

文章编号: 2095-2163(2021)06-0173-05

中图分类号: TN911.73

文献标志码: A

基于改进的 RFBNet 行人检测算法

黎国斌, 张 剑, 林向会, 谢本亮

(贵州大学 大数据与信息工程学院, 贵阳 550025)

摘 要: 深度学习方法在行人检测领域取得了不错的成绩,但还存在一些问题需进一步解决,例如遮挡、难负样本等问题。本文提出基于注意力机制的 RFBNet 行人检测算法,针对部分遮挡问题,可以取得更好的检测效果。在注意力机制的引导下,网络更加倾向于可见部分的行人信息,抑制背景信息,从而避免其误导网络训练,进一步降低负样本误检为正样本的概率。为了将模型能够部署在轻量级设备上,本文使用参数量更少的轻量网络模型。在 PASCAL VOC 行人数据集上实验结果显示,通过增加通道注意力机制,检测平均精确率增加了 0.51%;当模型参数量裁剪为 0.9M 和 3.1M 时,行人检测精度仍然能够达到 78.04% 和 80.01%,而模型参数量压缩为原来的约 10%。本文算法可以提高行人检测的精度,且具有良好的鲁棒性和适用性。

关键词: 注意力机制; RFBNet; 轻量网络模型; 行人检测; 部分遮挡

Based on the improved RFBNet pedestrian detection algorithm

LI Guobin, ZHANG Jian, LIN Xianghui, XIE Benliang

(College of Big Data & Information Engineering, Guizhou University, Guiyang 550025, China)

[Abstract] Great success has been achieved in the field of Pedestrian detection with wide usage of deep learning algorithm, but there are still some problems to be further solved, such as occlusion, difficult sample etc. We propose a pedestrian detection algorithm using RFBNet where attention mechanism is applied, which can solve the partial occlusion problem well to some extent. Under the guidance of the attention mechanism, the visible part of pedestrians will be more inclined to be focused by the network, and the background information restrained, so as to avoid the background information misleading network training and further reducing the probability of negative samples being mistakenly detected as positive samples. In order to run on lightweight devices, network has been compressed by reducing layers as many as possible in this work. Experiments on PASCAL VOC data set which is of pedestrian, show that mAP increase by 0.51% by using SENet channel attention, and the mAPs maintain 78.04% and 80.01%, while model parameter variables being reduced by 0.9M and 3.1M respectively. The proposed algorithm can increase the accuracy of pedestrian detection and it has great robustness and applicability.

[Key words] attention mechanisms; RFBNet; lightweight Network model; pedestrian detection; partial occlusion

0 引 言

深度学习技术在计算机视觉领域已经取得较为满意的成绩,例如图像分类、目标检测、车导线检测等。2017 年的 ILSVRC 图像分类挑战大赛中, Hu j 等人提出 SENet 并成功夺取第八届 ImageNet 图像分类冠军。

深度卷积神经网络在计算机视觉领域运用之前,传统的目标检测算法主要是基于人工提取特征,例如 SIFT、LBP、SURF、HOG,以及 Haar 等特征描述因子,其中 HOG 特征在行人检测中运用最为广泛。特征提取后输入至传统的分类器进行学习,进一步确定是行人还是背景,传统的分类器有 boosting 或

者 SVM。传统的目标检测过程需要人为提取特征,过程繁杂。

卷积神经网络在计算机视觉相关任务成功运用之后,神经网络特征提取能力得到进一步的推广。基于深度学习的检测算法分为一阶段检测算法和二阶段检测算法。二阶段目标检测算法,例如 R-CNN、Fast R-CNN、Faster R-CNN,先是生成目标候选框,然后将候选框通过分类和回归网络得到相应的目标类别、位置以及预测置信度。一阶段目标检测算法,例如 YOLOv1、YOLOv2、YOLOv3 等算法、SSD 以及 DSSD 算法,取消候选框生成阶段,直接将图像特征输入回归网络得到相应的检测结果,在速度上领先于二阶段检测算法,但在检测精度上稍逊

基金项目: 国家自然科学基金(61562009); 贵州大学引进人才科研项目(贵大人基合字(2015)29 号); 半导体功率器件教育部工程研究中心开放基金项目(ERCMEKFJJ2019-(06))。

作者简介: 黎国斌(1993-),男,硕士研究生,主要研究方向:目标检测与识别; 谢本亮(1978-),男,博士,副教授,主要研究方向:图像处理、图像分类、目标检测与识别。

通讯作者: 谢本亮 Email: blxie@ gzu. edu. cn

收稿日期: 2021-02-21

色于两阶段检测算法。

行人检测是一种特殊的目标检测任务,其目标尺度不一,算法需同时处理多尺度特征。RFBNet^[1]算法引入 Inception^[2]模块的多尺度分支,最后将不同尺度的特征进行融合,从而提高不同尺度行人的检测准确率。

为了进一步提高行人的检测精准率,本文在 RFBNet 基础上引进通道域注意力机制。通道域注意力机制可以凸显关键信息,忽略无关信息。在注意力机制下,网络模型可以更倾向可见部分行人信息,忽略被遮挡信息,提高行人与背景的分度。

1 相关工作

1.1 注意力机制

注意力机制与大脑神经相似,可以更倾向特定特征,忽略无关信息。注意力机制,利用掩码(mask)信息与网络提取的信息进行加权整合,得到更重要的信息。掩码信息是网络生成的一层权重,特征图上重要信息得到更大的权重,无关信息得到更小的权重。注意力机制分为软注意力机制、强注意力机制,两者最大的不同是,前者更加注重空间或者通道,并且是确定性的注意力,可以通过网络训练进行调整,并且可微分,通过网络不断的前向和反向传播迭代,得到最佳的注意力权重,而后者更加关注点,图像上每个点都可以延伸出注意力,同时强注意力是一个随机的预测过程,更强调动态变化。强注意力是不可微分,训练过程只能通过强化学习完成。

注意力机制分为通道域注意力、空间域注意力、以及混合注意力。本文重点阐述通道域注意力,其中在 RFBNet 算法基础上引入 SENet 通道域注意力。

图像经过深度卷积神经网络提取特征后,由于图像当中只有部分区域是存在目标,其它都为背景信息,输出各个通道的特征响应理应会有所不同,但神经网络更多的是把控整体的信息,难以针对目标、背景施加不同响应。

Hu J 等提出 SENet 通道域注意力网络,该网络充分模拟卷积层特征不同通道之间的关系,允许网络可以适当地进行特征校正^[3]。SENet 网络结构,如图 1 所示。

对于给定的输入 $X \in R^{H' \times W' \times C'}$ 通过任意的转换 F_{tr} (可以是一个卷积层或者卷积集合) 得到 $U \in R^{H \times W \times C}$ 。特征 U 首先通过一个压缩(squeeze)操作,压缩操作表示为全局池化层,即将 $H \times W$ 特征图转

化为一个通道描述因子。由于通道描述因子内嵌各个通道特征响应的全局分布,浅层的卷积层可以充分利用全局信息。压缩后的特征再通过扩张(excitation)操作,即通过带下采样因子 r (默认是 16) 的全连接层,得到的输出通道是 C/r , 之后添加非线性激活函数,接着通过带上采样因子 r 的全连接层,得到的输出通道为原来 C 。最后,扩张后的特征通过 sigmoid 激活函数,将权重限制在 $[0, 1]$ 。输入特征向量 U 对应各个通道的响应得到重新加权。

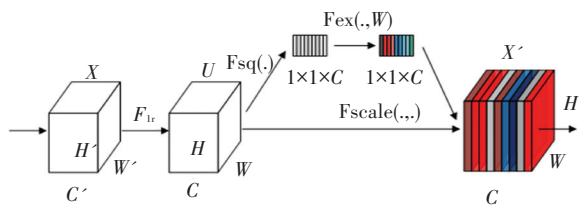


图 1 SENet 网络结构图

Fig. 1 SENet model structure

1.2 RFBNet 网络结构

RFBNet 网络结构如图 2 所示,分别为基础网络,额外层、头部(置信度、分类回归),RFBNet 的基础网络为 SSD,提取原始图像的特征,额外层对不同尺度的特征图应用 RFB 或者 RFB-s 模块,如图 3、图 4 所示,最后根据提取的特征进行分类回归。

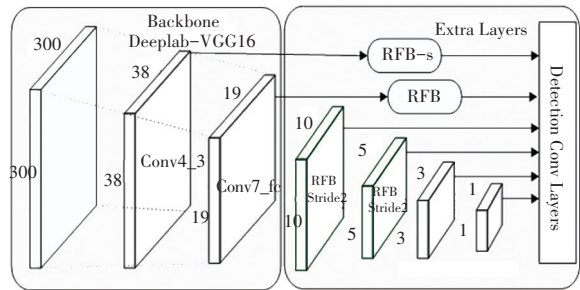


图 2 RFBNet 网络结构图

Fig. 2 RFBNet network structure diagram

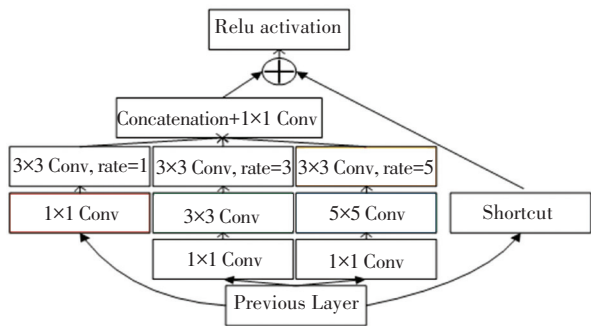


图 3 RFB 模块结构图

Fig. 3 RFB modular structure chart

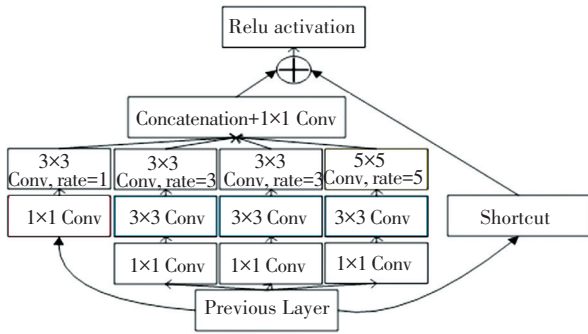


图 4 RFB-s 模块结构图

Fig. 4 RFB-s modular structure chart

RFB 模块主要是利用 Inception 模块的多尺度特征融合思想,并引入空洞卷积。空洞卷积在不增加参数的前提下,可以扩大卷积层的感受野,因而网络深层的语义信息更加丰富,提高网络对目标特征的提取能力。

1.3 轻量网络结构

随着硬件技术的发展,轻量设备可以运行现有的目标检测算法,例如英伟达的 Jetson tx2。尽管存储设备容量很大,但鉴于算法部署在硬件平台的限制,网络模型压缩已成为一种必要的趋势。基础网络可以采用移动端 mobilenet 作为特征提取网络,压缩模型的输出通道数,减少不必要的卷积层数量^[4]。

2 加入通道注意力机制的 RFBNet 算法

图像经过卷积神经网络提取的特征图,是泛化的,不具备针对性。行人被遮挡部分的背景信息参与网络训练,会误导网络判断,因此引入注意力机制。网络模型更加倾向可见部分的行人特征,抑制被遮挡部分的行人信息,加快网络模型往正确的方向进行训练。

本文引入的是 SENet 模块,该模块分为 3 部分:第一部分是 Squeeze,即先压缩;第二部分是 Excitation,即激励;最后一部分是特征重标定,即注意力。Squeeze-and-Excitation block 如图 1 所示。压缩的表达式如公式(1)所示:

$$Y = \frac{1}{h \times w} \sum_{i=0}^h \sum_{j=0}^w X(i, j). \quad (1)$$

公式(1)是图 1 中的 $F_{sq}(\cdot)$, 通道内特征做全局池化操作。 $h \times w$ 分别表示图像经过卷积神经网络得到的特征图的高度和宽度, $X(i, j)$ 表示特征图上位置 (i, j) 的特征信息,特征图上的信息经过压缩后输出一组通道数为 c 的特征信息。通道数为 c 的特征信息通过激励操作,对各个通道的相关性建模,激励操作的表达式如公式(2)所示:

$$S = Sigmoid(W_2 \cdot ReLU(W_1 Y)). \quad (2)$$

W_1 的维度是 $C' \times C$; W_2 的维度是 $C \times C'$; C' 为 $C * 1/4$ 。之后分别通过一个全连接层、激活函数、全连接层,对 Squeeze 的结果 Y 做非线性变换。最后通过 sigmoid 函数将值限定在 $[0, 1]$,输出的维度是 $1 \times 1 \times C$ 。最后是特征重标定,表达式如公式(3)所示:

$$X' = X \cdot S. \quad (3)$$

输入特征 X , Excitation 得到的结果作为权重,两者相乘得到各个通道调整后的特征,重要的特征可以获得更大权重,无关信息可以得到较低权重,因此可以对原特征进行重新标定。

RFBNet 网络基础上引入通道域注意力机制,分别在基础网络、RFB 模块或者 RFB-s 模块上添加,添加注意力机制后的网络结构如下图 5 所示。

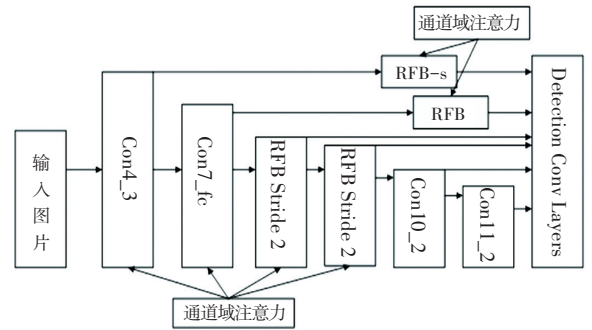


图 5 引入通道注意力的算法结构图

Fig. 5 An algorithm structure diagram that introduces channel attention

3 实验结果与分析

3.1 实验环境与数据集

本实验采用 Linux 操作系统 ubuntu16.04, 选用 HP inter(R) Core(TM) i7-9750 的 CPU 笔记本作为调试平台, TITAN V 显卡的服务器训练神经网络, 搭载 PyTorch 深度学习环境的开发平台。本实验采用 PASCAL VOC 行人数据集, VOC2007 和 VOC2012 的行人子集作为训练集, VOC2007-test 作为测试集。参与网络训练的行人数据集有 6 095 张, 测试集有 2 007 张。

3.2 实验设计及结果分析

目标检测主要使用均值平均准确率 (Map) 作为衡量算法的检测精度。本文实验主要分为 4 组, 第一组是压缩模型参数并测试行人检测的精度, 第二组是 0.9 M 模型参数量下引入注意力机制, 第三组是 3.1 M 模型参数量下引入注意力机制, 第四组是 RFBNet300-person 模型引入注意力机制, 分别定义 0.9 M、3.1 M、33.8 M 模型为 RFBNet_S、RFBNet_M、RFBNet300-

person。最后展示实验结果并分析原因。

3.2.1 第一组实验

本次实验主要在 RFBNet 网络基础上进行压缩,减少卷积层的数量、减小网络模型输出的通道数,因此参数量大幅度下降。实验设置迭代次数 100 epochs, *batch_size* 为 32,单 GPU 训练,学习率为 0.008,骨干网络 (backbone) 使用去掉全连接层的 VGG16,训练得到最后的网络模型。VOC2007 测试集上行人检测结果对比见表 1。

表 1 VOC2007 测试集上行人检测结果对比

Tab. 1 Comparison of pedestrian detection results on VOC2007

| 方法 | backbone | Map/ % | parameters |
|------------------|----------|--------|------------|
| RFBNet300-person | vgg | 82.44 | 33.8M |
| RFBNet_M | vgg | 80.01 | 3.1M |
| RFBNet_S | vgg | 78.04 | 0.99M |

RFBNet300 输入图像分辨率为 300,通过 VGG16 骨干网络提取行人的特征,从 conv4_3 特征图输出后输入至 RFB-s 模块提取不同尺度的特征,将不同尺度的特征进行拼接,最后输出至检测层。RFBNet 算法中分别从 conv4_3、fc_7、conv8_2、conv9_2、conv10_2、conv11_2 特征图输出不同尺度不同宽高比的锚框,在最后的检测层拼接所有的结果,之后通过 NMS 剔除重复的边界框。该算法虽然从多个卷积层提取不同尺度的行人特征,但弊端是通过多个特征图进行特征提取,会带来大量的计算,模型参数量大幅度增加。因此本文削减部分卷积层,降低卷积层的输出通道数,并减少每一个特征图锚框的个数。Conv1_2 输出通道数 64 改为 32,conv2_2 输出通道数 128 改为 64,conv3_3 输出通道数 256 改为 128,conv4_3 输出通道数 512 改为 258,conv5_2 输出通道数 512 改为 258,并削减 conv5_3 卷积层,同时削减 conv8_2 和 conv9_2 层的 RFB 模块,模型为 RFBNet_M。RFBNet_M 基础上将 conv4_3、conv5_2、conv6_2、conv7_2、conv8_2 的输出通道数 256 改为 128,并 conv9_2 的输出通道数 256 改为 64,得到的模型为 RFBNet_S。由表 1 知,RFBNet_M 算法平均精准率仍然为 80.01%,模型参数量压缩为原来的约 10%。RFBNet_S 模型参数量为 0.99M,平均精准率为 78.04%。RFBNet_M 模型削减 conv8_2 和 conv9_2 特征图的 RFB 模块,特征图并没有得到多尺度特征融合,产生漏检,检测精度稍微下降。另一原因是压缩部分卷积层和降低输出通道数,降低模型的非线性表达,因此模型对行人特征提取能力下

降。

本文在上述模型分别引入注意力机制,增加模型的非线性表达,可以增大模型的特征提取能力。注意力机制可以实现网络更倾向可见部分行人特征,减少被遮挡部分行人信息对模型的干扰。

3.2.2 第二组实验

本次实验是在 RFBNet_S 引入注意力机制。实验迭代次数设置为 100 epochs, *batch_size* 为 32,单 GPU 训练,学习率为 0.008。引入注意力机制的实验结果见表 2。

表 2 引入注意力机制的实验结果

Tab. 2 Experiment results of attention mechanism is introduced

| Backbone | RFB | Backbone | Map/ % | parameters |
|----------|-----|----------|--------|------------|
| | √ | vgg | 77.9 | 0.99M |
| √ | | vgg | 78.1 | 0.99M |
| √ | √ | vgg | 77.5 | 0.99M |

表 2 中√表示在骨干网络 (Backbone) 或者在 (RFB) 模块中引入注意力机制。从表 2 知,RFBNet_S 在骨干网络引入注意力机制效果更佳。在骨干网络的基础上引入注意力机制,检测精度提升 0.06%。通道域模块加入到骨干网络,会增加网络的非线性表达能力,通道域注意力计算出各个通道权重,行人特征对应通道得到较大权重,背景信息对应通道得到较低权重,因此提高模型对行人特征的提取能力。在 RFB 模块引入注意力机制,失去多尺度行人特征融合的作用,模型检测性能会略微下降。

3.2.3 第三组实验

本次实验是在 RFBNet_M 引入注意力机制。实验迭代次数设置为 100 epochs, *batch_size* 为 32,单 GPU 训练,学习率为 0.008。引入注意力机制的实验结果见表 3。

表 3 引入注意力机制的实验结果对比

Tab. 3 Experiment results of attention mechanism is introduced

| Backbone | RFB | Backbone | Map/ % | parameters |
|----------|-----|----------|--------|------------|
| | √ | vgg | 79.6 | 3.1M |
| √ | | vgg | 80.2 | 3.1M |
| √ | √ | vgg | 79.7 | 3.1M |

表 3 中√表示对应的模块引入注意力机制。从表 1、表 3 实验结果可知,RFBNet_M 在骨干网络引入注意力机制效果更佳。骨干网络添加注意力模块,可以增加网络的非线性,提高模型对行人特征的提取能力。通道域注意力可以对不同通道施加不同的权重,行人特征对应通道得到更大权重,背景信息对应通道得到较小权重,网络可以更倾向有用的行人

特征,忽略无关信息,因此行人检测精度提高。

3.2.4 第四组实验

本次实验是在 RFBNet300-person 引入注意力机制。实验迭代次数设置为 100 epochs, *batch_size* 为 32,单 GPU 训练,学习率为 0.008。引入注意力机制的实验结果见表 4。

表 4 引入注意力机制的实验结果

Tab. 4 Experimental results with attention mechanisms

| Baseline | RFB | RFB-s | Backbone | Map/ % | parameters |
|----------|-----|-------|----------|--------|------------|
| RFBNet | | | vgg | 82.44 | 33.8M |
| | √ | √ | vgg | 82.95 | 33.8M |

表 4 中√表示相应的模块引入注意力机制。RFBNet300-person 在骨干网络引入注意力机制效果更佳,检测精度提升 0.51%。通道域注意力可以实现网络更倾向行人的关键信息,忽略无关信息,精准训练网络。SENet 结构专注于卷积核各个通道的响应,图像经过卷积核提取特征后,图像上不同的目标在卷积核对应通道会得到不同的响应,SENet 可以加强有用行人特征,降低无用特征对应通道响应,因而网络对行人特征识别能力更强,进一步提升行人检测的性能。

(上接第 172 页)

信,使得消息数据更加简单,数据解析更加高效。为了应对不同的消费场景,采用可扩展的消费者集群的算法来实现消费者数量的动态变化。

目前,本文设计的中间件系统已经在智慧宿管平台的实际业务场景中应用,提高了响应速率,解决了数据丢失等问题,系统运行效果良好。但是由于硬件设备以及实际业务场景的局限性,更高并发场景下该消息中间件的表现与改进还有待进一步的研究与实践。

参考文献

- [1] 陈海明,石海龙,李勤,等. 物联网服务中间件:挑战与研究进展[J]. 计算机学报,2017,40(8):1725-1749.
- [2] 王鹏,周岩. 面向高性能应用的 MPI 大数据处理[J]. 计算机应用,2018,38(12):3496-3499,3508.
- [3] 陈瑶,李洋磊. 基于 ActiveMQ 的数据传输框架性能优化[J]. 现代信息科技,2019,3(16):128-130.
- [4] SOMMER P, SCHELLROTH F, FISCHER M, et al. Message-oriented Middleware for Industrial Production Systems[C]//2018 IEEE 14th International Conference on Automation Science and Engineering, 2018: 1217-1223.
- [5] 郭宗怀. Kafka 消息系统可靠性研究[D]. 重庆:重庆邮电大学,2018.
- [6] 叶姣姣. 基于消息中间件技术的智慧园区解决方案[J]. 电信科学,2018,34(S2):185-191.
- [7] 陈思媛,易国洪,金旺. 智能停车系统中消息中间件的设计与实

4 结束语

为了将 RFBNet 算法部署在轻量设备上运行,本文对 RFBNet 模型进行进一步压缩。削减卷积层,降低模型输出通道数,削减 conv8_2 和 conv9_2 特征图的 RFB 模块,经过实验,模型参数量压缩约为原来的 1/10,行人检测精度仍然保持为 80.01%。骨干网络引入通道域注意机制的 RFBNet 算法,提高行人检测的精度,且具有良好的鲁棒性和适用性。

参考文献

- [1] LIU S, HUANG D. Receptive field block net for accurate and fast object detection[C]// Proceedings of the European Conference on Computer Vision (ECCV). 2018: 385-400.
- [2] SZEGEDY C, VANHOUCHE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.
- [3] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
- [4] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861,2017.
- [5] 现[J]. 武汉工程大学学报,2020(2):224-230.
- [8] 吴斌烽. 基于微服务架构的物联网中间件设计[J]. 计算机科学,2019,46(S1):580-584,604.
- [9] NGUYEN C N, LEE J, HWANG S, et al. On the role of message broker middleware for many-task computing on a big-data platform[J]. Cluster Compute, 2019, 22: 2527-2540.
- [10] Georgios Lilis, Maher Kayal, A secure and distributed message oriented middleware for smart building applications, Automation in Construction, 2018, 86: 163-175.
- [11] 唐颖淳. 利用 Kafka 实现大话语并发数据流的吞吐系统[J]. 电信科学,2018,34(S2):134-139.
- [12] 赧伟,金蓓弘,陈海彪,等. 基于发布/订阅中间件的时空事件检测研究[J]. 计算机科学,2012,39(10):99-103,135.
- [13] 吴晗. 基于 AMQP 的消息中间件的设计和实现[D]. 南京:东南大学,2019.
- [14] 张娜,史佳炳,吴彪等. 基于 Netty 和 Kafka 的 IOT 终端服务系统设计方案[J]. 浙江理工大学学报(自然科学版),2020,43(2):240-245.
- [15] 夏斐. 基于 Netty 的消息中间件的研究与实现[D]. 成都:电子科技大学,2018.
- [16] 徐进,黄勃,冯炯. 基于消息通信的分布式系统最终一致性平台[J]. 计算机应用,2017,37(4):1157-1163.
- [17] M GARCÍAVALLS, GARRIDO D, M DÍAZ. Impact of Middleware Design on the Communication Performance [M]. 2017.
- [18] 葛东遥. 消息中间件管理和监控系统的设计与实现[D]. 北京:北京交通大学,2017.
- [19] 王重楠,王宗陶,鲍忠贵,等. 发布/订阅模式测控消息中间件系统设计[J]. 计算机应用,2015,35(3):878-881.