▶ 现代计算机

XIANDAI JISUANJI 第30卷第8期(总第800期) 半月刊(1984年创刊) 2024年4月25日出版

| 主管单 | 自位 | 中山大学 | | | |
|------|----|--------------------------|--|--|--|
| 主办单 | 包位 | 广州中山大学出版社有限公司 | | | |
| 出版单 | 包位 | 广东现代计算机杂志社有限公司 | | | |
| 发 | 行 | 广东省报刊发行局(全国公开发行) | | | |
| 印 | 刷 | 广州市友盛彩印有限公司 | | | |
| 社 | K | 黄少伟 | | | |
| 主 | 编 | 石玉珍 | | | |
| 编 | 委 | 邹岚萍 熊锡源 李 文 石玉珍 梁嘉珠 | | | |
| 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 | 邮发代码: 46-121 |
|----------------|--------------|
| CN 44-1415/TP | 定价: 30.00元 |





2024年4月

2024年**4**月 第30卷 第8期 (总第800期)

第30卷 第8期 (总第800期

M

现代计算机

• 中山大學出版社 主办





MOL

ISSN 1007-1423 CN 44-1415/TP

N

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

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

版权声明

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

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

3. 对所投稿件,本刊编辑有权根据刊物的需要进行删改或调整。

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

现代计算机

XIANDAI JISUANJI

第30卷 第8期(总第800期)2024年4月25日出版

目 次

研究与开发

| 基于改进YOLOv4的道路病害实时检测模型 黄艳国, 李 罗, 曾东红, | 王丽宁 | (1) |
|--|-----|------|
| 基于信息熵的二维局部二值模式静脉识别 张云飞,李江美, | 陈 熙 | (9) |
| 基于改进 YOLOv7-Tiny的高速公路入口两轮车辆闯入检测 ················ 王 宏, | 田恬 | (17) |
| 基于 FPGA 的语义分割算法设计与实现 | 陶青川 | (24) |
| 基于改进YOLOv5的车辆检测方法研究 ···································· | 王 哲 | (31) |
| 基于 Transformer 的 PM ₂₅ 浓度预测方法 ······ 耀, | 严 华 | (38) |
| 基于深度学习的对抗攻击发展研究 | | |
| ······································ | 蒋正锋 | (44) |
| 一种用于医学图像分割的混合卷积网络 | 谢隆腾 | (50) |
| 基于多尺度注意力机制的单幅图像超分辨率重建 | 严 华 | (56) |
| 基于YOLOv5算法的学生课堂行为识别研究 马瑞珵, 陈 继, 王炳怀, 龙俊丞, | 刘宇 | (62) |
| PyTorch框架下的复杂场景目标识别方法研究 ······ | 张进军 | (66) |
| 一种轻量化的排球自垒姿态检测算法 | 陶青川 | (72) |

实践与经验

| 基于主动学习的实体关系抽取的方法研究 | • • • • • • • • • • • • • • | 孙 | 涵 | (77) |
|-------------------------------------|-----------------------------|----|---|-------|
| 多模态学习投入测评的研究现状与启示——基于CiteSpace可视化分析 | 郑晨虹, | 张海 | 博 | (84) |
| 疫情防控背景下基于人脸识别及测温技术的智能门禁应用 | 左 楷, | 唐耀 | 平 | (89) |
| 基于时空特征交叉融合的网络流量预测 | 黄冬妹, | 宁 | 芊 | (95) |
| 基于神经网络的文本情感分类模型研究 司靖梓, | 邢建川, | 肖 | 鑫 | (100) |
| 面向用水结构研究领域的多阶段实体关系联合抽取方法 | | 陶天 | 然 | (103) |
| 基于S-G滤波算法的航空器着陆阶段滑行轨迹修正研究 | 何 昕, | 王经 | 义 | (108) |

开发案例

| 基于LabVIEW的心电信号与多数据采集分析系统设计 | ••••• | 陈亚华, | 张凯淇, | 马俊 | (112) |
|----------------------------|-------|------|------|-----|-------|
| 基于微信小程序的预约挂号系统的设计与实现 | | 王相喜, | 周晓波, | 倪甜弟 | (116) |

Modern Computer

(Vol. 30, No. 8; Apr. 25, 2024)

CONTENTS

Research and Development

| Real-time road disease detection model based on improved YOLOv4 | (1) |
|---|------|
| Information entropy-based vein recognition for two-dimensional local binary patterns | (9) |
| Two wheeled vehicle intrusion detection at highway entrance based on improved YOLOv7-Tiny | (17) |
| Design and implementation of FPGA-based semantic segmentation algorithm | (24) |
| Research on vehicle detection method based on improved YOLOv5 | (31) |
| A Transformer-based model for PM _{2.5} concentration prediction | (38) |
| Research on the development of adversarial attacks based on deep learning | (44) |
| A hybrid convolutional network for medical image segmentation | (50) |
| Single image super-resolution reconstruction based on multi-scale attention mechanism | (56) |
| Research on student classroom behavior recognition based on YOLOv5 algorithm | (62) |
| Research on object recognition methods in complex scenes under the PyTorch framework | (66) |
| A lightweight posture detection algorithm for volleyball self-digging | (72) |

Practice and Experience

| Research on entity relation extraction based on active learning | (77) |
|---|-------|
| Research status and insights on multimodal learning engagement measurement | |
| —based on CiteSpace visualization analysis ····· | (84) |
| Intelligent access control application based on face recognition and temperature measurement technology | |
| under the background of epidemic prevention and control····· | (89) |
| Network traffic prediction based on spatial-temporal features cross fusion | (95) |
| Research on text sentiment analysis based on neural network | (100) |
| Multi-stage entity and relation joint extraction method for water usage structure research field | (103) |
| Research on aircraft taxi trajectory correction during landing stage based on S-G filter algorithm | (108) |

Development Solution

| Design of ECG signal and multiple data acquisition and analysis system based on LabVIEW | (112) |
|---|-------|
| Design and implementation of appointment registration system based on WeChat mini program | (116) |

研究与开发

文章编号:1007-1423(2024)08-0001-09

DOI: 10.3969/j.issn.1007-1423.2024.08.001

基于改进YOLOv4的道路病害实时检测模型

黄艳国*,李 罗,曾东红,王丽宁

(江西理工大学电气工程与自动化学院,赣州 341000)

摘要:针对道路中存在多类、尺度不一的病害类型导致其检测精度低、检测速率慢以及漏检等问题,提出一种基于改进YOLOv4的道路病害实时检测模型。在网络模型中,首先优化卷积块中的归一化方式,采用组归一化来代替批量归一化,避免因Batch Size 过小而导致检测效果不佳的情况发生;同时对卷积块进行优化,使用深度可分离卷积块来替代原有卷积块,量化网络模型的参数计算量,提高检测速度;最后在检测头中使用自适应非极大值抑制算法,解决非极大值抑制固定阈值引起的小目标漏检和误检问题。实验结果表明,改进后的YOLOv4算法在道路病害检测的检测精度mAP值高达88.64%,检测速度可达37.90帧/秒;与原YOLOv4算法相比,改进后的算法在检测精度上提高了2.89个百分点,同比检测速度增加了10.60帧/秒,且有效解决了漏检现象,进一步提高了在道路病害检测中的实用性。

关键词:道路病害检测; YOLOv4; 组归一化; 深度可分离卷积; 自适应非极大值抑制

0 引言

建设交通强国是我国长期战略任务之一, 护养道路是保障出行安全、推动道路交通高质 量发展的关键。然而随着车流和外界因素影响, 道路路面出现的裂缝、车辙和坑槽等病害类型 也日益增多。道路病害的产生不仅影响行车安 全,增加交通事故风险,还会加剧道路的损坏, 导致道路护养面积和资金的增加,对此采取有 效的目标检测方法对道路病害进行高效、准确 的检测具有重大意义^[1]。

目标检测领域最初使用传统的图像处理方法,通过手动提取图像特征训练分类器。随着深度学习的兴起,目标检测算法不断向优化演进,算法分为One-stage和Two-stage两类。One-stage算法直接从输入图像生成目标的边界框和类别,如SSD^[2]和YOLO系列^[34]等,实时性好,但精度相对Two-stage算法较低。Two-stage算法分解成

候选区域,再进行分类和位置回归,如RCNN^[5]、 Fast RCNN^[6]和Faster RCNN^[7]等,精度优异,但 检测速度较慢,不能满足实时性要求。

在传统图像处理方面,赵芳等^[8]基于 Canny 算法并引入多尺度形态学和双边滤波结合的去噪 方法,优化算子和阈值,提高了检测效率和降低 了误检率,但这种检测方法存在效率一般、成本 高的缺点。在 One-stage 算法方面,Zhang 等^[9]基 于YOLOv3算法,引入了 MLAB多层次注意力机 制,实现了高精度的道路病害检测;安学刚等^[10] 基于YOLOv4算法,引入深度可分离卷积和注意 力机制,优化特征融合模块和损失函数,提高 检测精度和速度;罗晖等^[11-12]基于YOLOv4算法 提出了两种改进方法,其中包括引入深度可分 离卷积和可变形卷积等技术,进一步提高了检 测精度和检测速度。这类改进的 One-stage 检测 算法整体上检测效率和检测精度较好,但存在

收稿日期: 2023-10-25 修稿日期: 2023-11-30

基金项目:国家自然科学基金资助项目(72061016);江西省教育厅科学技术研究项目(GJJ170554);江西省大学生创新创业训练计划项目(S202210407032)

作者简介:*通信作者:黄艳国(1973—),男,湖北武汉人,博士,教授,研究方向为交通信息处理、智能交通、图像处理 及智能控制,E-mail:zscvszjj@163.com; 李罗(1997—),男,江西萍乡人,硕士研究生,研究方向为智能控制、图像处理与目 标检测;曾东红(1987—),男,江西赣州人,硕士,讲师,研究方向为智能控制、超声波电源;王丽宁(2003—),男,河北石家 庄人,本科,研究方向为无人机目标检测

一定的漏检问题。在Two-stage算法方面,Wang 等^[13]调整了Faster RCNN算法模型参数并引入了 数据增强技术,提高了检测精度,但整体检测 效率较低。

上述方法在道路病害检测精度中表现出色, 但在实际环境中道路病害的信息往往模糊不清, 同时尺度差异较大,导致检测模型存在漏检和 效率低下等问题。针对上述问题,本文提出一 种基于改进YOLOv4的道路病害实时检测模型, 通过引入组归一化和深度可分离卷积来提高检 测模型的检测精度和检测速度,并使用自适应 非极大值抑制解决漏检问题。

1 YOLOv4 检测网络

YOLOv4 是一种基于深度学习的目标检测 网络,可实现对图像或视频的实时目标检测和 跟踪。该算法在物体和行人检测等方面性能优 越,随着计算机硬件和算法技术的不断提升, YOLOv4 网络在实时目标检测领域具有广泛的应 用前景。

YOLOv4检测网络结构如图1所示,其中卷 积块(CBM/CBL)作为基础单元,由卷积层、归 一化层和激活函数组成,并被应用于多个模块 中。主干网络由CBM块和多个卷积稀疏编码网 络模块(CSPn)组成,其CSP模块采用多层残差 模块(Res unit)和跨阶段部分的连接方式,主干 网络用于将输入图像进行特征提取。Neck由空 间金字塔池化(SPP)和路径聚合网络(PANet)两 部分构成,其SPP模块通过对特征图进行网格 划分并进行相应的池化操作来提取不同尺度的特征信息;而PANet采用了特殊的上、下采样策略,将三个尺度的特征图融合为三种不同尺度的特征图。Head部分将卷积特征图转换为检测框和类别预测结果,同时使用非极大值抑制(NMS)^[14]算法对预测边界框进行筛选,最终输出检测结果。

2 YOLOv4网络改进

2.1 组归一化

在卷积块中,往往卷积层后会紧随批量归 一化(batch normalization,BN)^[15],来缓解梯度 消失问题并提高模型的训练效率和精度。BN层 是对当前层的输入进行规范化处理,使其输入 在 Batch Size 的维度上呈现出均值为0,标准差 为1的分布状态。然而 BN 层归一化效果受其 Batch Size 的大小影响,采用较大的 Batch Size 可 以改善 BN 层的归一化效果,但会占用更多显存 空间,这会给需要更多显存空间的目标检测算 法带来一定的困难。

引入组归一化(group normalization, GN)^[16], 与其 BN 层的不同之处在于, GN 层沿通道方向 计算均值和方差,避免了对 Batch Size 的依赖, 并且将通道分组,每组包含 C/G 个通道。BN 层 和 GN 层的计算过程如图 2 所示,类似于 BN 层, 不过 GN 层在计算均值和方差时是对每个通道组 进行[C/G, H, W]计算,而不是像 BN 层一样沿着 通道维度计算[N, H, W]的均值和方差。



图 1 YOLOv4网络结构



图 2 BN 层和 GN 层计算过程

GN层的计算公式为

$$\begin{cases} y_i = \gamma \frac{x_i - u_G}{\sigma_G} + \beta \\ \mu_G = \frac{G}{C} \sum_{i=1}^G x_i \\ \sigma_G = \sqrt{\frac{G}{C} \sum_{i=1}^G (x_i - u_G)^2} \end{cases}$$
(1)

式中: $i \in [1, G]$, $C 和 G \in$ Channel 数和 Group 数, $\mu_c \pi \sigma_c \in \mathbb{R}$ 输入 x_i 的均值和标准差, $\gamma \pi \beta \in$ 可学习的参数,用于调整归一化后的输出 y_i 。

在卷积层后紧随着 GN 层, 主要优点是在小 批量训练时的表现更加稳定, 同时也能够适应 不同的数据分布和网络架构, 实现更加灵活的 归一化处理。

2.2 深度可分离卷积

在YOLOv4网络结构中,使用1×1和3×3 标准卷积组合的方式来构建基础模块,其标准 卷积的计算过程如图3所示,对输入的特征图进 行卷积操作以得到输出特征图。若输入特征图 为 $C_i \times H \times W$,输出特征图为 $C_o \times H \times W$,卷积 核为3×3,padding为1,则所需卷积核参数量 为 $C_i \times H \times W \times C_o$,在不考虑标准卷积过程中 的加法运算时,其标准卷积的参数计算量 F_s 公 式为

 $F_s = C_i \times 3 \times 3 \times H \times W \times C_o$ (2) 式中: $C_i \oplus C_o$ 分别是输入通道数和输出通道数, H和W分别为输入特征图的高和宽。



图 3 标准卷积计算过程

引入 MobileNet^[17-18]提出的深度可分离卷积 方法,将卷积操作划分为逐通道卷积和逐点1×1 卷积。逐通道卷积的计算过程如图4所示,使用 3×3的卷积核,对输入图像进行逐通道点乘求 和且不改变通道深度,假设当前特征图为 $C_i \times$ $H \times W$, padding为1,所需卷积核参数量为 $C_i \times$ $H \times W$ 。逐通道卷积的参数计算量 F_d 公式为



图 4 逐通道卷积计算过程

由于逐通道卷积的通道间缺少特征融合, 且无法改变通道数,在逐通道卷积后连接一个 逐点1×1卷积。将逐通道卷积输出的特征图作 为1×1卷积的输入特征图,输出的特征图为 *C_o×H×W*,故逐点1×1卷积的参数计算量*F*₁ 公式为

F₁ = C_i×1×1×H×W×C_o (4) 对比深度可分离卷积与标准卷积的参数计 算量,两者之比值为r,由公式(5)可知在输出 通道数较大时,总体参数计算量约等于标准卷 积的九分之一,极大减少了卷积过程中的参数 计算量,并且在不牺牲特征提取性能的同时还 能提高目标检测速度。

$$r = \frac{F_d + F_1}{F_s}$$

$$= \frac{C_i \times 3 \times 3 \times H \times W + C_i \times H \times W \times C_o}{C_i \times 3 \times 3 \times H \times W \times C_o}$$

$$= \frac{1}{C_o} + \frac{1}{9}$$
(5)

2.3 自适应非极大值抑制

在目标检测的最后阶段,通常采用非极大 值抑制(NMS)算法来筛选预测边界框,以去除 重叠或冗余的边界框,只留下最佳效果的边界 框。NMS的基本思想是对预测框按置信度得分 排序,置信度得分高的框将置信度得分低的框 进行抑制。 NMS算法抑制函数表达式为

$$S_i = \begin{cases} S_i, & IoU(M, b_i) < TH \\ 0, & IoU(M, b_i) > TH \end{cases}$$
(6)

式中: S_i表示每个预测边界框的得分, IoU表示 两个边框的重合程度, M表示置信度最高的预 测边界框, b_i表示其余预测边界框中某一个, TH 为设定的阈值。

由于在道路病害检测中存在病害种类多样 且分布广泛的情况,可能会导致不同类型的病 害相互交叉或密集分布。如果采用NMS算法,则 可能会出现得分较低的、位于不同病害密集区域 内的病害被抑制的情况,从而造成漏检的情况发 生。在NMS算法的基础上引入动态阈值想法,提 出一种自适应非极大值抑制(A-NMS)^[19]算法,其 计算表达式为

$$TH = TH_{\min} + \frac{TH_{\max} - TH_{\min}}{e^{U \cdot I}}$$
(7)

式中: *TH*_{max}、*TH*_{min}为设置的最大、最小阈值, *U*和*I*分别是当前预测边界框与其他预测边界框 的并集面积和交集面积。

当目标稀疏时, U-I值较大,可以选择较小的阈值TH以剔除冗余的边界框;而当目标密集时,U-I值较小,选择较大的阈值TH可以保留正确的边界框。使用A-NMS算法可以根据实际情况来调整阈值,提高检测准确率,降低误检率。

2.4 改进后的网络模型

对YOLOv4网络模型进行改进,具体如图5

所示。在YOLOv4网络模型中,使用GN层替换 卷积块(CBM/CBL)中的BN层,DSC1和DSC2是 采用不同激活函数的深度可分离卷积块,由逐 通道卷积、组归一化、激活函数和逐点卷积组 成,而CDC3和CDC5则是由卷积块和深度可分 离卷积块组合而成的多次卷积模块。在主干网 络中,使用深度可分离卷积块替代了Resunit单 元模块中的第二个卷积块,同时将CSPn模块中 用于高度和宽度压缩的步幅为2的卷积块替换为 深度可分离卷积块。在Neck网络中,使用 CDC3或CDC5模块替换CCL×n模块,同时输 出端的卷积块和卷积层也被深度可分离卷积块 和卷积层所代替。

3 实验结果与分析

3.1 数据集处理

数据集分两部分来源,一部分来自2022年 IEEE大数据国际会议大数据杯组织的全球道路 损伤检测挑战赛(GRDDC)的数据集,选取在中 国境内通过摩托车及无人机拍摄的两个模块共 4878张;另一部分来自自主收集和拍摄国内道 路病害数据集共1012张,提高数据集的完整性。 为全面检测道路状态,将道路护养检测纳入病 害检测范畴,以避免道路护养不到位的情况。 道路病害检测类型包括:路面裂缝(纵向、横 向、鳄鱼)、其他病害(车辙、凹凸不平、坑洼、 分离)和道路养护,见表1。



图 5 改进后的 YOLOv4 网络模型

表 1 道路病害检测类型

| 检测类型 | 详细 | 类别简称 |
|------|---------------|--------|
| 纵向裂缝 | 纵向裂缝的方位 | D00 |
| 横向裂缝 | 横向裂缝的方位 | D10 |
| 鳄鱼裂缝 | 局部路面、整体路面 | D20 |
| 其他病害 | 车辙、凹凸不平、坑洼、分离 | D40 |
| 道路养护 | 道路的保养和维护 | Repair |

两个数据集共5890张图片,其道路病害检测各类型样本数据见表2。为了进行模型训练和 验证,将从这些图片中选取10%作为测试集, 剩下的图片划分为训练集和验证集。其中训练 集和验证集的比例为9:1,即训练集包含4771张 图片,验证集包含530张图片,测试集包含589 张图片。

表 2 道路病害数据集样本分布 单位:张

| 粉捉隹 | | 井井 | | | | |
|----------------|------|------|------|-----|--------|-------|
| 奴 ////未 | D00 | D10 | D20 | D40 | Repair | - 共月 |
| GRDDC | 4104 | 2359 | 934 | 321 | 1046 | 8764 |
| 自制数据 | 622 | 437 | 358 | 239 | 345 | 2001 |
| 总计 | 4726 | 2796 | 1292 | 560 | 1391 | 10765 |

3.2 实验环境及评价指标

3.2.1 实验环境

硬件环境为: 12th Gen Intel(R) Core(TM) i7-12700处理器, NVIDIA GeForce RTX1080Ti显 卡, 11 GB显存和16 GB内存容量。软件环境为: Windows 10操作系统, Python 3.6, CUDA 12.1, PyTorch 1.10.2+cu113。

3.2.2 评价指标

综合考虑模型的漏检、精度和速度这三个 方面,召回率(Recall)可以用于衡量检测模型的 漏检情况,即模型能够正确找到的正例比例;平 均准确率(AP)可用于衡量每个类别的精度表现, 而平均准确率均值(mAP)综合考虑所有类别的性 能表现;速度方面,每秒处理图像数量(FPS)是 评估模型实时性和响应速度的重要指标。

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

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

$$\begin{cases}
AP = \int_{0}^{1} P(R) dR \\
mAP = \frac{\sum_{i=1}^{c} AP_{i}}{c}
\end{cases}$$
(10)

式中: TP、FP、FN分别为真正例、假正例、 假负例, Recall(R)为召回率, Precision(P)为准 确率, c为分类数。

3.3 模型训练

为了提高检测算法的收敛速度并选择更适合 道路病害尺寸的先验框,使用K-means++算法对 道路病害数据集中真实框的宽高尺寸进行重新聚 类,聚类后结果如图6所示。根据道路病害尺寸 特点设定相应的先验框,并将其分配给16×16、 32×32和64×64三种尺寸的特征图,见表3。

表 3 聚类后的先验框尺寸

| 特征图尺寸 | 16 × 16 | 32 × 32 | 64×64 |
|-------|-----------|----------|----------------|
| | 54 × 422 | 34 × 170 | 60×22 |
| 先验框尺寸 | 145 × 265 | 95 × 93 | 24 × 65 |
| | 367 × 257 | 330 × 62 | 158 × 30 |



通过对主干网络模型进行修改,原预训练 权重无法使用,利用 Mosaic 数据增强方法增强 数据,同时设置 600 轮的训练轮数,并采用标签 平滑(label smoothing)技术减少真实样本标签的 类别在计算损失函数时的权重,有效抑制过拟 合,使得网络模型从零开始训练获得显著提升 的效果。 在网络模型训练中,输入数据大小为512× 512,使用 Batch Size为4,采用随机梯度下降 (SGD)优化器,最大学习率为0.01,最小学习率 为0.0001,并使用余弦退火(COS)的学习率下降 方式。同时应用标签平滑为0.005和CloU Loss 目标检测回归损失函数来提高对小目标物体的 检测精度和鲁棒性,使得模型更加稳定地进行 训练。根据图7的损失值曲线所示,训练损失值 和验证损失值随着训练轮数的增加而逐渐减小, 模型的训练效果逐渐体现,当训练轮数达到550 时,模型已基本收敛,通过对比训练损失值和 验证损失值的趋势来评估模型的性能,表明该 模型具有较好的学习效果。



图 7 训练损失值及验证损失值变化曲线

3.4 实验结果与分析

3.4.1 阈值调节

为了验证使用 A-NMS 算法对道路病害数据 集检测性能的影响,在保证精度的同时,通过 调节 *TH*_{max} 和 *TH*_{min} 阈值参数大小,提高模型的召 回率。通过实验测试,对使用 A-NMS 算法和 NMS 算法的召回率进行对比分析,在 *TH*_{max} 和 *TH*_{min} 阈值分别为 0.7 和 0.2 时,两者召回率曲线 对比如图 8 所示。比较同一置信度门限下的召回 率,使用 A-NMS 算法在不同病害类型中表现较 好,除 D20 类型略有下降外,其他类型病害的 召回率均有所提高,表明使用 A-NMS 算法可以 更准确地检测出病害区域,降低模型的漏检率。



图 8 使用 NMS 和 A-NMS 的 召回率曲线

3.4.2 消融实验

为了验证组归一化(GN)、基于深度可分离 卷积(DSC)的网络结构和自适应非极大值抑制 (A-NMS)对YOLOv4算法整体性能的影响,进行 了一系列消融实验,并将结果汇总于表4。通过 比对实验结果,使用了GN、DSC和A-NMS后的 YOLOv4算法,在不同指标上均有不同程度的性 能提升。其中,GN较好地解决了BN在小批量 数据下的效果退化问题,其检测精度提升了 2.03个百分点,由于需要对每个组计算均值和 方差,导致检测速度下降2.28帧/秒;DSC有效 地减少了网络参数数量和计算量,检测精度下 降0.90个百分点,检测速度提高了11.93帧/秒; A-NMS进一步提高了目标检测的精度,与传统 的非极大值抑制算法相比,A-NMS能够更好地 适应不同道路病害类型分布不均匀的情况。

| 标号 | GN | DSC | A-NMS | mAP/% | $FPS/(f \boldsymbol{\cdot} s^{-1})$ | |
|----|--------------|--------------|--------------|-------|-------------------------------------|--|
| 1 | × | × | × | 85.75 | 27.30 | |
| 2 | \checkmark | × | × | 87.78 | 25.02 | |
| 3 | × | \checkmark | × | 84.85 | 39.23 | |
| 4 | × | × | \checkmark | 86.70 | 27.21 | |
| 5 | \checkmark | \checkmark | × | 87.07 | 37.95 | |
| 6 | \checkmark | × | \checkmark | 88.92 | 24.87 | |
| 7 | × | \checkmark | \checkmark | 85.77 | 39.07 | |
| 8 | \checkmark | \checkmark | \checkmark | 88.64 | 37.90 | |

表 4 消融实验结果对比

注:"√"表示使用了对应的方法,"×"表示未使用 对应的方法。

3.4.3 对比实验

为比较不同网络模型在道路病害数据集上的 检测性能,选择了Faster RCNN、SSD、YOLOv3 和YOLOv4等具代表性模型与改进YOLOv4网络 进行对比。考虑到数据集的偶然性影响,每次随 机按9:1比例划分训练集和测试集,并重复了三 次独立实验,表5是三次实验结果的平均值。

表 5 不同网络模型在道路病害数据集上的检测性能对比

| | | | | 评价指标 | ŝ | | |
|----------------|-------------------|-------------------|-------------------|-------------------|----------------------|-----------|---|
| 模型名称 | AP (D00) /% | AP (D10) /% | AP (D20) /% | AP (D40) /% | AP (Repair) /% | mAP /% | $_{\rm FPS}^{\rm FPS}_{\rm /(f\cdot s^{-1})}$ |
| Faster RCNN | 89.04 | 85.15 | 81.25 | 95.50 | 88.67 | 87.92 | 13.50 |
| SSD | 83.23 | 78.41 | 82.17 | 74.48 | 89.26 | 81.51 | 32.14 |
| YOLOv3 | 81.11 | 75.95 | 81.34 | 71.99 | 87.66 | 79.61 | 34.95 |
| YOLOv4 | 86.28 | 83.43 | 80.56 | 89.06 | 89.43 | 85.75 | 27.30 |
| 改进 YOLOv4 | 90.95 | 87.00 | 84.83 | 91.03 | 89.39 | 88.64 | 37.90 |

对比实验结果, Faster RCNN 网络模型检测的 mAP 和 FPS 值分别为 87.92%、13.50 帧/秒, 检测精度远优于 SSD 和 YOLOv3,以牺牲检测速 率提高检测精度; YOLOv4 作为 YOLOv3 的加强 版,其检测精度提高了 6.14个百分点,却牺牲了 7.65 帧/秒的检测速度; 而改进 YOLOv4 网络模型 检测的 mAP 和 FPS 分别为 88.64%、37.90 帧/秒, 综合检测精度相较于YOLOv4和Faster RCNN分 别提高了2.89和0.72个百分点,同时在各类别 检测精度上也较为优异,且检测速率获得进一 步的提升。

3.4.4 检测效果

为了证明改进 YOLOv4 网络模型的优越性, 在不同场景下对比了三种网络模型的检测效果, 结果如图9~图11所示。改进的 YOLOv4 网络模 型在应对小尺度坑洼、斑马线上不明显的裂缝 和道路多病害且密集场景等情况下表现出了优 异的检测效果,能够准确地检测出小尺度的裂 缝和坑槽,同时避免了小尺度道路病害漏检情 况。此外,模型还能对每一类病害进行精准的 分类和定位。



(0) 1942 - 01

图 9 小尺度坑洼检测对比图



图 10 斑马线上不明显的裂纹检测对比图



图 11 道路多病害且密集场景对比图

4 结语

为解决道路中存在多类、尺度不一的病害

类型导致检测精度低、检测速度慢及漏检问题, 影响道路病害维护,研究在YOLOv4网络模型基础上进行改进。通过引入深度可分离卷积,并 对归一化和NMS进行优化,从而提高了检测的 精度和速率。方法在GRDDC数据集上的表现优 于其他算法,验证了方法的有效性。

考虑到道路养护的实际需求,该方法可应 用于智能化的道路检测系统。通过在道路检查 车载摄像头获取图像,利用该方法可以实时检 测各类道路病害。系统后续可根据检测结果指 导维修工作,实现道路病害的智能化、精确化 养护。这不仅大大提高了道路养护质量,也降 低了维护成本。

参考文献:

- [1] 秦美香.公路路面裂缝养护技术分析[J].运输经理 世界,2022,647(1):116-118.
- [2] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C] //Proceedings of the 14th European Conference on Computer Vision, Amsterdam, The Netherlands. Springer International Publishing, 2016:21-37.
- [3] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. arXiv:1804.02767,2018.
- [4] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. arXiv:2004.10934,2020.
- [5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014:580-587.
- [6] GIRSHICK R. Fast R-CNN [C] //Proceedings of the IEEE International Conference on Computer Vision, 2015:1440-1448.
- [7] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39 (6) : 1137-1149.
- [8] 赵芳,周旺辉,陈岳涛,等.改进的Canny算子在裂 缝检测中的应用[J].电子测量技术,2018,41(20):

107-111.

- [9] ZHANG Y, ZUO Z, XU X, et al. Road damage detection using UAV images based on multi-level attention mechanism [J]. Automation in Construction, 2022,144:104613.
- [10] 安学刚,党建武,王阳萍,等.基于改进YOLOv4的 无人机影像路面病害检测方法[J].无线电工程, 2023,53(6):1285-1294.
- [11] 罗晖, 贾晨, 李健. 基于改进 YOLOv4 的公路路面 病害检测算法[J]. 激光与光电子学进展, 2021, 58 (14): 336-344.
- [12] 罗晖,余俊英,涂所成.基于深度学习的公路路面病 害检测算法[J].科学技术与工程,2022,22(13): 5299-5305.
- [13] WANG W, WU B, YANG S, et al. Road damage detection and classification with faster R-CNN [C] //Proceedings of the 2018 IEEE International Conference on Big Data(Big data), 2018:5220-5223.
- [14] NEUBECK A, VAN GOOL L. Efficient non-maximum suppression [C] //Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06),2006,3:850-855.
- [15] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C] //Proceedings of the International Conference on Machine Learning, 2015: 448-456.
- [16] WU Y, HE K. Group normalization [C] // Proceedings of the European Conference on Computer Vision (ECCV), 2018:3-19.
- [17] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications [EB/OL]. arXiv: 1704.04861,2017.
- [18] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: inverted residuals and linear bottlenecks
 [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4510-4520.
- [19] LIU S, HUANG D, WANG Y. Adaptive NMS: refining pedestrian detection in a crowd[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019:6459-6468.

(下转第37页)

文章编号:1007-1423(2024)08-0009-09

DOI: 10.3969/j.issn.1007-1423.2024.08.002

基于信息熵的二维局部二值模式静脉识别

张云飞,李江美,陈 熙*

(贵州师范大学大数据与计算机科学学院,贵阳 550025)

摘要:基于现有LBP算法及其变体无法提取图像高维特征的问题,提出一种基于信息熵的二维局部二值模式识别算法。此方法首先利用统一局部二值模式(ULBP)对图像进行低维特征的提取,随后将图像信息熵与统一局部二值模式 图谱进行结合获取熵值加权的统一局部二值模式图谱(EULBP),并利用滑动窗口实现对局部区域内模式间共现特征信 息的统计,以其结果作为图像特征表达。并以直方图交叉距离为基础构建模式分类器,验证其识别性能。实验结果表 明,在SDUMLA-HMT数据集以及马来西亚理工大学指静脉数据集(FV-USM)中,提出的算法能取得99.94%和98.84% 的平均识别率。

关键词:二维共现局部二值模式;信息熵;旋转不变;方向特征

0 引言

生物特征识别技术是一种基于人体生理或 行为特征的识别技术,常见生物特征识别系统 包括基于指纹^[1]、掌纹^[23]、手部几何形状^[4]、 虹膜^[5]、面部图像^[6]等。静脉识别是基于人类身 体中的静脉所形成的特殊网状结构来进行识别 的,它具有唯一性、普适性、永久性等优点。 并且静脉成像需要近红外(NIR)光来显示人体皮 肤内部复杂的血管结构,相对于基于人体外在 生物特征的识别方法具有更强大的抗篡改能力, 因此受到越来越多的学者关注。

Ojala 等^[7]提出的局部二值模式(LBP)是一 种有效的纹理描述符,它不仅在人脸、纹理等 外在生物特征具有较好的识别性,Zhang等^[8]在 静脉识别之中也证明其具有优良的表现。然而 传统的局部二值模式对于旋转等几何变化表现 敏感。因此越来越多的抗噪声、抗旋转的变体 被提出,如基于编码之中旋转不变性信息的等 价局部二值模式(ULBP^[7])、具有多维信息维度 的完全局部二值模式(CLBP^[9])、基于对称像素 灰度差值的中心对称局部二值模式(CS-LBP^[10])、 通过改变阈值化函数而进行编码的局部三元模 式(LTP^[11])、基于Kirsch算子获取方向响应度参 数进行阈值编码的局部方向模式(LDP^[12])以及 基于线方向的局部线方向模式(LLDP^[13])。然而 这些变体均是基于自体出发的局部二值模式, 它们对于尺度变化的适应性较低,而共现模式 可以很好地解决这一问题,通过对不同模式共 现信息的提取可以增大纹理描述野,丰富提取 的特征信息、增强纹理描述符的识别能力。 Nosaka 等^[14]提出的共现局部二值模式(CoLBP) 通过设定不同灰度纹理形状进行编码,并利用 共现矩阵实现在局部区域内共现信息的提取, 提高了纹理描述符的感受野。Qi等[15]提出了成 对旋转不变共现局部二值模式(PRICoLBP),实 现了共生对的旋转不变特性。Bi 等^[16]提出的二 维局部二值模式(2D-LBP)则是以LBP模式为基 础建立的共现模式,并且其还具有通用性,在

• 9 •

收稿日期: 2023-11-08 修稿日期: 2024-01-09

基金项目:国家自然科学基金资助项目(61762022);湖南省教育科学"十四五"规划2023年课题(XJK23BJC011)

作者简介:张云飞(1999—),男,河南周口人,硕士研究生,主要研究方向为图像处理、模式识别;李江美(1999—), 女,贵州仁怀人,硕士研究生,主要研究方向为图像处理、模式识别;*通信作者:陈熙(1976—),男,湖南益阳人,副教授,硕导, 博士,主要研究方向为模式识别、生物特征识别,E-mail:13164352363@163.com

LTP、ULBP 编码之上也能进行共现模式的计数 操作。然而这种方式只是简单地将共现模式进 行加权计数统计,不具有旋转不变性。共现梯 度方向信息是模式分类之中重要的区分信息, 缺少共现梯度方向信息极易使得不同的纹理结 构产生相同的共现信息,导致分类错误,如图1 所示。



图 1 具有相同模式对的不同纹理结构

因此,本文在此基础上做出改进,通过利 用等价局部二值模式(ULBP)进行原始图像的基 础编码,提取原始灰度图像之中的旋转不变特 性,并在特征图谱之上利用三维共现矩阵进行 共现模式的统计。这种方法通过加入共现梯度 方向增强纹理描述符的描述能力,并在特征提 取过程中通过信息熵函数实现对分块图像特征 的加权联立。在SDUMLA-HMT数据集以及马来 西亚理工大学指静脉数据集(FV-USM)上进行了 仿真实验,证明本文提出的方法具有较高的识 别性能以及较强的抗噪声能力。

1 LBP算法及其变体的相关回顾

1.1 局部二值模式(LBP)

Ojala 等^[7]提出的局部二值模式(local binary patterns, LBP)是一种有效的灰度不变纹理描述 符。它通过阈值化中心像素与邻域像素的灰度差 实现对中心像素的LBP编码,进而实现对局部纹 理特征的提取,最后通过LBP模式直方图进行图 像的特征表达。其数学描述如式(1)所示:

$$LBP_{N,R} = \sum_{i=1}^{N-1} S(g_i - g_c) 2^i, S(x) = \begin{cases} 1, x \ge 0\\ 0, \text{ otherwise} \end{cases}$$
(1)

其中: g_e 表示中心像素灰度, g_i ($i = 1, \dots, N$)是 邻域像素灰度值, N是以R为半径的邻域内邻域 像素的数量。 $S(\cdot)$ 为阈值化函数。

1.2 等价局部二值模式(ULBP)

传统LBP模式在*R* = 1的邻域空间内会产生 256种LBP模式编码,而大多数的LBP在图像之 中出现的频次极低,因此Ojala等^[7]提出一种等 价LBP模式(*LBP*^{*}_{N,R})编码方式。它将传统LBP模 式二值编码首尾相接,计算环形编码之中邻近 两位编码的跳变次数,并将跳变次数小于等于 两次的LBP编码划分为"等价模式",其余的划 分为"混合模式"。通过这种方式,LBP模式编 码从 256种锐减至 59种。其中LBP模式编码跳 变次数计算式如式(2)所示:

$$U(LBP_{N,R}) = \sum_{i=0}^{N} |S(g_i - g_c)| - |S(g_{i+1} - g_c)| \quad (2)$$

其中: $g_n = g_{00}$ 、为了增强对图像旋转变换的鲁 棒性, Ojala 等^[7]基于 LBP 模式编码之中的旋转 不变信息提出了旋转不变等价局部二值模式 ($LBP_{N,R}^{m}$)。其数学描述如式(3)所示:

$$LBP_{N,R}^{nu} = \begin{cases} \sum_{i=0}^{N} S(g_i - g_c), & U(LBP_{N,R}) \leq 2\\ 8R + 1, & \text{otherwise} \end{cases}$$
(3)

等价模式通过将编码之中的二进制数字进行累加获取*LBP^m*_{N,R}编码,而混合模式则直接以 8*R* + 1 作为其模式编码。通过这种编码方式实现了LBP模式的降维,将256类LBP编码缩减至 10类编码,在保留大量信息的基础上很好地提取到了编码之中的旋转不变信息。

1.3 二维局部二值模式(2D-LBP)

2D-LBP特征提取算法是Bi等^[16]提出的一种 能够提取图像高维特征的提取方法,其特征提 取流程如下。首先,对目标图像使用传统LBP 算子进行特征提取,获取对应的特征图谱。为 获取局部区域内的空间上下文信息,Bi等^[16]提 出了模式对的概念,即将局部区域内处于不同 位置的两个LBP模式视作一组模式对,以局部 区域内不同模式对的出现频次作为信息的表达。 其模式对的定义如式(4)所示:

 $P_{co}(t, \Delta t) = (LBP_r^{riu2}(v_t), LBP_r^{riu2}(v_{t+\Delta t}))$ (4) 为方便算法的描述,目标图像被定义为图像*I*, 大小为 $M \times N_{\circ} P_{\circ\circ}(t, \Delta t) \in \{(0, 0), \dots, (8r+1, 8r+1)\}$ 则表示所有可能产生的模式对的组合情况。t = (x, y)表示像素的空间位置坐标,其中 $x \in [r+1, M-r]$, $y \in [r+1, M-r]$ 。 $\Delta t = (\Delta x, \Delta y)$ 表示在设定的 $a \times a$ 大小的局部区域内LBP模式对之间的空间位置增量,其中 $\Delta x \in (-a/2, a/2)$, $\Delta y \in (-a/2, a/2)$ 。随后通过全局遍历LBP特征图的每一个位置,实现对每个像素点及其周围区域 P_{co} 出现频次的统计,其数学描述如式(5)所示:

$$2DLBP_{r}^{I}(P_{1}, P_{2}) = \sum_{(x, y) \in I} f(P_{c0}, P)$$
(5)

其中: $P = (P_1, P_2) \in \{0 \le P_1 \le (8r+1), 0 \le P_2 \le (8r+1)\},$ 而对于 $f(\cdot)$, 其定义如公式(6)所示: $f(P_{co}, P) = \begin{cases} 1, LBP_r^{riu2}(v_i) = P_1, LBP_r^{riu2}(v_{i+\Delta i}) = P_2 \\ 0, \text{ otherwise} \end{cases}$

(6) 通过上述方式,可以实现对不同模式出现 频次的统计。随后将统计的计数矩阵一维展开 转化为直方图,用以对目标图像的信息表达。

2D-LBP只是简单地将模式对的出现频次进 行统计,而未对模式对的空间位置信息进行表 达。因此本文提出了一种基于共现梯度方向进 行分类的二维局部二值模式。通过将不同空间 位置的模式对分别进行计数来增强纹理描述子 的描述能力。

2 相关工作

2.1 二维局部二值共现模式

为增强对图像旋转变换的鲁棒性,我们提出了一种基于方向特征的二维局部二值模式

(2DCOLBP)。以各个模式对共现梯度方向为索引,分别统计其模式对的共现频次,以增强对 于方向的敏感性,提高特征描述能力,其结构 如图2所示。

首先,将一张大小为M×N的图像定义为图 像I。传统图像极易受到光照的影响,因此我们 通过对图像I进行具有灰度不变性的LBP特征的 提取,获取具有灰度不变性的模式图谱I-lbp。 然而传统LBP对于旋转等尺度变换表现出较强 的敏感性,并且存在一定的编码冗余。为增强 特征描述符对于图像旋转不变性特征的提取, 减少运算时间,我们采用旋转不变等价局部二 值模式代替原有的LBP编码,并获取对应的模 式图谱I-ulbp。

传统 LBP 及其基于自身的变体多只能从自 身出发,提取小范围的纹理特征,我们以目标 像素点为中心建立一个大小为 $l \times l$ 的矩形区域, 增大描述符的感受野,丰富局部特征信息,增强 特征描述符的描述能力。为进行局部区域内的共 现信息的提取,我们将目标像素点位置设定为主 模式,以 P_m 表示,其余邻居像素点位置设定为主 模式,以 P_m 表示。其中 $D_{ij}(0 < i < 1, 0 < j < 1)$ 表 示在矩形区内所有像素点对应的模式。以水平 向右方向为其标准正方向,其矢量表达为 $(\overline{1,0})$ 。 另外,将连接主模式与副模式方向与水平方向 形成的夹角视为共现梯度方向,记 θ_{MD_i} 为其数学 表达,计算如式(7)所示:

$$\theta_{MD_{ij}} = \arccos\left(\frac{\left[x_{1} - x_{2}, y_{1} - y_{2}\right] \cdot \left[\begin{bmatrix}1\\0\end{bmatrix}\right]}{len\left(P_{m}, P_{D_{ij}}\right)}\right) \quad (7)$$



图 2 总体结构示意图

其中: (x_1, y_1) 为主模式 P_m 的空间位置坐标, (x_2, y_2) 为 P_{D_y} 的空间位置。 $len(P_m, P_{D_y})$ 为计算两 坐标之间的空间距离。通过标准单位正方向进 行矩阵运算获取副模式所处的空间位置方向。 最终一组完整的模式对表达为 $[P_m, P_{D_y}, \theta_{MD_y}]$,其 数学表达式如式(8)所示:

$$P_{D_{ij}}(t,\Delta t,\theta) = \left(ULBP_{r}^{riu2}(v_{t}), ULBP_{r}^{riu2}(v_{t+\Delta t}), \theta_{MD_{ij}}\right)$$
(8)

其中:
$$P_u = (P_F, P_s, \theta_m) \in \{0 \le P_F, P_s \le (8r+1), \theta_m \in Si\}$$
, 对于公式 $f(\cdot)$ 的定义如式(9)所示:

$$f(P_{MV}, P_u) = \begin{cases} |l_{MV}|, & P_{MV} = P_u \\ 0, & \text{otherwise} \end{cases}$$
(9)

当 $ULBP_{r}^{riu2}(v_{t}) = P_{F}$, $ULBP_{r}^{riu2}(v_{t+\Delta t}) = P_{s}$, $\theta_{MV} = \theta_{m}$ 时,即意味着 $P_{MV} = P_{m}$ 。此时将主副模 式空间位置距离作为权重比进行加权计数,而 $|l_{MV}| = (\Delta x^{2} + \Delta y^{2})^{1/2}$ 。

2.2 方向标准化策略

通过将不同共现梯度方向上的模式对分别 记录可以增强纹理描述符的特征描述能力,然 而随着局部区域L的增大,其中所涵盖的空间方 向也愈多,如当L = 3时, \mathbf{R}_{θ} 包含8种方向;当 L = 5时, \mathbf{R}_{θ} 包含18种方向;L = 7时, \mathbf{R}_{θ} 包含 了 32种方向。较大的局部空间有利于局部信息 的提取,但随着L的增大,图像特征向量的数据 维度也随之产生爆炸式增长。因此,我们设计 了一种方向标准化策略,将复杂纷乱的多方向 共现矩阵归一化为标准方向上的共现矩阵。

首先定义水平向右方向为标准正方向,即 $V_1 = (1,0)$,以原点为轴心,逆时针旋转每经过 45° 定义一个线方向向量,即 $V_1 = (1,0)$, $V_2 = (1,1)$, $V_3 = (0,1)$, $V_4 = (-1,1)$, $V_5 = (-1,0)$, $V_6 = (1,-1)$, $V_7 = (0,-1)$, $V_8 = (1,-1)$ 。将以主模式 为起点,副模式为终点所连接而成的模式对向 量定义为 $V_{mD_g} = (x_1 - x_2, y_1 - y_2)$ 。

$$m_{i} = \frac{V_{1} \cdot V_{mD_{i}}}{|V_{1}|} \quad i = 1, 2, \cdots, 8$$
(10)

将模式对向量与线方向向量做内积获取在 不同方向上的响应度*m_i*(*i* = 1, 2, …, 8),将方向 响应度按照从大到小顺序进行排序,获取最大 响应度索引编号 i_a 与次最大相应度索引编号 i_b 。 对应的 $\theta_f = i_a \times \pi/4$ 为模式对矢量最近邻方向, $\theta_s = i_b \times \pi/4$ 为模式对矢量次近邻方向。而模式 对矢量 V_{mD_g} 方向处于两者形成的夹角空间之中。 利用向量的平行四边形分解机制,将 $|V_{mD_g}|$ 作为 权重进行分解,以此来实现对方向信息的数据 降维。

$$f_{l}(P_{MD}, P_{u}) = \begin{cases} |l_{MD}|, & P_{MV} = P_{u}, \theta_{s} = \theta_{f} \\ |l_{MD}| \times G(\theta_{s}, \theta_{f}, \theta_{MD}), & P_{MV} = P_{u}, \theta_{s} \neq \theta_{f} \\ 0, & \text{otherwise} \end{cases}$$

$$(11)$$

$$G(\theta_{s}, \theta_{f}, \theta_{MD}) = \cos(|\theta_{f} - \theta_{MD}|) + \cos(|\theta_{s} - \theta_{MD}|)$$

$$(12)$$

其中: f₁为新模式对权重计量方式。当模式对向 量处于标准方向时,直接将其向量模长作为计 量权重;若处于某个标准方向区间时,则将其 权重进行向量分解。

2.3 信息熵加权的二维局部二值共现模式

在对图像进行模式识别之中,通常将图片 进行分块处理,不同的分块会对最终的识别结 果造成一定的波动。其原因在于,图像经过分 块之后,在不同的分块之中包含不同的纹理结 构,有些信息丰富,有些信息稀少。信息熵是 克劳德·香农在十九世纪四十年代根据热力学 定理提出的^[17]。它描述的是信息源的不确定性, 当某段信息出现的频率较高时,其信息量就越 大,信息熵也就越高,通过衡量不同分块的信 息熵的大小为不同分块赋予不同的权重,可以 提高重要特征的表达,提升识别率。其信息熵 与其权值的计算公式如式(13)所示:

$$w_i = \frac{E_i}{\sum_{i=1}^{m} E_i}$$
(13)

其中: w_i 为块i的权重, E_i 为块i的信息熵值,m为图像的分块数。

以8个分块为例,我们将原始图像进行分 块之后分别求取其2DCoLBP的特征向量,并计 算其对应的信息熵值,将信息熵作为此块对应 的系数得到2DECoLBP的直方统计图,最后将 各块的特征进行归一化级联之后得到图像的 2DECoLBP特征,其特征提取过程如图3所示。



图 3 2DECoLBP特征提取示意图

2.4 基于二维共现局部二值模式的静脉图像 识别

手指静脉图像的识别过程主要包括训练过 程与识别过程两部分,训练过程首先将图像通 过特征提取算法提取对应的特征向量并保存于 训练数据库之中,随后对待识别图像进行特征 提取,并将其和训练数据库之中的数据进行一 一比对,识别出待识别图像属于训练数据库中 的哪一类,其流程图如图4所示。



图 4 识别算法流程图

首先对静脉图像进行平滑滤波以及ROI区 域分割等预处理操作,并归一化为120×240大 小,随后根据归一化的图像进行切割分块,将 图像分成大小相同的m个子块,并对每个子块 进行二维局部二值特征的提取,并根据式(13) 获取对应的信息熵加权系数,将系数矩阵与对 应块的特征相乘之后获得信息熵加权 2DCoLBP 特征,最后利用基于欧氏距离的模式分类器进 行相似度计量,相似度最大的样本类即为对应 的类。

3 实验结果及分析

为了分析和评估本文所提出的基于信息熵 的二维共现局部二值模式在静脉识别中的各项 性能,我们设计实现了多组对比仿真实验,其 中包括最优参数化调整、经典纹理方法对比、 噪声环境下算法的鲁棒性测试。本次实验选取 了SDUMLA手指静脉数据库和马来西亚理工大 学指静脉数据集(FV-USM)作为实验样本。

3.1 SDUMLA数据库

SDUMLA-HMT手指静脉数据库是山东大学 机器学习与数据挖掘实验室采集的一个开放数 据库,它采集了106位志愿者的双手食指、中指 以及无名指的图像。每个手指重复采集6次,得 到6个手指静脉图像,共计3816幅图像,每幅 图像大小为320×240像素。

3.2 马来西亚理工大学指静脉数据集(FV-USM)

FV-USM数据集采集于123名志愿者,其中 包括83名男性40名女性,其年龄从20~52岁不 等,每个研究对象提供四个手指:左食指、左 中指、右食指和右中指,共获取492个手指类型 的5904个图像,其大小为640×480像素。

3.3 手指图像预处理

手指静脉图像的识别,主要以手指静脉网 的特异性作为主要区分特征, 而采集的图像存 在许多手指未覆盖的干扰区域,为提升算法的 识别率我们对原始图像进行了预处理操作,如 图5所示。首先对图像进行均衡化处理以增强手 指静脉的显像度。而在图像采集过程中志愿者 的手指均出现了不同程度的不规则扭转。然后 我们使用滤波、结构元腐蚀等手段清除图像之 中的干扰元素, 使得图像边缘更加明显, 并利 用 Sobel 边缘检测算法对图像进行边缘化检测, 获取图像的边缘骨骼化图像。为求取手指图像 扭转程度,我们将骨骼化图像细化处理,并剔 除较短的干扰细线,利用手指上下边缘细线求 取手指中线扭转度。最后,利用上下边界和左 右边界的频数分布图实现对原始图像感兴趣区 域(region of interest, ROI)的精准选取,最终归

 (a) 原始图片
 (b) 滤波
 (c) 均衡化

 (d) 结构元腐蚀
 (e) 二值图像
 (f) 区域划分

 (g) ROI图
 (g) ROI图

图 5 ROI提取示意图

3.4 识别率评价方法

为评价不同纹理图像分类方法的分类精度, 我们使用文献[18]所使用的正确分类百分比 (CCPs)对其性能进行测评。其公式如式(14)所示:

 $Accuray(CCPs) = \frac{TP + TN}{TP + TN + FN + FP} \times 100\%$ (14)

其中: TP表示被正确分配到目标类的样本数, TN表示未被正确分配给分类样本的数量, FN 表示错误分配给目标类的样本数, FP表示错误 分配给其他类的样本数。

在实验中,为了得到具有统计学意义的实验结果,我们对实验所用数据集进行了50次的随机划分,将一部分图像样本用于训练,一部分图像用于测试。并将这50次独立划分所产生实验结果作均值运算,以均值结果作为最终识别结果。

3.5 参数优化

本文所提出的二维局部二值共现模式,通 过对局部区域内不同共现梯度方向上的等价模 式进行加权统计,实现对图像特征的提取,并 且基于信息熵为各个分块进行了权值附加,因 此不同的分块情况会影响最后的识别结果,我 们通过设计不同的分块方式实现对分块参数的 优化,以寻找最优的分块情况和训练样本比例。

从图6可以看出,随着对图像的分块更加精 细、训练样本更加丰富,图像的识别也更加准 确。本文设置了9种不同的分块,处于1×1分块 下的准确率最低。但随着横向分割更加精细, 图像识别率呈现一个上升的态势,并在3×3的情 况下达到了95.85%的识别率。经过观察可以发 现,图像在2×1、3×1的分块之下存在明显的识 别率下降。与其前方的临近分块方案相对比, 不难发现其横向分块增加、纵向分块减少导致 了其识别率下降。但其对于1×2、1×3分块方案 的总子块数量却没有发生改变,由此我们可以 得到以下结论。



图 6 分块性能对比图

手指静脉的走向在所采集的数据集中呈现 横向分布,横向的切分会破坏静脉的整体特征。 由于静脉血管的狭长走向,图像纵向的切分有 助于静脉特征的分离,增强纹理描述符的特征

一化为120×240大小图像。

提取能力,提高图像识别精度。所以在同等切 块数量的情况下,尽可能多的竖向切片有助于 提高算法的识别率。

3.6 SDUMLA数据集下的噪声鲁棒性检测

图像在生成与传输的过程中经常受到各种 噪声的干扰,这对于后续的图像识别会产生一 些不利的影响。因此,一个有效的图像识别方 法应该对于噪声攻击具有较强的鲁棒性,为测 试本文所提出的2DECoLBP在各种噪声攻击下的 鲁棒性和分类性能,我们在两个数据库上进行 了相关实验。

实验中我们分别为原始图像添加了信噪比 (SDR)为10的高斯噪声与瑞利噪声,通过表1 可以看出,在不同噪声的干扰下,当训练量较 小时本文所提出方法劣于其他对比方法;当训 练量超过总样本的30%时,本文所提出的方法 超越其余方法,最终达到99.94%的识别率。并 且本文所提出的方法具有较强的鲁棒性,在不 同的噪声干扰下均能够拥有较高的识别率。

| 噪声 类型 | Methods | 1 Train examples | 2 Train examples | 3 Train examples | 4 Train examples | 5 Train examples |
|----------|---------|---------------------|---------------------|---------------------|---------------------|---------------------|
| | LBP | 0.8520 | 0.9178 | 0.9603 | 0.9740 | 0.9860 |
| 原始 | CLBP | 0.8536 | 0.9105 | 0.9467 | 0.9815 | 0.9870 |
| 图像 | 2D-LBP | 0.8082 | 0.8920 | 0.9637 | 0.9770 | 0.9890 |
| Ours | Ours | 0.8562 | 0.9253 | 0.9840 | 0.9975 | 0.9994 |
| | LBP | 0.8430 | 0.9128 | 0.9570 | 0.9680 | 0.9840 |
| 高斯 | CLBP | 0.8428 | 0.9140 | 0.9450 | 0.9775 | 0.9810 |
| 噪声 | 2D-LBP | 0.8322 | 0.8923 | 0.9483 | 0.9624 | 0.9830 |
| Ours | Ours | 0.8398 | 0.9103 | 0.9733 | 0.9845 | 0.9970 |
| | LBP | 0.7330 | 0.8238 | 0.8800 | 0.9025 | 0.9360 |
| 瑞利 | CLBP | 0.7242 | 0.8383 | 0.8320 | 0.9085 | 0.9430 |
| 噪声 | 2D-LBP | 0.7634 | 0.8544 | 0.8870 | 0.9225 | 0.9532 |
| Our | Ours | 0.7802 | 0.8928 | 0.9320 | 0.9615 | 0.9610 |

表 1 不同噪声下的识别率对比

3.7 识别率对比与识别时间对比

为验证本文算法与目前算法识别效果的优劣 以及识别速度,我们选取了一些经典算法与之进 行对比,并以SDUMLA-HMT数据集及FV-USM 数据集作为实验数据集。 表2展示了在两个数据集上本文所提基于信 息熵的二维局部二值共现模式(2DECoLBP)与其 余经典算法之间的对比效果,表3则展示了算法 的识别时间。本文所提方法在识别率方面对比其 余算法拥有较为突出的表现,在SDUMLA-HMT 数据集上达到了99.94%的高识别率。在识别时 间方面,2DECoLBP以8方向为索引分别建立了 共现矩阵,其数据维度达到了800维,对比LBP 其余算法,虽然提高了识别率但对于数据维度 的增长也是巨大的。因此在同等设备之下对于 其余算法拥有较长的识别时间,但随着硬件设 备性能的提升可以有效解决这一问题。

表 2 不同算法下的识别效果(%)

| Madaala | 识别率 | | | | |
|----------------------------|---------------|------------|--|--|--|
| Methods | SDUMLA-HMT数据集 | FV-USM 数据集 | | | |
| 2DECoLBP | 99.94 | 98.84 | | | |
| 2DCoLBP | 98.54 | 97.75 | | | |
| LBP ^[7] | 98.72 | 98.50 | | | |
| LCCP ^[19] | 96.54 | 96.24 | | | |
| CLCCP ^[12] | 97.40 | 96.52 | | | |
| CLBP-S/M/C ^[12] | 98.35 | 97.94 | | | |
| $LTP^{[10]}$ | 96.43 | 95.52 | | | |
| CS-LBP ^[9] | 96.30 | 96.12 | | | |
| | | | | | |

表 3 不同算法下的识别时间

单位:s

| Mr. I | 识别时间 | | | | |
|----------------------------|---------------|-----------|--|--|--|
| Methods | SDUMLA-HMT数据集 | FV-USM数据集 | | | |
| 2DECoLBP | 6.0262 | 5.9276 | | | |
| 2DCoLBP | 5.4234 | 5.0274 | | | |
| LBP ^[7] | 3.6789 | 3.6654 | | | |
| LCCP ^[19] | 3.9594 | 3.9243 | | | |
| CLCCP ^[12] | 4.0356 | 4.0024 | | | |
| CLBP-S/M/C ^[12] | 3.7248 | 3.7052 | | | |
| LTP ^[10] | 3.4352 | 3.4456 | | | |
| CS-LBP ^[9] | 3.6029 | 3.6145 | | | |

4 结语

本文提出了一种新的基于信息熵加权的二 维局部共现二值模式的静脉识别方法。利用等 价局部二值模式获取图像中的旋转不变编码, 以模式对的相对共现梯度方向作为索引,提取 局部区域内的模式对共现信息作为图像特征。 解决了现有算法无法提取高维特征的缺陷,弥 补了2DLBP算法无法对方向性信息进行提取的 缺点。使用SDUMLA手指静脉数据库和马来西 亚理工大学指静脉数据集(FV-USM)进行仿真实 验,结果证明本文所提出的方法相对于其余对 比算法具有较高的识别率,对于噪声干扰拥有 较强的鲁棒性。

参考文献:

- [1] VAN T H, LE H T. An efficient algorithm for fingerprint reference-point detection [C] //Proceedings of the 2009 IEEE-RIVF International Conference on Computing and Communication Technologies, Danang, Vietnam, 2009: 1-7.
- [2] KONG A W K, ZHANG D. Competitive coding scheme for palmprint verification [C] //Proceedings of the 17th International Conference on Pattern Recognition, Cambridge, UK, 2004: 520-523.
- [3] VAN H T, TAT P Q, LE T H. Palmprint verification using GridPCA for Gabor features [C] //Proceedings of the 2nd Symposium on Information and Communication Technology, Hanoi, Vietnam, 2011: 217-225.
- [4] SHARMA S, DUBEY S R, SINGH S K, et al. Identity verification using shape and geometry of human hands[J]. Expert Systems with Applications, 2015, 42 (2):821-832.
- [5] EL-TARHOUNI W, ABDO A, ELMEGREISI A. Feature fusion using the local binary pattern histogram fourier and the pyramid histogram of feature fusion using the local binary pattern oriented gradient in iris recognition [C] //Proceedings of the 2021 IEEE 1st International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering MI-STA, Tripoli, Libya, 2021:853-857.
- [6] KARANWAL S, DIWAKAR M. OD-LBP: orthogonal difference-local binary pattern for face recognition[J]. Digital Signal Processing, 2021, 110: 102948.
- [7] OJALA T, PIETIKAINEN M, MAENPAA T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7):971-987.

- [8] ZHANG Z, WANG M. Multi-feature fusion partitioned local binary pattern method for finger vein recognition [J]. Signal, Image and Video Processing, 2022,16(4):1091-1099.
- [9] GUO Z, ZHANG L, ZHANG D. A completed modeling of local binary pattern operator for texture classification [J]. IEEE Transactions on Image Processing, 2010, 19(6): 1657-1663.
- [10] HEIKKILÄ M, PIETIKÄINEN M, SCHMID C. Description of interest regions with local binary patterns[J]. Pattern Recognition, 2009, 42(3): 425-436.
- [11] TAN X, TRIGGS B. Enhanced local texture feature sets for face recognition under difficult lighting conditions [J]. IEEE Transactions on Image Processing, 2010, 19(6):1635-1650.
- [12] ZHANG B, GAO Y, ZHAO S, et al. Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor [J]. IEEE Transactions on Image Processing, 2009, 19 (2): 533-544.
- [13] LUO Y T, ZHAO L Y, ZHANG B, et al. Local line directional pattern for palmprint recognition [J]. Pattern Recognition, 2016, 50:26-44.
- [14] NOSAKA R, OHKAWA Y, FUKUI K. Feature extraction based on co-occurrence of adjacent local binary patterns [C] //Advances in Image and Video Technology: 5th Pacific Rim Symposium, PSIVT 2011, Gwangju, South Korea, Proceedings, Part II 5. Springer Berlin Heidelberg, Tripoli, Libya, 2012: 82-91.
- [15] QI X, XIAO R, LI C G, et al. Pairwise rotation invariant co-occurrence local binary pattern [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(11):2199-2213.
- [16] BI X, YUAN Y, XIAO B, et al. 2D-LCoLBP: a learning two-dimensional co-occurrence local binary pattern for image recognition [J]. IEEE Transactions on Image Processing, 2021, 30:7228-7240.
- [17] 李均,王志诚,吴雨轩,等. 熵概念的延拓:从热熵到 信息熵[J]. 大学物理,2020,39(10):29-33.
- [18] WITKIN A P. Scale-space filtering [M]. Morgan Kaufmann: Readings in Computer Vision, 1987: 329-332.
- [19] CHEN X, ZHOU Z H, ZHANG J S, et al. Local convex-and-concave pattern: an effective texture descriptor[J]. Information Sciences, 2016, 363: 120-139. (下转第61页)

DOI: 10.3969/j.issn.1007-1423.2024.08.003

基于改进YOLOv7-Tiny的高速公路入口两轮车辆闯入检测

王宏,田恬*

(西安石油大学计算机学院,西安 710065)

摘要:近年来,浙江、福建等省区相继出台相关地方性法规,禁止两轮车辆(摩托车、电动车等)通行高速公路。 针对高速公路入口工作人员难以实时检测到两轮车辆闯入的问题,提出一种改进YOLOv7-Tiny的两轮车辆闯入检测算 法。首先,从VOC2005 中提取摩托车图片并增补了带有入口背景的图片后形成新数据集;其次基于YOLOv7-tiny,引 入ECA 注意力机制,使模型更加聚焦训练摩托车相关目标特征。使用 ssFPN 网络,对小目标特征信息进行增强;采用 基于动态非单调机制的 WIoU损失函数,提高对于小物体检测的准确性;使用 Adam 优化器,提升回归过程的收敛速度 和准确性。改进后的算法, mAP、Precision、Recall 分别提高了 2.63、4.01、13.92 个百分点, F1提高 0.10,表明该方法 具有显著的有效性。

关键词:两轮车辆闯入检测; YOLOv7-tiny; ECA注意力机制; ssFPN; WIoU

0 引言

近年来,浙江、福建、四川等省区相继出 台相关地方性法规,禁止两轮车辆(摩托车、电 动车等)通行高速公路。当前,大部分高速公路 采用人工拦截或加长车道栏杆臂等方式进行闯 入检测^[1]。但上述方法都存在一定的问题。

现如今,目标检测技术发展迅猛,已广泛 应用于各行各业,例如物体识别、智能监控以 及交通安全等。在此背景下,如何高效地将该项 技术同高速公路入口两轮车辆闯入检测相结合, 以实现提升工作人员的工作效率和节约不必要的 人力成本等,变得尤为重要。近年来,研究人员 提出了许多将深度卷积神经网络应用于目标检测 场景的算法,总体而言可以概括为两大类:一类 是两阶段检测算法(Two-stage),其中最具代表性 的就是Faster-RCNN算法。张杰^[2]采用Resnet50 残差网络替换原网络,并引入MC-FPN的方式对 传统的Faster-RCNN算法进行改进,使模型对于 小目标信息的检测能力得到提升。另一类就是 单阶段检测算法(One-stage),其中最具代表性 的就是 YOLO^[3]系列算法。吴明杰等^[4]通过对 YOLOv5s 算法引入双层路由注意力机制, Focal-EIoU损失函数和结合注意力机制的动态目 标检测头,从而提高模型对小型目标的检测精 度。以上方法通过不同的改进,直接或间接使小 目标的检测效果得到了提高,但仍有一些不足。

本文首先对公共数据集进行补充,从VOC2005 数据集中提取摩托车图片并增补了带有高速公 路入口背景的图片后形成新数据集,使模型更 加聚焦于训练高速公路背景下的两轮车辆目标 特征,捕捉细节信息。然后对当前目标检测算 法中十分有效的YOLOv7-Tiny算法进行了适当 改进,引入ECA注意力机制^[5],使模型更加聚焦 训练摩托车相关目标特征;使用ssFPN网络,对 小目标特征信息进行增强;采用基于动态非单调 机制的WIoU损失函数^[6],提高对于小物体检测的 准确性;使用Adam优化器,提升回归过程的收 敛速度和准确性。综上,改进后的YOLOv7-Tiny 模型能够显著提升高速公路入口两轮车辆闯入 检测的精度,降低漏检风险。

收稿日期: 2023-11-07 修稿日期: 2023-11-29

作者简介:王宏(1968—),男,陕西西安人,副教授,研究方向为图形图像处理、大数据技术;*通信作者:田恬(1995—), 女,陕西西安人,硕士研究生,研究方向为目标检测,E-mail:897083594@qq.com

1 基于 YOLOv7-Tiny 算法的检测模型

YOLOv7系列目标检测算法^[7-8]是由YOLOv4^{9]} 作者于2022年7月推出的,它的性能十分强大, 主干部分和加强特征提取部分都使用了创新的多 分支堆叠结构进行特征提取,也就是原网络结构 中所提到的ELAN模块和ELAN+模块,使模型的 跳连接结构更加密集。在SPP结构中引入了CSP 结构,增加模型的感受野,并提高模型的准确 性。在训练时,每一个真实框可以由多个先验框 负责预测,加快了模型的训练效率。同时借鉴了 RepVGG的结构,在网络的特定部分引入Rep-Conv,在推理阶段将多个计算模块合并为一个, 提高模型的效率和性能。其中YOLOv7-Tiny 网 络结构如图1所示。



图 1 YOLOv7-Tiny 网络结构

2 改进 YOLOv7-Tiny 的网络模型

2.1 注意力机制(ECA)

注意力机制(efficient channel attention, ECA)^[10] 起源于研究人员对人类视觉的研究。将注意力 集中在目标关键区域从而获取目标的关键信息, 使网络获得更多的上下文信息。为了加强目标特 征信息的利用,本文引入了效率更高的通道注意 力模块 ECA, ECA 为 SE(squeeze and excitation) 模块^[11]的改进,相较于SE,更加高效且轻量。 SE模块的主要原理为通过学习每个通道的重要 性权重,在卷积时根据重要性权重加强重要的 特征通道,抑制相对不重要的特征通道。既可 以加强各通道之间的相关性,也能获取更丰富 的特征信息,提高了目标检测的准确度。SE模 块由压缩部分和激励部分组成,如图2(a)所示。

SE模块的压缩部分如图 2(a)中的 F_{sq} 所示, 其原理为将卷积后的特征图 $H \times W \times C$ 进行全 局平均池化处理,如图 2(b)中 Global Average Pooling 模块所示,得到一个 $1 \times 1 \times C$ 的通道评 价向量,其中每个通道均会获得对应的权重。 SE模块的激励部分由两个全连接(FC)层、一个 ReLU和一个 Sigmoid 函数组成。其连接顺序如 图 2(b)中间部分所示。首先经过第一个 FC 将通 道评价向量的维度减为 c/r,后接 ReLU 激活函 数,然后使用第二个 FC 将通道数量增加到原始 通道尺寸大小(C),最后采用 Sigmoid 进行权重 归一化,得到各个通道的评价权重值与 SE 模块 输入的特征图对应的输出,如图 2(a)中后半部 分或图 2(b)中后半部分所示。

式(1)和式(2)分别为ReLU函数表达式和 Sigmoid函数表达式。

$$f(x) = \begin{cases} 0, x < 0\\ x, x \ge 0 \end{cases}$$
(1)

$$f(x) = \frac{1}{1 + e^{(-x)}}$$
(2)



图 2 SE 网络结构图

然而实验研究表明,SE模块中激励部分的 通道降维操作会出现数据信息丢失和对权重学 习产生消极影响,不仅获取依赖关系效率低而 且不必要。为此,ECA采用了如图3中虚框部分 所示的自适应局部跨通道一维卷积操作来增加 局部跨通道信息交互以及跨通道信息的获取。 图中*K*(卷积核大小)的大小表示有*K*个邻居通道 参与该通道的注意力权重预测。分组卷积研究 表明,一位卷积核*K*大小应与通道维度*C*成正 比,由于通道数通常为2的整数倍,*K*与*C*的映 射可如式(3)所示。

$$C = \phi(K) = 2^{(\alpha K - \beta)} \tag{3}$$

上式进一步推导得:

$$K = \Phi(C) = \left\lceil \frac{\log_2 C + \beta}{\alpha} \right\rceil_{ODD}$$
(4)

由式(4)可知, *K*为当α和β的值给定时, 通道数*C*取对数后一维运算最小奇数向上取整 的结果。综上可知, ECA通过一维卷积和共享 参数的机制,使通道间信息的交互得到了加强。 本文改进了YOLOv7-tiny网络特征提取阶段的结 构,将ECA注意力模块嵌入到上采样和下采样 之后,此更改可进一步加强对经主干网络特征 提取之后的特征图质量,且可在降低参数量的 同时极大提升网络对于特征的利用。



图 3 ECA网络结构图

2.2 ssFPN(Scale Sequence Feature Pyramid Networks)

FPN^[12]网络结构是一种多尺度特征融合网络结构,主要解决物体检测中的多尺度问题, 网络结构如图4所示。自底向上是CNN的前向 提取特征的过程,横向的连接是融合深层的卷 积层特征和浅层卷积特征的过程。自顶向下的 特征融合过程就是将顶层的特征和底层的特征 相融合的过程,这样底层的特征得到了来自顶 层的特征的加强,然而由于顶层的特征经过1*1 降维,势必会造成信息损失。引入了一种新的 FPN尺度序列(*S*²)^[13]特征提取方法,以增强底层 的特征信息。



将 FPN结构视为尺度空间,在 FPN的横轴上 通过 3D 卷积提取尺度序列(S²)特征。网络结构如 图 5 所示。获取 P3、P4、P5 三个特征融合层,以 P3(小目标的特征层)作为基准,首先将 P4和 P5 Resize 到 P3 的大小,接着使用 Unsqueeze 函数在 P3、P4、P5 层均增加一个维度,并在此维度上 将 Resize 后的三个特征层 Concat 起来,构成一 个 General View 模块,对 General View 模块进行 3D 卷积、批归一化等操作来提取其特征,最后 将处理后的 General View 模块和 P3 进行 Concat, 利用 1*1 的卷积调整通道后送入后续检测任务。



图 5 ssFPN 网络结构

2.3 损失函数WIOU

YOLOv7-Tiny的损失函数包括置信度预测 损失、类别损失和定位损失三部分,本文对 YOLOv7-Tiny的前两种损失函数不做改变,依旧 采用原网络中的二元交叉熵损失函数,而对于 定位损失,YOLOv7-Tiny采用的定位损失函数 CloU Loss^[14]如式(5)所示。

$$CIoULoss = 1 - IoU(B, B^{gt}) + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v$$
(5)

式中: B和 Bst 分别代表预测框和真实框,

 $\frac{\rho^{2}(b, b^{st})}{c^{2}}$ 为惩罚项, $b = b^{st}$ 分别表示 $B \pi B^{st}$ 的中心 点, ρ^{2} 为欧氏距离, $c \Rightarrow B \pi B^{st}$ 的最小外界矩形的 对角线距离, α 是一个正平衡参数, 表示 gt = 5预测 框长宽比的一致性。对于参数 $\alpha \pi v$, $f: \alpha =$ $\frac{v}{(1 - IoU) + v}$, $v = \frac{4}{\pi^{2}} \left(\arctan \frac{w^{st}}{h^{st}} - \arctan \frac{w}{h} \right)^{2}$, 式中: $w = w^{st}$ 表示 B跟 G的宽度, $h = h^{st}$ 表示 B跟 G的高度。由上式可知, CIoU考虑了边界框 回归的中心点距离、纵横比和重叠面积, 使得 目标框的回归更加稳定。但其未考虑预测框和 真实框的长宽比线性问题, 当长宽比成线性时, 上式中相对比例惩罚项一项将会退化为0, 从而 降低模型的识别精准度。

本文采用WIoUv3来替代CIoU,以解决小目标重叠遮挡或模糊不清等问题。下文中WIoU统 一指WIoUv3,其表达式如式(6)和式(7)。

$$\beta = \frac{L_{\text{lou}}^*}{\overline{L_{\text{lou}}}} \in [0, +\infty)$$
(6)

$$L_{\rm WIoUv3} = r L_{\rm WIoUv1}, r = \frac{\beta}{\delta \alpha^{\beta^{-\delta}}}$$
(7)

2.4 Adam 优化器

Adam 是一种基于梯度下降算法的自适应学 习率优化器。它可以根据训练过程中每个参数 的历史梯度和更新情况来自适应地调整每个参 数的学习率,从而加速神经网络的训练。它的 核心思想是在每个时间步骤中计算移动平均梯 度和平均平方梯度,并使用它们来更新模型参 数^[15]。改进后的模型如图6所示。



图 6 改进后的 YOLOv7-Tiny 网络结构

3 实验及结果分析

3.1 数据集与实验环境

本文实验使用自建数据集,从VOC2005中 提取519张摩托车图片并增补了276张带有高速 路入口背景的图片形成新的数据集,自建数据 集共794张图片。按8:1:1的比例将其划分 为训练集、验证集及测试集。实验结果表明, 使用增补后的数据集,对于高速公路入口两轮 车辆的闯入检测结果明显优于原始数据集的检 测结果。

实验环境所用操作系统为Windows10, 内存为16 GB, CPU为Inteli7-12700H,采用 PyTorch2.0.1的深度学习框架。

3.2 评价指标

本 文 使 用 Precision、 Recall、 F1-score 和 mAP作为模型的评价指标。

$$Precision = \frac{TP}{TP + FP}$$
(8)

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

$$F1 = \frac{2TP}{2TP + FP + FN}$$

(10)

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{11}$$

Precision 衡量的是精度,表达式如式(8)。 TP指的是预测值和真实值同时都为正例的个数; FP指的是预测值为正例,实际值为负例的个数。 Precision 值越大,代表模型的误检越少。Recall 指的是召回率,表达式如式(9)。FN是预测值 为负例,实际值为正例的个数。Recall 值越大, 漏检越少。F1-score 指的是精确率和召回率的调 和平均数,表达式如式(10)。通过F1-score 的方 式来权衡 Precision 与 Recall,可以有效避免短板 效应。mAP是对所有类别的 AP 值求平均值,表 达式如式(11)。AP 可以反映每个类别预测的准 确率,mAP 反映整个模型的准确率。

3.3 模型训练

本文在相同硬件设施及参数情况下对改进 前后模型进行分析,以训练后的Loss曲线为代 表解析实验结果,在模型训练过程中,网络参 数设置见表1。

图片尺寸

 表 1
 网络参数设置

 参数
 数值

 迭代轮数
 100

 批量大小
 16

 最小学习率
 (1E-3)*0.01

 权重衰減
 0

Loss 值对比如图 7 所示,由图 7 可直接得出 改进后算法的 Loss 值比原始算法的 Loss 值更小, 数值下降了 0.013,改进后算法的 Loss 值稳定在 0.019 上下,说明改进后的算法模型鲁棒性更 好,预测效果更佳。

640*640



图 7 Loss 值对比如图

3.4 对比实验

为了进一步证明本文所提算法对于高速公路 入口两轮车辆闯入检测的有效性与优越性,在自 建数据集下,分别使用YOLOv4、YOLOv5-s、 YOLOv7-Tiny、主干网络是VGG16的Faster CRNN和本文算法进行实验,结果见表2。

| 表 | 2 | 与当前主流检测算法对比结果 |
|---|---|---------------|
|---|---|---------------|

| 检测算法 | mAP/% | F1 | Recall/% | Precision/% |
|-------------|-------|------|----------|-------------|
| YOLOv4 | 60.80 | 0.49 | 32.91 | 94.55 |
| YOLOv5-s | 87.33 | 0.83 | 75.95 | 92.31 |
| YOLOv7-Tiny | 86.23 | 0.80 | 71.52 | 90.40 |
| Faster-RCNN | 75.14 | 0.57 | 81.65 | 43.88 |
| Ours | 88.86 | 0.90 | 85.44 | 94.41 |

由对比结果可知,改进之后算法的F1值达

到0.90,相对于目前主流检测算法,分别高出 0.41、0.07、0.10和0.33。综合来看,本文算法 对比于其他现有检测算法具有明显的优势,在 各项指标上反应效果比较良好,改进方法的优 势得到证实。

3.5 消融实验

为验证本文提出的改进算法对于原网络模型的有效性,以F1作为实验评估指标,以原 YOLOv7-Tiny模型作为基准进行消融实验,以 此充分评测所做各项改进对算法的性能影响, 实验结果见表3。

表 3 消融实验对比

| ECA | ssfpn | WIoU | Adam | mAP/% | F1 | Recall/% | Precision/% |
|--------------|--------------|--------------|--------------|-------|------|----------|-------------|
| × | × | × | × | 86.23 | 0.80 | 71.52 | 90.40 |
| \checkmark | × | × | × | 87.39 | 0.84 | 76.58 | 93.08 |
| × | \checkmark | × | × | 85.80 | 0.82 | 72.78 | 92.74 |
| × | × | \checkmark | × | 88.56 | 0.84 | 75.95 | 93.02 |
| × | × | × | \checkmark | 85.79 | 0.85 | 79.11 | 91.24 |
| \checkmark | \checkmark | × | × | 84.72 | 0.82 | 73.42 | 92.80 |
| \checkmark | × | \checkmark | × | 88.02 | 0.81 | 70.89 | 94.12 |
| \checkmark | × | × | \checkmark | 85.47 | 0.87 | 84.18 | 90.48 |
| × | \checkmark | \checkmark | × | 86.66 | 0.82 | 72.15 | 95.80 |
| × | × | \checkmark | \checkmark | 87.29 | 0.86 | 82.91 | 90.34 |
| × | \checkmark | × | \checkmark | 83.46 | 0.85 | 79.75 | 90.65 |
| \checkmark | \checkmark | \checkmark | × | 89.22 | 0.83 | 73.42 | 95.08 |
| \checkmark | × | \checkmark | \checkmark | 86.35 | 0.88 | 84.81 | 90.54 |
| \checkmark | \checkmark | × | \checkmark | 83.29 | 0.84 | 77.85 | 91.11 |
| × | \checkmark | \checkmark | \checkmark | 84.56 | 0.86 | 82.28 | 90.91 |
| \checkmark | \checkmark | \checkmark | \checkmark | 88.86 | 0.90 | 85.44 | 94.41 |

注:"√"表示实验中采用该模块,"×"则为不使用 该模块。

根据上述的实验结果,我们可以很直观地 分析出,在进行单个改进上,算法在不同程度上 均具有提升,进一步验证了算法的可行性。在 2~3个组合改进中,各项实验数据在至少3项 指标上都得到了提升,指标至少提高了1个百分 点。整体来看,4处改进共同叠加得到的效果最 好,其在各项指标上相对于原算法的mAP提高了 2.63个百分点,Precision提高了4.01个百分点, Recall提高了13.92个百分点,F1值提高了0.10 个百分点,改进后的算法优于原始算法。

3.6 检测效果分析

为了能够更加直观地了解改进前后算法的检测 效果,使用自建数据集对改进前后YOLOv7-Tiny 模型进行检测,得出的检测效果如图8所示。其 中,左侧是使用原始YOLOv7-Tiny算法检测的 效果图,右侧是使用改进后的YOLOv7-Tiny算 法检测的效果图。



图 8 检测结果图像对比

在图 8(a)中, 左图在远距离目标检测中存 在误检的现象, 错误的将行人识别成了摩托车, 右图不存在误检的情况。在图 8(b)中, 左侧远 距离存在树木遮挡的情况, 未能成功检测出摩 托车, 右图则准确识别出待检测物体。在图 8 (c)中, 左图在交叉密集目标检测中存在漏检现 象, 两辆摩托车只成功检测一辆, 右图则全部 检出, 不存在漏检现象。在图 8(d)中, 左图在 远距离小目标检测中,更远距离的目标未能成 功检测出,右图则成功检测出远距离摩托车。 由对比图可以看出,改进后的算法在有遮挡、 密集交叉情况或者远距离小目标等情况下检测 结果皆优于原始算法。

4 结语

本文提出了一种改进 YOLOv7-Tiny 的高速 公路入口两轮车辆闯入检测算法,将深度学习 算法应用于高速公路入口两轮车辆闯入检测。 为了提高目标检测算法在此过程中的检测准确 度,降低漏检率,我们引入了ECA注意力模块, 捕捉细节信息, 使模型更加聚焦于训练摩托车 相关目标特征。使用ssFPN网络,以增强小目 标的特征信息。采用更为先进的定位损失函数 WIoU Loss 计算损失,不仅考虑了预测框和真实 框的重叠部分,而且考虑两者之间的区域,提 高模型训练过程中预测框的收敛速度及效率。 同时,使用了Adam 优化器,提升了回归过程的 收敛速度和准确性。最终实验表明,本文提出 的改进 YOLOv7-Tiny 的高速公路摩托车检测算 法相对于原算法mAP的数值提高了2.63个百分 点, Precision的数值提高了4.01百分点, Recall 的数值提高了13.92个百分点,F1的数值提高了 0.10个百分点,并且与其他主流检测算法对比 同样具有明显优势,表明该方法具有显著的有 效性。由于距离、天气以及光线等原因,本文 算法在雨天或夜间场景以及小目标检测条件下 的检测精度还有很大的提升空间,在后续的工 作中将对此进行进一步深入的研究。

参考文献:

- [1] 覃东明.基于视频识别+轮轴检测技术的高速公路 摩托车闯入预警系统[J].西部交通科技,2023(2): 175-177.
- [2] 张杰.基于改进Faster-RCNN的小目标检测[J].现 代计算机,2023,29(14):14-18.
- [3] REDMON J, DIVVALA K S, GIRSHICK B R, et al. You only look once: unified, real-time object detection [C] // Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2016:779-788.
- [4] 吴明杰,云利军,陈载清,等.改进YOLOv5s的无人 机视角下小目标检测算法[J/OL].计算机工程与应

用:1-12[2023-10-18]. http://kns.cnki.net/kcms/detail/11.2127.tp.20230920.1219.042.html.

- [5] 谢溥轩,崔金荣,赵敏.基于改进YOLOv5的电动 车头盔佩戴检测算法[J]. 计算机科学,2023,50
 (S1):420-425.
- [6] 许晓阳,高重阳.改进YOLOv7-Tiny的轻量级红外 车辆目标检测算法[J/OL].计算机工程与应用:
 1-13[2023-10-07]. http://kns.cnki.net/kcms/detail/
 11.2127.TP.20230811.1703.022.html.
- [7] CHUNMING W, HUANYU G, JIANHENG L, et al. Application of improved YOLOv7-based sugarcane stem node recognition algorithm in complex environments [J]. Frontiers in Plant Science, 2023, 14: 1230517-1230517.
- [8] 杨成佳,钱明.基于小目标检测的YOLO算法研究 综述[J].吉林工程技术师范学院学报,2023,39 (3):92-96.
- [9] BOCHKOVSKIY A, WANG C-Y, LIAO H-Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. arXiv:2004.10934,2020.

- [10] 祁宣豪,智敏.图像处理中注意力机制综述[J/OL].计算机科学与探索:1-20[2023-10-16].http://kns.cnki.net/kcms/detail/11.5602.TP.20230629.1447.002.html.
- [11] 袁增千,曹景胜,杨传江.基于改进YOLOv3的前 车检测算法的研究[J]. 仪器仪表与分析监测,2023 (1):37-43.
- [12] 郭宝鑫,谢晓尧,刘嵩.改进ResNet50和FPN的多 尺度目标检测算法研究[J/OL].贵州师范大学学报 (自然科学版):1-9[2023-10-07].http://kns.cnki. net/kcms/detail/52.5006.N.20230925.1017.006.html.
- [13] HYEJIN P, JIWOO K, BYUNGGYU K. ssFPN: scale sequence (S2) feature-based feature pyramid network for object detection[J]. Sensors(Basel, Switzerland), 2023, 23(9):4432-4432.
- [14] 刘雄彪,杨贤昭,陈洋,等.基于CloU改进边界框 损失函数的目标检测方法[J].液晶与显示,2023, 38(5):656-665.
- [15] 马进, 王超. 基于改进 YOLOv4-tiny 的印刷电路板 缺陷检测研究[J]. 电子测量技术, 2022, 45(23): 99-106.

Two wheeled vehicle intrusion detection at highway entrance based on improved YOLOv7-Tiny

Wang Hong, Tian Tian^{*}

(School of Computer Science, Xi'an Shiyou University, Xi'an 710065, China)

Abstract: In recent years, provinces such as Zhejiang and Fujian have successively introduced relevant local regulations prohibiting two wheeled vehicles (such as motorcycles, electric vehicles, etc.) from passing through highways. A modified YOLOv7-Tiny two wheeled vehicle intrusion detection algorithm is proposed to address the issue of real-time detection of two wheeled vehicle intrusion by highway entrance workers. Firstly, motorcycle images were extracted from VOC2005 and images with entrance backgrounds were added to form a new dataset. Secondly, based on YOLOv7-Tiny, an ECA attention mechanism was introduced to make the model more focused on training motorcycle related target features. The ssFPN network was used to enhance small target feature information, and a WIoU loss function based on dynamic non monotonic mechanism was used to improve the accuracy of small object detection. Finally, use the Adam optimizer to improve the convergence speed and accuracy of the regression process. The improved algorithm improves mAP, Precision, Recall by 2.63, 4.01 and 13.92 percentage point, respectively, and improves F1 by 0.10, indicating significant effectiveness of the method.

Keywords: two wheeled vehicle intrusion detection; YOLOv7-Tiny; ECA attention mechanism; ssFPN; WIoU

文章编号:1007-1423(2024)08-0024-07

DOI: 10.3969/j.issn.1007-1423.2024.08.004

基于FPGA的语义分割算法设计与实现

汤镇铭,陶青川*

(四川大学电子信息学院,成都 610065)

摘要:为解决边缘端水利检测场景中水面情况复杂且具有实时性要求的问题,对BiSeNet网络进行了改进,得到了一种基于FPGA边缘设备的轻量化的实时语义分割算法。该算法通过Vitis AI对模型进行量化、Vitis软件平台软硬件协同优化设计定制了深度学习处理单元DPU,实现了基于FPGA的语义分割算法的部署。在自制水利数据集上保持较低精度损失的同时实现了良好的性能,实验结果表明,改进后的网络模型在精度上损失3.7%的情况下,在ZCU104设备上实现了31.06 FPS的推理速度,相对于ARM设备获得了18.9倍的加速推理效果。所提出的方法能够满足低功耗需求下实时水利场景分割任务的要求。

关键词:深度学习;语义分割; FPGA; BiSeNet; DPU

0 引言

随着深度学习技术的快速发展,基于神 经网络的语义分割模型正逐步运用于边缘端 设备中^[1],如目前的人像分割、自动驾驶、 智慧城市等。相比于目标检测而言,语义分 割不仅能识别出目标,还可以标记出目标的 边界,更适合于水利场景对于目标标定的需 求。然而,目前较为热门的语义分割网络如 SegNet^[2]、DeepLab v3+^[3]、SegFormer^[4]等网络 模型不够轻量化,在边缘设备上直接部署和使 用神经网络时存在网络参数量过大、实时性不 强的问题。这些困难极大地限制了将分割网络

现阶段将语义分割网络部署在边缘设备的 研究中均难以满足实时性场景大的需求,如文 献[5]中将U-Net部署于ARM-cortex-A9设备中的 推理时间为16.635 s; 文献[6]中将Multi-Net部 署于英伟达GTX TITAN XP GPU设备中的推理 时间为0.116 s。同时使用上述设备进行部署时 有着功耗高、散热困难、不够便捷等问题,而 FPGA 相对上述设备则具有更低的功耗以及更高 的并行度,更适合于作为边缘端设备进行语义 分割网络的部署。 针对上述情况,本文提出了一种基于 ARM+FPGA结构的低功耗语义分割加速器的设 计方法,首先利用轻量化网络MobileNet v3^[7]作 为BiSeNet^[8]的上下文路径主干,减少网络的参 数量与运算量。其次,在PL端基于Xilinx DPU IP进行硬件架构的设计,通过PYNQ^[9]平台实现 加速算子调用。本文设计加速器在功耗仅为 3090Ti GPU设备12%的情况下实现了31.06 FPS 的分割速度,相比ARM CPU设备取得了18.9倍 的加速效果。

1 语义分割网络

1.1 算法原理

传统的语义分割网络通常由编码器和解码器组成,编码器负责将图像进行特征提取,将空间信息转化为高维特征,解码器则将特征图进行逐层上采样和特征融合,生成与输入相同大小的分割结果。通常采用卷积神经网络(CNN)作为编码器来提取特征,采用反卷积层和跳跃连接(Skip connection)作为解码器进行上采样和特征融合。

然而,编解码结构的分割模型通常有较大的计算和存储需求,而边缘设备的计算资源和

收稿日期: 2023-11-02 修稿日期: 2024-03-21

作者简介:汤镇铭(1998—),男,云南曲靖人,硕士,研究方向智能信息系统;*通信作者:陶青川(1972—),男,四川 南充人,博士,副教授,硕导,研究方向为模式识别与智能信息系统,E-mail:taoqingchuan@scu.edu.cn 存储空间往往有一定限制,无法满足较大参数 量模型的部署需求。为解决这一问题,本文选 用具有轻量高效结构的BiSeNet网络进行部署, 该网络采用了双向路径来保留空间信息并扩大 感受野,相比于 SegNet、DeepLab v3+等网络在 速度和计算量方面均有较大提升。为进一步提 高推理效率,通过对模型进行结构轻量化,将 主干替换为更轻量化的网络,使其更适合在边 缘设备上部署应用。



图 1 BiSeNet网络结构示意图

1.2 BiSeNet 算法改进

BiSeNet采用双边分割结构,包含空间路径 (Spatialpath)和上下文路径(Context path)部分, 分別用来解决空间信息缺失和感受野不足的问题,此外还设计了如图2所示的注意力优化^[10] 模块(attention refinement module, ARM)和特征 融合模块(feature fusion module, FFM)来优化和 融合两个路径的输出。





文献[11]中提到,BiSeNet的网络结构和优 化策略使其在多个语义分割数据集上取得了较 好的性能,并能在GPU硬件平台上实现实时推 理。但以ResNet^[12]和 Xception^[13]为主干,虽特 征提取能力强,但参数量较大,并不适用于边 缘端部署。本文采用的改进方法是,将上下文 路径替换为轻量化的MobileNet v3 网络,在保留 特征提取能力的同时减小了网络的整体参数量。 改进后的网络结构如图 3 所示。





改进后网络将 MobileNet v3 的 16 倍和 32 倍 下采样结果作为 ARM 模块的输入特征。之后将 输出进行拼接并与空间路径进行特征层的融合。 MobileNet v3采用 DW conv 和 PW conv 函数代替传 统卷积函数,保留特征信息提取能力的同时极 大减小参数量,同时引入了 SE^[14]模块。传统的 大小为K的卷积输入通道为M,输出通道为N, 计算得到大小为 $W \times H$ 特征图时,卷积核的尺 寸是 $K \times K \times M$,一共N个,每一个都要进行 $W \times H$ 次运算,相关参数量与计算量分别为

$$Params = K \times K \times M \times N \tag{1}$$

$$Madds = K \times K \times M \times N \times W \times H \tag{2}$$

若选择深度可分离卷积进行该操作,得到 大小为W×H的特征图时,深度卷积的卷积核尺 寸K×K×M,一共要做W×H次乘加运算;逐点 卷积的卷积核尺寸为1×1×M,共有N个,一共 要做W×H次乘加运算。故其所需参数和计算 量分别为

 Params = K×K×M+1×1×M×N
 (3)

 Madds = K×K×M×W×H+M×N×W×H
 (4)

 通过上述公式可知,传统卷积与深度可分

离卷积的计算量之比为

$$\frac{1}{N} + \frac{1}{K^2} \tag{5}$$

利用通道可分离卷积代替传统卷积,显著 地减少网络的参数与运算次数,使其更加适合 在FPGA和其他低功耗的、算力有限的边缘上进 行推理加速。

改进后的网络上下文结构特征提取模块见表1,其中K为卷积核大小,IN为输入通道数, EXP为倒残差结构的扩张率,OUT为输出通道数,S为步长。其中,Bneck结构如图4所示。

表 1 改进后网络上下文模块主干结构(Small版本)

| Name | Input | OP | K | IN | EXP | OUT | S | SE |
|---------|------------|--------|---|-----|-----|-----|---|----|
| | 224×224×3 | conv2d | 3 | 3 | _ | 16 | 2 | _ |
| Stage 1 | 112×112×16 | bneck | 3 | 16 | 16 | 16 | 2 | SE |
| | 56×56×16 | bneck | 3 | 16 | 72 | 24 | 2 | _ |
| St. 2 | 28×28×24 | bneck | 3 | 24 | 88 | 24 | 1 | _ |
| Stage 2 | 28×28×24 | bneck | 5 | 24 | 96 | 40 | 2 | _ |
| | 14×14×40 | bneck | 5 | 40 | 240 | 40 | 1 | SE |
| | 14×14×40 | bneck | 5 | 40 | 240 | 40 | 1 | SE |
| | 14×14×40 | bneck | 5 | 40 | 120 | 48 | 1 | SE |
| Stage 3 | 14×14×48 | bneck | 5 | 80 | 480 | 48 | 1 | SE |
| | 14×14×48 | bneck | 5 | 112 | 672 | 96 | 2 | SE |
| | 7×7×96 | bneck | 5 | 160 | 672 | 96 | 1 | SE |
| | 7×7×96 | bneck | 5 | 160 | 960 | 96 | 1 | SE |



图 4 Bneck 模块示意图

2 语义分割加速器实现

2.1 加速器总体结构设计

加速器整体结构如图 5 所示,加速器由 PS 和 PL 两部分构成,其中 PS 部分主要由 ARM CPU和PYNQ平台构成,负责语义分割模型的数

据读取以及前后处理事务;PL端则主要由DPU IP构成,通过对其进行配置,可在Vitis AI中将 神经网络模型转换为DPU指令。在启动后DPU 将从片外存储器上获得指令控制计算模块的资 源调度。DPU中的计算模块采用流水线设计, 其中含有多个处理单元PE,每个处理单元由乘 法器、加法器和累加器等组成。DPU在片上存储 器BRAM中存储模型数据、缓存输入输出和中间 数据来减少外部存储器的带宽。PS和PL端通信 主要由AXI总线负责,其中AXI-lite接口负责获 取指令,两个AXI-Full负责PS端数据的传输。



图5 加速器整体结构

2.2 DPU IP综合与设计

本实验采用的 DPU 架构为 DPUCZDX8G, 是专为Zynq UltraScale+ MPSoC嵌入式设备设计 的深度学习处理单元。为对 DPU IP 资源用量进 行了解评估,实验在配置前先在 Vivado 中按照 传统 FPGA 开发流程对 IP 进行了综合与实现。 DPU 配置以及连接参考 DPU-PYNO 配置中的 DPU Hardware Design 流程,选择适配于ZCU104 的B4096的架构,该架构的DPU在像素和输入 输出通道方面具有显著的并行性,单周期内的 运算次数可达4096次。同时资源消耗测试过程 中,首先将 IP 核添加到项目工程中,根据实际 需要配置参数后,即可与Zynq的PS系统连接。 DPU 中 DPU0_M_AXI_DATA0 和 DPU0_M_AXI_ DATA1 分别与 PS 端的 S AXI HPO FPD 和 S_AXI_HP1_FPD 相连进行数据传输,端口 S AXI HP2 FP用于进行指令读取。配置时钟和 复位完成后,生成整体Block design 如图6所示,



图 6 DPU Block design 结构示意图

并通过默认策略进行综合布局与实现。为实现 基于 PYNQ 平台的加速器,与上述流程稍有不 同的是,需要在 Vitis 平台中对 DPU IP进行硬件 设计的构建,主要分为环境配置和硬件配置两 个步骤,首先配置 Vitis 和 XRT 环境,之后进行 硬件环境编译,之后得到 PYNQ 平台运行所需 的 dpu.bit、dpu.hwh 以及 dpu.xclbin 文件,即可 得到加速器硬件平台。

综合后资源消耗见表2,Bram_18k利用率 表示用于存储神经网络权重和图像缓存的资源 使用率,DSP的使用则表示使用DSP进行乘加 运算的资源消耗,本实验中选择了高性能的 DPU配置,故上述两项均具有较高利用率。

表 2 加速器整体资源消耗

| Item | Bram_18k | LUT | FF | DSP |
|---------------|----------|--------|--------|-------|
| Total | 210 | 103380 | 198613 | 1100 |
| Available | 312 | 230400 | 460800 | 1728 |
| Utilization/% | 67.31 | 44.87 | 43.10 | 63.66 |

2.3 模型量化与部署

为将语义分割模型映射至FPGA设备中进行 加速,需要对模型权重进行量化处理^[15]。在利 用DPU进行推理加速的流程中,需要将模型利 用量化器从FP32模型转换为INT8甚至更低精度 的模型。在进行量化前,因DPU中算子的限制, 需要对网络结构进行简单更改:主要是将逐点 卷积的激活函数设置为LeakyReLU,深度卷积 的激活函数设置为ReLU;更改MobileNet v3中 的SE模块,添加一层卷积并减少avgpool层的通 道数。减小量化流程如图7所示,首先将输入的 浮点模型进行预处理,主要将无用的节点等信 息进行清除,然后将网络权重和偏置信息量化 到定点位宽。之后为提高模型精度,利用数据 集中的部分图片进行校准,之后将校准后模型 转换为可在DPU中部署的xmodel模型文件。



图 7 模型量化流程

3 实验结果与分析

3.1 软硬件环境

本实验模型训练的操作系统是 64 位 Ubuntu 22.04, CPU 型号为 Intel Core i7,显卡型号为 NVIDIA GeForce GTX 3090Ti 24 GB,深度学习 框架为 PyTorch。硬件环境基于 Xilinx ZCU104设 备开发,该设备配置 4 核 ARM Cortex-A53 处理 器,拥有 504 K 个系统逻辑单元、1728 个 DSP 单元、双核 Cortex-R5™实时处理器、Mali-400™ MP2 图形处理单元、支持 4KP60 的 H.264/H.265 视频编解码器以及 16 nm 的 FinFET+可编程逻 辑。加速器 IP 核基于 Xilinx DPU IP 构建,构建 前利用 Vivado 2022.2 进行资源的测试评估,之 后基于 Vitis Platform进行加速器的综合实现,使 用 PYNQ 对 IP 核进行加速效果测试与网络构建 实现。

3.2 模型训练数据集

3.2.1 数据集

本文算法主要用于对水利场景进行语义分 割,数据集主要通过网上爬取、成都九眼桥河道 拍摄和成都河长制水利局拍摄得到等方式共获取 2867 张水利监控场景图片,图片中包含多个地点 和天气条件下的水利场景。使用 Labelme 软件对 原始图像进行分类标注,生成对应标签文件, 并通过脚本将同一类的图像区域标记为同一颜 色,并将得到的样本按照9:1划分为训练集和 验证集,数据集及其标注情况如图8所示。



图 8 数据集及标注情况

3.2.2 评价指标

实验中使用 Flops 和 Params 来评估模型复杂 度和存储资源需求。Flops 反映了网络的复杂 性,较高的 Flops 值表示更复杂的结构。Params 表示模型的大小,即需要学习的权重和偏置数量。我们以FPS作为模型的推理速度标准, FPS代表每秒处理的图像帧数。为全面评估模型性能和适用性,本文在Intel i5 CPU和Nvidia 3090Ti GPU上测试了模型的推理速度。

实验采用平均交并比 MIoU (mean intersection over union)评价模型分割效果。采用 Flops、 Params和 Size 作为模型参数量的判断依据,同 时对网络的推理速度进行测试,评价指标为 FPS。在语义分割任务中, IoU 计算真实值和预 测值两个集合的交集和并集之比, MIoU 将每一 类 IoU 累加并取平均。具体如下所示:

$$MIoU = \frac{1}{n} \sum_{i=1}^{n} \frac{P_{ii}}{\sum_{j=1}^{n} P_{ij} + \sum_{j=1}^{n} P_{ji} - P_{ii}}$$
(6)

其中: n为分割的类别数, P_{ii}为预测为第 i 类且 与真实值相同的像素点, P_{ij}为预测为第i类但真 实值为第j类的像素点, P_j为预测为第j类但真 实值为第i类的像素点。

3.3 结果分析

3.3.1 网络结果对比

实验分别对表3中的网络在自制的水利数据 集上进行性能和效果的对比,具体性能对比结 果见表3。

根据表3结果可知,与FCN和Deeplabv3等 网络相比,BiSeNet结构的参数量更少,计算量也 有降低。当将主干替换为轻量级的MobileNetv3 时,参数进一步减少,使BiSeNet结构更适合于 在边缘设备上部署。在模型精度方面,经过自制 水利场景数据集的实验,BiSeNet网络的精度与 编解码结构的网络相差不大,即使在替换主干网 络的情况下,精度损失仅为3.7%,同时在量化 后,模型的精度也没有较大损失。综合考虑计算 资源和网络精度,使用MobileNetv3_Small作为 主干网络更适合于基于FPGA的语义分割任务。

3.3.2 硬件性能指标

使用定点量化后的 BiSeNet-MobileNet v3_Small模型在ZCU104设备中进行推理计算, 将其与其他平台的推理结果在推理速度、平均 功耗等方面进行了对比和分析,结果见表4。

| Model | Backbone | MIoU/% | Flops/G | Params/M | Size/MB | $\mathrm{FPS}_1/\!(\mathrm{f}\!\cdot\!\mathrm{s}^{-1})$ | $\mathrm{FPS}_2/\!(\mathbf{f}\boldsymbol{\cdot}\mathbf{s}^{-1})$ |
|---------------|-------------------|--------|---------|----------|---------|---|--|
| FCN-32s | Vgg16 | 73.8 | 80.50 | 15.30 | 61.2 | 2.210 | 62.044 |
| Deeplabv3 | Resnet50 | 75.8 | 171.05 | 39.76 | 167.6 | 0.870 | 26.441 |
| BiSeNet | resnet18 | 73.9 | 13.03 | 12.80 | 51.3 | 14.074 | 89.892 |
| BiSeNet | Mobilenetv3_Large | 71.5 | 1.80 | 1.45 | 6.1 | 32.234 | 92.074 |
| BiSeNet | Mobilenetv3_Small | 70.2 | 0.78 | 0.56 | 2.4 | 59.085 | 98.507 |
| BiSeNet(int8) | Mobilenetv3_Large | 71.3 | — | 1.45 | 1.5 | — | — |
| BiSeNet(int8) | Mobilenetv3_Small | 70.0 | — | 0.56 | 0.6 | — | — |

表 3 网络改进前后性能对比

注: FPS₁表示在 CPU上的结果, FPS₂表示在 GPU上的结果。

| 测试设备 | 测试平台 | Frequency/GHz | Time/ms | Power/W | |
|------------------|---------|---------------|---------|---------|--|
| CPU (arm) | PyTorch | 1.20 | 608.100 | 8.85 | |
| CPU (i5-11th) | PyTorch | 1.71 | 16.920 | 53.20 | |
| GPU (3090) | PyTorch | 3.20 | 10.152 | 60.40 | |
| ZCU104 | PYNQ | 0.30 | 32.200 | 7.63 | |

表 4 各平台推理结果对比

根据表4结果,本文加速器相较于功耗仅相差 1.22 W的ARM结构 CPU,取得了 18.9 倍的加速效果。同时,与 x86 CPU和 GPU 设备相比,在满足实时分割的前提下,网络推理所需功耗 仅为它们的 14% 和 12%。这表明,经过 Vitis AI 量化后的神经网络加速器具有低功耗和速度快的优势,更适合于在边缘端部署。

4 结语

本文对 BiSeNet 网络进行改进,得到一种轻 量化的实时语义分割算法。充分利用 FPGA 低功 耗特性,通过 Xilinx Vitis AI 对模型进行量化压 缩,利用 Vitis 软件平台对 DPU 进行配置和综合 实现,得到了基于 FPGA 的语义分割加速器。在 精度仅损失 3.7% 的条件下在 ZCU104 设备上实 现了 31.06 FPS 的推理速度。相比于传统 ARM 设 备,获得了 18.9 倍的加速效果,为水利场景的 实时分割任务提供了性能保障。该方法有助于 对低功耗需求下的实时分割网络进行部署应用, 也为后续在边缘计算和嵌入式系统中部署高性 能语义分割提供了有力支持。

参考文献:

- [1] LOPEZ-MONTIEL M, LOPEZ D A, MONTIEL O. JetSeg: efficient real-time semantic segmentation model for low-power GPU-embedded systems [EB/ OL]. arXiv:2305.11419,2023.
- [2] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12):2481-2495.
- [3] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C] // Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany. 2018:801-818.
- [4] XIE E, WANG W, YU Z, et al. SegFormer: simple and efficient design for semantic segmentation with transformers [C] //Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS 2021), 2021, 34:12077-12090.
- [5] KORDESTANI M, ZANJ A, ORCHARD M E, et al. A modular fault diagnosis and prognosis method for hydro-control valve system based on redundancy in multisensor data information [J]. IEEE Transactions on Reliability, 2018, 68(1): 330-341.
- [6] WU D,LIAO M W,ZHANG W T, et al. YOLOp: you only look once for panoptic driving perception
 [J]. Machine Intelligence Research, 2022, 19 (6): 550-562.
- [7] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3 [C] //Proceedings of the IEEE/ CVF International Conference on Computer Vision, Seoul, Korea, 2019:1314-1324.

- [8] YU C, WANG J, PENG C, et al. BiseNet: bilateral segmentation network for real-time semantic segmentation [C]//Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 2018:325-341.
- [9] WANG L, AO T, FU L, et al. Design of a YOLO model accelerator based on PYNQ architecture [C] //Proceedings of the 2022 International Conference on Machine Learning and Intelligent Systems Engineering(MLISE), Guangzhou, China, 2022:15-18.
- [10] CHEN L C, YANG Y, WANG J, et al. Attention to scale: scale-aware semantic image segmentation [C]
 //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA,2016:3640-3649.
- [11] TAVERA A, MASONE C, CAPUTO B. Reimagine BiSeNet for real-time domain adaptation in semantic segmentation[EB/OL]. arXiv:2110.11662,2021.

- [12] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016: 770-778.
- [13] CHOLLET F. Xception: deep learning with depthwise separable convolutions [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA, 2017: 1251-1258.
- [14] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018:7132-7141.
- [15] 陈辰,柴志雷,夏珺.基于Zynq7000 FPGA 异构平 台的 YOLOv2 加速器设计与实现[J]. 计算机科学 与探索,2019,13(10):1677-1693.

Design and implementation of FPGA-based semantic segmentation algorithm

Tang Zhenming, Tao Qingchuan*

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

Abstract: To address the intricacies and real-time imperatives inherent in water surface detection within edge water resources monitoring scenarios, enhancements are made to the BiSeNet network, yielding a streamlined real-time semantic segmentation algorithm tailored for FPGA edge devices. This algorithm first undergoes model quantization via Vitis AI, followed by collaborative optimization design on the Vitis software platform, leading to the customization of a Deep Learning Processing Unit(DPU) for facilitating the deployment of the semantic segmentation algorithm on FPGA. Evaluation on a proprietary water resources dataset demonstrates the algorithm's ability to maintain commendable performance while incurring only a nominal loss in accuracy. Experimental findings reveal that the refined network model attains an inference speed of 31.06 frames per second(FPS) on the ZCU104 device, with a marginal accuracy decrement of 3.7%. This translates to an inference acceleration of 18.9 times compared to ARM devices. In summary, the proposed methodology satisfies the real-time segmentation requisites of water resources monitoring tasks within low-power constraints.

Keywords: deep learning; semantic segmentation; FPGA; BiSeNet; DPU

DOI: 10.3969/j.issn.1007-1423.2024.08.005

基于改进 YOLOv5 的车辆检测方法研究

赵月爱,王 哲*

(太原师范学院计算机科学与技术学院,晋中 030619)

摘要:基于YOLOv5s对交通道路上前方车辆的检测进行了研究,为提高车辆检测精度,对原YOLOv5s网络模型进行了改进。首先,将Mish激活函数应用于YOLOv5s模型之中代替原有的ReLU函数,使用更平滑的Mish激活函数有效 避免梯度消失的问题。其次,采用BiFPN作为特征金字塔,增加了特征传递的信息通道,提升模型的感知能力和上下 文信息的关联能力。最后,引入EloU作为损失函数的一部分,准确地表示预测框和真实框之间的位置关系,间接地使 收敛速度提升。在D²-City数据集上的实验表明,改进后的YOLOv5s平均精度mAP₀₅为84.2%,比原始YOLOv5s算法提 升了2个百分点。

关键词: YOLOv5s; 车辆检测; Mish; BiFPN; EloU

0 引言

随着经济的快速发展,城市化进程不断加 快,2022年我国汽车保有量4.17亿辆,驾车出 行已经成为当今最主要的交通方式之一。但随 之而来的是日益突出的交通安全问题,交通事 故的不断发生,给人们带来巨大的生命威胁和 财产损失。如何有效减少交通事故是我国道路 交通安全方面的一个重大难题。

近年来,随着人工智能和计算机视觉技术 的迅速发展,汽车辅助驾驶系统在一定程度上 可减少交通事故的发生。谷歌"无人车之父" Sebastian Thrun认为现在虽然大多数都在利用雷 达系统做无人驾驶汽车,其实用摄像头才是最 好的方式,因为人开车是用眼睛看的,这样更 接近现实,同时成本也更低。借助摄像头实现 无人驾驶,首先要完成对前方视野内的车辆进 行目标检测。

目标检测是近二十多年来快速崛起的技术, 通过计算机对图像和视频进行分析处理,最终 实现识别定位和标注物体功能。目标检测技术 最初使用的是传统机器学习方法,这种方法的 步骤是先进行特征提取:其次,使用分类器如 向量机(SVM)、随机森林(Random Forest)、朴 素贝叶斯分类器(Naive Baves)等方法对每个滑 动窗口内的目标进行检测分类。如文学志等[1] 使用扩展类 Haar^[2]特征描述车辆边缘及结构特 征获得多个弱分类器;然后,通过 AdaBoost^[3] 算法得到的弱分类器组合成一个强分类器进行 识别车辆任务。但是该算法在处理复杂场景和 非刚性物体的检测时有一定的局限性。随着深 度学习方法的发展,目标检测主要分为两阶段 (Two-stage)和单阶段(One-stage)两种常见方法, 两阶段的典型算法有 SPP-Net^[4]、Fast R-CNN^[5]、 Faster R-CNN^[6]等,虽然在检测精度上有着不错 的表现,但是模型网络结构复杂、速度较慢。 单阶段的典型目标检测算法是 YOLO 系列, YOLO算法可以有效解决上述算法问题。张富凯 等^[7]提出 DF-YOLOv3 使用深度残差网络提取车 辆特征在残差网络之后添加多个卷积层将其分 成6个分支,形成多尺度预测网络的方法,但是, 该方法加深网络结构导致检测速度变化,不利于 实时检测。邝先验等^[8]在YOLOv5s算法模型网络 结构上加上了 CBAM 注意力机制, 使得在针对

收稿日期: 2023-10-28 修稿日期: 2023-11-27

作者简介:赵月爱(1974—),女,山西祁县人,教授,从事机器学习、智能信息处理和网络安全相关技术的研究;*通信 作者:王哲(1998—),男,山西交城人,硕士研究生,研究方向为深度学习、计算机视觉研究,E-mail:www3182567@163.com

复杂场景的车辆检测任务中精度提升4.6%,与 原网络相比精度均有所提高。张青春等^[9]把原 始YOLOv5模型的损失函数CloU替换为α-loU, 用DWConv(depthwise conv)替代传统卷积,最终 实验结果map@0.5达到89.44%,算法识别速度 较原始算法具有较大提升,但检测精度还有提 升空间。

本文采用 YOLOv5s 模型进行车辆检测研究 并进行改进,首先,用 Mish^[10]函数替换原始算 法中的 ReLU^[11]激活函数,有效避免了梯度消 失问题;其次,用 BiFPN (bidirectional feature pyramid network)^[12]加权双向特征金字塔网络代替 原先网络的 FPN+PAN结构,使网络具有更多层级 特征融合、动态权重调整和双向连接的优势;最 后,在输出端 (Prediction)使用 EIoU (Efficient Intersection over Union)^[13]作为损失函数,相比 原模型在精确度和平均精度上均有所提升。

1 YOLOv5网络模型

YOLOv5有YOLOv5x、YOLOv5l、YOLOv5m 和YOLOv5s四个版本,最小版本是YOLOv5s, 模型大小为27.3 MB,最大版本YOLOv5x的大小 是142.8 MB。虽然YOLOv5s模型较小,但其依旧 保留着非常高的检测精度和检测速度。YOLOv5 网络结构主要分为输入端、Backbone、Neck和 Prediction四个部分^[14]。

YOLOv5 在输入端使用了 Mosaic 数据增强, 通过随机选择四张图片进行裁剪、缩放和拼接 变换,通过要求模型从局部视图识别物体,提 升了对小目标^[15](长宽 0×0~32×32 的物体)的 检测。模型还加入了自适应锚框(auto learning bounding box anchors, ALBBA),通过学习输入 图像和目标分布动态生成的锚框来自动调整预 测锚框的合适大小和所在位置,运用自适应图 片缩放提升模型鲁棒性。

输入图像经过输入端的预处理后进入到模型的Backbone阶段进行特征提取,YOLOv5借鉴 CSPDarknet53^[16]网络结构,设计了两种带有残 差模块^[17]的CSP结构,其中一种CSP1_X结构与 SPP(spatial pyramid pooling)结构^[18]一起构成 Backbone层,另一个CSP2_X应用在Neck层中。 CSP1_X结构包含两个CSP层,输入特征图在每 个 CSP 层中分别进入两个通道进行不同处理, 最后将两个通道输出,结果拼接结构如图 1 所 示。CSP1_X结构在减少模型复杂度的同时,有 效提取图片的特征信息,其中 CBL 是由卷积模 块(COV)、归一化处理(batch normalization, BN) 和 Leaky ReLU 激活函数组合的模块。



图 1 CSP1_X结构示意图

特征融合层(Neck)是由路径聚合网络结构 (path aggregation network, PANet)和特征金字塔 结构(feature pyramid network, FPN)组成,是将 不同层级的特征图进行融合和整合,能更好地 捕捉不同尺度目标的信息,使得网络能够检测 出不同尺寸和比例的物体,实现高质量的目标 检测结果。

Prediction阶段对检测结果进行非极大值抑制(NMS)操作,确保每个目标保留一个最优的预测框,通过损失函数 GIoU(generalized intersection over union)^[19]提供预测框与真实框的位置评估,经过边界框裁剪、类别筛选、概率值筛选等后处理获得最终目标检测结果。

2 改进的 YOLOv5s 网络模型

2.1 Mish激活函数

YOLOv5s原网络模型中使用的激活函数是 ReLU激活函数,如图2(a)所示,这个激活函数 在许多深度学习任务中表现良好,可以有效解 决梯度消失的问题。但是ReLU函数在小于零的 区间输出恒为零,这就会导致神经元死亡现象。 Mish激活函数是一种非线性的激活函数,如图2 (b)所示,它可以解决上述出现的输出恒为零的 情况。并且ReLU函数在原点处不可导,而Mish 函数是一个连续可导的,这样模型在训练中更 容易收敛,将输入值映射到输入与输出之间不
存在比例关系结果的非线性变化中,提高模型 对目标的识别能力和解决梯度消失问题,如公 式(1)。



2.2 特征金字塔

随着网络层数的加深,会造成一定程度的 特征信息丢失。因此多尺度融合被广泛应用于 目标检测网络中,提高网络模型的检测性能。 例如,原YOLOv5s网络模型中的FPN通过进行 不断的下采样,拥有一些具有高语义内容的特 征层,然后使用PAN重新进行上采样,使得特 征层的长宽重新变大,用大的特征图去检测小 目标。但是FPN+PAN的网络结构复杂,缺少不 同层级之间的连续性和信息流动性。本文引入 了特征金字塔网络(BiFPN)进行动态尺度调整, 学习不同特征的重要性,对不同的输入特征有 区别地融合,在不增加太多成本的情况下可融 合更多特性,解决原模型网络复杂、缺乏信息 传输流动的问题。

BiFPN的核心是将特征金字塔分为顶部和 底部两部分,底部是一个自下而上的特征金字 塔,通过上采样操作将底层特征与上一层特征 融合,生成保留复杂细节信息的特征图。顶部 部分是一个自上而下的特征金字塔,使用下采 样操作将顶部特征与下一层特征融合,增强特 征的感受野,获得更有表征能力的特征图。为 了将不同层次之间的特征图进行拼接,在底部 和顶部的特征金字塔之间添加多路双向连接, 每个连接都由两个卷积层组成,一个用于水平 连接,一个用于竖直连接。网络结构如图3所 示,虚线框内表示的是一个BiFPN特征网络层。



图 3 BiFPN 结构

2.3 损失函数

在 YOLOv5 网络模型中,使用 GIoU 作为损 失函数的一部分,但是当真实框包含预测框的 位置关系时,导致损失无法体现出区别。本文 使用鲁棒性更强的 EIoU 作为损失函数。

IoU及其损失函数计算公式如式(2)、(3)所示,GIoU及其损失函数的计算如式(4)、(5)所示,EIoU的损失计算公式如式(6)所示。

$$loU = \frac{A \cap B}{A \sqcup B} \tag{2}$$

$$L_{IoU} = 1 - IoU \tag{3}$$

$$GIoU = IoU - \frac{\left|C - (A \cup B)\right|}{\left|C\right|} \tag{4}$$

$$L_{GloU} = 1 - GIoU \tag{5}$$

$$L_{EloU} = 1 - IoU - \frac{\rho^2(b, b^{gt})}{c^2} - \frac{\rho^2(w, w^{gt})}{c_w^2} - \frac{\rho^2(h, h^{gt})}{c_h^2}$$
(6)

其中: *A*表示预测框, *B*表示真实标注, *C*为预测框和真实框的外接矩形框的区域, *b*和*b*st分别

MCMIP

表示预测框和真实测量框的中心点, ρ 为计算欧 氏距离,c表示为预测框和真实框外接矩形的对 角线长度, c_w 和 c_h 是外接矩形的宽度和长度,w和h分别代表框的宽和长。当A和B完全不相交 时,无论真实框和预测框之间是怎样的位置关 系, $IoU都为0, L_{IoU}都为1。而 GIoU 通过加入C$ 作为判断,当C越大(两个框的距离越远)GIoU $越小,<math>L_{GoU}$ 损失越大,通过损失函数的值获取两 个框之间位置关系信息。

但是,当预测框和真实标注框为包含关系时,如图4所示。C的区域即为AUB,使用 GloU作为损失函数,图中两种情况仅仅通过计 算GloU作为损失,损失相同无法反映两个框之 间的位置信息。而本文使用EloU作为损失函数 的一部分,来作为真实标注框和预测框之间匹 配的衡量标准,损失值作为反向传播更新模型 的参数。EloU将重叠面积、中心点距离、长损 失和宽损失都作为影响因子,解决上述问题, 增加了模型的鲁棒性。



图 4 真实框完全包含预测框

3 实验结果与分析

3.1 实验数据集

D²-City数据集是滴滴出行提供的各种不同 交通场景下的数据集,数据大多为驾驶过程中 行车记录仪捕获的视频,其中包括经过抽取视 频获得的一张张图片。本实验关注点更多在于 目标框对车辆宽度的获取情况,目前,大部分 车辆检测整个汽车轮廓,检测框包含的面积较 广,且没有区分汽车的局部部分,而车辆可以 被作为一个拥有六个面的长方体。为了准确估 算车距,本文检测车辆前后保险杠所在整个面 作为车辆预测框,图像中的预测框宽度与实际 车宽的比值等于相机焦距与车距的比值,这样可以估算出与前方车辆的距离。在D²-City中选取了3892张图片,将图片中所有车辆分为小型车(轿车,SUV)和大型车(公交车,货车)两个类别,使用LabelImg工具对车辆数据集的汽车预测框进行标注。如图5所示。



图 5 检测车辆预测框示例

3.2 实验环境

本实验平台采用Windows11操作系统,显卡 为NVDIA GeForce RTX 3090(24 GB),处理器为 12th Gen Intel(R) Core(TM) i9-12900k(24CPUs)~ 3.2 GHz,语言为Python3.7,深度学习框架是 PyTorch, CUDA平台为11.6,深度神经网络库 为cuDNN8.3。

3.3 模型评价指标

本文采用精确率(Precision, *P*)和平均精度 均值(mean Average Precision, *mAP*)两个指标作 为衡量模型好坏的标准。P表示模型检测到的 正确的目标占数据集中标注的所有目标的占 比, P的计算公式如下:

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

式(7)中: TP(True Positive)为所有正确检测出 的车的数量, FP(False Positive)为所有没有识别 出的车的数量加上识别错误车的数量之和。召回 率R(Recall)计算公式如式(8)所示, AP(Average Precision)的计算公式如式(9)所示。

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

$$AP = \int_{0}^{1} PR dR \tag{9}$$

其中: FN为模型检测结果中没有检测出的车辆的个数。mAP是所有类别的AP值取平均数值,

用来反应模型在不同类别上的检测的准确性, 公式如下:

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

3.4 实验与结果分析

本文在 D²-City 数据集上,根据数据中不同 路况来识别车辆是小型车或是大型车。图6为原 YOLOv5s 和改进后 YOLOv5s 模型训练的准确率 P、召回率R、平均精度均值 mAP的对比图。在 改进后的 YOLOv5s 模型中,在训练 200 轮的情 况下, P、R和mAP都有不同程度的提高。



表1显示原YOLOv5s模型和改进YOLOv5s

模型的实验结果比较。改进 YOLOv5s 模型在召回率提升 2.7 个百分点的情况下,准确率和 mAP 分别提升了 2.4 和 2 个百分点。

表1 实验结果对比(%)

| 算法 | 准确率 | 召回率 | mAP |
|-------------|------|------|------|
| 原YOLOv5s模型 | 82.1 | 79.4 | 82.2 |
| 改进YOLOv5s模型 | 84.5 | 82.1 | 84.2 |

图 7 的左右两部分分别是原 YOLOv5s 模型 和改进 YOLOv5s 模型在 D²-City 数据集上的输出 结果,将改进前后进行对比。从图 7 可以看出, 在原 YOLOv5 模型检测中会出现遗漏目标的情 况。如图 7(a)左边图中有一个目标车辆未能识 别到,但在右图经过改进的 YOLOv5s 算法被识 别出来。并且检测到的目标置信度也会高于原 YOLOv5s 模型,如图 7(b)中可以看出检测的同 一目标概率数值,右图改进模型高于左图原模 型结果。

YOLOv5s算法 改进YOLOv5s算法



(b)

图 7 车辆检测结果

3.5 消融实验

为了更好地验证本文改进的YOLOv5s模型对 原YOLOv5s模型优化的有效性,采用YOLOv5s、 YOLOv5s+Mish、YOLOv5s+BiFPN、YOLOv5s+ EloU 和改进YOLOv5s模型五组网络模型在相同 的数据集上进行比较实验,结果见表2。

表 2 各模型训练结果对比(%)

| 模型 | 准确率 | 召回率 | mAP |
|---------------|------|------|------|
| 原 YOLOv5s | 82.1 | 79.4 | 82.2 |
| YOLOv5s+Mish | 80.1 | 79.5 | 82.6 |
| YOLOv5s+BiFPN | 80.4 | 80.4 | 84.1 |
| YOLOv5s+EIoU | 81.4 | 82.2 | 83.2 |
| 改进YOLOv5s | 84.5 | 82.1 | 84.2 |

从表2可以看出,YOLOv5s+Mish与原 YOLOv5s相比,在召回率提升0.1个百分点的 前提下,平均精度提高了0.4个百分点。YOLOv5s+ BiFPN和原YOLOv5s对比,在召回率提升1个百 分点的前提下,平均精度提升了1.9个百分点。 可以看出,在模型Neck部分使用的BiFPN根 据特征的重要性程度进行加权处理,可以提高检测 精度。YOLOv5s+EloU和原YOLOv5s对比,在召 回率提升了2.8个百分点的前提下,平均精度提 升1个百分点,可以看出,用EloU函数作为回 归损失函数可以准确描述真实框和预测框之间 的损失。而融合了上述三种方法的本文改进的 YOLOv5s,在精确度、召回率和平均精度方面 均高于原模型。

4 结语

本文旨在改进YOLOv5s目标检测模型对车辆进行检测。通过采用Mish激活函数、BiFPN特征金字塔网络和GIoU损失函数,在D²-City数据集上进行车辆检测实验,与原始YOLOv5s相比,在召回率提升2.7个百分点的情况下,准确率和mAP分别提升了2.4和2个百分点。今后将继续优化YOLOv5s网络,在保持精确度不变的情况下对模型进行剪枝轻量化,方便将模型部署到硬件Jeston nano上进行车辆测距。

参考文献:

- [1] 文学志,方巍,郑钰辉.一种基于类Haar特征和改进AdaBoost分类器的车辆识别算法[J].电子学报,2011,39(5):1121-1126.
- [2] LIENHAR R, MAYDT J. An extended set of Haar-like features for rapid object detection [C] // Proceedings of the IEEE International Conference on Image Processing, Rochester, NY, USA, 2002: 900-903.

- [3] DOOLEY D, MCGINLEY B, HUGHES C, et al. A blind-zone detection method using a rear-mounted fisheye camera with combination of vehicle detection methods [J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 17(1):264-278.
- [4] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9):1904-1916.
- [5] GIRSHICK R. Fast R-CNN [C] //Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015: 1440-1448.
- [6] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39 (6) : 1137-1149.
- [7] 张富凯,杨峰,李策.基于改进YOLOv3的快速车辆检测方法[J]. 计算机工程与应用,2019,55(2):
 12-20.
- [9] 张青春,吴峥,周玲,等.基于改进YOLOv5的车辆 与行人目标识别方法[J].中国测试,2023,49(7): 82-88.
- [10] MISRA D. Mish: a self regularized non-monotonic activation function [EB/OL]. arXiv: 1908.08681, 1908.
- [11] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks [J]. Journal of Machine Learning Research, 2011, 15: 315-323.
- [12] TAN M, PANG R, LE Q V. EfficientDet: scalable and efficient object detection[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020: 10778-10787.
- [13] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[EB/OL]. arXiv:2101.08158,2021.
- [14] 谢嘉飞,赵月爱.基于改进YOLOv5的电动摩托车 驾驶人头盔检测方法[J].太原师范学院学报(自然 科学版),2023,22(1):24-31.
- [15] KISANTAL M, WOJNA Z, MURAWSKI J, et al. Augmentation for small object detection [EB/OL]. arXiv:1902.07296,2019.
- [16] BOCHKOVSKIY A, WANG C Y, LIAO H Y M.

YOLOv4: optimal speed and accuracy of object detection[EB/OL]. arXiv:2004.10934,2020.

- [17] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] //Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016:770-778.
- [18] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual rec-

ognition [J]. IEEE Transactions on Pattern Analysis And Machine Intelligence, 2015, 37(9):1904-1916.

[19] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: a metric and a loss for bounding box regression [C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019:658-666.

Research on vehicle detection method based on improved YOLOv5

Zhao Yue'ai, Wang Zhe*

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

Abstract: Based on YOLOv5s, the detection of vehicles ahead on traffic roads is investigated, and the original YOLOv5s network model is improved in order to improve the vehicle detection accuracy. First, the Mish activation function is applied to the YO-LOv5s model instead of the original ReLU function, and the smoother Mish activation function is used to effectively avoid the problem of gradient disappearance. Secondly, BiFPN is used as the feature pyramid, which increases the information channel for feature transfer and improves the model's perceptual ability and the ability to correlate contextual information. Finally, EIoU is introduced as part of the loss function to accurately represent the positional relationship between the predicted and real frames, which indirectly makes the convergence speed increase. Experiments on the D²-City dataset show that the improved YOLOv5s has an average accuracy mAP@0.5 of 84.2%, which is a 2 percentage point improvement over the original YOLOv5s algorithm.

Keywords: YOLOv5s; vehicle inspection; Mish; BiFPN; EIoU

(上接第8页)

Real-time road disease detection model based on improved YOLOv4

Huang Yanguo*, Li Luo, Zeng Donghong, Wang Lining

(School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou 341000, China)

Abstract: A real-time detection model of road diseases based on improved YOLOv4 is proposed with the aim of addressing the issues of low detection accuracy, slow detection rate, and missed detection caused by many types and varying scales of road diseases. To prevent the poor detection effect brought on by the short Batch Size, the network model first optimizes the normalizing approach in the convolution block and uses group normalization rather than batch normalization. In order to increase detection speed and quantify the quantity of parameters the network model calculates, the convolution block is optimized simultaneously and replaced by a deep separable convolution block. Finality, the detection head use the adaptive non-maximum suppression algorithm to address the issue of false and missing identification of tiny targets resulting from the fixed non-maximum suppression threshold. The enhanced YOLOv4 algorithm has a detection accuracy mAP value of 86.64% and a detection speed of 37.90 frames per second in road disease detection, according to the testing data. The enhanced method, when compared to the original YOLOv4 algorithm, enhances detection speed by 10.60 frames per second, improves detection accuracy by 2.89 percent point, and successfully resolves the missed detection phenomena, all of which contribute to the increased practicability of road illness detection.

Keywords: road disease detection; YOLOv4; group normalization; deeply separable convolution; adaptive non-maximum suppression

文章编号:1007-1423(2024)08-0038-06

DOI: 10.3969/j.issn.1007-1423.2024.08.006

基于Transformer的PM_{2.5}浓度预测方法

叶耀,严华*

(四川大学电子信息学院,成都 610065)

摘要: 在深度学习领域中,通常采用循环神经网络(RNN)等方法对PM₂₅的浓度进行预测研究,但传统方法在捕捉 多站点数据之间的时空相关性方面存在一定困难。为了解决这一问题,基于Transformer 网络模型对PM₂₅浓度数据预测 进行研究。Transformer 采用多头自注意力机制,能够更好地捕捉PM₂₅浓度的时空依赖性。其通过模型中编码器提取特 征信息,通过模型中解码器处理特征中的依赖关系,输出未来时刻的PM₂₅浓度,在真实数据集上的实验表明,Transformer 网络模型具备更好的预测能力。

关键词: PM25; Transformer; 时空相关性; 深度学习

0 引言

随着社会的不断进步和工业化的不断发展, 空气污染对环境和气候带来了显著而深远的负 面影响,而 PM_{2.5}这种典型的空气污染物可能还 会导致呼吸系统疾病、脑血管疾病等健康问 题^[1-2]。因此,精确预测 PM_{2.5}未来时刻的浓度对 于预防和减轻空气污染带来的危害至关重要。

在以往的研究中,空气污染物的预测方法通 常分为非深度学习方法和深度学习方法两种。确 定性方法作为一种典型的非深度学习方法,其预 测原理主要是基于大气物理学和化学反应来模拟 空气污染物的生成、扩散、转化和清除过程^[34]。 典型的确定性方法包括CMAQ^[5]和WRF^[6]等。统 计方法作为另一种非深度学习方法,具备更良好 的数据驱动和统计建模技术表现^[78]。

但一般来讲,非深度学习方法对于非线性 条件下空气污染物的分析能力仍然存在一定的 局限性。于是,一些深度学习技术相继被研究 并运用到空气污染物预测领域。循环神经网络 (RNN)^[9]、长短时间记忆网络(LSTM)^[10]、门控 神经网络(GRU)^[11]等网络模型表现出了良好的 对于空气污染物特征的捕捉能力。但是这些模 型对于长距离的时间依赖性捕捉仍存在一定的 困难,同时针对多变量空气污染物数据时,预 测效果也表现出不足。

针对上述问题,本文将 Transformer 运用在 空气污染物预测领域,完成PM,5浓度预测相关 工作的研究。自2017年Vanilla Transformer模型 被提出后,这种完全基于自注意力机制的网络 模型在自然语言处理(NLP)和计算机视觉(CV) 等领域都得到了迅速的研究与运用,并取得了 不俗的表现^[12-13]。与此同时, Transformer模型也 为空气污染物浓度预测提供了一种新的思路和 方法,其采用的多头注意力机制能够有效处理 时间序列数据中的长期依赖关系,并且能够从 多个维度分析提取数据特征进而有效捕捉多变 量空气污染物之间的潜在相关性。与传统的 Transformer 模型不同,本文为了减少模型训练 与预测过程的时间复杂度和网络信号的空间复 杂度,采用了概率稀疏的注意力算法思想[14]。 在模型预测过程中只关注涉及主导查询并仅关 注最显著的头部部分,可以更有效利用好硬件 资源同时提高模型的训练效率。总体上看,本 文提出的预测方法较之前的深度学习方法在真 实数据集上针对PM,,浓度的预测表现有显著的 提升,验证了Transformer的有效性。

作者简介: 叶耀(1999—),男,四川富顺人,在读硕士研究生,研究方向为基于深度学习的大气污染物时空分析与浓度预测; *通信作者:严华(1971—),男,四川渠县人,博士,教授,研究方向为智能信息系统,E-mail;yanhua@scu.edu.cn

收稿日期: 2023-11-07 修稿日期: 2024-03-20

1 方法

1.1 整体框架概述

本文提出一种基于 Transformer 的 PM_{2.5} 预测 方法,方法的总体框架如图 1 所示。



PM2.5历史数据

图 1 基于 Transformer 预测模型总体框架图

具体方法流程为,首先收集了目标城市和 60个邻域城市的大量PM₂₅浓度数据,经过数据 预处理,与目标数据最相关的N个特征数据被 筛选出。这些数据表示为输入特征矩阵*C*,如下 所示:

 $\boldsymbol{C} = \left[C_1(t), C_2(t), \cdots, C_N(t) \right] =$

$$\begin{bmatrix} C_{1}(1) & C_{2}(1) & \cdots & C_{N}(1) \\ C_{1}(2) & C_{2}(2) & \cdots & C_{N}(2) \\ \vdots & \vdots & \ddots & \vdots \\ C_{1}(T-1) & C_{2}(T-1) & \cdots & C_{N}(T-1) \end{bmatrix}$$
(1)

其中:*T*表示时间维度的总长,*N*表示输入特征 个数。

随后对于每一个批次的训练,按照时间维度,输入特征被分为*X*,*Y*,*Z*这三部分数据, 其中*X*和*Y*表示两组连续的历史数据,用于 Transformer模型学习时间序列数据之间的依赖 关系和相关性。*Z*部分数据用一组空白数据填 充,用于表示未来时刻时间序列的占位符。经 过模型的训练学习后,这些空白数据为被替换 为模型预测的未来时刻特征值。

基于Transformer的预测模块主要包括:编

码器、解码器和输出层。其中,编码器主要由 多个多头概率稀疏的自注意力机制组成,用于 学习时间序列数据的多维度特征和相关性,输 出一个数据特征矩阵;解码器由多个注意力机 制组成,用于解析编码器输出的特征矩阵并进 一步学习时间依赖性,最终配合输出层得到预 测的未来时刻的时间序列数据。

最后,经过编码器和解码器的特征提取和 学习后,再经过一个全连接层的处理,就可以 得到未来Z个时间步长的PM_{2.5}浓度的未来时刻 数据。

1.2 数据预处理

在使用基于Transformer模型对PM₂₅进行浓 度预测的实验过程中,我们发现一些冗余的数 据会对实验结果产生一定程度的影响。一方面 是由于基于Transformer的预测模型是完全基于 自注意力机制对输入的每一个元素分配权重, 一些不相关的数据往往会导致模型的注意力被 分散,从而降低预测的精确性。另一方面,一 些冗余的输入数据会增加模型的计算时间和空 间复杂度,降低计算效率同时消耗一些硬件资 源。为了解决这个问题,首先需要对数据进行 相关性分析,去除一些影响实验效果的不相关 数据,最大信息系数(MIC)估计器被采用,作为 输入数据的预处理。

互信息(MI)是信息论中一种重要的度量工 具,用于量化两个变量之间所共享的有用信息 量^[15]。其计算公式如下:

$$I(x; y) = \int p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} dxdy$$
(2)

其中: p(x,y)是随机变量x,y之间的联合概率, p(x)和p(y)分别是x和y的质量概率。

由于在 MI 的计算过程中,联合概率的计算 过程十分繁琐。为了解决这个问题, MIC 采用 在二维空间中执行沿 x、y方向的方形格点中散点 落入的统计,替换了联合概率的直接计算。其 方程如下:

$$mic(x; y) = \max_{a^{*b} < B} \frac{I(x; y)}{\log_2 \min(a, b)}$$
(3)

其中: *a*和*b*分别表示*x*和*y*方向的方格数, *B*是一个经验值,其大小约为数据体积的0.6次方。

我们选择目标城市的 PM_{2.5}浓度作为基准, 所有邻域城市的浓度数据与其进行相关性评估。 只有相关性大于阈值β的特征数据才被选择为进 一步在网络中处理的输入。在第2节实验中,我 们通过一系列实验确定了预处理选用的阈值β。

1.3 特征嵌入模块

在经过相关性估计器后,与目标序列最相 关的特征数据被筛选出。但是由于自注意力机 制本身无法感知时间序列所携带的时间信息和 位置信息,我们无法直接将这部分特征数据输 入到编码器中提取数据之间的依赖。所以我们 需要在将特征数据输入到编码器之前完成嵌入 操作,这其中就包括特征嵌入VE、时间嵌入TE 和位置嵌入PE。

VE是一个输入维度为d_{in}、输出维度为d_{model}的卷积层(Conv1d)。TE将数据的时间轴嵌入为单独的组件,包括分钟、小时、天、星期和月份等信息。在PE中,pos表示当前位置的索引, i表示位置嵌入向量的维度索引,sin和cos函数 可以在不同值的维度中生成不同的位置嵌入向量,并将数据标准化到[-1,1]的范围内。嵌入 模块的计算公式如下:

 $VE = Conv1d(X_s^t)$

$$TE = minute + hour + day + weekday + month$$

$$\begin{cases}
PE(pos, 2i) = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \\
PE(pos, 2i + 1) = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right)
\end{cases}$$
(4)

最后各个嵌入操作得到的嵌入数据通过加 权作为下一个模块的输入。

1.4 编码器模块

自注意力机制处理时间序列预测问题的核 心思想在于建立序列中每个元素之间的依赖性 权重,允许模型对输入的不同部分分配不同的 注意力,从而为模型在处理时间序列的过程中 赋予了捕捉更长距离的依赖关系的能力。其计 算过程中,主要由查询(Query)、键(Key)和值 (Value)矩阵,通过计算查询与键之间的相似 度,借助Softmax函数获取注意力权重,将这些 权重应用于值,以得到输出。这种方式允许模型在不同位置对输入序列进行加权计算,有助于更好地理解全局关系,并得到更好的表示形式。其计算流程如图2所示。



图 2 自注意力机制算法示意图

在编码器模块中,为了提高自注意力机制 的训练效率,同时节约资源。我们采用了概率 稀疏的自注意力算法的思想。这种方法通过主 导查询,仅计算注意力中最重要的部分,而忽 略其余部分。计算公式见式(7):

$$A(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \operatorname{Softmax}\left(\frac{\overline{\boldsymbol{Q}} \boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_{\mathrm{model}}}}\right) \boldsymbol{V} \qquad (5)$$

其中: \overline{Q} 是从查询矩阵 Q 导出的概率系数矩阵, 用于计算当前位置元素与其他位置元素之间的 关联程度。K 是关键矩阵,提供其他位置序列 的信息。V 是值矩阵,存储值信息。 d_{model} 指的是 模型中的特征数量。所有分数都经过 Softmax 函 数进行归一化,以获得与查询元素对应的每个 关键元素的权重。

编码器接受了一个固定长度的历史数据。 通过学习数据,输出一个特征矩阵。编码器中 的三个多头ProbSparse自注意力模块可以从不同 角度捕捉多变量时间序列之间的依赖性,防止 信息的丢失。

1.5 解码器模块

与编码器不同,解码器的任务是基于历史 数据和编码后的特征矩阵输出目标序列的未来 数据。因此我们在解码器中引入了掩码操作, 以使网络在训练过程中无法感知未来数据的存 在。解码器的输入数据包括编码器输出的特征 矩阵,还有一组起到填充作用的空白数据和历 史嵌入数据组成的矩阵。

在整个训练过程中,编码器和解码器之间 的信息流是单向的,解码器通过不断生成时间 步的预测值,逐步构建出目标序列。在训练后, 填充数据被替换为预测的数据。最后再经过一 层全连接层,我们获得了包含预测信息的 PM₂₅ 浓度序列。

2 实验

2.1 数据集

为了验证模型的有效性,我们选择了北京 地区的PM_{2.5}浓度作为目标序列,同时华北平原 上主要区域的城市PM_{2.5}浓度被选为邻域城市的 数据,如天津、石家庄等^[16]。所有城市的PM_{2.5} 浓度数据都来源于中华人民共和国生态环境部 (MEE)。这些数据的时间跨度为2015年1月1日 到2018年12月31日期间。我们以每三小时的 时间间隔,收集了来自包括北京及周边地区在 内的40个城市的PM_{2.5}浓度数据,这一过程共得 到了每个城市包含11688条记录的数据集。

在进行实验时,我们将整个数据集分成了 训练、验证和测试集,比例为8:1:1。这种方 法有助于高效训练网络的最佳参数,并防止过 拟合。数据集概述见表1。

表 1 数据集的构成与相关统计信息

| Feature | Unit | Min | Max | Mean |
|-------------------|-------------------|--------|--------|--------|
| PM _{2.5} | ug/m ³ | 0 | 900 | 53.22 |
| Longitude | o | 28.44 | 40.96 | 34.47 |
| Latitude | ٥ | 111.66 | 121.62 | 118.18 |

2.2 对比实验

在这项工作中,我们在四个不同的深度学 习模型上基于相同的真实世界数据集完成了对 比试验,然后对这些模型与基于Transformer的 预测模型进行性能评估:

MLP, 多层感知器, 作为一种基本的前馈 神经网络, 具有强大的非线性拟合能力。

RNN,循环神经网络,其具有循环连接, 允许信息在不同时间步之间传递,从而可以捕 捉序列数据中的时间依赖关系。但使用时可能 存在梯度消失和梯度爆炸等训练问题。 LSTM,长短期记忆网络是常用的递归神经 网络,用于时间序列建模,可以解决长序列训 练的梯度消失和爆炸问题。

GNN-GRU,一种基于门控循环单元(GRU) 和图神经网络(GNN)的混合模型,可以有效提 取时空特征并预测污染物浓度。

2.3 评估指标

我们对所有的实验模型采用同样的实验指标进行性能评估,包括均方根误差(RMSE)、平均绝对误差(MAE)和R²。其中RMSE和MAE通过度量预测值与实际观测值之间的差异来评估预测效果,计算公式如下:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_i^* - y_i)^2}{N}}$$
(6)

$$MAE = \frac{\sum_{i=1}^{N} |y_i^* - y_i|}{N}$$
(7)

R²是一种常用的回归分析指标,用于评估 回归模型的拟合程度。R²的范围从0到1,其中 0表示模型没有任何预测能力,而1表示模型完 美拟合数据,其计算式如下:

$$R^{2} = 1 - \frac{\sum_{i=1}^{N} (y_{i}^{*} - \overline{y})^{2}}{\sum_{i=1}^{N} (y_{i} - \overline{y})^{2}}, \overline{y} = \frac{1}{N} \sum_{i=1}^{N} y_{i}$$
(8)

其中: y 是N个真实数据点的平均值。

2.4 实验设置

为了探究不同相关性的输入特征对目标序 列预测效果的影响,我们完成了一组附加实验 以确定数据预处理的最佳阈值,并选择MAE和 RMSE作为附加实验结构的评估指标。附加实验 的结果见表2。

表 2 不同阈值对应的实验结果

| Threshold value $\boldsymbol{\beta}$ | MAE/(ug·m ⁻³) | RMSE /(ug·m ⁻³) |
|--------------------------------------|---------------------------|-----------------------------|
| 0 | 10.23 | 14.42 |
| 2200 | 9.34 | 13.56 |
| 4500 | 8.19 | 13.05 |
| 6000 | 10.63 | 13.06 |

当阈值设定为0时,来自40个城市的PM₂₅ 浓度数据都添加到模型训练。从结果来看,这

些用于训练的数据中除了与目标城市强相关的 城市数据外,另外一些不太相关的冗余数据干 扰了基于Transformer预测模型的实验结果。当 阈值设定为2200时,一些冗余数据被过滤掉后 可以观察到模型的预测效果有所提升。当阈值 设定到6000时,可以观察到由于属于的特征太 少导致了模型预测性能的降低。因此,我们将 阈值设定到4500,尽可能多地选择与目标城市 PM₂₅数据最相关的城市数据,同时也滤除一些 干扰实验结果的冗余数据。被筛选出的与目标 城市最相关的邻域城市将作为后续模型的输入。

2.5 实验结果

实验和对比试验的预测结果见表3。

| Method | MAE/(ug·m ⁻³) | RMSE /(ug·m ⁻³) | \mathbb{R}^2 |
|-------------|---------------------------|-----------------------------|----------------|
| MLP | 11.13 | 15.46 | 0.89 |
| RNN | 11.07 | 14.85 | 0.91 |
| LSTM | 10.43 | 14.53 | 0.92 |
| GNN-GRU | 9.02 | 13.50 | 0.93 |
| Transformer | 8.14 | 12.91 | 0.95 |

表 3 不同方法的预测结果对比

可以看出, MLP由于受到其结构设计的限 制,无法捕捉PM25浓度数据之间的长期依赖关 系,加之学习和收敛的速度较差,在所有对比 实验中表现出了最差的预测效果。RNN作为一 种常用的处理时间序列的神经网络, 其预测效 果优于 MLP。在处理长序列时,由于LSTM 能够 解决RNN所存在的潜在梯度消失和梯度爆炸问 题,所以在时间序列预测能力上的表现比RNN 更好。GNN-GRU作为一种混合模型,能够通过 GNN对不同地区的PM2.5数据进行空间相关性感 知,也能基于GRU在时间维度学习数据的长期 依赖性,在预测效果上也呈现出不俗的效果。 但总的来看,基于Transformer的预测模型在所 有评估指标上都取得了最佳的预测效果,这得益 于其完全不同于传统神经网络的完全基于多头注 意力机制的编码器和解码器模型结构。这种结构 可以让基于Transformer的预测模型在感知时空序 列依赖性时不受元素位置和距离的影响,更好地 理解序列的全局和局部模式,提高了预测的精确 性。同时也说明了基于 Transformer 的模型在空 气污染物时序数据的预测任务中具有巨大潜力。

3 结语

本文提出了一种基于 Transformer 的 PM25 预 测方法,能够利用多地区PM2.5浓度数据对目标 序列未来时刻的PM,5浓度值进行预测。总的来 讲,首先我们对所有的数据进行预处理,以消 除与目标序列相关性不大且影响模型预测性能 的数据,然后利用基于概率稀疏的多头自注意 力机制的编码器框架对多地区的历史数据进行 特征提取和学习,最后经过基于多头注意力机 制的解码器和一个全连接层完成对目标地区 PM25浓度未来数据的预测。在真实的数据集上, 验证了基于 Transformer 预测模型的有效性,这 对政府机构和居民在处理大气环境污染问题上 具有一定的意义。在未来的工作中,我们将进 一步深入基于Transformer方法针对空气污染物 的相关研究,致力于在模型内部结构的深度和 混合模型的广度等方向探索出更多行之有效的 方法。

参考文献:

- PEACOCK J L, ANDERSON H R, BREMNER S
 A, et al. Outdoor air pollution and respiratory health in patients with COPD [J]. Thorax, 2011, 66 (7) : 591-596.
- [2] SANTIBAÑEZ D A, IBARRA S, MATUS P, et al. A five-year study of particulate matter (PM_{2.5}) and cerebrovascular diseases[J]. Environmental Pollution, 2013,181:1-6.
- [3] BAKLANOV A, MESTAYER P G, CLAPPIER A, et al. Towards improving the simulation of meteorological fields in urban areas through updated/advanced surface fluxes description [J]. Atmospheric Chemistry and Physics, 2008, 8(3): 523-543.
- [4] GENG G, ZHANG Q, MARTIN R V, et al. Estimating long-term PM_{2.5} concentrations in China using satellite-based aerosol optical depth and a chemical transport model [J]. Remote Sensing of Environment, 2015, 166:262-270.
- [5] CHOI Y J, FERNANDO H J S. Implementation of a windblown dust parameterization into MODELS-3/ CMAQ:application to episodic PM events in the US/ Mexico border[J]. Atmospheric Environment, 2008, 42(24):6039-6046.

- [6] GRELL G A, PECKHAM S E, SCHMITZ R, et al. Fully coupled "online" chemistry within the WRF model[J]. Atmospheric Environment, 2005, 39(37): 6957-6975.
- [7] LI R, MA T, XU Q, et al. Using MAIAC AOD to verify the PM_{2.5} spatial patterns of a land use regression model [J]. Environmental Pollution, 2018, 243: 501-509.
- [8] LEONG W C, KELANI R O, AHMAD Z. Prediction of air pollution index (API) using support vector machine(SVM)[J]. Journal of Environmental Chemical Engineering, 2020, 8(3): 103208.
- [9] LOY-BENITEZ J, VILELA P, LI Q, et al. Sequential prediction of quantitative health risk assessment for the fine particulate matter in an underground facility using deep recurrent neural networks[J]. Ecotoxicology and Environmental Safety, 2019, 169: 316-324.
- [10] NAVARES R, AZNARTE J L. Predicting air quality with deep learning LSTM: towards comprehensive models[J]. Ecological Informatics, 2020, 55:101019.
- [11] HUANG G, LI X, ZHANG B, et al. PM_{2.5} concentration forecasting at surface monitoring sites using

GRU neural network based on empirical mode decomposition [J]. Science of the Total Environment, 2021,768:144516.

- [12] HAN K, WANG Y, CHEN H, et al. A survey on vision transformer [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45 (1): 87-110.
- [13] HAN X, ZHANG Z, DING N, et al. Pre-trained models: past, present and future [J]. AI Open, 2021, 2:225-250.
- [14] ZHOU H, ZHANG S, PENG J, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting [C] // Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35 (12): 11106-11115.
- [15] ROSS B C. Mutual information between discrete and continuous data sets[J]. PloS One, 2014, 9(2): e87357.
- [16] KARIMIAN H, LI Q, LI C, et al. Spatio-temporal variation of wind influence on distribution of fine particulate matter and its precursor gases [J]. Atmospheric Pollution Research, 2019, 10(1):53-64.

A Transformer-based model for PM_{2.5} concentration prediction

Ye Yao, Yan Hua^{*}

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

Abstract: In the field of deep learning, the recurrent neural networks (RNNs) are often used to predict the concentration of $PM_{2.5}$. However, traditional methods encounter challenges in capturing the spatiotemporal correlations among multi-site data. To address this issue, research is conducted on predicting $PM_{2.5}$ concentration using a Transformer-based network model. The Transformer employs a multi-head self-attention mechanism that better captures the spatiotemporal dependencies of $PM_{2.5}$ concentration indices across various locations. The model's encoder extracts feature information, while the decoder handles dependencies in the input features to output future $PM_{2.5}$ concentrations. Experimental results on real datasets demonstrate that the Transformer network model possesses enhanced predictive capabilities.

Keywords: PM25; Transformer; spatiotemporal correlation; deep learning

文章编号:1007-1423(2024)08-0044-06

DOI: 10.3969/j.issn.1007-1423.2024.08.007

基于深度学习的对抗攻击发展研究

卢彦利1,石雪莹1,刘光晓1,柳雪飞1,文小慧2,李章敏3,蒋正锋1,4*

(1. 广西民族师范学院数理与电子信息工程学院,崇左 532200; 2. 江西财经大学统计与数据科学学院,南昌 330013;3. 威宁县岔河镇新发小学,毕节 553105; 4. 武汉大学计算机学院,武汉 430072)

摘要:随着深度学习在各领域的广泛应用,对抗攻击问题引起学术界与工业界的关注。首先概述了对抗攻击的背景,包括对抗攻击的定义、分类以及与传统的机器学习安全问题的区别。然后讨论了对抗样本生成及攻击策略,以及 白盒攻击和黑盒攻击等攻击手段。最后总结了对抗攻击的意义,并展望未来研究方向,期待通过研究和探索提高深度 学习模型的安全性和可靠性。

关键词:深度学习;对抗攻击;数据攻击;模型攻击;防御策略

0 引言

近年来,机器学习技术取得了重大进展, 并在许多领域取得了显著的应用成果。随着深 度学习的迅速发展,机器学习领域再次迎来新 的繁荣时期,极大地推动了人工智能的进步, 并在众多应用领域取得了卓越的成果,包括目 标检测[1-3]、计算机视觉[4-6]、语音识别[7-9]、自然 语言理解和恶意软件检测等。2014年, Szegedy 等^[10]在原始样本上施加了肉眼难以察觉的扰动, 引发了模型的错误输出,这些样本被称为对抗 样本。2015年, Goodfellow等^[11]利用一张著名的 熊猫图片,首次提出了对抗攻击的观念。自此 研究人员开始聚焦于探究和解析不同模型的潜 在漏洞,以及提升各类模型的稳健性。研究表 明,深度神经网络的图像识别系统容易受到对 抗样本的攻击,即通过在输入样本中加入轻微 扰动,诱导神经网络分类器对原始图像进行错 误预测,且被攻击的模型对错误预测的输出具 有高度的信任,这种对抗样本可应用于现实世

界。随着深度学习应用领域的不断深入和扩展, 深度学习暴露的安全问题受到了更为广泛的关 注。目前的研究主要集中在对抗攻击方法和对 抗防御研究两个方面,其中对抗攻击方法的研 究主要关注如何生成有效的对抗样本以及如何 提高对模型的攻击成功率,而对抗防御研究则 主要集中在如何检测基于对抗样本的攻击和提 高模型的鲁棒性。

对抗攻击是指在输入中加入人眼无法察觉 的微小扰动,使优秀的学习模型在加入扰动后 对输入产生错误预测。在传统机器学习中,手 工生成特征的机器学习系统是对抗攻击的主要 目标,如垃圾邮件过滤器、欺诈检测等,如通 过添加特殊修改字符攻击过滤器,使得垃圾邮 件无法被检测到。Roli等^[12]评估了几种主动防 御,以提高机器学习模型的安全性。2006年, Barreno等^[13]提供了一个回答"机器学习安全 吗?"问题的框架,对不同类型的攻击进行了分 类,讨论了针对这些攻击的各种防御措施,并给 出攻击者工作函数的下界的分析模型。2010年,

基金项目: 国家级大学生创新创业项目(202210604038)

收稿日期: 2023-10-07 修稿日期: 2024-03-20

作者简介: 卢彦利(2001—),女,广西灵山人,本科,研究方向人工智能; 石雪莹(2003—),女,广西北海人,本科,研究方 向为人工智能; 刘光晓(1996—),男,海南临高人,本科,研究方向为人工智能; 柳雪飞(1984—),男,湖北武汉人,硕士,讲师, 研究方向为大数据技术、智能优化算法; 文小慧(1999—),女,广西贵港人,硕士研究生,研究方向为数据挖掘; 李章敏 (1999—),女,贵州毕节人,本科,三级教师,研究方向为信息技术教育、数学教育、数据分析; *通信作者:蒋正锋(1979—),男, 广西桂林人,硕士,副教授,研究方向为最优化理论与应用、机器学习、量子计算,E-mail;jiangzhengfeng@gxnun.edu.cn

Barreno 等^[14]研究了关于机器学习安全性的问题,将针对机器学习系统的攻击分为三个维度, 分别为攻击是否会影响训练数据,对抗攻击是 假阳性还是假阴性,攻击的目标是特定的还是 广泛的。2013年 Biggio等^[15]提出了一种高效且 易于实施的基于梯度的对抗攻击方法,可对线 性分类器、支持向量机、神经网络等分类器进 行攻击,并系统性地评估分类算法的安全性。 通过增加攻击者对学习器信息的了解程度和修 改样本的限制大小,模拟不同的分类器攻击场 景,展现了广泛的应用前景和重要的实践价值。 本文介绍了对抗攻击的定义、分类以及与传统 机器学习安全问题的区别,然后探讨了对抗样 本的生成及攻击策略以及白盒攻击和黑盒攻击 等攻击手段,最后展望了未来的研究方向。

1 研究现状存在问题及分析

经典的对抗攻击方法主要从攻击时对模型 信息获取的程度和攻击模式两方面对当前的对 抗攻击方法进行分类和概述,并对研究现状进 行了总结和分析。尽管目前的对抗攻击技术已 经相对成熟,很多攻击方法都取得了不错的成 果,但对抗攻击技术仍然存在以下问题:

(1)黑盒攻击的查询次数高且成功率低,这 是目前存在的一个问题。在攻击方式上,白盒 攻击需要获取被攻击模型的所有结构和参数, 而在实际情况中,攻击者通常只能观察到模型 的输出结果,因此黑盒攻击更加真实。从攻击 算法的原理来看,基于梯度的攻击或分数式攻 击需要依赖模型的输出概率,但攻击者往往难 以获得这种信息,因此在现实情况下,这些攻 击方式往往是不切实际的,因此人们开始关注 基于决策的攻击方式。尽管在黑盒攻击方面已经 做了很多工作,但目前一些黑盒攻击的平均查询 次数仍然很高。通常情况下,最有效和最复杂的 攻击方式仍然需要数万或数十万次查询,并生成 不同数量的对抗样本,因此如何实现更高效的黑 盒攻击方法仍然是一个亟待解决的问题。

(2) 攻击的数据类型较为单一。从攻击数据 的角度来看,大多数攻击方法主要针对图像数 据进行攻击,这种攻击方式已经非常流行。然 而,也有一些研究开始尝试攻击其他类型的数 据,如流量数据、音频数据和文本数据,特别 是针对文本数据的攻击,已经引起了越来越多 NLP研究者的关注。由于文本数据的离散性,生 成对抗样本更具挑战性,同时也为相关研究提供 了更广阔的空间。攻击不同类型的数据有助于发 现对抗样本对不同深度学习模型的潜在威胁,从 而促使人们开发出更多更有效的防御方法。

(3) 攻击方法的执行时间较长。从攻击效率 的角度来看,传统的基于梯度的攻击方法通常 需要较长时间来执行,构造对抗样本耗时久, 并且通常需要大量的计算资源,因此实际应用 价值不高。然而,一些黑盒优化技术可以用来 优化对抗攻击的过程,如贝叶斯优化算法和进 化算法等。最新的优化算法可以与经典攻击方 法相结合,形成更高效、更实用的攻击方法。

(4) 对抗样本的可迁移性较差。从对抗样本 迁移的角度来看,大多数对抗样本的生成方法 只能针对少数特定的模型起作用,生成的对抗 样本无法应用到其他模型中。此外,攻击方法 的性能受到模型差异的较大影响。因此,在提 高攻击性能的同时,更应该关注对抗样本的可 迁移性,研究如何生成更具通用性的对抗样本 对于提高攻击的适应性和有效性至关重要。

2 深度学习对抗攻击

2.1 对抗样本

2014年,Szegedy 等^[10]第一次提出并总结了 利用对抗样本来欺骗深度神经网络的方法。在 神经网络的输入中添加一个微小的扰动,将产 生一个高置信度的错误输出,而人眼无法察觉 其中的差异,这种扰动被称为对抗扰动。我们 可以这样定义:输入图像为x,对抗样本为x', 距离度量函数为 $D(\bullet)$,图像分类网络为F(0,z), 分类网络输出的预测标签y(a) = F(0,z)。通常, 对抗样本满足以下条件:

 $D(x, x') < \varepsilon \coprod y(x) \neq y(x') \tag{1}$

对抗样本的一个典型例子^[10]如图1所示, 分类模型认为原始图像是熊猫,置信度为 57.7%。在原始图像上添加轻微扰动后,同一模 型将其归为长臂猿,置信度为99.3%,而人眼完 全无法区分原始图像和对抗图像的区别。



图 1 大熊猫加上微小扰动被误判为长臂猿

当前研究中,对抗样本存在的原因尚未有 公认的明确解释,可能归结于模型过拟合或正 则化不足。然而,通过在正则化模型中加入扰 动,Goodfellow等^[11]发现,这并不能显著提高对 抗样本的有效性,并认为产生对抗样本的根本 原因是网络的高维空间是非常线性的。这种观 点已经得到了更广泛的认同。由于对抗样本具 有不稳定性特征,这给攻击者构建稳健的对抗 样本带来了挑战,同时也给在实际场景中部署 对抗样本带来了困难。

2.2 对抗样本攻击方法

Goodfellow等^[11]认为,深度神经网络在高维 空间下的线性行为是导致存在对抗样本这一问 题的根本原因,提出了一种可以快速生成对抗 样本,并有效地计算对抗扰动的方法,如公式 (2)~(4)所示。

 $x^{\wedge'} = x + \varepsilon^* sgn(\nabla_x L(\theta, x, y))$ non-target (2)

$$x^{\Lambda'} = x - \varepsilon^* sgn(\nabla_x L(\theta, x, y)) target \qquad (3)$$

 $\eta = \varepsilon^* sgn\big(\nabla_x L(\theta, x, y)\big) \tag{4}$

其中: x为原始图像, η 为扰动, ε 表示控制扰 动大小的自定义参数, L为损失函数, sgn为符 号函数。

比较经典的FGSM(fast gradient sign method) 方法是通过让扰动方向与梯度方向一致,使损 失函数值变化最大,进而使分类器分类结果变 化最大,即通过将代价函数相对于输入的梯度 与输入进行合适的相加,得到的样本只要能成 功地使模型将其分类为与真实类别不同的对抗 类别,即可以称为对抗样本,如公式(5)和(6)。

$$\bar{X} = X + \eta \tag{5}$$

$$\eta = \varepsilon^* sgn(\nabla_x J(\theta, x, y)) \tag{6}$$

其中: X输入特征, y是标签, θ是模型变量参

- 数, η 是扰动步长, ε 是超参数,J是损失函数。
- FGSM优点为只需一步迭代就能生成对抗样本,且通过控制参数 *c* 可以生成任意范数距离的对抗样本,但也存在一些缺点,其扰动自身的抗干扰能力并不强,容易受到其他噪声的影响,而且输入与损失函数之间并非完全线性关系,表明该算法生成的对抗样本扰动并非是最优扰动。

上述攻击方法各有特点,需根据实际情况 选择合适的方法,也可以结合使用,以达到更 好的攻击效果。

2.3 针对模型的攻击

随着深度学习技术的广泛应用,其模型在 各领域扮演着日益重要的角色,然而攻击者采 用多种手段生成对抗样本,对模型的预测结果 产生干扰和误导。本节将分别阐述白盒攻击和 黑盒攻击这两种攻击深度学习模型的方法。

2.3.1 白盒攻击

白盒攻击^[16]是一种针对深度学习模型的具 有高针对性的攻击方式。攻击者拥有模型的完 全访问权限,可以深入了解模型的内部结构和 参数信息,从而利用这些信息制造出有效的对 抗样本,主要有:

(1)基于梯度的方法是常用的方法之一,攻 击者利用模型参数的梯度信息来生成对抗样本。 一种比较常见的攻击方法是FGSM(fast gradient sign attack),它通过计算模型损失函数对输入特 征的梯度,并乘以一个符号函数来生成对抗样本。 另一种比较常见的攻击方法是PGD(projected gradient descent),它通过多次迭代投影梯度下 降算法来生成对抗样本。这些攻击方法可以有 效地干扰或误导深度学习模型的预测结果,尤 其是对于图像分类和自然语言处理等任务。

(2) 基于 Jacobian 的方法通过最大化输入到 输出之间的 Jacobian 矩阵的范数生成对抗样本。 Jacobian 矩阵描述了输入微小变化对模型输出的 影响程度,因此,通过最大化 Jacobian 矩阵的范 数,可以找到那些微小的输入变化能够导致模 型做出严重错误预测的样本。该方法的优点是 直观且简单,但计算量较大。

(3)利用神经网络特定性质来创建对抗样

本,经由激活函数的性质或过滤器的响应来生成对抗样本。考虑到ReLU激活函数在负数区域 值为0的特性,可把输入朝负数方向调整,从而 形成对抗样本。其他攻击方法也可利用卷积神 经网络中过滤器响应的特性来生成对抗样本, 这些攻击方法常常能够避免常见的防御策略, 如数据增强和过滤器响应最小化等。

2.3.2 黑盒攻击

黑盒攻击^[16]是一种针对深度学习模型的间接 攻击方式,攻击者无法获取模型的内部结构和参 数信息,只能依赖模型的输出结果进行攻击。

(1)基于梯度上升的方法通过不断调整输入 样本的微小扰动,以逼近对抗样本。攻击者首 先生成一个接近原始输入的对抗样本,然后通 过多次迭代逐渐接近目标类别。这种方法需要 多次查询模型,因此比较耗费计算资源,但同 时也可能绕过一些防御措施,如模型沙箱和过 滤器响应最小化等。

(2) 基于遗传算法的方法将生物进化中的自 然选择和遗传机制引入到对抗样本的生成过程 中。攻击者首先生成一组初始对抗样本,然后 通过多次迭代的选择、交叉和变异操作来逐渐 接近目标类别。这种方法需要多次查询模型, 计算量较大,但在某些情况下,它可以生成更 具欺骗性的对抗样本。

(3) 基于强化学习的方法通过训练一个强化 学习代理(例如Q-learning)来生成对抗样本。攻 击者首先将原始输入作为状态输入到强化学习 代理中,并定义一个奖励函数来引导代理生成 对抗样本。然后,攻击者将生成的对抗样本作 为新的状态输入到代理中,并重复此过程直到 代理能够自动生成对抗样本为止。这种方法需 要大量的计算资源和时间,但是可以生成更具 欺骗性的对抗样本。

介绍了深度学习模型的两种攻击方法:白 盒攻击和黑盒攻击。这些攻击方法具有不同的 特点和挑战,需要根据特定的深度学习模型和 场景选择合适的方法。同时,由于这些攻击方 法的存在,我们需要更加重视深度学习模型的 安全性和可靠性,并加强对相关领域的研究和 探索。

3 研究对抗攻击的意义

研究深度学习对抗攻击的意义主要可以从 以下几个方面进行阐述:

(1)保护人工智能系统的安全性。深度学习 模型在各个领域都有广泛的应用,如自然语言 处理、计算机视觉、专家系统等。然而,对抗 攻击的研究表明,深度学习模型的准确性和鲁 棒性容易受到对抗样本的影响。攻击者可以通 过向良性数据中添加特定的扰动,生成对抗样 本,从而欺骗深度学习模型。因此,研究深度 学习对抗攻击有助于提高人工智能系统的安全 性,防止恶意攻击。

(2)推动神经网络模型的发展。对抗样本及相关生成算法的研究对于提升模型鲁棒性、推动神经网络模型在实际生产环境中的实施和部署具有重要积极意义。通过研究对抗攻击,可以不断优化和完善神经网络模型,提高其鲁棒性和泛化性能,从而推动神经网络模型的发展。

(3)促进人工智能技术的进步。对抗攻击研 究是攻击和防御的博弈。一方面,研究者不断 在已有的攻击算法基础上改进优化,提出更加 高明的对抗攻击算法,并尝试建立更加完善的 防御策略;另一方面,防御策略的进步反过来 推动对抗攻击技术的发展。这种攻防博弈促进 了人工智能技术的不断进步和发展。

(4)保障数据隐私和安全。对抗攻击可能会 威胁到数据隐私和安全。攻击者可以通过分析 深度学习模型的输出来推断输入数据的特征, 从而窃取敏感信息。研究深度学习对抗攻击有 助于保障数据隐私和安全,防止恶意攻击者利 用深度学习模型泄露敏感数据。

(5)提高模型的鲁棒性和泛化性能。对抗攻 击可以视为一种特殊的噪声,通过在输入数据中 添加微小的扰动,使得模型无法正确地处理这些 特殊情况,从而影响模型的鲁棒性和泛化性能。 通过对抗攻击的研究,可以更好地了解模型的弱 点,从而提高模型的鲁棒性和泛化性能。

(6)促进深度学习在安全领域的应用。随着 互联网和物联网的普及,网络安全问题越来越 受到人们的关注。深度学习在安全领域具有广 泛的应用前景,例如网络入侵检测、恶意软件 分析等。研究深度学习对抗攻击有助于推动深 度学习在安全领域的应用,提高网络安全防护 水平。

(7) 探索人工智能的边界。深度学习对抗攻 击的研究涉及到人工智能领域的多个方面,如 机器学习、神经网络、自然语言处理等。通过 对抗攻击的研究,可以不断拓展人工智能的边 界,探索其新的应用和发展方向。

研究深度学习对抗攻击对于提高人工智能 系统的安全性、推动神经网络模型的发展、促 进人工智能技术的进步、保障数据隐私和安全、 提高模型的鲁棒性和泛化性能、促进深度学习 在安全领域的应用以及探索人工智能的边界都 具有重要的意义。随着人工智能技术的不断发 展和应用,深度学习对抗攻击的研究将变得更 加重要和紧迫。

4 结语

近两年来,针对深度学习算法的对抗攻击 和防御技术迅速发展,然而,一些理论问题, 如对抗样本的起源和一般鲁棒边界的存在性还 没有找到答案,需要进一步深度研究。随着互 联网技术的不断发展以及人工智能理论研究、 相关方法的推广应用,有关对抗样本生成的方 法将会层出不穷,对抗性攻击技术的研究不仅 可以促进深度学习模型可解释性的发展,进一 步完善对抗性攻击防御技术,而且还可以促进 相关领域的研究。不仅如此,在实际安全应用 中,还没有一套有效且通用的对抗防御技术框 架与方法,目前的对抗性训练防御技术,在实 际部署中计算成本仍然太高。许多启发式防御 仍然缺乏针对自适应白盒攻击者的进一步验证。 总之,要实现有效防御目标,不仅需要在深度 学习算法安全性理论方面得到突破,还需要从 系统框架、安全测试、环境适配等多个方面与 安全技术相结合,才能推动深度学习对抗性安 全的跨越式发展。

关于对抗攻击的研究已经持续了很长时间, 但对其深层原理仍然没有普遍可信的解释。深 度学习的不可解释性为对抗攻击的理论研究带 来了很大的困难。因此,在未来的研究工作中, 对深度神经网络的内部运行机制的研究将直接 推动对抗攻击和防御方法的进步,此外,由于 深度学习的广泛应用,有必要从更多的角度了 解对抗样本,有针对性地研究不同的任务背景 下的攻击手段和防御手段。

参考文献:

- [1] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] //Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington D C, USA: IEEE Press, 2014:580-587.
- [2] GIRSHICK R. Fast R-CNN [C] //Proceedings of the IEEE International Conference on Computer Vision. Washington D C, USA: IEEE Press, 2015: 1440-1448.
- [3] REN S,HE K,GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Washington D C, USA: IEEE Press, 2015: 91-99.
- [4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6):84-90.
- [5] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C] //Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington D C, USA: IEEE Press, 2016:770-778.
- [6] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington D C, USA:IEEE Press, 2017:4700-4708.
- [7] GRAVES A, MOHAMED A, HINTON G. Speech recognition with deep recurrent neural networks
 [C] // Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing.
 Washington D C, USA: IEEE Press, 2013:6645-6649.
- [8] GRAVES A, JAITLY N. Towards end-to-end speech recognition with recurrent neural networks [C]//Proceedings of the International Conference on Machine Learning. Washington D C, USA: IEEE Press, 2014: 1764-1772.
- [9] ZHANG Y, PEZESHKI M, BRAKEL P, et al. To-

wards end-to-end speech recognition with deep convolutional neural networks [EB/OL]. arXiv: 1701. 02720,2017.

- [10] SZEGEDY C, ZAREMBA W, SUTSKEVER I, et al. Instriguing Properties of Neural Networks [EB/OL]. arXiv:1312.6199,2014.
- [11] GOODFELLOW I J, SHLENS J, SZEGEDY C. Explainingand harnessing adversarial examples [EB/OL]. arXiv:1412.6572,2015.
- [12] ROLI F, BIGGIO B, FUMERA G. Pattern recognition systems under attack [C] //Proceedings of the Iberoamerican Congress on Pattern Recognition. Springer, Berlin, Heidelberg, 2013:1-8.
- [13] BARRENO M, NELSON B, SEARS R, et al. Can

machine learning be secure?[C]//Proceedings of the 2006 ACM Symposium on Information, Computer and Communications Security, 2006:16-25.

- [14] BARRENO M, NELSON B, JOSEPH A D, et al. The security of machine learning[J]. Machine Learning,2010,81(2):121-148.
- BIGGIO B, CORONA I, MAIORCA D, et al. Evasion attacks against machine learning at test time [C]
 //Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Berlin, Heidelberg, 2013: 387-402.
- [16] 潘文雯,王新宇,宋明黎,等.对抗样本生成技术综述[J].软件学报,2020,31(1):67-81.

Research on the development of adversarial attacks based on deep learning

Lu Yanli¹, Shi Xueying¹, Liu Guangxiao¹, Liu Xuefei¹, Wen Xiaohui², Li Zhangmin³, Jiang Zhengfeng^{1, 4*}

(1. College of Mathematics, Physics and Electronic Information Engineering, Guangxi Minzu Normal University, Chongzuo 532200, China;

2. School of Statistics and Data Science, Jiangxi University of Finance and Economics, Nanchang 330013, China;

3. Xinfa Primary School, Chahe Town, Weining County, Bijie 553105, China;

4. School of Computer Science, Wuhan University, Wuhan 430072, China)

Abstract: With the widespread application of deep learning in various fields, the issue of adversarial attacks has attracted attention from both academia and industry. Firstly, the background of adversarial attacks is outlined, including the definition, classification, and differences from traditional machine learning security issues. Then we discussed adversarial sample generation and attack strategies, as well as attack methods such as white box and black box attacks. Finally, the significance of adversarial attacks was summarized, and future research directions were looked forward to improving the security and reliability of deep learning models through research and exploration.

Keyword: deep learning; adversarial attacks; data attacks; model attacks; defense strategies

文章编号:1007-1423(2024)08-0050-06

DOI: 10.3969/j.issn.1007-1423.2024.08.008

一种用于医学图像分割的混合卷积网络

张眉芳*,李其铿,谢隆腾

(福建卫生职业技术学院公共卫生与健康管理学院,福州 350101)

摘要:从医学图像中鲁棒性地分割器官是医学图像分析用于疾病诊断的关键技术之一。U-Net是一种用于医学图像 分割的鲁棒结构。然而,U-Net采用连续的下采样编码器来捕捉多尺度特征,高层语义特征恢复不足,从而导致上下 文信息丢失,无法充分恢复待分割器官特征。对编码器难以捕捉多尺度特征、高层语义特征恢复不足导致上下文信息 丢失进行研究,提出了一种新的混合卷积网络来捕捉更多的上下文信息和高级语义特征。混合卷积网络的主要思想是 利用提出的混合卷积连接模块从特征编码器提取更多的上下文信息和高级语义特征。多尺度特征提取模块用于连接编 码器和解码器子网络,以获得更丰富的多尺度特征图。将提出的方法与最先进的方法在 CHASEDB 数据集和 FRSA 数据 集上进行了比较。实验结果表明,提出方法的分割效果优于其他分割方法。

关键词: 医学图像分割; 混合卷积; 多尺度特征; 上下文信息

0 引言

从医学图像中进行稳健的器官分割对医学图 像分析至关重要。其关键任务是分割医学图像中 感兴趣的对象(如病变或器官)。然而,由于医学 图像信噪比低、质量差,医学图像分割仍是一项 非常棘手的任务。此外,感兴趣的器官往往被埋 藏在肌肉、血管、颅骨等复杂的背景中。

在过去几十年中,研究者提出了许多传统的 医学图像分割方法。Smeets等^[1]提出用于肝脏肿 瘤分割的组合方法,其包括螺旋扫描方法和统计 像素分类方法的半自动水平集方法。Lam等^[2]结 合了线形凹度测量和局部归一化凹度测量,提 出了多腔建模血管分割方法。Azzopardi等^[3]对 视网膜血管图像分割提出了可训练的COSFIRE 过滤器。传统监督学习图像分割算法的缺点是 依赖于精确而复杂的特征提取。这些特征的建 模和选择不仅需要专业知识,而且很难获得。 由于特征模型的质量直接影响到图像分割和分 类的效果,从而影响算法的通用性和准确性。

随着机器学习在图像和视频分析领域的发 展,使用卷积神经网络(CNN)特征自动学习已 成为医学图像分割的一种可行方法。与传统的 像素分类方法通常利用手工选择的图像特征不 同, CNN 方法可以学习图像特征并解决手工选 择问题。U-Net^[4]是著名的医学图像分割网络结 构之一, 在神经元结构分割和细胞分割方面表 现出色。U-Net是一种编码器-解码器模型,通 过跳跃连接将不同阶段的特征图连接。通过使 用编码器,可以逐步降低特征图的空间维度, 获得更多高级语义特征。解码器用于查找医学 图像中的细节信息和空间维度信息。Wang等^[5] 建立了一个高效的双编码器 U-Net(DEU-Net),以 提高视网膜血管的像素到像素分割性能。Wu等^[6] 提出了一种 Vessel-Net, 它结合了感知模型和残 差模型进行医学图像分割。Xu等^[7]将多尺度和 全局信息聚合在一起,开发了一种用于分割医

基金项目: 福建省科技计划引导项目(2021H01010037);福建卫生职业技术学院2022年卫生信息管理专业教学创新团队项目(JG2022105)

作者简介:*通信作者:张眉芳(1985—),女,福建建瓯人,硕士,讲师,研究方向医学信息处理、卫生信息管理, E-mail:mfzhang2017@163.com; 李其铿(1973—),男,福建闽侯人,硕士,教授,研究方向为智慧医院信息系统、卫生信息管理; 谢隆腾(1993—),男,福建连江人,硕士,助教,研究方向为文本信息分析、卫生信息管理

收稿日期: 2023-10-05 修稿日期: 2023-12-14

学图像的新方法。在该方法中,全局信息由一 个新的自适应的用于捕捉全局上下文融合模块 组成。

这些基于 CNN 的方法取得了良好的效果。 然而,U-Net 的连续下采样编码器及其变种的网 络在上采样过程中会导致上下文信息的丢失和 高级语义特征的恢复不足,不利于医学图像分 割与处理。Khan 等^[8]开发了一种深度混合网络 (DH-Net),它将 DenseNet与 U-Net 结合在一起, 对卫星图像中的土地覆盖进行分类。Khan 等^[9] 提出了一种编码器-解码器网络(EE-Net),用于 检测卫星图像中的建筑物足迹。这两种方法在 遥感图像分割方面取得了良好的效果,但在医 学影像分割方面的效果还需要进一步验证。此 外,Zhang等^[10]提出了多分支混合注意力特征解 码器来恢复高级语义特征。Sun等^[11]提出了U型 上下文残差网络,从编码器中恢复全局和空间 注意力信息。

针对U-Net及其改进网络由于连续下采样编 码器导致上下文信息的丢失和高级语义特征恢 复不足的问题,提出了一种新的混合卷积网络 用于医学图像分割。提出的方法主要包括一个 带有预训练ResNet 的编码器模块、特征解码器 模块、混合卷积连接模块和多尺度特征提取模 块。由于解码器中的连续卷积和池化操作, U-Net缺乏整合多尺度上下文的能力。使用混合 卷积连接模块连接不同阶段的特征图,以从特 征编码器恢复更多的上下文信息。将多尺度特 征提取模块作为连接编码器和解码器的桥梁, 以恢复更多的多尺度空间信息。主要贡献如下:

首先,提出一种新颖的混合卷积连接模块, 以从特征编码器恢复更多的上下文信息。

其次,提出多尺度特征提取模块,以恢复 更多高级语义特征。

再次,提出了一种用于医学图像分割的混 合卷积网络,它包含一个特征编码器、一个特 征解码器、一个连接不同阶段的特征图的混合 卷积连接模块以及一个连接编码器和解码器的 多尺度特征提取模块,提出了多分支混合注意 力特征解码器。

最后,对CHASEDB数据集和FRSA数据集的实验结果进行了分析。结果表明,与标准分割方法相比,提出的方法效果更好。

1 混合卷积网络

图1为提出的混合卷积医学图像分割网络。 它包括一个特征编码器模块、特征解码器模块、 连接特征编码器和解码器的多尺度特征提取模 块和混合卷积连接模块。接下来,将详细介绍 所提出混合卷积网络的各个部分。

1.1 特征编码器模块

U-Net^[5]中的编码器块由两个 3×3 卷积、一 个 ReLU和2×2 最大池化组成。沿用了 CE-Net^[12], 采用预训练的 ResNet-34^[13]作为特征编码器块。 在 ImageNet 数据集上预训练的权重参数被用作 初始化参数。

1.2 特征解码器模块

在U-Net^[4]中,解码器模块包含一个上采样 操作、一个连接操作和两个用 ReLU 进行的卷 积。不过,与U-Net不同的是,所提出的解码器 由多个解码器块组成,每个解码器块包含一个 转置卷积、与相应的经过混合卷积模块的裁剪 特征图相加、两个3×3卷积、一个通道注意力, 如图1所示。将18×18×512的特征图输入所提出 的特征解码器。首先,使用转置卷积对特征图 进行上采样,得到36×36×512大小的特征图。 然后,将该特征图与经过混合卷积模块的特征 编码器中的相应特征图连接起来,得到一个36× 36×768大小的特征图。两个 3×3 卷积和通道关 注来选择性地加权每个通道的重要性,从而得 到一个36×36×256大小的特征图。经过其他三 个特征解码器块后,得到一个大小为576×576× 32的特征图。最后通过两次3×3的卷积运算得 到分割结果。

1.3 多尺度特征提取模块

为从医学影像中获取更多的多尺度空间信息和高级语义信息,提出了多尺度特征提取模块用于连接特征编码器和解码器。这一模块是连接编码器和解码器子网络底部的桥梁。图2为所提出的多尺度特征模块。如图2所示,该模块四个级联分支。与ResNet中的快捷机制类似,将提取的四层特征图直接添加到原始特征中。 组卷积大小G可通过以下公式求得^[14]:



图 1 所提出的混合卷积医学图像分割网络

$$G = \begin{cases} 1, & g = 3\\ \frac{g-1}{2}, & g > 3 \end{cases}$$
(1)

其中:g=2×i+1,i=1,…,M为卷积核的大 小。卷积核的尺度为3、5、7和9,对应分组卷 积核的大小为1、4、8和16。值得注意的是, 大核卷积提取的是粗粒度信息,而小核卷积则 提取细粒度信息。通过融合不同核大小的卷积, 多尺度特征提取模块可以在更细的粒度水平上 提取多尺度空间信息。



图 2 多尺度特征提取模块

1.4 混合卷积连接模块

U-Net^[4]和CE-Net^[12]使用连接操作从编码器 中捕捉上下文信息。本节将提出混合卷积连接 模块,以从编码器中获取丰富的上下文特征。 它可以实现多尺度上下文信息并保留编码器中 的详细特征。图3为混合卷积连接模块。它包含 一个1×1卷积、两个混合空洞卷积(hybrid dilated convolution, HDC)、将1×1卷积、两个HDC的 输出进行连接并将其通过SE块(Squeezeand-Excite block)。采用两个HDC来扩大感受野,从编码 器中提取更丰富的上文下信息。通过将串联三 个不同的卷积的输出特征图,混合卷积模块可 以减轻在编码过程中逐渐被稀释的全局上下文 和语义特征。最后,将提取更丰富的上下文信 息的特征图传递给特征解码器模块。



图 3 混合卷积连接模块

2 实验与讨论

为了验证提出的混合卷积网络在医学图像 分割方面的性能,在CHASEDB数据集^[15]和股腘 动脉管腔内支架X射线影像FRSA数据集^[11]上对 其进行了测试。CHASEDB数据集包含大小为 999×960×3的28幅视网膜图像。训练数据集包 括前14幅图像,测试数据集包括其余图像。在 CHASEDB中,每张图像的大小被调整为896× 896×3。FRSA数据集包括尺寸为800×800×3的 92幅图像,其中76幅作为训练图像,其余26幅 作为测试图像。

采用二元交叉熵作为损失函数。将数据集的批量大小设为4。采用初始学习率为 0.0001

的 Adam 优化器来优化所有模型。为 CHASEDB 数据集设置了 300 个 epoch,为 FRSA 数据集设置了 100 个 epoch。

2.1 CHASEDB数据集实验

将提出的方法与U-Net^[4]、R2U-Net^[16]、 CE-Net^[12]、DEU-Net^[17]、RSAN^[18]、SCS-Net^[19]和 MHA-Net^[10]进行了比较。使用三个常见的性能 指标,即接收器下面积操作特征曲线(AUC)、 灵敏度(Sen)和准确度(Acc)作为视网膜图像分 割的评估标准。表 1为不同医学图像分割方法 的AUC、Sen和Acc值。从表 1可以看出,本文 方法在AUC、Sen和Acc方面的分割性能最佳。 例如,本文方法的AUC值为0.9905,超过其他 七种方法的AUC值。Sen和Acc的值亦最高。

表 1 不同分割方法在CHASEDB数据集上的 AUC、Sen和Acc的值

| 对比方法 | AUC | Sen | Acc |
|---------|--------|--------|--------|
| U-Net | 0.9869 | 0.8744 | 0.9869 |
| R2U-Net | 0.9815 | 0.7756 | 0.9815 |
| CE-Net | 0.9896 | 0.8729 | 0.9896 |
| DEU-Net | 0.9812 | 0.8037 | 0.9812 |
| RSAN | 0.9894 | 0.8486 | 0.9894 |
| SCS-Net | 0.9867 | 0.8365 | 0.9867 |
| MHA-Net | 0.9897 | 0.8790 | 0.9895 |
| 本文方法 | 0.9905 | 0.8845 | 0.9907 |

图4为不同方法对CHASEDB图像的分割结

果。



CHASEDB数据集上的分割结果

图4中的第1列是两幅原始视网膜图像。图4 中的第2~4列分别是U-Net、CeNet和本文方法 提供的相应分割结果,第5列显示了标注的真实 分割结果。仔细观察图4,本文方法所提供的分割结果优于U-Net和Ce-Net。例如,如图4中红 色方框区域所示,本文方法的分割精度略高于 U-Net和CE-Net。

2.2 股腘动脉管腔内支架X射线影像FRSA 数据集实验

将本文方法与U-Net^[4]、CE-Net^[12]、RSAN^[18] 和 SA-UNet^[20]进行了比较。将 Intersection Ratio (IoU)、Dice coefficient (Dice)、Sen 和 Acc 这四 个常见的性能指标作为评估 FRSA 数据集图像分 割的标准。

表2列出了不同分割方法下 FRSA 图像的 IoU、Dice、Sen和Acc的值。从表2可知,本文 方法在IoU、Dice、Sen和Acc数值方面提供了 最佳的分割性能。本文方法的IoU得分为 0.9355,超过了其它四种方法提供的值。Dice、 Sen和Acc的得分亦最高。

图5第1列是两幅原始FRSA图像。第2~6 列显示了U-Net、CE-Net、RSAN、本文方法和 标注的分割结果。结果表明,本文方法的分割 效果优于U-Net、CE-Net和RSAN。例如,本文 方法提供了更好的分割精度,如图5红色方框区 域所示。

表 2 不同分割方法在FRSA数据集上的 IoU、Dice、Sen和Acc的值

| _ | | | | | |
|---|---------|--------|--------|--------|--------|
| | 对比方法 | IoU | Dice | Sen | Acc |
| | U-Net | 0.8678 | 0.9264 | 0.9533 | 0.9980 |
| | CE-Net | 0.9276 | 0.9623 | 0.9880 | 0.9990 |
| | RSAN | 0.8848 | 0.9389 | 0.9543 | 0.9985 |
| | SA-UNet | 0.8525 | 0.9204 | 0.9274 | 0.9980 |
| | 本文方法 | 0.9355 | 0.9664 | 0.9900 | 0.9991 |



2.3 消融实验

为了评估提出的方法中各个组件的有效性, 在FRSA数据集上进行了消融实验。将没有多尺 度特征提取模块和混合卷积连接模块的网络作 为基线。表3显示加入多尺度特征提取模块或 混合卷积连接模块,可以改善IoU、Dice和Sen 指标。这表明所提出的多尺度特征提取模块捕 捉到了丰富的多尺度特征,混合卷积连接模块 从特征编码器提取更丰富的上下文信息。将多 尺度特征提取模块和混合卷积连接模块相结合, 网络可以获得最佳性能,可以获得更高级的语 义信息。

表 3 在 FRSA 数据集上对多尺度特征提取模块和 混合卷积连接模块进行消融实验

| 方法 | IoU | Dice | Sen |
|--------------|--------|--------|--------|
| 基准 | 0.9172 | 0.9620 | 0.9889 |
| 基准+多尺度特征提出模块 | 0.9300 | 0.9635 | 0.9895 |
| 基准+混合卷积连接模块 | 0.9326 | 0.9650 | 0.9899 |
| 本文方法 | 0.9355 | 0.9664 | 0.9900 |

3 结语

提出了新的混合卷积连接模块和多尺度特 征提取模块,并构造混合卷积网络用于医学图 像分割。提出的混合卷积连接模块可以从编码 器提取更丰富的上下文信息,提出的多尺度特 征提取模块能够获取更多的多尺度空间信息和 高级语义信息。在CHASEDB数据集和FRSA数 据集上的实验结果表明提出的方法优于其他分 割方法。

参考文献:

- [1] SMEETS D, LOECKX D, STIJNEN B, et al. Semi-automatic level set segmentation of liver tumors combining a spiral-scanning technique with supervised fuzzy pixel classification [J]. Medical Image Analysis, 2010, 14(1):13-20.
- [2] LAM B S Y, GAO Y S, LIEW A W-C, et al. General retinal vessel segmentation using regularization-based multiconcavity modeling [J]. IEEE Transaction on Medical Imaging, 2010, 29(7):1369-1381.
- [3] AZZOPARDI G, STRISCIUGLIO N, VENTO M, et al. Trainable cosfire filters for vessel delineation

with application to retinal images [J]. Medical Image Analysis, 2015, 19(1):46-57.

- [4] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C] //Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 2015:234-241.
- [5] WANG B, QIU S, HE H. Dual encoding U-Net for retinal vessel segmentation [C] //Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 2019:84-92.
- [6] WU Y C, XIA Y, SONG Y, et al. Vessel-Net: retinal vessel segmentation under multi-path supervision [C] //Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention, Shenzhen, China, 2019:264-272.
- [7] XU R, WANG C, XU S, et al. DC-Net: dual context network for 2D medical image segmentation [C] //Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Strasbourg, France, 2021: 503-513.
- [8] KHAN S D, ALARABI L, BASALAMAH S. Deep hybrid network for land cover semantic segmentation in high-spatial resolution satellite images [J]. Information, 2021, 12(6):230.
- [9] KHAN S D, ALARABI L, BASALAMAH S. An encoder-decoder deep learning framework for building footprints extraction from aerial imagery [J]. Arabian Journal for Science and Engineering, 2023, 48 (2):1273-1284.
- [10] ZHANG M F, SUN Q, CAI F G, et al. MHA-Net: a multibranch hybrid attention network for medical image segmentation [J]. Computational and Mathematical Methods in Medicine, 2022:8375981-1-7.
- [11] SUN Q, DAI M Y, LAN Z Y, et al. UCR-Net: u-shaped context residual network for medical image segmentation [J]. Computers in Biology and Medicine, 2022, 151:106203(1-10).
- [12] GU Z, CHENG J, FU H, et al. CE-Net: context encoder network for 2D medical image segmentation
 [J]. IEEE Transactions on Medical Imaging, 2019, 38 (10):2281-2292.
- [13] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // Proceedings of

the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016: 770-778.

- [14] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018:7132-7141.
- [15] OWEN C G, RUDNICKA A R, MULLEN R, et al. Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (CAIAR) program [J]. Investigative Ophthalmology & Visual Science, 2009, 50 (5) : 2004-2010.
- [16] ALOM M Z, HASAN M, YAKOPCIC C, et al. Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation[EB/OL]. arXiv:1802.06955,2018.
- [17] WANG B, QIU S, HE H. Dual encoding U-Net for retinal vessel segmentation [C] // Proceedings of the

International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 2019:84-92.

- [18] GUO C, SZEMENYEI M, YI Y, et al. Residual spatial attention network for retinal vessel segmentation [C] //Proceedings of the International Conference on Neural Information Processing, Bangkok, Thailand, 2020; 509-519.
- [19] WU H S, WANG W, ZHONG J, et al. SCS-Net: a scale and context sensitive network for retinal vessel segmentation [J]. Medical Image Analysis, 2021, 70: 102025.
- [20] GUO C, SZEMENYEI M, YI Y, et al. SA-Unet:spatial attention U-Net for retinal vessel segmentation [C] //Proceedings of the 25th IEEE International Conference on Pattern Recognition (ICPR), Milan, Italy, 2021:1236-1242.

A hybrid convolutional network for medical image segmentation

Zhang Meifang^{*}, Li Qikeng, Xie Longteng

(School of Public Health and Health Management, Fujian Health College, Fuzhou 350101, China)

Abstract: Robust segmentation of organs from medical images isone of the key techniques in medical image analysis for disease diagnosis. U-Net is a robust structure for medical image segmentation. However, U-Net uses a continuous downsampling encoder to capture multi-scale features, and the high-level semantic features are not sufficiently recovered, which leads to the loss of contextual information and fails to adequately recover the organ features to be segmented. In this paper, a new hybrid convolutional network is proposed to capture more contextual information and high-level semantic features. The main idea of the hybrid convolutional network is to extract more contextual information and high-level semantic features from the feature encoder using the proposed hybrid convolutional connectivity module. The multi-scale feature extraction module is used to connect the encoder and decoder sub-networks to obtain richer multi-scale feature maps. The proposed method is compared with the state-of-the-art methods on CHASEDB dataset and FRSA dataset. The experimental results show that the proposed method outperforms other segmentation methods.

Keywords: medical image segmentation; hybrid convolutional; multi-scale feature; contextual information

文章编号:1007-1423(2024)08-0056-06

DOI: 10.3969/j.issn.1007-1423.2024.08.009

基于多尺度注意力机制的单幅图像超分辨率重建

阿火黄军,严 华*

(四川大学电子信息学院,成都 610065)

摘要:近年来,深度卷积神经网络(CNN)在单幅图像超分辨率重建中取得了明显的进展。在此基础上,提出了一 个校准多尺度通道空间注意网络(CMCSAN)。CMCSAN由两个关键模块组成:校准多尺度模块(CMSM)和通道空间注 意模块(CSAM)。CMSM从不同尺度提取特征信息,自适应调整信息特征;CSAM模块可以自动鉴别不同通道的特征信 息,有效调整空间的位置权重。实验结果表明,CMCSAN显著增强了挖掘中间特征信息的能力,在单幅图像超分辨率 重建中表现出良好的性能。

关键词: 单幅图像; 校准多尺度模块; 通道空间注意模块; 超分辨率重建

0 引言

图像超分辨率(super-resolution, SR)是计算 机视觉领域的研究课题之一,目前广泛应用于 医学成像、交通监控及卫星图像等领域。SR 的 关键任务是找到一个从低分辨率图像到高分辨 率图像之间的映射关系,这个映射关系不是唯 一的,所以SR 是一个不适定问题。为了解决这 个问题,研究人员提出了基于插值的方法^[1]、 基于稀疏编码^[2]的方法及基于学习等方法,其 中基于学习方法是目前主流的方法。

近年来,基于深度卷积神经网络的单幅图 像超分辨率重建任务得到了广泛的研究。Dong 等^[3]提出的SRCNN是第一个将CNN引入到超分 辨率任务中的方法,它使用双三次插值来提高 低分辨率的图像,得到了比传统方法更好的高 分辨率图像。Kim等^[4]提出了VDSR算法,并引 入了残差网络,使网络层数达到20层,充分利 用了中间特征,性能优于SRCNN。EDSR^[5]设计 了集成残差、去除BN层的网络,进一步提高了 传统残差网络在SR的性能。Zhang等^[6]提出一 种极深残差通道注意网络RCAN,降低了网络学 习难度。虽然现有的基于CNN的单幅图像超分 方法取得了进展,但是这些模型仍然存在一些 局限。

针对以上问题,我们设计了一种CMCSAN, 能充分获取全局和局部特征信息,增强特征信 息的提取,从而恢复出质量更高的重建图像。 CMCSAN主要由CMSM和CSAM组成。在CMSM 中使用多条分支提取不同尺度信息,提升了多 尺度表示能力。CSAM可以鉴别学习不同通道特 征信息进一步挖掘中间特征信息。我们的网络 可以更好地提取丰富的特征信息,这些特征信 息使我们的模型能够更准确地恢复高分辨率图 像,提升了重建图像质量。

1 相关工作

1.1 多尺度特征提取块

多尺度特征表示已经广泛应用于计算机图 视觉任务中,图像的特征在不同尺度下会表现 出不同的特征,这为提高图像超分辨率重建性 能提供了潜在条件。GoogLeNet^[7]使用核大小不 同的并行滤波器来充分利用多尺度特性。之后, Li等^[8]提出了一种多尺度特征提取方法 MSRB, 旨在有效地利用不同尺度的图像特征。该方法

收稿日期: 2023-11-07 修稿日期: 2023-12-03

作者简介: 阿火黄军(1995—),男,四川雷波人,在读硕士研究生,研究方向为基于深度学习的单幅图像超分辨率重建; *通信作者:严华(1971—),男,四川渠县人,博士,教授,研究方向为智能信息系统,E-mail:yanhua@scu.edu.cn

将多个卷积核集成到一个块中,以自适应地捕获不同尺度的图像特征,有效提升了SR性能。

然而,以上方法存在一定的限制,未能充 分利用前一层的图像特征,导致局部特征难以 在不同层次之间传递,从而抑制了多尺度特征 的有效传递,同时也增加了计算负担。因此, 需要进一步探索一种更为高效的多尺度特征信 息提取块,以进一步优化SR的性能。

1.2 注意力机制

注意力机制将深度神经网络的重点聚焦到 信息更多的区域。它会忽略与网络不相关的信 息而专注于重要的信息。近年来,注意力机制 被成功地应用于基于 CNN 的图像增强方法中, 它试图模拟人类的视觉系统来聚焦输入信号的 基本区域。Hu等^[9]提出了挤压激励块来建模通 道方向的关系,通过显式建模信道间的相互依 赖,自适应地重新校准信道特征响应。受 SENet^[9]启发, Zhang等^[6]提出了一种残差通道注 意网络RCAN,把通道注意机制引入到残差块 中,通道关注机制采用了全局平均池化来提取通 道的统计信息,并根据这些信息为不同的通道分 配不同的权重。这样,网络可以更有针对性地关 注那些包含信息丰富特征的通道。Dai 等^[10]提出 了一种精确图像重构的深度二阶注意力网络 SAN, 用于捕获远距离依赖关系和学习更具鉴 别性的表示。Liu 等^[11]提出一种增强空间注意模 块, 该模块的设计目的在于有效地整合特征残 差信息。通过引入跳跃连接, 该模块能够在每 个局部残差分支上传播特征,使得神经网络更 加集中关注关键的空间内容。

Niu 等^[12]提出 HAN 网络,设计了一种通道 空间注意模块用于学习每个通道位置的置信度, 选择性地捕获更多信息特征。在本文中我们引 入了一个通道空间注意模块,用来增强网络中 特征信息间的相互依赖关系,更有效地捕捉和 整合特征信息,使网络能够充分获取图像细节 信息,从而提高重建图像的质量。

2 校准多尺度通道空间注意网络

在本节中,我们将对提出的校准多尺度通 道空间注意网络整体结构进行一个概述,然后 分别对网络框架中的核心特征增强部分校准多 尺度模块和通道空间注意模块进行详细介绍, 最后对网络中用到的损失函数进行说明。

2.1 网络总体结构

校准多尺度通道空间注意网络如图1所示, 包括三个部分,浅层提取部分、特征增强部分和 上采样重建部分。增强特征块(enhanced feature block, EFB)是网络的核心,它由一个CMSM和一 个CSAM组成。CMSM从不同尺度提取特征信息, 进行自适应调整,然后将提取的特征信息输入到 CSAM中,进一步挖掘特征信息的空间位置的相 关性。在整个网络中,多个EFB以链式结构进行 堆叠,这种模式能有效增加网络的深度,使网络 有更大的感受野捕捉高级特征信息。



图1 校准多尺度通道空间注意网络框架

浅层提取部分(shallow feature extractor, SFE), 给定一张LR图像 I_{LR} 作为网络的初始输入,使用 一个卷积层从LR中提取浅层特征 F_0 ,用于后续 工作,可以表示为

$$F_0 = H_{\rm SFE}(I_{\rm LR}) \tag{1}$$

其中: $H_{SFE}(\cdot)$ 表示浅层特征提取函数,它是一个卷积核为3×3的卷积层, F_0 是提取到的浅层特征。

然后将浅层特征 F₀送入特征增强部分,特 征增强部分包含20个串联的中间高级特征提取 模块 EFB。EFB 能够充分获取不同尺度特征信 息,然后提升通道和空间特征信息的依赖关系, 最后生成如下的高级特征。

 $F_{m}^{i} = H_{EFB}^{i}(F_{m}^{i-1}), i = 1, 2, ..., N$ (2) 其中: $H_{EFB}^{i}(\cdot)$ 是增强特征部分第i个 EFB, N是 EFB 的数量, F_{m}^{i} 表示特征增强中的第i个 EFB 的 输出,这样就从浅层特征中提取出了高级特征 信息。最后将提取的高级特征 F_{m}^{N} 通过上采样模 块,再与经过双线性插值处理的原始 I_{LR} 进行求 和,得到最终重建图像 I_{SR} .

$$I_{\rm SR} = \operatorname{Conv}\left(H_{\rm UP}(F_m^N)H_{\rm Int} \oplus (I_{\rm LR})\right) \tag{3}$$

其中: F_m^N 表示最后的高级特征, $H_{up}(\cdot)$ 表示上 采样恢复函数,Conv表示卷积核为3×3的卷积 层, $H_{lnt}(\cdot)$ 表示插值函数, \oplus 表示上采样特征 信息和原始 I_{LR} 插值信息进行融合。

2.2 校准多尺度模块 CMSM

为了利用多尺度信息,MSRN^[13]被引入到超 分辨率任务中来检测不同尺度上的特征,以提 升重建超分辨率图像的准确性。然而,MSRN的 模型比较复杂,训练和计算成本较高。

受文献[13]、[14]和[15]的启发,我们将 MSRN^[13]中的密集分支简化为与RFB^[14]相同的四 条分支,然后加上一条残差校准分支,得到一 个如图2所示的校准多尺度模块。校准多尺度模 块的各分支先对浅层提取特征信息F₀进行处理, 然后引入膨胀卷积来提升模块的感受野,上面 三条分支都由一个膨胀卷积和两个卷积核分别 为1×1和3×3普通卷积层组成,三条分支从上到 下的扩张率分别为2、3和5,用MS(·)表示, 第四个分支仅由一个核为3×3的卷积构成。然 后,为了充分获取特征信息,将上一分支的输 出与下一条分支的输出进行融合,直到处理完 全部分支,此过程可表示为

$$FM_{i} = \begin{cases} Conv(F_{0}), i = 1\\ MS(F_{i}), i = 2, 3, 4 \end{cases}$$
(4)

其中: Conv表示卷积核为3×3的卷积层, FM_i表示不同分支的输出信息。



图 2 校准多尺度模块

然后,再将四条分支提取的信息进行全局 融合。表达式如下:

 $F_{u} = FM_{1} + FM_{2} + FM_{3} + FM_{4}$ (5) 其中: F_{u} 表示四条分支融合后的输出。 最后,为了进一步将底层信息融合到已经 捕获的多尺度信息中,我们在多尺度模块中加 上一条卷积核为3×3的残差校准分支。

$$F_{\rm MS} = F_u + \rm{Conv}(3) \tag{6}$$

其中: F_{MS}表示最终的输出, Conv(3)表示卷积 核为3的校准分支。

2.3 通道空间注意模块CSAM

通道空间注意模块包含来自特征映射各个 维度的响应,主要是用来增强网络对不同位置 特征的表达能力。通过在不同位置之间进行关 注,使得网络能够更好地挖掘出不同位置之间 的相关性,提高对图像空间结构的理解能力。 在超分辨率任务中,通道空间关注模块可以帮 助网络更好地提取纹理和边缘等高频信息,从 而提高SR性能。

通道空间注意模块如图3所示。首先将经 过校准多尺度模块的输出特征输入到一个步长 为1、卷积核大小为3×3×3的三维卷积层中,然 后使用Sigmoid激活函数对三维空间中相邻特征 映射进行卷积,得到一个三维关注矩阵,该矩 阵可以对相邻特征之间的相关性进行建模。最 后将输入特征乘以注意力矩阵,以加强相邻特 征之间的层次关系。这个过程可以表示为

 $F_{CSAM} = (F_{MS} \otimes (H_{*}(H_{3D}(F_{MS})))) \oplus F_{MS}$ (7) 其中: $H_{3D}(\cdot)$ 表示 3D 卷积函数, $H_{*}(\cdot)$ 表示 Sigmoid 激活函数, $F_{CSAM}(\cdot)$ 表示 CSAM 的输出。 CSAM 模块能自适应地学习通道间和通道内的特 征信息,在超分辨率任务中,CSAM 可以帮助网 络更好地提取纹理和边缘等高频信息,从而提 高网络对输入图像的感知能力。





3 实验结果与分析

3.1 数据集及评价指标

在图像重建技术中经常使用 DIV2K 数据集 对模型进行训练,它是一个高质量的图像数据 集,其中包含 800 个训练图像、100 个验证图像 和100个测试图像。我们使用DIV2K数据集中的800幅训练图像作为训练集。并用Set5、Set14、Urban100、B100和Manga109五个基准数据集对模型的性能进行了验证。

本文的客观评价指标主要采用峰值信噪比 (PSNR)和结构相似指数度量(SSIM)。PSNR和 SSIM值越高,SR重建图像质量越好。我们用双 三次线性插值(BI)和模糊降阶(BD)退化模型进 行实验。在变换后的YCBCR颜色空间的亮度Y 通道上计算PSNR和SSIM,使用PSNR和SSIM 来评估SR的结果性能。

3.2 实施细节

我们使用 PyTorch 平台实现了所提出的网络,在训练模型中,通过对 HR 图像进行双三次线性插值下采样得到退化图像。然后使用水平翻转、垂直翻转和随机旋转90°的方法对 DIV2K数据集中800幅训练图像进行数据增强。以小批量作为网络的输入,每个小批量由16个尺寸为32×32的 LR 色块组成。用 ADAM 优化器对模型

进行训练,优化器参数分别设置为β1=0.9,β2= 0.999,初始学习率设置为*lr*=0.0001,每经过训 练总数 epoch 的 0.5 倍后,将学习率更新为初始 值的 0.1 倍。训练和测试环境使用的是 NVIDIA GeForce RTX 2080 GPU。

3.3 实验对比

实验过程中分别以比例因子为×2、×3、×4 双三次插值对原始图像进行下采样退化,生成 实验原始低分辨率图像。然后对重建图像在 YCbCr 空间上的Y通道上进行PSNR和SSIM的 测量,以此来衡量不同比例因子下的网络重建 图像效果。

我们在Set5、Set14、Urban100、B100和 Manga109五个标准数据集上进行了测试,然后 将本文方法与Bicubic^[1],SRCNN^[3],LapSRN^[16], MemNet^[17],EDSR^[5],ARRFN^[18]几种常见方法 进行了对比。表1展示了不同网络在比例因子 为×2、×3、×4时重建图像的PSNR和SSIM平 均值。

表 1 比例因子为×2、×3、×4时不同方法在五个基准数据集上的结果对比

| 一十十 | 计 十日乙。 | S | et5 | Se | et14 | В | 100 | Urba | an100 | Man | ga109 |
|---------|---------------------------|-------|--------|-------|--------|-------|--------|-------|--------|-------|--------|
| 刀伝 | 成 八凶丁 [。] | PSNR | SSIM |
| Bicubic | ×2 | 33.66 | 0.9299 | 30.24 | 0.8688 | 29.56 | 0.8431 | 26.88 | 0.8403 | 30.80 | 0.9339 |
| SRCNN | ×2 | 36.66 | 0.9542 | 32.45 | 0.9067 | 31.36 | 0.8879 | 29.50 | 0.8946 | 35.60 | 0.9663 |
| LapSRN | ×2 | 37.52 | 0.9591 | 32.99 | 0.9124 | 31.80 | 0.8952 | 30.41 | 0.9103 | 37.27 | 0.9740 |
| MemNet | ×2 | 37.78 | 0.9597 | 33.28 | 0.9142 | 32.08 | 0.8978 | 31.31 | 0.9195 | 37.72 | 0.9740 |
| EDSR | ×2 | 37.99 | 0.9604 | 33.57 | 0.9175 | 32.16 | 0.8994 | 31.98 | 0.9272 | 38.54 | 0.9769 |
| ARRFN | ×2 | 38.01 | 0.9606 | 33.66 | 0.9179 | 32.20 | 0.8999 | 32.27 | 0.9295 | — | — |
| Our | ×2 | 38.03 | 0.9603 | 33.71 | 0.9183 | 35.31 | 0.9444 | 31.86 | 0.9264 | 38.76 | 0.9810 |
| Bicubic | ×3 | 30.39 | 0.8682 | 27.55 | 0.7742 | 27.21 | 0.7385 | 24.46 | 0.7349 | 26.95 | 0.8556 |
| SRCNN | ×3 | 32.75 | 0.9090 | 29.30 | 0.8215 | 28.41 | 0.7863 | 26.24 | 0.7989 | 30.48 | 0.9117 |
| LapSRN | ×3 | 33.81 | 0.9220 | 29.79 | 0.8325 | 28.82 | 0.7980 | 27.07 | 0.8275 | 32.21 | 0.9350 |
| MemNet | ×3 | 34.09 | 0.9248 | 30.00 | 0.8350 | 28.96 | 0.8001 | 27.56 | 0.8376 | 32.51 | 0.9369 |
| EDSR | ×3 | 34.37 | 0.9270 | 30.28 | 0.8417 | 29.09 | 0.8052 | 28.15 | 0.8527 | 33.45 | 0.9439 |
| ARRFN | ×3 | 34.38 | 0.9272 | 30.36 | 0.8422 | 29.09 | 0.8050 | 28.22 | 0.8533 | — | — |
| Our | ×3 | 35.65 | 0.9514 | 31.27 | 0.8745 | 30.23 | 0.8357 | 30.17 | 0.8916 | 33.97 | 0.9523 |
| Bicubic | ×4 | 28.42 | 0.8104 | 26.00 | 0.7027 | 25.96 | 0.6675 | 23.14 | 0.6577 | 24.89 | 0.7866 |
| SRCNN | ×4 | 30.48 | 0.8628 | 27.50 | 0.7513 | 26.90 | 0.7101 | 24.52 | 0.7221 | 27.58 | 0.8555 |
| LapSRN | ×4 | 31.54 | 0.8852 | 28.19 | 0.7720 | 27.32 | 0.7275 | 25.21 | 0.7562 | 29.09 | 0.8900 |
| MemNet | ×4 | 31.74 | 0.8893 | 28.26 | 0.7723 | 27.40 | 0.7281 | 25.50 | 0.7630 | 29.42 | 0.8942 |
| EDSR | ×4 | 32.09 | 0.8938 | 28.58 | 0.7813 | 27.57 | 0.7357 | 26.04 | 0.7849 | 30.35 | 0.9067 |
| ARRFN | ×4 | 32.22 | 0.8952 | 28.60 | 0.7817 | 27.57 | 0.7355 | 26.09 | 0.7858 | — | — |
| Our | ×4 | 32.27 | 0.8974 | 28.76 | 0.7895 | 29.73 | 0.8029 | 26.04 | 0.7854 | 31.75 | 0.9284 |

注:黑体为最好结果。

从表1可以看出,最基本的基于深度学习的 方法 SRCNN 也优于传统双三次插值方法。 SRCNN虽然有效改进了传统方法,但是从表中 可以看出,与其他方法相比还存在较大的改进 空间。比例因子为×2时,本文方法与ARRFN方 法比较,在Set5上的PSNR提高了0.02 dB。比 例因子为×4时,本文方法在Set14上的PSNR值 比ARRFN方法高0.16 dB。比例因子为×3时, 本文方法比其他算法都表现出更好的重建效。

4 结语

本文提出了一种基于校准多尺度通道空间 注意网络来改进单幅图像超分辨率重建技术。 具体而言,方法的核心是提出了一个增强特征 块EFB,它能充分利用多尺度信息并增强中间 特征的相关性。EFB 主要由校准多尺度模块 CMSM 和通道空间注意模块 CSAM组成,CMSM 通过从不同尺度提取特征信息并自适应地调整 这些信息,捕获图像的局部和全局特征信息。 CSAM的引入有助于识别不同通道和空间特征之 间的相互依赖关系,进一步增强了中间特征的 表示能力,获取更多细节信息。我们使用公共 数据集训练网络,并在五个不同的公共数据集 上进行测试,实验结果表明,本文提出的校准 多尺度通道空间注意网络在图像重建视觉效果 上取得了较好的效果。

参考文献:

- ZHANG L, WU X. An edge-guided image interpolation algorithm via directional filtering and data fusion
 IEEE Transactions on Image Processing, 2006, 15
 (8):2226-2238.
- [2] YANG J, WRIGHT J, HUANG T S, et al. Image super-resolution via sparse representation [J]. IEEE Transactions on Image Processing, 2010, 19 (11) : 2861-2873.
- [3] DONG C, LOY C C, HE K, et al. Learning a deep convolutional network for image super-resolution [C] //Proceedings of the 13th European Conference on Computer Vision, Zurich, Switzerland. Springer International Publishing, 2014:184-199.
- [4] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of the IEEE Conference on

Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016:1646-1654.

- [5] LIM B, SON S, KIM H, et al. Enhanced deep residual networks for single image super-resolution [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 2017: 136-144.
- ZHANG Y, LI K, LI K, et al. Image super-resolution using very deep residual channel attention networks
 //Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 2018: 286-301.
- [7] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 2015:1-9.
- [8] LI J, FANG F, MEI K, et al. Multi-scale residual network for image super-resolution [C] // Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 2018:517-532.
- [9] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018:7132-7141.
- [10] DAI T, CAI J, ZHANG Y, et al. Second-order attention network for single image super-resolution [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019:11065-11074.
- [11] LIU J, ZHANG W, TANG Y, et al. Residual feature aggregation network for image super-resolution [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020:2359-2368.
- [12] NIU B, WEN W, REN W, et al. Single image super-resolution via a holistic attention network [C] //Proceedings of the 16th European Conference on Computer Vision, Glasgow, UK. Springer International Publishing, 2020: 191-207.
- [13] LI J, FANG F, MEI K, et al. Multi-scale residual network for image super-resolution [C] // Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 2018:517-532.
- [14] LIU S, HUANG D. Receptive field block net for accurate and fast object detection [C] // Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 2018: 385-400.

- [15] LIU J J, HOU Q, CHENG M M, et al. Improving convolutional networks with self-calibrated convolutions [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020:10096-10105.
- [16] LAI W S, HUANG J B, AHUJA N, et al. Deep laplacian pyramid networks for fast and accurate super-resolution [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recogni-

tion, Honolulu, HI, USA, 2017:624-632.

- [17] TAI Y, YANG J, LIU X, et al. Memnet: a persistent memory network for image restoration [C]//Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017:4539-4547.
- [18] QIN J, ZHANG R. Lightweight single image super-resolution with attentive residual refinement network[J]. Neurocomputing, 2022, 500:846-855.

Single image super-resolution reconstruction based on multi-scale attention mechanism

Ahuo Huangjun, Yan Hua*

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

Abstract: In recent years, deep convolutional neural network (CNN) has made significant progress in single-image super-resolution reconstruction. Building upon this foundation, a Calibrated Multi-scale Channel-Space Attention Network (CMCSAN) is introduced. CMCSAN comprises two crucial modules: the Calibrated Multi-scale Module (CMSM) and the Channel-Space Attention Module (CSAM). CMSM extracts features from various scales and adaptively adjusts feature information. The CSAM module automatically distinguishes different channel features and effectively adjusts spatial positional weights. Experimental results demonstrate that CMCSAN significantly enhances the ability to extract intermediate features and exhibits promising performance in single-image super-resolution reconstruction.

Keywords: single-image; calibrated multi-scale module; channel-space attention module; super-resolution reconstruction

(上接第16页)

Information entropy-based vein recognition for two-dimensional local binary patterns

Zhang Yunfei, Li Jiangmei, Chen Xi*

(School of Big Data and Computer Science, Guizhou Normal University, Guiyang 550025, China)

Abstract: Based on the problem that the existing LBP algorithm and its variants are unable to extract high-dimensional features from images, a two-dimensional local binary pattern recognition algorithm based on information entropy is proposed. This method first extracts the low-dimensional features of the image using the unified local binary pattern (ULBP), then combines the image information entropy with the unified local binary pattern atlas to obtain the entropy-weighted unified local binary pattern atlas (EULBP), and realizes the statistics of the co-occurring feature information among the patterns in the local area using the sliding window, and uses the result as the image feature expression. And the pattern classifier constructed on the basis of histogram cross distance is used to verify its recognition performance. The experimental results show that the proposed algorithm can achieve an average recognition rate of 99.94% and 98.84% in both the SDUMLA-HMT dataset as well as the Universiti Teknologi Malaysia finger vein dataset(FV-USM).

Keywords: two-dimensional co-occurrence of local binary patterns; information entropy; rotational invariance; directional features

文章编号:1007-1423(2024)08-0062-05

DOI: 10.3969/j.issn.1007-1423.2024.08.010

基于YOLOv5算法的学生课堂行为识别研究

马瑞珵*,陈 继,王炳怀,龙俊丞,刘 宇

(西南民族大学电子信息学院,成都 610225)

摘要:随着智慧课堂和教育大数据挖掘的普及,创建一种智慧课堂中学生学习数据的自动检测和分析方法成为可能。基于YOLOv5模型实现两类学习数据:班级同学计数、学习行为识别。该模型对低头写字、低头看书、抬头听课、转头、举手、站立、小组讨论七种学生课堂行为进行识别,以此辅助教师判断学生学习情况并做出教学决策。研究表明,检测结果精确度达到97.92%。

关键词:学生行为识别;YOLOv5;计算机视觉技术

0 引言

人工智能学习分为机器学习与深度学习两 方面,而卷积神经网络(CNN)则是深度学习的 代表之一。目前国内外已发表了很多基于视频 的人体行为识别算法,许多相应技术也在不断 进步,识别的准确率也在逐年提高。但基于 CNN模型的学生课堂行为识别技术大多依旧处 于理论探索阶段,尚未建成健全的智慧课堂系 统。随着人工智能不断发展和教育信息化不断 推进,"智慧教育"从理论走向实际^[1]。在国家 《中国教育现代化2035》等政策推动下,各大高 校正积极探索智慧课堂应用的新理念、新路径。 在传统课堂中,教师往往为一对多的"共享式" 模式,缺少对学习者个性化数据的分析获取。

对于人体行为识别而言,人体特征提取是 图像识别的基础。国内外对特征提取方面的研 究已有很多。赵海勇等^[2]以人体模板特征为基 准,通过前景与背景相减获取人体轮廓;朱煜 等^[3]则认为运动中的人体形状具有一定的柔性, 不能用简单的数学模型来描述运动过程中的人 体形状,同时在复杂环境下通过运动检测无法 准确获得人物位置;李长海^[4]认为光流法与特征跟踪等传统识别方法所提取的特征适用范围 过窄,在复杂的多人环境下识别准确效果较低, 而深度学习网络可以从数据集中学习到特定的 模式,在人体行为识别的应用上有显著的效果。

将人体行为识别技术应用至课堂行为识别 方面在近年来都是比较热门的内容,现在应用 于课堂的学习数据挖掘和分析技术已有一定发 展基础,目前学界已有大量学者提出对教育大 数据的挖掘有助于发现学习规律、支持教学决 策、进行个性推荐,从而充分发挥技术优势赋 能教学的效果和效益提升。学术界已经提出了 许多新型模型与算法,提高目标检测的精确性 与效率,如魏艳涛等^[5]采用基于VGG16网络模 型的卷积神经网络,相较于传统卷积神经网络拥 有了更多层数,特征学习能力更强;黄勇康等[6] 采取了目标检测模型加目标跟踪算法,再结合能 够获取深层次行为时空特征的深度残差3D卷积 行为识别算法来解决课堂场景中的学生课堂行为 识别问题。本文拟基于YOLOv5创建一种智慧课 堂中学生学习数据的自动检测和分析方法,并运 用于智慧课堂中,助力教育智慧化转型。

基金项目:四川省大学生创新创业训练计划项目(X202210656224)

收稿日期: 2023-11-07 修稿日期: 2023-12-07

作者简介: *通信作者:马瑞珵(2002—),男,云南昆明人,在读本科,研究方向为电子信息工程,E-mail:3102055376@qq.com; 陈继(2002—),男,四川达州人,在读本科,研究方向为计算机科学与技术; 王炳怀(2002—),男,河北沧州人,在读本科,研 究方向为电子信息工程; 龙俊丞(2003—),男,四川达州人,在读本科,研究方向为网络工程; 刘宇(2002—),男,河北 承德人,在读本科,研究方向为电子信息工程

· 63 ·

1 目标检测算法

1.1 目标检测算法

传统的目标检测算法分为单阶段(One Stage) 与两阶段(Two Stage)两类,其中以YOLO算法 与SSD算法为代表的单阶段算法通过快速识别 图像,对图像检测区域进行预估并采集,具有 实时性的特点,缺点则是识别精度不够准确; 两阶段算法对数据集的识别精度相较于单阶段 算法更高,缺点则是识别速度偏慢。

1.2 YOLOv5算法模型

YOLO系列算法是一种运用于目标检测领域 的单阶段算法,截至2023年10月19日,其版本 已经前后更新并推出至YOLOv8版本,而在其以 往版本中又以YOLOv5最为出名,被广泛运用于 教学场景中。YOLOv5继承了第四代的基本结 构,同时根据不同通道的尺度缩放,依照模型 从小到大构建出YOLOv5-N/S/M/L/X五种模型^[7]。 YOLOv5也引入了自适应技术,针对不同的检测 目标图片实行裁剪、缩放等以实现识别目标图 片规格的统一。作为单阶段算法,YOLOv5也有 足够快的运算速度,可以实现实时监测。因此, 本研究选择了YOLOv5算法作为研究设计之用, 其流程设计如图1所示。



图 1 YOLOv5算法流程图

1.3 目标行为检测评价指标

TP(True Positives): 正样本被正确识别为正 样本。

TN(True Negatives): 负样本被正确识别为 负样本。

FP(False Positives): 负样本被错误识别为 正样本。

FN(False Negatives):正样本被错误识别为 负样本。

精确率(Precision):所有被识别为正样本的

样本中实际为正样本的概率, 计算公式为

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

召回率(Recall):在实际为正的样本中被识 别为正样本的概率,计算公式为

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

PR曲线(Precision-Recall Curve): Precision-Recall曲线越接近(1, 1),代表模型的性能越好。

AP(Average Precision): PR曲线的线下面面积。

mAP(mean Average Precision): 各类别 AP 的平均值。

2 模型训练

2.1 数据集处理分类

数据处理对深度学习的重要性不言而喻, 由于各教育机构对学生隐私的保护,目前课堂 行为识别目标检测的相关数据集并未完全建立。 除去目前已有的公共数据集外,为确保拥有足 量训练数据并且完全模拟出摄像头的拍摄角度 以分析视线阻碍等情况,通过网络公开课截取 图片并进行处理标记,以扩大数据集,确保训 练模型适用于各种场合下的学生课堂学习行为 分析。本文主要选取中小学课堂数据为研究对 象,以更好地获得各类姿态行为模型。部分图 片数据如图2所示。



图 2 数据集样例

将学生课堂行为分为低头写字(dx)、低头 看书(dk)、抬头听课(tt)、转头(zt)、举手(js)、 站立(zl)、小组讨论(xt)七类;考虑到授课教师 对数据的影响,补充教师识别(jz)这一类型。完 成分类后,为其编号,见表1。

表 1 各类别编码及数目

| 类别 | dx | dk | tt | zt | js | zl | xt | jz |
|----|-------|-------|--------|------|------|------|------|-----|
| 编码 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 个数 | 72462 | 58932 | 117528 | 5339 | 4183 | 4101 | 4663 | 680 |

处理好的数据集分布情况如图3所示。



图 3 数据集分布情况

2.2 算法训练

实验选择训练设备 CPU 为第 12 代英特尔酷 睿 i7-12700H处理器,在 RTX3060 显卡支持下配 置完成 YOLOv5s 模型。根据所使用硬件配置,设 置训练参数:训练次数 300 次, batch-size 为 8。 YOLOv5 网络运行结果如图 4 所示。

YOLOv5部分训练结果如图5所示,部分验证结果如图6所示。

根据上文表2、表3训练日志记录数据显示,经过训练以后,训练集与验证集的损失函数均有所减小,在250轮之后损失值数值趋于平



图 4 YOLOv5网络运行结果



图 5 部分训练集识别结果

图 6 部分验证集识别结果

表 2 第0、99、199、299轮训练数据(1)

| 训练轮数 | train/box_loss | train/obj_loss | train/cls_loss | metrics/precision | metrics/recall | metrics/mAP_0.5 | metrics/mAP0.5:0.95 |
|------|----------------|----------------|----------------|-------------------|----------------|-----------------|---------------------|
| 0 | 0.073945 | 0.23269 | 0.0439460 | 0.62092 | 0.26123 | 0.21678 | 0.10391 |
| 99 | 0.030562 | 0.12208 | 0.0078262 | 0.92529 | 0.91466 | 0.96047 | 0.80648 |
| 199 | 0.028398 | 0.11124 | 0.0065061 | 0.94828 | 0.92736 | 0.97182 | 0.82759 |
| 299 | 0.027524 | 0.10566 | 0.0055784 | 0.95975 | 0.94097 | 0.97921 | 0.84226 |

表 3 第0、99、199、299轮训练数据(2)

| 迭代次数 | val/box_loss | val/obj_loss | val/cls_loss | x/lr0 | x/lr1 | x/lr2 |
|------|--------------|--------------|--------------|-----------|-----------|-----------|
| 0 | 0.050587 | 0.184570 | 0.0325020 | 0.0033316 | 0.0033316 | 0.0700160 |
| 99 | 0.023986 | 0.106030 | 0.0058159 | 0.0078311 | 0.0078311 | 0.0078311 |
| 199 | 0.023463 | 0.101660 | 0.0050206 | 0.0033321 | 0.0033321 | 0.0033321 |
| 299 | 0.023273 | 0.098422 | 0.0043732 | 0.0010010 | 0.0010010 | 0.0010010 |

稳,各类AP平均值达到97.92%,模型训练精度 较高。

3 结语

3.1 总结

通过YOLOv5算法训练,本文所设计模型检测精度较高,在学生课堂行为识别中有一定出 色表现;同时算法为单阶段算法,运用YO-LOv5s模型满足课堂行为识别中的实时性需求。

此次研究仍存在一定的不足:首先本研究 模型仅依托于YOLOv5算法的目标检测功能,与 实际上学生行为的具体识别还存在一定误差, 对人体姿态的识别能力有限;同时,仅依靠姿 态识别,不对学生所交互的对象进行识别分类, 很难判断同一姿态下的学生的具体行为。其次 在数据集处理方面,当处理视频为可识别的图 片作为数据集时,对视频进行逐帧截取时各帧 间距过短,导致数据集里存在许多相似度高的 图片;同时在验证集选择方面,选择训练集与 验证集比例为7:1,验证集数量占比偏低。

3.2 下一步改进方案

在课堂行为识别反馈方面,赵春等^[8]提出 了"学习行为投入度"理念,将学生课堂行为通 过识别后,实行投入度测量与量化分析,使得教 师与学生间互动得以增强。"投入度"这一概念 的提出,让教师能更客观地获取学生课堂表现, 从而对自身教学方式、教学计划做出相应调整。 因此下一步研究除了将选择通过 OpenPose 实现 人体关键点检测,同时通过 YOLO 实现目标交互 物体检测,以更好地判断学生课堂行为,提高识 别准确性以外,也会尝试将学生面部表情等考量 因素补充为判断学生课堂行为的标准之一。

参考文献:

- [1] 马瑞祾,徐娟.语言智能赋能国际中文智慧教育:现 实境况与未来路向[J].国际中文教育(中英文), 2023,8(2):43-52.
- [2] 赵海勇,刘志镜,张浩.基于模板匹配的人体日常行为识别[J].湖南大学学报(自然科学版),2011(2): 88-92.
- [3] 朱煜,赵江坤,王逸宁,等.基于深度学习的人体行 为识别算法综述[J].自动化学报,2016(6): 848-857.
- [4] 李长海.基于深度学习的人体行为识别算法研究
 [D].成都:电子科技大学,2021.
- [5] 魏艳涛,秦道影,胡佳敏,等.基于深度学习的学生 课堂行为识别[J].现代教育技术,2019(7):87-91.
- [6] 黄勇康,梁美玉,王笑笑,等.基于深度时空残差卷 积神经网络的课堂教学视频中多人课堂行为识别
 [J].计算机应用,2022(3):736-742.
- [7] 邵延华,张铎,楚红雨,等.基于深度学习的YOLO
 目标检测综述[J].电子与信息学报,2022,44(10):
 3697-3708.
- [8] 赵春,舒杭,顾小清.基于计算机视觉技术的学生课 堂学习行为投入度测量与分析[J].现代教育技术, 2021(6):96-103.

(下转第71页)

文章编号:1007-1423(2024)08-0066-06

DOI: 10.3969/j.issn.1007-1423.2024.08.011

PyTorch框架下的复杂场景目标识别方法研究

张进军*

(安徽警官职业学院信息管理系,合肥 230031)

摘要:人工智能框架(PyTorch+OneDNN)在复杂场景目标识别上,易陷入网络参数梯度超规模的问题。为此,提 出一种基于PyTorch框架的复杂场景目标识别方法。引入MPI中的Ring Allreduce 算法,优化PyTorch框架,以此实现 复杂场景特征迭代提取过程中的超规模数据同步和规约处理。以优化后的PyTorch框架为基础,考虑到复杂场景目标特 征与背景特征之间的交叉性,构建增强多尺度特征层输出的目标特征之间的关联。借助反卷积特征融合操作和残差融 合操作的优势,依据上述关联性,实现目标自动识别。测试结果表明:所提方法的整体错误识别数量为113个,整体未 识别数量为107个,证明了所提方法具有较优的自动化识别效果。

关键词: PyTorch框架; 复杂场景; 目标自动化识别; Ring Allreduce 算法; Scatter Reduce 操作; AllGather 操作

0 引言

目标识别是指在图像或视频中找出感兴趣 的目标并进行分类,而目标分割则是将图像或 视频中的目标从背景中分离出来,形成准确的 目标边界。在杂乱的场景中,目标可能被其他 物体、阴影、噪声等干扰,使得识别和分割变 得更加困难。因此,复杂场景目标识别的研究 目的是提高在复杂场景下目标识别和分割的准 确性和稳定性^[1]。

随着信息时代的发展,图像信息成为人们 获取外界信息的重要来源之一。人类通过视觉 获取的信息量占比很高,因此计算机视觉领域 的研究逐渐受到重视。目标识别和分割作为计 算机视觉领域的重要分支,具有广泛的应用价 值。例如,在智能交通系统中,可以通过识别 车辆和行人来优化交通流,提高交通安全;在 安防领域,可以通过识别异常行为来及时发现 和处理安全问题;在医学领域,可以通过识别 病变组织和器官来帮助医生进行精准诊断和治 疗。因此,复杂场景目标识别的研究对于推动 计算机视觉领域的发展和提高实际应用价值具 有重要意义。

相关学者在这一领域不断取得进展:例如, 文献[2]提出了一种以改进ResNeXt50残差网络 为基础的目标自动化识别方法,其充分考虑了 背景特征与目标特征之间的交叉关系,在对 ResNeXt50残差网络进行适应性改进之后,实现 对目标特征的深度学习,以此实现对目标的精 准识别,并降低了识别的漏检率。文献[3]提出 了一种以改进 CvcleGAN 和注意力增强迁移学习 为基础的目标自动化识别方法,该方法以图像 自身的属性特征为基础对 CvcleGAN 进行适应性 改进后,在具体的目标识别过程中引入了注意 力增强迁移学习机制,以此提高了识别结果的 可靠性。结合上述对复杂场景目标自动化识别 方法的分析可以发现,虽然已经取得了一定的 成果,但是其均是以PyTorch框架自主学习为基 础,仍存在进一步提升的空间[4]。主要是当前 方法受困于深度学习领域在复杂场景目标识别 上的缺陷, 识别框架会陷入超规模数据同步困 难和规约过多的问题,这使得在 PyTorch 框架

基金项目: 2023年度安徽省高校自然科学研究重点项目(2023AH052757)

作者简介: *通信作者:张进军(1979—),男,安徽泗县人,本科学士,讲师,主要研究方向为计算机教育、数据挖掘、数据分析,E-mail:a19855182628@126.com

收稿日期: 2023-10-31 修稿日期: 2024-01-10

上,研究复杂场景下支持深度学习的目标识别 方法,成为当下无比紧迫的任务。

本文以PyTorch框架为基础,提出复杂场景 目标自动化识别研究,并通过对比测试的方式 验证了设计方法的实际应用效果。

1 复杂场景目标自动化识别方法设计

1.1 PyTorch框架优化设计

为了保障 PyTorch 框架能够最大限度适应 复杂场景下的目标自动化识别需求,本文首先 为 PyTorch 框架构建了优化机制。对于原始的 PyTorch 框架而言,在数据并行计算的过程中, 随着每一次训练迭代的进行,对于分配给不同 节点的数据,每个运算节点都会利用其持有的模 型副本进行独立计算^[5-6],即所有运算节点在依照 模型结构进行正向和反向计算时,输出的结果包 括网络参数梯度和中间数据梯度两部分^[7]。而实 际上,对于复杂场景下的目标自动化识别需求而 言,上述过程极容易出现网络参数梯度超规模的 情况,此时需要对 PyTorch 框架中的超规模数据 进行同步处理,同时进行规约^[8]。

在 PyTorch 框架自适应优化过程中,本文借助 MPI提供的 MPI_Allreduce 算法作为 PyTorch 框架的规约函数,并将神威国产超算平台引入到 MPI_Allreduce 算法的环形通信结构 RinglIreduce 中。具体的处理过程除了最基础的初始化处理 外,本文借助 Scatter Reduce 优化 PyTorch 框架中 节点间的交换数据机制^[9-10],借助 AllGather 对 PyTorch 框架中节点建立数据块交换机制^[11]。具体的实现步骤如下:

步骤1:初步划分PyTorch框架节点中的数据,使其以N个较小的块形式存在,需要注意的是,此时的数据块划分数量与MPI_Allreduce 算法的环形通信结构中节点数量一致。

步骤2:对于完成初步划分的PyTorch框架 节点数据块,对其进行Scatter Reduce,具体的 执行次数比PyTorch框架节点数据块数量少一 次。按照上述方式,上一个节点发送的不同的 数据块会在每次操作中汇集至当前执行Scatter Reduce操作的节点^[12]。其中,图1为Scatter Reduce操作的具体执行流程示意图。



图 1 Scatter Reduce 操作具体执行流程示意图

步骤 3: PyTorch 框架中的所有节点均完成 Scatter Reduce 操作后,再以 PyTorch 框架中的初 始节点为起点,逐个对其进行 AllGather 操作, 此时节点执行的动作包括数据发送、数据接收、 数据覆盖,当 PyTorch 框架中的所有节点数据均 完成更新后,结束迭代。

按照上述所示的方式,结合复杂场景下的 目标自动化识别需求实现对PyTorch框架自适应 优化处理,其中,Scatter Reduce 操作的主要作 用是在PyTorch框架节点数据块与数据块交换 前,对具体的节点状态数据进行相加操作,通 过上述方式进一步提升PyTorch框架对复杂场景 下的目标特征计算分析的性能。AllGather_ring 操作主要作用是在PyTorch框架节点中对重复数 据进行覆盖处理,以此降低复杂场景下的目标 特征匹配阶段的重复计算量。

1.2 优化 PyTorch 框架下复杂场景目标识别 方法设计

复杂场景中存在大量的目标物体,它们可能 以不同的尺度、姿态、光照条件等多样化的方式 出现。借助1.1部分构建的优化PyTorch框架,使 识别方法更好地适应上述复杂变化,并实现更准 确的识别。因此,本文在开展复杂场景目标自动 化识别时,主要是根据目标特征实现的。

首先,本文考虑到复杂场景下的目标特征 与背景特征之间的交叉性,构建PyTorch框架中 增强多尺度特征层输出的目标特征之间的关联。 在具体的执行过程中,借助了反卷积特征融合 操作和残差融合操作的优势,以此最大限度保 障在不丢失PyTorch框架中增强多尺度特征层输 出结果的前提下,实现对目标多尺度特征的深度 融合。其中,具体的特征融合方式可以表示为

$$X_i = Concat \left| b_i(x_i) \right| \tag{1}$$

其中: X_i表示 PyTorch 框架中增强多尺度特征层 输出目标特征的深度融合结果, Concat 表示目 标特征的关联操作函数, b_i表示 PyTorch 框架中 增强多尺度特征层构成的卷积组合, x_i表示复杂 场景下目标的多尺度特征参量。需要特别注意 的是,一般情况下,复杂场景下目标的多尺度 特征参量为恒值,但是受背景以及图像自身质 量的影响,不同维度下的特征参量可能会出现 不同程度的波动,此时以高频特征作为计算过 程中的执行参量。

在此基础上,对于目标的识别问题就可以 转化为对复杂场景下目标融合特征的匹配问题, 具体的识别方式可以表示为

$$L(x, c, l, g) = \max \frac{L_{\text{class}}(x, c) + aL_{\text{loc}}(x, l, g)}{N} \rightarrow X_i$$
(2)

其中: L(x,c,l,g)表示复杂场景下目标融合特征的匹配结果, x表示输入到 PyTorch 框架中的复杂场景图像信息,c表示识别目标类别信息, l表示 PyTorch 框架中的特征候选框,g表示真实的特征参数,L_{class}表示复杂场景下目标特征的分 类损失误差,L_{loc}表示复杂场景下目标特征的位 置损失误差,a表示权重系数,N表示识别目标 类别信息总数。

按照上述方式,实现复杂场景下目标的自 动化识别,最大限度提高识别的可靠性。实现 的源代码如下:

import torch

import torch.nn as nn

```
class MultiScaleFeatureFusion(nn.Module):
    def __init__(self):
    super(MultiScaleFeatureFusion, self).__init__()
```

```
self.conv_blocks = nn.Sequential(
```

```
#这里添加卷积组合:
```

```
nn.Conv2d(in_channels=3, out_channels=64,
```

```
kernel_size=3, stride=1, padding=1),
```

```
nn.ReLU(inplace=True),
```

#添加更多的卷积层...

```
)
```

self.deconv_blocks = nn.Sequential(

#这里添加反卷积组合如:

```
nn.ConvTranspose2d(in_channels=64, out_channels=
32, kernel_size=3, stride=2, padding=1, output_
padding=1),
```

```
nn.ReLU(inplace=True),
```

#添加更多的反卷积层...

```
)
```

self.residual_blocks = nn.Sequential(#这里添加残差组合: nn.Conv2d(in_channels=32, out_channels=32,

kernel_size=3, stride=1, padding=1),

nn.ReLU(inplace=True),

#添加更多的残差层...

)

)

self.classifier = nn.Sequential(

nn.Linear(in_features=32, out_features=10),

```
#这里的输出特征数量需要根据实际需求调整
```

```
def forward(self, x):
x = self.conv_blocks(x)
x = self.deconv_blocks(x)
x = self.residual_blocks(x)
x = x.view(x.size(0), -1) # flatten the tensor before
classification
x = self.classifier(x)
return x
# 实例化模型并进行训练...
model = MultiScaleFeatureFusion()
```

对于损失函数的部分,可以根据实际需求自 定义损失函数。可以使用PyTorch的内置损失函 数,如nn.CrossEntropyLoss()进行分类损失的计 算,nn.MSELoss()进行位置损失的计算。然后,

可以将这两个损失结合起来,形成一个总的损失,如下所示:

| criterion = nn.CrossEntropyLoss() | #分类损失函数 |
|--|----------|
| criterion_loc = nn.MSELoss() | #位置损失函数 |
| # 在训练循环中 | |
| <pre>outputs = model(inputs)</pre> | |
| loss_class = criterion(outputs, labels) # 计算分类损失 | |
| <pre>loss_loc = criterion_loc(outputs_loc, labels_loc)</pre> | |
| | # 计算位置损失 |
| $loss = loss_class + loss_loc$ | # 计算总损失 |
| | |

2 应用测试

2.1 测试设置

在分析本文设计的基于 PyTorch 框架的复杂 场景目标自动化识别方法实际应用效果时,设 定参与测试的对照组分别为文献[2]提出的以改
进 ResNeXt50 残差网络为基础的目标自动化识 别方法,以及文献[3]提出的以改进 CycleGAN 和注意力增强迁移学习为基础的目标自动化识 别方法。在具体的测试过程中,使用的数据集主 要包括两个来源,分别为美国国防高级研究计划 署 (Defense Advanced Research Project Agency, DARPA)提供的 MSTAR (Moving and Stationary Target Acquisition and Recognition)图像,以及空 军研究室 (Air Force Research Laboratory, AFRL) 提供的 MSTAR 图像。

为了保证实验准确性,首先对上述数据集 图像进行预处理。本文对MSTAR数据中的灰度 图像进行转化,使其以伪彩色图像的形式存在。 结合MSTAR数据自身的属性特征,本文从提升 训练样本信息完善程度和样本丰富程度的角度 出发,采用不同角度的旋转方式对训练样本集 进行预处理。对于每一幅原始测试图片,对应 的旋转角度分别为90°、180°和270°,选取中心 大小为64×64的区域的下采样结果作为最终的 识别对象。其中,具体的测试MSTAR数据集设 置情况见表1。

| 表 1 | 测试MSTAR数据集设置情况 |
|-----|----------------|
| | |

22400

35600

30700

34500

测试图像原始大小

32×32

32×32

32X32

32×32

训练数据规模 测试数据规模

4560×4

2260×4

3020×4

2160X/

| | 2100.4 | | 5252 | |
|------|--------|----------------|------------------------|-----|
| 信号灯 | 3390×4 | 26100 | 32×32 | |
| 指示牌 | 2560×4 | 21200 | 32×32 | |
| 自行车 | 2300×4 | 23600 | 32×32 | |
| 以表 | 長1所示的测 | 试数据准备 | 情况为基础, | 为 |
| | | | 1149 6 94 ± 449 | / 4 |
| 了降低车 | 交大的斑点 | 噪声对于最 约 | 冬识别结果的 | 影 |
| 响,在测 | 则试前对图值 | 象整体进行》 | 虑波处理, じ | 此 |
| 改善测试 | 武样本的质量 | 遣 。在目标⊧ | 自动化识别达 | t程 |
| 中,通常 | 常会进行预算 | 业理。数据 到 | 页处理后, 通 | 自常 |
| 色彩的音 | 移分损失对目 | 目标识别结果 | 是造成影响不过 | 大, |
| 但是需要 | 要注意某些颜 | 领色在彩色植 | 莫式下与背景 | ł的 |
| | | | | |

对比明显,而转灰度后反而不明显的情况,结



(a)RGB模式

图 2本文方法自动化识别测试

(b)灰度模式

由图2可知,本文所提方法在复杂场景目标 自动化识别中,受背景影响较小,可获取较为 精准的识别效果。

2.2 测试结果对比与分析

针对复杂场景图像中的目标识别问题,现 采用改进 ResNeXt50 残差网络识别方法、改进 CycleGAN 和注意力增强迁移学习识别方法和本 文设计识别方法对图 3(a)初始图像展开识别处 理。待识别目标为一辆白色 SUV,目标识别测 试结果如图 3(b)~(d)所示。







(e) 改进 ResNeXt50 残差 网络识别方法

(d) 改进 CycleGAN 和注意力 增强迁移学习识别方法

图 3 三种方法的识别效果

根据图3所呈现的测试结果可以发现,所提 方法相比于其他两种方法的目标识别效果更好, 边缘更清晰,表明所提方法对复杂场景图像的 目标识别具有较高的有效性。

在上述测试环境的基础上,对不同方法的 识别结果进行分析时,本文分别将错误识别目 标数量以及未识别目标数量作为衡量目标自动 化识别效果的评价指标。其中,不同方法的具 体测试结果见表2。

目标

大巴

行人

吊车

垃圾桶

果如图2所示。

表 2 不同方法目标自动化识别效果对比

单位:个

| 改进ResNeXt50残差网络识别方法 | | 改进CycleGAN和注意力增强迁移学习识别方法 | | 本文设计识别方法 | | |
|---------------------|--------|--------------------------|--------|----------|--------|-------|
| 日1小 - | 错误识别数量 | 未识别数量 | 错误识别数量 | 未识别数量 | 错误识别数量 | 未识别数量 |
| 大巴 | 37 | 23 | 30 | 46 | 18 | 13 |
| 行人 | 40 | 24 | 30 | 49 | 15 | 16 |
| 吊车 | 41 | 21 | 31 | 53 | 17 | 16 |
| 垃圾桶 | 39 | 25 | 32 | 46 | 16 | 17 |
| 信号灯 | 41 | 25 | 30 | 49 | 14 | 15 |
| 指示牌 | 45 | 20 | 29 | 52 | 18 | 16 |
| 自行车 | 47 | 28 | 31 | 55 | 15 | 14 |
| 合计 | 290 | 166 | 213 | 350 | 113 | 107 |

结合表2所示的测试结果对三种不同目标自 动化识别方法的性能进行分析,其中,在改进 ResNeXt50残差网络识别方法下,对于目标的错 误识别数量较多,整体错误识别数量达到了290 个,自行车错误识别数量最高,达到了47个, 错误识别数量最低也达到了37个(大巴),对应 的未识别数量相对偏低,对于不同类型目标的 未识别数量均在30个以下,整体未识别数量为 166个。在改进CycleGAN 和注意力增强迁移学 习识别方法的测试结果中,对于目标的未识别 数量较多,整体未识别数量达到了350个,未识 别数量最高的目标类型同样为自行车,对应的 数量达到了55个,未识别数量最低的目标类型 为大巴,但是也达到了46个,对应的错误识别 数量与之相对有所降低,对于不同类型目标的 错误识别数量稳定在29~32个之间,整体错误 识别数量为213个。相比之下,在本文设计方法 的测试结果中,整体错误识别数量仅为113个, 分别低于改进ResNeXt50残差网络识别方法177 个,低于改进CycleGAN 和注意力增强迁移学习 识别方法100个;整体未识别数量仅为107个, 分别低于改进ResNeXt50残差网络识别方法59 个,低于改进CycleGAN 和注意力增强迁移学习 识别方法243个。以具体的识别对象为基础对本 文设计方法的识别效果进行分析,其中,错误 识别数量最高仅为18个(指示牌和大巴),未识 别数量最高也仅为17个(垃圾桶)。

综合上述测试结果可以得出结论,本文设 计的以PyTorch框架为基础的复杂场景目标自动 化识别方法可以实现对目标的有效识别。

3 结语

面对越来越复杂的场景环境,目标识别面 临着越来越高的挑战,如何提高目标识别的可 靠性成为了进一步提高图像深度利用与信息挖 掘的重要基础。为此,本文提出PyTorch框架下 的复杂场景目标自动化识别研究,切实实现了 对不同场景下目标对象的有效识别,降低了识 别结果的漏检率和误检率,具有良好的实际应 用价值。借助本文对于目标自动化识别方法的 研究与设计,希望能够为实际的复杂场景图像 信息挖掘与利用提供有价值的帮助。

参考文献:

- [1] 李明珏,张高兴.基于YOLO目标检测的边缘到终端的道路损伤检测系统设计[J].机电工程技术, 2023,52(9):291-295.
- [2] 王军龙,宣魁,熊海涛,等.基于改进ResNeXt50残
 差网络的锦鲤选美方法[J].农业机械学报,2023, 54(S1):330-337.
- [3] 刘世晶,刘洋春,钱程,等.基于改进CycleGAN和 注意力增强迁移学习的小样本鱼类识别[J].农业 机械学报,2023,54(S1):296-302.
- [4] 周天,司吉坤,杜伟东,等.采用GAF-D3Net深度学 习网络的水下目标有源识别方法[J]. 声学学报, 2023,48(5):950-958.
- [5] 李月琴,张红莉,张维,等.基于改进CNN的 HRRP目标识别方法[J]. 兵器装备工程学报, 2022,43(8):265-274.
- [6] 罗毅智,陆华忠,周星星,等.基于YOLOV5-MobilenetV3和声呐图像的鱼类识别轻量化模型[J]. 广东农业科学,2023,50(7):37-46,36.
- [7] 于雅俐,熊涛,康俊,等.基于各向异性二维半导体

GeSe的双波段实时目标识别图像传感器(英文) [J]. Science Bulletin, 2023, 68(17):1867-1870.

- [8] 汪泽,张朝阳,尹健.基于魂芯五号HXAI100的 YOLOV3模型目标识别应用实现[J].中国集成电路,2023,32(9);15-19.
- [9] 郑伟,于洋,刘砚菊.骨架形状特征的目标识别算法
 [J].沈阳理工大学学报,2022,41(1):14-19.
- [10] 简泽明,赵旭辉,胡君豪,等.基于改进 Shuffle-

NetV2模型的声目标识别方法研究[J]. 传感器与微系统,2023,42(8):43-45,49.

- [11] 姚璐,韩磊,杨磊,等.用于HRRP多类目标识别的 D距离分类器[J].北京理工大学学报,2022,42 (11):1144-1149.
- [12] 孙亚杰,梁轲,马孟新,等. 弹体侵彻多层混凝土靶板的引信层目标识别方法研究[J]. 测控技术, 2023,42(7):95-101.

Research on object recognition methods in complex scenes under the PyTorch framework

Zhang Jinjun^{*}

(Department of Information Management, Anhui Vocational College of Police Officers, Hefei 230031, China)

Abstract: The artificial intelligence framework (PyTorch+OneDNN) is prone to the problem of network parameter gradient superscale in complex scene target recognition. To this end, a complex scene object recognition method based on the PyTorch framework is proposed. Introducing the Ring Allreduce algorithm in MPI, optimizing the PyTorch framework to achieve super scale data synchronization and reduction processing in the iterative feature extraction process of complex scenes. Based on the optimized Py-Torch framework, considering the intersection between complex scene target features and background features, we construct an enhanced correlation between target features output by multi-scale feature layers. By leveraging the advantages of deconvolution feature fusion and residual fusion operations, and based on the above correlation, automatic target recognition can be achieved. The test results show that the overall number of errors identified by the proposed method is 113, and the overall number of unrecognized errors is 107, proving that the proposed method has better automation recognition performance.

Keywords: PyTorch framework; complex scene; target automation identification; Ring Allreduce algorithm; Scatter Reduce operation; AllGather operation

(上接第65页)

Research on student classroom behavior recognition based on YOLOv5 algorithm

Ma Ruicheng*, Chen Ji, Wang Binghuai, Long Juncheng, Liu Yu

(School of Electronic Information, Southwest Minzu University, Chengdu 610225, China)

Abstract: With the popularization of smart classrooms and educational data mining, our project aims to create an automatic detection and analysis method for student learning data in a smart classroom. Based on the YOLOv5 model, two types of learning data are implemented: classmate counting and learning behavior recognition. This model recognizes seven types of student classroom behaviors, which assists teachers in judging student learning situations and making pedagogical decisions, including writing with head down, reading with head down, listening with head up, turning, raising hand, standing, and group discussion. Research shows that the detection accuracy of the model reaches 97.921%.

Keywords: student behavior recognition; YOLOv5; computer visual technology

文章编号:1007-1423(2024)08-0072-05

DOI: 10.3969/j.issn.1007-1423.2024.08.012

一种轻量化的排球自垒姿态检测算法

凌 驯*, 陶青川

(四川大学电子信息学院,成都 610065)

摘要:边缘设备有限的存储和处理能力,在实际应用中难以部署较为复杂的YOLOv7pose姿态检测模型。对YOLOv7pose进行了一系列轻量化处理,使用FasterNet的主干网络重构YOLOv7pose的特征提取网络,将特征提取后的输出应用CBAM注意力机制来弥补精度上的损失,最后对冗余的多尺度检测头进行删减,实验表明改进后的轻量化网络较原网络的参数量下降了2/3,计算速度提升了2.5倍,精度仅减少了3.8%,能够满足边缘设备实时检测排球对墙自垒过程中的人体姿态情况。

关键词: 姿态估计; 轻量化; 边缘设备; YOLOv7pose

0 引言

排球作为一种受众广泛的团队运动,在全 世界范围内有着众多参与者和巨大的影响力^[1], 与足球、篮球一起并称为"三大球"。对墙自垒 是锻炼排球接发球基本技术动作的最佳方式, 通过将球对着墙进行回球来改进接球和传球技 能。如今,排球对墙自垒已经入选全国中学体 育考试的球类选考项目。然而在练习过程中, 练习者无法对自身的姿势进行观测,因此难以 发现和复盘一些错误或不标准的动作姿势。计 算机视觉姿态估计的发展,为解决这样的问题 提供了方案。

计算机视觉领域中的姿态估计是指从图像 或者视频中确定人体关节点的空间位置和姿态 的过程。按照实际应用中检测的人体数目分为 单人姿态估计和多人姿态估计^[2]。

在单人姿态估计中,待检测的目标唯一, 需要先确定目标框的位置,然后利用坐标回归 或者热图的方式来预测出所有关节点坐标信息。 DeepPose^[3]作为典型的回归方式,利用 CNN 提 取特征图之后对关节点坐标进行高质量回归, 这种非线性的映射方式缺乏一定的鲁棒性;而 利用热力图的方式,如 RePose^[4]、UniPose^[5]等 基于统计概率,能够获得对应关节点的概率分 布和位置信息,从而在精度上表现更好,但相 应的网络结构也更加复杂。

在多人姿态估计中,由于待检测的目标不 止一个,因此包括定位和检测阶段。按照这两 个阶段的执行先后顺序可以分为自顶向下和自 底向上的方法。以CPM^[6]、HRNet^[7]为代表的自 顶向下的方式,对输入图像进行人体目标检测 后再对每一个人体框进行单人姿态检测;与之 相反,以OpenPose^[8]为代表的自底向上的方式, 首先检测出图像中的所有关节点,然后依照一 定的策略进行聚类,从而将属于同一人体关节 点结合在一起。

1 轻量化的姿态检测网络

1.1 YOLOv7pose网络简介

无论是自顶向下还是自底向上的方式都不 是端到端的检测方式,作为YOLOv7系列中的姿态估计分支,YOLOv7pose是一个真正的端到端 网络,能在检测出人体检测框的同时预测出对 应人体的关节点信息^[9],因此克服了两阶段检 测速度方面的不足。

YOLOv7pose主要由Backbone、Neck和Head 三部分组成,如图1所示。

收稿日期: 2023-11-03 修稿日期: 2023-11-29

作者简介: *通信作者:凌驯(1998—),男,四川眉山人,硕士,研究方向为模式识别与智能系统,E-mail:892469973@ qq.com; 陶青川(1972—),男,四川南充人,硕士生导师,副教授,研究方向为模式识别与智能系统



图 1 YOLOv7pose 结构图

输入的是 3×640×640 的 RGB 图像, 经过 Backbone 网络进行特征提取。Backbone 主要由若 干相邻的CBS模块和ELAN模块构成。CBS模块对 特征图大小减半, 通道数加倍, ELAN如图2(a)、 ELAN-W如图2(b)所示, 是高效的层聚合模块, 通过设计最短梯度路径和最长梯度路径来保证 信息的有效传播。



经过Backbone进行特征提取后得到了图像的高维度语义信息,Neck中的SPPCSPC能匹配不同分辨率的图像,此外,YOLOv7pose的Neck借鉴了PANet¹⁰⁰的方式,进行高低语义信息融合,大大减少了信息传递的层数。最终得到256×80×80、512×40×40、768×20×20、1024×10×10四种尺寸的特征图,用于小目标、中目标、较大目标和大目标的检测。将这四个特征图输入到检测头中,同时预测出人体框和17个二维关节点的坐标信息。

1.2 FasterNet Block介绍

在部署设备算力和存储空间有限的情况下,

一个具有低延时、高吞吐量的神经网络能够给用 户带来更好的体验。定义神经网络从输入得到输 出的时间间隔为*Delay*,网络模型的计算复杂度, 即总浮点运算次数为*FLOPS*,网络每秒进行的浮 点运算次数为*fs*,三者的关系如公式(1)所示:

$$Delay = \frac{FLOPS}{fs} \tag{1}$$

用深度可分离卷积来代替常规卷积提取图像特征被应用于轻量化网络,如MobileNetV3^[11]、ShuffleNetV2^[12]、GhostNet^[13]等,从而具有更低的参数量和更少的网络FLOPS,然而在实际中这些结构中的级联、shuffle、池化等额外的数据操作,以及内存访问增加的副作用,使得网络的*fs*减小,最终这些轻量化的网络表现得不够"快速",不利于实时性要求较高的应用场景。

FasterNet 中提出了 PConv 的概念^[14]。在深 度可分离卷积的 DWConv 中,需要在 DWConv 后 紧跟着 PWConv,进行通道维度的扩充来补偿精 度的下降。假设输入的图像形状为 *C×H×W*,卷 积核的大小为*K*,输出图像形状不变,常规卷积 的计算量和内存访问量如公式(2)、(3)所示:

$FLOPs = H \times W \times K^2 \times C^2 \tag{2}$

 $MACs = H \times W \times 2C + K^2 \times C^2 \approx H \times W \times 2C \qquad (3)$

在深度可分离卷积中,由于会进行输出通道的扩展,即*C*变大造成了额外的内存访问*MACs*增加,实际*Delay*不一定降低。Pconv如图3(a)所示,仅仅对输入通道中第一个或者最后一个连续 C_p 个通道上应用常规卷积操作,而剩余输入通道保持不变,同时减小了冗余计算和内存访问。一般情况,当 $C_p = C/4$ 时,FLOPS只有常规卷积的1/16。通过在PConv卷积后附加2个1×1的PWconv层,先提升通道数后再降低,最后短接输入构成倒残差结构,得到FasterNet Block如图3(b)所示。



利用高效的网络结构,无论是在 GPU、 CPU 还是 ARM 处理器上,FasterNet Block 都具 有较低的网络延迟和较高的 fs。

1.3 CBAM注意力机制

目前,注意力机制已经成为图像分类、目标检测、图像分割等深度学习领域的重要工具, 注意力机制能够强化关节数据,减少冗余信息, 提高网络的特征提取效率,增强模型的表征和 泛化能力,改善网络的性能。CBAM 是一种轻 量化的注意力模块^[15],能同时融合特征通道和 特征空间的信息,如图4所示。



图 4 CBAM 注意力机制

CAM模块如图5所示,输入特征图形状为 C×H×W,经池化后得到两个C×1×1的特征张量, 接着通过一个共享的两层MLP,对通道数先减 后增,使得通道数前后保持不变,把经过MLP 后的两部分输出进行元素叠加,再经sigmoid激 活,赋予重要通道中元素更高的权重。



图 5 CAM 模块

SAM模块如图6所示,首先作用通道层次的全局和平均池化,得到两个1×H×W的特征张量,并将这两个特征作通道维度的拼接形成2×H×W的权重张量,通过卷积融合这两部分特征,并降维为1×H×W,最后再经sigmoid激活,赋予同一通道中重要位置元素更大的权重。





CBAM模块将注意力集中于输入数据中特 定通道中的特定区域,从而强化了重要的特征, 减少不相关信息的干扰,提高了模型的精度。

1.4 改进后的轻量化姿态检测网络

综上,改进的轻量化姿态检测网络如图7所 示,Backbone部分参考FasterNet四个stage的设 计,分别堆叠1,2,8,2个基础FasterNetBlock, 在每个stage之前,利用CBS使特征图长宽减半, 通道增加;在Backbone和Head之前加入了 CBAM注意力机制以提高模型检测的准确性; 此外,考虑到在排球对墙自垒的实际应用中, 希望检测全身人像的17个关节点,因此所检测 人体目标应该大部分属于中等大小,所以删去 原网络中的大型目标检测头,既符合实际应用 又减少了网络参数。



图 7 改进后的姿态检测网络

2 实验与分析

2.1 数据集准备

通过在网上爬取不同场景、不同视角下, 身形不同的训练者对墙垒球的视频,对视频进 行间隔抽帧得到原始图片,之后进行翻转对称、 剪切、拼接、马赛克等数据增强,共6686张图 片,用labelme软件进行人体17个关节点和人体 检测框的标注,最终用脚本转化为txt格式,按 7:2:1的比例划分训练集、验证集和测试集。

2.2 实验平台和评价指标

本实验训练和测试的PC平台软硬件参数见表1。

表 1 实验 PC 平台参数

| 软硬件 | 即 置信自 |
|--------------|---|
| | |
| 处理器 | 15 vCPU AMD EPYC 7542 32-Core Processor |
| 内存 | 80 GB |
| 操作系统 | ubuntu18.04 |
| GPU | RTX 3090(24 GB) |
| Python | 3.8.18 |
| torch | 1.10.0+cu111 |
| torch vision | 0.11.0+cu111 |

实际部署的边缘设备采用比特大陆的 SE5 设备,该设备搭载了基于第三代的 TPU 芯片 BM1684,峰值算力可达 17.6 TOPS,内存 12 GB, ARM 架构 8核 A53 处理器,主频 2.3 GHz。

考虑到人体大小的差异,采用OKS指标来 度量检测对象和目标对象关节点的相似性,如 公式(4)所示:

$$OKS = \sum_{i} \frac{\exp\left(-\frac{d_{i}^{2}}{2s^{2}\sigma_{i}^{2}}\right)\delta(v_{i} > 0)}{\sum_{i}\delta(v_{i} > 0)}$$
(4)

其中: d_i 是检测关节点和对应真值之间的欧氏 距离; s^2 等于该人体框 groundtruth 所占的面积; σ_i 为第i个关节点的归一化因子,表征该关节点 的标注难度; v_i 取0,1,2,分别表示该关节点 存在,存在但被遮挡,存在且可见。

根据OKS设定不同阈值(0.5~0.95),可计 算出对应超过阈值的OKS数目占总数目的比 例,最后将这些比例求平均得到最终的评价指标mAP。

2.3 实验结果及分析

表2展示了用典型的轻量化网络主干替换 YOLOv7pose主干,并删去一个冗余检测头后模 型的大小、计算量、检测精度和在PC平台GPU 上的推理速度。

| 表 2 | 2 不 | 同主干 | 网络的 | 实验结果 | 对比 |
|-----|-----|-----|-----|------|----|
|-----|-----|-----|-----|------|----|

| 主千 | 模型大小/M | 计算量/GFLOPS | mAP/% | Fps |
|--------------|--------|------------|-------|------|
| YOLOv7pose | 161.2 | 102.2 | 96.0 | 51.7 |
| MobileNetV3 | 53.4 | 53.0 | 91.9 | 62.5 |
| ShuffelNetV2 | 49.1 | 46.6 | 92.3 | 66.7 |
| GhostNet | 53.2 | 51.6 | 92.1 | 58.8 |
| FasterNet | 58.4 | 59.0 | 91.3 | 85.7 |

表2结果表明,用FasterNet Block 重构的网络模型虽然损失了一定精度,但参数量较低且 具有更低的网络延时,更符合边缘设备实时检 测的场景。在此重构化网络的基础上,为进一 步提升检测精度,添加CBAM注意力模块并进 行了消融实验,结果见表3。

表 3 CBAM 添加位置对比

| CBAM模块添加位置 | mAP/% | 模型大小/M | Fps |
|-----------------|-------|--------|------|
| 未添加 | 91.3 | 58.4 | 85.7 |
| Backbone和Neck之间 | 92.2 | 58.5 | 85.5 |
| Neck 和 Head 之间 | 91.6 | 58.8 | 85.1 |

由此可得,将CBAM模块置于Backbone和Neck之间,在带来很少参数量增加的情况下, mAP提高了0.9个百分点。最后,将原YOLOv7pose 模型和最终的轻量化模型部署在SE5边缘设备 上,推理速度和检测结果见表4。

表 4 实际部署推理速度

| 网络模型 | Fps |
|------------|------|
| YOLOv7pose | 19.3 |
| 轻量化姿态检测网络 | 68.4 |



图 8 实际部署检测效果

3 结语

本文针对边缘设备实时检测排球自垒时的 人体姿态,提出了一种轻量化的姿态检测网络, 使用 FasterNet 中的 Block 对 YOLOv7pose 的 Backbone 进行重构,并在 Backbone 和 Neck 之间加入 了 CBAM 注意力机制。实验表明,在保证一定 精度的条件下,改进后的网络参数量大幅下降, 推理速度增加,能较好地应对实时检测的场景, 具有实用价值。

参考文献:

 杨珊珊,古松.我国群众排球发展现状综述和可持 续发展对策[J].运动,2015(22):146-147.

- [2] 王珂,陈启腾,陈伟,等.基于深度学习的二维人体 姿态估计综述[J/OL]. 郑州大学学报(理学版):
 1-10 [2023-11-02]. https://doi. org/10.13705/j.
 issn.1671-6841.2022334.
- [3] TOSHEV A, SZEGEDY C. Deeppose: human pose estimation via deep neural networks [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, Ohio, USA: IEEE, 2014: 1653-1660.
- [4] ISACK H, HAENE C, KESKIN C, et al. RePose: learning deep kinematic priors for fast human pose estimation[EB/OL]. arXiv:2002.03933,2020.
- [5] ARTACHO B, SAVAKIS A. UniPose: unified human pose estimation in single images and videos [C]
 //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020:7033-7042.
- [6] WEI S E, RAMAKRISHNA V, KANADE T, et al. Convolutional Pose Machines [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 4724-4732.
- [7] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation
 [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2019: 5686-5696.
- [8] CAO Z, HIDALGO G, SIMON T, et al. OpenPose: realtime multi-person 2D pose estimation using part affinity fields[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(1):172-186.
- [9] MAJI D, NAGORI S, MATHEW M, et al. YOLO-pose: enhancing YOLO for multi person

pose estimation using object keypoint similarity loss [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). New Orleans, LA, USA: IEEE, 2022:2636-2645.

- [10] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, UT, USA:IEEE,2018:8759-8768.
- [11] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetV3 [EB/OL]. arXiv: 1905.02244, 2019.
- [12] MA N, ZHANG X, ZHENG H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design [C] // Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: Springer, Cham, 2018.
- [13] HAN K, WANG Y, TIAN Q, et al. Ghostnet: more features from cheap operations [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020:1577-1586.
- [14] CHEN J ,KAO S H ,HE H ,et al. Run,don't walk: chasing higher FLOPS for faster neural networks [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver,BC,Canada:IEEE,2023:12021-12031.
- [15] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C] //Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: Springer, Cham, 2018: 3-19.

A lightweight posture detection algorithm for volleyball self-digging

Ling Xun*, Tao Qingchuan

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

Abstract: Due to the limited storage and processing capabilities of edge devices, it is difficult to deploy a more complex YO-LOv7pose pose detection model in practical applications. The article conducted a series of lightweight processing on YOLOv7pose, using the backbone network of FasterNet to reconstruct the feature extraction network of YOLOv7pose. The output of the feature extraction was applied with a CBAM module to compensate for the loss of accuracy. Finally, redundant multi-scale detection heads were deleted. Experiments have shown that the improved lightweight network reduces the parameter quantity by 2/3 compared to the original network, increases the computing speed by 2.5 times, and reduces the accuracy by only 3.8%. It can meet the real-time detection of human posture during the process of volleyball self-digging by edge devices.

Keywords: pose estimation; lightweight; edge devices; YOLOv7pose

实践与经验

文章编号:1007-1423(2024)08-0077-07

DOI: 10.3969/j.issn.1007-1423.2024.08.013

基于主动学习的实体关系抽取的方法研究

孙 涵*

(太原开放大学网络服务中心,太原 030027)

摘要:关系分类是NLP中提取实体间关系的一项重要任务。介绍一种用于大规模的中文信息抽取数据集的方法, 该方法将BERT合并到一个新的框架,并将主动学习应用于联合实体关系抽取中。这种模型从四个方面完善了现有的方 法。第一,可以解决多个实体属于多个三元组的问题。基于概率图的思想设计了该框架,并研究出一种新的"头尾" 标记方法;第二,提出了一种将主动学习应用于关系抽取问题的创新方法;第三,为了在主、谓、宾三种实体之间传 输信息,提出了一种新的规范化方法,称为条件层规范化;第四,设计了一个新的损失函数,以避免类不平衡。实验 证明,增强了模型的信息提取能力,在单个模型的测试集上的F1-score达到0.840,在用完整数据训练的情况下与原始 深度模型相比,用更少的数据取得了更好的性能。

关键词: BERT; 主动学习; 联合实体关系抽取

0 引言

实体和关系抽取的任务是发现实体及其语 义关系。这是信息提取和知识库建设的中心问题。给定一个句子和一个预定义的关系集,目 的是从句子中识别出所有正确的关系实体三元 组^[1]。其中三元组包含两个实体(主语S和宾语 O)及其谓语P,例如(S:人,P:出生地,O: 地点),(S:个人,P:丈夫,O:个人)。

为了解决这个问题,传统框架采用流水线 方法^[2-3],将此任务分解为两个步骤:首先识别 实体,然后执行实体对的关系识别。这类方法 具有高度的灵活性,而且易于实现。然而,它 们忽略了实体识别和关系分类之间的相关性, 并且实体识别的效果将影响关系分类的最终结 果,容易导致错误的不断累积^[4]。为了克服上 述问题,一个解决方案就是使用单一联合模型 整合实体识别和关系提取的方法^[5-6],大多数早 期联合方法依赖于需要复杂特征工程的手工特 征。另一种解决方案是基于神经网络的方法, 该方法也广泛用于SPO提取问题^[7-8] 语言表示模型,如 BERT^[9]、GPT^[10]、Ro-BERTa^[11],已被证明有助于改善许多 NLP 存在 的问题。在所有预处理模型中,BERT 的提出引 起了极大的关注。它可以被转移到多个 NLP 任 务中,并得到在分类、序列标注等方面最好的 表现。

通常,用于标注问题的标注数据远远少于 所有的可用数据,尤其是用于实体和关系提取 的标注数据。这需要多个步骤,需要训练有素 的注释器^[12-13]。因此,获取带注释的数据变得非 常重要。在这种情况下,需要主动学习来减少 标注数据,从而实现更高的效率^[14]。主动学习 提出了一种选择标记样本的有效方法。据我们 所知,主动学习尚未应用于实体和关系提取。

这项任务也有一些困难。如图1所示,可以 注意到一个实体可能存在于多个关系中,并且 实体跨度之间存在重叠。大多数句子中标记的 三元组是"一个S,多个(P和O)"的形式,即 一个主语对应多个宾语。也有一些数据具有 "多个S,一个(P和O)""多个S,多个(P与O)"

收稿日期: 2023-11-13 修稿日期: 2023-11-27

基金项目:山西省教育科学"十四五"规划课题(GH-21105);山西省现代远程教育学会2024年度课题(SXYJ202403) 作者简介:*通信作者:孙涵(1993—),女,硕士研究生,副主任,讲师,主要研究方向为软件工程、远程教育,E-mail: s_532313026@qq.com

和"一个S,一个O,多个P"的情况。

"text":"布丹是出生于1824年的法国画家"

"spo<u>list</u>":("布丹","出生日期","1984"),("布丹","国籍","法国")

图 1 数据集例子

本文的实用主动学习算法将利用 BERT 模型来从文本中提取出 SPO 三元组,其主要贡献如下:

(1)该方法可以解决多个实体属于多个三元组的问题。基于概率图的思想设计了这个框架,并开发了一种新的"头尾"标记方法。

(2)提出了一种创新的方法,将主动学习应 用于关系抽取问题。

(3)为了在主语实体、谓语实体和宾语实体之间传输信息,提出了一种新的规范化方法,称为条件层规范化。

(4)设计了一个新的损失函数,以避免类不 平衡问题。

1 相关工作

实体关系抽取的目标是从句子中抽取出 SPO 三元组。解决该任务的两种主要方法是流 水线方法和联合学习方法。流水线方法将实体 识别和关系识别作为两个后续步骤,而联合模 型旨在同时解决这两个问题。

早期大多数研究使用流水线方法,首先识别命名实体(NER)^[15],然后对实体对之间的关系进行分类^[16]。传统的NER方法是基于线性模型的,并依赖手工特征提取,如HMM^[17]和CRF^[18]。现在,CNN和RNN等深度神经网络已成功用于NER^[19-20],并取得很好的效果。关系分类也有两种方法,分别是基于手工特征提取^[21]和神经网络^[22-23]。

联合模型^[5.24]提出可以同时提取实体和识别 关系。早期的联合学习模型是基于 feature-based 的,这些方法需要用到外部数据预处理、复杂 的特征工程和可靠的 NLP 工具。同时,需要基 于神经网络的联合模型,包括 CNN 和 RNN 来处 理这项任务。

Zheng等^[25]发现了一种新标注方案,用于处理标记任务中的实体关系提取问题。Katiyar等^[26]

提出了 RNN,并注重解决问题。但这些方法无 法处理多个关系和多个实体在同一句话中重叠 的情况。Zeng 等^[27]提出了一种具有复制机制的 端到端的神经联合模型,但其性能取决于分词 的精度。Bekoulis 等^[28]提出的联合模型使用 CRF 层来重新识别实体,并通过多头选择方法提取 关系。

语言表示模型^[29-30]已被证明有助于改善许多 NLP问题。Wu等^[31]将BERT应用于关系分类任 务,并实现了很高的效率,但该方法不能直接 提取SPO三元组。Huang等^[32]基于Bekoulis等^[28] 的框架,使用BERT来替代BiLSTM,并提出了一 个软标签嵌入层来连接两个步骤。然而,这两种 方法无法解决主语实体有重叠的情况。Xue等^[33] 将BERT语言模型集成到中医数据集的实体关系 扩展问题中。

主动学习可以减少标注数据从而实现更高的效率,因此非常适合许多自然语言处理问题^[34-35]。Settles等^[36]将主动学习和分析最佳策略应用于序列标注上。Shen等^[37]将主动学习与命名实体识别的深度学习相结合,与25%的原始培训数据相匹配。

实体关系抽取的主动学习受到的关注较少。因此,可以将主动学习算法应用于预处理BERT 模型,以提取SPO三元组。

2 实验方法

2.1 模型架构

该方法的架构如图2所示。

对于一个输入字符句,我们首先使用 BERT 方法得到包含上下文信息的编码向量。经过层 归一化处理后,采用两个密集连接分类器,得 到由两个分别标记头部和尾部位置的向量组成 的主语实体。然后以主语实体向量为条件,通 过条件层归一化对 BERT 编码序列进行归一化 处理。对于每种类型的谓词,构造一个密集层 来预测对应的对象实体的开始和结束位置,从 而同时预测对象和谓词。为了解决多个实体属 于多个三元组的问题,提供了一种"头尾"标 记方法和一个 Sigmoid 层。"头尾"标记法是指 用两个 0/1 向量分别标记实体的起始位置和结 束位置。



图 2 总体架构

2.2 主动学习

为了制作电子稿件,必须使用Adobe的便携 文档格式(PDF)。PDF文件通常是使用pdflatex命 令从LATEX生成的文件。如果用LATEX版本生 成Postscript文件,可以使用ps2pdf或dvipdf将其 转换为PDF。在Windows上,还可以使用Adobe Distiller生成PDF。为了减少标注数据,主动学习 算法必须评估标注哪个实例。设L为一个标签集 合,U为无标签池,查询策略为O(x),x为实例, 整个学习过程包括以下几个步骤:

第一步,随机初始化模型,设L为有标签 集合。

第二步,利用模型选择信息实例的前N个。

第三步,将前N个实例从U移到L。

然后进行下一轮学习,直到U中没有实例。 不确定性抽样^[38]是衡量信息量最常用的策略之 一。Shen等^[37]探索了一种基于MNLP的实体关 系提取问题的方法。选择模型中置信度最低的 实例W0on,相当于以下分数:

$$\max \prod_{i=1}^{n} P\left(y_{i} \middle| y_{i}, \cdots, y_{n-1}, \left\{X_{ij}\right\}\right) \longleftrightarrow$$
$$\max \sum_{i=1}^{n} \log P\left(y_{i} \middle| y_{i}, \cdots, y_{n-1}, \left\{X_{ij}\right\}\right) \tag{1}$$

这个分数的取得是因为选择了较长的句子, 因为这样会包含单词的总和。将置信度分数进 行归一化会得到:

$$\max \frac{1}{n} \sum_{i=1}^{n} \log P\left(y_i \middle| y_i, \cdots, y_{n-1}, \left\{X_{ij}\right\}\right) \qquad (2)$$

2.3 BERT用于主题实体提取

BERT是一种新的语言表示模型,它可以使 用双向编码器对大型无标记语料库进行预训练, 并在其他任务上对预训练的模型进行微调。 BERT已被广泛应用,并在各种自然语言处理任 务中显示出极大的改进,如分词、命名实体识 别、情感分析和问题回答等方面。在这里可以 使用BERT提取上下文特征。

首先,'[CLS]'会被添加到每个序列的开始, 作为序列的第一个标记。对应于第一个令牌的 Transformer输出的最后隐藏状态被用作分类任 务的句子表示。如果一个任务中有两个句子, 则使用"[SEP]"分隔这两个句子。

BERT通过使用预训练目标,基于掩盖的语 言模型(MLM)对模型参数进行预训练,该模型 随机对输入序列当中的某些位置进行屏蔽,并 设置优化目标来根据上下文预测屏蔽单词的原 始词汇表id。与从左到右的语言模型预训练不 同,MLM能够使得模型编码得到的结果同时包 含上下文的语境信息,从未被标记的文本中预 训练出一个深度双向的表示模型。

除了基于掩盖的语言模型,BERT还训练了 一个"下句预测"任务,该任务联合预训练文 本对进行表示。

2.4 条件层归一化

分层归一化方法克服了批处理归一化的缺点。

$$LN(x_i) = \gamma^* \frac{x_i - \mu_L}{\sqrt{\sigma_L^2 + \varepsilon}} + \beta$$
(3)

其中: μ_L 和 σ_L 表示均值和标准方差。一层中所 有的隐单元都有相同的归一化项 μ 和 σ 。定义 γ 和 β 为 $LN(x_i)$ 同维的偏置和增益参数。

为了从主体实体处理阶段调整整个联合实体-关系提取的过程,我们提出了条件层归一 化。该方法的关键思想是通过主体实体向量预 测层归一化的γ和β。我们可以使用两种不同的 变换矩阵将输入条件,即主体实体向量变换到 与β和γ相同的维数,然后将两种变换结果分别 加到β和γ上。为了防止原始预训练的权值干 扰,两个变换矩阵都应初始化为0。因此,在初 始状态下,模型与原始预训练模型一致。

$$BLN(x_i) = \gamma(s) * \frac{x_i - \mu_L}{\sqrt{\sigma_L^2 + \varepsilon}} + \beta(s)$$
(4)

这些更新的 $\gamma(s)$ 和 $\beta(s)$ 是从主体实体向量 推导出来的。

2.5 损失函数调节

如果利用"头尾贴标签"的方法,这样会 出现类别不平衡的问题。通常,目标实体词比 非目标词少得多,因此标签1将比标签0少得 多。处理不平衡的常规方法可以使用,如损失 函数或手动调整类的权重。但是,还可能会遇 到阈值的问题。我们提出了一种新的方法,即*n* 次方概率。

具体来说,原来的输出是一个概率值p,表示第一类的概率为p,我们将其改为 p^n ,表示第一类的概率为 p^n 。损失函数L(p)仍然用交叉熵损失函数。由于 $0 \le p \le 1$, p^n 将更接近于0。从而使初始状态满足目标分布,最终可以加快收敛的速度。

假设标签 $t \in \{0, 1\}$,则原始损失为

 $L(p) = -t \log p - (1 - t) \log(1 - p)$ (5) n次方之后的损失变成:

 $L(p) = -t \log p^{n} - (1 - t) \log (1 - p^{n})$ (6)

当标签为1时,那么-tlogpⁿ = -ntlogp,相 当于放大损失的权重。当标签为0时,损失量 log(1 - pⁿ)会更小。因此,我们的方法可以自适 应地调整损失权重。与损失函数或手动调整类的 权重相比,该方法的优点是在不改变内积分布的 情况下,使分布更接近目标,有利于模型的优化。

3 实验过程

3.1 数据集和评价指标

我们在本次实验中使用SKE数据集进行评估。SKE数据集是业内最大的中文信息抽取数据集,SKE数据集中所有句子均来源于百度百科和百度新闻提要。该数据集分为一个训练集(170000句)、一个开发集(20000句)和一个测试集(20000句)。训练集和开发集将用于培训,并可免费下载。测试集分为两部分:测试集1用于自我验证,测试集2在实验结束前一周发布,用于最终评估。

3.2 超参数和实现细节

最大序列长度设置为128,关系抽取的全连

接层设置为1, 批大小设置为64。

在训练过程中,我们使用Adam优化器,学 习率为2E-5,丢弃率为0.1。该模型在5轮次内 收敛。

为了使训练过程更加稳定,对所有可训练 变量采用指数移动平均,衰减率为0.99。模型 的损失函数是二元分类的交叉熵损失函数,因 为实体的从头到尾矢量预测是一个二分类问题。 主实体的预测是两个二分类问题,关系类型*p*和 目标实体*o*的预测问题是100个二分类问题。在 训练过程中,学习首先经过一个热身步骤,从 零开始慢慢增加,然后慢慢减少,以防止模型 发散过多而不收敛。

3.3 结果

SKE数据集的结果见表1。基线模型基于谷歌BERT,使用硬标签嵌入,只在SKE数据集上进行训练。如表1所示,结合条件层归一化, F1得分从0.820增加到0.833。损失函数中的*n*次幂概率也提高了F1得分。综合所有这些贡献,我们在开发集中的单一模型上获得了F1得分为0.840。

表 1 联合模型在 SKE 数据集上的性能

| Model | Precision | Recall | F1-score |
|------------------------------|-----------|--------|----------|
| Multi-Head Selection | 0.821 | 0.855 | 0.837 |
| Baseline | 0.802 | 0.838 | 0.820 |
| Baseline+CLN | 0.815 | 0.851 | 0.833 |
| Baseline+N-th weight | 0.812 | 0.846 | 0.829 |
| Baseline+all | 0.823 | 0.857 | 0.840 |
| Baseline+all+active learning | 0.842 | 0.862 | 0.852 |

该方法与SKE数据集最近发布的"基于 BERT的多头选择"^[39]方法的结果进行了比较。 可以看出,我们的模型优于以往的方法。 R-BERT的F1值为0.840,这比该数据集上的前 一个解决方案更好。

此外,我们还对主动学习的性能进行了评价, 并对MNLP(maximum normalized log-probability)和 RAND(random select)两种策略进行了比较。使 用1%的原始训练数据初始化模型。在每一轮 中,每个算法从剩余的训练数据中选择句子, 添加到训练池中,直到选择17万句。如图3所 示, MNLP算法的效果明显优于随机方法。这 里, 主动学习方法仅使用全部训练数据的 48.5%, 就能取得与原始模型相同的性能。



图 3 测试数据集上 F1 分数与对应于标注的句子数量

4 结语

本文将BERT纳入到一个新的框架中,并将 主动学习应用于联合实体关系提取中。该模型 从四个角度扩展了现有的方法。第一,该方法 可以解决多个实体属于多个三元组的问题。基 于概率图的思想设计了该框架,并开发了一种 新的"头尾"标记方法;第二,提出了一种将 主动学习应用于关系抽取问题的创新方法;第 三,为了在主体实体、谓词实体和客体实体之 间传递信息,我们提出了一种新的归一化方 法——条件层归一化;第四,设计了一个新的 损失函数来避免类的不平衡。实验证明,我们 增强了模型的信息提取能力,在单模型的测试 集上F1-score达到了0.840,在比原始的全数据 训练的深度模型少得多的数据下取得了更好的 性能。

参考文献:

- [1] PRANJAL A, BALCAN M F, LONG P M. The power of localization for efficiently learning linear separators with noise [EB/OL]. arXiv: 1307.8371, 2013.
- [2] CHAN S Y,ROTH D. Exploiting syntactico-semantic structures for relation extraction [C] // Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, Oregon, USA, 2011:551-560.

- [3] MAKOTO M, MIYAO Y, SÊTRE R, et al. A rich feature vector for protein-protein interaction extraction from multiple corpora [C] //Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, Singapore: Association for Computational Linguistics, 2009:121-130.
- [4] QI L, JI H. Incremental joint extraction of entity mentions and relations[C]//Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Baltimore, Maryland: Association for Computational Linguistics, 2014:402-412.
- [5] MAKOTO M, SASAKI Y. Modeling Joint Entity and Relation Extraction with Table Representation [C] //Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar: Association for Computational Linguistics, 2014: 1858-1869.
- [6] REN X, WU Z Q, HE W Q, et al. CoType: joint extraction of typed entities and relations with knowledge bases[C]//Proceedings of the 26th International Conference on World Wide Web, 2017.
- [7] LI X Y, FAN Y, SUN Z J, et al. Entity-relation extraction as multi-turn question answering [C] //Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy: Association for Computational Linguistics, 2019: 1340-1350.
- [8] MAKOTO M, BANSAL M. End-to-end relation extraction using LSTMs on sequences and tree structures [C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany: Association for Computational Linguistics, 2016:1105-1116.
- [9] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C] //Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota: Association for Computational Linguistics, 2019:4171-4186.
- [10] RADFORD A, NARASIMHAN K, SALINANS T, et al. Improving language understanding by generative pre-training [EB/OL], 2018. https://www.semanticscholar.org/paper/Improving-Language-Understanding-by-Generative-Radford-Narasimhan/cd18800-

a0fe0b668a1cc19f2ec95b5003d0a5035.

- [11] LIU Y H, OTT M, DU J F, et al. RoBERTa: a robustly optimized BERT pretraining approach [EB/ OL]. arXiv: 1907.11692,2019.
- [12] DOJCHINOVSKI M, REDDY D, KLIEGR T, et al. Crowdsourced corpus with entity salience annotations [C] //Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC). Portoro\vz, Slovenia: European Language Resources Association(ELRA),2016:3307-3311.
- [13] PASCAL H, KRZYSTOF J, JULIA B G, et al. Editorial of the special issue on latest advancements in linguistic linked data [J]. Semantic Web, 2022, 13(6): 911-916.
- [14] SANJOY D, KALAI A T, MONTELEONI C.
 Analysis of perceptron-based active learning [C]
 //Proceedings of the International Conference on Computational Learning Theory, 2005:249-263.
- [15] DAVID N, SEKINE S. A survey of named entity recognition and classification[J]. Lingvisticæ Investigationes. International Journal of Linguistics and Language Resources, 2007, 30(1): 3-26.
- BRYAN R, HARABAGIU S. UTD: classifying semantic relations by combining lexical and semantic resources [C] //Proceedings of the 5th International Workshop on Semantic Evaluation, Uppsala, Sweden: Association for Computational Linguistics, 2010: 256-259.
- [17] ALEXANDRE P, KUMAR V, MCCALLUM A. Lexicon infused phrase embeddings for named entity resolution [C] // Proceedings of the Eighteenth Conference on Computational Natural Language Learning, Ann Arbor, Michigan: Association for Computational Linguistics, 2014:78-86.
- [18] LUO G, HUANG X J, LIN C Y. Joint entity recognition and disambiguation [C] //Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Lisbon, Portugal: Association for Computational Linguistics, 2015:879-888.
- [19] CHIU J, J E. Named entity recognition with bidirectional LSTM-CNNs[J]. Transactions of the Association for Computational Linguistics Nichols, 2016, 4: 357-370.
- [20] HUANG Z H, XU W, YU K. Bidirectional LSTM-CRF models for sequence tagging [EB/OL]. arXiv:1508.01991,2015.
- [21] BRYAN R, HARABAGIU S. UTD: classifying se-

mantic relations by combining lexical and semantic resources [C] //Proceedings of the 5th International Workshop on Semantic Evaluation, Uppsala, Sweden: Association for Computational Linguistics, 2010: 256-59.

- [22] RICHIRD S, CHEN D Q, D.MANNING C, et al. Learning new facts from knowledge bases with neural tensor networks and semantic word vectors[EB/OL]. arXiv:1301.3618,2013.
- [23] ZENG D J, LIU K, LAI S W, et al. Relation classification via convolutional deep neural network [C] //Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers, Dublin, Ireland: Dublin City University and Association for Computational Linguistics, 2014: 2335-2344.
- [24] REN X, WU Z Q, HE W Q, et al. CoType: joint extraction of typed entities and relations with knowledge bases[C]//Proceedings of the 26th International Conference on World Wide Web, 2017: 1015-1024.
- [25] ZHENG S C, HAO Y X, LU D Y, et al. Joint entity and relation extraction based on a hybrid neural network[J]. Neurocomputing, 2017, 257:59-66.
- [26] KATIYAR A, CARDIE C. Going out on a limb: joint extraction of entity mentions and relations without dependency trees [C] // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Vancouver, Canada: Association for Computational Linguistics, 2017:917-928.
- [27] ZENG X R, ZENG D J, HE S Z, et al. Extracting relational facts by an end-to-end neural model with copy mechanism [C] //Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia: Association for Computational Linguistics, 2018:506-514.
- [28] BEKOULIS I, DELEU J, DEMEESTER T, et al. Joint entity recognition and relation extraction as a multi-head selection problem [J]. Expert Systems With Applications, 2018:114:34-45.
- [29] JACOB D, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C] //Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, Minnesota: Association for

Computational Linguistics, 2019: 4171-4186.

- [30] LIU Y H, OTT M, GOYAL N, et al. RoBERTa; a robustly optimized BERT pretraining approach [EB/ OL]. arXiv: 1907.11692, 2019.
- [31] WU S C, HE Y F. Enriching Pre-trained language model with entity information for relation classification[EB/OL]. arXiv:1905.08284,2019.
- [32] HUANG Z H, XU W, YU K. Bidirectional LSTM-CRF models for sequence tagging [EB/OL]. arXiv:1508.01991,2015.
- [33] XUE K, ZHOU Y M, MA Z Y, et al. Fine-tuning BERT for joint entity and relation extraction in Chinese medical text[C]//Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine, 2019:892-897.
- [34] ZHANG T, FRANK J O. Text categorization based on regularized linear classification methods [J]. Information Retrieval, 2001, 4(1):5-31.
- [35] ZHU X J, D. LAFFERTY J, GHAHRAMANI Z. Combining active learning and semi-supervised learn-

ing using gaussian fields and harmonic functions [J]. American Journal of Hematology, 2003, 74 (3) : 179-181.

- [36] SETTLES B, CRAVEN M. An analysis of active learning strategies for sequence labeling tasks [C]
 //Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, Honolulu, Hawaii: Association for Computational Linguistics, 2008:1070-1079.
- [37] SHEN Y Y, YUN H K, LIPTON Z, et al. Deep active learning for named entity recognition [C]//Proceedings of the 2nd Workshop on Representation Learning for NLP, Vancouver, Canada: Association for Computational Linguistics, 2017:252-56.
- [38] LEWIS, DAVID D. A sequential algorithm for training text classifiers [J]. ACM SIGIR Forum, 1995, 29 (2):13-19.
- [39] HUANG W P, CHENG X Y, WANG T F, et al. BERT-based multi-head selection for joint entityrelation extraction[EB/OL]. arXiv:1908.05908,2019.

Research on entity relation extraction based on active learning

Sun Han^{*}

(Network Service Center, Taiyuan Open University, Taiyuan 030027, China)

Abstract: Relation classification is an important NLP task to extract relations between entities. In this paper, we report our method for a largest schema-based Chinese information extraction dataset. We incorporate BERT into a new framework and apply active learning for joint entity relation extraction. This model extends existing approaches from three perspectives. First, our method could solve the problem that multiple entities belongs to multiple triplets. We design this framework based on the idea of probability graph and develop a new "head-tail" labeling method. Second, we proposed an innovative approach that apply active learning on relation extraction problem. Third to transmit information between subject entities, predicate and object entity, we propose a new normalization method called conditional layer normalization. fourthly, a new loss function is designed to avoid class imbalance. Therefore, we enhance the information extracting ability of the model and achieve F1-score 0.840 on test set with a single model, and achieve better performance with much less data than the original deep models trained by full data.

Keywords: BERT; active learning; joint entity relationship extraction

文章编号:1007-1423(2024)08-0084-06

DOI: 10.3969/j.issn.1007-1423.2024.08.014

多模态学习投入测评的研究现状与启示 ——基于CiteSpace可视化分析

郑晨虹*,张海博

(贵州师范大学教育学院,贵阳 550025)

摘要:多模态技术突破传统单一模态数据的局限,实现对学习投入精准、科学全面的测评。以Web of Science数据 库收录的204篇文章为样本,运用CiteSpace对多模态学习投入测评研究进行了可视化计量分析,较为全面地展现了主要 研究国家、期刊、高被引论文、关键词聚类特征等。未来,多模态学习投入测评可以进一步拓宽应用场景,深化研究 主题; 重视优化算法,提高数据分析精准度; 介入生理数据,加强多模态数据协同分析。

关键词: 多模态技术; 学习投入; 知识图谱

0 引言

学习投入主要包含行为投入、认知投入、 情感投入和社交投入^[1],是影响学习绩效和学 习动机的重要因素,对其进行客观准确的测评 可以准确反映学习者更加真实的学习状态。由 于学习投入本身的复杂性,以往学习投入的表 征多采用传统测量、自陈量表和行为编码等单 模态数据评价方法,不仅难以实现科学、准确、 真实、客观的评价要求,而且学习的内在发生 机制和过程仍属于一个"黑箱",难以进行有效 测量。近年来,多模态技术因其可以动态捕获 两种以上的过程数据,实现对复杂评价过程的 客观表征,受到了学习投入研究者的广泛关注。

多模态学习投入测评致力于运用不同模态 数据的互补性优势,对测评结果的准确性作了 进一步的提升,包含外显可直接观察到的动作 型数据和伴随式内隐的生理型数据。其中,以 心电、皮肤电、肌电为代表的内隐生理数据在 经过人工智能技术处理后,可以分析出学习者 的情感投入,从而形成更加完整的情绪画像^[2]。 Noroozi等^[3]提出将自我报告与生理数据相融合 的多模态分析技术,用以全面、客观地刻画学 生的学习投入。Andrade等^[4]构建了包含身体、 凝视、手势、模糊语、言语五大类数据识别行 为集群的分析框架,即使用多模态学习分析法 为学生的行为进行建模,探究行为与学习之间 的深层关系。

由此可见,国外多模态学习投入测评研究 在理论探索和实证检验方面都取得了一定的经 验,以国外多模态学习投入测评研究为鉴,把 握多模态学习投入测评研究现状和发展脉络, 挖掘潜在的研究价值,可以为我国学者开展基 于多模态技术的学习投入评价研究提供更多的 参考和启示。

1 研究设计

1.1 研究方法

科学知识图谱可以实现对多模态学习投入 相关研究作者、地域、发表年份、关键词等进 行量化分析和可视化,具有标准化的研究流程, 可以帮助研究者更加形象直观地了解研究领域 发展状况。因此,本研究运用 CiteSpace 软件

收稿日期: 2023-11-05 修稿日期: 2023-11-27

作者简介: *通信作者:郑晨虹(1996—),女,浙江金华人,在读硕士研究生,研究方向为教育数据挖掘、教师专业发展,E-mail:zheng_ch66@163.com; 张海博(1996—),男,河南周口人,在读硕士研究生,研究方向为信息技术应用、教学评价研究

(5.8.R3版)对多模态学习投入测评研究开展可 视化数据分析,采用聚类分析、共现分析等方 法梳理多模态技术的发展现状,包括主要发文 国家和期刊,挖掘具有影响力的作者和文章, 展现领域研究热点。

1.2 数据来源

科学引文数据库(web of science, WoS)收录 了全球范围内大量权威期刊,其中的学术性文 章具有一定的代表性和前沿性,并且可以通过 独特的引文检索,追溯研究领域的起源和发展 历史。另外,考虑到CiteSpace文献分析软件是 基于WoS数据格式开发的,在对WoS数据库文 章进行共现分析、共被引分析时具有更强的适 应性。因此,为了有效获取国际上多模态学习 投入研究领域内的高水平研究文章,本研究在 WoS核心合集数据库中以"multimodal learning engagement"作为主题词进行文章检索,共获得 487篇国际相关研究文章。

1.3 样本筛选

为了提高文献计量分析的准确性和科学性, 精准揭示多模态学习投入研究的现状和未来趋势,本研究参考系统性文献综述法中基础的文 章筛选标准并做出适应性调整^[5],针对初步检 索出的487篇文章制定了5条文章纳入/排除标 准,如下所示:

(1) 英文文章/非英文文章;

(2)文章来源是期刊论文/书稿、会议论文、 报告等;

(3) 文章内容至少包含三页/少于三页的海 报、短论文或简介等;

(4) 文章标题只出现一次/重复性标题;

(5)研究方向为教育类研究/研究方向侧重 计算机科学、心理学等。

后续依据该标准开展严格的计量分析文献 样本筛选过程(如图1所示),最终确定纳入204 篇多模态学习投入相关研究文章作为本研究的 样本数据。



图 1 文献样本筛选流程图

2 相关研究的计量分析

2.1 主要发文期刊和国家

从发文国家来看(如图2所示),美国是最早 将多模态数据应用到学习投入测评和分析上的, Skinner等于2008年发表的"Engagement and disaffection in the classroom: part of a larger motivational dynamic?"论文开启了领域研究。同时, 美国也是目前为止发文总量最多的国家,以88 篇的总数排名第一,远远超过位于第二的澳大 利亚(26篇)。另外,美国发文的中介中心性为 0.42(>0.1),说明其是非常重要的关键节点,对 领域的发展和连接起着重要作用,有着举足轻 重的地位。综合分析发现,受美国影响,澳大 利亚、挪威、英国等国家研究者在2010年后陆 续开展了多模态学习投入的相关研究。

我国的相关研究始于2017年,虽然相对于 美国晚了近十年,但是在国际高水平期刊上, 我国研究者陆续发表了13篇学术论文,已处于 领域内研究前沿。但是,各国在多模态学习投 入研究领域上的连接线段较少且稀疏,可见彼 此之间仍缺乏紧密的联系,远未形成稳定合作 的学术共同体(如图2所示)。





为了进一步追溯多模态学习投入测评的研究情况,本研究对文章高被引的期刊进行可视 化分析,生成了一幅具有443个节点、1245条连 线的关系图。同时,为了清晰展现被引用期刊之 间的关系,本研究只提取被引频次大于25的期 刊,并在节点显示期刊名称,结果如图3所示。



图 3 被引期刊关系图

可以看出,节点最凸显的是《Comput Edue》 期刊,作为该领域内具有较大影响力和领导力 的权威期刊,早在2007年初开始刊登多模态学 习投入主题的相关文章,被引共48次,点度中心 度达到0.46。紧随其后的是《J Adolesc Adult Lit》 期刊,被引次数达到45次,但通过对该节点周围 连线的进一步解读,可以发现:尽管该期刊拥有 相对较高的被引量,但与另外其它期刊之间的联 系不够紧密。除此之外,《Harvard Educ Rev》期 刊被引数为39次,《Brit J Educ Technol》和 《Rev Educ Res》期刊的被引次数也达到了32 次。从被引期刊关系图可以明显看出,多模态学 习投入的研究虽在小范围内形成了一定的影响 力,但是论文的被引数量和联系仍然不足,尚未 形成多模态学习投入测评的专业学术研究团体。

2.2 文章共被引分析

在发文数量上, Candance Doerrstevens、Kay Kimber和Kristina Kumpulainen居于榜首,均发 表了三篇相关文章。另外Sharma、Kajamaa、 Wyattsmith、Chenlee、Gourlay和Pantaleo研究者 均发表了两篇文章。从现有的文章发表数据上 看,领域研究者的发文数量仍相对较少,因此 还需要研究者们更加深入的探讨和持久的关注, 增强彼此之间的交流和联系,形成强有力的学 术共同体,共同推进多模态学习投入研究的深 入发展。

由于文章的共被引数据可以直观显现出具 有代表性的作者和文章,因此对多模态学习投 入研究论文进行了共被引数据分析。从表1可以 发现,Andrade、Sharma、Grawemeyer、Blikstein、Giannakos研究者发表的论文,被引次数 达到4次,是目前领域内被引次数最多的。其 中,Sharma无论从发文总量,还是从共被引次 数上看,都是领域内重要的研究者,为促进多 模态学习投入测评研究发展做出了一定的贡献。 此外,领域内具有较大影响力的论文大多产生 于2015年之后,可见多模态学习投入相关研究 起步时间较晚,作为信息技术发展下的新生研 究领域,未来极具深度挖掘的价值。

表 1 文章共被引分析表

| 年份 | 作者 | 篇名 |
|------|------------|--|
| 2016 | Andrade | Using multimodal learning analytics to model student behaviour: a systematic analysis of behavioural framing |
| 2019 | Sharma | Building pipelines for educational data using AI and multimodal analytics : a "grey-box" approach |
| 2017 | Grawemeyer | Affective learning: improving engagement and enhancing learning with affect-aware feedback |
| 2016 | Blikstein | Multimodal learning analytics and education data mining: using computational technologies to measure complex learning tasks |
| 2019 | Giannakos | Multimodal data as a mean to understand the learning experience |

2.3 关键词聚类分析

研究发现排名前十的关键词分别为:student、 engagement、 education、 language、 literacy、 knowledge、 technology、 classroom、 environment 和 EEG。可以看出:首先,多模态学习投入研 究侧重于探究学习者在智能辅导系统,智慧课 堂环境下对识字、语言学习、知识理解等学习 活动的参与程度和注意力保持。其次,功能性 近红外光谱成像装置(fNIRS)、运动传感器、眼 动仪等现代智能分析装备与自然语言处理、计 算机视觉等技术广泛应用于学习测评研究。再 次,智能化采集设备和便携式传感器的广泛使 用,使得学生的脑电、心电、肌电、皮肤电等 生理数据获取更加便捷,为研究者开展学习投 入测评提供了基本工具和研究角度。总之,多 模态数据是推动测评从结果性到过程性转变的 关键,但目前针对学习投入的研究仍处于初步 探索阶段。

关键词聚类可用来发现当前多模态学习投 入测评的研究热点。因此,选择 MI 算法对关键 词进行聚类,提取出主要11个研究区块(如图4 所示)。聚类#0 coherence,是对多模态数据进行 数据处理, 增强数据的可读性和可解释性, 从 而能够有效实现数据的保存和共享。聚类#5 artificial intelligence 是多模态数据协同分析常用 的技术和方法,在人机协同的发展大环境下, 有效运用机器深度学习、卷积神经网络算法实现 数据的自动化分析。尽管在研究方法上, 日趋强 调数据密集分析下的定量分析,但同样不可缺少 定性分析。聚类#4 narrative inquiry<research methodology 叙事调查作为一种研究方法在多模 态学习投入测评中依然被广泛使用。聚类#6 multiliteracy 和聚类#8 art education 强调多模态学 习投入测评具体的研究领域,在语言学习和艺 术教育上得到了初步应用探索。另外,聚类#10 elf learning 是指学生在电子环境下开展学习活 动,该状况下的学习投入会深刻影响学习者的 学习绩效,多模态数据可以实现对其进行科学测 评。聚类#1 being and becoming、聚类#3 children preferences 和聚类#9 learner initiatives 则侧重对 学习过程中学习者的学习偏好和状态的监测, 最终希望能够提高学生的主动性和积极性。聚 类#7 empowering teachers 是多模态学习投入测评 最终指向,给予教师反馈,提高教师教学能力。 通过聚类分析结果可知,多模态学习投入测评 是聚类#2 multimodal learning多模态学习分析中 的重要方向,其本质是借助数据科学发现隐藏 在学习背后的教育规律和原理,在实践中发挥 指导意义和操作化价值。





3 研究结论与启示

3.1 研究结论

研究筛选了国外多模态学习投入为主题的 204篇研究文章,通过对发文期刊和国家、文章 共被引频次、关键词聚类等数据可视化分析, 得到以下三条结论。

(1)从主要发文期刊和国家看,世界各国和 主要发文期刊在多模态学习投入研究领域上缺 乏紧密的联系,尚未形成稳定的学术共同体。

(2)从文章发表数量及共被引频次来看,多 模态学习投入测评研究总体起步时间较晚,是 现代技术发展下的新生研究领域,与人工智能、 算法技术等领域密切较高,具有深度挖掘的必 要性。

(3)从关键词排名和聚类看,多模态数据是推动学习测评转变的关键技术,但数据类型采集、数据融合分析等仍处于初步探索阶段。

3.2 研究启示

在分析国外多模态学习投入研究的基础上, 对我国未来多模态学习投入测评研究方向可以 重点考虑以下三点:

(1)拓展应用场景,深化研究主题。多模态 数据支持的学习投入研究已然成为了教育评价 领域的研究热点,尤其在学习分析领域^[6]。多 模态数据作为洞察教育过程规律的新范式^[7],可以全面、客观地测评不同场景下的学习投入, 蕴藏着巨大的应用价值。多模态学习投入测评 作为一个新兴的研究领域,一方面可以持续探 索在不同教学模式下的应用,如STEM教育、在 线学习、混合同步课堂等,实现对学生的学习 投入的全面测评。另一方面,随着理论基础的 不断深入和实证研究的广泛开展,未来基于多 模态数据的学习投入必将逐渐从多模态学习分 析领域中分化出来,形成相对独立且具有成熟 研究基础的研究领域。此外,多模态学习投入 测评需要不断地进行丰富数据指标细化、分析 模型构建和实证应用探究,以更好应对复杂的 教育过程,适应新时代教育评价改革的要求。

(2) 重视算法优化,提高数据分析精准度。 面对学习过程的海量数据,数据驱动的学习投 入测评需要不断地进行优化算法,提高数据处 理分析的精准度。在机器学习和深度学习等技 术的推动下,现有的学习投入测评采用卷积神 经网络算法的频次较多,这种算法可以有效保 留特征,实现参数共享,减少数据的训练量。 然而, 大规模的多模态数据研究和分析面临着 训练数据缺少的难题,过度依赖人工标注,结 果的精确度还有很大的提升空间。赵亮等[8]针 对多模态数据的低质性问题,提出了一种新的 多模态鲁棒特征学习方法,主要由引入模态误 差矩阵来有效降低噪声数据对融合结果的影响, 并通过实验对比证明,实现对多模态大数据共 享特征的准确学习。可见,持续优化相关数据 分析算法,是提高基于多模态数据的学习投入 测评精确性的有效路径。

(3) 介入生理数据,协同分析多模态数据。 随着可穿戴传感器成本降低和便携特性凸显, 研究者可以高效获取和收集学习者的生理数据。 由此,多模态数据分析不再局限于外显行为, 更多结合了皮肤电、肌电等人体生理数据。从 文献计量分析中可以发现,脑电波(EEG)作为 突出节点,成为研究领域内重点关注的数据采 集工具。脑电波借助脑电仪可以记录大脑活动 时产生的电波变化,不仅反映人体生理和疾病 数据,也包含学习者的思维活动信息,给学习 投入的测评研究带来了新角度。大量研究也表 明,基于海量多模态数据进行测评分析通常更 具准确性^[9]。生理数据的介入成为可以预见的必 然。已经有越来越多的研究开始采用三种及以 上不同模态数据来实现协同分析,有望突破教 育数据挖掘关键技术,推动多模态数据融合表 征,全面洞悉学习者在学习活动过程中的投入 程度,揭示学习规律。

参考文献:

- FREDRICKS J A, BLUMENFELD P C, PARIS A
 H. School engagement: potential of the concept, state of the evidence[J]. Review of Educational Research, 2004,74(1):59-109.
- [2] BANDARA D, SONG S, HIRSHFIELD L, et al. A more complete picture of emotion using electrocardiogram and electrodermal activity to complement cognitive data [C] // Proceedings of the 10th International Conference on Foundations of Augmented Cognition: Neuroergonomics and Operational Neuroscience, 2016: 287-298.
- [3] NOROOZI O, PIJEIRA-DÍAZ H J, SOBOCINSKI M, et al. Multimodal data indicators for capturing cognitive, motivational, and emotional learning processes: a systematic literature review[J]. Education and Information Technologies, 2020, 25: 5499-5547.
- [4] ANDRADE A, DELANDSHERE G, DANISH J A. Using multimodal learning analytics to model student behaviour: a systematic analysis of behavioural framing
 [J]. Journal of Learning Analytics, 2016, 3 (2): 282-306.
- [5] 李新,李艳燕.基于系统性文章综述的国外学习投入实证研究分析[J].现代远程教育研究,2021,33
 (2):73-83,95.
- [6] 张琪,武法提,许文静.多模态数据支持的学习投入 评测:现状、启示与研究趋向[J].远程教育杂志, 2020,38(1):76-86.
- [7] 黄琰,赵呈领,赵刚,等.教育过程挖掘智能技术:研究框架、现状与趋势[J]. 电化教育研究,2020,41
 (8):49-57.
- [8] 赵亮,张洁,陈志奎.基于双图正则化的自适应多模态 鲁棒特征学习 [J/OL].计算机科学:1-12
 [2022-05-09]. http://kns. cnki. net/kcms/detail/50.1075.TP.20211207.0005.002.html.
- [9] 骆方,田雪涛,屠焯然,等.教育评价新趋向:智能化测评研究综述[J].现代远程教育研究,2021,33
 (5):42-52.

(下转第94页)

文章编号:1007-1423(2024)08-0089-06

DOI: 10.3969/j.issn.1007-1423.2024.08.015

疫情防控背景下基于人脸识别及测温技术的智能门禁应用

左 楷,唐耀平*

(湖南科技学院理学院,永州 425199)

摘要:通过深入研究人脸识别和测温技术,实现了门禁系统的智能化,旨在为疫情防控提供更高效、更安全的解决方案。该应用系统将人脸识别技术和红外测温技术相结合,在快速准确识别出入人员身份信息的同时实现体温快速 检测,并对通行人员进行记录和统计,为疫情防控提供更加全面的数据支持。

关键词: 人脸识别; 红外测温; 门禁系统; 智能防控

0 引言

在疫情防控阶段,门禁管理是实现工作、 学习环境封闭管理的重要手段。疫情防控有着 严格的进出人员的测温需求,因而基于人脸识 别及测温技术的智能门禁系统得到了广泛应用, 其可以实现快速准确地测温,有效保障了场所 的安全,符合现代化安全管理的需求。此外, 智能门禁应用还可以为用户提供便捷的出入管 理服务,提高了管理效率。

本文旨在研究基于人脸识别及测温技术的 智能门禁应用,设计并实现一个高效、精准、安 全的门禁系统。具体研究内容包括:人脸识别算 法的研究和优化、测温技术的集成和优化、门禁 系统的设计和实现等。通过研究,为实现更加安 全、高效的门禁管理提供可行性解决方案。

1 人脸识别及红外测温系统的分析与设计

1.1 人脸识别设计与应用

人脸识别作为现在常用的身份鉴别手段, 被逐渐引入到人们的学习、工作和生活当中^[1]。 人脸识别是一个具有发展潜质的领域,也是一 个具有极大挑战力的领域^[2]。本设计在人脸识 别的基础上做出了一系列符合当代环境下的拓 展。下面将列出对人脸识别的流程与现代识别 的几种方法与选择。

1.1.1 人脸识别流程

(1)数据采集:通过摄像头采集人脸图像。

(2) 预处理:对采集到的人脸图像进行预处 理^[3],包括图像尺寸调整、人脸定位、人脸对 齐等操作,以便于后续的特征提取和匹配。

(3)特征提取:从预处理后的人脸图像中提取出一组关键特征向量,用于表示这张人脸图像的独特特征。常用的特征提取算法包括PCA、LDA、SIFT、SURF、Deep Learning等。

(4)特征匹配:将特征向量与已有的人脸特 征库中的特征进行匹配,计算出它们之间的相 似度或距离,确定最佳匹配结果。

(5) 判定识别结果:根据匹配结果和系统预 设的匹配阈值,对识别结果进行判定,如果匹 配度超过阈值,则认为识别成功。

(6) 反馈处理:根据识别结果,执行相应的 反馈处理操作,例如开门、拒绝进入、报警等。



基金项目: 2021年度湖南省大学生创新创业训练计划项目(湘教通[2021]197号-3575)

作者简介:左楷(2002—),男,湖北黄冈人,学生,专业为信息与计算科学;*通信作者:唐耀平(1973—),男,湖南 永州人,教授,硕士,主要从事数值代数及群体决策研究,E-mail:typ_01@163.com

收稿日期: 2023-11-06 修稿日期: 2023-11-27

在识别过程中,如果遇到人脸图像质量不 佳,或者特征向量与存储的特征向量差距较大, 可能导致识别结果不够准确。因此,在设计和 实现人脸识别系统时,应该考虑到图像采集、 特征提取、特征匹配等方面的技术与算法,并 进行充分测试与优化,以提高准确度和稳定性。

1.1.2 人脸识别设计

(1) 基于特征脸算法(Eigenfaces)。基于特 征脸算法是一种传统的人脸识别方法,该算法 的实现首先是对图像进行预处理,对采集到的 人脸图像进行裁剪、缩放、灰度化等处理。然 后再进行特征提取,特征提取使用主成分分析 (PCA)将图像转换为对应的特征向量,并选择 前N个特征向量作为特征脸。最后进行人脸识 别,将待识别的人脸图像与特征脸进行比对, 选取最相似的人脸作为识别结果。该算法的优 点是准确性高,对光照、姿势等因素不敏感, 但缺点是对图像中干扰因素较为敏感,容易出 现误识别。

(2) 基于 Fisher 线性判别分析算法(Fisherfaces)。该算法的实现与特征脸算法类似,主要 区别在于使用线性判别分析(LDA)进行特征提 取。该方法使用 LDA 将图像投影到一个新的低 维空间中,得到投影后的特征向量。该算法的 优点主要是它可以对光照和表情变化进行内插 和外推,有研究表明,在预处理时与主成份分 析(PCA)方法结合,Fisherfaces算法准确度可达 到93%。

(3) 基于神经网络的人脸识别算法(Convolutional Neural Network)。基于神经网络的人脸 识别算法是一种使用神经网络来进行人脸识别 的算法,它通过对人脸图像进行特征提取和表 示,然后使用神经网络来进行学习和分类。该 算法的优点是对图像中的变形和干扰因素的影 响较小,且具有较高的识别准确率,但缺点是 需要大量的训练数据来进行学习和分类,且模 型的训练和调参比较复杂。

1.1.3 人脸识别算法选取

本项目基于OpenCV计算机视觉库实现人脸 识别。该库提供多种人脸检测和识别算法的实 现,包括 Haar Cascade、LBP Cascade、Eigenface、Fisherface、LBPH等。本项目环境为Windows 11, Python 3.7.3, OpenCV-python 3.4.2.16, OpenCV-contrib-python 3.4.2.16。本项目采用局 部二值模式直方图(Local Binary Patterns Histogram)的方法,通过调用OpenCV中的cv2.face. LBPHFaceRecognizer_create()来创建识别器模 型,并用cv2.face_FaceRecognizer.train()函数进行 训练, 最后用 cv2.face FaceRecognizer.predict()进 行预测。该算法分为训练人脸数据集和人脸识 别两大步骤。在训练人脸数据集阶段,首先加 载人脸数据集中的人脸图像,进行预处理和特 征提取,使用Haar分类器^[4]进行人脸部位检测, 然后根据特征值和对应标签进行分类器的训练。 在人脸识别阶段,首先使用 Haar 分类器定位摄 像头捕捉的图像中的人脸部位,进行人脸对齐, 然后读取分类器文件,再调用识别器中的predict()函数进行人脸识别。根据 confidence 值与 之前设置的阈值对比,判断是否匹配成功,成 功则输出对应的人名;否则,输出unknown。算 法结构如图2所示。



图 2 人脸识别算法结构

1.1.4 实验数据集及测试结果

本项目数据采用 Labeled Faces in the Wild (LFW), LFW 是一个常用的人脸识别数据集, 包含超过13000张人脸图像,涵盖多种种族、年 龄和姿态,被广泛用于测试人脸识别算法的性 能^[5]。将该数据按照4:1分成训练集和测试集。 将训练好的分类器,用于测试数据,然后通过 识别器的预测函数,即predict函数得到的预测 id 与测试数据本身的id比较,如果不匹配且将 预测函数得到的相似度与阈值比较,如果小 于阈值,则认为匹配失败,并计数。通过调 整阈值重复上述操作。由于阈值过小会造成 相似度低的也能匹配成功,阈值过高,会造 成对于同一对象的不同角度照片匹配不成功, 因此在训练阶段通过随机梯度法调整阈值到 合适位置可以使匹配准确度收敛于最佳。上 述算法中,在置信度值为0.68时,匹配准确 率最大,为95.8%。

在上述参数的基础上,对一群人的不同环 境下照片进行训练,然后进行人脸识别,部分 识别结果如图3所示。



图 3 人脸识别测试

1.2 红外测温器设计

1.2.1 红外测温传感器的硬件选择

选择合适的红外测温传感器是门禁系统的 关键。在选择传感器时,需要考虑测温范围、 准确度、响应时间、分辨率等因素。在门禁系 统中,建议选择可以测量人体温度范围的红外 测温传感器,例如-20°C~50°C或30°C~45°C 范围内的传感器。此外,需要确保传感器具有 足够的准确度和响应时间,并具有高分辨率以 获得更精确的温度读数。

1.2.2 测温点的选择

在门禁系统中红外测温器安装位置,即测 温点的选择至关重要,在选择测温点时应该考 虑测量温度数据的准确性和用户体验。门禁系 统通常测量访问者的手腕温度或额头温度,考 虑到与前面人脸识别相结合,测量手腕对于访 问者来说,体验感并不太好。因此采用测量额 头位置,而且额头的温度数据比手腕更准确。 此外, 红外测温器应该安装在稳定的表面, 防 止阳光直射, 以及其他热源的干扰造成误差。

1.2.3 温度阈值设置及错误处理

为了有效遏制疫情扩散,门禁系统中需要 设定温度阈值。不同地区应根据当地防疫政策 或相关部门建议来确定。红外测量可能存在误 差,需定期校准设备,采用已知温度物体进行 校准,并多次测量以提高准确性。设备误差可 通过调整传感器参数来降低,而环境影响可通 过重复测量来提高准确性。

1.3 通行日志设计

通行日志是本项目中非常重要的部分,它 记录着每个通行者在系统中的进出记录,对于 管理者来说,可以通过这些记录进行安全管理, 后期也可以通过深度学习和大数据分析研究出 更有价值的数据,从而做出更有利于学校发展 的决策。

1.3.1 数据库设计

一个好的数据库设计可以对系统性能、数据可靠性、数据安全性等方面带来很多好处。 本项目中的通行日志使用关系型数据库MySQL。 数据库需要设计以下几个表:

(1)用户表。用户表是用来记录门禁系统中 的用户信息,包括用户ID、姓名、部门信息。 用户基本信息见表1。

表1 用户表

| 字段名称 | 类型 | 长度 | 约束 | 描述 |
|-----------|-------------|-------|--------|--------|
| UserId | varchar(11) | 11 | 主键(非空) | 用户ID |
| UserName | varchar(30) | 30 | 非空 | 用户姓名 |
| UserDepId | varchar(10) | 10 | 非空 | 用户部门编号 |
| FaceId | text | 65535 | 非空 | 用户人脸信息 |

(2) 门禁位置表。门禁位置表的设计是用于 记录某通行人员具体是从哪一个门通行的,其 中记录门禁位置ID、名称、位置信息。门禁详 情信息见表2。

| 表 2 | 门禁位置表 |
|-----|-------|

| 字段名称 | 类型 | 长度 | 约束 | 描述 | |
|-----------|-------------|----|--------|------|--|
| DoorId | varchar(7) | 7 | 主键(非空) | 门禁ID | |
| DoorName | varchar(20) | 20 | 非空 | 门禁姓名 | |
| DoorLocal | varchar(40) | 40 | 非空 | 门禁位置 | |

(3) 通行日志表。通行日志表记录门禁系统 中每一个通行记录,包括通行记录ID、通行人 员ID、通行时间、通行门禁位置ID、通行温度 和通行类型。其中通行类型表示该通行者是进 入还是出行,可以在数据库中用1和0表示。通 行日志基本信息见表3所示。

表 3 通行日志表

| 字段名称 | 类型 | 长度 | 约束 | 描述 |
|-------------|-------------|----|--------|-------------------------|
| PassLogId | varchar(40) | 40 | 主键(自增) | 通行记录编号 |
| UserId | varchar(11) | 11 | 外键(非空) | 通行人员编号 |
| PassTime | datetime | 8 | 非空 | 通行时间 |
| DoorId | varchar(7) | 7 | 外键(非空) | 通行门禁编号 |
| PassType | varchar(1) | 1 | 非空 | 通行类别(1表 示进,0表示 出) |
| Temperature | float(3,1) | 7 | 非空 | 通行者的体温 |

1.3.2 数据的采集

用户表数据可以与学校学生和教职工共享, 避免数据冗余。FaceId可通过学校照片收集并 打标签编号存入数据库,便于模型训练及阈值 设置。门禁位置表数据从已有门禁位置入库, 每添加一处门禁即加入表中。通行日志表数据 通过识别成功后返回的FaceId查找UserId,将 DoorId、时间和通行类别插入表中。这些数据有 助于访问控制、安全管理和数据统计。

1.3.3 数据传输和存储

门禁系统需要将采集到的数据传输到服务 器进行处理和存储。为了确保数据的安全性和 及时性,需要采用加密传输和备份策略等措施。 同时,需要设置合理的数据存储策略,例如数 据压缩和归档等,以便更好地管理和查询数据。

1.3.4 数据导出与分析

管理人员可通过后台查询通行记录,筛选

条件包括时间段、门禁位置、人员和体温等, 结果可导出为Excel表格。通行日志数据可用于 访问控制、安全管理和数据统计。开发应用程 序可更好地利用数据,例如安全事件报警程序。

1.4 系统可行性分析

人脸识别和红外测温相结合的门禁系统是 一种高度自动化和智能化的解决方案,可实现 快速准确的身份识别和温度检测。对系统可行 性进行分析,包括以下几个方面。

(1)技术可行性:人脸识别和红外测温技术 已成熟,结合用于门禁系统可实现高效、精准 的人脸识别和体温检测,提高安全性并减少病 毒传播风险。

(2)安全可行性:为提高校园安全系数和防止疫情传播,需提高人脸识别和红外测温准确度。数据可使算法更精确,设备定期校准可提高准确度。本项目可降低人员接触,降低疫情传播风险。

(3) 经济可行性:本项目利用开源的 OpenCV库来实现人脸识别,大大降低了软件开 发的经济支出,也降低了长期需要大量门卫保安 等人员的人力支出。后续可以通过研究通行日志 得出更有价值的数据,这是人工难以达到的。

2 实验测试与结果分析

2.1 实验数据收集

选取该门禁系统记录了两周的数据,共计 500个通行者,每个通行者平均访问了5次。在 每次通行之前,访客需要通过人脸识别和红外 温度检测来通过门禁。在导出数据时,选择了 通过门禁的时间和检测到的体温两个字段。

2.2 数据分析

采用Python进行数据分析得到以下结论。

2.2.1 通过门禁时间分布

绘制进出人员数量随时间的变化曲线来表示每个通行者通过门禁的时间分布,由图4可以看出,11点~12点和18点~20点进出门禁最为频繁,图中灰色柱形标记人流量最密集的时刻,14点~16点之间也非常活跃。



图 4 通行者每天时间分布图

2.2.2 通行者温度分布情况

绘制盒图来显示访客的体温分布情况,结 果如图5所示。



图 5 通行者体温分布图

由图5可以看出,通行者的体温分布情况较为正常,中位数约为36.2°C,标准偏差约为0.3°C,图5中有四个异常点,其中三个温度过高,一个温度过低。门禁系统能够准确地检测到这三个体温过高的参与者,并阻止了他们进入。这表明门禁系统的红外测温仪器在检测体温方面具有较高的准确性和可靠性。

2.2.3 访问次数分布情况

绘制一张带有直方图的密度图来显示每个 访客访问次数的分布情况,结果如图6所示。

通过图6可以看出,大部分人进出两周通行 次数在20次左右,然而还发现有些人通行次数 频繁,通行次数甚至超过了35次。



2.3 实验结果

基于上述数据,可以得出:①在11点~12点 和18点~20点进出门禁最为频繁;②红外温度 检测仪的准确率比较高,且大部分温度属于正 常,有少数人员温度异常,这可能与环境因素 有关;③从两周数据情况来看,大部分人员进出 次数在20次左右,即平均每天进出一次。此外, 注意到了一些潜在的问题:极个别人员进出频 繁,甚至两周内超过35次。由于统计数据量较 小,后续可以对进出频繁者进行更多关注,可以 进一步提高校园安全系数。此外,也可以设计好 的算法分析数据,增强门禁系统管理。

3 结语

随着科技进步,基于人脸识别和红外测温的门禁系统正在逐渐取代传统的门禁系统,成 为更高效和智能的选择。本文设计了一个人脸 识别系统结合红外测温的门禁系统,该系统运 用开源的OpenCV实现人脸数据集的训练与识 别,再用深度学习的方法寻找最优阈值来提高 人脸匹配准确率,进而提高门禁系统的安全。 在红外测温仪器的设计上使用现成的仪器。最 后就是通行记录的设计,本项目采用MySQL技 术来实现数据库的存储与查询功能,通过对数 据库的设计,降低该系统中数据冗余现象。该 项目可以推广到其他各个行业,如医院、火车 站、商场,等等。

参考文献:

[1] 王婉清.基于人脸识别的考勤系统设计[D].南京:

南京邮电的大学,2016.

- [2] 宋嘉程.人脸识别技术的现状和发展[J].电子技术 与软件工程,2017(9):172-178.
- [3] 汪成龙,孙培宜,林晓鹏,等.基于 SeetaFace 的人脸 识别门禁系统[J].制造业自动化,2018,40(8):

111-112,150.

- [4] 王洋,郑佳春.基于 OpenCV 人脸检测技术的研究 及实现[J].物联网技术,2018(3):81-83.
- [5] 袁小童.基于深度学习的人脸属性识别算法研究
 [D].北京:北京邮电大学,2023.

Intelligent access control application based on face recognition and temperature measurement technology under the background of epidemic prevention and control

Zuo Kai, Tang Yaoping*

(College of Science, Hunan University of Science and Engineering, Yongzhou 425199, China)

Abstract: Through in-depth research on face recognition and temperature measurement technology, the intelligence of the access control system has been realized, aiming to provide a more efficient and safer solution for epidemic prevention and control. The application system combines face recognition technology and infrared temperature measurement technology to realize rapid temperature detection while quickly and accurately identifying the identity information of people entering and leaving, and record and statistics of people passing through, providing more comprehensive data support for epidemic prevention and control

Keywords: face recognition; infrared thermometer; door security; intelligent prevention and control

(上接第88页)

Research status and insights on multimodal learning engagement measurement—based on CiteSpace visualization analysis

Zheng Chenhong^{*}, Zhang Haibo

(Institute of Education, Guizhou Normal University, Guiyang 550025, China)

Abstract: Multimodal technology breaks through the limitations of traditional single-modal data to realize accurate, scientific and comprehensive assessment of learning engagement. This paper takes 204 articles included in Web of Science database as samples, and uses CiteSpace to visualize and analyze the multimodal learning engagement assessment research, which comprehensively shows the main research countries, journals, highly cited papers, keyword clustering features, etc. In the future, we can further broaden the application scenarios, deepen the research topics, and emphasize the optimization of multimodal learning input assessment. In the future, multimodal learning engagement measurement can further broaden the application scenarios and deepen the research themes; pay attention to optimizing algorithms to improve the accuracy of data analysis; intervene in physiological data and strengthen the collaborative analysis of multimodal data.

Keywords: multimodal technology; learning engagement; knowledge graph

文章编号:1007-1423(2024)08-0095-05

DOI: 10.3969/j.issn.1007-1423.2024.08.016

基于时空特征交叉融合的网络流量预测

黄冬妹,宁 芊*

(四川大学电子信息学院,成都 610065)

摘要:精确的网络流量预测对网络资源合理分配、提高通信质量有着重要作用。然而网络流量存在着复杂的时空 依赖性,呈现高度非线性、复杂性,这给流量预测带来了困难。经过对现有的网络流量预测文献进行研究,分析网络 流量的时间性质和空间性质,提出时空特征交叉融合的网路流量预测模型STCFusion。并在三个公开的数据集 ABILENE、GEANT和CERNET进行充分的实验,实验结果表明提出的STCFusion有明显效果。

关键词:网络流量预测;自注意力机制;时空特征

0 引言

随着互联网的发展,通信网络中用户数量、 网络模式、终端设备等越来越多,使得网络结 构变得高度复杂,网络流量呈爆炸式增长^[1]。 这给网络资源管理带来了巨大挑战。精确的网 络流量预测能有效地帮助流量监控、流量拥塞 控制、流量控制等,是智能化、数字化网络资 源管理的发展方向重要技术之一^[2]。

网络流量预测方法大致可以分为两种:线 性预测和非线性预测^[3]。线性预测方法主要在 传统的序列预测任务或局部简单预测任务上, 而非线性的方法则主要是针对复杂的、基于数 据驱动的任务上。在过去,研究人员根据网络 流量的时间特性,利用ARIMA方法进行网络流 量预测^[4]。但随着网络结构越来越复杂,流量 预测精度已不能用简单的线性方法来解决。研 究人员开始使用循环神经网络(RNN)、长短期 记忆循环神经网络(LSTM)、门控循环单元 (GRU)对网络流量进行预测^[5]。根据网络流量 的空间特性,研究人员使用图卷积神经网络 (GCN)进行流量预测^[6]。

基于以往的研究,本文同时考虑网络流量的

时间性质和空间性质,提出一种时空特征交叉融 合的网络流量预测方法STCFusion。STCFusion主 要由GRU、GCN,以及两个基于自注意力机 制^[7]的时空融合模块和全连接预测模块组成。 GRU主要捕获网络流量的动态变化,即时间特 征;GCN主要捕获网络流量的拓扑特性,即空 间特征。基于自注意力机制的时空特征交叉融 合模块实现了时间特征和空间特征的交互、融 合,确保了网络流量的特征完整性,提高预测 精度;全连接预测模块使用全连接层进行网络 流量的预测。

1 问题描述

网络流量预测任务是根据历史的流量数据 来预测未来的流量变化的任务。

在通信网络中,网络拓扑结构具有多种形 式,例如星型拓扑、环形拓扑、网状拓扑、树 状拓扑和混合拓扑等^[8]。不同拓扑结构对应的 网络流量分布也不同,进行流量预测就需要考 虑到网络流量的空间性质。网络流量的时间性 质通常体现在特定时间内流量数据的变化趋势, 如网络流量的突发性、季节性、周期性等特

收稿日期: 2023-11-06 修稿日期: 2023-11-29

作者简介:黄冬妹(1998—),女,广西百色人,硕士,研究方向为深度学习、时空预测;*通信作者:宁芊(1969—),女, 四川成都人,博士,教授,研究方法为智能控制、物联网、水利信息化等,E-mail:ningq@scu.edu.cn

征^[9]。网络流量的时间性质和空间性质相互影响、相互作用,形成了复杂的时空依赖性。

将通信网络的拓扑结构表示为无向的图 G(V, E, A)。V代表网络拓扑中的节点集合;E代 表网络链路边的集合;A代表邻接矩阵,组成元 素为0和1,当网络中两个节点之间存在通信链 路表示为1,否则为0。使用源到目的的流量矩 阵 $X^{N \times N \times T'}$ 代表网络流量,N是网络拓扑中的节 点数,T'是历史流量时间总步数。且 $A \in \mathbf{R}^{N \times N}$ 。

因此可以将网络流量预测任务建模,具体 公式表示为

 $[X^{t+1}, X^{t+2}, \cdots, X^{t+T}] = f [G, (X^{t-T'-1}, X^{t-T'-2}, \cdots, X^{t})]$ (1)

其中: *f*(•)为本文的网络流量预测模型, *T*为预测的未来时间总步数。

2 研究方法

2.1 GRU

GRU是RNN的变体,常用于序列预测任务。GRU对RNN进行了改进,以解决传统RNN存在的梯度消失问题。

GRU模型可表示为

$$u_{t} = \delta \left(\boldsymbol{W}_{z} \cdot \left[x_{t}, h_{t-1} \right] + b_{z} \right)$$
(2)

$$r_{t} = \delta \left(\boldsymbol{W}_{r} \cdot \left[\boldsymbol{x}_{t}, \boldsymbol{h}_{t-1} \right] + \boldsymbol{b}_{r} \right)$$
(3)

$$c_{t} = \tanh\left(\boldsymbol{W}_{c} \cdot \left[\boldsymbol{x}_{t}, \boldsymbol{r}_{t} \odot \boldsymbol{h}_{t-1}\right] + \boldsymbol{b}_{c}\right)$$
(4)

$$h_{t} = u_{t} \cdot h_{t-1} + (1 - u_{t}) \cdot c_{t}$$
(5)

其中: u_t 为更新门, r_t 为重置门, c_t 为t时刻候选 隐藏状态, h_t 为t时刻的隐藏状态, h_{t-1} 为上一 时间步长的隐藏状态。 W_2 、 W_r 、 W_e 分别代表更 新层、重置层、候选隐藏层的权重矩阵; b_2 、 b_r 、 b_e 分别代表更新层、重置层、候选隐藏层的 偏置。

2.2 GCN

网络流量的拓扑结构用无向图来表示,使用GCN来提取网络流量的空间特征。GCN用来处理图数据结构,通过计算特征矩阵和邻接矩阵获得最后结果。特征矩阵包含网络流量特征,邻接矩阵包含网络拓扑信息,通过图卷积获得通信网络流量空间特征。

GCN具体计算公式如下:

$$GCN(\boldsymbol{X}, \boldsymbol{A}) = \delta \left(\hat{A} \cdot \operatorname{ReLU}(\hat{A} \boldsymbol{X} \boldsymbol{W}_0) \boldsymbol{W}_1 \right) \quad (6)$$

其中: $\hat{A} = D^{\frac{1}{2}}AD^{\frac{1}{2}}$, W_0 、 W_1 分别表示第0层、 第1层的权重矩阵。

2.3 自注意力机制

自注意力机制是用于处理序列到序列任务 的方法,常用于语言翻译。自注意力机制关注 输入序列之间的相关性,解决了在长序列信息 丢失、梯度消失问题。自注意力机制引入三个 矩阵查询矩阵 Q、键矩阵 K、值矩阵 V。其中 $Q = W_{Q}X$ 、 $K = W_{\kappa}X$ 、 $V = W_{\nu}X$, W_{Q} 、 W_{κ} 、 W_{ν} 分别是查询矩阵 Q、键矩阵 K、值矩阵 V的权重 矩阵。

自注意力机制的公式如下:

Attention
$$(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \operatorname{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{T}}{\sqrt{d_{k}}}\right)\boldsymbol{V}$$
 (7)

其中: d_k 是自注意力机制模型的维度,一般为 W_k 的维度。

2.4 总体模型

STCFusion主要由四部分组成: GRU、GCN、时空交叉融合模块和全连接预测模块。GRU提取时间特征,GCN提取空间特征,时空交叉融合模块获取时空依赖,全连接预测模块预测未来的网络流量。网络总体结构如图1所示。

首先,在历史*T*′时间步下,有流量网络矩阵(*X*^{*i*-*T*′-1},*X*^{*i*-*T*′-2},…,*X*^{*i*})和通信网络拓扑结构 图*G*(*V*,*E*,*A*)。

流量矩阵经过GRU提取时间特征:

 $F_{\text{time}} = GRU(X^{t}) = (1 - r_{t}) \odot h_{t-1} + r_{t} \odot c_{t}$ (8) GCN 提取空间特征:

$$\boldsymbol{F}_{\text{spatial}} = GCN(\boldsymbol{X}^{t}, \boldsymbol{A}) = \delta(\hat{\boldsymbol{A}} \cdot \text{ReLU}(\hat{\boldsymbol{A}}\boldsymbol{X}^{t}\boldsymbol{W}_{0})\boldsymbol{W}_{1})$$
(9)

时空交叉融合模块基于自注意力机制实现, 时间-空间特征信息TS-Attention以网络流量的 时间特征信息为查询矩阵,以网络流量的空间 特征信息为键矩阵和值矩阵。具体公式如下:

$$\boldsymbol{A}_{\text{TS}} = \text{softmax} \left(\frac{(\boldsymbol{W}_{Q} \boldsymbol{F}_{\text{time}}) (\boldsymbol{W}_{K} \boldsymbol{F}_{\text{spatial}})^{T}}{\sqrt{d_{k}}} \right) (\boldsymbol{W}_{V} \boldsymbol{F}_{\text{spatial}})$$
(10

空间-时间特征信息ST-Attention以网络流量

的空间特征信息为查询矩阵,以网络流量的时间 特征信息为键矩阵和值矩阵。具体公式如下:

$$\boldsymbol{A}_{\rm ST} = \operatorname{softmax}\left(\frac{\left(\boldsymbol{W}_{Q}\boldsymbol{F}_{\rm spatial}\right)\left(\boldsymbol{W}_{K}\boldsymbol{F}_{\rm time}\right)^{T}}{\sqrt{d_{k}}}\right)\left(\boldsymbol{W}_{V}\boldsymbol{F}_{\rm time}\right)$$
(11)

将得到的时间-空间特征信息和空间-时间 特征信息求和 $S = A_{TS} + A_{ST}$,得到完整的网络流 量特征。

最后使用全连接预测模块预测未来*T*时间 步的网络流量。具体公式如下:

$$\left[X^{t+1}, X^{t+2}, \cdots, X^{t+T}\right] = FC(S) = W_{FC}S + b_{FC}$$
(12)

其中: W_{FC}和b_{FC}分别为全连接层的权重矩阵和 偏置。



图 1 STCFusion 网络总体结构

3 实验结果

3.1 实验设计

本文的实验使用深度学习框架版本为PyTorch 1.10.2,编程语言版本为Python 3.6.13。使用的 操作系统为64位Ubuntu 16.04,GPU为NVIDIA GTX 3090。Batch_size设置为64,学习率为 0.0001。按照8:2的比例划分实验训练集和测试 集。本文使用了AdamW优化器和L1损失函数进 行梯度计算。假设 \hat{Y} 为预测值,Y为实际值,n为总的预测时间步数,L1损失函数具体公式如 下:

$$loss(\hat{Y}, Y) = \frac{1}{n} \sum_{1}^{n} |\hat{Y} - Y|$$
 (13)

本文使用的数据集为ABILENE、GEANT和 CERNET。每个数据集的详细信息见表1。

表 1 数据集概述

| 名称 | 节点数 | 边数 | 地区 |
|---------|-----|----|----|
| ABILENE | 12 | 30 | 北美 |
| GEANT | 23 | 38 | 欧洲 |
| CERNET | 14 | 32 | 中国 |

3.2 评价指标

本文实验使用平均绝对误差(MAE)、均方 根误差(RMSE)、准确性(ACCURACY)和可解释 方差分数(VAR)作为评价指标,以评估模型性 能。MAE和RMSE评估网络流量的预测值与真 实值之间的绝对差异,数值越小说明模型性能 越好。ACCURACY表示模型预测的准确性,数 值越大说明模型性能越好。VAR表示预测模型 能够解释预测的网络流量方差的比例,取值范 围通常在0到1之间,值越接近1表示模型性能 越好。

具体公式如下:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |Y_i - \hat{Y}_i|$$
 (14)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}$$
(15)

$$ACCURACY = 1 - \frac{\left\|Y - \hat{Y}\right\|_{F}}{\left\|Y\right\|_{F}}$$
(16)

$$VAR = 1 - \frac{var(Y - \hat{Y})}{var(Y)}$$
(17)

3.3 结果分析

本文选取的基线模型分为两种: 传统线性 方法和非线性方法。传统的线性方法有 ARIMA, 非线性方法为GRU、GCN、T-GCN^[10]。T-GCN 进行了时空特征融合,但没有使用到自注意力 机制。T-GCN使用GRU先提取时间特征,后用 GCN 提取空间特征,已达到时空特征融合目的。 实验预测未来15 min、30 min、45 min的网络流 量,实验结果见表2。由表2可知,相比于其他 基线模型, STCFusion的评价指标在三个数据集 上都得到了最好的效果。ARIMA的效果最差, 这可能是因为ARIMA无法捕获网络流量中非线 性信息,只捕获了网络流量的序列信息。GRU 和GCN 都是基于神经网络、非线性的方法,两 者都能捕获网络流量中非线性的信息。从表2可 以看出, GRU总体性能比GCN好, 这可能是因 为网络流量的时间性质大于空间性质,或可能 流量信息包含的空间特征相对于时间特征更难 获取,或者网络流量的时间信息比空间信息更 丰富。T-GCN 同时提取了网络流量的时间特征 和空间特征,比提取单一特征模型GRU和GCN 性能更好。说明进行网络流量预测,网络流量 的时间特征和空间特征都应考虑。引入了自注 意力机制的STCFusion,性能最好。这是因为时 空交叉融合模块实现了时间特征和空间特征的 交互、融合,提取了时间-空间特征和空间-时 间特征,不仅保证了网络流量特征完整性,而 且也在一定程度上避免了梯度消失。

将基于 GEANT、ABILENE、CERNET 数据

集的15min流量预测可视化,分别如图2、图3、 图4所示。由此可知,STCFusion相对于其他基 线模型的预测时延更低,且也最好地拟合了网 络流量的峰值和低谷值。说明STCFusion能更好 地预测突发流量。STCFusion整体网络流量变化 趋势最好地拟合了实际的网络流量变化趋势, 说明STCFusion的整体预测性能最好。

| Datasets | Algorithms | MAE (15min/30min/45min) | RMSE (15min/30min/45min) | ACCURACY (15min/30min/45min) | VAR (15min/30min/45min) | |
|----------|------------|----------------------------|-----------------------------|---------------------------------|----------------------------|--|
| | ARIMA | 0.02124/0.021241/0.02124 | 0.03798/0.03798/0.037982 | 0.10167/0.10117/0.10158 | 0.01036/0.00984/0.00997 | |
| | GRU | 0.00700/0.00772/0.007988 | 0.02407/0.02696/0.027532 | 0.64191/0.59891/0.59042 | 0.86022/0.82469/0.81712 | |
| GEANT | GCN | 0.00896/0.010231/0.01097 | 0.02436/0.02792/0.029784 | 0.63744/0.58446/0.55676 | 0.85670/0.81177/0.78585 | |
| | T-GCN | 0.00907/0.01025/0.009361 | 0.02107/0.02529/0.026728 | 0.68651/0.62371/0.60224 | 0.89824/0.85012/0.82840 | |
| | STCFusion | 0.00551/0.00722/0.00799 | 0.02002/0.02449/0.02636 | 0.70223/0.63567/0.60779 | 0.90352/0.85583/0.83318 | |
| ABILENE | ARIMA | 0.04683/0.046831/0.04683 | 0.07094/0.07094/0.070943 | 0.31916/0.31916/0.319153 | 0.00028/0.00028/0.00027 | |
| | GRU | 0.01909/0.021047/0.02115 | 0.03809/0.04112/0.042165 | 0.73336/0.71211/0.704843 | 0.87480/0.85403/0.84614 | |
| | GCN | 0.02013/0.026128/0.02702 | 0.03542/0.04675/0.047847 | 0.75200/0.67276/0.665066 | 0.89166/0.81136/0.80238 | |
| | T-GCN | 0.0165/0.019393/0.021214 | 0.03025/0.03575/0.038782 | 0.78822/0.74992/0.728524 | 0.92100/0.88982/0.87016 | |
| | STCFusion | 0.01378/0.01711/0.01899 | 0.0284/0.03472/0.03811 | 0.80118/0.75699/0.73329 | 0.9304/0.89605/0.87490 | |
| CERNET | ARIMA | 0.12913/0.129132/0.12916 | 0.15531/0.15531/0.15534 | 0.40861/0.40858/0.40848 | 0.00652/0.00648/0.00614 | |
| | GRU | 0.01961/0.027456/0.03302 | 0.03734/0.05093/0.05924 | 0.86406/0.81456/0.78434 | 0.96886/0.94206/0.92162 | |
| | GCN | 0.03416/0.044437/0.04615 | 0.04826/0.06516/0.06781 | 0.8243/0.76278/0.753123 | 0.94960/0.90672/0.89875 | |
| | T-GCN | 0.01910/0.028272/0.03386 | 0.02906/0.04668/0.05592 | 0.89221/0.83004/0.79640 | 0.98045/0.95133/0.93018 | |
| | STCFusion | 0.01321/0.02354/0.02939 | 0.02548/0.04459/0.05430 | 0.90717/0.83759/0.80222 | 0.98549/0.95562/0.93424 | |

| 表 2 | 15 min/30 | min/45 | min | 下的评价指标 |
|-----|-----------|--------|-----|--------|
|-----|-----------|--------|-----|--------|



图 2 15 minGEANT的流量变化



图 3 15 minABILENE 的流量变化



图 4 15 min CERNET 的流量变化

4 结语

本文从网络流量的时间性质和空间性质出 发,设计了时空特征交叉融合模型STCFusion。 其中GRU提取时间特征,GCN提取空间特征, 时空特征交叉融合模块则提取了时间-空间特征 和空间-时间特征。在数据集ABILENE、 CERNET和GEANT上进行了实验,实验结果表 明STCFusion比单一特征提取方式、T-GCN效果 更好。说明进行网络流量预测要同时考虑时间 和空间特征,并且引入注意力机制能有效融合 两种特征。综上,STCFusion在一定程度上提升 以往网络流量的预测效果,为网络流量预测任 务提供了一定的参考依据。

参考文献:

- [1] LEE J, LEE H. The computer-mediated communication network: exploring the linkage between the online community and social capital [J]. New Media & Society, 2010, 12(5):711-727.
- [2] HUSSAIN F, HASSAN S A, HUSSAIN R, et al. Machine learning for resource management in cellular

and IoT networks: potentials, current solutions, and open challenges [J]. IEEE Communications Surveys & Tutorials, 2020, 22(2):1251-1275.

- [3] TIAN Z, SONG P. A novel network traffic combination prediction model [J]. International Journal of Communication Systems, 2022, 35(7): e5097.
- [4] LANER M, SVOBODA P, RUPP M. Parsimonious fitting of long-range dependent network traffic using ARMA models [J]. IEEE Communications Letters, 2013,17(12):2368-2371.
- [5] FAN J, MU D, LIU Y. Research on network traffic prediction model based on neural network [C]//Proceedings of the 2019 2nd International Conference on Information Systems and Computer Aided Education(ICISCAE),2019:554-557.
- [6] GUI Y, WANG D, GUAN L, et al. Optical network traffic prediction based on graph convolutional neural networks [C] //Proceedings of the 2020 Opto-Electronics and Communications Conference (OECC), 2020:1-3.
- [7] CORDONNIER J B, LOUKAS A, JAGGI M. On the relationship between self-attention and convolutional layers[EB/OL]. arXiv:1911.03584,2019.
- [8] LANGA C, TARWIREYI P, ADIGUN M. Evaluating named data networking forwarding strategies in different IoT topologies [C] //Proceedings of the 2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems(icABCD), 2020:1-6.
- [9] TEDJOPURNOMO D A, BAO Z, ZHENG B, et al. A survey on modern deep neural network for traffic prediction: trends, methods and challenges [J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(4):1544-1561.
- [10] ZHAO L, SONG Y, ZHANG C, et al. T-GCN: a temporal graph convolutional network for traffic prediction [J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(9):3848-3858.

Network traffic prediction based on spatial-temporal features cross fusion

Huang Dongmei, Ning Qian*

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

Abstract: Accurate network traffic prediction plays an important role in the rational allocation of network resources and the improvement of communication quality. However, network traffic has complex spatial-temporal dependencies, presenting a high degree of nonlinearity and complexity, which brings difficulties to traffic prediction. After studying existing literature on network traffic prediction, analyzing the temporal and spatial properties of network traffic, a network traffic prediction model STCFusion based on the cross fusion of spatial-temporal features is proposed. And sufficient experiments were conducted on three publicly available datasets, ABILENE, GEANT, and CERNET, and the experimental results showed that the proposed STCFusion had significant effects.

Keywords: network traffic prediction; self-attention mechanism; spatial-temporal features

文章编号:1007-1423(2024)08-0100-04

DOI: 10.3969/j.issn.1007-1423.2024.08.017

基于神经网络的文本情感分类模型研究

司靖梓1,邢建川1,2*,肖 鑫1

(1. 西藏大学信息科学技术学院, 拉萨 850000; 2. 电子科技大学计算机科学与工程学院, 成都 610000)

摘要:情感分类作为近年来自然语言处理领域的热门研究方向,旨在识别文本中的情感态度,如积极、消极或者 中立等,对社交媒体、新闻、评论和用户反馈等大量文本数据进行挖掘并分析其情感极性对于研究者和政府来讲具有 十分重要的意义。传统的情感分类算法通常使用基于统计的特征提取方法,如词袋模型,再结合机器学习算法,如支 持向量机(SVM)和朴素贝叶斯分类器等来进行分类。在基于对神经网络的研究下实现文本情感分析,对文本数据集进 行预处理操作后建立情感分析模型,使用Keras框架搭建循环神经网络以识别情感倾向,定义相关函数后进行模型训练, 并采用一系列方法指标来进行模型评估检验模型性能,比较传统机器学习算法提高了情感分析的精度和效率。

关键词: 自然语言处理; 循环神经网络; 词向量训练; 深度学习框架

0 引言

社交媒体、新闻、评论及用户反馈等大量 文本数据的产生使信息得到迅速的传播与交 互^[1],理解这些用户的情感并利用这些信息来 进行分析预测对于社会舆情的处理以及商业销 售等都具有重大的意义,因此情感分类逐渐成 为自然语言处理领域的热门研究方向。

纵观情感分类的研究现状和发展动态, 尧欢 欢^[2]等提出了一种融合预训练语言模型(ERNIE) 和深层金字塔神经网络结构(DPCNN)/双向门控 循环单元-注意力机制(BiGRU-Attention)的双通 道文本情感分类模型。Bae^[3]等使用一种基于词 汇的方法,在Twitter数据集上进行情感分类, 以了解公众对名人微博的情绪。Kolog^[4]等研究 了关于学生生活的情感分析和社会现象。他们 通过分发问卷收集了学生生活的数据集,使用 K-means算法对学生生活数据进行聚类划分。此 外,他们还比较了几种算法的分类结果,包括 序列最小优化、多项朴素贝叶斯和决策树。

国外的研究人员对文本的情感分析研究开 始较早,但是由于英文文本和中文文本在语法

收稿日期: 2023-11-03 修稿日期: 2023-11-27

基金项目:大学生创新训练项目(S202310694051)

结构上存在极大的差异性以及中文文本的复杂 性、多样性,再结合当前的文本信息大多短小 精悍数据稀疏,传统的情感词典是不能直接套 用的。

本文基于神经网络实现文本情感分析,再 对公开数据集进行收集、整理以及分析,包括 清洗、分词、去停用词等预处理操作后,建立 情感分析模型,使用Keras深度学习框架搭建神 经网络,并采用精准率、召回率、F1值来评价 模型性能。

1 实验准备

1.1 实验数据来源

使用网络爬虫技术获取某社交平台评论信息约12万条,将获取到的数据保存在本地CSV格式文件中,作为本次研究的数据集。对文本数据集进行标注,用数字代表情感倾向,1为积极的情感,0为消极的情感。"####"用作特殊符号,将所标注的数字和后面的文本数据隔开,具体示例如下:

作者简介:司靖梓(2003—),女,山东德州人,在读本科,研究方向为自然语言处理、人工智能、实时计算机应用、云计算与大数据;*通信作者:邢建川(1974—),男,河南漯河人,博士,副教授,研究方向为云计算与大数据、物联网、人工智能、 复杂性科学、信息科学与其他学科交叉研究等,E-mail:xingjianchuan@sina.com; 肖鑫(2003—),男,湖南邵阳人,在读本科, 研究方向为软件工程

1 #### [送花]//@类四而已-: 生活顺利宝 我们都要 好好的!!!

0 #### [挠头] [挠头] //清蒸小鱼:太罪恶了吧,应 该严惩!

1.2 数据预处理

1.2.1 数据清洗

因为是中文的文本情感分类,所以在清洗 的时候将中文文本数据留下,其余的文字符号 删掉。定义相关函数用于判断一个unicode是否 是汉字并进行保留汉字的操作,对文本数据集 进行清洗,包括删除中文标点符号等留下中文 文本数据,对不在标注中的样本以及其他数据 做剔除处理。采用正则表达式匹配特殊字符和 无用信息并将其删除使文本更规整、干净。

停用词在文本情感分类任务中的研究极其 有限,因此本次研究考虑将其删除,基于百度 停用词表以及此社交平台评论的特点^[5],在停 用词库中额外增加了一些与文本情感分类关系 较小的词汇^[6],例如"展开"等。对每一行进 行逐行的读取并返回读取到的内容。定义一个 处理数据集的函数并利用jieba分词工具和去停 用词库进行对比等完成去停用词以及删除异常 数据例如空白文本的操作。定义两个数组备用, 一个用于存放内容,一个用于存放标签。如果 文本内容<=0,那进行剔除操作。原始数据进行 分词、过滤后以数组的形式保留下来,至此数 据清洗完成。具体示例如下:

生活顺利呀宝我们都要好好的→['生活','顺 利','我们','都','要','好','好']

1.2.2 数据集划分及词向量化

基于已标注数据划分的数据集将大约14万 条数据进行划分:60%作为训练集,用于训练 模型以及确定参数;20%作为验证集,用于确 定网络结构以及调整模型的超参数;剩余20% 用作训练集,用于检验模型的泛化能力。对数 据集的标签数据进行 one-hot 编码,例如标签1 对应的编码为[1,0],标签0对应的编码为 [0,1],然后对文本进行词向量化。根据词频 使用 Tokenizer 对词组进行编码,词频越大,编 号越小。词频即文本中每个词出现的频率,频率 越高的词越能反映文本的主题,计算公式如下:

式中,ω指待计算词语。

根据循环神经网络的特点将每个序列调整 为相同的长度(用0填充),输入到神经网络中。

2 实验过程

2.1 模型搭建

神经网络一般包括输入层、隐藏层、承接 层和输出层。整个网络的运行过程为序列数据 由输入层传入网络,隐藏层对数据进行信息提 取、降维、非线性映射等操作,所获得信息由 承接层将其进行延时操作与下一时刻的输入值 共同输入网络参与下一时刻数据处理,沿序列演 进方向如此往复行进。输出层输出的信息必将是 当前时刻和之前所有历史输入值的综合信息,从 而使循环神经网络具有记忆功能。前向传播和随 时间反向传播是一般循环神经网络的两个训练阶 段。循环神经网络的一般训练过程为:①初始化 网络参数;②输入数据由隐藏层处理后经前向 传播得到输出值,同时将输出值传入承接层与 下一时刻数据再次输入网络进行数据处理,沿 序列演进方向如此往复行进;③计算输出值与 实际值之间的差值;④当误差大于阈值时,求 得网络各参数对于总误差的梯度,并参考梯度 的大小对各个参数进行调整; ⑤返回步骤②继 续训练,直到误差小于等于阈值时,网络训练 过程结束[7]。

逐行进行模型搭建,首先进行 Embedding嵌 入,然后使用简单的循环神经网络,之后再进 行全连接和50%的随机丢弃,随后重复进行一 次全连接。因为数据分为积极和消极两种,所 以全连接最后是输出两类,使用 softmax 输出概 率。定义损失函数和优化方法,至此模型搭建 完成,代码如下:

inputs = Input(name='inputs, shape=[max_len]) ## Embedding(词汇表大小,batch大小,每个新闻的 词长)

layer = Embedding(max_words+1,128,input_length= max_len)(inputs)

layer = SimpleRNN(128)(layer)

layer = Dense(128, activation-"relu", name="FC1")

(layer) layer = Dropout(0.5)(layer) layer = Dense(2, activation="softmax", name="FC2") (layer) model = Model(inputs=inputs, outputs=layer) model.summary() model.compile(loss="categorical_crossentropy", optimizer="RMSprop", metrics=["accuracy"])



图1 RNN结构图

2.2 模型训练

对搭建好的模型进行模型训练,将batch_size 设置为128,epochs设置为10,即每次训练 128个数据,共训练10轮,最后将训练的日志 用TensorBoard进行保存。根据训练过程统计得 出准确率在逐步增加,在第6轮训练时,准确率 达到90%以上;损失率也在逐步降低。具体数 值如表1所示。

表 1 训练过程中模型准确率和损失率

| epoch | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| accuracy | 0.870 | 0.884 | 0.887 | 0.898 | 0.900 | 0.905 | 0.905 | 0.910 | 0.915 | 0.920 |
| loss | 0.300 | 0.280 | 0.265 | 0.255 | 0.246 | 0.240 | 0.235 | 0.230 | 0.225 | 0.200 |

2.3 模型评估

以精准率P(Precision)、召回率R(Recall)、 F1值作为相关指标来评价模型性能。使用已经 搭建好的模型来预测测试集,输出一个测试的 标签。以实际标签进行一个对比,输入到混淆 矩阵里,取得精准率、召回率以及F1值,对其 进行输出,最后输出混淆矩阵。

精准率、召回率、F1值计算公式如下:

$$R = \frac{ 正确预测为该类别的样本数}{ 应该预测为该类别的样本数}$$
(3)

$$F1 = \frac{2*P*R}{P+R} \tag{4}$$

其中: F1值是基于精准率与召回率的调和平均 定义的,它能综合考虑分类器的准确率和召回 率,因此F1值越高表示分类器的性能越好。

混淆矩阵的对角线表示预测的准确值,其 余位置则表示预测错误。最后得出准确率为 0.887,精准率为0.88,召回率为0.899,F1值为 0.884。说明模型具有一定的性能。

3 结语

循环神经网络是递归神经网络中的一类特殊类型,该模型节点均采用链式连接,而且沿着序列的演进方向进行递归。该网络的记忆功能可有效地将历史信息与当前信息进行综合处理。一般神经网络都是同时对输入数据进行处理的,无法高效地处理具有序列特征的数据。 RNN 网络结构非常适合处理序列数据,可从序列输入数据中提取有效的特征信息。RNN 在时序分析、自然语言处理、机器翻译以及语音识别等领域都有着非凡的表现。

本文在针对情感分类模型的研究中比较传 统机器学习算法的差异,通过情感词典的扩充 构建以及一种改进的神经网络模型实现了一种 性能较好的文本情感分类算法,增强了模型的 准确性、鲁棒性以及泛化性等表现,有效利用 文本内在结构提高了情感分析的精度和效率。

参考文献:

- [1] 刘园园.基于图神经网络的方面级情感分析算法研 究与应用[D].兰州:西北师范大学,2023.
- [2] 尧欢欢,朱小栋.基于ERNIE和融合双通道特征的 文本情感分类模型[J].软件工程,2023,26(2): 33-39.
- [3] BAE Y, LEE H. Sentiment analysis of twitter audiences: measuring the positive or negative influence of popular twitterers [J]. Journal of the American Society for Information Science and Technology, 2012, 63 (12):2521-2535.
- [4] KOLOG E A, MONTERO C S, TOIVONEN T. Using machine learning for sentiment and social influence analysis in text[C]//Proceedings of the International Conference on Information technology & Systems, Springer, Cham, 2018.
- [5] 王伟贤.基于情感词典和深度学习的微博文本情感 分析[D].扬州:扬州大学,2023.
- [6] 崔连超. 互联网评论文本情感分析研究[D]. 济南: 山东大学,2016.
- [7] 黄东瑞,毛克彪,郭中华,等.几种神经网络经典模型综述[J].高技术通讯,2023,33(8):860-871.

(下转第107页)

文章编号:1007-1423(2024)08-0103-05

DOI: 10.3969/j.issn.1007-1423.2024.08.018

面向用水结构研究领域的多阶段实体关系联合抽取方法

陶天然*

(华北水利水电大学信息工程学院,郑州 450046)

摘要:以往的知识抽取模型忽略了实体关系间的内在语义关联,并且在处理具有复杂关系的用水数据集时会产生 大量的冗余信息。针对以上问题,提出一种融合语义信息的实体关系联合抽取模型。模型包括三个阶段:第一阶段, 将经过BERT-wwm编码后的文本信息投影到关系检测空间中,过滤掉关系集合中的冗余数据;第二阶段,利用多头注 意力机制将关系信息融合进文本编码,获取对应关系下的头实体和尾实体集合;第三阶段,引入融合上下文语义信息 的实体相关矩阵,完成对三元组的准确提取。实验结果表明,所设计的模型在用水结构研究数据集上取得了较好的实 体关系抽取效果。

关键词:用水结构研究;联合抽取;多阶段抽取;语义信息;知识图谱

0 引言

在建设节水型社会的大环境下,用水结构 作为与水足迹、水资源管理一体的综合性问题, 是重点研究对象。目前,用水结构研究涉及到 的数据多来源于水利部、流域机构以及地方水 利部门,存在信息资源分散、数据共享难的问 题。利用知识图谱技术[1-2]对数据进行建模、提 取和存储,对实现领域内海量多源异构数据的 集成管理具有重要意义。实体关系抽取作为将 海量数据知识化的关键,是构建知识图谱的核 心任务之一^[3]。当前通用领域的实体关系抽取 技术发展迅速。但现有方法[47]在解码阶段通常 将实体识别任务转化为对应于预定义关系集合 的标签分类或序列标记问题, 会产生大量对应 于错误关系的冗余实体,不适用于处理关系复 杂型的用水数据。此外,这些方法所设计的三 元组抽取策略对上下文语义信息利用不足,使 得其无法解决汉语歧义多的问题。

针对上述挑战,本文设计了一种多阶段的 实体关系联合抽取模型。通过潜在关系提取模 块,实现对冗余关系的过滤,使模型适用于具 有复杂关系的数据集。在实体抽取模块,引入 特定于关系的多头注意力机制,将候选关系 与句子编码相融合,获取实体和关系之间的 内在语义关联,实现对句子中实体的准确识 别。此外,在模型中使用双向长短期记忆网 络(bi-directional long short-term memory, BiLSTM) 进一步挖掘词向量间的语义信息,并构建实体 相关矩阵,对歧义三元组进行剪枝,使模型能 够适应更复杂的语境,实现了对用水结构研究 数据集中三元组的准确抽取。

1 方法

本文模型主要包括四个模块:文本序列编 码器、潜在关系提取模块、实体识别组件以及 实体匹配模块。图1给出了本文模型的总体架 构,本节将详细介绍各模块的实现方法。

1.1 文本序列编码器

对于给定长度为n的文本序列 $X = [w_1, w_2, ..., w_n]$,需要先将X转换为BERT-wwm模型^[8]输入 需要的嵌入向量 $T = [t_1, t_2, ..., t_n]$,该向量由词 嵌入 W_T 、分割嵌入 W_S 和位置嵌入 W_P 相加而得, 计算式如公式(1)所示。然后,将嵌入向量T输 入 BERT-wwm进行编码,得到富含语义信息的

收稿日期: 2023-11-03 修稿日期: 2023-11-29

作者简介:*通信作者:陶天然(1998—),女,河南商丘人,硕士,研究方向为知识图谱,E-mail:taotianrans@163.com



图 1 本文模型的总体架构

输出序列 $H = [h_1, h_2, \dots, h_n]$,其计算式如式(2) 所示:

$$\boldsymbol{T} = \boldsymbol{W}_{\mathrm{T}} + \boldsymbol{W}_{\mathrm{S}} + \boldsymbol{W}_{\mathrm{P}} \tag{1}$$

$$H = BERT(T) \tag{2}$$

其中: $h_i \in \mathbf{R}^d$, $d \neq \text{BERT-wwm}$ 模型的嵌入维度。

1.2 潜在关系提取

首先,对输出序列*H*进行平均池化,得到 句子嵌入的压缩表示:

$$H^{\text{avg}} = avgpool(H) \tag{3}$$

其中: $avgpool(\cdot)$ 表示平均池化操作, $H^{avg} \in \mathbb{R}^{d \times 1}$ 。

其次,将H^{avg}投影到关系检测空间中,从而 将关系预测任务转换为多标签二值分类,得到 句子中包含的所有潜在关系。具体公式如下:

$$P_{\rm rel} = \sigma(W_r H^{\rm avg} + b_r) \tag{4}$$

其中: $\sigma(\cdot)$ 表示 sigmoid 函数, r表示关系类别的 总数。

1.3 实体识别

为了提升实体识别效果,本文利用关系感 知注意力使模型能够根据特定的关系信息自适 应地关注不同部分的文本,从而更准确地捕捉 实体的位置和边界。具体公式如下:

$$Q_j = W_j^q H + b_j^q \tag{5}$$

$$\boldsymbol{K}_{j}^{i} = \boldsymbol{W}_{j}^{k} \boldsymbol{e}_{i} + \boldsymbol{b}_{j}^{k}$$
(6)

$$\boldsymbol{V}_{i}^{i} = \boldsymbol{W}_{i}^{v}\boldsymbol{e}_{i} + \boldsymbol{b}_{i}^{v} \tag{7}$$

$$head_j^i = \text{Softmax}\left(\frac{Q_j K_j^{i^{\prime}}}{\sqrt{d}}\right) V_j^i$$
 (8)

 $r_i = Concat(head_1^i, head_2^i, \dots, head_h^i)W^o$ (9) e_i 表示关系向量,其中 $i \in \{1, 2, \dots, m\}$, m表示 句子中潜在关系的个数,h表示注意力头的数 量。

然后,将处理后的关系信息与词向量相结 合,生成基于关系的句子表示。最后,利用两 个三元分类器分别预测句子中头实体和尾实体 的开始位置、内部位置和外部位置。具体公式 如下:

$$\boldsymbol{h}_{i,i} = Concat(\boldsymbol{h}_i, r_i) \tag{10}$$

$$P_{i,j}^{\text{sub}} = \boldsymbol{\sigma}(W_s \boldsymbol{h}_j^{r_i} + b_s) \tag{11}$$

$$P_{i,j}^{\text{obj}} = \boldsymbol{\sigma}(W_o \boldsymbol{h}_j^{r_i} + b_o) \tag{12}$$

 $h_i \in \mathbb{R}^d$ 表示句子中的词向量,其中 $j \in \{1, 2, \dots, n\}$,
n表示句子长度, $P_{i,j}^{\text{sub}}$, $P_{i,j}^{\text{bbj}} \in \mathbf{R}^3$ 中的三个分量, 分别表示句子中的第j个单词被预测为实体开始 位置、内部位置或外部位置的概率。

1.4 实体匹配

利用实体识别组件得到了特定于关系的头实体集合和尾实体集合后,需要进一步实现集合中头实体和尾实体的匹配。首先,利用BiLSTM将上下文语义信息编码进句子的词嵌入中,这有助于更准确地判断头尾实体之间的联系。具体公式如下:

$$\vec{H} = \overrightarrow{LSTM} (H) \tag{13}$$

$$\dot{H} = \overleftarrow{LSTM} (H) \tag{14}$$

$$W = Concat \left[\vec{H}, \vec{H} \right]$$
(15)

其中: $\vec{H} \in \mathbf{R}^{n \times \frac{d}{2}}$ 为由前向后的上文语义信息, $\tilde{H} \in \mathbf{R}^{n \times \frac{d}{2}}$ 为由后向前的下文信息。其次,将句 子序列中的词嵌入进行两两组合,并将其输入 全连接层计算出实体间的匹配分数,得到实体 相关矩阵。最后,检查实体矩阵中每个实体对 对应的分数,如果该值超过阈值 λ_e ,则在最终 结果中保留该三元组,否则将其过滤掉。具体 公式如下:

 $F_{i,j} = \sigma (WConcat(\boldsymbol{w}_i, \boldsymbol{w}_j) + b)$ (16) 其中: $\boldsymbol{w}_i, \boldsymbol{w}_j \in \mathbf{R}^d$ 表示经过 BiLSTM 编码后的词 向量。

2 实验

2.1 数据集

本文面向用水结构研究领域涉及到的政策 文件、数据信息和水利基础知识,选取2012— 2021年的《河南省水资源公报》《河南省用水条 例》以及水利部官网中的水知识数据和水利术语 作为原始语料。共标注了原始语料中的5665条句 子,形成了一个包含17549个三元组、285个关 系类型和9个实体类型的用水结构研究领域数 据集。

2.2 评价指标

使用准确率P、召回率R和F1值作为模型 的评价指标。并且,只有当关系类型和实体对 都与真实三元组相匹配时,才认为该三元组被 正确识别。评价指标具体的计算公式如下:

$$P = \frac{N_{\text{correct}}}{N_{\text{corr}}} \tag{17}$$

$$R = \frac{N_{\text{correct}}}{N_{\text{all}}} \tag{18}$$

$$F1 = \frac{2PR}{P+R} \tag{19}$$

其中: N_{correct}表示模型抽取出的正确三元组数量, N_{out}表示模型抽取出的三元组总数, N_{all}表示数 据集中所有三元组的总数。

2.3 实验结果及分析

为了验证本文模型的有效性,本文将其与 六个表现优秀的基线模型进行了比较:Novel-Tagging将实体关系抽取统一为一个序列标注任 务^[9];CopyRe为一种基于复制机制的联合关系 抽取模型^[10];SPointer是一种利用双指针模块的 端到端的联合抽取模型^[11];CASREL则是一种 用于解决关系重叠问题的级联式二元标注框 架^[5];TPLinker将实体关系联合抽取任务转化为 令牌对链接问题^[4];PRGC模型探索了一个全局 对应矩阵来对齐实体^[12]。

本文模型与基线模型在用水结构数据集上 的性能对比结果见表1。可以看出,本文模型在 准确率、召回率和F1值三个评测指标上均显著 优于 NovelTagging、CopyRe 和 SPointer, 证明了 模型的优越性。与CasRel、PRGC模型相比, 本文模型的查准率分别提升了4.26%和2.08%, 召回率分别提升了2.27%和2.02%,这是因为 CasRel 模型回避了对关系信息的抽取和利用, 导致其难以处理拥有大量复杂关系的领域数据 集。而PRGC模型虽然将关系信息和句子编码 进行了融合,但忽略了实体和关系间的潜在语 义联系,在实际构建用水结构领域知识图谱的 应用中表现不佳。与TPlinker模型相比,本文模 型的查准率提高了5.25%,而召回率有所降低。 这是因为本文模型对抽取到的实体关系三元组 集合进行了剪枝操作,这虽然大大减少了集合 中的错误三元组,但是也会误删部分正确的实 体关系三元组,因而模型的查准率较高,而召 回率偏低。从整体看,本文模型在处理用水结 构研究领域涉及到的文本数据时有更优异的表 现,提升了知识抽取的准确率。

| 利に出来 | | | | |
|--------------|-------|-------|-------|--|
| 模型 | Р | R | F1 | |
| NovelTagging | 0.295 | 0.209 | 0.245 | |
| CopyRe | 0.424 | 0.412 | 0.418 | |
| SPointer | 0.347 | 0.335 | 0.341 | |
| CASREL | 0.846 | 0.839 | 0.842 | |
| Tplinker | 0.838 | 0.874 | 0.856 | |
| PRGC | 0.864 | 0.841 | 0.852 | |
| 本文模型 | 0.882 | 0.858 | 0.870 | |

表 1 本文模型与基线模型在 WUSR 数据集上的

为了进一步验证模型对复杂数据的处理能 力,本文设置N为语句中存在三元组的个数, 将数据集划分为N=1,N=2,N=3,N=4,N≥5 共五种类型,并在五种不同类型的数据集上分 别进行了训练。实验结果见表2。可以看出,在 这五类数据上,本文模型均取得了最好的结果, 这表明在充分整合了文本中的潜在语义信息后, 本文模型能够适应更复杂的情况,对于包含有 多个三元组的语句的处理能力更强。

表 2 本文模型与基线模型在不同三元组数量数据集上 F1 值的对比结果

| 模型 | N=1 | N=2 | N=3 | <i>N</i> =4 | <i>N</i> ≥5 |
|----------|-------|-------|-------|-------------|-------------|
| CopyRe | 0.481 | 0.494 | 0.421 | 0.409 | 0.271 |
| CASREL | 0.842 | 0.848 | 0.861 | 0.863 | 0.774 |
| TPLinker | 0.844 | 0.860 | 0.862 | 0.878 | 0.814 |
| PRGC | 0.836 | 0.858 | 0.865 | 0.873 | 0.827 |
| 本文模型 | 0.855 | 0.869 | 0.884 | 0.889 | 0.852 |

3 结语

本文提出了一种面向用水结构研究领域的 多阶段实体关系联合抽取模型,模型使用 BERT-wwm作为文本序列编码器,整体上包括 关系提取、实体识别以及实体匹配三个阶段。 模型的整体架构适用于关系复杂型数据集。此 外,模型利用关系感知的多头注意力将实体信 息和关系信息进行合理融合,从而获取到关系 嵌入与文本编码间的关联信息,提升了实体识 别效果;引入BiLSTM进一步挖掘文本序列中的 深层上下文语义信息,并将其用于实体匹配模 块,提升了模型抽取三元组的准确率。在用水 结构研究领域数据集上的实验结果表明,本文 提出的模型相较于基线模型,在面向用水结构 研究领域的知识抽取任务中达到了更好的应用 效果,这表明其适用于处理这类具有密集实体 信息和复杂关系集合的领域数据。

参考文献:

- [1] YAN J H, WANG C Y, CHENG W L, et al. A retrospective of knowledge graphs[J]. Frontiers of Computer Science, 2018, 12(1):55-74.
- [2] PENG C Y, XIA F, NASERIPARSA M, et al. Knowledge graphs: opportunities and challenges [J]. Artificial Intelligence review, 2023, 56: 13071-13102.
- [3] JI S X, PAN S R, CAMBRIA E, et al. A survey on knowledge graphs: representation, acquisition, and applications[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(2):494-514.
- [4] WANG Y C, YU B W, ZHANG Y Y, et al. TPLinker: single-stage joint extraction of entities and relations through token pair linking [C] //Proceedings of the 28th International Conference on Computational Linguistics, 2020:1572-1582.
- [5] WEI Z P, SU JL, WANG Y, et al. A novel cascade binary tagging framework for relational triple extraction [C] // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020:1476-1488.
- [6] SHANG Y M, HUANG H Y, MAO X L. OneRel: joint entity and relation extraction with one module in one step [C] //Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 26 (10) : 11285-11293.
- [7] LIU X, DU W R, WANG X Y, et al. A mutually-exclusive binary cross tagging framework for joint extraction of entities and relations [J]. PLoS ONE, 2022, 17(1):e0260426.
- [8] CUI Y M, CHE W X, LIU T, et al. Pre-training with whole word masking for Chinese BERT [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2021, 29:3504-3514.
- [9] ZHENG S C, WANG F, BAO H Y, et al. Joint extraction of entities and relations based on a novel tagging scheme [C] //Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics(Volume 1:Long Papers), 2017:1227-1236.
- [10] ZENG X R, ZENG D J, HE S Z, et al. Extracting

relational facts by an end-to-end neural model with copy mechanism [C] // Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics(Volume 1:Long Papers), 2018;506-514.

- [11] BAI C Y, PAN L M, LUO S L, et al. Joint extraction of entities and relations by a novel end-to-end model with a double-pointer module [J]. Neurocomputing, 2020, 377: 325-333.
- [12] ZHENG H Y, WEN R, CHEN X, et al. PRGC:potential relation and global correspondence based joint relational triple extraction [C] //Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1:Long Papers), 2021:6225-6235.

Multi-stage entity and relation joint extraction method for water usage structure research field

Tao Tianran^{*}

(School of Information Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450046, China)

Abstract: Previous knowledge extraction models overlooked the intrinsic semantic relationships between entities and, in parallel, resulted in a significant amount of redundant information when dealing with water usage datasets with complex relationships. To address these issues, this paper proposes a novel entity-relation joint extraction model that integrates semantic information. The model consists of three stages: In the first stage, text information encoded through BERT-wwm is projected into the relation detection space to filter out redundant data in the relation set. In the second stage, a multi-head attention mechanism is employed to fuse relation information into text encoding, obtaining sets of head and tail entities corresponding to each relation. In the third stage, an entity-related matrix incorporating fused contextual semantic information is introduced to accurately extract triplets. Experimental results demonstrate that the designed model achieves a notable entity-relation extraction performance on a water usage structure research dataset.

Keywords: water usage structure research; joint extraction; multi-stage extraction; semantic information; knowledge graph

(上接第102页)

Research on text sentiment analysis based on neural network

Si Jingzi¹, Xing Jianchuan^{1,2*}, Xiao Xin¹

(1. School of Information Science and Technology, Tibet University, Lhasa 850000, China;2. School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610000, China)

Abstract: As a popular research direction in the field of natural language processing in recent years, sentiment classification aims to identify the emotional tendency in text, such as positive, negative or neutral, etc., and it is of great significance for researchers and governments to mine and analyze the emotional polarity of a large amount of text data such as social media, news, comments and user feedback. Traditional sentiment classification algorithms usually use statistical-based feature extraction methods, such as bag-of-words models, combined with machine learning algorithms, such as support vector machines (SVMs) and naïve Bayes classifiers. Based on the research of neural network, the text sentiment analysis is realized, and the sentiment analysis model is established after the text data set is preprocessed, use the Keras framework to build a recurrent neural network to identify emotional tendencies, after defining the relevant functions, model training is performed, and use a series of methods to test the model performance. Compared with traditional machine learning algorithms, the accuracy and efficiency of sentiment analysis are improved.

Keywords: NLP; recurrent neural networks; word vector training; deep learning frameworks

文章编号:1007-1423(2024)08-0108-04

DOI: 10.3969/j.issn.1007-1423.2024.08.019

基于S-G滤波算法的航空器着陆阶段滑行轨迹修正研究

何 昕,王经义*

(中国民用航空飞行学院空中交通管理学院, 广汉 618307)

摘要:对航空器着陆阶段运行轨迹分析是客观评价飞行员技术水平与航司运行品质的重要手段,在保障航空器安 全高效运行等方面具有重要意义。通过对QAR数据中经纬度参数在地图软件中打点分析发现,航空器在跑道上减速滑 行阶段轨迹出现明显的曲折、错位与偏移现象。采用S-G滤波算法,构建航空器着陆阶段轨迹修正模型,对航迹进行修 正处理。结果表明,S-G滤波算法修正模型可对原始航迹进行良好修正。

关键词:航迹修正; S-G滤波; QAR数据

0 引言

QAR (Quick Access Recorder)数据是航空器 在实际运行中由机载快速存取记录器记录的超 过 600小时原始飞行数据,其包含航空器的高 度、地速、空速、经纬度、加速度、驾驶员操作 数据、空地电门等上千种飞行参数,涵盖了航空 器运行的各个阶段^[1]。航空公司通常依据相关记 录数据开展飞行技术检查、安全评估、不安全事 件调查等工作^[24],以此发现机组操作、发动机工 况以及航空器性能等方面的问题,从而避免事故 隐患,保障飞行安全,提高运营效率。

然而由于机载设备误差、卫星定位误差、 系统延时误差等因素,通过地图软件打点分析 发现,航空器着陆阶段航迹存在一定的曲折、 错位与偏移现象。为保证对航空器运行轨迹的 准确分析,对QAR数据中经纬度参数所形成 的航迹进行误差修正处理成为一项重要工作。 张召悦等^[5]通过建立飞行计算与航迹识别模 型,结合实际运行数据通过程序仿真,实现 了对ADS-B航迹的修正。李楠等^[6]针对由速度 引起的航空器轨迹聚类的轨迹点不匹配的问 题,基于航空器速度的轨迹点采集方法提出速 度修正系数,有效改进了航空器轨迹的聚类效 果。徐开俊等^[7]利用三角窗函数对机载SD卡所记 录的航迹数据进行分段处理,消除了GPS单点定 位误差,实现对飞行员飞行技术的精确评估。 Averyanova等^[8]考虑到起飞、着陆以及航路的天 气因素,结合机载雷达系统提供的气象资料, 实现航空器飞行轨迹的短期修正。

国内对于航迹修正的研究主要集中于航空 器空中运行阶段,对于着陆阶段的研究较少。 着陆阶段作为整个运行过程中最易发生不安全 事件的阶段,对该过程航迹进行修正处理,可 对机组操作水平进行更客观合理的评价,因此 存在较高的研究意义。针对以上问题,本文首 先对航空器着陆过程分阶段分析,再利用S-G (Savitzky-Golay)滤波算法对航空器着陆阶段运行 轨迹进行修正处理,减少各类误差对着陆阶段航 迹的影响,以确定更贴近于实际的航迹数据。

1 着陆过程分析

航空器着陆过程可分为五个阶段,分别为 拉平阶段(Flare)、接地阶段(Touchdown)、制动 阶段(Braking)、滑行阶段(Coasting)和退出阶段

收稿日期: 2023-11-07 修稿日期: 2024-03-22 基金项目: 民航局空管局委托项目(H2021-61)

作者简介:何昕(1979—),女,浙江浦江人,硕士,教授,研究方向为空中交通管理、空管运行新技术、通用航空飞行技术与安全;*通信作者:王经义(1999—),男,四川德阳人,硕士研究生,研究方向为空管运行新技术、空中交通管理, E-mail:1535711082@qq.com

(Exists),如图1所示。



图 1 航空器着陆各阶段示意图

(1) 拉平阶段: 航空器从跑道头离地 50 ft 运行至主轮接地过程。该阶段航空器处于空中, 机身纵轴应对准跑道中线。

(2) 接地阶段: 航空器主轮接地至前轮接地 过程。该阶段航空器处于地面,接地点位置应 在跑道中线上。

(3)制动阶段:航空器在跑道上做减速运动。该阶段航空器主要做刹车减速工作,飞行员应保证机身运动轨迹为直线且处于跑道中线。

(4) 滑行阶段:航空器已减速至脱离速度并 保持该速度沿跑道中线滑行至最近脱离道口。

(5)退出阶段:航空器转弯脱离跑道并进入 滑行道。该阶段航空器运行轨迹为一条平滑的 曲线。

通过对QAR数据中经纬度数据在Bigemap-GIS Designer软件中进行打点分析后发现,航空 器部分运行轨迹存在曲折、错位与偏移现象, 如图2所示,实线为QAR数据所呈现的航空器 着陆阶段航迹。



图 2 着陆航迹打点图

通过对航空公司飞行手册与相关法规查阅, 航空器着陆过程各阶段应始终沿跑道中线运行, 虽然不同阶段的速度存在差异,但理论上前四 个阶段的运行轨迹在不遭受强大外力冲击的情 况下,连续经纬度点间的连线应为一条直线且 与跑道中线重合,退出阶段理论上应为一条圆 滑的弧线。

2 S-G滤波算法相关理论

2.1 算法简介

Savitzky-Golay(S-G)滤波器^[9]是一种特殊的 低通滤波器,又称Savitzky-Golay平滑器。低通 滤波器的主要用途是平滑噪声数据,消除所有 具有较大误差的数据点,或者从图形中做出初 步的简单参数估算。S-G滤波算法是一种信号平 滑方法,可将原始数据中的噪声和抖动平滑化, 同时保留数据的趋势与特征,减少对原始数据 趋势的畸变。该算法在时域内基于局部多项式, 通过移动窗口利用最小二乘法进行拟合,可直 接处理来自时间域内数据的平滑问题,最大程 度地减少原始数据失真的情况。

2.2 算法流程与公式

S-G滤波算法利用多项式进行数据平滑,基 于最小二乘法,能够保留原始轨迹的整体运动 趋势,平滑偏移量过大的经纬度点,在航迹打 点图中的直观体现为将航迹中的"毛刺"去掉, 整个航迹在前四个阶段中更趋近于一条直线, 退出阶段为平滑的曲线,具体步骤如下:

第一步,设曲线方程为

 $y = a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m$ (1)

窗口大小为 $2n + 1(n = 0, 1, 2, \cdots)$, 多项式的 阶数为 $m(m = 0, 1, 2, 3, \cdots)$, 各测量点 $x = (-p, -p + 1, \cdots, 0, 1, p)$ 。其中, x为原始数据, y为拟 合后数据, $a_i(i = 0, 1, 2, 3, \cdots)$ 为待求解参数。

第二步,将 n 个上述多项式并列,可得到 下列矩阵运算:



第三步,简化超定方程组为

$$Y_{(2p+1)\times 1} = X_{(2p+1)\times k} \cdot A_{m\times 1} + E_{(2p+1)}$$
(3)
第四步, *A*的最小二乘解为

$$\hat{A} = (X^{\mathrm{T}} \cdot X)^{-1} \cdot X^{\mathrm{T}} \cdot Y \tag{4}$$

第五步, 滤波值 Y 为

$$\hat{Y} = X \cdot A = X \cdot (X^{\mathrm{T}} \cdot X)^{-1} \cdot X^{\mathrm{T}} \cdot Y = B \cdot Y \quad (5)$$

$$B = X \cdot \left(X^{\mathrm{T}} \cdot X\right)^{-1} \cdot X^{\mathrm{T}} \tag{6}$$

3 航迹修正

航空器着陆轨迹修正包括QAR原始数据处 理、着陆阶段航迹提取、基于S-G滤波算法修 正等方面。航迹修正流程图如图3所示。



图 3 着陆阶段航迹修正流程图

3.1 QAR数据处理

由于QAR原始数据无法被后续研究中的程序 所识别,故首先对QAR原始数据进行解码处理, 提取数据中的经纬度参数并按时间顺序排列。

通过人工遍历的方式对数据进行分析,发 现部分经纬度点出现缺失现象。针对缺失数据, 本文首先对机场跑道进行平面坐标建模,再将 经纬度坐标转化为平面坐标,利用线性插值方 程对缺失数据进行填充,公式如下所示。

$$Y = \left[\frac{(X - X_1)(Y_2 - Y_1)}{X_2 - X_1}\right] + Y_1$$
(7)

其中: (*X*, *Y*)为缺失值; (*X*₁, *Y*₁)(*X*₂, *Y*₂)为与缺 失值相邻的两坐标点。

再通过对航空器着陆阶段运行过程的分析, 发现航空器占用跑道的时间集中于45~80 s的 区间,同时QAR数据中经纬度数据的记录频率 为1 Hz,故本研究选取航空器高度50 ft对应的 经纬度点为起始点,向后截取连续的100个经 纬度点,并将其按时间顺序组成的连线确定为 该航空器着陆阶段运行轨迹。

3.2 模型参数设置

在利用S-G滤波算法对航空器着陆航迹进 行修正时,可以使用诸如Python、MATLAB、 JAVA等平台。由于Python具有众多开源机器学 习库,便于程序编写,故本文采用Python平台 搭建航空器着陆轨迹修正模型,模型参数设置 如下:①轨迹修正算法采用Savitzky-Golay滤波 算法;②滑动窗口大小设置为11;③多项式阶 数设置为2。

3.3 修正结果

本文采用 Python 中的可视化模块,对航空 器着陆阶段航迹通过 S-G 滤波算法的修正效果 与原始航迹进行对比,结果如图4所示,细节 展示如图5所示。



图 4 平滑处理后对比图

从图4可以看出,对比修正处理后的航迹 与QAR原始数据中经纬度所组成的原始航迹, 两者总体运动趋势一致,保持原始航迹的方向 与运动趋势不变,未出现发散现象。



图 5 平滑处理细节展示图

通过图 5 局部细节的展示,平滑处理后的 航迹(虚线部分)与原始航迹(实线部分)相比, 曲折、错位等现象明显减少,符合航空器着陆 阶段运动学原理,证明经过 S-G 滤波航迹修正 模型处理,有效减少了航空器着陆阶段运行轨 迹误差。

4 结语

QAR数据作为评价飞行员技术水平与航司运行品质的重要依据,具有重要的研究价值。本研究针对QAR数据中经纬度数据的误差,提出了基于S-G滤波算法的航迹修正模型,通过Python中可视化模块的展示,证明该方法可有效修正QAR数据中的异常经纬度点,使航迹更加贴近于真实轨迹,增加对飞行员飞行技术评估的客观性和合理性。

参考文献:

- [1] 谢玉冰. 基于机器学习的航空器着陆跑道占用时间 预测研究[D]. 广汉:中国民用航空飞行院,2023.
- [2] 罗凤娥,赵强,王波,等.基于QAR数据的飞行安全 评价综述[J].科技和产业,2021,21(9):192-197.

- [3] WANG L, WU C X, SUN R S. An analysis of flight quick access recorder (QAR) data and its applications in preventing landing incidents [J]. Reliability Engineering and System Safety, 2014, 127:86-96
- [4] 王天明.基于QAR数据的飞行安全模型研究[D].天津:中国民航大学,2009.
- [5] 张召悦,杜海龙,高春燕.基于基本飞行模型的 ADS-B航迹修正研究[J].中国民航大学学报, 2013,31(5):1-4.
- [6] 李楠,邓人博,刘朋,等.基于速度修正的终端区离场航空器的轨迹谱聚类研究[J].武汉理工大学学报,2017,39(11):27-31,60.
- [7] 徐开俊,李成,林浩冬,等.基于飞行数据的航迹精度的优化与分析[J].现代计算机,2022,28(2):
 91-96.
- [8] AVERYANOVA Y, RUDIAKOVA A, YANOVSKY F. Aircraft trajectories correction using sharing operative meteorological radar information [C] //Proceedings of the 2020 21st International Radar Symposium (IRS), Warsaw, Poland, 2020:256-259.
- [9] 蔡天净,唐瀚. Savitzky-Golay 平滑滤波器的最小 二乘拟合原理综述[J]. 数字通信,2011,38(1):
 63-68,82.

Research on aircraft taxi trajectory correction during landing stage based on S-G filter algorithm

He Xin, Wang Jingyi*

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

Abstract: Analyzing the operational trajectory of aircraft during landing is an important means to objectively evaluate the technical level of pilots and the operational quality of airlines, and is of great significance in ensuring the safe and efficient operation of aircraft. Through dot analysis of the longitude and latitude parameters in QAR data in map software, it was found that the trajectory of the aircraft during deceleration and taxiing on the runway showed significant twists, dislocations, and deviations. This article uses the S-G filtering algorithm to construct a trajectory correction model for the aircraft landing phase, and performs trajectory correction processing. The results indicate that the S-G filtering algorithm can effectively correct the original trajectory by modifying the model.

Keywords: aircraft trajectory correction; S-G filter; QAR data

现代计算机 Modern Computer

开发案例

文章编号:1007-1423(2024)08-0112-05

DOI: 10.3969/j.issn.1007-1423.2024.08.020

基于LabVIEW的心电信号与多数据采集分析系统设计

陈亚华,张凯淇,马 俊*

(长沙医学院信息工程学院,长沙 410203)

摘要:近年来心血管疾病的发病率和死亡率不断攀升。为了给人们提供一个较为准确的心电信号分析结果,该系统以心电信号为感知节点,结合虚拟仪器技术,打造一个准确性更高、成本更低、更可靠的心电信号分析系统。它的 优点在于会结合使用者的各种因素,如:所处气象、饮食、地理位置、身体状况和生活习惯等因素,并结合已经发展 较为成熟的心电信号数据分析手段,在对连续的心电图数据实时采集处理与分析功能方面,此设备具有数据实时性、 精度较高等特点。结果表明,结合了多数据的分析结果,比单一分析系统给出的结果更为准确,也更加能够满足人们 的需要,这将对预防和治疗心血管疾病产生重要作用。

关键词:虚拟仪器技术;心血管疾病;系统设计;心电信号

0 引言

随着医学、计算机科学技术的发展和心电 信号监测技术研究的深入,人们对于心电信号 监测分析的准确度要求也越来越高。传统的心 电检测方式已经逐渐满足不了人们的需求。而 虚拟仪器技术的发展和应用能够提供一个良好 的心电信号处理环境,利用虚拟仪器技术开发 的系统只需要提前利用平台提供的各种函数和 特定的信号处理功能进行提前设定,而不需要 用户进行复杂的操作和设计,就可以实现自己 想要的功能。但是在采集心电信号的过程中, 为了达到一定的采集精度,不得不考虑对心电 信号造成影响的诸多原因,例如:用户和用户 之间的身体情况各不相同,同一用户在不同时 期也各有相同,还有生活中的各种行为习惯等 等,这些都是影响心电信号检测不可忽视的因 素。但是人们对心电信号的准确性和可参考性 的要求又越来越高,所以一份准确度高的心电 信号数据分析报告显得尤为重要。

针对这些问题,本文利用LabVIEW开发环 境,使用模块化分析处理方法,建立影响心血 管疾病患者的多种因素与心电分析之间的关联, 为患者提供更为准确的心电信号分析结果^[1]。 本文针对心血管疾病,利用最新LabVIEW开发 环境设计一种低成本、方便携带和操作的心电 信号采集和分析系统,针对多种数据对心电信 号的影响进行了分析,也对心电信号数据的采 集和QRS 波群各波段的定位分析进行了研究, 总结了心电信号数据采集分析的方法和优势, 系统设计开发的难点以及优缺点,为系统设计 的实现提供了支持。表1介绍了本文所需要用到 的开发环境的特点。

收稿日期: 2023-10-31 修稿日期: 2023-11-27

基金项目: 2022年度国家级大学生创新训练项目(教高司函[2022]10号-s202210823028); 湖南省大学生创新创业 训练计划项目(湘教通[2022]174号-4642)

作者简介: 陈亚华(2001—),男,湖南永州人,在读本科,研究方向为计算机科学与技术; 张凯淇(2003—),女,湖南 岳阳人,在读本科,研究方向为计算机科学与技术; *通信作者:马俊(1979—),男,湖南常德人,硕士,副教授,研究方向为 计算机网络信息安全,E-mail:5166618@qq.com

表 1 LabVIEW 编程语言的特点

| 特点 | 具体内容 |
|-------|---|
| 所见即所得 | LabVIEW使用所见即所得的可视化技术建立人 机界面,以便用户可以方便地将现有的控制对 象改成适合自己需要的控制对象 |
| 函数库丰富 | LabVIEW 平台提供了大量的现成函数库供用户 调用,几乎涵盖了仪器设计中所需要的所有函 数,并且LabVIEW 还支持用于网络通讯的协议 等应用软件标准 |
| 运行速度快 | LabVIEW采用编译方式运行32/64位应用程序, 解决了其它图形编程平台速度慢的问题 |
| 支持跨系统 | LabVIEW支持多种系统平台运行使用 |
| 调试简便 | LabVIEW的强大功能归因于它的层次化结构, 使得整个虚拟仪器调试起来相对的简便、快捷 |

1 相关工作

多种因素数据对心电信号的影响分析如下。

(1)气象因素:21世纪以来,极端天气频 发,对人体健康构成威胁。根据研究可知,在 发生极端天气时,心血管疾病患者会感到明显 的不适,而且对气温骤变的感知也很敏感,这 种因素对心电信号检测都会产生影响^[2]。

(2) 饮食因素: 生活中的一些饮食习惯也是 心血管疾病病发和死亡的重要因素之一。我们 应该增加膳食纤维和其他营养素的摄入量,从 而降低血清胆固醇和血脂来达到降低心脑血管 疾病的发病率和死亡率,可以对心血管疾病起 到预防作用,如果已经患有心血管疾病就要降 低对高脂肪含量的食物摄入^[3]。

(3)地理位置:有研究显示,当气压在 101.6 kPa 时发病率最低,当气压小于或大于此 值时发病率都会有所增加。

(4)身体状况:在心血管疾病的产生中衰老和一些基础性疾病是心血管疾病产生的主要原因,表现为一定的组织病变,器官衰老及其功能、适应能力和抵抗能力的减退。同时还受到 多方面因素的影响,如营养、卫生条件、遗传和环境等。

(5)生活习惯:主要是在日常的生活中有无 抽烟和喝酒的习惯。相对于吸烟来说,饮酒对 血压的影响要和缓一些,不过,饮酒同样也可 导致心率加快、血管收缩、血压升高,增加脑 出血的发病概率,有研究发现酒精具有扩张血 管的作用,所以饮酒后血压会有所降低,这也 是在进行心电信号分析的时候需要考虑和注意 的。

本文利用 LabVIEW 平台提供的函数设计一 个数据采集输入模块,如图1所示,用于用户输 入这些数据,而输入的数据将会由模块自动保 存至后台方便进行调用。



图 1 数据采集输入模块

用户通过"文本对话框"模块录入自己的 数据,在对用户的心电信号数据采集完毕之后 会存储到"写入测量模块"文件存储模块中, 如需调用录入的数据可以通过调用系统中的 "测量文件模块",此模块将会自动完成调用, 并根据设定的程序自动进行组合分析,最终系 统会根据设定的程序,通过独特的分析算法, 结合测量到的心电信号数据分析,最终提供给 用户一个具有个性化特点的分析报告,用户可 以通过PC界面查看分析结果报告,也可以进行 保存或者分享给其他人。分析报告帮助用户了 解自己身体目前的情况,并且会提供参考建议 供用户选择,极大地满足用户的需求。

2 心电信号处理研究

心电信号属于生理信号,具有低频、低幅、 非平稳等特征,因此心电信号抗干扰能力弱, 在信号采集过程中极易受到周围电极设备和人 体不自主运动,如骨骼肌颤动、呼吸运动等影 响。常见的心电信号噪声主要有工频干扰、肌 电干扰、基线漂移等^[45]。噪声对于心电信号数 据的分析具有重要的影响,因此在对心电信号 进行采集时对其固有的噪声滤除和抑制显得十 分重要。近几年针对ECG的去噪方法主要有数 字滤波法、小波变换法、经验模态分解法等, 本文对在数据采集过程中遇到的干扰噪声进行 分析,同时也对去噪方法进行列举,并采取相 对合适的一种去噪方法进行应用分析。

表 2 噪声的种类及分布

| 种类 | 产生原因 | 噪声特征 | |
|------|------------------------------------|--------------------------|--|
| 工频干扰 | 城市电力系统产生的电磁波 所造成的干扰以及电极的 接触 | 50 Hz, 会影响后期 对心电信号的分析 | |
| 肌电干扰 | 人体运动产生的皮肤电势 | 范围较宽, 20~ 5000 Hz | |
| 基线漂移 | 在心电信号采集过程中人体 的微动或呼吸导致电极接触 不良 | 超低频信号,1Hz | |

在进行心电信号检测时可以采用多种方法, 由于心电信号的识别是对心电周期波形中的P 波、QRS波群和T波进行识别和检测,且它们的 波频都集中在0.05 Hz~150 Hz范围。其中研究 最为深入突出的识别波形之一是心脏QRS波群, 它包含着人体最为隐秘、丰富复杂的心脏活动 信息。

由于在心电采集检测的过程中很容易受到 噪声的干扰,所以对于收集到的心电信号进行 预处理是很有必要的,在除去噪声的心电信号 基础之上,再通过算法处理分析提取特征值, 这是进行信号检测的一贯做法。随着研究学者 对于心电信号的降噪和QRS波群特征点检测识 别的方法研究的深入,目前主要有:数字滤波 器法、人工神经网络法、模板匹配法、小波阈 值变换法等方法^[6]。

近年来随着研究学者对心电信号检测识别 的深入研究,发现小波阈值变换方法在心电信 号的检测和降噪处理上能够达到很好的效果。 此方法的核心思想是将目标信号通过一系列的 带通滤波后,再根据信号所处各个波形上分频 率的不同将其分解到不同的尺度层次上。在实 践中利用不同种的信号在不同的尺度上分布不 同,把QRS 波群的特征点和噪声干扰信号区分 开,该方法大大降低了干扰对QRS 波群特征点 检测识别的影响。

小波阈值变换法与上述方法在对比消除噪 声和特征检测等方面具有独特的优势和特点。也 有大量实验和文献表明,小波阈值变换法在心电 信号检测和识别上具有很大的优势。本文所研究 的心电信号分析系统就是运用小波阈值变换的方 法对心电信号进行降噪和特征检测处理。

3 基于小波阈值变换的滤波算法研究

3.1 小波阈值变换降噪的原理

在信号检测方面,学者 Donnho 和 Johnstone 提出了一种应用于降噪方面的算法供学习和参 考,这种算法就是小波阈值去噪算法[7]。这种算 法在进行降噪之前会对含有噪声的心电信号进 行多尺度分解。分解会产生近似系数和细节系 数,因为不同尺度的采样频率不同,所以尺度 分布在不同的频段上,并且互不相交。含有噪 声的心电信号经过变换后,有用信号和噪声就 会分布在不同的频域,并且各种表现也会不同。 例如:噪声信号的小波系数较小且幅值较低, 而有用信号则相反,幅值大、数量多,占主导 地位。所以此时便可以利用阈值算法对其进行 处理, 使得心电信号尽可能保留有用的信号, 以达到降噪的目的。经过研究我们发现,采用 阈值对心电信号进行去噪处理能够满足均方误 差最小,同时含有噪声的信号在经过去噪处理 后能够得到较为纯净的信号,这种方法计算量 较小,满足系统精确度和效率的需求。其流程 如图2所示。



图 2 模拟降噪流程图

3.2 基于小波变换模极值对的 R 波检测

为了能够更加准确地判断出 R 波在 QRS 波 群中的位置信息,可以通过小波变换的方法, 对采集到的心电信号进行分解和重建,检测在 多尺度分解中极大值对的信息,这样可以降低 噪声干扰信号以及 QRS 波群中 P 波和 T 波对定位 R 波信号的影响,这样就能够让 R 波的位置更加 突出也更容易被捕捉到^[89]。然后再将那些冗余 的极大值点进行清除,留下正确的极大值对, 就能够准确地识别到 QRS 波群中的 R 波了,具 体流程如图 3 所示。

3.3 R波检测实验结果

根据我们的理论推论,并从MIT-BIT心电数 据库中取得了部分数据来进行验证,选取了 119.dat、122.dat还有212.dat这三个心电信号数 据进行了实验。经过实验得知,该算法对于一 些信号波很弱的信号,甚至是一些病态信号, 依然可以较为准确地检测到R波的波峰位置, 完成实验设想。



图 3 R 波检测流程图

3.4 QRS波群定位检测

鉴于前面的算法成功对R波进行了准确的 定位,因此就可以继续对Q波和S波进行定位检 测。由于这两个波段的特征出现的时间幅度、 幅值和波动方向都类似,所以对这两个波段的 定位一般都是同时进行的。QRS波群的起点是 一个往下震荡的Q波,终点也是一个往下震荡 的S波,它们分布在R波的两侧,出现的时间很 短、幅值也很短,所以我们根据Q波和S波多出 现在多尺度小波分解的小尺度等位置这个特点, 对QRS波群的起点和终点的检测定位主要在分 解尺度较低值的位置进行,这样不仅可以得到 较为精准的信号,还避免了低频基线漂移的干 扰。具体流程如图4所示。



图 4 QRS 波检测流程图

4 结语

本文以 MIT-BIH^[10]心电数据库为对象,考 虑了多种数据对心电信号检测分析的影响,由 于心电信号的提取和检测会不可避免地存在噪 音,我们在对心电信号进行提取的时候会清洗 和滤除不需要的噪音,结合两方面的研究分析 保证最终结果的准确性。

随着心血管疾病发病率的升高,人们对自 身健康度的重视,一份快捷且准确的心电信号 分析报告越来越被人们所需要。本文重点分析 了气候、饮食以及一些个人的生活习惯会对身 体造成的影响,参考这一点通过高准确度的心电 信号分析,给用户提供一份分析结果。用户可以 参考这份结果报告调整自己的生活习惯或者及时 就医,以保证生命安全。当然本文所设计的系统 也会随着心电信号检测技术的提高和人们对心电 信号影响因素的研究深入而不断更新升级,以提 高检测的准确度,满足用户的需求。

参考文献:

- [1] 于璐,张勇德,沙宪政. Labview和 Matlab 混合编程 方法在生物医学信号分析中的应用[J]. 中国医疗 器械杂志,2011,35(1):15-18.
- [2] 杜佩阳,王昭俊.气象因素和室温对心脑血管疾病 影响研究综述[J].煤气与热力,2022,42(10):1-7.
- [3] 李倩楠, 贾思侬, 吴楠. 膳食模式对于心脑血管疾病 的影响[J]. 科学大众(科学教育), 2017(4):154.
- [4] 王润.心电信号的预处理算法分析[J].现代计算机 (专业版),2018(7):33-36.
- [5] 杜鹏飞.心电信号的预处理及特征提取算法研究[D].郑州:郑州大学,2015.
- [6] 徐洁,王阿明,郑小锋.基于小波阈值去噪的心电信
 号分析[J].计算机仿真,2011,28(12):260-263.
- [7] 杨柳.基于虚拟仪器平台的心电信号采集与处理系统研究[D],南昌:东华理工大学,2017.
- [8] 李露.心电信号ST段分析研究[D].重庆:重庆大学,2007.
- [9] 王玲利. 心电信号检测与分类算法的研究[D]. 西安:陕西科技大学, 2019.
- [10] 崔英善,何晓艳,鲜于丹谦. MIT-BIH心电信号实 验系统在 LabVIEW 中的应用[J]. 中国仪器仪表, 2017(1):50-53.

(下转第120页)

文章编号:1007-1423(2024)08-0116-05

DOI: 10.3969/j.issn.1007-1423.2024.08.021

基于微信小程序的预约挂号系统的设计与实现

王相喜*,周晓波,倪甜弟

(西京学院计算机学院,西安 710123)

摘要:微信小程序具有方便、快捷、简单的特点,它不需要下载,也不占据太大的内存空间,操作简单,而且成本低廉。因此,使用微信小程序进行线上预约挂号已经成为一种趋势。项目采用 eclipse 微信开发者工具、JSP 技术、Java 技术和 MySQL 等,运用这些工具和技术开发出基于微信小程序的预约挂号系统。该系统可以直接远程实现预约挂号功能,成为一种便捷的医疗服务工具。

关键词: 微信小程序; 微信开发者工具; Java; JSP; MySQL

0 引言

随着计算机技术的不断发展,特别是小程 序的出现,给人们的生活提供了相当大的便利。 此前,人们看病只能去医院排队挂号,对于一 些老人或者出行不便的人来说极其不便。如今, 在各方面都迅速发展的中国^[1],线下挂号这一 方式已经愈来愈少了。在科技的更迭和人们群 众生活方式改变的时代潮流下,网上挂号的方 式逐渐成为主流。预约挂号系统作为一种便捷 的医疗服务工具,给患者和医院都带来了极大 的便利。

1 开发工具及技术介绍

1.1 系统开发技术介绍

1.1.1 JSP技术简介

JSP实质上就是Java,对于用户来说JSP会 以HTML的形式反馈给用户。因此,JSP与Java 有很多相似的地方,比如,它们都有良好的工 具支持,都具有语言的相对易用性,都具有跨 平台优势。JSP主要用于处理并产生动态网页, 也就是说,它是前端和后台的桥梁,可以使得 那些采用Web开发的项目变得快捷和容易。JSP 可以直接反馈给客户一个直接的文本,用户可 以选择任意浏览器来浏览。

1.1.2 MySQL简介

数据库就是收集数据的结构。MySQL占用 内存较小,运行起来速度也比较快。对于一些 不太庞大的数据,完全可以使用MySQL。而且, MySQL由于其是开源的,成了大多数公司和用 户的选择。此外,MySQL的操作非常简单,对 于初学者来说,不需要掌握更多知识,很容易 就可以使用。

1.2 系统开发工具介绍

1.2.1 eclipse

eclipse 是一个常用的后台开发平台,由于 它本身的特点,多被用于开发Java语言。作为 开发Java语言的常用平台,eclipse 拥有很多的 优点。首先,它拥有丰富的插件,基本可以满 足用户和企业的需求;其次,eclipse有中文版, 可以便于编程人员进行开发和维护。eclipse 界 面非常简洁,无特别复杂的结构,用户的使用 体验非常好。本系统使用的eclipse 内置了jdk 和

收稿日期: 2023-11-06 修稿日期: 2024-03-25

作者简介: *通信作者:王相喜(1997—),男,河南洛阳人,在读硕士研究生,研究方向为大数据技术分析与应用, E-mail:1762193309@qq.com;周晓波(1982—),女,陕西西安人,硕士生导师,副教授,研究方向为智能信息处理、大数据技 术及信息安全与决策支持;倪甜弟(1999—),男,安徽蚌埠人,在读硕士研究生,研究方向为大数据技术分析与应用

tomcat,可以不用配置环境变量,比较方便。而 且,由于 eclipse 是完全免费的,它也是大多数 用户的选择。

1.2.2 Navicat

Navicat 就是一套可靠、简洁、安全的管理 数据信息的一种工具,它的功能可以满足个人 用户、企业用户的基本需求,而且Navicat 可以 不断地更新,不断扩展它本身的功能。使用 Navicat 去访问远端服务器,不会受到安全影响。 除此之外,Navicat 还可以支持多种数据格式, 也可以登录到数据库中,进行各种的操作。 Navicat 界面非常简洁,无特别复杂的功能,这 保证了其工作效率。

2 系统设计

2.1 系统功能概述

在本次设计开发的系统中,总共分为两个 模块,分别是用户端模块以及后台管理模块。 用户端模块主要包括查找医生、用户挂号、医 院概况和医院导航。后台管理模块主要包括后 台管理登录、医生管理、预约挂号管理和登录 日志功能。具体的功能概括如图1所示。



图 1 系统功能概括图

2.2 用户端功能设计

用户可以查询到医院医生信息,然后选择 一位医生,填写自己的姓名和联系方式进行预 约挂号。挂号成功后,用户可以查询到自己的 挂号信息。此外,用户也可以查询医院概况, 可以更加具体地了解医院的相关信息^[2]。通过 医院导航功能,用户可以通过地图导航的方式 直接查询到医院的具体位置。用户登录小程序 的流程如图2所示。



图 2 用户登录小程序流程图

2.3 后台管理功能设计

后台管理模块,主要有医生管理、管理员管理、预约查询等功能。管理员可以修改医生的名称、资料等,也可以添加和删除医生信息^[3]。

2.4 数据库设计

2.4.1 数据库功能概述

为了满足本次系统开发的要求,也为了用 户能够及时了解系统的必要信息,本系统至少 拥有以下功能:

(1)建立数据表保存用户登录小程序时所需 要的数据。

(2)建立数据表保存预约挂号信息的数据, 如用户名称、预约价格等信息。

(3)建立数据表保存所有的医生,如管理员 名称等信息。

2.4.2 数据表设计

对于数据库的数据表来说,数据表就是字 段名、字段的数据类型、基本的属性和约束, 也就是某一张表的内部字段之间的关系和其他 几张表之间的字段和字段之间的关系^[4]。这里 以用户表和医生信息表为例详细描述其表结构。

(1) 普通用户表

普通用户信息数据表主要用来保存普通用 户的基本信息。该表的结构见表1。

表 1 普通用户表

| 列名 | 数据类型 | 长度 | 描述 | 是否主键 |
|-----------------------|----------------|------------|--------------|--------|
| user_id | int | 30 | 用户id | 是 |
| user_name | varchar | 100 | 用户名 | 否 |
| password | int | 100 | 登录密码 | 否 |
| type_name | varchar | 200 | 用户类型 | 否 |
| password type_name | int varchar | 100 200 | 登录密码 用户类型 | 否 否 |

(2) 医生信息表

医生信息数据表用来保存医院医生的所有 信息,该表的结构见表2。

| 列名 | 数据类型 | 长度 | 描述 | 是否主键 |
|------------------|---------|-----|-------|------|
| doctor_id | int | 50 | 医生id | 是 |
| doctor_name | varchar | 150 | 医生名称 | 否 |
| doctor_introduce | varchar | 100 | 医生介绍 | 否 |
| doctor_picture | varchar | 200 | 医生照片 | 否 |
| doctor_title | varchar | 100 | 医生职称名 | 否 |

表 2 医生信息表

3 系统的实现

3.1 用户端功能实现

3.1.1 用户挂号功能的实现

当用户点击挂号功能时,系统会显示医院 医生的详细信息,如医生的姓名、挂号价格、 简介等。首先在微信小程序里的page目录下建 立文件夹,在wxml里构建出预约挂号的页面结 构^[5];然后,在form表单里设计页面主要结构, 并且添加bindsubmit和bindreset,使bindsubmit 可以将表单form中的数据成功触发submit事件; 再在js页面中创建data数据,将wxml的所有数 据放到data里,利用cheliangminginput方法获取 医生信息,设置好method值为post,可以保证 中文不乱码,最后在CSS页面中设置好页面样 式。医生信息运行界面如图3所示。



图 3 医生信息运行界面

3.1.2 医院概况功能的实现

首先在 banner.wxml 里在 button 按钮里设置 bindtap 的函数,函数名自拟;然后在 banner.js里 设计函数方法,在这个方法里设置 wx.navigateto 属性;最后再设置好 URL 地址,也就是设置好 要跳转的页面,采用 bindtap 调用这个函数。对 于布局,利用 view标签和相关的属性以及 flex 布 局,这个页面有一个轮播图的效果,利用了 swiper标签,在 swiper标签里首先设置 circular 衔 接滑动属性,使得轮播图可以滑动,用 interval 设 置自动时间间隔。这里包含了医院简介、医院 分布、值班时间等功能。医院概况运行界面如 图4所示。



图 4 医院概况运行界面

3.1.3 医院导航功能的实现

首先,在page里新建一个文件map,下载 SDK提供直接调用接口来使用腾讯地图,或自 行采用URL进行封装请求并返回自己的数据, SDK封装时会调用某些接口,而在示例中会多 时间内调用同一接口,会造成频繁调用接口的 错误。定位用到wx.chooseLocation函数,成功后 会返回五个参数值,包括latitude、longitude、 scale、width和height,因此可直接在latitude和 longitude设计好医院的详细位置信息,利用 wxml和wxss设置好医院导航界面的布局。地图 导航界面如图5所示。



图 5 地图导航运行界面

3.2 后台管理功能的实现

3.2.1 后台登录功能的实现

这里给出了数据库中两个现有的字段,分 别是用户名和密码,设置了随机验证码来对机 器人进行登录功能的限制。在登录功能的JSP页 面,分别设置了三个属性,密码、用户名称以 及登录验证码,在用户名和密码的input标签中 的 name属性值必须与数据库和实体类中的字段 名一样。输入密码与用户名,也就是向系统输 入了两个值,这两个值将会被传到后台的 controller类中。在登录和注册功能的模块中有一个 基于 MD5 加密的功能。当系统管理员登录系统 时,系统又会根据数据库的值进行匹配,并将 用户名和密码登录时传入后台,负责登录功能 的模块调用 dao 类中的方法,并将这两个值当作 参数传入参数列表。登录界面如图6所示。



图 6 登录运行界面

3.2.2 医生管理功能的实现

对于医生管理模块,系统管理员可以直接

对医生信息进行修改,添加、删除添加医生的 相关信息,比如医生的名称、照片等。实现的 过程如下:首先,在Eclipse里创建guahao的项 目,在Script里定义客户端脚本,并引入在 HTML页面中插入已经设计好的JavaScript。在 head设计最基本的必需的元素,并且在head里 的link标签里用rel规定好文档与链接文档之间 的关系是StyleSheet,用href规定好链接文档位 置,在Style设置好医生管理页面的各种样式, 比如页面背景、页面标题、页面表格等。

3.2.3 预约挂号管理功能的实现

用户挂号成功后,管理员通过预约挂号管 理功能查询到用户的挂号信息^[6]。首先在后端 Eclipse 创建对应的项目,在JSP页面中 body标 签和Style 标签设计出预约挂号管理的界面。然 后,再利用微信小程序挂号功能模块 yisheng.js 页面中的wx.request 去和后端服务器 API接口进行 交互,但为了数据安全性,采用 https 去访问 API 接口,然后在后端新建一个getServerName类用来 获取 URL,再用微信小程序封装好的请求 API 访 问此 URL,使得后端获取到用户的挂号信息。

4 结语

本次开发的微信小程序线上挂号系统,利 用微信小程序快捷、高效的特点,帮助人们在 微信小程序上预约挂号^[7]。微信小程序与APP 不同,它不需要下载,不占据多大的内存,用 户在需要时可以随时使用,而且,它所包含的 功能相当丰富,可以根据用户的实际需求不断完 善自身的功能^[8]。除此之外,目前基本上所有的 医院都是通过微信小程序进行线上问诊、预约挂 号,降低了运营成本。因此,此款微信小程序具 有一定的现实意义,是一款便民服务类工具。

参考文献:

- [1] 王秋颖,李昂,张爱莹.门诊预约挂号管理系统应用 微信小程序的实践与展望[J].中国医药导报, 2019,16(28):170-173.
- [2] 周雨萌,邱晶,张会苹,等.微信小程序的健康卫生 交流平台的设计与实现[J].电子世界,2021(14): 192-193.
- [3] 欧阳斌,陈瑞志,冯子祥,等.基于微信的校医院师 生服务小程序设计[J].电脑知识技术,2021,17

(33):49-52.

- [4] 郭若楠,梁雅丽,王正禹,等.基于微信小程序的通 达学院疾病防控平台[J].电脑知识技术,2021,17
 (19):61-64.
- [5] 周雅娟,赵亚丽,王燕华,等.微信小程序在门诊预 约挂号的应用效果探究[J].甘肃科技,2020,36
 (13):121-123.
- [6] 陈哿菲,李静丹,季美芬."你说我改"门诊就诊者投 诉微信小程序的设计与应用[J]. 护理与康复, 2022,21(2):52-53.
- [7] 郭小磊, 帅宏业. 医生随访微信小程序的设计与开发[J]. 电脑与电信, 2020(12): 40-44.
- [8] 李安妮.基于微信小程序的口腔诊所服务平台设计 与实现[J].软件工程,2019,22(4):40-43.

Design and implementation of appointment registration system based on WeChat mini program

Wang Xiangxi^{*}, Zhou Xiaobo, Ni Tiandi

(School of Computer Science, Xijing University, Xi'an 710123, China)

Abstract: WeChat mini programs have the characteristics of convenience, speed, and simplicity. They do not require downloading or occupying too much memory space, are easy to operate, and are cost-effective. Therefore, using WeChat mini programs for online appointment and registration has become a trend. The project developed this time uses eclipse, WeChat developer tools, JSP technology, Java technology, MySQL, and other tools and technologies to develop a WeChat mini program based appointment and registration system. This system can directly and remotely implement the appointment registration function, becoming a convenient medical service tool.

Keywords: WeChat mini program; WeChat developer tools; Java; JSP; MySQL

(上接第115页)

Design of ECG signal and multiple data acquisition and analysis system based on LabVIEW

Chen Yahua, Zhang Kaiqi, Ma Jun*

(School of Information Engineering, Changsha Medical College, Changsha 410203, China)

Abstract: In recent years, the incidence rate and mortality of cardiovascular diseases have been rising. In order to provide people with a more accurate ECG signal analysis result, the system uses ECG signal as the sensing node and combines virtual instrument technology to create a more accurate, low-cost and reliable ECG signal analysis system. Its advantage lies in its ability to combine various factors of the user; For example, due to factors such as weather conditions, diet, geographical location, physical condition, and lifestyle habits, combined with mature ECG signal data analysis methods, this device has the characteristics of real-time data acquisition, processing, and analysis for continuous ECG data, with high accuracy. The results indicate that combining multiple data analysis results is more accurate and can better meet people's needs than a single analysis system, which will play an important role in preventing and treating cardiovascular diseases.

Keywords: virtual instrument technology; cardiovascular disease; system design; electrocardiogram signal