Journal of Computer Applications
计算机应用,2020,40(12):3679－3686

ISSN 1001-9081
CODEN JYIIDU

2020-12-10
http://www.joca.cn

文章编号:1001-9081(2020)12-3679-08

DOI:10.11772/j.issn.1001-9081.2020071084

基于YOLO-tiny-RFB模型的电站旋钮开关状态识别

史梦安[1*],陆振宇[2,3]

(1.苏州大学 应用技术学院,江苏 苏州 215325; 2.南京信息工程大学 人工智能学院,南京 210044;
3.江苏省大气环境与装备技术协同创新中心(南京信息工程大学),南京 210044)
(*通信作者电子邮箱 mashi@suda.edu.cn)

摘 要:针对多类别目标检测在特定场景中数据样本有限的情况,为进一步提高机器人系统中轻量级神经网络对小型物体识别的准确率和稳定性,提出了一种基于机器人操作系统(ROS)的目标状态识别模块。首先,考虑到嵌入式设备的算力限制,目标识别模型采用轻量级的网络YOLO-tiny作为主要架构,并在YOLO-tiny中引入RFB,提出了YOLO-tiny-RFB模型。随后,基于MobileNet对旋钮开关的多种状态实现精准分类。最后,设计数据关联规则,通过图像配准及交并比(IOU)计算等算法使识别模块完成同一场景多次识别结果的融合,从而使用户能够对不同时刻各表计的状态进行追踪。实验结果表明,相较于YOLO-tiny,YOLO-tiny-RFB模型在少量增加模型计算量的情况下,在构建的电站仪器识别数据集上,其目标识别平均精度均值(mAP)提升了17.9%,达到了82.4%。在旋钮数据分布极端不均衡的情况下,通过引入多种数据增广方法使模型的平均准确率达到了90.7%。所提出的目标检测模块和状态识别网络模型能够有效、准确地完成各类仪器的状态识别,同时能够对仪器状态的识别结果在时间跨度上进行融合。

关键词:机器人操作系统;目标检测;图像分类;轻量级神经网络;数据增广;YOLO-tiny;RFB;旋钮开关状态

中图分类号:TP391.4 **文献标志码:**A

Power station rotary switch status recognition based on YOLO-tiny-RFB model

SHI Meng'an[1*], LU Zhenyu[2,3]

(1. *Applied Technology College, Soochow University, Suzhou Jiangsu* 215325, *China*;
2. *School of Artificial Intelligence, Nanjing University of Information Science & Technology, Nanjing Jiangsu* 210044, *China*;
3. *Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology*
(*Nanjing University of Information Science & Technology*), *Nanjing Jiangsu* 210044, *China*)

Abstract: Data samples were always limited for multi-category object detection in some specific scenes. In order to improve the stability and accuracy of the light-weight neural networks for small object recognition in robotic system, an object status recognition module based on Robotic Operating System (ROS) was designed. Firstly, considering the computing power limitation of embedded devices, a lightweight network YOLO-tiny was used as the main architecture of object recognition model, then the Respective Field Block (RFB) was introduced in YOLO-tiny, so as to construct the YOLO-tiny-RFB model. Secondly, MobileNet was employed to conduct an accurate classification of multiple statuses of rotary switches. Finally, the data association rules were designed, and algorithms such as image alignment and Intersection Over Union (IOU) calculation were used to make the recognition module complete the fusion of multiple recognition results of the same scene, so that users were able to track the statuses of each meter at different times. Experimental results show that on the constructed power station instrument recognition dataset, compared with the YOLO-tiny, the YOLO-tiny-RFB model increases the object recognition mean Average Precision (mAP) by 17.9%, which is achieved to 82.4% with a small increase in computational load of model. In the case of extremely unbalanced rotary switch data distribution, the average accuracy of model reaches 90.7% by introducing various data enhancement methods. The proposed object detection module and status recognition network model can complete the status recognition of all kinds of instruments effectively and accurately, meanwhile they can fuse the recognition results of instrument status at different times.

Key words: Robotic Operating System (ROS); object detection; image classification; light-weight neural network; data argumentation; YOLO-tiny; Respective Field Block (RFB); rotary switch status

0 引言

在我国庞大的电力系统当中,存在大量的仪器设备需要人员进行定期的巡检维护。但电站地处偏远,因而可能长期处于无人值守的状态。如果能够借助于自动巡检机器人和计算机视觉技术实现仪器的实时检测,那么电力系统就能够减少人力成本的投入,提升管理效率。

收稿日期:2020-07-22;**修回日期:**2020-10-09;**录用日期:**2020-10-19。

基金项目:江苏省高等学校自然科学研究项目(19KJB520051);国家自然科学基金资助项目(61773220)。

作者简介:史梦安(1985—),男,江苏淮安人,讲师,硕士,CCF会员,主要研究方向:机器学习、基于视觉的同步定位与地图构建方法;陆振宇(1976—),男,江苏常州人,教授,博士,主要研究方向:人工智能、智能控制、模式识别。

机器人操作系统(Robot Operating System, ROS)是开源的,被很多机器人项目所采用,其模块式的设计使得用户能够根据自身的需求定制该系统。遵照ROS定义的标准信息表达和通信机制,开发者的功能模块能够轻松地和ROS进行无缝对接。由于ROS能够在C++和Python环境中运行,因此将深度学习模型融合到机器人系统当中也非常方便。

近十年,深度学习在图像处理的多个领域,如图像分类、语义分割、目标识别等领域都取得了重大的突破。每一年都会有新的网络结构在ImageNet、Pascal VOC、MS COCO等大型数据集上取得突破。观察这些模型的发展过程,可以发现模型的表现与模型的深度有着很大的关联。从2012年提出的VGG16[1]到2016年的ResNet(Residual Network)[2],网络模型的深度从当初的十几层扩展到了一百多层。模型准确率提升和深度的拓展也意味着模型参数量的增加。对于移动设备而言,这些网络的结构过于庞大,移动设备无法满足其计算要求。因而,为了能够在移动设备上使用神经网络这一强大的技术,就需要在模型的结构上做出优化裁剪。为寻找到计算量和准确率之间的平衡点,很多研究者对较深的网络模型在深度和宽度上进行裁剪来达到减少计算量的目的。如YOLO(You Only Look Once)系列[3-5]的轻量版本YOLO-tiny。而有些研究者则是对轻量级的模型加以特殊的结构设计,如MobileNet[6-7]、ShuffleNet[8]、RFB(Respective Field Block)[9]等,通过注入更多的先验知识,使得轻量级模型也能取得较好的识别准确率。其中RFB模块的设计,融合了人眼感受野和视网膜折射率的关系函数这一先验知识。RFB通过网络结构模拟人眼的观察事物的特征,提升模型的准确率。

由于RFB的结构轻便、计算量小,能够很方便地结合在其他识别模型的顶层,因而本文提出了一种基于轻量级神经网络YOLO-tiny和RFB模块的仪器状态识别模块,希望结合两者,在小幅增加模型计算量的同时大幅提升模型的识别准确率。对于细粒度的旋钮状态分类,本文采用了MobileNet为基础模型实现旋钮8个状态的判断。针对旋钮状态分布的极度不均衡,本文还通过翻转等数据增广方式平衡各类别以抵消数据分布不均的影响。由于是电站巡检任务,不同时刻统一表计的状态追踪非常重要。所提出的模块还采用数据跨时间关联机制,将不同时间点的数据联合在一起。

本文的主要工作如下:

1)对YOLO-tiny进行结构调整,融合YOLO-tiny和RFB提出了YOLO-tiny-RFB模型。在保证模型轻量化的同时,大幅提升目标检测的准确率。

2)设计了基于ROS的仪器状态识别系统,将目标检测模型与分类模型嵌入到机器人操作系统(ROS)当中。作为机器人的仪器状态识别模块,其赋予机器人定点仪器的识别能力。

3)由于机器人巡检过程中会对同一巡检点进行多次监测,本文还提出了跨时间的目标检测数据关联机制。该机制能使同一监测点下不同时间点的识别结果有效地整合,从而赋予机器人仪器状态追踪的能力。

1 相关工作

1.1 机器人操作系统

ROS是一个开源的、专为机器人开发、模块化的分布式操作系统。其中模块化的好处在于用户可以自由地选择装载工具库,亦或是手动实现功能。ROS定义了一系列标准的机器人数据表述,如机器人的几何关系转换、位姿表示、传感器数据表示、里程计等定位信息的表示。只要开发者通过标准的信息表述进行功能开发,就能与ROS生态进行无缝的对接。

ROS底层提供了信息传递的接口,该接口会提供进程间的通信功能。通信功能包括:1)发布或订阅匿名消息;2)记录或回溯信息;3)请求响应、远程调用;4)分布式参数系统。

除了进程间的通信功能,ROS系统还提供了基础的机器人工具库,其包含有导航定位、机器人姿势调整、机器人硬件信息诊断等。这些基础功能使一个机器人快速地运作起来。

1.2 目标监测模型

基于深度学习的目标检测方法在实际中已经有了相当广泛的应用,在行人[10]、车辆[11]、船舶[12]等各类目标的监测任务当中都有着良好的效果。基于深度学习的目标检测主要可以分成两大类:1)多阶段处理实现目标检测;2)单模型实现目标检测。

多阶段的目标检测算法有RCNN(Regions with Convolutional Neural Network features)[13]。首先通过Selective Search[14]算法,在输入图像内寻找所有可能包含物体的候选框。随后,利用卷积神经网络对各候选框内的物体进行特征提取。最后,基于各候选框内物体特征,使用分类器支持向量机(Support Vector Machine, SVM)进行框内物体类别的判断。另外,该阶段还会使用一个回归模型对候选框的大小进行调节。可以发现,RCNN实现目标检测主要分为3个阶段,其识别过程包含多个模型的协作。因而RCNN不管是模型的训练过程还是预测过程,都较为复杂。Fast RCNN[15]则将分类模型和回归模型直接整合到了卷积神经网络当中,大大地提升了模型的预测速度和检测精度。Faster RCNN[16]则更进一步,提出了Region Proposal Network,以此来替换Selective Search等检测算法产生候选框,Faster RCNN相较于Fast RCNN在预测速度和精确度上进一步提升。

单模型实现目标检测的模型有SSD(Single Shot multibox Detector)[17]、YOLO等。相较于多阶段处理利用单独的模型寻找候选框,YOLO和SSD都是端到端模型,即将一幅图像输入到模型,模型会直接输出在图像内识别到的物体的位置和该位置的物体类别。该类方法通过在特征图上设置"锚"点,避免了使用单独的模型或算法产生候选框的过程。"锚"即一系列设定好大小和长宽比的候选框。对于卷积神经网络输出的特征图,模型会在特征图的每一个像素位置都预设N个设定好的候选框("锚")。模型会根据输入图像的特征,对该特征点处是否包含物体进行判断,并对包含物体的"锚"的大小进行调整。

端到端的模型依靠单一模型实现目标检测,因而在计算复杂度上具有一定优势,因此本文采用端到端的目标检测模型完成嵌入式设备中的目标检测任务。

2 系统结构

本文所提出的仪器状态识别模块结构如图1所示。该模块能够对仪器状态进行实时检测,并整合同一检测点K在不同时间点的识别结果。本章首先将对数据关联机制作一个详细地介绍,随后对提出的目标检测模型YOLO-tiny-RFB的结构进行解释。

2.1 识别模块的数据关联机制

巡检机器人在日常巡检中的某一段时间内的执行流程如

下：某一时刻t，机器人执行巡检任务依靠定位系统来到检测点K前。机器人调用云台相机对仪器柜内的仪器进行拍摄，得到图像G_t。随后，通过调用目标识别模型完成G_t内仪器状态的识别，结果记作r_t。下一时刻$t+1$，机器人执行下一次巡检任务，再次来到检测点K前，拍照后得到仪器图像。同样调用目标识别模型完成G_{t+1}内仪器状态的识别，结果记作r_{t+1}。

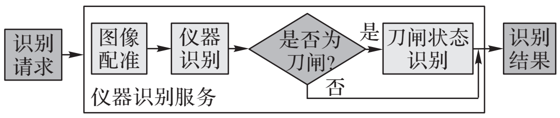


图1　仪器状态识别模块结构

Fig. 1　Structure of instrument status recognition module

虽然是同一检测点K，但由于机器人的定位并非百分之百精准，对于同一个面板，各时间点间拍摄到的图像会存在位移和角度偏差。如图像G_t和G_{t+1}，虽然拍摄的都是相同的一组仪器，但是各仪器在两张图像内的位置会存在偏差。因此识别结果r_t和r_{t+1}内所包含的相同仪器的坐标并不一致，就无法直接对仪器状态的识别结果进行整合。将r_t和r_{t+1}内的识别结果关联起来，对于仪器状态的全局检测非常重要。

为实现跨时间的监测数据关联，首先，拍摄监测点K处的开关柜的标准图像G_s并标注标准图像G_s中的仪器位置；随后，以标准图像G_s为“锚点”，让相同场景下，不同时间点拍摄的图像G_t都和标准图像G_s进行图像配准；接着，对配准后的图像G_t^a进行目标识别；最后，让识别结果r_t^a和标注标准的仪器位置进行相似度的计算，使得识别结果都和标准图像中的仪器关联起来，从而对各时间点的数据完成整合。其过程如图2所示。

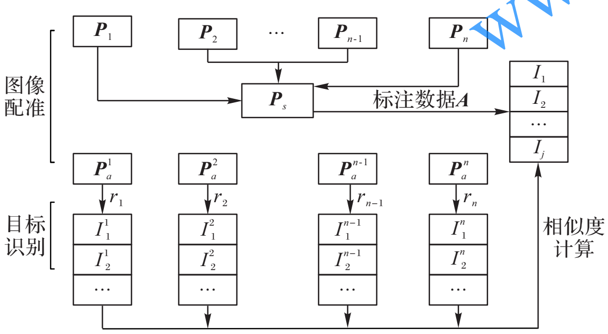


图2　跨时间检测数据关联流程

Fig. 2　Flowchart of detecting data association across time

识别前，为仪器面板K拍摄一幅标准的仪器面板图G_s，并对G_s上所有的仪器位置进行标注，标注数据用矩阵$\boldsymbol{A}$表示。$\boldsymbol{A}$包含G_s上各仪器I_j的坐标，用$(x_{j1},y_{j1},x_{j1},y_{j2})$四个值表示。$(x_{j1},y_{j1})$为包含仪器的矩形框的左上角像素坐标，$(x_{j2},y_{j2})$为矩形框的右下角像素坐标。标准图像$G_s$和标注信息$\boldsymbol{A}$将分别存储在图像数据库和仪器面板数据库当中。

2.1.1　特征点提取

当前机器人巡检拍摄的图像记作G_t，将G_t和G_s进行图像配准。首先，要在两幅图像中寻找相似的特征点。本文通过ORB（Oriented FAST and Rotated BRIEF）图像特征检测算法在两幅图像G_t和G_s中寻找特征点坐标向量$\boldsymbol{P}_i^t$、$\boldsymbol{P}_i^s$，以及特征点的描述特征$\boldsymbol{f}_i^t$、$\boldsymbol{f}_i^s$。

ORB算法通过FAST（Features from Accelerated Segment Test）算法来寻找具有旋转不变性的特征点。FAST算法假设，当一个像素与其周围大部分的像素差别较大时，它应该是一个特征点。算法会比较中心像素与周边像素的灰度值，当大部分周边像素的灰度值和中心像素灰度值之差大于某一阈值时，该中心点就是一个候选特征点。由于只进行像素值的比较，FAST算法的计算效率很高。

随后，ORB用BRIEF（Binary Robust Independent Elementary Features）描述子对特征点的周边像素进行特征描述，从而得到各特征点的特征。BRIEF是一种二进制描述子，即它的描述特征的向量由许多个0和1组成，这里的0和1编码了特征点附近两个像素点（例如说q_1和q_2）的大小关系：如果q_1的像素值比q_2大，则此维度的特征取1；反之就取0。如果选取了特征点附近128个q_1、q_2对，最后就得到128维由0、1组成的特征向量。

最后，根据特征$\boldsymbol{f}_i^t$、$\boldsymbol{f}_i^s$，计算G_t和G_s各特征点间的距离。根据距离排序，设定阈值筛选出一定数量的相似点对$(\boldsymbol{P}_i^t,\boldsymbol{P}_i^s)$。

2.1.2　单应矩阵推断及数据联合

根据多个特征点对$(\boldsymbol{P}_i^t,\boldsymbol{P}_i^s)$求解单应矩阵$\boldsymbol{H}$。$\boldsymbol{H}$为一个$3\times3$的矩阵，表示了两幅图像间各像素坐标的转换关系。由于$\boldsymbol{H}$具有8个自由度，所以至少需要在配准的两幅图像之间找到4对特征点，才能求解出单应矩阵$\boldsymbol{H}$。单应矩阵$\boldsymbol{H}$的求解过程如下：

$$\boldsymbol{P}_i^s=\boldsymbol{H}\boldsymbol{P}_i^t \tag{1}$$

$$\begin{bmatrix} p_{i1}^s \\ p_{i2}^s \\ 1 \end{bmatrix}=\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}\begin{bmatrix} p_{i1}^t \\ p_{i2}^t \\ 1 \end{bmatrix} \tag{2}$$

将式(2)写成联立的方程组，可得：

$$\begin{cases} h_{11}p_{i1}^t+h_{12}p_{i2}^t+h_{13}-p_{i1}^s=0 \\ h_{21}p_{i1}^t+h_{22}p_{i2}^t+h_{23}-p_{i2}^s=0 \\ h_{31}p_{i1}^t+h_{32}p_{i2}^t+h_{33}=1 \end{cases} \tag{3}$$

由于式(3)中的第3项等于1，将其代入前两项，得：

$$\begin{cases} h_{11}p_{i1}^t+h_{12}p_{i2}^t+h_{13}-(h_{31}p_{i1}^t+h_{32}p_{i2}^t+h_{33})p_{i1}^s=0 \\ h_{21}p_{i1}^t+h_{22}p_{i2}^t+h_{23}-(h_{31}p_{i1}^t+h_{32}p_{i2}^t+h_{33})p_{i2}^s=0 \end{cases} \tag{4}$$

进而，整理式(4)，再次整理成矩阵形式，得式(5)：

$$\boldsymbol{A}_i\boldsymbol{h}=\boldsymbol{0} \tag{5}$$

其中：

$$\boldsymbol{A}_i=\begin{bmatrix} p_{i1}^t & p_{i2}^t & 1 & 0 & 0 & 0 & -p_{i1}^t p_{i1}^s & -p_{i2}^t p_{i1}^s & -p_{i1}^s \\ 0 & 0 & 0 & p_{i1}^t & p_{i2}^t & 1 & -p_{i1}^t p_{i2}^s & -p_{i2}^t p_{i2}^s & -p_{i2}^s \end{bmatrix}$$

$$\boldsymbol{h}=[h_1\quad h_2\quad h_3\quad h_4\quad h_5\quad h_6\quad h_7\quad h_8\quad h_9]^{\mathrm{T}}$$

式(4)中为一个2×9矩阵。每一对点都对应一个矩阵$\boldsymbol{A}_i$。将四对点的矩阵进行堆叠，即会得到8×9的矩阵。对$\boldsymbol{A}$进行奇异值分解（Singular Value Decomposition，SVD），即可求得$\boldsymbol{H}$。

$$\boldsymbol{A}\boldsymbol{h}=\begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{bmatrix}\boldsymbol{h}=\boldsymbol{0} \tag{6}$$

利用$\boldsymbol{H}$对G_t进行转换，可将G_t转换到G_s的“视角”。完成图像的转换也分为两个步骤：1）通过$\boldsymbol{H}$将G_t的各像素坐标$\boldsymbol{p}_i^t$转换为$\boldsymbol{p}_i'$，计算过程中，需先将$\boldsymbol{p}_i^t$转换为齐次坐标的形式；

2)利用插值算法如近邻插值法，计算变化后的坐标$\boldsymbol{p}_i'$处的像素值。最后得到配准之后的图像G_a^t。

将图像G_a^t输入到仪器识别模型。仪器识别模型将会输出面板上所有仪器I_i^t在图像中所在的像素坐标位置$(x_{i1}^t,y_{i1}^t,x_{i2}^t,y_{i2}^t)$及其类别$C_i^k$(其中$i$表示单张面板中仪器的索引，$k$表示仪器的类别索引)。如果仪器类别$i$为旋钮，模块会调用旋钮状态识别模型对其状态进行判断。

计算G_a^t中识别到的仪器I_i^t的所处位置$(x_{i1}^t,y_{i1}^t,x_{i2}^t,y_{i2}^t)$和$G_s$中标定的仪器$I_j$的位置$(x_{j1},y_{j1},x_{j1},y_{j2})$的交并比(Intersection Over Union, IOU)，其计算式如式(7)。根据IOU值排序，选择与仪器I_j重合度最高的I_i^t作为仪器I_j此次的识别结果。最终，确定每个仪器I_j的状态。

$$IOU=\frac{Overlap}{Area_a+Area_s-Overlap} \tag{7}$$

其中：

$$Area_a=\left(y_{i2}^t-y_{i1}^t\right)\times\left(x_{i2}^t-x_{i1}^t\right)$$

$$Area_s=\left(y_{j2}-y_{j1}\right)\times\left(x_{j2}-x_{j1}\right)$$

$$Overlap=\left(\min\left(x_{j2},x_{i2}^t\right)-\max\left(x_{j1},x_{i1}^t\right)\right)\times\left(\min\left(y_{j2},y_{i2}^t\right)-\max\left(y_{j1},y_{i1}^t\right)\right)$$

2.2 仪器识别模型

目标检测模型的结构如图3所示。左侧的虚线框架内为YOLO-tiny模型的下采样结构。本文提出的模型采用了与YOLO-tiny相同的下采样结构。实验时对下采样结构的参数进行“冻结”，即该部分参数不会进行梯度的更新。该部分的参数为YOLO-tiny在COCO数据集上的预训练的参数。采用这种方式：一是为了防止模型的过拟合；二是提升模型的识别准确率。

下采样过程的每一个模块的结构均类似，下采样过程由3×3的卷积核、Batch Normalization层、非线性变换Leakey RELU(Leakey REctified Linear Unit)(未在图3中画出)和Max-pooling层构成。需要注意的是，本实验去除了YOLO-tiny模型下采样的倒数第二个模块中，步长为1的Max-pooling层。

图3右侧为模型的上采样过程。由于在进行目标识别时，图像中的物体尺度变化较大。因此，在上采样过程中，模型采用了多尺度特征融合的方式。即在不同尺度的特征图进行上采样操作之后，与来自下采样过程的同样大小的特征图在通道维度上进行叠加。这种跨连接方式在U-net[18]上已经得到证明，它可以减少模型对于训练数据数量的要求。在YOLO-tiny模型中，网络只对两个尺度的特征图进行特征融合，而本文提出的模型则增加了一个尺度，即在三个特征尺度上进行目标检测。特征图的上采样方法采用的是近邻插值法，上采样方法会将特征图插值到原来的两倍。

RFB模块的结构如图4所示。由于生物学的研究表明，人眼的感受野是关于视网膜折射率的函数。它由多个分支构成，与Inception[19]的结构类似。RFB模块通过各分支的前端卷积核模拟不同大小的人眼感受野；利用后端的空洞卷积[20]模拟人眼的感受野和视网膜的折射率的关系，从而实现模拟人眼提取事物特征的过程。在RFB模块的最右侧还引入了ResNet的跨连接结构。由于RFB模块结构参数量较少，且和其他模型的结合方式简便，可直接衔接其他识别模型的顶端。因此本文引入RFB到YOLO-tiny模型当中，提出YOLO-tiny-RFB。

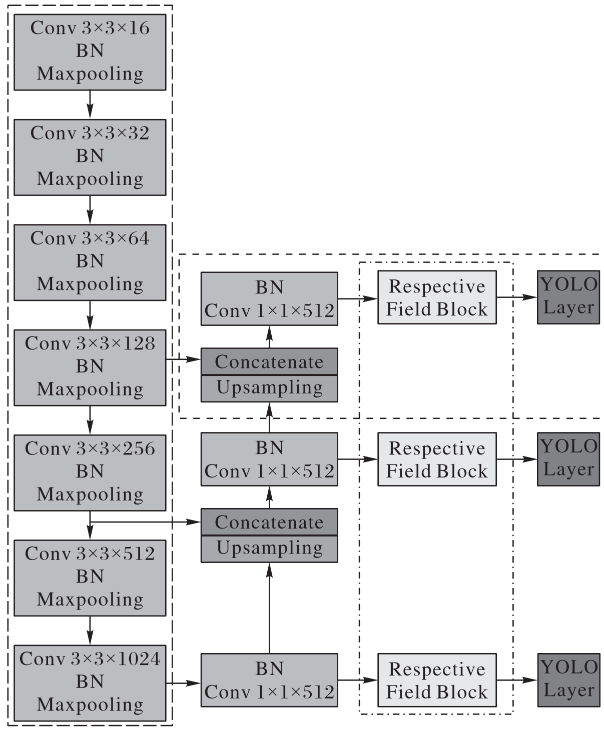


图3 YOLO-tiny-RFB模型结构

Fig. 3 Structure of YOLO-tiny-RFB model

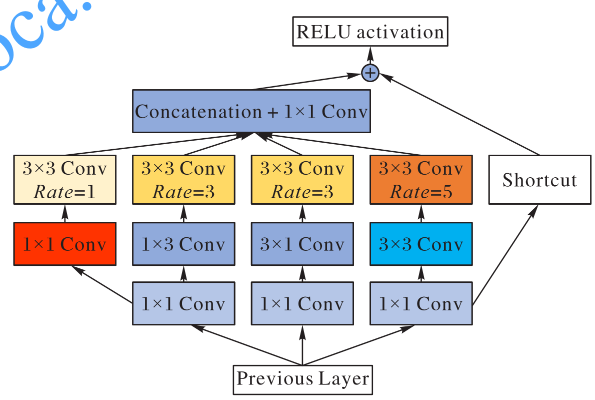


图4 RFB结构

Fig. 4 Structure of RFB

本文所提出的模型YOLO-tiny-RFB与YOLO-tiny有两处不同：

1)由于电站的仪器在尺度上的变化较大，YOLO-tiny-RFB会在三个不同的尺度上进行特征融合，YOLO-tiny则是仅仅在两个特征尺度上进行特征的融合。图3右上角的虚线框中的结构即为新增的特征尺度提取模块。模型还去除了YOLO-tiny在下采样过程中的最后一个步长为1的池化层。由于该步长为1的池化层对YOLO-tiny并未起到减小计算量的作用，而Max-pooling会破坏图像的细粒度特征，因此本实验去除了该层。

2)在进行跨连接和特征融合之后，YOLO-tiny通过相同的1×1卷积核对两个尺度的特征图进行特征提取并直接输入给YOLO Layer。而YOLO-tiny-RFB则在特征输入YOLO Layer之前，引入了RFB模块，以不同大小的卷积核对特征图进行提取，帮助YOLO Layer对相似的物体进行更好的类别判

断。图3中的右下虚线框内为模型新增的RFB结构。

2.3　旋钮开关识别模型

对于旋钮的8个子状态的判断，本文使用了MobileNetV2，其结构如表1所示。为了提升模型的表现，实验使用了MobileNetV2在ImageNet上的预训练参数，随后在旋钮状态分类任务上继续训练。

表1　MobileNetV2结构

Tab. 1　Structure of MobileNetV2

输入	结构	t	c	n	s
$128^2 \times 3$	Conv2d	—	32	1	2
$64^2 \times 32$	Bottleneck	1	16	1	1
$64^2 \times 16$	Bottleneck	6	24	2	2
$32^2 \times 24$	Bottleneck	6	32	3	2
$16^2 \times 32$	Bottleneck	6	64	4	2
$8^2 \times 64$	Bottleneck	6	96	3	1
$8^2 \times 96$	Bottleneck	6	160	3	2
$4^2 \times 160$	Bottleneck	6	320	1	1
$4^2 \times 320$	Conv2d	—	1 280	1	1
$4^2 \times 1280$	Avg-pool	—	—	1	—
$1 \times 1 \times 8$	Conv2d	—	8	—	

MobileNetV2的主要结构由多个Bottleneck Residual Block组成，其同样引入了残差的概念。但与ResNet中的残差块的不同之处在于，Bottleneck Residual Block引入了拓展因子(expansion factor) t，通过改变t的值可以控制模块中的通道数，进行模型的裁切。表1中，c表示模块的输出通道，n表示模块的重复次数，s则表示模块的步长。

3　实验设置

本章将会对实验数据的采集过程中，各物体类别的具体分布、保持各类平衡的方法以及数据的增广方式做一个详细的介绍。

3.1　实验数据

本文用于模型训练的仪器图片均采自于高压变电站，仪器柜的图像如图5所示。

图5　仪器柜图像

Fig. 5　Image of instrument cabinet

数据集中各类仪器类型及出现的次数如表2所示。本文一共采集了1 300张高压屏柜的仪器面板图。在一幅图片中，如果出现了某种类型的仪器，则其数量计数增加1，多次出现会进行多次计数。表2中的统计结果表明数据集中各类的分布较为均衡。本文也考虑到了保持类内各仪器状态分布相对平衡的重要性，采集时也尽量使各类内状态保持平衡，避免模型在训练过程中倾向于某一大类。

在进行数据类别标注时，本文将同一类型仪器的不同状态视作不同类别。如断路器有"开"和"关"两种状态，在标注时会以不同的标签来表示。类似的压板仪器、指示灯、指针表计、按钮的各自的子状态都视作一个单独的类别。这么做的目的是让目标检测模型在实现仪器类型判断的同时也进行仪器状态的判断。

表2　数据集中的仪器类型及分布

Tab. 2　Categories and distribution of instruments in dataset

仪器类型	仪器状态	出现次数
断路器	开、关	740
压板	开、关、闲置	869
指示灯	亮、灭	725
指针表计	无	755
按钮	无	643
旋钮开关	8类状态	438

唯一例外的是旋钮开关，旋钮开关包含8个子状态，即可能会被置于8个不同的方向(左、右、上、下、左上、左下、右上、右下)。其子状态数量较多，但各子状态的图片却较少。因此在实验时，会首先将所有旋钮开关视作一类，不论其处于什么状态。对于旋钮开关的子状态的进一步判断，本文则通过训练一个轻量级分类模型来完成。

从表3可以发现，在采集到的旋钮开关图片中，旋钮开关的子状态主要集中在左上、上、右上3个方向，各状态下旋钮的具体形态可以参考图6，而其他的子状态出现的情况较少。为保持各类均衡，防止分类模型判断偏向于某一方向的判断，本文通过翻转旋钮开关图像的方式来平衡各类数据。例如，旋钮开关置于"下"状态的图像可以通过垂直翻转置于"上"状态的旋钮图像得到。类似地，出现次数较少的旋钮状态，也均可通过水平或垂直翻转出现次数较多的旋钮状态图像来填充。从而，使得各子状态下的旋钮数量达到均衡，均衡后的数量如表3所示。

表3　旋钮开关各状态的数量分布

Tab. 3　Quantity distribution of different statuses of rotary switches

旋钮开关朝向	初始数量	增广后数量
上	121	121
下	52	112
左	12	113
右	23	115
左上	82	113
左下	21	114
右上	114	114
右下	3	114

3.2　数据增广和迁移学习

虽然本文采用的均是较为轻量级的模型，但模型参数总量依旧远多于实验样本的数量。因而，在实验中引入数据增广防止模型的过拟合是有必要的。本文采用的数据增广方式有两种：1)随机地左右翻转图像；2)随机改变图像的对比度和光照强度。采用左右翻转图像的数据增广方式，是因为各类仪器基本都呈现左右对称，且左右翻转也不会改变仪器的状态。而随机改变图像的对比度和光照强度则是为了让机器人适应在电场中的工作环境，使模型在各种的照度下都能有好的识别结果。

迁移学习[21]是目前众多基于深度学习的项目采用的提升模型准确率的方式。迁移学习旨在通过利用类似领域的数据

集让模型学习到泛化的特征，从而减少当前任务的数据量需求。因为，在众多的实际应用中，数据集的采集和标注是相当昂贵费时的过程，即使花费时间精力，采集到的数据数量对于深度学习而言也远远不够。较少的数据集也往往会导致模型的过拟合。

因而，为了在相对较小的数据上取得好的识别结果，首先让模型在大型的数据集，如ImageNet上进行训练；随后，保留并固定模型前部的模型参数用以提取模型的特征，只对模型后部的网络参数进行迭代更新，可以使得模型在少量数据上也取得较好的效果。本文以YOLO-tiny在COCO数据集上训练的目标检测模型的参数为基础，在仪器识别的任务上进行参数微调(Fine tuning)。因为同为目标识别任务，在COCO数据集上进行预训练能够使本文的识别模型学习到泛化的图像特征提取，从而提升模型的识别精度。

图6　旋钮开关的8种置位方向

Fig. 6　Eight setting directions of rotary switch

4　实验与结果分析

4.1　实验设置

模型训练所采用的设备为RTX2080TI，CPU i7-9700K，内存16 GB，系统Ubuntu16.04，编程语言Python3.7，深度学习框架Pytorch1.3。对于目标检测模型的训练，实验对各模型分别进行了2×10^5次迭代，模型选取的学习率$lr=3\times10^{-4}$，正则化项的参数$\lambda=1\times10^{-5}$。为实现数据增广，对于每一幅输入图像，程序有30%的概率会对其进行翻转，有50%的概率会对其进行亮度和对比度的变换。对于旋钮开关状态的识别，实验采用了MobileNetV2，模型选取的学习率$lr=1\times10^{-3}$，正则化项的参数$\lambda=1\times10^{-5}$，一共迭代了5×10^4次。所采用的数据增广方式为50%概率对图像亮度和对比度进行变换。

训练集、验证集及测试集按照7∶1∶2的比例在各类别内进行随机分割，从而保证各类数据在训练集、验证集及测试集中均保持相对均衡。

由于YOLO-tiny-RFB模型除了引入RFB模块，比YOLO-tiny还多了一个尺度的特征图。因此，为了验证加入RFB模块对于模型表现的提升是确实有效的，而不是仅仅依赖于多引入的特征尺度。实验去除了图3结构中的Respective Field Block模块，保留其他的修改，即仅仅去除图3中右下虚线框当中的结构，并将该模型记作YOLO-tiny-modified。以YOLO-tiny-modified作为一个参照模型。

实验使用了平均精度(Average Precision，AP)和平均精度均值(mean Average Precision，mAP)来评估各模型的目标识别表现，使用Precision和Recall来评估旋钮状态分类模型的表现。

4.2　目标识别模型对比

目标识别模型的实验结果如表4和表5所示。从表4中可以发现，YOLO-tiny-modified和YOLO-tiny-RFB分别在4个类(旋钮、断路器-开、指示灯-关、压板-开)和5个类别(断路器-关、指示灯-开、压板-关、压板-闲置、按钮)上取得了最好的识别结果。在这九种类别的识别中，YOLO-tiny-modified和YOLO-tiny-RFB的表现均远优于YOLO-tiny。对于电压、电流表类的表计识别，YOLO-tiny的表现最好，不过3种模型识别准确度的差距并不明显。

表4　各类仪器状态识别准确度

Tab. 4　Status recognition accuracy of different instruments

模型	旋钮	电压、电流表	断路器-关	断路器-开	指示灯-关	指示灯-开	压板-开	压板-关	压板-闲置	按钮
YOLO-tiny	0.683	0.995	0.757	0.744	0.918	0.732	0.753	0.661	0.571	0.184
YOLO-tiny-modified	0.826	0.971	0.851	0.901	0.973	0.801	0.874	0.751	0.761	0.339
YOLO-tiny-RFB	0.811	0.985	0.895	0.863	0.962	0.825	0.777	0.845	0.771	0.513

表5　各神经网络模型识别结果对比

Tab. 5　Recognition result comparison of different neural network models

模型	mAP	参数量/10^6	计算时间/ms
YOLO-tiny	0.699	8.28	88
YOLO-tiny-modified	0.806	8.65	91
YOLO-tiny-RFB	0.824	9.66	110

表5表明，在所有类别的识别表现上，YOLO-tiny-RFB是最好的，其mAP为0.824。YOLO-tiny-modified在参数量较YOLO-tiny增加0.37×10^6的情况下，mAP提升了0.107。YOLO-tiny-RFB相较YOLO-tiny增加了1.38×10^6的参数量，mAP提升了0.125。YOLO-tiny-modified只增加了4.5%的参数，就获得了最有效的性能提升，表明多特征尺度的输入以及取消步长为1的池化层，使得模型拥有了更加丰富和准确的特征用作类别判断。对于按钮识别的表现，YOLO-tiny-RFB则明显较YOLO-tiny-modified和YOLO-tiny要好得多，可见，RFB模块的加入进一步提升了模型对于相似物体的判别能力。YOLO-tiny-RFB模型的识别结果如图7所示。

4.3　旋钮开关识别结果

MobileNetV2模型的旋钮状态识别结果如表6所示。MobileNetV2模型旋钮开关识别的平均准确率为0.907，基本能够对旋钮的状态进行准确的分类。通过观察各类的Recall和Precision，可以发现MobileNetV2模型对于“上”“下”“右”“右上”的状态判断要比其他类别准确，两项指标均高于或接近0.9。

表3中旋钮各状态的原始分布中，“上”“下”“右上”在未进行过人工增广的原数据集占有的比重很大。“上”“下”“右上”在进行训练时，包含的人工数据是较少的。因而MobileNetV2模型能够从真实的数据中学习到该状态下的旋钮特征。而其他状态的旋钮则几乎都是通过这几个状态旋转变换得到的。相比之下，MobileNetV2模型对于“左”“左上”“右下”等状态的判断要稍差一些，Recall和Precision均低

于0.9。

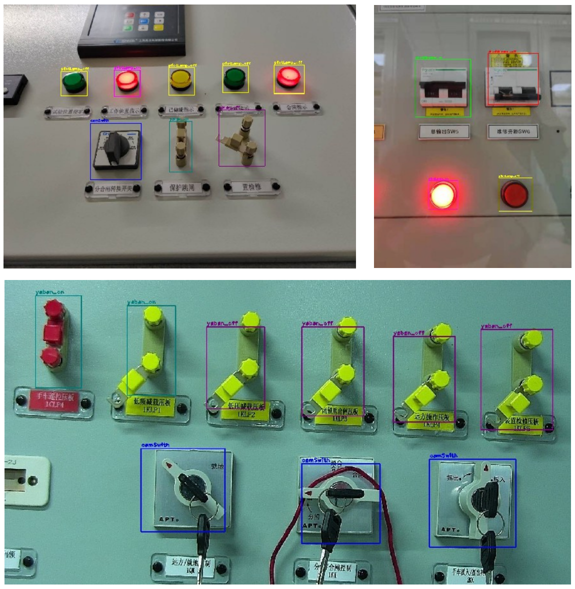


图7　YOLO-tiny-RFB的仪器识别结果

Fig. 7　Instrument recognition results of YOLO-tiny-RFB

表6　MobileNetV2模型的旋钮开关状态分类结果

Tab. 6　Status classification results of rotary switch based on MobileNetV2 model

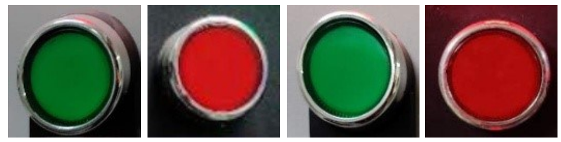
旋钮开关朝向	Recall	Precision
上	0.974	0.902
下	0.921	0.972
左	0.895	0.829
右	0.921	0.921
左上	0.865	0.865
左下	0.892	1.000
右上	0.946	0.897
右下	0.838	0.886

4.4　结果讨论

通过表4和表5的实验结果可以发现，YOLO-tiny-modified相较于YOLO-tiny的参数量提升了4.5%，mAP取得了15.3%的性能增幅；相对的，YOLO-tiny-RFB相较于YOLO-tiny，在参数量增加了16.7%的的情况下，mAP获得了17.9%的表现增幅。虽然YOLO-tiny-RFB看起来收益较小，但是YOLO-tiny-RFB却在较为困难的按钮识别中远优于其他模型。RFB模块的加入使YOLO-tiny能够捕捉到更多的细节，对于较为细粒度的特征更加地敏感。从结果上综合来看，YOLO-tiny-RFB比另外两种模型表现更好。

三种模型对于按钮的状态判断都较差，因此本文进一步分析测试集中数据集的相关特点以寻找原因。本文发现，按钮和指示灯在外观上是非常相似的，其特征如图8所示。它们外观上的区别在于：指示灯在熄灭时，其外围为黑色；按钮的外围为一层银色的外包。当指示灯亮起时，指示灯和按钮则比较容易区分。两者的相似性是三种模型在区分按钮和指示灯时表现不佳的原因。模型无法准确判断两者之间的特征差别。

对旋钮状态判断模型进行进一步的分析，观察测试集中分类错误的图像，发现模型对于图像的细粒度特征有所忽视。另外图像的拍摄角度也对模型的判断有较大的影响。图9(a)、(b)两幅图像均偏离正视视角，开关具有一定形变，此时本文模型无法准确判断旋钮的正确方向。图9(c)、(d)两幅图像中，模型对开关状态做出了相反的判断。作为人类，可以通过"手把"上的颜色准确地判断出旋钮分别是朝向"左上"和"左下"的状态；但模型没有捕捉到这一特征，作出了相反的判断。



(a) 按钮

(b) 指示灯

图8　按钮和指示灯的外观特征对比

Fig. 8　Appearance feature comparison of button and light

(a) 左上1　(b) 右上　(c) 左上2　(d) 左下

图9　旋钮状态分类错误数据

Fig. 9　Rotary switch status misclassification data

5　结语

针对算力资源有限情况下的多类别小型物体状态识别，本文通过在轻量级模型YOLO-tiny中引入RFB结构，以此提出了YOLO-tiny-RFB模型。该模型在小幅度提升模型参数量的情况下，大幅提升了在自建的仪器识别数据集上的精度。本文提出的YOLO-tiny-RFB较YOLO-tiny增加了1.38×10^6的参数量，mAP提升了17.9%。通过数据的增广，本文也较好地完成了在极度不均衡数据集上的分类任务，对于旋钮状态的判断，本文模型的平均准确率可达90.7%。另外，本文还考虑到仪器巡检场景中，同一仪器在不同时间段的状态会被多次监测，因而设计了仪器状态数据的跨时间联合机制。该机制确保同一仪器在不同时刻下的状态能够得到准确追踪。

在随后的工作中，一方面工作的重点将继续着力于控制模型的参数总量，例如加入图像通道的注意力机制[22]。因为网络捕获到的特征众多，但对于识别任务，并不是所有的特征均通道对于最终的识别结果都是有效或贡献相同的，让模型自动去学习各通道之间的权重关系，或许是更好的选择；亦或是引入MobileNetV2的扩张系数(expansion factor)，更好地控制模型的宽度。另一方面，会继续提升模型对于细粒度特征的敏感度，加入新的数据增广方式，如仿射变换、随机图像切割、随机图像噪声等[23]，提升模型的泛化能力。

参考文献(References)

[1] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2019-04-10]. https://arxiv.org/pdf/1409.1556.pdf.

[2] 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. Piscataway: IEEE,

2016: 770-778.

[3] REDMON J, DIVVALA S, GIRSHICK 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. Piscataway: IEEE, 2016: 779-788.

[4] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 6517-6525.

[5] REDMON J, FARHADI A. YOLOv3: an incremental improvement [EB/OL]. [2020-06-08]. https://arxiv.org/pdf/1804.02767.pdf.

[6] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2020-06-08]. https://arxiv.org/pdf/1704.04861.pdf.

[7] SANDLER M, HOWARD A, ZHU M, et al. MobileNetV2: inverted residuals and linear bottlenecks [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 4510-4520.

[8] ZHANG X, ZHOU X, LIN M, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 6848-6856.

[9] LIU S, HUANG D, WANG Y. Receptive field block net for accurate and fast object detection [C]// Proceedings of the 2018 European Conference on Computer Vision, LNCS 11215 . Cham: Springer, 2018: 404-419.

[10] 朱繁,王洪元,张继. 基于改进的Mask R-CNN的行人细粒度检测算法[J]. 计算机应用,2019,39(11):3210-3215. (ZHU F, WANG H Y, ZHANG J. Fine-grained pedestrian detection algorithm based on improved mask R-CNN [J]. Journal of Computer Applications, 2019, 39(11): 3210-3215.)

[11] 张家晨,陈庆奎. 基于YOLO的道路车辆拥堵分析模型[J]. 计算机应用,2019,39(1):93-97. (ZHANG J C, CHEN Q K. Road vehicle congestion analysis model based on YOLO [J]. Journal of Computer Applications, 2019, 39(1): 93-97.)

[12] 聂鑫,刘文,吴巍. 复杂场景下基于增强YOLOv3的船舶目标检测[J]. 计算机应用,2020,40(9):2561-2570. (NIE X, LIU W, WU W. Ship detection based on enhanced YOLOv3 under complex environments [J]. Journal of Computer Applications, 2020, 40(9): 2561-2570.)

[13] 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. Piscataway: IEEE, 2014: 580-587.

[14] UIJLINGS J R R, VAN DE SANDE K E A, GEVERS T, et al. Selective search for object recognition [J]. International Journal of Computer Vision, 2013, 104(2): 154-171.

[15] GIRSHICK R. Fast R-CNN [C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015: 1440-1448.

[16] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]// IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[17] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector [C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9905. Cham: Springer, 2016: 21-37.

[18] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation [C]// Proceedings of the 2015 International Conference on Medical Image Computing and Computer-Assisted Intervention, LNCS 9351. Cham: Springer, 2015: 234-241.

[19] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 1-9.

[20] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40 (4) : 834-848.

[21] ZHUANG F, QI Z, DUAN K, et al. A comprehensive survey on transfer learning [EB/OL]. [2020-07-05]. https://arxiv.org/pdf/1911.02685.pdf.

[22] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.

[23] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning [J]. Journal of Big Data, 2019, 6: Article No. 60.

This work is partially supported by the Natural Science Research Foundation of Jiangsu Higher Education Institutions (19KJB520051), the National Natural Science Foundation of China (61773220).

SHI Meng'an, born in 1985, M. S., lecturer. His research interests include machine learning, simultaneous localization and mapping method based on vision.

LU Zhenyu, born in 1976, Ph. D., professor. His research interests include artificial intelligence, intelligent control, pattern recognition.