

面向设备开关图像识别的改进 Faster R-CNN^①



宋旭峰¹, 蒋梦姣², 周怡伶², 吉俊杰², 陆晓翔³

¹(国网新能源控股有限公司, 北京 100032)

²(华东宜兴抽水蓄能有限公司, 宜兴 214200)

³(河海大学 计算机与信息学院, 南京 211100)

通信作者: 陆晓翔, E-mail: luxx0824@163.com

摘要: 在大型工业厂房中, 由于设备控制开关种类繁多、数量庞大, 在日常的运维过程中, 操作规程的繁杂性和人为判断的主观性可能导致操作失误, 造成严重后果. 为辅助操作人员准确判断设备开关状态是否正确, 提出了面向设备开关状态识别的改进 Faster R-CNN. 首先, 使用膨胀残差网络作为特征提取网络, 在 ResNet50 中引入多分支膨胀卷积, 融合不同感受野的信息; 其次, 改进特征金字塔网络, 在原网络上增加一条自底向上的特征增强分支, 融合多尺度的特征信息; 然后, 使用 K-means++ 算法对开关边界框聚类, 设计适合设备开关的候选框尺寸; 最后, 使用 Soft-NMS 代替非极大值抑制算法 NMS 来降低开关重叠对检测效果的影响, 增强抑制重叠候选框的能力. 在开关状态数据集上, 改进 Faster R-CNN 的均值平均精度 (*mAP*) 达到了 91.5%, 并且已实际应用于抽水蓄能电站日常运维的设备开关状态辅助识别, 满足复杂场景下的智能监管需求.

关键词: Faster R-CNN; 残差网络; 膨胀卷积; 特征金字塔网络

引用格式: 宋旭峰, 蒋梦姣, 周怡伶, 吉俊杰, 陆晓翔. 面向设备开关图像识别的改进 Faster R-CNN. 计算机系统应用, 2022, 31(10): 211-224. <http://www.c-s-a.org.cn/1003-3254/8753.html>

Improved Faster R-CNN for Recognition of Device Switch Images

SONG Xu-Feng¹, JIANG Meng-Jiao², ZHOU Yi-Ling², JI Jun-Jie², LU Xiao-Xiang³

¹(State Grid Xinyuan Holdings Co. Ltd., Beijing 100032, China)

²(East China Yixing Pumped Storage Power Co. Ltd., Yixing 214200, China)

³(College of Computer and Information, Hohai University, Nanjing 211100, China)

Abstract: In large industrial plants, due to a wide variety and a large number of equipment control switches, the complexity of operating procedures and the subjectivity of human judgment may lead to operational errors and cause serious consequences in the daily operation and maintenance process. To assist operators in accurately judging whether the state of an equipment switch is correct, an improved Faster R-CNN algorithm is proposed for state recognition of equipment switches. Firstly, the dilated residual network (ResNet) is used as the feature extraction network, and the multi-branch dilated convolution is introduced into ResNet50 to fuse the information of different receptive fields. Secondly, the feature pyramid network is improved by the addition of a bottom-up feature enhancement branch to the original network, which is used to integrate multi-scale feature information. Then, the K-means++ algorithm is applied to cluster bounding boxes of switches, and the size of proposals for equipment switches is designed. Finally, the non-maximum suppression (NMS) algorithm is replaced with Soft-NMS to reduce the influence of switch overlap on the detection effect and enhance the performance of suppressing the overlapping proposals. On a switch state dataset, the mean average precision (*mAP*) of the improved Faster R-CNN reaches 91.5%. Moreover, it has been applied to assist state recognition of equipment switches in the daily operation and maintenance of pumped-storage power stations to meet the needs of intelligent

① 基金项目: 国网新能源控股有限公司科技项目 (SGXY2000074)

收稿时间: 2021-12-10; 修改时间: 2022-01-10, 2022-01-27; 采用时间: 2022-03-10; csa 在线出版时间: 2022-07-07

supervision in complex scenarios.

Key words: Faster R-CNN; residual network (ResNet); dilated convolution; feature pyramid network (FPN)

在大型工业厂房中,通常部署着各种各样的机械电子设备,设备控制所涉及的开关种类繁多、数量庞大,给日常的运行维护工作带来挑战,例如抽水蓄能电站、光伏发电站等.在日常的运维过程中,工作人员需要在复杂的环境下对各种设备上众多的开关进行操作和检查,保证开关状态的正确性,确保工厂安全稳定地运行.

以抽水蓄能电站为例,作为电力系统中的重要电力来源,抽水蓄能电站在电力调峰、调压调相、事故备用等方面发挥着重要作用.因此,安全稳定地运行、持续可靠地供电对于抽水蓄能电站至关重要^[1,2].日常运维监管作为抽水蓄能电站安全生产与管理的重要手段,对抽水蓄能电站机组设备的检修维护、实时监督、状态评估与预测等各个方面均有重要作用^[3].然而,抽水蓄能电站工况多、维护操作复杂的特点加大了监管的难度,传统完全依赖人工监督的方式对于运行维护和突发事件的处置是不完全可靠的,一旦错误操作没能及时发现与纠正,将会造成设备损坏和人员伤亡的严重后果,同时影响下游电力用户的生产和生活.

因此,为了有效监控上述场景下的人工运维过程,本文引入基于深度学习的图像识别技术来智能化地辅助操作人员判别机组设备的开关状态,确保操作结果的正确性.

对于设备开关状态的识别问题,目前已有一些研究工作.主要分为如下3类.

第1类是基于传统数字图像处理的形态学与特征分析^[4-10],比如周凯等人^[4]提出一种基于非下采样剪切波变换(NSST)图像融合的开关图像识别方法,胡聪等人^[6]提出利用粒子群优化算法对图像进行分割.这种基于传统图像处理技术的方法容易受到操作环境的影响,只能局限于固定的操作场景.

第2类是基于传统机器学习的模式识别方法^[11-13],比如黄钊鹏^[11]提出利用图像处理的相关方法,冷从林^[12]提出基于SVM的开关状态识别算法.这类方法的缺点在于图像特征需要人为定义和提取,一方面提出的特征必须适用于开关的分类,否则会影响分类的效果,另一方面这种特征提取方法很难迁移到其他领域.

第3类则是基于深度学习的图像处理算法^[14-21].

比如史梦安等人^[14]提出基于YOLO-tiny-RFB模型的电站旋钮开关状态识别方法,杨雨等人^[15]提出基于YOLOv5的变电站薄厚硬压板检测技术,李岩等人^[16]、苏杨等人^[17]提出通过卷积神经网络识别开关状态,邓乐武等人^[18]提出了基于深度学习的座舱开关识别方法,高露^[19]、Khan^[20]、张礼波等人^[21]通过数据增强、特征融合和候选框的设计改进Faster R-CNN模型.上述方法均是针对单一的开关类型,只能应用到相近的领域,限制了方法的普适性.

相较于上述文献中所述的设备开关操作场景,本文研究的操作场景更加复杂,具体表现为设备控制所涉及的开关种类繁多、同类开关形状与大小也各异.从开关的类型及其状态划分,总共分为18类开关,36种开关状态;从图像数据分析,图像中的开关密集分布,形状相似,数量较多.这些都给开关识别带来了很大的困难.针对这些挑战,本文提出了一种面向设备开关状态识别的改进的Faster R-CNN.

相比于原始Faster R-CNN^[22],改进之处在于以下4点:选择残差网络为基础网络,将该网络中卷积层的残差块改为多分支膨胀卷积残差块,融合不同感受野的信息;在原特征金字塔网络^[23]的基础上新增一条自底向上特征融合的分支,让塔顶的特征图拥有底层特征图的细节信息;根据开关尺寸信息,聚类出适合开关数据集的候选框尺寸;使用Soft-NMS替代原来的NMS算法.试验表明,改进模型在精度上优于YOLO^[24]、SSD^[25]等常用的目标检测模型,在开关状态数据集上比传统的Faster R-CNN检测精度提升了约7%.此外,该算法已实际应用于抽水蓄能电站日常运维的设备开关状态辅助识别,实践表明可以满足这一复杂场景下的智能监管需求.

1 基于改进Faster R-CNN的设备开关状态识别

相比传统的工厂,抽水蓄能电站部署的机械电子设备更多,设备控制所涉及的开关种类繁多、数量庞大,更加符合本文研究的复杂场景,所以,以抽水蓄能电站为例,研究开关状态的识别问题.

本文的实验数据主要来自华东地区某抽水蓄能电

站,通过对抽水蓄能电站日常操作视频的抽帧,剔除走动、模糊数据,图像增强等技术构建图像数据集.经过图像的分析 and 专家提供的信息,将开关一共分为 18 种,每种开关虽然有多个状态,但是日常操作仅涉及 2 种状态,

所以每种开关均设置 2 种状态. 18 种开关用 1 到 18 的整数编号,同时为了图像标注方便,将“关”“off”“拉开”“现地”“Local”这类开关状态统一记为 0,将“开”“on”“合上”“远程”“Auto”记为 1. 部分开关的状态如表 1 所示.

表 1 部分开关类型和状态

数据集开关编号	开关名称	状态0	状态1
1	直流起励电源开关		
5	集尘器电源开关		
6	中性点铡刀开关		
7	上导油盆加热器控制开关		
10	高压注油泵启动盘电源开关		
11	电压开关		

除了上述开关类型及状态较多的问题之外,还存在以下的难点问题:

- (1) 在不同的设备上存在形状相似的开关.
- (2) 单张图像中存在密集排列且众多的小型开关.
- (3) 不同设备上开关的大小不一样.

这些问题的处理都直接影响到开关的整体识别结果.

针对上述问题,本文提出了改进的 Faster R-CNN. 针对问题 (1), 在 ResNet50 中加入多分支膨胀卷积, 融合不同感受野的信息, 提取更加明显的特征, 便于区分相似的开关; 针对问题 (2), 改进了特征金字塔网络, 新增一条自底向上的特征增强分支, 融合不同尺度的信息, 提高对于密集型开关和小型开关的识别能力; 针对问题 (3), 用 K-means++ 算法^[26] 对设备开关的边界框进行聚类, 设计出适合开关状态数据集的 RPN 候选框尺

寸, 能够检测出不同形状的开关; 最后, 使用 Soft-NMS 替代原 RPN 的 NMS, 提升抑制重叠候选框的效果, 提高候选框选择的正确率; 改进后模型整体结构如图 1 所示.

图 1 最左侧为提出的膨胀残差网络, 将 C3、C4、C5 卷积模块中最后一个残差块变为多分支膨胀卷积; T2、T3、T4、T5 为原特征金字塔网络多尺度融合输出的特征图, D2、D3、D4、D5 为新增的自底向上特征融合分支的输出特征图; P2、P3、P4、P5 为改进后的特征金字塔网络输出的多尺度特征图. 将这些特征图送入 RPN 中, 根据 K-means++ 算法选出的候选框进行目标检测; 最后, 将每个特征矩阵通过 ROI pooling 层缩放到 7×7 大小的特征图, 接着通过一系列全连接层分类回归, 最终通过多次训练达到预期效果.

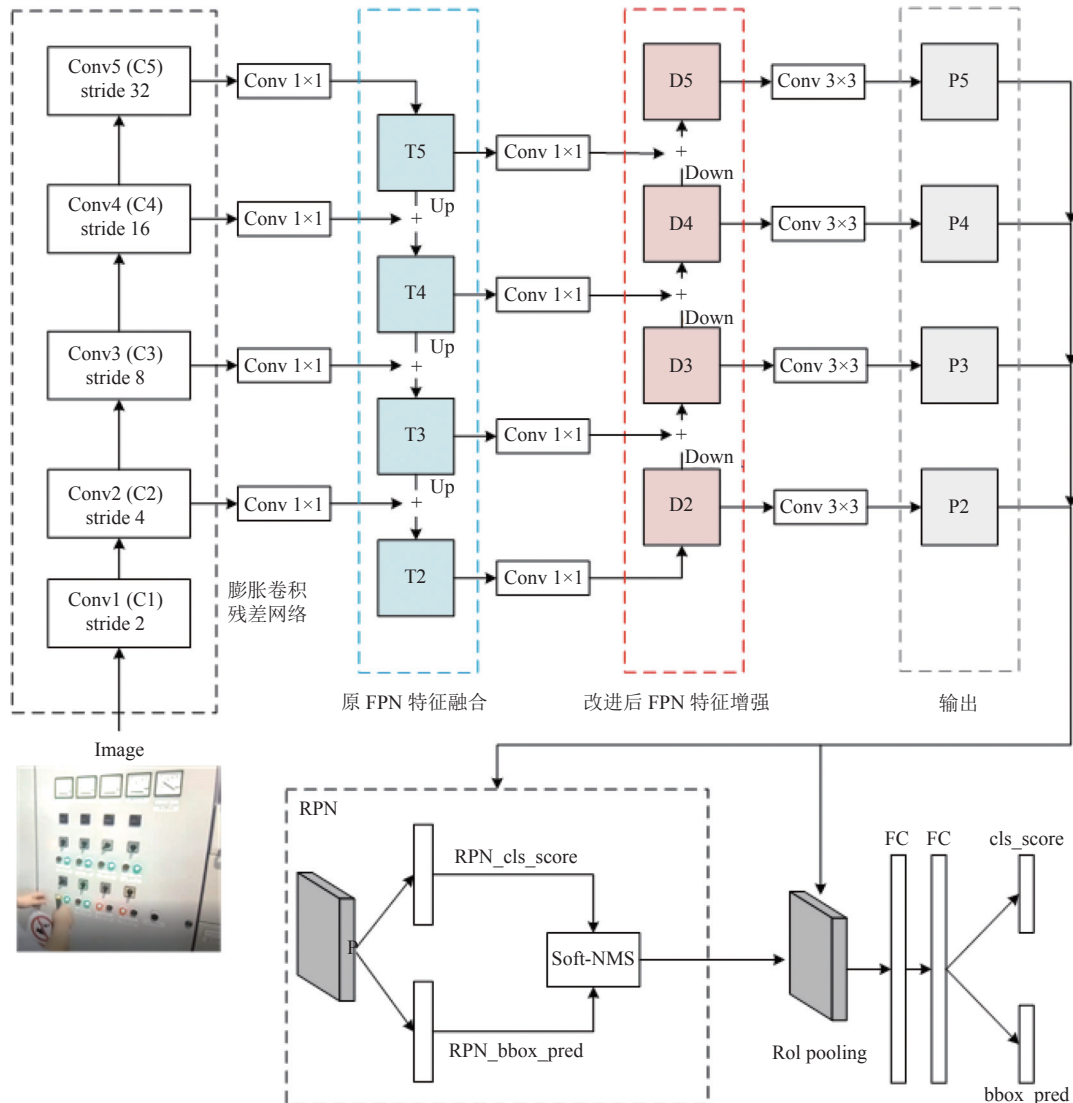


图1 改进 Faster R-CNN 整体模型结构

如图1所示, Faster R-CNN 的损失主要包括 RPN 损失和 ROI 损失, 损失函数如下:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*)$$

其中, i 是批量数据中 anchor 框的序号, p_i 为 anchor 预测为目标的概率, p_i^* 是真实的标签信息 (在 RPN 损失中是二分类, 在 ROI 损失中是多分类); $t_i = \{t_x, t_y, t_w, t_h\}$ 表示预测框的位置信息, t_i^* 表示目标框的位置信息; $L_{cls}(p_i, p_i^*)$ 表示目标和非目标的对数损失, 针对二分类问题, 形式如下:

$$L_{cls}(p_i, p_i^*) = -\log[p_i^* p_i + (1 - p_i^*)(1 - p_i)]$$

$L_{reg}(t_i, t_i^*)$ 是回归损失, 一般采用 smooth L1 函数

计算, 形式如下:

$$smooth_{L1}(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases}$$

$p_i^* L_{reg}(t_i, t_i^*)$ 表示只有预测到目标时才计算损失.

RPN 损失包括分类损失和回归损失, 分类损失是二分类交叉熵损失, 回归损失是对 anchor 的位置进行回归微调. ROI 损失和 RPN 损失相似, 不同的是, ROI 的分类损失是多分类交叉熵损失, ROI 的回归损失是对 ROI 的位置进行回归微调.

1.1 膨胀卷积残差网络

在真实抽水蓄能电站操作场景中, 某些设备的开关排列较为密集, 如果拍摄角度偏离设备, 那么一些开关的部分图像信息会被其他开关遮挡, 没有明显分隔边界.

开关状态识别任务中开关种类有 18 种, 部分类型开关的图像相似, 如图 2 所示, 3 种开关外观相似, 导致开关状态在语义上很难有统一且清晰的定义. 因此现有的深度模型要实现这样的深层语义理解是十分困难的.

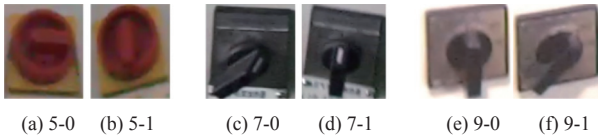


图 2 3 种开关及其状态

卷积神经网络在处理图像特征图时, 为了提高网络训练效率, 会在一些卷积层后使用下采样操作, 但是下采样会丢失很多空间语义信息和细节信息, 因此在层数较深的网络模型中, 深层网络会失去很多浅层网络的空间语义信息, ResNet^[27] 中引入捷径的目的就是将浅层网络的空间信息能够传递到深层网络, 提高了浅层信息的利用率. 开关图像识别需要提取更多细节信息, 所以本文使用残差网络 ResNet50 作为骨干网络来取代传统的 VGG16 网络^[28], 网络参数如表 2 所示. 网络在卷积 conv5 输出之后, 再通过 1×1 的卷积实现通道聚合, 降低参数量, 最后通过全连接层实现分类.

在深度网络中, 特征图空间信息的丢失容易导致小目标信息的丢失. 目前, 这类问题无法完全解决, 但是可以使用膨胀卷积^[29] 在不改变卷积核大小的情况下, 通过改变扩张率 (dilated rate) 扩大感受野^[30] 来缓解上述问题. 卷积核被膨胀卷积处理后感受野大小的计算公式为:

$$n = r \times k + (r - 1) \quad (1)$$

其中, n 为膨胀卷积后感受野, k 为原始卷积核大小, 例如卷积核大小为 3×3, k 就是 3, r 为扩张率.

表 2 ResNet50 网络结构

layer name	output size	ResNet50
conv1	112×112	7×7, 64, stride2
		3×3 max pool, stride 2
conv2	56×56	block1: $\begin{bmatrix} 1 \times 1 & 64 \\ 3 \times 3 & 64 \\ 1 \times 1 & 256 \end{bmatrix} \times 3$
conv3	28×28	block2: $\begin{bmatrix} 1 \times 1 & 128 \\ 3 \times 3 & 128 \\ 1 \times 1 & 512 \end{bmatrix} \times 4$
conv4	14×14	block3: $\begin{bmatrix} 1 \times 1 & 256 \\ 3 \times 3 & 256 \\ 1 \times 1 & 1024 \end{bmatrix} \times 6$
conv5	7×7	block4: $\begin{bmatrix} 1 \times 1 & 512 \\ 3 \times 3 & 512 \\ 1 \times 1 & 2048 \end{bmatrix} \times 3$

图 3 分别表示一个 3×3 的卷积核在 r 取 1、2、4 时的感受野. 图 3(a) 的卷积核的扩张率 r 取 1, 卷积核

感受野没有变化, 依然为 3×3. 图 3(b) 的扩张率 r 取 2, 感受野扩大为 7×7, 由于空洞中填充的是 0, 不参与实际的计算, 所以计算量没有改变. 图 3(c) 的扩张率 r 取 4, 感受野扩大为 15×15.

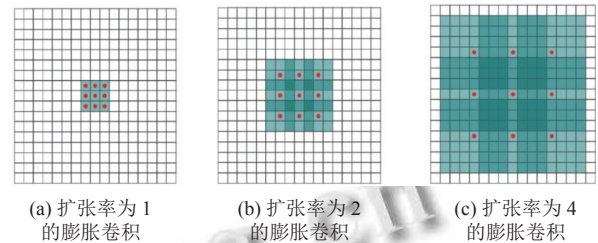


图 3 3 种扩张率下卷积核的感受野

特征图经过膨胀卷积处理后大小变为:

$$o = \frac{W + 2p - r \times k + (r - 1)}{s} \quad (2)$$

其中, W 为输入特征图的大小 (宽或高), p 为 padding 值, s 为步长 stride.

使用膨胀卷积不可避免地出现网格效应 (gridding effect)^[31], 即膨胀卷积的卷积核在滑动时覆盖到的有效区域呈现网格状, 丢失临近点的有关信息, 扩张率 r 越大这种问题越严重. 所以, 既要扩大感受野, 又要降低网格效应的影响, 本文采用综合多个扩张率的膨胀卷积方式, 在感受野扩大的同时, 减少信息的丢失. 具体方法是将一组不同扩张率的卷积进行并联, 均处理相同输入的特征图, 把各个膨胀卷积输出的特征信息融合在一起, 那么多分支膨胀卷积模块的最终输出可以包含不同尺度、不同感受野的信息, 小扩张率的卷积核提取小型开关目标的局部特征, 大扩张率卷积核提取中、大型开关目标的局部特征.

图 4 展示了多分支膨胀卷积模块, 图中模块一共设置了 3 条分支, 扩张率分别为 1、2、3, 并将 3 条分支输出相融合. 图 5 是对不同扩张率膨胀卷积输出特征图的融合, 左侧从上到下分别是扩张率为 1、2、3 的膨胀卷积, 右侧为输出的融合后的特征图.

Faster R-CNN 模型的基础网络使用的是 ResNet50, 具体参数见表 2. 在 ResNet50 基础上使用膨胀卷积进行改进, 主要是将残差块中的 3×3 普通卷积变为 3×3 的多分支膨胀卷积, 改进前后对比如图 6 所示.

使用膨胀残差网络能够在不扩大卷积核尺寸的基础上扩大感受野, 同时减少丢失的信息. 不同尺度的目标能够在不同的感受野下提取特征, 丰富了残差块输出特征图的信息, 同时也能够让小型开关目标在深层

特征中有丰富的语义信息. 而且, 各个膨胀卷积分支之间权值共享, 相较于原始残差块计算量不会增加, 同时可根据不同的数据集调整多分支膨胀卷积残差块在卷积网络中的位置来获得适合数据集的网络模型.

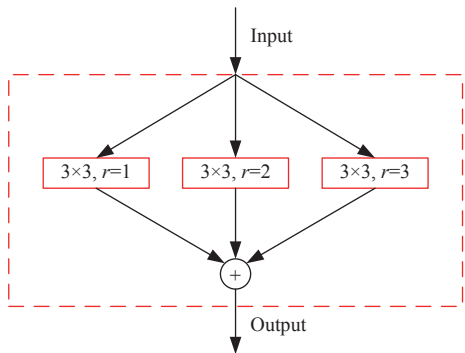


图4 多分支膨胀卷积模块图

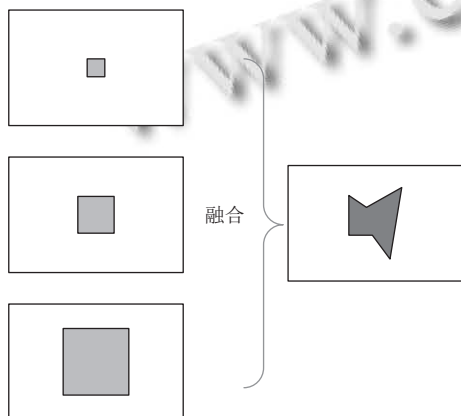
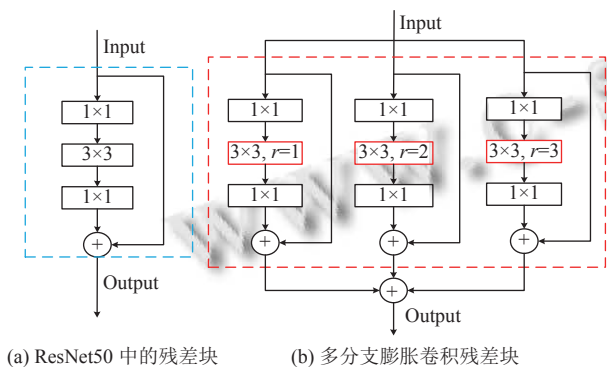


图5 多分支特征图的融合



(a) ResNet50 中的残差块 (b) 多分支膨胀卷积残差块

图6 残差块改进前后对比

1.2 改进特征金字塔网络

卷积神经网络在特征提取时, 底层网络提取的是图像的细节信息, 层数越深的网络提取的特征越抽象. 图7是包含多个开关的设备图像, 由于拍摄距离较远, 所以图像中开关的面积较小.

使用训练好的 ResNet50 对图7进行特征提取, 并给出具有代表性的4个卷积层输出的特征图, 这些特征图如图8所示.



图7 多个开关的设备图像

图8中越亮的地方, 训练时网络越会关注, 从图8(a)和图8(b)可知, 浅层网络较关注图像的纹理和细节信息, 图8(d)的特征图已经很抽象了, 看不出开关的细节.

由第2.2.2节可知, 在拍摄距离较远的情况下, 一些小型开关在图像中较小, 此时开关目标的检测和识别对细节信息要求较高. 然而部分开关是中型的, 也有大型开关, 例如大阀门、闸刀等, 需要兼顾高层的语义信息. 所以特征提取网络既要保留底层细节信息, 还要兼顾高层语义信息. 因此在膨胀卷积残差网络上加入特征金字塔网络, 融合多维信息特征进行开关状态的检测和识别.

图9所展示的是特征金字塔网络结构, 该结构一共两个分支, 左侧自底向上的分支为特征提取, 不断下采样输出多个尺度特征, 每一个方框都是一个尺度的特征图, 右边分支的高层特征不断进行2倍上采样. 这种结构融合了高层特征和底层特征, 让每个尺度的特征图都有丰富的语义信息. 原Faster R-CNN模型使用单一的特征提取网络, 如VGG-16、ResNet等, 仅在网络输出层做预测, 由于特征提取网络的卷积层较多, 在卷积过程中丢失底层语义信息, 不利于小目标开关的检测. 使用FPN后, 可以在底层输出的融合特征图上进行预测, 有利于关注细节信息的小目标开关的检测.

但是特征金字塔网络多尺度融合的特征图中只有金字塔底部包含底层信息较多^[32], 顶部的特征图经过多次卷积后已经丢失了很多底层信息, 特征融合只是自顶向下融合, 让下层拥有上层的特征信息. 为了解决这个问题, 对FPN结构进行改进, 在原FPN右侧分支增加一个自底向上的特征增强分支, 更加充分地融合高层和底层的语义信息, 让高层特征图也有底层语义信息, 改进后的FPN模型结构如图10所示.

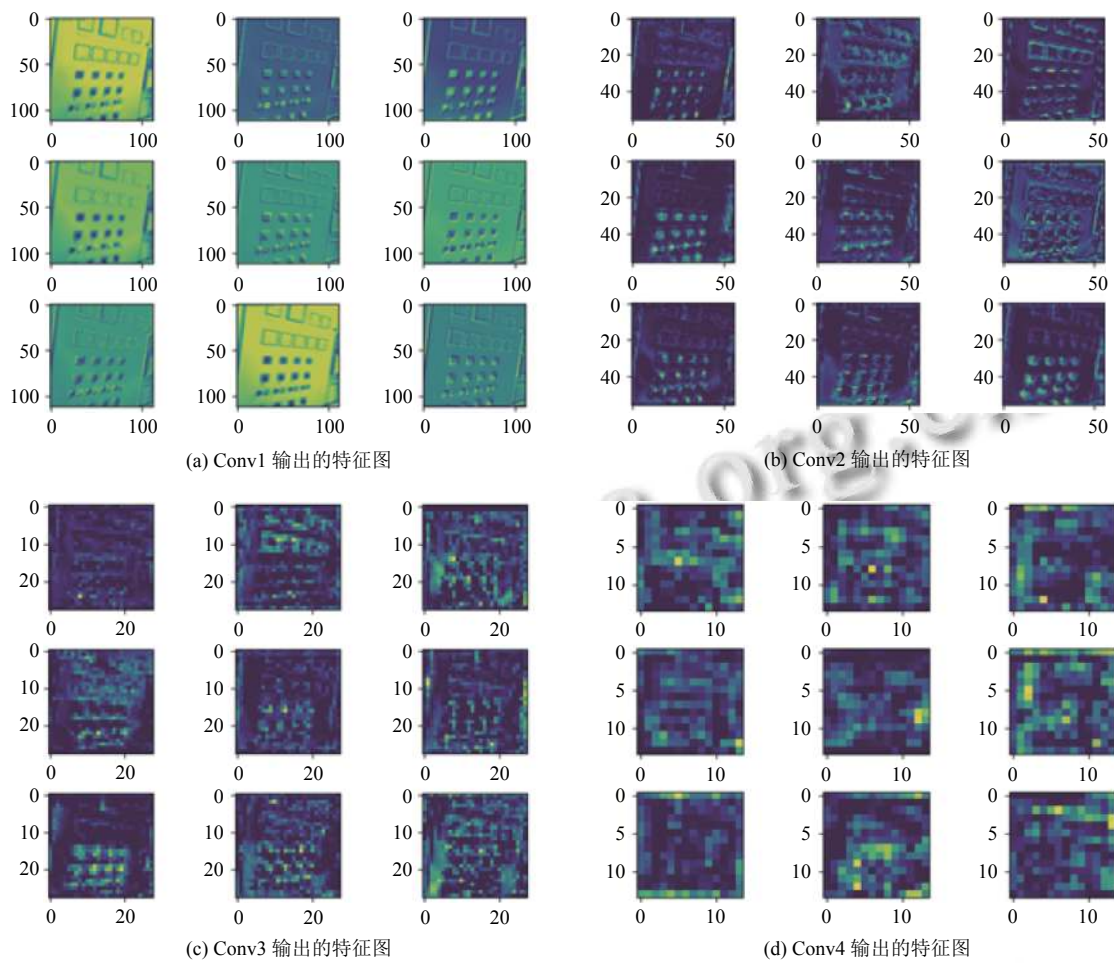


图8 各卷积层输出的特征图

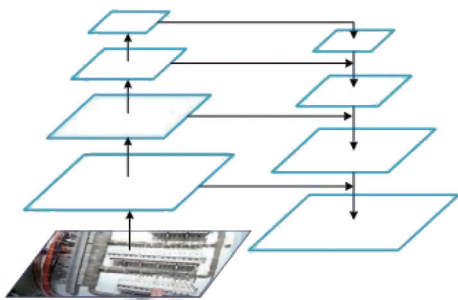
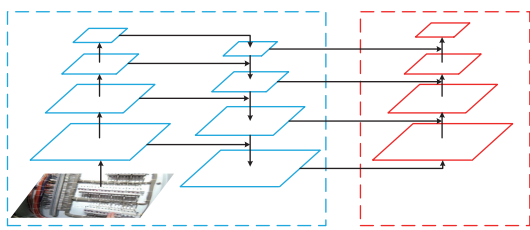


图9 多尺度特征金字塔网络结构



(a) FPN 主干 (b) 新加的增强分支

图10 改进的多尺度特征金字塔网络

FPN 改进后, 因为从特征提取网络的底层到特征增强分支顶层路径上较短, 卷积操作较少, 所以在特征融合过程中底层信息丢失较少, 特征增强分支的各尺度均含有底层信息. 改进后的 FPN, 各层所包含的语义信息要比原始网络多, 尤其是底层以上的各层均含有更丰富的底层信息, 更加有利于小型开关目标的检测与识别.

1.3 基于 K-means++ 的候选框尺寸设计

在目标检测中, 合适的候选框尺寸将决定最后分类回归的效果. 因为开关状态数据集与公开的目标检测数据集存在很大差别, 主要表现在开关的形状相似, 分布密集. 所以 Faster R-CNN 原模型中的候选框尺寸不一定适合开关状态识别任务, 因此需要根据所有开关的真实边界框尺寸来合理设计候选框大小和长宽比例.

本文使用 K-means++ 聚类算法对开关的真实边界框大小和宽高聚类, 得到更适合开关的候选框尺寸, 使模型候选框的尺寸与真实边界框尽量吻合.

为了评估 K-means++ 聚类的效果, 使用轮廓系数

(Silhouette coefficient, SC) 作为开关真实边界框的聚类评价指标, SC 可以同时考虑簇内凝聚度和簇间分离度两方面, 从不同的 K 中选择最优的结果. 假设第 i 个样本与同簇内其他样本的平均距离为 a_i , 到其他簇 C_j 的所有样本的平均距离为 b_{ij} , 设 $b_i = \{b_{i1}, b_{i2}, \dots, b_{iK}\}$; a_i 衡量簇内凝聚度, b_i 衡量簇间分离度, 根据 a_i 、 b_i 定义第 i 个样本的轮廓系数 SC_i , 如式 (3) 所示:

$$SC_i = \begin{cases} 1 - \frac{a_i}{b_i}, & a_i < b_i \\ 0, & a_i = b_i \\ \frac{b_i}{a_i} - 1, & a_i > b_i \end{cases} \quad (3)$$

a_i 比 b_i 越小, SC_i 的值越接近于 1, 表示样本聚类越合理, 反之 SC_i 接近于 -1, 则表明样本 i 与其他簇的样本更相似, 如果接近于 0, 说明样本 i 在两个簇边界. 数据集内所有样本的 SC_i 的平均值就是在取 K 时聚类算法的轮廓系数 SC , 公式如下:

$$SC = \frac{\sum_{i=1}^n SC_i}{n} \quad (4)$$

其中, n 为开关目标总数.

K-means++ 聚类实验前要先设置 K 的值, 根据标注开关的经验预测为 K 在 3 到 6 之间, 实验尝试了 K 取 3 到 12, 使用 SC 作为评价指标, 其结果如图 11 所示.

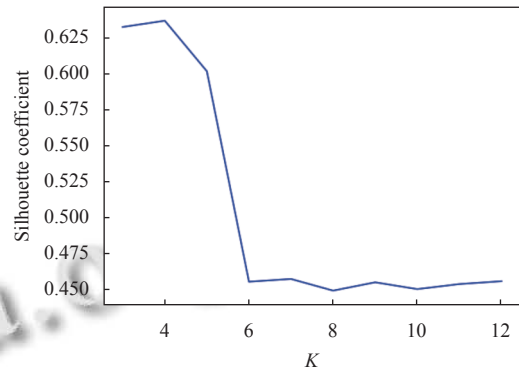


图 11 不同 K 值情况下的 SC 值

由图 11 知 K 取 4 时达到最佳效果, 最高的 SC 值为 0.636 94, K 取 4 也在经验预测的范围内. 图 12 展示了 K 取 3 到 6 的实验结果图.

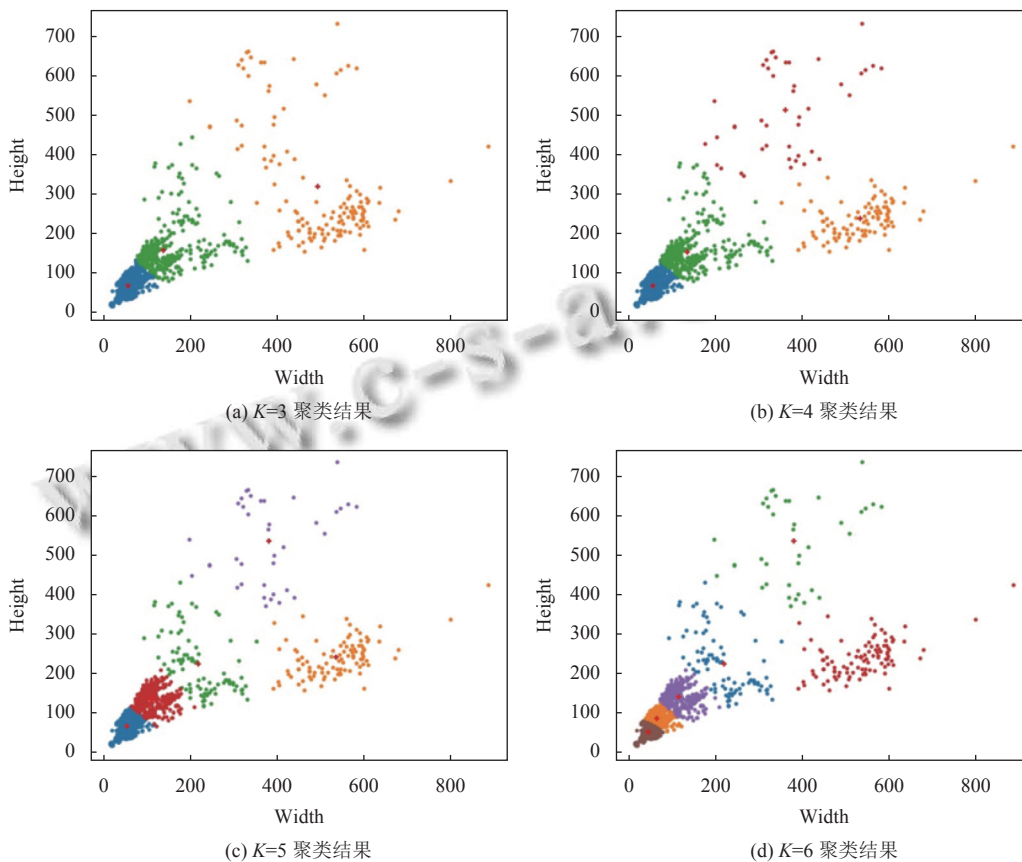


图 12 K-means++ 算法取不同 K 值时的聚类结果

图 12(b) 为最佳聚类效果, 4 个簇的聚类中心分别为 (135, 153)、(56, 68)、(535, 238)、(363, 512), 不难发现, 聚类中心 (135, 153) 密集程度最高, 聚类中心 (363, 512) 密集程度最低. 因此, 对于开关状态的识别可以更加关注中小型目标. 计算 4 个聚类中心点代表的面积, 分别为: 20 655、3 808、127 330、185 856, 对这些面积开平方取近似值得: 144、62、357、431, 计算聚类中心横纵坐标之比也是宽高比取近似值得: (1:1.2)、(1:1.2)、(1:0.5)、(1:1.5). 最终, 本文候选框大小为 62×62、144×144、357×357、431×431, 宽高比设置为 (1:1.2)、(1:0.5)、(1:1.5).

1.4 Soft-NMS 抑制重叠候选框

本文模型在检测开关时, 由于特征图的每个滑动窗口上都会生成多个候选框, 所以相邻的窗口会有多个候选框交叉重叠, 使用非极大值抑制算法 (non-maximum suppression, NMS) 抑制候选框, 删去重叠的多余的候选框, 找到检测效果最佳的候选框, 从而提高模型性能. NMS 算法使用前后的检测窗口变化如图 13 所示.

从图 13(a) 可以看出, 在对采集到的开关图像进行检测时, 开关上有很多交叉重叠的候选窗口, 但并不是所有窗口都可以完整地包含目标开关, 多数窗口只能检测到一部分. 此时, 使用 NMS 算法从图 13(a) 中的候选框中选择最优的窗口, 如图 13(b) 所示, 这样可以提高开关检测的准确度, 模型效果也会更好.



(a) NMS 前开关检测窗口

(b) NMS 后开关检测窗口

图 13 使用 NMS 前后候选框图示结果

假设某张图像在检测过程中, 图像内的一个开关上生成了 n 个候选框, 将候选框集合记为 $B = \{B_1, B_2, \dots, B_n\}$, 候选框对应的得分集合记为 $S = \{S_1, S_2, \dots, S_n\}$, 得分是候选框内有目标的概率大小, S_i 的计算公式如下:

$$S_i = \begin{cases} S_i, & IoU(M, B_i) < p \\ 0, & IoU(M, B_i) \geq p \end{cases} \quad (5)$$

其中, IoU 为两个候选框的交并比.

显然, 从式 (5) 中发现, NMS 算法会直接丢弃所有

超过阈值 p 的候选框, 导致相邻的候选框的分数为 0. 如果相邻候选框内含有类似大小的目标, 那么两个候选框 IoU 可能会高于 p , 相邻目标的检测会丢失一个目标候选框.

针对 NMS 暴力抑制候选框的问题, 本文使用了改进的非极大值抑制算法 Soft-NMS^[33]. Soft-NMS 提出了两种对候选框分数的加权, 分别为线性加权和高斯加权, 本文采用高斯加权方式, 如式 (6) 所示.

$$S_i = \begin{cases} S_i, & IoU(M, B_i) < p \\ S_i \times e^{-\frac{IoU(M, B_i)^2}{\sigma}}, & IoU(M, B_i) \geq p \end{cases} \quad (6)$$

2 实验与结果分析

2.1 实验设计

实验所使用的数据集是第 1 节中经过数据增强的设备开关状态数据集, 一共 1 644 张图像, 开关约有 3 600 个, 分为 18 种开关 36 类状态. 以 7:3 的比例划分训练集与测试集, 训练集 1 151 张图像, 测试集 493 张图像.

在参数设置方面, 优化器选择 SGD, 学习率 lr 设置为 0.005, 衰减系数 weight_decay 设置为 0.000 5, 动量参数 momentum 设置为 0.9, 批大小 batch_size 设置为 2, 迭代次数 epoch 设置为 50 次.

在抽水蓄能电站真实的操作场景下, 本文对目标检测模型 Faster R-CNN 进行了 4 个方面的改进, 分别为: 设计了膨胀残差网络作为骨干网络、使用改进后的特征金字塔网络进行多尺度信息融合、使用 K-means++ 算法聚类出候选框尺寸、使用 Soft-NMS 代替 NMS. 为了验证改进的有效性, 设计了以下 5 组对比试验:

(1) 不同多分支膨胀卷积使用策略的对比实验: 先验证膨胀卷积对提升识别效果的有效性, 然后对膨胀卷积的层组合进行实验, 得出最合适的膨胀卷积的层组合.

(2) 特征金字塔结构改进前后的对比实验: 使用膨胀残差网络作为骨干网络, 对比加 FPN 与不加 FPN 模型测试效果.

(3) 使用原模型的候选框尺寸与 K-means++ 算法聚类出的候选框之间的对比实验: 验证设计的候选框尺寸可以提升模型精度.

(4) Soft-NMS 与 NMS 的对比实验: 使用两种筛选候选框的算法, 测试最终模型效果.

(5) 本文模型与其他目标检测模型之间的对比实验: 将改进后的 Faster R-CNN 与原 Faster R-CNN 模

型、SSD、YOLO 进行实验对比,原模型特征提取网络均使用 VGG16. 该组实验主要是为了验证改进的有效性且将改进的 Faster R-CNN 与其他目标检测算法进行对比实验.

衡量一个目标检测模型的精准度需要一些评价指标来表示,对于目标检测模型,常用的度量方法有精确率、召回率、平均精度、均值平均精度、交并比等.

(1) 精确率 (Precision)

精确率表示检测出开关并正确识别出的开关状态的数量占测试集中所有识别出的开关状态的比例,简记为 P , 计算公式如下:

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

其中, TP 是检测出开关并正确识别出开关状态的数量, FP 是检测出开关但识别状态错误的数量.

(2) 召回率 (Recall)

召回率表示正确检测出的开关数量占测试集中所有开关总数的比例,简记为 R , 计算公式如下:

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

其中, TP 是检测出开关状态的数量, FN 是没有检测出开关的数量.

(3) F1 分数 (F1 score)

F1 分数是 Precision 和 Recall 的调和平均值,简记为 $F1$, 公式如下:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

(4) 准确率 (Accuracy)

一般用来评估模型的全局准确程度,简记为 Acc , 计算公式如下:

$$Acc = \frac{TP + TN}{\sum N_{total}}$$

其中, TP 是检测出开关并正确识别出开关状态的数量, TN 表示检测出非开关状态的数量, $\sum N_{total}$ 表示所有类开关的数量之和.

(5) 平均精度 (average precision, AP)

平均精度 AP 常被用来衡量模型对某一单个类别的检测精度. 事实上, AP 是 $P-R$ 曲线下的面积, 也就是单个类别图像的精度 $Precision$ 和所有图像数量的比值, 通常 AP 越高, 表示模型对这个类别的检测性能就越好. 单个开关状态 AP 的公式如下:

$$AP_{switch} = \frac{\sum Precision_{switch}}{N_{total}}$$

其中, AP_{switch} 表示单个类别开关状态的平均精度, $\sum Precision_{switch}$ 是这一类别开关状态的精度和, N_{total} 是单个类别开关的总数量.

(6) 均值平均精度 (mean average precision, mAP)

均值平均精度 mAP , 它用来衡量模型某类图像的检测精度. 通常 mAP 的值越高, 表示此模型的检测性能越好. mAP 的计算公式如下:

$$mAP_{switch} = \frac{\sum AP_{switch}}{N_{classstotal}}$$

其中, AP_{switch} 表示某类开关状态的平均精度值, $\sum mAP_{switch}$ 是所有类开关状态的精度和, $N_{classstotal}$ 是开关状态类别数.

2.2 实验结果分析

开关状态识别效果评价指标主要使用 Precision、F1 score、Accuracy 和 mAP (在实验中分别简记为 P , $F1$, Acc 和 mAP), IoU 均设置为 0.5. 将从第 2.1 节的 5 个方面进行实验并对比分析, 验证本文改进算法的有效性.

(1) 不同多分支膨胀卷积使用策略的对比实验

首先, 通过一组实验验证膨胀卷积的有效性, 将 ResNet50 中的 conv4 的残差块换成膨胀卷积, 其结果如表 3 所示.

表 3 使用不同扩张率膨胀卷积的对比实验 (%)

扩张率	P	$F1$	Acc	mAP
1	64.23	67.59	88.47	85.69
2	65.41	68.51	89.23	86.14
3	65.37	68.42	88.92	86.12

由表 3 结果可知, 在开关数据集上, 仅改变残差块的感受野大小可以提升模型检测的效果.

从图 8 的特征图中可以看出, conv1、conv2 输出的特征图与原图像相比信息丢失较少, 所以本文在 conv3、conv4、conv5 这 3 个模块使用膨胀卷积, 然后对 conv3、conv4、conv5 使用多分支膨胀卷积的组合形式进行了实验, 结果如表 4 所示.

由表 4 可知, 在 ResNet50 中的 conv3、conv4、conv5 使用多分支膨胀卷积在开关状态数据集上效果最好, 比原模型 mAP 高了 6%.

(2) 改进的特征金字塔结构使用前后的对比实验

本文改进了 FPN 结构, 在原 FPN 基础上新加一条自底向上特征增强的分支, 做了两组对比实验, 分别在

骨干网络为 ResNet50 和膨胀残差网络的 Faster R-CNN 模型上对比 FPN 改进前后模型的精度. 实验结果如表 5 所示.

表 4 多分支膨胀卷积不同层组合的对比实验 (%)

使用膨胀卷积的层组合	<i>P</i>	<i>F1</i>	<i>Acc</i>	<i>mAP</i>
不使用	64.23	67.59	88.47	85.69
conv3, 4	65.38	68.42	88.92	86.12
conv3, 5	65.02	68.38	88.58	85.76
conv4, 5	65.41	68.51	89.23	86.14
conv3, 4, 5	65.47	68.54	89.27	86.27

表 5 特征金字塔结构使用前后的对比实验 (%)

网络模型	<i>P</i>	<i>F1</i>	<i>Acc</i>	<i>mAP</i>
ResNet50+FPN	66.02	68.44	89.24	89.79
ResNet50+改进的FPN	70.89	74.02	92.06	90.62
膨胀残差网络+FPN	70.43	73.72	91.91	90.48
膨胀残差网络+改进的FPN	71.36	75.21	92.43	91.02

由表 5 可知, 在两个骨干网络上使用 FPN 的效果都相当明显, 以 ResNet50 为骨干网络的模型精度提升了 4.1%; 使用改进后的 FPN, 提升了对小目标的检测能力, 比改进前提升约 0.8%, 膨胀残差网络作为骨干网络的模型提升了约 0.5%. 说明改进的 FPN 对开关状态的识别有作用.

(3) 候选框聚类前后的对比实验

以上一组实验为基础, 在 Faster R-CNN 的 RPN 网络中设置适合数据集候选框大小和宽高比例, 将使用原 RPN 的候选框尺寸和使用 K-means++ 算法得出的尺寸进行对比实验, 结果如表 6 所示.

表 6 候选框聚类使用前后的对比实验 (%)

网络模型	<i>P</i>	<i>F1</i>	<i>Acc</i>	<i>mAP</i>
ResNet50+FPN	66.02	72.54	90.06	89.79
ResNet50+FPN+聚类	72.04	74.02	91.06	90.28
膨胀残差网络+FPN	71.36	75.21	92.43	91.02
膨胀残差网络+FPN+聚类	74.18	77.24	93.69	91.43

由表 6 可知, 在开关状态数据集上候选框聚类对模型精度的识别有提升效果, 主要原因是开关形状都是规则的, 大多数的开关具有相似的大小和高宽比例. 所以候选框聚类对提升模型识别精度有效.

(4) 使用 Soft-NMS 与使用 NMS 的对比实验

原 Faster R-CNN 模型使用的是传统的 NMS 算法来抑制重叠的候选框, 物体重叠时, NMS 的暴力丢弃容易导致丢失目标, 本文改进的模型使用 Soft-NMS, 提高了模型抑制候选框的效果, 从而提升模型精度.

由表 7 可知, Soft-NMS 对模型的识别效果有提升作用. 因为数据集中含有较多的电压开关, 且这些开关紧密排列, 在检测这些开关时, 各个目标的多个候选框容易重叠, Soft-NMS 比传统的 NMS 更有效地抑制重叠的候选框.

表 7 使用 Soft-NMS 与 NMS 的对比实验 (%)

网络模型	<i>P</i>	<i>F1</i>	<i>Acc</i>	<i>mAP</i>
膨胀卷积残差网络+FPN+聚类	74.18	77.24	93.69	91.43
膨胀卷积残差网络+FPN+聚类+Soft-NMS	74.36	77.61	93.84	91.50

(5) Faster R-CNN 与其他目标检测模型之间的对比实验

本节主要介绍改进前后的 Faster R-CNN 与 YOLOv3、SSD 的比较实验, 其中, YOLOv3、SSD 均使用原始的骨干网络, 实验比较结果如表 8 所示.

表 8 主流模型与本文模型的对比实验 (%)

网络模型	特征提取网络	<i>P</i>	<i>F1</i>	<i>Acc</i>	<i>mAP</i>
Faster R-CNN	VGG-16	64.32	63.26	83.27	84.47
Faster R-CNN	ResNet50	65.01	66.31	86.30	85.61
SSD	VGG-16	65.82	67.94	89.46	87.53
YOLOv3	Darknet-53	71.24	73.36	90.97	90.02
改进的Faster R-CNN	膨胀残差网络	74.36	77.61	93.84	91.50

本文模型在开关状态数据集上比传统的 Faster R-CNN 检测精度提升了约 7%. 而且模型能较好地适用于抽水蓄能电站设备的开关状态识别, 满足了预期的识别精度要求, 能够辅助监督抽水蓄能电站日常运维操作.

不难发现, 模型的精确率 *Precision* 和 *F1 score* 的数值较低, 这是由于部分类没有检测出目标, 识别出开关状态, 经过平均后数值较低.

为了进一步分析和验证模型训练的稳定性以及收敛性, 对于改进的 Faster R-CNN 模型, 给出了模型的损失收敛曲线和精度收敛曲线. 结果如图 14、图 15 所示.

从图 14 中的曲线可以看出, 模型在迭代 4 000 step 之后损失基本收敛, 从图 15 中的曲线可以看出, 模型精度在训练 40 epoch 之后基本收敛. 所以模型的训练过程稳定, 收敛速度较快.

最后, 对于改进的 Faster R-CNN, 测试模型在其他开关数据集上的检测效果. 改进后 Faster R-CNN 对抽水蓄能电站设备开关状态识别的效果图, 如图 16 所示. 可以发现, 模型可以很好地检测出图像中的开关, 并且对开关状态的分类准确率很高, 能够起到辅助监督的作用.

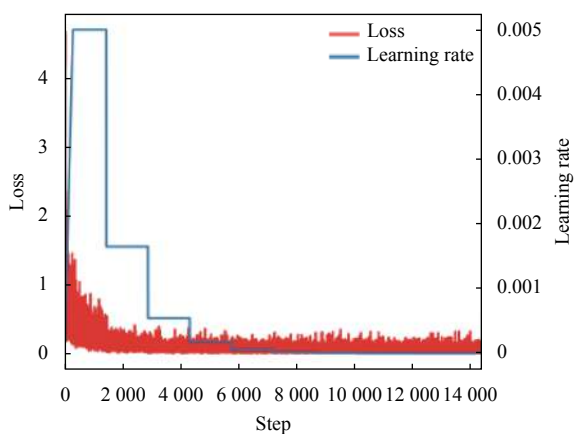


图 14 改进模型的损失收敛曲线

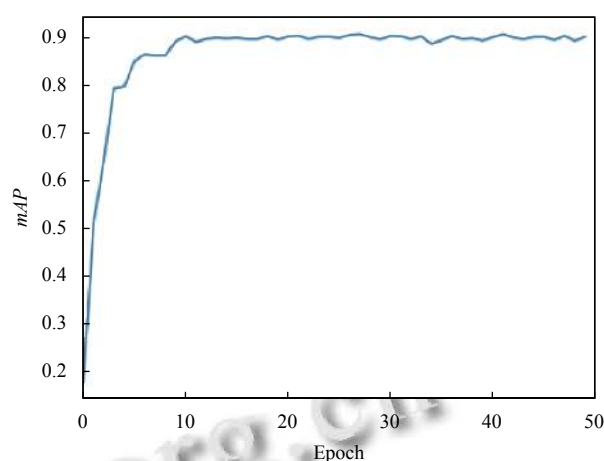


图 15 改进模型的 mAP 收敛曲线

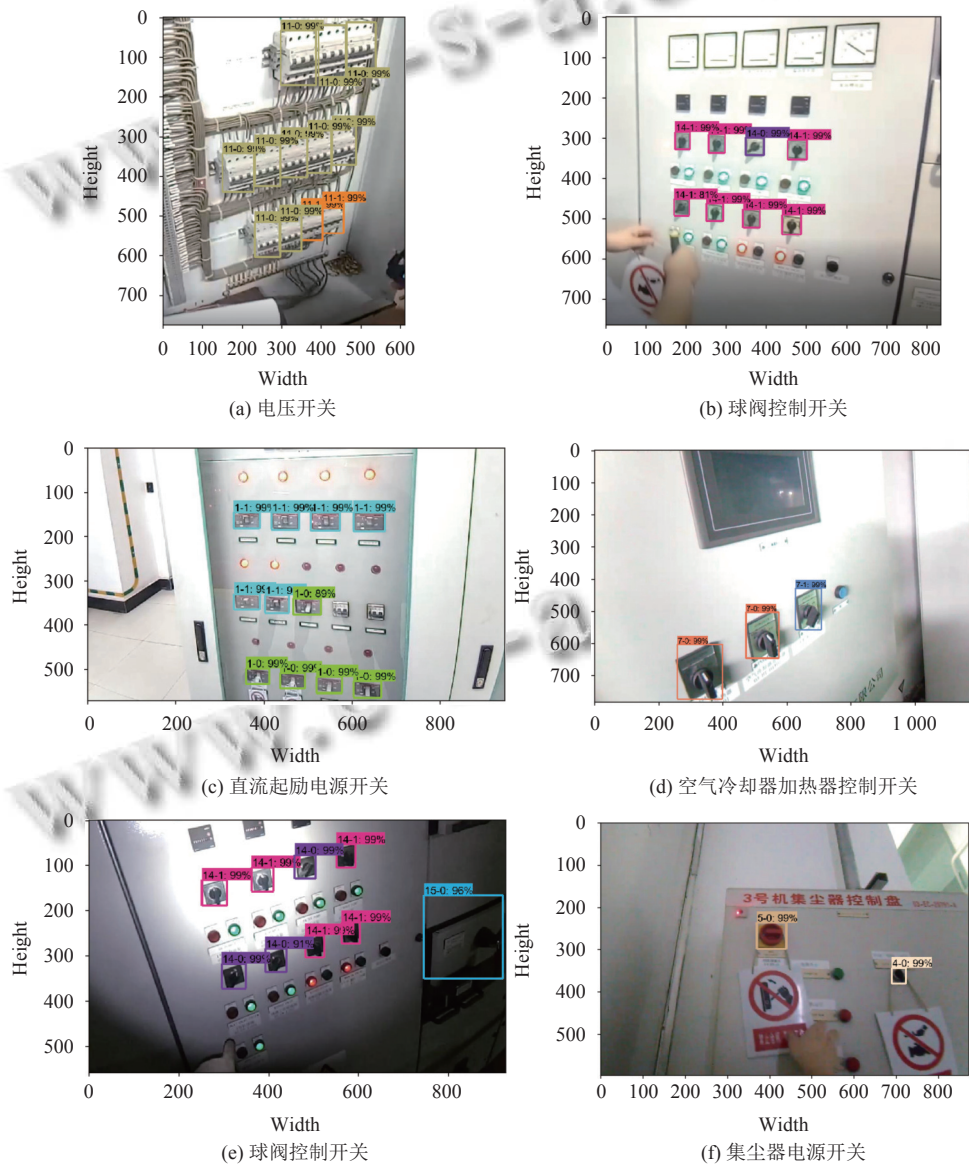


图 16 改进的 Faster R-CNN 检测识别效果

3 总结与展望

针对大型厂房中设备开关状态识别问题,提出了改进的 Faster R-CNN: 设计了膨胀残差网络作为骨干网络,提升了模型特征提取的性能;改进特征金字塔网络进行多尺度融合,提升小型开关的检测效果;使用 K-means++ 算法对开关的真实边界框进行聚类,设计出更适合设备开关状态数据集的候选框尺寸;使用 Soft-NMS 代替 NMS,降低了 NMS 暴力丢弃候选框对开关检测的影响。最终,模型的 mAP 达到 91.5%,并且已实际应用于抽水蓄能电站日常运维的设备开关状态辅助识别,满足复杂场景下的智能监管需求。

在设备开关状态识别问题上,实际应用中由于拍摄角度可能严重偏离设备开关,从而影响采集的开关状态图像的识别精度,这一问题仍然值得进一步研究。

参考文献

- 1 彭小圣, 邓迪元, 程时杰, 等. 面向智能电网应用的电力大数据关键技术. 中国电机工程学报, 2015, 35(3): 503–511.
- 2 Li B, Li T, Xu NW, *et al.* Stability assessment of the left bank slope of the Baihetan Hydropower Station, Southwest China. *International Journal of Rock Mechanics and Mining Sciences*, 2018, 104: 34–44. [doi: 10.1016/j.ijrmms.2018.02.016]
- 3 Zhu YM, Chen SJ, Ma GW, *et al.* Complementary operation of a small cascade hydropower station group and photovoltaic power stations. *Clean Technologies and Environmental Policy*, 2020, 22(7): 1565–1578. [doi: 10.1007/s10098-020-01896-x]
- 4 周凯, 唐墨, 胡志坚, 等. 基于 NSST 图像融合的变电站开关状态识别. 高压电器, 2021, 57(10): 50–58.
- 5 孟令枫, 杨兴, 于晓春, 等. 基于直线检测法的变电站开关状态图像识别系统的研究. 电子质量, 2017, (4): 5–8. [doi: 10.3969/j.issn.1003-0107.2017.04.002]
- 6 胡聪, 施保华, 王俊. 智能变电站隔离开关状态图像识别新方法. 电力学报, 2019, 34(5): 498–504.
- 7 汪洋, 黎恒焜, 鄂士平, 等. 基于深度学习的变电站硬压板状态检测与识别算法. 沈阳工业大学学报 2020, 42(6): 676–680.
- 8 周克, 杨倩文, 王耀艺, 等. 一种改进的压板状态识别 SSD 算法. 电测与仪表, 2021, 58(1): 69–76.
- 9 许根养, 翁士状, 孙长翔, 等. 基于相位特征的保护压板状态的准确识别方法. 安徽大学学报 (自然科学版), 2020, 44(3): 38–42.
- 10 阮浩浩, 王春亚, 王辉, 等. 基于图像识别的保护压板状态识别技术研究与应用. 电力设备管理, 2020, (1): 139–140.
- 11 黄钊鹏. 变电站保护压板中开关的图像识别方法研究 [硕士学位论文]. 武汉: 武汉理工大学, 2019.
- 12 冷从林. 基于机器视觉的变电站压板开关状态识别系统研究 [硕士学位论文]. 武汉: 武汉理工大学, 2019.
- 13 李琛. 基于图像识别的电力智能巡检系统设计与实现 [硕士学位论文]. 南京: 南京大学, 2018.
- 14 史梦安, 陆振宇. 基于 YOLO-tiny-RFB 模型的电站旋钮开关状态识别. 计算机应用, 2020, 40(12): 3679–3686.
- 15 杨雨, 陈服, 陈柏青, 等. 基于深度学习的变电站保护硬压板检测与状态识别技术研究. 科技创新与应用, 2021, 11(24): 25–29.
- 16 李岩, 李丹, 李建, 等. 基于机器视觉的电力设备监测系统的研究. 中小企业管理与科技, 2021, (11): 194–196.
- 17 苏杨, 余萱. 基于卷积神经网络的机房柜门开关状态识别研究. 工业仪表与自动化装置, 2019, (5): 22–25, 47. [doi: 10.3969/j.issn.1000-0682.2019.05.005]
- 18 邓乐武, 成金涛, 曾苏凡. 基于深度学习的座舱开关状态识别研究. 测控技术, 2021, 40(8): 54–57.
- 19 高露. 基于深度学习的变电站开关状态识别系统的设计与实现 [硕士学位论文]. 沈阳: 中国科学院大学, 2019.
- 20 Khan D. 基于深度学习的电力开关检测与状态识别 [硕士学位论文]. 成都: 电子科技大学, 2018.
- 21 张礼波, 雷鸣, 周怡君. 基于目标跟踪的变电站隔离开关状态图像识别方法. 机械与电子, 2020, 38(1): 36–39. [doi: 10.3969/j.issn.1001-2257.2020.01.009]
- 22 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149. [doi: 10.1109/TPAMI.2016.2577031]
- 23 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu: IEEE, 2017. 936–944.
- 24 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv: 1804.02767, 2018.
- 25 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot MultiBox detector. *Proceedings of the 14th European Conference on Computer Vision*. Amsterdam: Springer, 2016. 21–37.
- 26 Arthur D, Vassilvitskii S. K-means++: The advantages of careful seeding. *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*. New Orleans Louisiana: Society for Industrial and Applied Mathematics, 2007. 1027–1035.

- 27 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778.
- 28 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, 2014.
- 29 Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. arXiv: 1511.07122, 2015.
- 30 Li YH, Chen YT, Wang NY, *et al.* Scale-aware trident networks for object detection. Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 6053–6062.
- 31 Liu ST, Huang D, Wang YH. Receptive field block net for accurate and fast object detection. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 404–419.
- 32 Yang X, Sun H, Fu K, *et al.* Automatic ship detection in remote sensing images from Google Earth of complex scenes based on multiscale rotation dense feature pyramid networks. Remote Sensing, 2018, 10(1): 132. [doi: [10.3390/rs10010132](https://doi.org/10.3390/rs10010132)]
- 33 Bodla N, Singh B, Chellappa R, *et al.* Soft-NMS-improving object detection with one line of code. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 5562–5570.

(校对责编: 孙君艳)