

基于视频图像的弓网接触位置动态监测方法^{*}

王恩鸿 柴晓冬 钟倩文 李立明 张乔木

(上海工程技术大学城市轨道交通学院, 201620, 上海//第一作者, 硕士研究生)

摘 要 针对列车人工检测受电弓状态效率低下的问题, 提出基于视频图像的弓网接触位置动态监测方法。首先, 将运用帧间差抓取受电弓视频中得到的包含弓网的图像作为原始训练数据集; 然后, 对复杂的受电弓图像背景进行分割处理, 采用超像素分割获得最大特征区域, 结合特征图像的 HOG (方向梯度直方图) 获得最大特征 ROI (感兴趣区域), 形成训练数据集, 并设计标签; 最后, 运用改进的 YOLO v3-tiny-strong 网络结构检测分类器, 用训练的权重对视频目标进行监测。结果表明: 该动态监测方法能够在每 1 帧图像中精确标记出受电弓与接触网的接触点位置, 并且能够持续对受电弓的运动状态进行捕捉, 有效获取接触点与受电弓的相对坐标位置, 从而达到对受电弓的监测目的, 其对弓网视频的检测精度可达 98%。

关键词 受电弓监测; 视频目标识别; 特征提取; 超像素分割; YOLO v3-tiny-strong 网络结构

中图分类号 U264.3+4

DOI:10.16037/j.1007-869x.2021.07.044

Dynamic Monitoring Method of Pantograph-Catenary Contact Position Based on Video Image

WANG Enhong, CHAI Xiaodong, ZHONG Qianwen, LI Liming, ZHANG Qiaomu

Abstract Aiming at the problem of low efficiency of train manual detection of pantograph status, a dynamic monitoring method of pantograph-catenary contact position based on video image is proposed. First, the image containing the pantograph-catenary obtained from the pantograph video captured by frame difference is used as the original training data set; then, the complex pantograph image background is segmented, and the super pixel segmentation is used to obtain the largest feature area, combining the HOG (Histogram of Directional Gradients) of the feature image to obtain the largest feature ROI (region of interest), to form the training data set, and to design the label; finally, the improved YOLOv3-tiny-strong network structure is used to detect the classifier and trained weight is used to

monitor video targets. Results show that the dynamic monitoring method can accurately mark the position of the contact point of pantograph and catenary in each frame of image and can continuously capture the movement state of the pantograph, and effectively obtain the contact point and the pantograph relative coordinate position, so as to achieve the purpose of monitoring the pantograph, of which the detection accuracy can reach 98%.

Key words pantograph monitoring; video target recognition; feature extraction; super pixel segmentation; YOLOv3-tiny-strong network structure

Author's address Shanghai University of Engineering Science, School of Urban Railway Transportation, 201620, Shanghai, China

0 引言

对受电弓及接触网的视频监控是保障列车高速、安全运行的关键手段, 其中, 对监测视频的分析与识别是受电弓监测的核心。目前对受电弓的监测主要是人工操作, 效率低、耗时长^[1], 并且对受电弓与接触网接触线位置关系方面的研究鲜有涉及。受电弓与接触网的触点若在安全区域内列车可安全运行, 否则将会导致列车运行故障, 严重时甚至影响列车的运行安全, 所以必须对受电弓与接触网的触点进行有效监测。

目前主要是运用力学传感器采集受电弓运动过程中受力的变化来监测受电弓的状态, 并采用图像处理技术对受电弓的磨损进行监测。文献[2]运用图像侧连法对受电弓滑板的磨损部位进行检测, 运用传感器测量法检测受电弓与接触网的接触受力部位, 采集受力信息; 文献[3]通过监测列车高速运行时受电弓与触网之间的受力情况, 对受电弓进行动态监测; 文献[4]采用边缘检测、闭运算、边缘

^{*}上海市地方院校能力建设项目(18030501300); 上海工程技术大学科研启动基金项目(0240-E3-0507-19-05081)

强度和模板匹配等方法对受电弓的图像特征进行提取,并对受电弓烧坏和裂纹等故障通过像素分析进行判别;文献[5]设计了基于机器视觉系统的高速列车弓网动态性能参数实时检测系统,对受电弓的拉出值进行提取,对燃弧状态进行检测,并对检测数据进行存储;文献[6]对受电弓滑板磨损和中心线偏移检测进行研究,运用先进边界区分噪声检测算法、改进均值滤波算法以及基于自适应的 Canny 滑板磨损边缘检测算法(以下简称“Canny 算法”)实现对受电弓的检测。

但是,在列车运行过程中,以上研究并不能通过车载摄像头采集的受电弓视频数据直接获得弓网接触相对位置,因此本文提出一种基于视频图像的弓网接触位置动态监测方法。首先,对列车摄像头采集到的视频数据进行受电弓特征帧图像提取,对视频图像设定初始的受电弓帧图像,并通过帧间差运算分离出视频中含有受电弓的帧,将此保存为图片数据;其次,为了去除图像中的复杂背景,获得主要目标图像信息并减少运算量,采用基于熵率的超像素分割技术对图片进行处理;再次,通过找到图像的最大特征 ROI(感兴趣区域),提取形成特征数据集;最后,采用优化后的 YOLO v3-tiny 网络结构对数据集进行训练,获得输出的权重。通过对权重的对比即可对受电弓与接触网的接触位置进行跟踪和识别,达到监测的目的。

1 受电弓视频数据预处理

图 1 为基于视频图像的弓网位置动态实时监测算法流程图。首先,通过帧间差提取视频中单帧包含受电弓的图像;其次,通过熵率的超像素分割算法获得最大导熵率区域,在受电弓图像 HOG(方向梯度直方图)中分割出最大特征的最小 ROI 区域,形成特征数据集;最后,利用优化的 YOLO v3-tiny-strong 网络模型训练出权重,对视频中的弓网接触位置进行识别。

1.1 视频特征帧提取

为了获得受电弓的特征帧,需要对受电弓的视频图像做特征帧提取。特征帧是包含所需目标图像最大特征的帧图像,在视频图像空间域中提取视频的每 1 帧,可得到一幅静态的图像。采用等量帧间差运算来分离列车高速运行时采集到的受电弓视频图像,可分离出受电弓视频中包含受电弓特征的特征帧,并且每张图像中都包含目标图像的最大

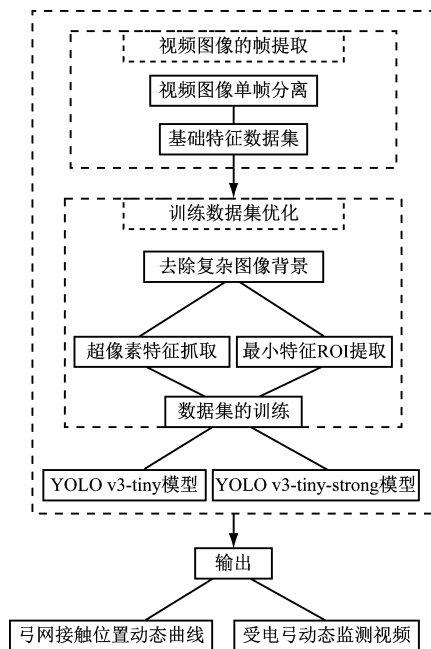


图 1 基于视频图像的弓网位置动态实时监测算法流程

特征。通过空间分量的帧提取方法,对列车在高速运行情况下采集到的受电弓视频进行处理,获得受电弓的图像数据。

1.2 受电弓图像特征提取

1.2.1 最大熵率的超像素特征提取

超像素是按照一定规律将图像中具有相似特征的像素点聚合成大小不一的像素块,使这些像素块表达一种视觉意义。基于熵率的超像素分割通过将聚合成的像素块进行处理,达到降低复杂度、提高计算速度的目的。

设离散的随机变量为 M ,熵率构造的目标函数 $H(M)$ 用于度量随机变量 M 的不确定性^[7],定义 φ 为 M 的支撑集,则有:

$$H(X) = - \sum_{X \in \varphi} P_X(X) \lg P_X(X) \quad (1)$$

式中:

X ——像素点;

$P_X(X)$ ——像素点在图像中出现的概率。

超像素熵率是通过图像的超像素建立拓扑结构,本质是寻找图像的拓扑结构,将图像拓扑定义为 A ,将所有的超像素块相连,其最大化目标特征函数 E 的计算公式为:

$$E = \max_A \{ H(A) + \lambda B(A) \} \quad (2)$$

式中:

$H(A)$ ——熵率项,用以形成相对规则的超像素;

λ ——平衡因子;

$B(A)$ ——平衡项,用以约束超像素的尺寸。

$H(A)$ 、 $B(A)$ 的构造式分别为:

$$H(A) = - \sum_i \mu_i \sum_j P_{i,j}(A) \lg[P_{i,j}(A)] \quad (3)$$

$$P_{i,j}(A) = \begin{cases} \frac{w_{i,j}}{w_i}, & i > j \text{ 且 } e_{i,j} \in A \\ 0, & i < j \text{ 且 } e_{i,j} \in A \\ 1 - \sum \frac{jw_{i,j}}{w_i}, & i = j \text{ 且 } e_{i,j} \in A \end{cases} \quad (4)$$

$$B(A) = - \sum_i P_{Z_A}(i) \lg[P_{Z_A}(i)] - N_A \quad (5)$$

$$P_{Z_A}(i) = \frac{|S_i|}{|v|}, \quad i = \{1, \dots, N_A\} \quad (6)$$

式中:

i, j ——自然数序列;

μ_i ——适应因子;

$P_{i,j}(A)$ ——像素点概率;

$w_{i,j}$ ——二维像素;

w_i ——一维像素;

$e_{i,j}$ ——图像子熵;

Z_A ——子图分布;

$P_{Z_A}(i)$ ——子图的分布律;

N_A —— G 中连通的子图个数(假设原图为 G);

S_i ——子图熵;

v ——总图熵。

为了获得最优的目标函数,利用贪婪算法进行处理,所得的分割结果 λ 为平衡因子参数。 λ 越大,分割超像素形状越不规则,但形状更均匀; λ 越小,超像素形状越规则,但大小不均匀。假设图像超像素的初始值为 ξ , K 为超像素数量, β 为动态因素,则:

$$\lambda = K\beta\xi \quad (7)$$

$$\beta = \frac{\max e_{i,j} H(e_{i,j}) - H(\phi)}{\max e_{i,j} B(e_{i,j}) - B(\phi)} \quad (8)$$

式中:

ϕ ——无效子图;

$H(e_{i,j})$ ——原图熵率项;

$H(\phi)$ ——支撑集熵率项;

$B(e_{i,j})$ ——原图平衡项;

$B(\phi)$ ——支撑集平衡项。

图2为不同初始值 ξ 所对应的分割结果示意

图。可以看出, ξ 越大,熵率超像素特征效果越好。

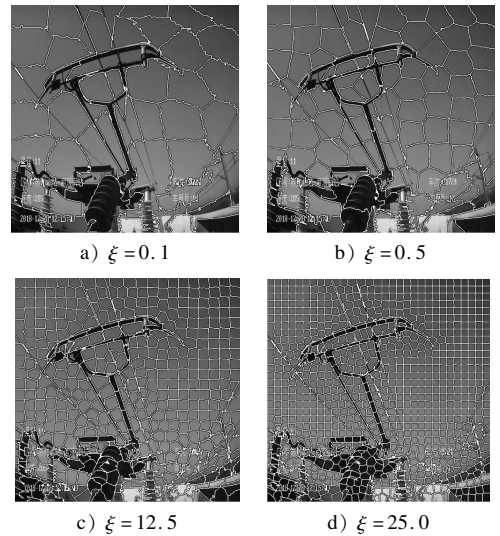


图2 不同初始值 ξ 的熵率超像素特征图

1.2.2 受电弓图像 HOG 特征提取

HOG 通过局部梯度或者目标边缘方向密度分布来描述图像局部区域的形状和纹理特征^[8]。首先将图像分成许多小的连通区域(又称为细胞单元),然后计算每个细胞单元的像素梯度或者边缘方向直方图,最后将所有细胞单元的 HOG 串联,用以描述图像的 HOG 特征,从而实现对 HOG 特征的提取。

1.2.2.1 图像归一化处理

将图像转化为 $I(m, n)$ 二维图,使用 Gamma 校正法对二维图像进行标准化处理,以减少图像中的不平衡因素及图像噪声的干扰,可获得 $U(m, n)$:

$$U(m, n) = I(m, n)^{\text{Gamma}} \quad (9)$$

1.2.2.2 计算图像梯度

分别计算二维图像的横向梯度和纵向梯度:

$$G_y(x, y) = I(x, y + 1) - I(x, y - 1) \quad (10)$$

$$G_x(x, y) = I(x + 1, y) - I(x - 1, y) \quad (11)$$

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \quad (12)$$

$$\theta = \arctan\left[\frac{G_y(x, y)}{G_x(x, y)}\right] \quad (13)$$

式中:

$G_y(x, y)$ ——图像纵向梯度;

$G_x(x, y)$ ——图像横向梯度;

$G(x, y)$ ——图像梯度;

θ ——方向角。

1.2.2.3 建立细胞单元

为每个细胞单元建立直方图,将输入图像分为

64 × 32 像素大小的图像单元,再次细分解 16 × 16 像素步长为 8 的单元,作为输入。

图像的梯度方向范围是 0° ~ 360°。计算每个细胞单元中的直方图后,将所有的直方图串联形成四堡单元梯度直方图,形成 HOG 的一维数组。

1.2.3 最大熵率与 HOG 的最小 ROI 特征定位

依据受电弓帧在复杂背景下的超像素熵率特征,通过对比受电弓目标区域的熵率最大特征,可提取受电弓目标图像的熵率最显著区域。将包含目标特征的图像进行熵率超像素分割,可获得图像最大熵率特征,并将图像的最大熵率特征与 HOG 特征进行结合,获得目标区域中包含这 2 个特征的最小范围,进而从图像中定位出最大特征的最小 ROI 范围,以此去除复杂的背景。

1.3 最大特征 ROI 数据集

经过超像素复杂背景分割,可得到目标区域最大特征的最小 ROI。在 ROI 中除去原始视频图像中的复杂背景,可减小其他非目标物体对监测造成的干扰,且可以保留监测目标物体的最大特征。将分割后的连续帧保存为去除复杂背景后具有最大特征的最小 ROI 的视频数据,然后对 ROI 进行 Canny 算法的边界提取,产生具有边界效应最大熵 ROI-Canny 的视频文件。将 ROI-Canny 边界处理技术作为测试训练的方法,对监测结果进行测试。把从原始视频数据提取得到的最大特征 ROI 视频和经过 Canny 算法处理后的 ROI-Canny 图像数据,建立最大特征数据集。图 3 为建立受电弓复杂背景分割最大特征数据集的流程图。

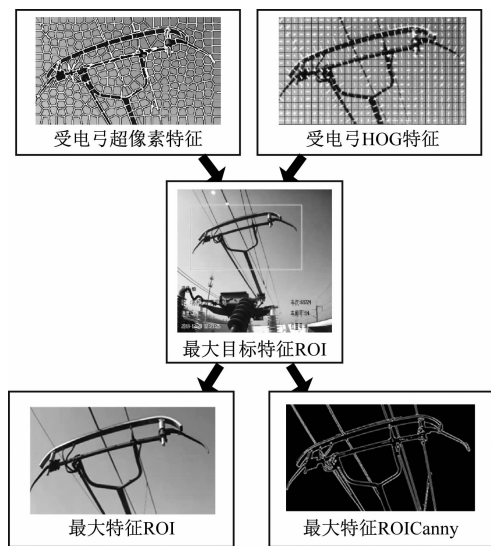


图3 建立受电弓复杂背景分割最大特征数据集的流程

2 视频图像中弓网接触位置的识别

2.1 YOLO v3-tiny 网络结构

与 YOLO v3 网络结构一样, YOLO v3-tiny 也采用了多尺度的检测机制。稍有不同的是 YOLO v3-tiny 只使用了 2 个尺度: 13 × 13 像素和 26 × 26 像素。此外, YOLO v3-tiny 的网络结构与其他网络结构类似, 主要由卷积层和池化层组成, 整个网络结构较为简单、整洁。YOLO v3-tiny 结构的主要特点是网络整体参数少、计算量小, 对计算机的性能要求低, 能够快速完成对输入信息的推理计算。其缺点也较为明显, 为了提升计算速度, YOLO v3-tiny 牺牲了检测精度。由于层结构简单, 该结构对于特征的提取并不充分。针对 YOLO v3-tiny 的实际情况, 结合本文的实际需求, 应在其检测速度降幅不大的情况下, 提升其检测精度。通过运用 YOLO v3-tiny 结构对 ROI-canny 进行训练, 可通过 canny 算法处理后的视频中识别跟踪得到受电弓与接触网的接触位置。

2.2 YOLO v3-tiny-strong 结构

为了更加适用于受电弓图的频识别, 本文在 YOLO v3-tiny 结构的基础上, 进一步对结构功能进行了优化。

首先, 对网络结构进行扩充。对于深度学习网络, 其浅层网络提取的特征包含了如物体的刚性结构、纹理特征等较为明显的特征^[9-12]。同时, 使用大的特征图进行卷积操作, 也能够提取到更为丰富的细节特征。在原 YOLO v3-tiny 网络中, 第 2 层就对特征图进行采样操作, 减少网络的计算量。针对这一情况, 在网络的第 1 层和第 2 层中间加入了 1 × 1 卷积层和 3 × 3 卷积层 2 层网络, 以加强对于浅层的特征提取功能。添加 1 × 1 卷积层的作用主要有 2 个: 一是起到特征提取的平滑过渡, 将第 1 层提取到的特征更加平滑地过渡到新添加的 3 × 3 卷积层中, 使特征在传递的过程中不会丢失; 二是减少网络的参数量。如果仅加入 1 层 3 × 3 网络, 其参数量过多, 会造成卷积层的计算量过大, 降低整体检测速度。在 3 × 3 卷积层前加入 1 × 1 卷积层, 可先减少整体网络参数, 然后再进行特征提取操作, 从而保证具有较高的检测速度。

然后, 在 26 × 26 像素尺度的检测前加入更多的卷积层, 进行特征提取。在原 YOLO v3-tiny 网络中, 26 × 26 像素尺度的检测只由第 19 层的 3 × 3 卷

积层和第 8 层的 3×3 卷积层进行路径合并后送入检测。虽然通过 route 合并功能将浅层特征和深层特征进行合并,但是由于深层特征只由单一的 1 层 3×3 卷积层进行特征提取,提取到的抽象特征较少。本文参考了 YOLO v3 多尺度检测的层结构设置功能,在深层网络原有 3×3 网络的基础上再加入 2 个 3×3 卷积层,用以增强对于深层特征的提取。

优化后的网络模型称为 YOLO v3-tiny-strong,虽然其检测速度与 YOLO v3-tiny 结构相比有所降低,但是通过增强网络对特征的提取能力,YOLO v3-tiny-strong 的检测精度得到了提升。将 YOLO v3-tiny-strong 与 YOLO v3-tiny、YOLO v3、SSD 网络模型和 Fast R-CNN 网络模型进行横向对比,通过平均精度这一参数对比各模型的输出精度,得到的结果如表 1 所示。

表 1 各网络模型的平均输出精度对比表	
网络模型	平均输出精度/%
YOLO v3-tiny-strong	98.0
YOLO v3-tiny	54.5
YOLO v3	86.5
SSD	81.7
Fast R-CNN	90.3

由表 1 可知, Fast R-CNN 网络由于其固有的二阶网络结构,其检测精度相比于 YOLO v3 网络、SSD 网络这类一阶的算法更好,特别是其候选窗推荐机制,增强对目标的检测能力;YOLO v3 网络由于采用了新的 darknet-53 网络结构,加深了网络层数,增强了整体的特征提取能力,并且采用了多尺度检测机制,提高了对小目标的检测精度。同时, YOLO v3 网络的下采样操作采用了步长为 2 的卷积层,能够防止下采样过程中的特征细节丢失。以上这些优化技术,使 YOLO v3 网络的综合检测精度有了较大的提升,其输出的平均精度与 Fast R-CNN 网络相接近。而通过对比发现,与 YOLO v3-tiny 网络相比, YOLO v3-tiny-strong 的检测精度有了大幅度的提升,增幅约为 15%。

在带有 NVIDIAGTX980GPU 的情况下对 5 个算法的检测速度进行了测试,可通过每秒传输帧数来对比模型处理速度,测试对比结果如表 2 所示。

通过表 1~2 的对比可知,修改后的 YOLO v3-tiny-strong 模型虽然检测速度有所下降,但检测精度有了大幅度提升,满足实时检测的条件值。而其

他网络模型中, Fast R-CNN 的网络结构较为复杂,检测速度最慢;YOLO v3 在 YOLO v1 的基础上经历了两代的更新,其网络结构相比于之前也更为简洁,因而检测速度比类似的同时为端到端网络的 SSD 模型更快。通过上述对比, YOLO v3-tiny-strong 模型的整体性能要优于 YOLO v3-tiny,其对目标特征提取的能力也更强,更适用于本次设计的运用场景。

表 2 网络模型每秒传输帧数对比表	
网络模型	每秒传输帧数/(帧/s)
YOLO v3-tiny-strong	47
YOLO v3-tiny	40
YOLO v3	26
SSD	19
Fast R-CNN	11

3 试验结果与分析

3.1 试验环境

为了对算法进行验证,本文搭建了试验环境,在 Linux 系统下对试验环境进行配置,试验中采用 Python3.6 语言、darknet 框架,相关硬件环境如表 3 所示,试验所使用的软件环境如表 4 所示。

表 3 弓网接触位置监测实现的硬件环境			
CPU 环境	GPU 环境	主板型号	内存大小
I510400	GTX980	GJB365	16G

注:CPU——中央处理器;GPU——图形处理器。

表 4 弓网接触位置监测实现的软件环境				
CUDA 版本	Python 版本	Opencv 版本	Ubuntu 系统	CUDNN 版本
9.2	3.6	3.1.4	16.04	7.1

3.2 试验结果

利用 YOLO v3-tiny-strong 模型对最大特征 ROI 数据集进行训练,迭代 4 215 次后得到损失函数曲线如图 4 所示。平均损失是用于衡量模型训练精度的指标,随着训练批次的增加,平均损失降低。由图 4 可知, YOLO v3-tiny-strong 模型的损失函数优于 YOLO v3-tiny 模型,所以 YOLO v3-tiny-strong 模型的训练效果优于 YOLO v3-tiny 模型。

通过以上的训练,可获得 YOLO v3-tiny-strong 网络输出权重。运用权重进行实时监测,能够从视频图像中识别出受电弓与接触网的稳定位置关系,并且能够输出接触点在图像中的坐标。

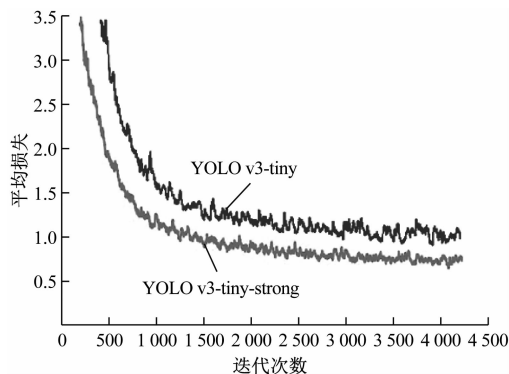


图4 YOLO v3-tiny 模型与 YOLO v3-tiny-strong 模型训练损失函数对比

在列车运行时对动态接触位置进行准确监测,其目标动态监测效果如图5所示。图中的中点位置指受电弓的中点位置,两旁安全域边界内的区域代表安全区域,框出部分为接触位置。



图5 受电弓的视频动态监测效果

图6中为受电弓动态参数偏移结果,图中数据在安全区域上边界和下边界范围内表示受电弓安全运行,超出上下边界部分表示受电弓超出安全位置。

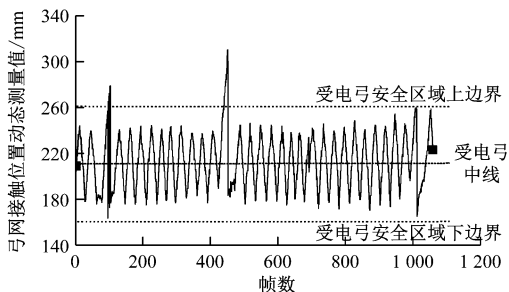


图6 受电弓接触点动态参数监测

4 结语

本文提出了基于视频图像的弓网接触位置动态监测方法,用以解决视频图像中受电弓与接触网位置的目标监测。首先,利用帧间差对受电弓视频数据进行提取,获得基础的数据集;其次,用熵率超像素与图像 HOG 特征去除图像中的复杂背景,提取受电弓图像的最大特征 ROI;最后,用改进后的 YOLO v3-tiny-strong 模型对受电弓最大特征 ROI 特征进行数据集训练,用训练的权重对视频目标进行监测。这种新型的受电弓视频目标监测方法能够在受电弓视频中实时标记出受电弓与接触网的接触位置,达到对受电弓与接触网动态位置的监测目的,可有效监测列车在运行时的弓网状态。

参考文献

- [1] 张健,陈维荣,吴峰,等. 受电弓接触网动态监测系统[J]. 铁道运营技术,2008(1):44.
- [2] 李科,仇广民,王克,等. 受电弓动态检测系统技术原理及应用分析[J]. 铁道机车与动车,2018(3):34.
- [3] 杜高峰,伍川辉. 受电弓动态检测方法研究[J]. 硅谷,2014(16):69.
- [4] 谢力. 基于图像处理的受电弓状态检测技术研究[D]. 成都:西南交通大学,2009.
- [5] 杨恒. 基于机器视觉的高速列车弓网动态性能检测方法研究[D]. 成都:西南交通大学,2017.
- [6] 姚小文. 基于图像处理的城轨列车受电弓检测技术研究[D]. 南京:南京理工大学,2017.
- [7] 韩常. 基于超像素分割的时空上下文模型视频追踪算法研究[D]. 兰州:甘肃政法学院,2017.
- [8] 贾磊. 基于 LBP 和 HOG 特征融合的人脸表情识别算法研究[D]. 太原:中北大学,2019.
- [9] 阮激扬. 基于 YOLO 的目标检测算法设计与实现[D]. 北京:北京邮电大学,2019.
- [10] 熊咏平. 基于 YOLO 的深度学习目标检测算法研究[D]. 武汉:武汉科技大学,2019.
- [11] 王强. 智能视频监控中的行人检测系统设计与实现[D]. 杭州:浙江理工大学,2019.
- [12] 窦笑. 基于深度学习的机器人环境感知研究[D]. 哈尔滨:哈尔滨工程大学,2019.

(收稿日期:2019-07-02)