校大型设备共享平台系统的设计和实现[J]. 信息系统工程,2022(4):117-120.

Design of new Web-based education service platform

Yu Wenqi, Yu Jie, Jia Xueke, Zhao Shuo, Fan Wenjing

(Department of Information Engineering, Hebei University of Environmental Engineering, Qinhuangdao 066100)

Absrtact: Education and training is a kind of organization or online learning system that has gradually emerged in recent years to informationize knowledge education resources. The platform design scheme of online course education and education training information acquisition is analyzed and discussed in detail. Through the analysis of the demand for educational resources and market prospects, this paper expounds the significance of building a platform, clarifies the goals, tasks and main functions of the platform design, which can provide students with a Web-based learning environment, enable students to easily screen out local high-quality educational resources, and provide secondary development opportunities for the education and training industry.

Keywords: online education; acquisition of education and training information; service platform

文章编号: 1007-1423(2023)01-0106-04 DOI: 10.3969/j.issn.1007-1423.2023.01.019

基于云平台的信息系统云研发设计模式探索

刘月灿[1*]，孙建刚[1]，窦祎楠[1]，李伟良[1]，彭 嫚[2]，田 涛[2]，王小平[2]

（1.国家电网有限公司信息通信分公司，北京 100761；2.北京中电普华信息技术有限公司，北京 102200）

摘要：随着信息系统业务洪峰瓶颈效应持续加大，传统的系统架构因功能强耦合、无法复用能力等问题，很难满足业务快速迭代、业务规模庞大的需求。因此，在信息系统研发模式引入云原生架构，利用云平台已具备的服务能力引入DevOps管理模式，将容器化与微服务架构结合，探索出云原生研发设计模式，满足业务并发量高、线上用户体量大、版本迭代快等需求，推动信息系统向云原生架构演进，有效应对技术架构管理、研发运维转型和安全挑战。

关键词：云原生；微服务；DevOps；持续集成

0 引言

随着科学信息技术的不断进步，云计算的发展在各行各业的应用日益普遍。云平台包括基础设施服务、平台服务以及软件服务。云平台为资源利用、成本控制、应用程序效能、能源管理等提供了一种新的服务模式。如何利用好云平台服务，集成到应用系统中，显著提升研发人员的工作效率，一直以来都是信息系统研发设计的关注焦点。以容器技术为基础结合微服务架构，为应用的快速迭代、自动部署和快速交付带来了全新、高效的体验。目前亟需探索以云原生应用架构为基础的研发设计模式，满足业务快速响应、业务敏捷迭代、进度管理透明及研发运维便捷等需求，为今后企业信息系统的建设指明发展方向。

1 研发设计模式演进

云原生技术的发展，为信息系统的研发设计模式演进带来了自上而下的创新。信息系统研发设计模式的核心架构也由单体架构演进为微服务架构，将进一步继续演进到云原生架构模式，如图1所示。

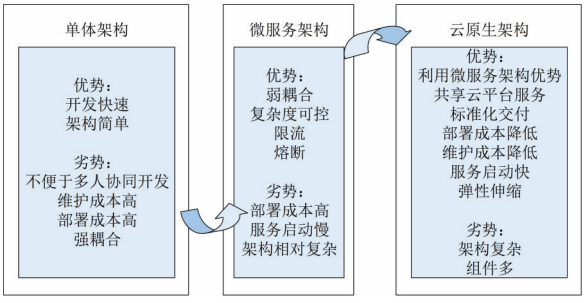


图 1 业界研发设计架构的演进

单体架构是将所有的业务服务都放在一个单体中的业务逻辑耦合的架构。DNS返回相应域名的IP地址，用户根据实际地址访问相应的服务器，最后服务器会做出响应。

微服务架构核心特点是服务相对独立，抽象力度小。微服务互相之间是独立的，当某个微服务出问题时，短时间内是不会影响其他微服务。微服务复杂度相对可控[1-2]。

云原生架构主要以容器化技术Docker和Ku-

收稿日期：2022-07-12 **修稿日期：**2022-08-17

作者简介：*通信作者：刘月灿(1993—)，女，北京人，硕士研究生，研究方向为云计算，E-mail：lycan120624@163.com；孙建刚(1992—)，男，北京人，硕士研究生，研究方向为云平台、数据中台；窦祎楠(1993—)，女，山西临汾人，硕士研究生，研究方向为云平台运维；李伟良(1979—)，男，北京人，硕士研究生，研究方向为云平台、数据中台、电力系统；彭嫚(1987—)，女，河北保定人，学士，研究方向为云平台、移动应用；田涛(1981—)，女，河北石家庄人，学士，研究方向为云平台、软件研发；王小平(1976—)，男，江西九江人，学士，研究方向为云平台、研发架构

bernetes为首的容器编排技术，在微服务架构的基础上，增加资源隔离、容器网络架构、资源自动调度、弹性伸缩等能力。

2　云平台的云原生组件

云原生技术为信息系统建设的灵活性提供了支撑。参照国际CNCF的云原生技术图谱，梳理信息系统建设所依赖的云原生组件。信息系统设计研发人员可以了解每种云原生技术功能，选择符合需求的底层技术、基础架构、协同管理、安全系统等。

2.1　容器基座

容器化技术是一种轻量级的虚拟化技术，本质是在同一个操作系统中进程隔离，可同时启动多个功能单一的任务，有效提升资源利用率。容器和虚拟机虽然都是计算单元，但容器具备秒级启动、快速创建和删除、快速分发和部署、依赖更少的资源、可以弹性伸缩等优点。提供动态调度容器计算资源，容器交付与部署功能，实现“一次构建、到处运行”和“增量更新、快速回退”。基于现实需要，提供容器多节点编排、集群联邦、随机端口映射网络、负载自发现等功能，具备应用弹性扩容、快速自愈、敏捷部署等能力，大幅提升研发交付效率[3-4]。

2.2　微服务架构

在云平台建设过程中，通过深度定制和大规模自主研发，微服务架构组件包含异构多注册中心、节点级服务发现、全链路灰度、同端口双协议、Spring Cloud框架、Dubbo框架，实现高内聚、松耦合、可复用、易扩容的微服务治理功能，具备十万级服务的注册发现和日均亿级的服务接入能力[5]。

2.3　监控与追踪

全链路监控与追踪的应用可以有效减少服务中断的次数、降低解决问题的时间，帮助技术人员尽快识别和解决故障，减轻故障对信息系统的影响。通过对应用程序全方位的监控分析，立即发现并纠正异常情况。包含日志组件、监控组件、服务跟踪组件、混沌工程组件四部分。其中日志组件实现获取有关进程信息的日志记录；监控组件收集指标、存储、监控及告警信息；服务跟踪组件对全链路追踪服务依赖关系以及流量在云平台中的走向追踪，包括延迟、错误等；混沌工程组件是一种软件测试工具，在缺陷影响服务交付之前对其进行识别并加以修复[6]。

3　云原生研发设计模式

研发设计模式的核心作用和特点是使信息系统研发保持统一性，形成统一的架构参考模型和协作管理模型[6]。本文探索以云原生技术为依托，形成基于云平台的云原生研发模式，将容器化、微服务架构技术、DevOps相结合，融合持续交付的理念，形成业务能力标准、服务开发规范、系统运行机制和运营服务的生态体系，为业务提供快速组合和低成本创新的研发设计能力，提高多团队协作效率[7-8]。云原生研发设计模式把云原生看作一种构建和运行应用程序的方法，云原生研发设计作为一组云计算应用程序研发的平台模式，包含敏捷交付流水线、容器化、微服务架构、持续调试与监控、安全扫描等能力，如图2所示。

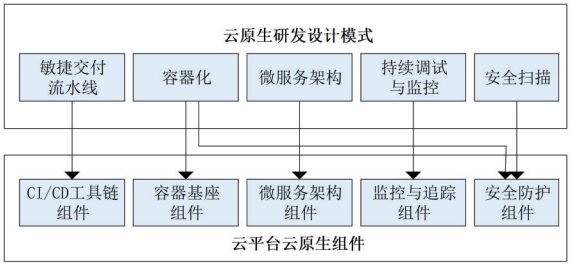


图2　云原生研发设计模式的能力模型

云原生研发设计模式可以帮助企业快速、持续、可靠、规模化地交付业务软件。采用云原生研发设计模式的信息系统研发在6个方面存在优势：

（1）可以使用云计算、云存储等云资源。

（2）可以采用统一云原生架构以及提供的各种工具、云平台服务，快速研发云原生应用。

（3）可以组建多个高度灵活的小团队组合，负责不同的微服务研发。

（4）通过研发与运维的结合（DevOps），实现对微服务有端到端的责任，提高持续交付的效率。

（5）通过全方位监控，快速定位故障。

（6）通过使用云平台云原生组件，实现容器基础设施安全、容器编排平台安全、镜像安全、微服务安全[9-10]。

3.1 容器与微服务结合设计

容器是一种新的软件交付方式，将应用和其运行环境以一个标准镜像格式打包，能保证应用及其运行环境的统一。微服务是应用软件架构设计模式[3]，推崇单一职责、服务自治、轻量通信和接口明确等原则。结合两者优势，容器可以较好地配合使微服务易于开发和维护等[4]。容器化和微服务架构是云原生研发设计的关键组成部分，前者相当于基石，后者相当于合作伙伴。微服务和容器化结合，主要有以下两个优势，首先采用微服务架构，使用容器作为基础设施，能够实现快速部署，快速迭代[5]；其次提供了一种配置中心和注册中心的标准，容器天然地为微服务架构提供了开发配置中心和注册中心，而无需再开发。从开发工程师的角度，关注的是微服务架构，以微服务为基础。从运维工程师的角度，关注的是容器化，以容器等基础设施环境为基础[11-12]。

3.2 DevOps研运一体管理

DevOps是开发运维一体化的方法论。DevOps实现了项目开发、运维管理、环境部署的一体化，为软件的在线开发创造了更好的环境。除了支持在线开发外，还实现镜像管理、进度调控等一系列功能。DevOps促进工作思维方式的改变，为业务价值快速提升提供支持[6]。DevOps强调共同目标，以用户价值为唯一评价标准，有利于技术开发、技术运维、质量保障等部门之间的沟通协作整合，保证产品功能及时实现，确保成功部署和稳定使用[13]。

3.3 云平台服务组件利用

云原生研发设计模式可以归纳为模块化、可观察、可部署、可测试、可替换、可处理这六个特质。云原生研发设计包括敏捷交付流水线、容器化、微服务架构、服务调试与分析、安全扫描等能力，充分利用了云平台的云原生组件，云原生研发设计是以云平台的平台服务作为基础支撑，敏捷交付流水线利用了CI/CD工具链组件，容器化利用了容器基座组件和安全防护组件，微服务架构利用了微服务架构组件，持续调试与监控利用了监控与追踪组件，安全扫描利用了安全防护组件，如图3所示。

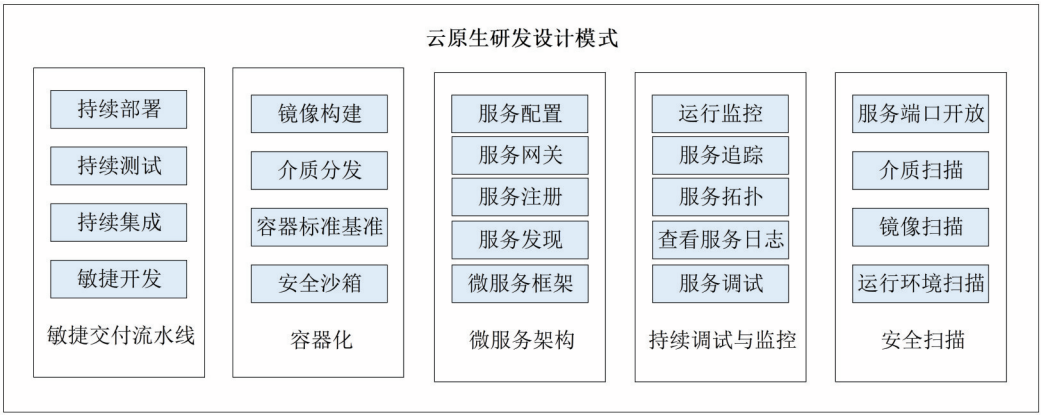


图3 云原生研发设计模式与云平台服务的关系

4 结语

科学技术的发展深刻地改变着人们的生活，在未来的发展道路中，这种变革会更加深入、更加迅速。依托DevOps流水线、容器化基础架构、微服务架构，打造支撑敏捷交付、研运一体化的高效研发能力；信息系统云原生研发设计模式在企业级架构建设中得到充分运用，一大批生于云、长于云的新一代应用系统陆续诞

生，为创新发展提供高水平科技支撑。

参考文献：

[1] 黄铮，周洁，朱学海. 一种基于Azure DevOps的面向嵌入式软件的持续集成及自动测试环境[J]. 工业控制计算机，2022，35(9)：17-19，23.

[2] 张亚辉，马海燕，陈森，等. DevOps模式下的CMDB研究[J]. 信息技术与信息化，2022(9)：127-130.

[3] 郝玉锴，戴小氐，吴姣. 安全的开发运维一体化及其在机载领域的应用[J]. 航空计算技术，2022，52(4)：77-80.

[4] 王强. 基于微服务架构的宜昌“智慧人社”建设实践[J]. 信息系统工程，2022(7)：96-99，104.

[5] 卞中杰. 一种基于DevOps的自动化测试及CI/CD流程的设计与实现[J]. 科学技术创新，2022(20)：58-61.

[6] 李林，张亚辉，陈森，等. 新集中化模式下的IT运维发展思路探讨[J]. 数字通信世界，2022(1)：162-164，167.

[7] 王鹏飞，唐灏. 基于DevOps的研发仿真环境构建[J]. 现代计算机，2021，27(32)：85-89.

[8] 范国伟. 基于容器的DevOps云平台设计及系统测试[J]. 科学技术创新，2021(32)：119-122.

[9] 刘帅华，王天青. 企业传统应用架构向微服务架构转型的一种流程设计[J]. 微型电脑应用，2021，37(10)：105-107.

[10] 王保中，姚文胜，梁奂，等. 开发外包场景下DevOps实施方法及实践[J]. 电信科学，2020，36(12)：113-122.

[11] 陈泳，姚文胜，陈靖翔，等. 基于云原生技术敏捷交付云网融合应用[J]. 电信科学，2020，36(12)：96-104.

[12] 张嘉杰. 基于JMeter、Ant和WeTest的佣金管理系统接口自动化测试程序开发[J]. 电子技术与软件工程，2020(24)：26-29.

[13] 居彬，王启威，谢承翰. 基于DevOps的流水线可视化自由编排方案与实践[J]. 电子技术与软件工程，2020(24)：50-51.

Exploration of cloud R&D design mode of information system based on cloud platform

Liu Yuecan[1*], Sun Jiangang[1], Dou Yinan[1], Li Weiliang[1], Peng Man[2], Tian Tao[2], Wang Xiaoping[2]

(1. Information and communication Branch of State Grid Co., Ltd., Beijing 100761;
2.Beijing Zhongdian Puhua Information Technology Co., Ltd., Beijing 102200)

Abstract: As the bottleneck effect of information system business continues to increase and the traditional system architecture can not meet the needs of rapid business iteration and large business scale due to strong functional coupling and inability to reuse. Therefore, the Cloud Native architecture is introduced into the information system R&D mode and Using the service capabilities of the cloud platform, the DevOps management mode is introduced. By combining containerization with the micro service architecture. The Cloud Native designing mode is explored to meet the needs of high business concurrency, large online users and fast version iteration, promote the evolution of the information system to the Cloud Native architecture, and effectively respond to the technical architecture management and R&D mode transformation and security challenges.

Keywords: cloud native; microservices; DevOps; continuous integration

文章编号: 1007-1423(2023)01-0110-04 DOI: 10.3969/j.issn.1007-1423.2023.01.020

基于天擎的容器应用

钟艳雯，夏正龙，朱　亮，吕冠儒，朱宏武

（湖南省气象信息中心，长沙 410118）

摘要： 研究Docker技术在天擎算法接入的应用。以湖南省水文数据同步算法为例，阐述了从镜像制作到算法调度运行的过程。利用水文降水量数据对算法进行性能测试，实验结果表明，同一算法在天擎算法容器云中同步效率高于虚拟机环境及自建容器服务器。

关键词： 天擎；Docker；算法

0　引言

随着气象业务的飞速发展，对气象系统的信息化、集约化、扁平化的要求越来越高，Docker容器凭借轻量级、启动快、资源占用少的优点[1]，在气象领域中得到了广泛的应用。近年来多省气象部门开展了Docker技术在业务中的研究与应用，利用相关技术搭建气象云平台服务、部署相关算法等[2-4]。2021年湖南省气象大数据云平台，即“天擎·湖南”（以下简称天擎）正式上线，提供数算一体的平台化服务，实现了数据管理、加工处理、应用服务的高度集约。2022年天擎新增了镜像资源管理与容器资源管理功能，允许用户在天擎环境中制作镜像，并在加工处理算法的容器云计算框架中加载运行。本文拟对基于天擎的容器应用进行研究，将Docker体积小、部署快等优点与天擎环境结合，在充分保证运行环境独立的同时，减少环境部署时间，降低系统资源消耗。

1　相关技术

1.1　Docker技术

Docker是dotCloud公司于2013年发布的轻量级容器引擎，得益于开源、高效和标准化实现，Docker已成为容器云平台的事实标准[5]。Docker机制包括镜像（image）、容器（container）和仓库（repository）。Docker镜像类似于虚拟机镜像，为只读模板，包含了文件系统。容器是从镜像创建的运行实例，容器间相互隔离，独立运行。镜像是创建容器的基础，仓库是Docker集中存放镜像文件的场所。Docker容器是在操作系统层面上实现虚拟化，直接复用本地主机的操作系统，对硬件的需求远低于传统虚拟机。从资源占用上来说，更加轻量级；从应用上来说虚拟移植性更强，灵活性更高[6-9]。

1.2　天擎容器云

天擎提供镜像资源管理与容器资源管理功能，制作并管理Docker镜像和容器；在算法管理中提供容器云计算框架，将用户镜像加载后供加工流水线调度运行。天擎提供两类基础镜像：一类为操作系统基础镜像，只包含操作系统及常用的基础软件；另一类为语言包基础镜像，除操作系统外，包含基础的语言环境（如Java、Python、C++等）。算法基于容器有两种运行模式：①算法在容器外的宿主机上，运行时

收稿日期： 2022-07-21　**修稿日期：** 2022-08-19

基金项目： 湖南省气象局2021年重点课题(XQKJ21A005)：省级多源气象数据采集与行业数据入云关键技术研发

作者简介： 钟艳雯(1976—)，女，广西柳州人，硕士，高级工程师，研究方向为气象信息技术应用；夏正龙(1980—)，男，江苏建湖人，硕士，高级工程师，研究方向为气象信息技术应用；朱亮(1986—)，女，湖南双峰人，硕士，高级工程师，研究方向为气象信息技术应用；吕冠儒(1987—)，男，甘肃金昌人，硕士研究生，工程师，研究方向为气象信息技术应用；朱宏武(1978—)，男，湖南常德人，博士，高级工程师，研究方向为气象数据处理

通过映射的方式在容器内运行；②算法装载到容器内，直接在容器内运行。模式①的优点是方便算法的调试，算法修改后无需重新打包镜像，缺点是不利于算法的管理，且需提供额外的存储空间。模式②可以实现对算法及资源的严格管理，但当算法更改时，需要重新打包镜像。天擎在算法开发测试阶段允许使用模式①，便于进行算法的调试。当算法完成测试后，在业务环境中需要将算法装载至容器内，以模式②运行。

2　容器应用

2.1　前期准备

首先需要申请天擎业务账户，激活数据存储、算法集成与运行权限；其次在资源申请页面中申请存储资源、算力资源、数据访问；最后按照天擎标准对算法输入输出、运行环境、监视告警信息等进行研发与改造，实现与天擎的适配。

2.2　镜像制作

用户通过在线和离线等方式，进行算法镜像的制作和上传提交申请，在管理员审核通过后，可以在天擎业务平台“算法注册”页面进行使用。镜像制作主要包含三种方式：

方法一：离线镜像制作。适用于用户本地无Docker环境且无Docker镜像制作基础，或算法环境依赖需要自定义制作时。使用该方式进行镜像制作时与虚拟机安装环境类似：用户申请容器资源；管理员审核通过后在指定服务器上创建容器并启动sshd服务；用户远程登录到容器内，进行算法依赖安装、算法测试；测试通过后，在离线制作镜像页面选择已调试后的容器，制作成镜像。

方法二：在线镜像制作。适用于用户有Docker镜像制作基础，并有算法包文件，依赖环境点单。用户在平台提供的基础镜像上，通过Docker指令在线制作镜像，包括添加算法指令包、指定执行目录、端口映射、挂载目录等。

方法三：镜像文件上传。用户将通过镜像下载页面保存镜像tar文件或者其他定制好的镜像tar文件，在镜像资源管理页面上传。

2.3　湖南省水文数据同步镜像制作

方法一、二制作的镜像简单、规范、安全，符合天擎审核要求。但是，在天擎提供的基础镜像无法满足算法运行环境需求时，方法一、二中用户无root权限，仅能以用户申请的天擎业务用户运行，且除家目录以外无写权限，在需要安装其它依赖及环境的情况下，建议采用方法三，在自建容器服务器中制作自定义镜像，打包后上传。下面以制作湖南省水文数据同步镜像为例，介绍离线镜像的制作方法。

（1）使用自建容器服务器制作镜像时，需要修改容器内用户最大线程数为65535。

修改方式为：vi /etc/security/limits.d/20-nproc.conf，如图1所示。

```
# Default limit for number of user's processes to prevent
# accidental fork bombs.
# See rhbz #432903 for reasoning.

*          soft    nproc     65535
root       soft    nproc     unlimited
~
~
~
```

图1　修改容器内用户最大线程数

（2）编写Dockerfile。

天擎要求Docker容器中的进程以业务账户执行，在用户镜像上传后会校验是否需要包含业务账户名称、用户Uid，用户目录是否符合规范/space/cmadaas/dpl/{业务账户名称}等。因此，选择编写Dockerfile文件，基于DSL语法构建镜像，如图2所示。

```
FROM python:3.8
# 标识作者
MAINTAINER {业务账户名称}
#创建用户
RUN mkdir -p /space/cmadaas/dpl && useradd -d /space/cmadaas/dpl/{业务账户名称} -u {用户Uid } -m {业务账户名称}
# 设置工作目录
WORKDIR /space/cmadaas/dpl/{业务账户名称}/
COPY . .
COPY libxugusql.so /usr/lib/
RUN pip install -r requirements.txt
```

图2　Dockerfile内容

其中，创建用户与设置工作目录关系到能否通过天擎的镜像校验，是必不可少的。业务账户名称与产品加工业务系统用户名称保持一致，用户Uid可以在镜像管理页面进行查看。

COPY命令复制当前目录下的算法及相关文件或者目录到容器指定目录。

“RUN pip install -r requirements.txt”命令安装requirements.txt中指定的Python软件包。requirements.txt文件内容为：

```
pymssql==2.2.5
requests==2.28.1
```

（3）生成Docker镜像。

```
docker build -t sqltoxg:v3 -f ./Dcokerfile.
```

（4）运行Docker容器。

```
docker run —name sqltoxugu3 -it sqltoxg:v3 /bin/bash
```

（5）调试成功后，将运行中的容器保存为docker镜像。

```
docker commit a346c822d116 sqltoxg:v3.1
```

（6）镜像打包为tar文件。

```
docker save -o docker_SqlToXugu_v3.tar sqltoxg:v3.1
```

（7）上传镜像并注册算法，将算法加载到加工流水线调度运行，算法运行界面如图3所示。

图3 算法运行界面

2.4 运行比较

湖南省水文数据同步算法主要任务是从ms-SQL数据库同步水文行业共享数据到天擎的虚谷数据库中。下面选取水文降水量数据，以相同的算法及参数运行在虚拟机环境、自建容器服务器(以下简称容器)、天擎算法容器云(以下简称算法)中，对虚拟机、容器、算法在同步效率方面进行测试。

（1）不同时间段对比。

选取2022年7月1日0时至1时、7月2日至3日，7月11日至20日的数据进行同步，分别进行小时、日、旬数据同步速度比较，结果见表1。

表1 同步速度测试结果

时间段	记录数	虚拟机速度（条/秒）	容器速度（条/秒）	算法速度（条/秒）
小时	914	554.923	544.343	612.791
日	21957	626.855	644.537	692.479
旬	196526	512.053	606.892	635.285

（2）不同步骤对比。

每次同步操作的主要步骤为：①查询SQL数据库；②去除虚谷数据库中重复记录；③更新虚谷数据库。选取2022年7月21日至30日逐日数据进行同步测试，比较3个步骤耗时情况，查询耗时、去重耗时和更新耗时结果依次如图4、图5和图6所示。

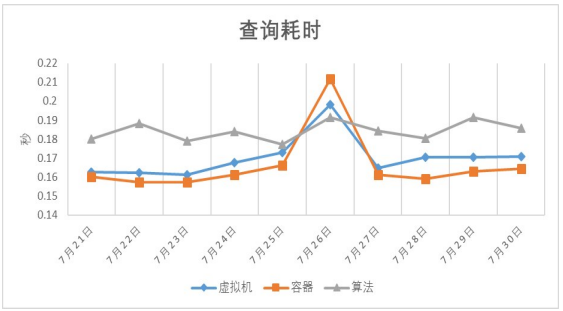


图4 逐日查询耗时

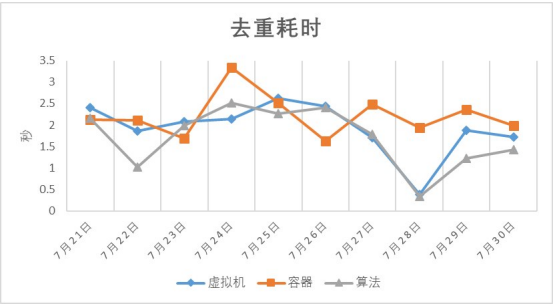


图5 逐日去重耗时

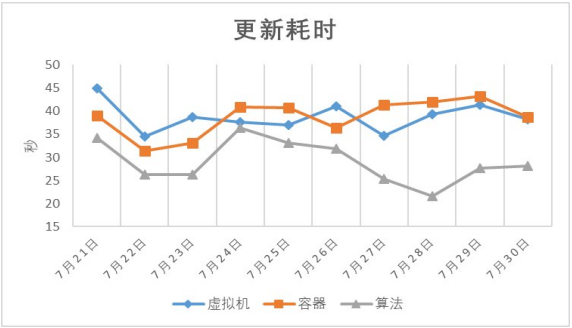


图6 逐日查询耗时

10天总记录数为219501条，每条在3个步骤中的平均耗时见表2。

表2 每条记录平均耗时 单位:毫秒

名称	查询平均耗时	去重平均耗时	更新平均耗时
虚拟机	0.0078	0.0875	1.7648
容器	0.0076	0.1009	1.7613
算法	0.0084	0.0780	1.3253

所有测试结果的不同时间段对比中，算法同步速度最快。不同步骤对比中，步骤1算法平均耗时、单日耗时稍长，虚拟机与容器耗时接近；步骤2算法平均耗时较短，容器耗时最长；步骤3算法单日和平均耗时最短，虚拟机与容器耗时接近。综上所述，虚拟机与容器同步效率接近，算法效率最高。

3　结语

天擎是目前气象部门的核心平台之一，各类算法融入气象大数据云平台是气象信息化、集约化发展的趋势与要求。Docker具有体积小、启动快，在保证运行环境独立的同时减少环境部署时间、降低系统资源消耗等优点，在天擎的算法接入方面有较大的优势。本文以湖南省水文数据同步算法为例，介绍了从镜像制作到算法调度运行的全流程，并对同步效率进行了测试。这对其他算法融入具有一定的参考、示范意义，对推进本省“云+端”的气象业务体系起着积极作用。

参考文献：

[1] 邱建新. 基于Docker容器技术的Linux在线实验环境设计[J]. 信息技术,2022(2):48-52,58.

[2] 朱辉,辛文鹏,冯勇,等. Docker容器技术在省级气象云服务中的应用[J]. 信息技术与信息化,2021(4):31-33.

[3] 杨永毅,赵芳,赵思亮. Docker技术在气象云平台服务中的研究[J]. 机械制造与自动化,2020,49(5):113-115.

[4] 夏正龙,傅承浩,朱亮,等. 基于“天擎”的模式产品云化技术研究与实现[J]. 现代计算机,2021,27(34):46-50.

[5] 应毅,刘亚军,任凯. 基于容器云的分布式深度学习实验平台构建[J]. 实验技术与管理,2022,39(3):147-152.

[6] 武志学. 云计算虚拟化技术的发展与趋势[J]. 计算机应用,2017,37(4):915-923.

[7] 张丽敏,高晶. 微服务环境下容器编排可视化实践研究[J]. 计算机工程与科学,2019(8):1366-1373.

[8] 方意,朱永强,宫学庆. 微服务架构下的分布式事务处理[J]. 计算机应用与软件,2019(1):152-158.

[9] 齐磊,张海峰. 基于容器技术的PaaS云平台方案[J]. 电信科学,2017(4):177-182.

Container application based on CMADaaS

Zhong Yanwen, Xia Zhenglong, Zhu Liang, Lyu Guanru, Zhu Hongwu

(Hunan Meteorological Information Center, Changsha　410118)

Abstract: This paper studies the application of Docker technology in CMADaaS algorithm integration. Taking the hydrological data synchronization algorithm of Hunan Province as an example, this paper describes the process from Image making to algorithm scheduling. The performance of the algorithm is tested by using hydrological precipitation data. The experimental results show that the efficiency of the same algorithm in CMADaaS container cloud is higher than that in virtual machine environment and self-built container server.

Keywords: CMADaaS; Docker; algorithm

文章编号: 1007-1423(2023)01-0114-04 DOI: 10.3969/j.issn.1007-1423.2023.01.021

基于SpringBoot+Vue前后端分离的高校实验室预约管理系统的设计与实现

王以伍，舒 晖*

（成都医学院现代教育技术中心，成都 610500）

摘要：针对高校实验室管理过程中存在的实验室利用率不高，仪器耗材投入产出不成比例，仪器耗材管理不善引发的管理安全问题，以及传统的实验预约效率低下，体验不佳等问题，设计了一套实验室管理系统，以信息化手段实现实验室统筹管理，实现仪器耗材从采购入库、申领、使用、废弃全生命周期的管理，提升了高校信息化水平。

关键词：预约；信息化；前后端分离；SpringBoot

0 引言

高校实验室管理系统是高校教学科研重要组成部分，也是开展实验教学和科研集中地。高校实验室建设与系统管理水平高下直接影响高等教育教学、科研、管理质量和水平。伴随着计算机信息技术飞速发展和高校校园网建设日益成熟，大量的信息化管理平台及信息化新技术运用到实验室管理领域。文献[1]实现了网上预约实验室及器材报修。文献[2]实现了试剂从订单管理到废弃物处理全生命周期的监督管理。文献[3]对实验室管理加入了物联网相关技术的有益尝试，对仪器设备管理更为精确快捷。文献[4]将实验室开放预约打通了微信接口，预约审批更便捷。对高校实验室进行规范化、科学化管理，实现高校实验室数字化、信息化，已经成为高校实验室管理发展必然要求和必然趋势[5]。

1 高校实验室管理存在的问题

实验室安全问题是高等学校最严重的现实威胁，实验室安全管理面临严峻的考验。特别是医学院校实验室设计到医学、药学、生物、化学等学科领域，既有潜在的生物性安全问题，也可能接触到多种危险化学药品，同样还会面临电气安全、消防安全等问题[6]。

实验教学是医学院校教学环节中的重要一环，实验室种类繁多，试剂和仪器种类多且价格昂贵，教师科研和研究生、本科生实验室开放管理任务特别繁重，各种信息和数据种类繁杂，给实验室的管理带来极大的困难[7]。各实验室归口不同的学院进行管理(使用)，没有一个高效的管理渠道对相关信息进行统计(如：实验室使用情况、仪器耗材(药品)入库、出库、使用率)，管理难以精细化。

从实验室管理者角度来看，安全问题是重中之重，部分严格管控的试剂从入库、出库、使用到最后的废弃处理全过程不易监管，对实验室管理带来安全隐患。

从使用者视角来看，师生预约实验室手续繁琐，线下审批流程得不到及时处理及反馈，人工处理相关手续工作效率低，用户体验有待提升。

收稿日期：2022-07-29 **修稿日期：**2022-08-19

基金项目：教育部科技发展中心新一代信息技术创新项目(2020ITA02047)；四川省教育信息技术研究“十四五”规划2021年度课题(川教馆[2021]230)；成都医学院教改项目(JG202017)

作者简介：王以伍(1982—)，男，四川成都人，硕士，工程师，研究方向为网络安全、软件工程、大数据分析；*通信作者：舒晖(1972—)，女，四川成都人，本科，高级工程师，研究方向为现代教育技术，E-mail:shuhui@cmc.edu.cn

2　系统平台实现

本文通过对某医学院校的师生、用户、管理者进行需求调研，旨在解决各类用户在实验室使用过程中的各种问题，以实现实验室基本信息管理、实验室预约管理、仪器耗材管理、使用记录、绩效报告等功能，使实验室资产、耗材使用、维修以及日常使用可以通过系统实现流程化控制，提高管理效率，降低管理成本，提高实验室管理的信息化水平，让数据多跑路，师生少跑腿。推动建立实验室建设项目申报标准，面临审计更简单、更高效。

2.1　系统技术框架

系统选用Java语言进行开发，后台数据库选用MySQL，基于SpringBoot+Vue的前后端分离的开发模式，将前端和后端的项目业务进行分离，可以做到更好的解耦合，前后端之间的交互通过XML或Json的方式，前端主要做用户界面的渲染，后端主要负责业务逻辑和数据的处理。前端只需要独立编写客户端代码，后端也只需要独立编写服务端代码提供数据接即可。前端通过Ajax请求来访问后端的数据接口，将model展示到view中即可。前后端开发者只需要提前约定好接口文档，定义好url、参数、数据类型等参数，然后分别独立开发即可，前端可以造假数据进行测试，完全不需要依赖于后端，最后完成前后端集成即可，真正实现了前后端应用的解耦合，极大地提升了开发效率。前后端分离开发模式如图1所示。

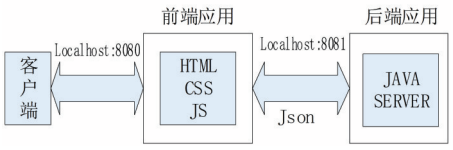


图1　前后端分离开发模式

2.2　系统功能设计

实验室预约管理系统功能设计主要分为两部分，一是系统基础管理部分，包含用户管理（权限、用户组）、消息管理、平台管理，主要实现用户组划分、权限设置、消息审计、创建编辑平台等功能。二是实验室管理模块，主要管理包含了实验室基本信息管理、实验室设置、仪器预约管理、耗材仓库管理、试剂耗材申购/领用等功能模块。功能模块设置如图2所示。

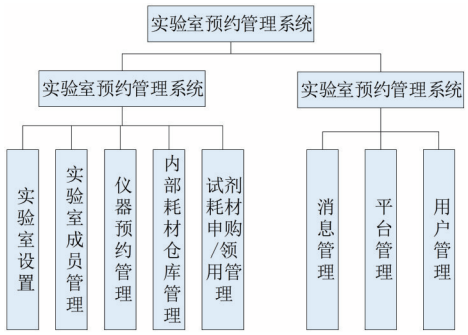


图2　实验室预约管理系统功能模块

2.2.1　仪器耗材管理

仪器耗材管理涉及到所属实验室平台、管理责任人、当前是否可用、是否开放共享、开放给课题组的仪器耗材也会将使用成本进行统计，字段设置尽量将相关属性考虑完整，方便后期统计分析（图3）。仪器耗材表设计如表1所示。

表1　仪器耗材表设计结构

字段	注释
Id	ID
platform_id	实验室ID
platformcat_id	仪器分类ID
title	仪器名称
serial	序列号
brand	品牌
model	仪器型号
chief	负责人
chief_email	负责人邮箱
chief_username	负责人账号
position	位置
img	图片
basefee	基本费
hourfee	小时费
samplefee	样本费用
extmark	仪器介绍
usetype	预约类型
status	状态
ispublic	共享
rules	操作说明
create_by	创建者
created_at	创建时间
update_by	更新者
updated_at	更新时间

图3 仪器耗材编辑

2.2.2 平台设置

平台设置主要是对平台进行分类管理，管理员根据实际情况，将各类实验室设置好实验室默认预约责任人、默认的采购/库存管理员、使用状态、授权类型，将仪器耗材采购权限、预约审批权限下发至相关实验室，实验室根据授权类型可以分为高级教授权限、基础使用授权(图4)。实验室平台表设计如表2所示。

表2 实验平台表设计结构

字段	注释
Id	ID
platform_type	国家级1，省级2，课题组3
name	平台名称
status	状态(0正常 1停用)
auth_type	授权类型
ended_at	授权截止日期
create_by	创建者
created_at	创建时间
update_by	更新者
updated_at	更新时间

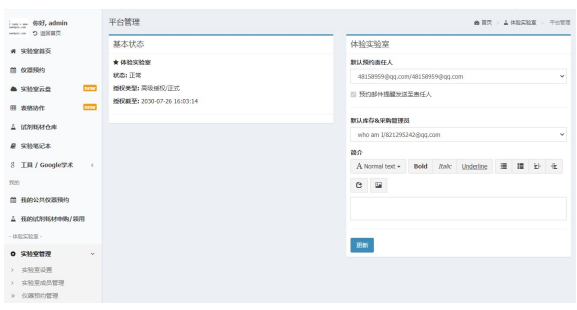


图4 实验平台编辑

2.2.3 预约管理

预约模块可以查看当前实验室平台的状态，进入实验室后查看所需仪器耗材的状态(可否预约)，选择实验小组的责任人，预约的时间段，系统根据预约时间自动预估所需费用，提交预约后由责任人负责审核(图5)。预约数据结构设计如表3所示。

表3 仪器耗材预约表设计结构

字段	注释
Id	ID
payuser_id	预约责任人ID(付费用户)
payuser_username	责任人用户名
payuser_email	责任人email
platform_id	实验室平台ID
user_id	用户id
user_name	用户名
user_email	用户email
equipment_id	仪器耗材id
equipment_title	仪器耗材名称
started_at	预约开始时间
ended_at	预约结束时间
hours	使用时长
sample_quantity	样本数量
fees	费用
est_hours	预估时长
est_sample_quantity	预估样本数量
est_fees	预估费用
desc	备注
create_by	创建者
created_at	创建时间
update_by	更新者
updated_at	更新时间

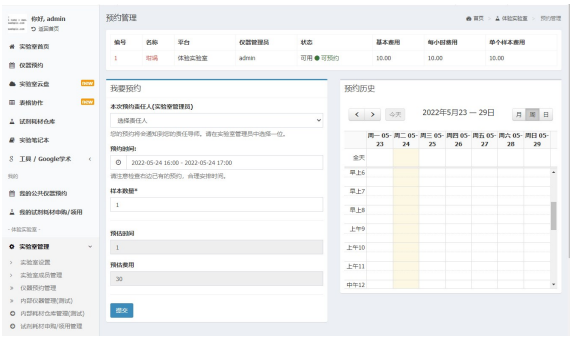


图5 仪器耗材预约界面

3　结语

本文描述了一种基于SpringBoot+Vue的前后分离开发的高校实验室预约管理系统，采用前后端分离的开发模式，将视图开发和业务开发更好地解耦，提高了开发效率。功能上通过信息化管理手段，将实验室平台统筹管理，各实验室通过分级权限下放，实现了仪器耗材归口管理，流程上可以追溯到各仪器耗材全生命周期，避免了重复购买，无度浪费，实验室预约功能的设计极大地方便师生进行实验室预约，提高了实验室管理效率。通过管理平台的建立，将所有实验资产纳入统一管理、调度、进行绩效考核，对实验室相关仪器设备耗材进行全生命周期精细化管理。建立起实验教学信息化管理方式和管理流程，积累的相关数据助力学校教学环境提升决策参考。

参考文献：

[1] 熊捷，崔斌，王玉磊，等. 高校开放式实验室管理平台的设计与实现[J]. 无线互联科技，2020，17(6)：50-52.

[2] 王赟，韩涛，徐四平，等. 高校实验室化学品管理平台的设计与实践[J]. 实验技术与管理，2019，36(12)：256-259.

[3] 姜成. 高校实验室综合管控平台的设计与实现[D]. 济南：山东大学，2018.

[4] 苏东伟，梁志坚，贺秋丽，等. 基于微信平台的高校开放实验室预约系统的设计与实现[J]. 信息通信，2018(1)：160-161.

[5] 牛亮. 实验室预约管理系统开发与应用[J]. 电子技术与软件工程，2018(23)：54-56.

[6] 陈如男，师文文，吴佳玲，等. 基于微信小程序构建开放护理实验室预约平台的需求分析[J]. 实用临床护理学电子杂志，2018，3(47)：170-171.

[7] 刘素转，吴卫江. 实验室开放预约系统的设计与开发[J]. 教育教学论坛，2018(16)：275-276.

Design and implementation of university laboratory reservation management system based on SpringBoot + Vue

Wang Yiwu, Shu Hui*

(Modern Educational Technology Center, Chengdu Medical College, Chengdu 610500)

Abstract: In view of the problems existing in the process of laboratory management in universities, such as low utilization rate of laboratory, disproportionate input-output of instrument consumables, management safety problems caused by poor management of instrument consumables, low efficiency of traditional experiment reservation and poor experience, a set of laboratory management system is designed to resolve the overall management of laboratory by means of information technology, and realize the purchase, warehousing, application, use and the management of the whole life cycle has improved the informatization level of universities.

Keywords: make an appointment; information; front and rear end separation; SpringBoot

文章编号: 1007-1423(2023)01-0118-03 DOI: 10.3969/j.issn.1007-1423.2023.01.022

基于OpenCV的儿童票身高识别

彭安娜，崔建一，张育玮

（辽宁科技大学电信学院，辽宁 114051）

摘要：在大数据时代，人们追求高效便捷的生活。在各个生活生产领域都进行了工业自动化的开发，而目前各个场所的儿童票检查却还依赖于人工。去人工的自动识别，可以大幅度的提高效率，方便数据的管理。为了实现自动化，此项目基于OpenCV对动态的人体进行像素识别，通过固定比例尺来计算人的身高。通过实验可知，该方法可行性高，基本能够实现身高识别。

关键词：机器视觉；OpenCV；图像处理；身高识别

0 引言

随着社会的发展，无人化成为了各个行业的发展方向。但在儿童票身高检测领域依旧使用人工的方式，为提高效率、减少劳动量和运营成本，本文介绍了一种可以应用于儿童票身高识别的方法，可以取代人工，达到判断儿童身高是否符合要求的目的，本方案虽然比不上人工的精准度，也无法达到百分之百的识别成功率，但综合分析来看，在儿童票使用频繁的地方，此方案是更加划算的。

1 实现高效的儿童票检验

根据国家相关法律规定，符合相关标准的儿童可在很多场合使用儿童票，享受相关的优惠政策，目前，不论是火车儿童票、飞机儿童票还是景区儿童门票都是以身高为标准，一般为1.5米以下可享受儿童票优惠。

视觉识别技术现在十分流行，被广泛地应用于生活的方方面面。近年来，火车站、机场以及部分景区都在门口使用上了视觉识别技术，人们买票之后，只需要刷票的同时扫描一下面部，即可完成检票工作，代替了之前的人工检票，既节省了人工又提高了效率。但是，在儿童票检票时，由于要检验儿童是否符合身高标准，所以还是需要人工来测量身高，十分麻烦。所以，本文从这个问题入手，制作一个身高检测系统，可快速检测出被测人的身高，使检票环节做到完全的去人工化，在人流量较大时，该系统优势明显，不但可以节省大量时间，同时还可以排除安全隐患，避免拥堵导致的不良事件发生。使检票流程高效快速，同时也可减少人力资源的浪费，节约运营成本。

2 使用方法

OpenCV是一个开源的计算机视觉开发包，它的优点在于拥有覆盖了机器视觉的大多应用领域的图像处理函数。它的算法都经过了很好的优化[1]。基于OpenCV来实现儿童票身高识别，具有成熟可靠的原理基础。

本方案使用Python下载OpenCV库供编程使用，共两套程序，使用时固定距离放置摄像机，连接电脑，儿童进入通道之后，第一套程序启动摄像机开始录像，将视频传送给电脑；第二套程序启动，电脑识别视频中的动态儿童，将

收稿日期：2022-07-17 **修稿日期：**2022-08-09

基金项目：辽宁科技大学创新创业计划资助项目（X202210146203）

作者简介：彭安娜(2001—)，女，安徽马鞍山人，大学本科，研究方向为通信工程；崔建一(2000—)，男，辽宁灯塔人，大学本科，研究方向为机械设计制造及其自动化；张育玮(2001—)，男，辽宁本溪人，大学本科，研究方向为电气工程及其自动化

图像转化为一个个像素点，待儿童在固定位置站好后，截取关键帧，测量最上方像素点与最下方像素点的距离，乘上对应的比例尺，将得到的数据与标准数据进行比较，既完成了判断过程，符合标准则无反应，不符合标准，则通过语音提醒，之后立马等待下一个目标，继续下一次动态识别。

在此实验过程中，运用的核心技术便是帧间差分法。帧间差分法主要应用场景为检测目标是相对固定的，而运动目标的位置是处于变化的情况下。在该前提下，让相邻两个视频帧进行相减，在理想状态下，背景部分差值为0，而运动区域的差值不为零。这时候选取一个合适的阈值把视频帧的背景过滤掉，从而提取到运动目标[2]。

基于OpenCV的儿童票身高识别具体流程如下：

第一步，进行图像预处理。基于OpenCV的儿童票身高识别，致力于便捷可靠。在实践运用中，取景摄像头为固定的，被检测儿童处于固定距离外。固定的取景位和角度可避免图像倾斜带来的误差。取到图像后，通过预处理突出重点，抑制背景的干扰。预处理操作包括图像灰度化、边缘检测、二值化和倾斜校正[3]，因为儿童多顽皮好动，可能给图像带来影响，所以倾斜校正是必不可少的一步。如图1所示。

第二步，寻找关键帧并求正外接矩形。儿童到达目标位置后站定，帧间分差背景部分差值变为0，截取关键帧。接下来求得人体正外接矩形。人体正外接矩形为平行于x，y轴的规矩图形，进而可得到人体高度像素[4]。如图2所示。

图1　图像预处理

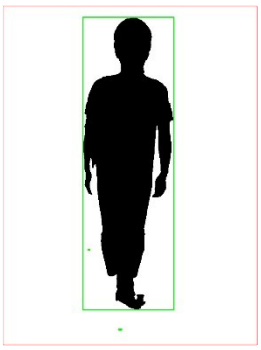


图2　人体正外接矩形

第三步，进行比例换算。根据处理器屏幕分辨率和屏幕尺寸，可得到PPI(每英寸像素的个数)，已知摄像头焦距和物距，可将得到的人体身高像素换算为实际身高。实际身高为(人体身高像素/PPI×25.4)×比例尺。

3　实验结果

对程序进行了多次模拟实验，不断地优化改进，基本可以满足实际的要求，可以证明，此方法可行性高，后期将针对儿童进行大量实验，争取可以早日将本方案投入到实际生产生活中。

身高识别技术目前还存在着许多问题，首先，对于视觉识别技术来说，不可能有百分百的成功率，识别过程中不可避免地会出现识别异常的现象，我们只能降低失败率，无法保证失败率彻底清零。在实验过程中，为了追求结果的理想化和数据的最小误差，实验环境整洁，背景为白墙。而在实际运用中，检票环境是杂乱的，除了被检测儿童以外，还有许多干扰元素，都会对数据结果造成影响。其次，我们服务的对象是儿童，儿童的自主能力较差，可能会出现不配合机器识别的情况，如果在识别过程中，儿童出现跑、跳等动作，机器识别的准确率与成功率都会大幅下降。

4　结语

本项目基于OpenCV对动态的人体进行像素帧识别，计算身高。随着科学技术的发展，数据趋于信息化，效率为生产发展的一大重点。火车站、地铁站等人流量巨大的区域，为了避免拥堵，对效率的要求更加严格。在不久的将来，儿童票身高识别机器的出现是必然的。

参考文献：

[1] 李瑶. Hadoop+OpenCV的集群化图像处理开发环境建设与测试研究[J]. 电脑知识与技术，2016，3(1)：203-205.

[2] 李娜. 行人检测跟踪技术的系统研究与实现[J]. 网络安全技术与应用，2022，4(15)：64-65.

[3] 岳辰. 基于OpenCV的水表读数智能识别系统的软件实现[D]. 成都：电子科技大学，2019.

[4] 孙付超. 基于OpenCV的图像预处理在仪表定位中的应用[D]. 济南：山东大学，2015.

OpenCV-based child ticket height recognition

Peng Anna, Cui Jianyi, Zhang Yuwei

(School of Electronic and Information Engineering, University of Science and Technology Liaoning, Liaoning 114051)

Abatract: In the era of big data, people pursue an efficient and convenient life. Industrial automation has been developed in various areas of life and production, and at present, the inspection of child tickets in various places still relies on manual labor. Automatic identification to manual can greatly improve efficiency and facilitate data management. In order to automate, the project is based on OpenCV, which performs pixel recognition on the dynamic human body and calculates the height of the person by fixing the scale. Through experiments, it can be seen that the method is highly feasible and can basically achieve height recognition.

Keywords: machine vision; OpenCV; image processing; height recognitionx

(上接第49页)

Research on irrigation area discharge prediction model based on grey wolf optimization algorithm

Shan Wuqian[1], Ning Qian[1,2*], Zhou Xinzhi[1], Zhang Renyuan[1], Luo Qiang[3]

(1. College of Electronics and Information Engineering, Sichuan University, Chengdu 610065;
2. School of Computer Science and Technology, Xinjiang Normal University, Urumqi 830054;
3. Chengdu Wanjiang Gangli Technology Co., Ltd., Chengdu 610064)

Abstract: Fast and accurate irrigation area flow forecast is an important basis for water resource allocation. Existing traffic prediction models generally use traditional methods or simple neural network models, and the above methods have defects such as complicated operations or low prediction accuracy. Therefore, this paper proposes to use the gray wolf optimization algorithm to optimize the BP model and design a new GWO-BP model. In order to verify the superiority of the model, the model was applied to the measured flow in Huahai Irrigation District of Gansu Province, and compared with the traditional empirical formula method and BP model. The results show that the GWO-BP prediction results are more stable and reliable, and can meet the engineering accuracy requirements of water resources measurement in irrigation areas.

Keywords: water resource; discharge; gray wolf optimization algorithm; artificial neural network