

## 基于改进 Faster R-CNN 的接触网管帽目标定位算法

顾桂梅<sup>1</sup>, 陈充<sup>1\*</sup>, 余晓宁<sup>1</sup>, 张存俊<sup>2</sup>, 仝甄<sup>3</sup>, 梅小芸<sup>1</sup><sup>1</sup>兰州交通大学自动化与电气工程学院, 甘肃 兰州 730070;<sup>2</sup>中国铁路兰州局集团有限公司, 甘肃 兰州 730030;<sup>3</sup>国网甘肃省电力公司庆阳供电公司, 甘肃 庆阳 745000

**摘要** 为了改善接触网管帽这类小尺度部件在故障检测过程中定位困难的情况, 提出一种基于改进 Faster R-CNN 的接触网管帽目标定位算法。通过 K 均值聚类算法(K-means)对 region proposal network (RPN) 层中生成 anchor boxes 的比例及面积进行改进, 所提算法在定位接触网管帽这类小部件上具有较好的表现。并通过比较 VGG16、resnet50、resnet101、resnet152 等 4 种特征提取网络在原始及改进的 Faster R-CNN 上定位管帽的准确率、召回率、准确率和召回率的调和平均  $F_1$ 、单张检测时间等指标来选择最优特征提取网络。实验结果表明, 基于 resnet50 的改进 Faster R-CNN 深度网络模型在接触网管帽定位中具有明显的优势, 召回率为 89.78%, 定位准确率可以达到 83.16%,  $F_1$  值为 86.34%, 单张检测时间为 0.283 s。

**关键词** 图像处理; 接触网管帽; 定位; Faster R-CNN; K 均值聚类算法; 深度学习

中图分类号 U225.4

文献标志码 A

doi: 10.3788/LOP202259.0410009

## Target Location Algorithm of Contact Network Pipe Cap Based on Improved Faster R-CNN

Gu Guimei<sup>1</sup>, Chen Chong<sup>1\*</sup>, Yu Xiaoning<sup>1</sup>, Zhang Cunjun<sup>2</sup>, Tong Zhen<sup>3</sup>, Mei Xiaoyun<sup>1</sup><sup>1</sup>School of electrical engineering and automation, Lanzhou Jiaotong University, Lanzhou, Gansu 730070, China;<sup>2</sup>China Railway Lanzhou Bureau Group Co., Ltd., Lanzhou, Gansu 730030, China;<sup>3</sup>Qingyang Power Supply Company of State Grid Gansu Electric Power Company, Qingyang, Gansu 745000, China

**Abstract** In order to improve the difficulty of locating small-scale components such as contact network pipe cap in the process of fault detection, a contact network pipe cap target location algorithm based on improved Faster R-CNN is proposed. The proportion and area of anchor boxes generated in the region proposal network (RPN) layer are improved by K-means clustering algorithm (K-means), the proposed algorithm has good performance in locating small components such as contact network pipe caps. The optimal feature extraction network is selected by comparing the accuracy, recall, accuracy, harmonic average of accuracy and recall  $F_1$ , and single sheet detection time of VGG16, resnet50, resnet101, and resnet152 feature extraction networks on the original and improved Faster R-CNN. The experimental results show that the improved Faster R-CNN deep network model based on resnet50 has obvious advantages in contact network pipe cap locating, the recall rate is 89.78%, the locating accuracy can reach 83.16%, the  $F_1$  value is 86.34%, and the single detection time is 0.283 s.

**Key words** image processing; pipe cap; positioning; Faster R-CNN; K-means clustering algorithm; deep learning

收稿日期: 2021-02-02; 修回日期: 2021-02-21; 录用日期: 2021-03-23

基金项目: 甘肃省自然科学基金(20JR10RA216)、甘肃省科技计划(20JR10RA216)

通信作者: \*1334065344@qq.com

## 1 引言

接触网管帽存在于接触网支持与定位装置中,是接触网安全运行不可或缺的组成部分。接触网中管帽起到保护空心腕臂钢管和定位管的作用,管帽缺失会使尘土、雨水和雪等易进入腕臂钢管与定位管中,使腕臂与定位管的使用寿命大大缩减,而且会导致腕臂的绝缘效果变差,从而影响列车的运行安全。相对于接触网支持与定位装置中的其他部件而言,管帽是一个较小的部件,但管帽缺失是发生频率较高的故障之一。管帽定位是后续的接触网管帽故障诊断的基础,具有较为重要的实际工程意义。

近些年,利用图像定位算法定位接触网部件由于具有安全性高及不影响运输秩序等特点,在接触网部件定位方面受到越来越多人的青睐<sup>[1]</sup>。目前接触网部件定位算法主要有基于机器视觉的接触网管帽及 U 型抱箍故障检测算法<sup>[2]</sup>、基于 scale-invariant feature transform (SIFT) 的旋转双耳定位与提取算法<sup>[3]</sup>、基于快速模糊匹配提取棒式绝缘子的目标区域算法<sup>[4]</sup>、基于定位线夹区域结构特征的目标检测算法<sup>[5]</sup>、基于 speeded up robust features (SURF) 特征匹配的旋转双耳状态检测算法<sup>[6]</sup>等。上述接触网部件定位算法都是传统的图像定位算法。基于机器视觉的接触网管帽及 U 型抱箍故障检测算法利用直线检测成功定位了接触网管帽,该方法虽然有着不俗的准确率,但是该方法只能定位含 3 个管帽的接触网,不能定位含 2 个管帽的接触网,而且采集到的含管帽的接触网图片不完整时,无法进行直线检测导致无法定位管帽。传统的接触网部件定位算法是由特征提取和分类两部分组成的,且两部分是分开进行设计的。特征提取算法需要研究人员针对不同的情况提出不同的特征提取算法,对设计算法的人具有极高的专业要求,而且在面对一些复杂情况时,往往很难找到优秀的特征提取算法,并且人工设计特征提取算法工作量大且鲁棒性不高,因此传统的接触网部件定位算法有着很大的局限性。使用深度学习来实现接触网部件的定位相对于传统的接触网部件定位算法可以大幅度地减少定位时间和设计特征提取算法时消耗的人力,并且基于深度学习的接触网部件定位算法相对于传统的接触网部件定位算法拥有更好的鲁棒性。相对于传统算法,基于深度学习的算法由于强大的神经网络可以更快、更精确地完成图像的

检测和分类。文献[7]提出了一种基于生成对抗网络(GAN)的等电线故障状态检测算法。文献[8]利用基于卷积神经网络(CNN)的对象检测算法来检测多个接触网支撑部件,并提出了一种统一的 CNN 架构,但该算法仅仅做了器件定位,而且并未对管帽进行研究。文献[9]设计了一种自动快速定位系统,该系统可以同时定位 12 类多尺度接触网支撑组件,但该文献中的定位并未涉及管帽。深度学习中有许多优质的网络和检测框架,如 visual geometry group (VGG)<sup>[10]</sup>、residual network (Resnet) 系列网络<sup>[11]</sup>和 Faster R-CNN<sup>[12]</sup>、single shot MultiBox detector (SSD)<sup>[13]</sup>、you only look once (YOLO)<sup>[14]</sup>等检测框架,其中 Faster R-CNN 为二阶分类器,SSD、YOLO 为一阶分类器,二阶分类器具有较高的准确率但是检测速度慢,一阶分类器具有较高的检测速度但是检测准确率较低。文献[15]用反卷积与捷径对 CNN 进行改进,用在线困难样本挖掘优化了 Faster R-CNN 的损失函数,提高了对毫米波图像的检测效果。文献[16]用基于 3 种特征提取网络的 Faster R-CNN 算法对油菜田间杂草进行识别。文献[17]用基于深度学习的算法实现了对车辆识别结果的有效编号及车位空缺的准确判断。文献[18]用建立特征融合层的方式改进了 YOLOv3 算法,提高了对小目标的检测效果。针对接触网管帽目标定位,本实验组选择二阶 Faster R-CNN 检测框架。

在接触网 1C~6C 系统中,4C 为接触网悬挂状态检测监测装置,主要功能是检测接触网悬挂部件技术状态及检测接触网几何参数,主要拍摄接触网接触悬挂、支持装置、定位装置、附加悬挂。所研究的管帽位于接触网支持装置和定位装置中,根据接触网的型号不同,接触网的管帽数量可能为 2 个或 3 个。由于接触网管帽目标较小,且 4C 通常都在夜间作业,在 4C 所采集到的图像中,部分黑色管帽由于黑色背景特征不明显。因此,精确定位管帽具有较高的难度。目前国内外对管帽定位的研究较少。由于小尺度目标在整个图像中所占的比例很小,以接触网管帽为例,本实验采用由 4C 装置拍摄的尺寸为 5120×5120 的含管帽图片,根据拍摄角度的不同,每个管帽在图中所占像素大小也不同,一个管帽在采集到的原始图像中所占的像素大小约为 280×280。以网络深度相对较浅的 VGG16 特征提取网络为例,为了更高效地训练模型及减小训练模型时所占的内存与显存,将采集到的分辨率为

5120×5120 的含有管帽的接触网原始图像进行压缩处理,将图片缩放到分辨率为 512×512,这样一来图片中的一个管帽所占的像素大小就变为 28×28 左右。随后将缩放过后的图片输入基于 VGG16 的 Faster R-CNN 深度学习网络进行特征提取,512×512 大小的图片再经过 VGG16 特征提取网络之后生成的特征图大小为 32×32,此时一个管帽所占的像素大小为 2×2 左右。网络从这些少量像素中提取到有效特征的能力非常有限,也导致生成的模型对接触网管帽这种小尺度部件的检测能力较弱。

针对上述问题,本文提出一种基于改进 Faster R-CNN 的接触网管帽目标定位算法。通过 K 均值聚类算法(K-means)对 region proposal network(RPN)层中生成 anchor boxes 的比例及面积进行改进。实验结果表明,所提算法在定位接触网管帽这种小尺度部件上具有较好的效果。

## 2 基于改进的 Faster R-CNN 接触网管帽目标定位算法

### 2.1 K-means 算法

K-means 算法是一种无监督的迭代求解的聚类分析算法,流程如下。

- 1) 将数据集  $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$  指定划分为  $k$  个簇,  $k$  是根据情况事先给定的聚类数。
- 2) 随机选取  $k$  个聚类中心  $\mu_1, \mu_2, \dots, \mu_k \in \mathbf{R}^n$ 。

3) 对于每一个样例  $i$ , 计算其应该属于的类, 表达式为

$$c^{(i)} = \arg \min_j \|x^{(i)} - \mu_j\|^2, \quad (1)$$

式中:  $c^{(i)}$  代表样例  $i$  与  $k$  个类中距离最近的类, 取值为  $[1, k]$ ;  $\mu_j$  为质心, 为同一类的样本中心点。

4) 对于每一个类样例  $j$ , 重新计算该类的质心, 表达式为

$$\mu_j = \frac{\sum_{i=1}^m 1\{c^{(i)}=j\} x^{(i)}}{\sum_{i=1}^m 1\{c^{(i)}=j\}}. \quad (2)$$

5) 重复上述步骤 3)~4), 直至收敛。

### 2.2 改进的 Faster R-CNN 网络模型

Faster R-CNN 深度学习模型由特征提取网络、RPN、分类、回归网络组成, 如图 1 所示, 其中圈出来的为使用 K-means 算法改进 anchor boxes 的部分。特征提取(feature extraction network)用于提取图像深层与浅层的特征, 进而生成特征图, 且与之后的 RPN 层共享卷积特征。以 VGG16 特征提取网络为例, VGG16 网络由 13 个卷积层与 3 个全连接(FC)层组成, 故得名 VGG16。基于 VGG16 的 Faster R-CNN 只用到 VGG16 网络的 13 层卷积层来提取图像特征, 由于在特征提取网络不使用 VGG16 的全连接层, 故对输入 VGG16 网络图像的大小并不做限制。VGG16 卷积层采用大小为 3×3 的卷积核, 激活函数采用非线性 ReLU 函数, 卷积层后接(2×2)最大池化层用于缩放图片。VGG16 特征提取网络参数如表 1 所列。

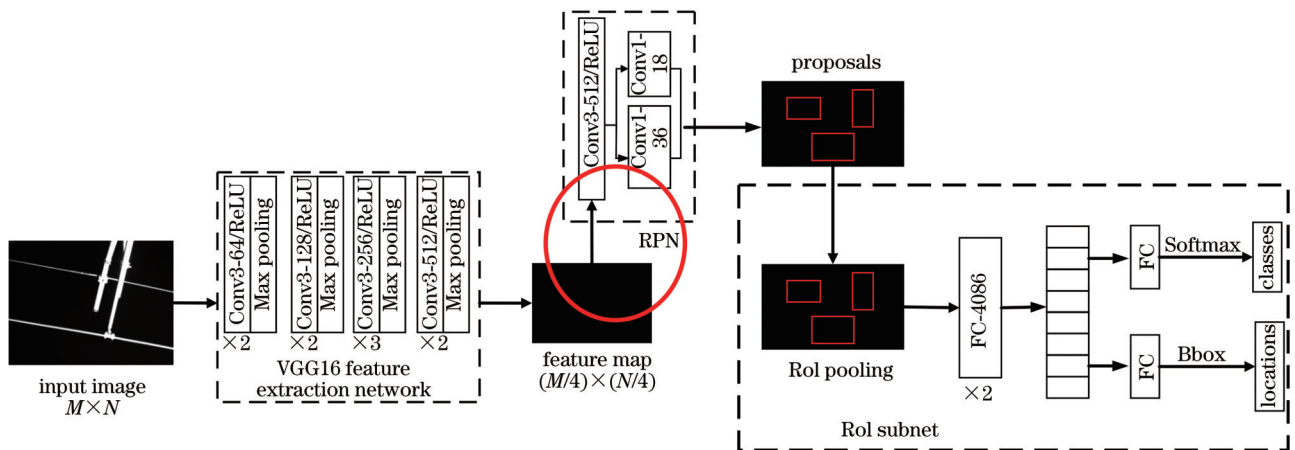


图 1 基于 VGG16 的 Faster R-CNN 深度学习模型

Fig. 1 Faster R-CNN deep learning model based on VGG16

RPN 为区域建议框生成网络, 是 Faster R-CNN 相对于 R-CNN 算法和 Fast R-CNN 算法的重要改进, Faster R-CNN 采用 RPN 代替 selective search

(SS) 算法, 从而大幅度提高了算法的运行速度。RPN 的本质为滑动窗口, 将特征提取网络生成的特征图作为输入, 在特征图上用一个大小为 3×3、步



表1 VGG16网络卷积过程参数

Table 1 Parameters of VGG 16 network convolution process

Layer	Input	Convolution kernel	Step length	Output
Conv3-64	$M \times N \times 3$	$3 \times 3$	1	$M \times N \times 64$
Conv 3-64	$M \times N \times 64$	$3 \times 3$	1	$M \times N \times 64$
Pool 1	$M \times N \times 64$	$2 \times 2$	2	$(M/2) \times (N/2) \times 64$
Conv3-128	$(M/2) \times (N/2) \times 64$	$3 \times 3$	1	$(M/2) \times (N/2) \times 128$
Conv3-128	$(M/2) \times (N/2) \times 128$	$3 \times 3$	1	$(M/2) \times (N/2) \times 128$
Pool 2	$(M/2) \times (N/2) \times 128$	$2 \times 2$	2	$(M/4) \times (N/4) \times 128$
Conv3-256	$(M/4) \times (N/4) \times 128$	$3 \times 3$	1	$(M/4) \times (N/4) \times 256$
Conv3-256	$(M/4) \times (N/4) \times 256$	$3 \times 3$	1	$(M/4) \times (N/4) \times 256$
Conv3-256	$(M/4) \times (N/4) \times 256$	$3 \times 3$	1	$(M/4) \times (N/4) \times 256$
Pool 3	$(M/4) \times (N/4) \times 256$	$2 \times 2$	2	$(M/8) \times (N/8) \times 256$
Conv3-512	$(M/8) \times (N/8) \times 256$	$3 \times 3$	1	$(M/8) \times (N/8) \times 512$
Conv3-512	$(M/8) \times (N/8) \times 512$	$3 \times 3$	1	$(M/8) \times (N/8) \times 512$
Conv3-512	$(M/8) \times (N/8) \times 512$	$3 \times 3$	1	$(M/8) \times (N/8) \times 512$
Pool 4	$(M/8) \times (N/8) \times 512$	$2 \times 2$	2	$(M/16) \times (N/16) \times 512$
Conv3-512	$(M/16) \times (N/16) \times 512$	$3 \times 3$	1	$(M/16) \times (N/16) \times 512$
Conv3-512	$(M/16) \times (N/16) \times 512$	$3 \times 3$	1	$(M/16) \times (N/16) \times 512$
Conv3-512	$(M/16) \times (N/16) \times 512$	$3 \times 3$	1	$(M/16) \times (N/16) \times 512$

距为 1 的滑动窗口进行滑动。滑动窗口的中心点称为 anchor, 每一个 anchor 在原图上对应  $K$  个候选框 (anchor boxes)。对于一张大小为  $1000 \times 600 \times 3$  的图像, 生成的特征图大小为  $60 \times 40 \times 512$ , 大约有  $60 \times 40 \times 9$  个候选框, 将跨越边界的候选框剔除, 大约剩余 6000 个候选框。由于 RPN 生成的候选框存在大量重叠, 基于候选框的 box-classification (cls) 得分, 采用非极大值抑制 (NMS) 将冗余的候选框筛选并删除, IOU 设为 0.7, 每张图片剩余大约 2000 个候选框。最后从剩余 2000 个候选框中选取 cls 得分最高的 300 个为区域建议生成框 (proposals), 并输入 RPN 网络的分类、回归层。RPN 网络结构如图 2 所示。

传统的 anchor boxes 有 3 种宽高比例 (1:2, 1:1, 2:1), 3 种面积 ( $128^2, 256^2, 512^2$ ), 每个位置在原图上都对应有  $3 \times 3$  个 anchor boxes。传统的 anchor boxes 是用来检测 pascal voc 数据集集中的汽车、人、马等大尺度物体, 对于接触网管帽这种小尺寸部件的适用度不高。针对该问题, 采用第 1.1 节的 K-means 算法, 对训练样本中的 xml 标注文件中标注的框宽高比进行聚类, 绘制出  $k$  值与聚类准确率的关系图像, 如图 3(a) 所示。从图 3(a) 中可以看出, 聚类准确率随着  $k$  值的增大而不断增大, 当  $k$  值为 0~12 时, 准确率随着  $k$  值增大而明显增长,  $k$  大于 12

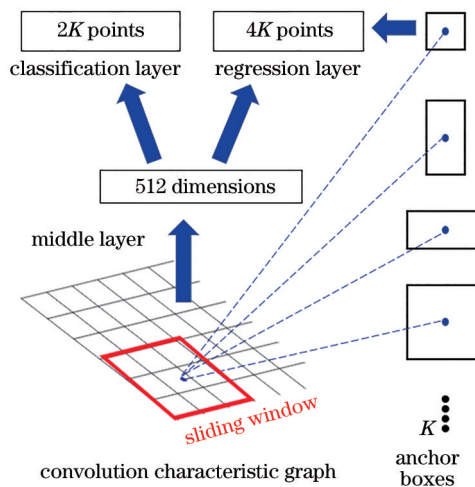
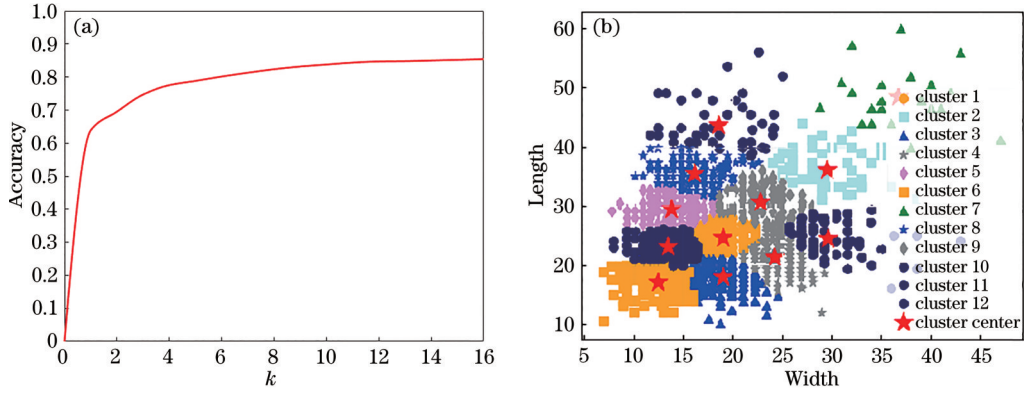


图2 RPN网络结构

Fig. 2 Structure of RPN network

之后, 准确率随着  $k$  值增大的趋势变缓。由于最终需要 3 种聚类结果,  $k$  值过大不利于 3 种聚类结果的确。综上所述, 取  $k=12$ ,  $k=12$  时的聚类结果如图 3(b) 所示。聚类所得的结果为 (0.42, 0.45, 0.47, 0.58, 0.73, 0.74, 0.76, 0.77, 0.82, 1.05, 1.13, 1.21) 12 种宽高比, 将聚类结果分为 3 组, 紧接着取每组的平均数并向上保留一位小数取整之后得到 (0.5, 0.8, 1.1) 3 种宽高比, 对应的比例为分别为 1:2, 4:5, 11:10, 对应的 anchor boxes 的 3 种面积分别为  $128^2, 204^2, 281^2$ 。

图3 聚类。(a)  $k$  值与准确率的关系；(b)  $k=12$  时的聚类结果Fig. 3 Clustering. (a) Relationship between  $k$  value and accuracy; (b) clustering result when  $k=12$ 

将RPN提取的proposals和特征图作为输入,输入感兴趣区域池化层(RoI pooling)将不同大小的RoI固定成大小为 $7 \times 7$ 的RoI特征池化图,之后通过连接的两个全连接层对RoI特征池化图层进行展平处理,将生成的4096维特征向量分别输入分类、回归网络(classification、bounding-box regression network),并在分类层中利用Softmax计算得分,在回归层中利用边框回归(Bbox regressor)得到位置偏移量,从而获得更加准确的目标检测框。

RPN的总体损失函数为

$$L(p_i, \mathbf{t}_i) = \frac{1}{N_{\text{cls}}} \sum_i L_{\text{cls}}(p_i, p_i^*) + \lambda \frac{1}{N_{\text{reg}}} \sum_i p_i^* L_{\text{reg}}(\mathbf{t}_i, \mathbf{t}_i^*), \quad (3)$$

式中: $p_i$ 表示第*i*个anchor预测为真实标签的概率,为正样本时为1,为负样本时为0; $N_{\text{cls}}$ 表示一个mini-batch中的所有样本数量,为256; $N_{\text{reg}}$ 表示anchor位置的个数,约为2400; $L_{\text{cls}}$ 表示分类损失; $L_{\text{reg}}$ 表示回归损失。 $L_{\text{cls}}$ 本质上是目标与非目标的对数损失,表达式为

$$L_{\text{cls}}(p_i, p_i^*) = -\text{lb}[p_i p_i^* + (1 - p_i)(1 - p_i^*)], \quad (4)$$

式中:lb表示对数损失函数。

$L_{\text{reg}}$ 表示回归损失,表达式为

$$L_{\text{reg}}(\mathbf{t}_i, \mathbf{t}_i^*) = \sum_i \text{smoothL1}(\mathbf{t}_i - \mathbf{t}_i^*), \quad (5)$$

$$\text{smoothL1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & |x| \geq 1 \end{cases}, \quad (6)$$

$$\begin{cases} \mathbf{t}_i = [t_x, t_y, t_w, t_h] \\ \mathbf{t}_i^* = [t_x^*, t_y^*, t_w^*, t_h^*] \end{cases}, \quad (7)$$

$$\begin{cases} t_x = (x - x_a) / w_a \\ t_y = (y - y_a) / h_a \end{cases}, \quad (8)$$

$$\begin{cases} t_w = (x^* - x_a) / w_a \\ t_h = (y^* - y_a) / h_a \end{cases}, \quad (9)$$

$$\begin{cases} t_x^* = (x^* - x_a) / w_a \\ t_y^* = (y^* - y_a) / h_a \end{cases}, \quad (10)$$

$$\begin{cases} t_w^* = \log(w^* - w_a) / w_a \\ t_h^* = \log(h^* - h_a) / h_a \end{cases}, \quad (11)$$

式中:smoothL1为光滑的最小绝对值偏差函数(LAD); $\mathbf{t}_i$ 表示预测第*i*个anchor的边界框中心点坐标及宽高; $\mathbf{t}_i^*$ 表示第*i*个anchor对应的真实边框的中心点坐标及宽高; $x, y, w, h$ 分别表示预测框的中心坐标及宽高;变量 $x_a, x^*$ 分别表示anchor、真实框的*x*坐标; $y, w, h$ 同理。

## 3 实验验证

### 3.1 实验平台

本实验使用在Windows 10企业版操作系统下,基于Intel(R)Core(TM)i7-8750H CPU, 2.2 GHz和NVIDIA GeForce GTX 1050Ti GPU的硬件设备搭建TensorFlow深度学习框架,使用Python语言实现接触网管帽目标定位模型的训练与测试。

### 3.2 数据集

本实验使用由4C采集到的1300张大小为 $5120 \times 5120$ 的含管帽的接触网图片进行训练、验证与测试,为了更高效训练、验证与测试模型,并减小训练、验证与测试模型时所占的内存与显存,将分辨率为 $5120 \times 5120$ 的1300张接触网原图进行压缩处理,将图片压缩到分辨率为 $512 \times 512$ 。根据机器

学习相关知识,将采集到的数据集的 80% 左右作为训练集和验证集,其余的作为测试集,训练集、验证集、测试集比例为 8:2:3,从压缩之后的 1300 张图片中随机抽取 300 张作为测试集,其余的按比例分为训练集和验证集。采用 labimg 对压缩之后的 1300 张图片进行标注,并生成 xml 文件。

### 3.3 训练参数

利用 pascal voc 数据集的预训练模型进行初始化参数设置。最大迭代次数设置为 50000 步,每隔 5000 步保存一次模型。模型的优化器选择 momentum 算法, momentum 优化算法的表达式为

$$v = \beta v + (1 - \beta)dw', \quad (12)$$

$$w' = w' - \alpha v, \quad (13)$$

式中:  $dw$  为原始梯度;  $v$  是用指数加权平均计算出来的梯度;  $\beta$  为动量因子,一般设置为 0.9;  $\alpha$  是学习率,设置为 0.01;  $w'$  为待训练的网络参数。利用测试集对模型效果进一步验证,并保留置信度大于 0.7 的目标,最后将保留的目标输出在测试集上。

### 3.4 模型评价指标

为评价所提模型解决接触网管帽目标定位时

的有效性,采用准确率、召回率和准确率和召回率的调和平均进行模型评价,相应公式为

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \times 100\%, \quad (14)$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \times 100\%, \quad (15)$$

$$F_1 = \frac{2PR}{P + R} \times 100\%, \quad (16)$$

式中:  $P$  为准确率;  $R$  为召回率;  $F_1$  为准确率和召回率的调和平均;  $N_{TP}$  为正确定位管帽的数量;  $N_{FP}$  为错误定位管帽的数量;  $N_{FN}$  为未定位出的管帽的数量。

### 3.5 模型训练、测试结果及分析

为了在接触网管帽目标定位的 Faster R-CNN 深度网络模型中选择最优的特征提取网络,分别利用 4 种不同深度、不同结构的特征提取网络进行训练,这 4 种特征提取网络分别为 VGG16、resnet50、resnet101、resnet152。随着迭代次数的增加,分别对基于不同特征提取网络的原始 Faster R-CNN 深度网络模型及用 K-means 聚类算法改进之后模型的总体损失曲线进行对比,实验结果如图 4 所示。

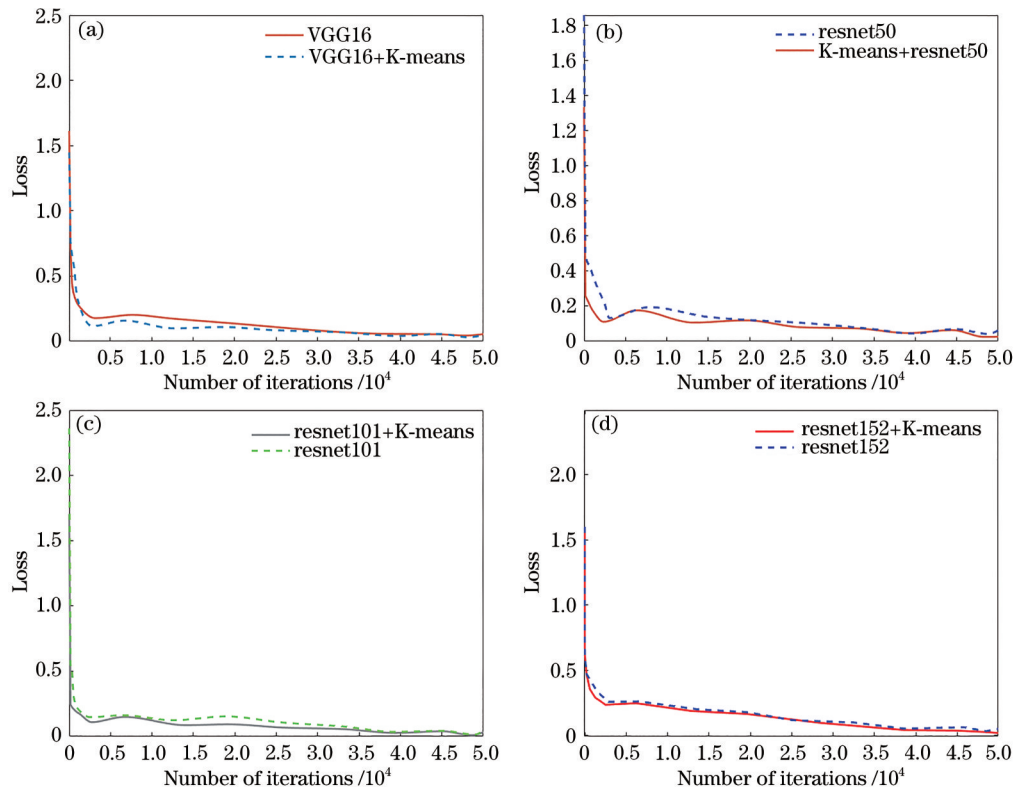


图 4 四种训练损失结果。(a) VGG16 模型; (b) resnet50 模型; (c) resnet101 模型; (d) resnet152 模型

Fig. 4 Results of four training losses. (a) VGG16 model; (b) resnet50 model; (c) resnet101 model; (d) resnet152 model

从图 4 中可以看出,相对于原始的 Faster R-CNN 接触网管帽目标定位深度学习模型,使用 K-means 算法改进的模型的收敛性有明显改善,其中基于 VGG16 和 resnet50 的模型在迭代次数达到约 25000 次时收敛速度明显提升,迭代次数 30000 次之后原始以及 K-means 算法改进之后的两条收敛曲线基本重合,其次基于 resnet101 和 resnet152 的模型在迭代次数约为 25000 次时,收敛速度也有小幅加快,其中基于 resnet101 的模型在迭代次数为 35000 次之后两条收敛曲线基本重合,基于 resnet152 的模型在迭代次数为 50000 次之后两条收敛曲线基本重合。为了更直观地比较 K-means 算法改进之后不同提取网络的收敛速度,将改进之后基于不同特征提取网络的 Faster R-CNN 接触网管帽目标定位模型训练损失曲线放入同一坐标系下,如图 5 所示。从图 5 中可以看出, resnet50+K-means 模型能快速收敛,迭代次数 12000 次之后基本趋于稳定,明显优于其他 3 种网络模型。VGG16+K-means 模型与 resnet101+K-means 模型收敛情况基本相同, resnet152+K-means 模型收敛速度最慢,30000 次之后才基本趋于稳定。

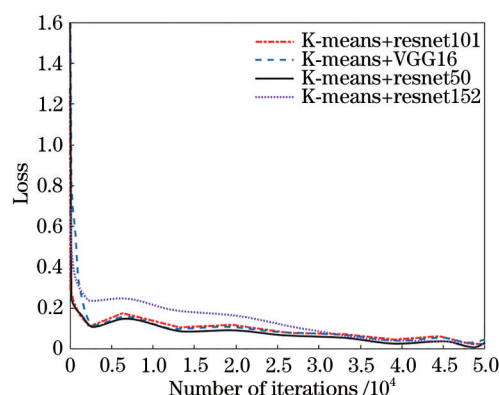


图 5 不同特征提取网络模型训练损失

Fig. 5 Training loss of different feature extraction network models

上述对比结果表明,基于 K-means+resnet50 的 Faster R-CNN 接触网管帽目标定位深度学习模型在训练过程中能够迅速完成收敛,从而可以大大降低模型的训练时间。为选出最优接触网管帽目标定位模型,将 4 种不同特征提取网络的 Faster R-CNN 原始模型及改进后模型在相同测试集上的准确率和召回率与迭代次数的关系绘制成曲线,如图 6 和图 7 所示。从图 6(a)中可以看出,在 VGG16

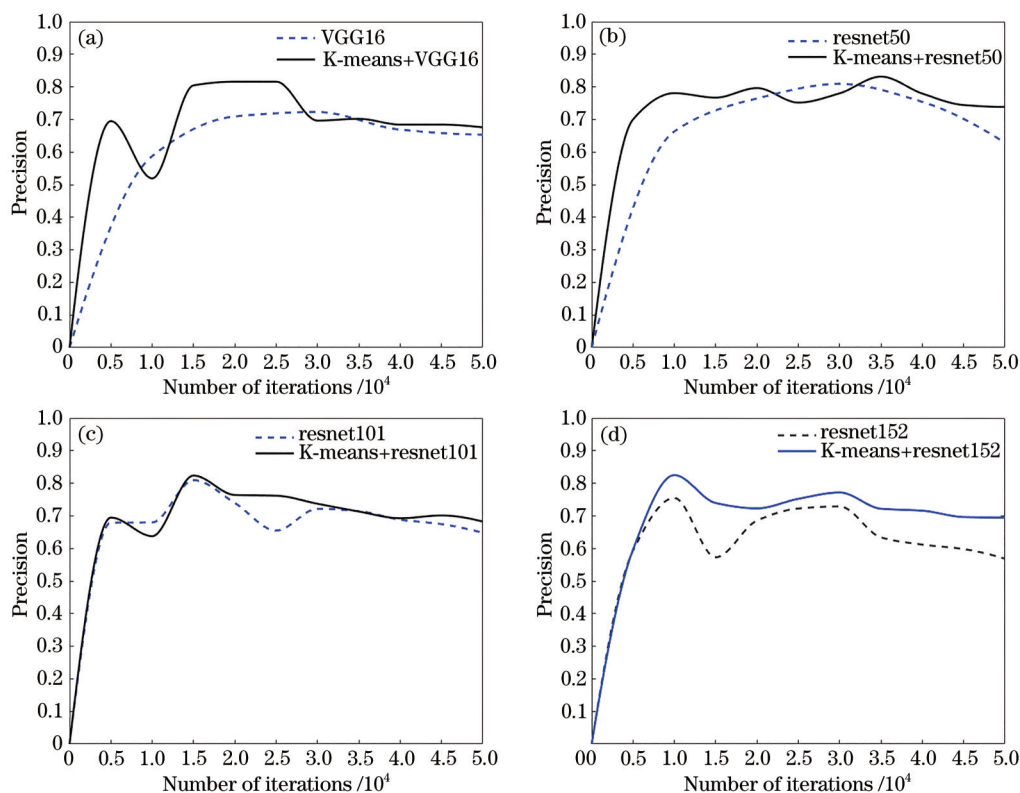


图 6 模型准确率与迭代次数的关系。(a) VGG16 模型;(b) resnet50 模型;(c) resnet101 模型;(d) resnet152 模型

Fig. 6 Relationship between model precision and number of iterations. (a) VGG16 model; (b) resnet50 model; (c) resnet101 model; (d) resnet152 model



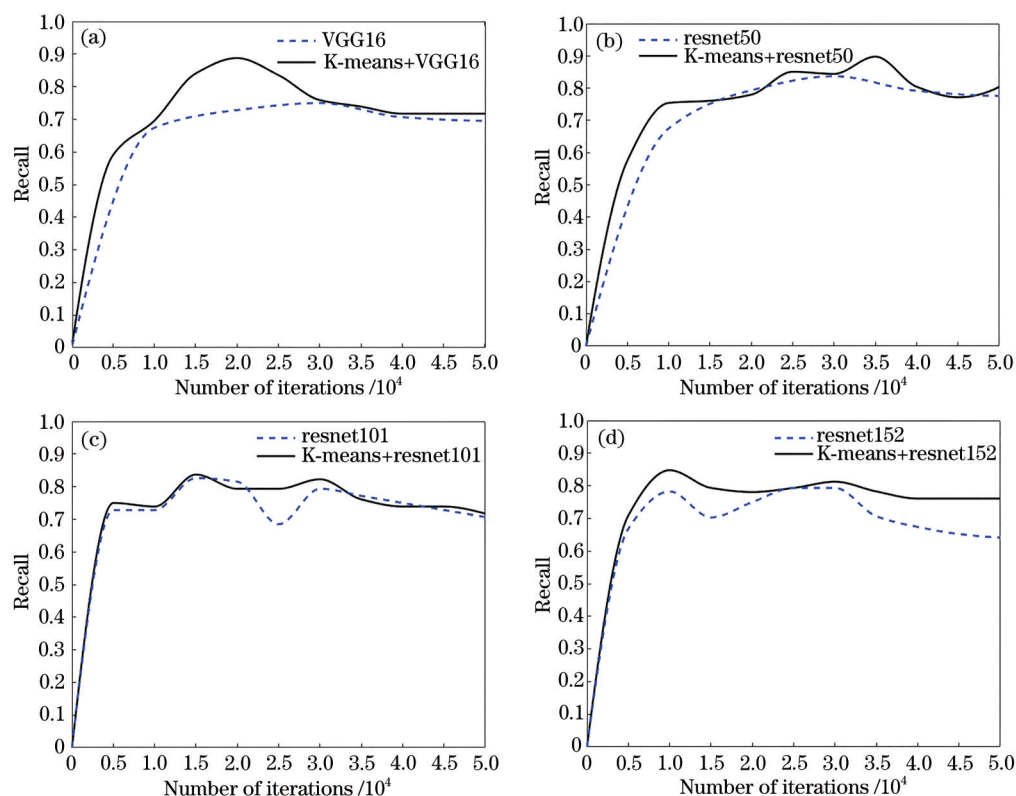


图7 模型召回率与迭代次数的关系。(a) VGG16模型;(b) resnet50模型;(c) resnet101模型;(d) resnet152模型

Fig. 7 Relationship between model recall and number of iterations. (a) VGG16 model; (b) resnet50 model; (c) resnet101 model; (d) resnet152 model

模型中,原始模型准确率随着训练迭代次数的增加呈现出不断上升的情况,当迭代次数达到20000之后基本达到平稳状态,最高准确率为72.29%,而改进后的模型准确率随着迭代次数的增长处于一个波动状态,准确率在迭代次数为25000次之后反而下降,主要原因是迭代次数过多导致模型过拟合,准确率反而下降。VGG16模型准确率在迭代次数为7000次以内及13000次至28000次时,准确率明显优于原始模型,当迭代步数大于28000次时准确率趋于平稳,最高准确率为81.60%。从图6(b)中可以看出,在resnet50模型中,原始模型准确率随着迭代次数的增加呈现出增长的趋势,在迭代次数约为30000次时有最高准确率,最高准确率达到80.92%,而改进之后的模型准确率明显高于原始模型,并且随着迭代次数的增加呈现出一种振荡增长的情况,当迭代次数约为35000时有最高准确率,为83.16%。由图6(c)可以看出,在resnet101模型中,原始模型准确率的变化趋势与改进后的模型基本一致,改进后模型的准确率略优于原始模型且两者都在迭代次数为15000时达到最高准确率,原始模型最高准确率为80.97%,改进后模型最高准确

率为82.37%。从图6(d)中可以看出,对于resnet152模型,改进后模型的准确率明显高于原始模型,两者变化趋势相似,改进后模型的最高准确率为82.51%,原始模型的最高准确率为75.49%。

图7为模型召回率与迭代次数的关系,从图7(a)至(d)能看出,改进后的模型召回率明显优于原始模型,最直观地体现在VGG16、resnet50、resnet152三个网络上, resnet101网络原始模型与改进模型召回率变化趋势基本一致,改进后模型召回率略优于原始模型。其中召回率最高的为resnet50特征提取网络,改进后在测试集上的最高召回率为89.78%。

依据模型评价指标,对4种不同的深度学习网络及改进前后的准确率、召回率、 $F_1$ 值、检测单张图片所需要的时间进行了对比,结果如表2所示。基于VGG16、resnet50、resnet101、resnet152特征提取网络的原始模型及用K-means改进的Faster R-CNN深度学习模型对接触网管帽进行目标定位。通过图8接触网管帽定位识别结果、图6模型准确率与迭代次数的关系、图7模型召回率与迭代次数的关系及表2中数据的对比可得,用K-means算法

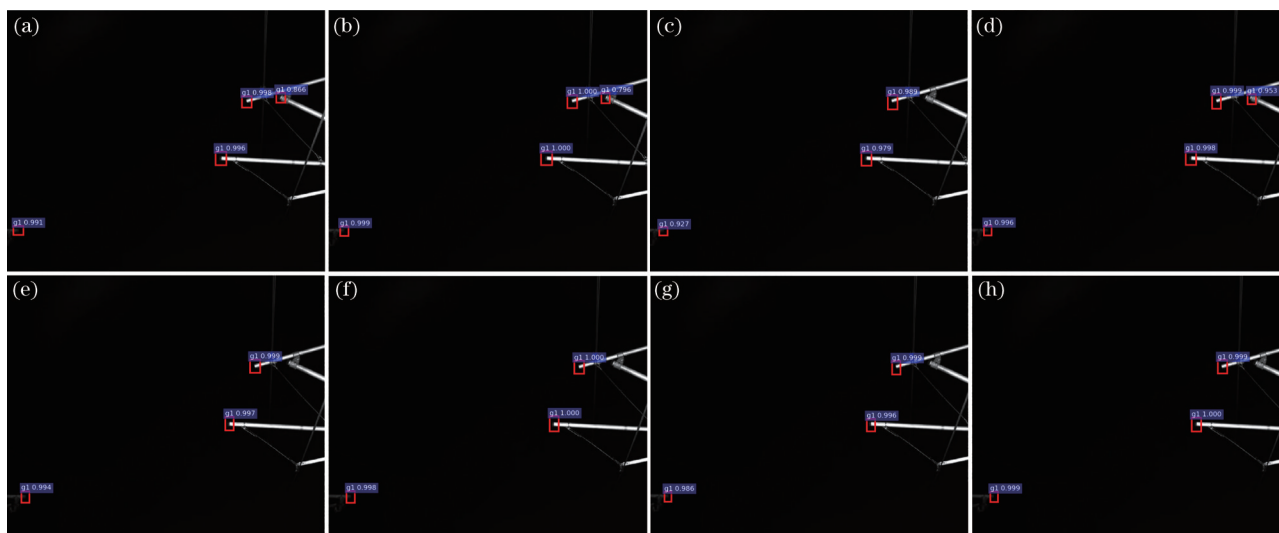


表 2 深度学习网络对比  
Table 2 Comparison of deep learning network

Extraction network	Accuracy / %	Recall / %	$F_1$ / %	Detection time / s
VGG16	72.29	75.00	73.62	0.215
K-means+VGG16	81.60	88.74	85.02	0.216
resnet50	80.92	83.70	82.28	0.283
K-means+resnet50	83.16	89.78	86.34	0.283
resnet101	80.97	82.61	81.78	0.339
K-means+resnet101	82.37	83.70	83.02	0.338
resnet152	75.49	79.35	77.37	0.395
K-means+resnet152	82.51	84.78	83.62	0.395

改变 anchor boxes 对于接触网管帽目标定位的效果有着明显的提升,其中改进后模型评价指标最优的模型为基于 K-means+resnet50 的 Faster R-CNN 深度学习模型,改进后模型的最大准确率为 83.16%,最大召回率为 89.78%,模型的  $F_1$  值为 86.34%,明显优于其他模型。从图 8 中可以看出,改进后的深度学习模型相较于原始模型定位接触网管帽的概率有明显提升( $g_1$  为接触网管帽),其中 resnet50 网

络在改进后可以定位出原始网络及 resnet101 和 resnet152 网络定位不到的部分管帽。虽然 VGG16 网络也可以达到类似的效果,但是综合定位管帽的概率, resnet50 网络具有明显的优势。在检测时间上来看,从 VGG16 至 resnet152 网络,随着特征提取网络层数的不断增加,检测单张的时间也逐渐增加,综合表 2 的数据,其中 resnet50 网络和最短用时的 VGG16 网络相差并不明显。



取的效果,导致模型的定位效果变差。改进后的 Faster R-CNN 算法中, resnet50 网络的效果仍然最优,但 VGG16 网络的改进后效果最明显,主要原因为 VGG16 的网络结构相对简单,在简单物体的特征提取上本就有不俗的表现,原始网络表现不佳是原始 anchor boxes 过大导致提取的干扰特征过多,有效特征较少。当用 K-means 改进 anchor boxes 后, VGG16 提取的有效特征的比例大大增加,从而大大提高了模型的定位效果。改进后 resnet50、resnet101、resnet152 网络定位效果均有不同的提高,但是由于复杂的残差网络结构和相对较深的网络层数,改进之后的效果不如 VGG16 改进前后明显,但 resnet50 由于相对 VGG16 在特征提取上有更优秀的表现,改进后的 resnet50 网络仍为所提算法最优特征提取网络,具有最优的  $F_1$  值。

从表 2 及图 8 中可以看出,将原始 Faster R-CNN 算法中生成的 anchor boxes 用 K-means 改进之后明显提高了管帽定位的准确率,原因是原始 Faster R-CNN 算法中生成的 anchor boxes 是用来检测 pascal voc 数据集中的汽车、人、马等大尺度物体,将原始较大的 anchor boxes 用来检测接触网管帽这种小部件时,较大的 anchor boxes 中除了被检测部件外还包含了太多无用特征和干扰特征。将 anchor boxes 的比例和大小用 K-means 算法改进为适合接触网管帽的大小时, anchor boxes 中提取的无用和干扰特征将大大减少,从而提高了算法的对接触网管帽的定位准确率。

## 4 结 论

基于 Faster R-CNN 深度学习的方法,利用 4C 装置所采集到的含接触网管帽的图片作为样本,以 TensorFlow 为学习框架,实现了接触网管帽的目标定位,有效地改善了接触网管帽定位困难的情况,对后续的接触网管帽故障诊断提供了研究基础,具有较为重要的实际工程意义。通过实验比较了原始及改进后的基于不同特征提取网络的 Faster R-CNN 深度学习模型的收敛曲线及在测试集上的定位准确率、召回率、 $F_1$  值、单张检测时间等指标。实验结果表明,改进后的网络相较于原始网络在接触网管帽目标定位效果上有着显著的提升,其中基于 K-means+resnet50 的 Faster R-CNN 模型为最优的深度学习模型。从实验结果中还可以看出,随着特征提取网络深度的不断增加,模型性能并没有表现

出所预期的结果,主要原因在于过深的卷积层提取的是目标的个性化特征,并不具有普适性,会导致模型的效果变差。因此,在实际生活中,针对不同问题选择与之匹配的特征提取网络是非常重要的。后续工作将研究如何简化模型并进一步提高模型定位准确率。

## 参 考 文 献

- [1] Wang X, Wang Z X, Chai X P. Research on the high-speed railway catenary systems online real-time testing[J]. Journal of Railway Engineering Society, 2014, 31(9): 97-101.  
王翔,汪正兴,柴小鹏. 高速铁路接触网在线实测研究[J]. 铁道工程学报, 2014, 31(9): 97-101.
- [2] Han S. Research on fault detection algorithm of contact network pipe cap and U-shaped holding hoop based on machine vision[D]. Shijiazhuang: Shijiazhuang Tiedao University, 2019.  
韩帅. 基于机器视觉的接触网管帽及 U 型抱箍故障检测算法研究[D]. 石家庄: 石家庄铁道大学, 2019.
- [3] Han Y, Liu Z G, Han Z W, et al. Fracture detection of ear pieces of catenary support devices of high-speed railway based on SIFT feature matching[J]. Journal of the China Railway Society, 2014, 36(2): 31-36.  
韩焯, 刘志刚, 韩志伟, 等. 基于 SIFT 特征匹配的高速铁路接触网支撑装置耳片断裂检测研究[J]. 铁道学报, 2014, 36(2): 31-36.
- [4] Zhang G N, Liu Z G, Han Y, et al. A fast fuzzy matching method of fault detection for rod insulators of high-speed railways[J]. Journal of the China Railway Society, 2013, 35(5): 27-33.  
张桂南, 刘志刚, 韩焯, 等. 接触网棒式绝缘子故障检测的快速模糊匹配方法[J]. 铁道学报, 2013, 35(5): 27-33.
- [5] Zhang C C, Wang S, Zhang D K, et al. Detection of catenary steady clamp based on image processing[J]. Journal of Railway Science and Engineering, 2015, 12(6): 1478-1484.  
张春春, 王嵩, 张冬凯, 等. 基于图像处理的接触网定位线夹目标检测[J]. 铁道科学与工程学报, 2015, 12(6): 1478-1484.
- [6] Yang H M, Liu Z G. Defective condition detection of rotary double ears of catenary support device of electrified railway based on feature matching of SURF[J]. Journal of the China Railway Society, 2016, 38(8): 28-34.  
杨红梅, 刘志刚. 基于 SURF 特征匹配的电气化铁路

- 接触网支撑装置旋转双耳不良状态检测[J]. 铁道学报, 2016, 38(8): 28-34.
- [7] Lü Y, Han Z W, Zhong J P, et al. A GAN-based anomaly detection method for isoelectric line in high-speed railway[C]//2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), May 20-23, 2019, Auckland, New Zealand. New York: IEEE Press, 2019: 1-6.
- [8] Liu W Q, Liu Z G, Núñez A, et al. Unified deep learning architecture for the detection of all catenary support components[J]. IEEE Access, 2020, 8: 17049-17059.
- [9] Liu Z G, Liu K, Zhong J P, et al. A high-precision positioning approach for catenary support components with multiscale difference[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(3): 700-711.
- [10] Albelwi S, Mahmood A. Automated optimal architecture of deep convolutional neural networks for image recognition[C]//2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), December 18-20, 2016, Anaheim, CA, USA. New York: IEEE Press, 2016: 53-60.
- [11] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [12] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [13] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [14] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [15] Hou B J, Yang M H, Sun X W. Real-time object detection for millimeter-wave images based on improved faster regions with convolutional neural networks[J]. Laser & Optoelectronics Progress, 2019, 56(13): 131009.  
侯冰基, 杨明辉, 孙晓玮. 基于改进 Faster RCNN 的毫米波图像实时目标检测[J]. 激光与光电子学进展, 2019, 56(13): 131009.
- [16] Zhang L, Jin X, Fu L Y, et al. Recognition method for weeds in rapeseed field based on Faster R-CNN deep network[J]. Laser & Optoelectronics Progress, 2020, 57(2): 021508.  
张乐, 金秀, 傅雷扬, 等. 基于 Faster R-CNN 深度网络的油菜田间杂草识别方法[J]. 激光与光电子学进展, 2020, 57(2): 021508.
- [17] Xu L X, Chen X J, Ban Y, et al. Method for intelligent detection of parking spaces based on deep learning[J]. Chinese Journal of Lasers, 2019, 46(4): 0404013.  
徐乐先, 陈西江, 班亚, 等. 基于深度学习的车位智能检测方法[J]. 中国激光, 2019, 46(4): 0404013.
- [18] Ju M R, Luo H B, Wang Z B, et al. Improved YOLO V3 algorithm and its application in small target detection [J]. Acta Optica Sinica, 2019, 39(7): 0715004.  
鞠默然, 罗海波, 王仲博, 等. 改进的 YOLO V3 算法及其在小目标检测中的应用[J]. 光学学报, 2019, 39(7): 0715004.