

基于小样本学习的高铁接触网鸟窝检测方法研究

王 犇¹, 田 野¹, 李 晨¹, 王俊平², 皮 魏¹, 孙木兰¹, 袁希文¹

¹株洲中车时代电气股份有限公司智能工控事业部, 湖南 株洲

²株洲中车时代电气股份有限公司数据与智能技术中心, 湖南 长沙

收稿日期: 2025年10月22日; 录用日期: 2026年5月21日; 发布日期: 2026年5月29日

摘 要

接触网作为电气化铁路供电系统的重要组成部分, 承担着电力传输的重任, 然而鸟窝的入侵可能造成接触网接地跳闸或绝缘击穿等故障, 对行车安全造成危害。由于人工对鸟窝巡检的效率偏低, 利用动车车载接触网运行状态检测装置(3C)采集的可见光全景图像, 并基于深度学习方法来识别鸟窝已经成为当今技术发展的一种趋势。现有深度学习框架的训练方式大多是基于大量带标注的训练数据, 而鸟窝在接触网中的稀缺性使得很难采集到充足的样本数据, 且鸟窝的目标较小, 容易受背景干扰。因此本文采用级联的方法对鸟窝进行识别, 第一级检测网络基于YOLOv8算法, 实现对鸟窝潜在区域(横梁、支柱和杆塔)的识别, 排除大面积无关背景的干扰, 第二级检测网络基于域适应小样本目标检测算法AcroFOD, 以潜在区域作为输入, 识别并定位鸟窝, 通过域感知增强和定向优化策略, 将源域的鸟窝特征迁移到目标域, 达到96.9%的检出率和2.4%的误报率。此外, 根据鸟窝区域的几何特征, 对算法进行后处理优化, 在保证相同检出率的基础上, 误报率降低至0.1%, 解决了工程化应用瓶颈。本文方法只需6张极少量的接触网鸟窝数据, 就能实现鸟窝的有效检测, 为后续接触网鸟窝的清除提供有力支撑。

关键词

鸟窝, 深度学习, YOLOv8, 领域自适应, 小样本目标检测, AcroFOD

Research on High-Speed Rail Catenary Bird Nest Detection Method Based on Few-Shot Learning

Ben Wang¹, Ye Tian¹, Chen Li¹, Junping Wang², Wei Pi¹, Mulan Sun¹, Xiwen Yuan¹

¹Intelligent Industrial Control Division, Zhuzhou CRRC Times Electric Co., Ltd., Zhuzhou Hunan

²Data and Intelligent Technology Center, Zhuzhou CRRC Times Electric Co., Ltd., Changsha Hunan

Abstract

As a crucial component of the electrified railway power supply system, the catenary undertakes the important task of power transmission. However, the intrusion of bird's nests may cause faults such as catenary ground tripping or insulation breakdown, posing risks to train operation safety. Due to the low efficiency of manual bird's nest inspection, using visible light panoramic images collected by the EMU-mounted catenary operation status detection device (3C) and identifying bird's nests based on deep learning methods has become a trend in current technological development. Most training methods of existing deep learning frameworks rely on a large amount of annotated training data. Nevertheless, the scarcity of bird's nests in the catenary makes it difficult to collect sufficient sample data; additionally, bird's nests are small targets and easily interfered by the background. Therefore, this paper adopts a cascaded approach to identify bird's nests. The first-stage detection network is based on the YOLOv8 algorithm, which realizes the identification of potential bird's nest areas (beams, pillars, and poles) and eliminates interference from large areas of irrelevant backgrounds. The second-stage detection network is based on AcroFOD, a domain-adaptive few-shot object detection algorithm. It takes potential areas as input to identify and locate bird's nests. Through domain-aware enhancement and directional optimization strategies, the bird's nest features from the source domain are transferred to the target domain, achieving a detection rate of 96.9% and a false alarm rate of 2.4%. Furthermore, according to the geometric features of bird's nest areas, post-processing optimization is performed on the algorithm. On the basis of ensuring the same detection rate, the false alarm rate is reduced to 0.1%, solving the bottleneck in engineering applications. The method proposed in this paper only requires a very small number of catenary bird's nest data (6 samples) to achieve effective detection of bird's nests, providing strong support for the subsequent removal of catenary bird's nests.

Keywords

Nest, Deep Learning, YOLOv8, Domain Adaptation, Few-Shot Object Detection, AcroFOD

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

高速铁路在我国经济社会发展中的占比越来越高，保证高铁的安全运行对于保障乘客生命财产安全有着重要意义[1]。接触网构建在高铁的运行线路上空，是一种向高速行驶的列车提供电能的输电线路。鸟窝时常出现在接触网的横梁、杆塔或支柱上，可能造成输电线短路和绝缘子击穿的风险，导致列车晚点和停运，故有必要对接触网的鸟窝入侵事件进行严格监控，来保证高速铁路的安全运行。

鉴于人工巡检方法的诸多弊端，越来越多的学者开始将机器视觉和传统图像处理技术运用到鸟窝的检测识别中。段旺旺[2]等人针对人工检测的低效性，提出一种基于关键区域 HOG 特征的接触网鸟巢检测方法，该方法先对疑似鸟窝的区域进行粗提取，再通过 SVM 分类器进一步甄选出鸟窝区域，获得了较好的识别率，但鲁棒性较差。徐晶[3]等人利用鸟窝的颜色和纹理特征来进行识别，该方法能在一定程度上排除背景的干扰，但存在较大局限性。

上述方法通常都是基于具有先验知识的特征工程，手工特征设计的好坏起着决定性作用，算法复杂度高且特征计算量大。相比之下，深度学习具有更好的鲁棒性，只要有充足的数据就能实现高精度检测。杨沛[4]用 Faster RCNN 网络提取支柱区域，再用双判别器生成对抗网络 DDGANs 来识别鸟窝，提高了模型性能，但值得注意的是，该方法默认鸟窝只存在于支柱区域，忽略了其他位置，模型难以泛化。王纪武[5]等人对 Faster RCNN 算法进行改进，用改善的 VGG16 提取特征，然后参考 RPN 网络获取目标初始建议区域，最后将特征映射层传入子网络中来检测鸟窝。

在大量数据的驱动下，传统的深度学习算法能够获得较高的检测精度，但巡检视频中含有鸟窝的图像十分稀缺，采集难度大，无法在模型训练过程中投喂大量数据来确保测试阶段中的高精度识别，导致模型性能受到极大限制。为缓解深度学习模型对大规模接触网鸟窝数据的依赖，实现低成本的高效训练，本文提出一种基于目标检测 - 域适应小样本学习 - 后处理优化的级联方法，实现高铁接触网鸟窝检测，先用 YOLOv8 深度学习算法定位鸟窝的潜在区域，再将域适应小样本目标检测算法 AcroFOD [6]用于鸟窝的检测识别中，该方法将输电线路领域的鸟窝图像作为源域，接触网鸟窝图像为目标域，经训练实现领域间的知识迁移。在此基础上，利用鸟窝区域的几何等特征进一步优化算法。实验表明，该方法在低误报率 0.1%的前提下实现了接触网鸟窝 96.9%的高精度检测。

2. 框架

本文提出的接触网鸟窝检测系统整体框架流程如图 1 所示，输入高铁 3C 系统可见光全景图像后，首先通过鸟窝感兴趣区域网络进行检测，算法采用目前比较前沿的 YOLOv8 目标检测算法，用于识别鸟窝在接触网场景中的潜在区域，包括横梁、杆塔和支柱，输出结果后进行判断，如果不存在兴趣区域，则说明该图像中没有鸟窝，结束检测，如存在潜在区域，则保存该区域，并将其输入 AcroFOD 网络中进行识别，该网络是一种域适应小样本目标检测方法，用于识别并定位潜在区域中的鸟窝，输出结果后再次进行判断，如不存在目标框，则证明潜在区域中没有鸟窝，结束检测，如存在，则再进行后续处理，输出目标框的坐标并根据优化后的算法更好的判断检出和误报情况。

3. 方法

3.1. 目标检测算法

目标检测是深度卷积神经网络在计算机视觉领域的重要应用之一。目前，深度学习的目标检测算法可以分为两类：基于区域候选网络的二阶段目标检测算法和基于回归的一阶段目标检测算法。二阶段检测算法首先使用特定的算法生成一系列样本候选框，并通过卷积神经网络对样本数据进行分类，最后对样本边界框的位置进行调修，这种方法虽然能有效提高识别精度，但是使用候选区网络影响了推理时间。

基于回归的一阶段目标检测算法，不需要产生候选框，而是直接将目标框定位问题转换为回归问题，通过对输入图像进行特征提取和预测，直接输出所有目标的类别和位置。这种方法提高了推理效率，使得检测算法更快速、更高效。这里采用 YOLOv8 [7]网络进行鸟窝分布区域的检测识别，图 2 展示了 YOLOv8 目标检测网络的结构图，该算法是一阶段检测算法的代表。网络由输入端 Input、主干网络 Backbone、Neck 模块和输出端这四个部分组成。输入层引入了 Mosaic、Mixup 等图像混合增强技术，丰富了数据的多样性，负责将待检图像送入网络；主干网络做出了一系列改进，首先对 Focus 的卷积结构进行改进，将其用 3×3 大小的卷积层替换，减小了计算量。其次，在之前 YOLO 系列的基础上提出 C2f 模块，该模块融合了 YOLOv5 中 C3 模块和 ELAN 模块的思想，利用分流的思想进行改进，遏制了深层结构的过度收敛，并缩减了大量参数，减少了计算量，负责对图像进行特征提取；Neck 部分提出 SPPF 模块代替了原本的空间金字塔池化模块，SPPF 对卷积核尺寸进行同等替换，加快了检测速度，进一步提高

了特征提取能力，负责将输入的特征层进行池化和特征融合；最后输出端输出结果，从图2中可以看出，网络对回归分支和预测分支进行了分离操作，这样可以使模型的收敛速度更快。

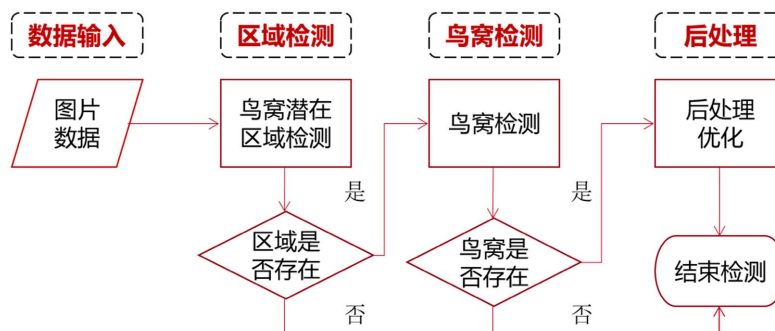


Figure 1. Overall flowchart of high-speed rail catenary bird nest detection
图1. 高铁接触网鸟窝检测整体流程图

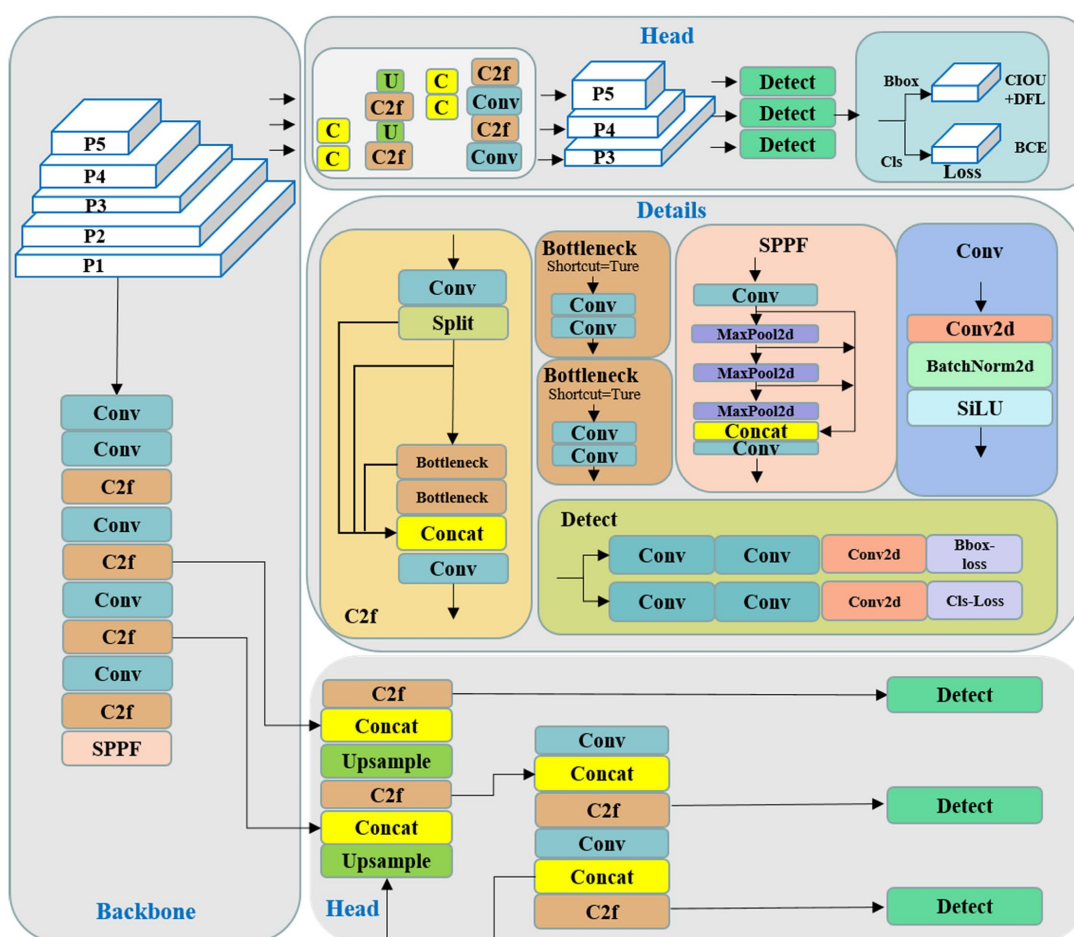


Figure 2. YOLOv8 object detection network structure diagram

图2. YOLOv8 目标检测网络结构图

3.2. 小样本学习算法

基于深度学习的目标检测模型虽然效果显著，但需要大量标注数据作为支撑。基于小样本学习的目

标检测方法旨在研究在样本规模有限的情况下如何提升模型的小样本分类和检测能力，通过少量标注样本就能有效学习到不同物体的特征表示和检测模型，且保持较高的精度。大多数方法都是以传统目标检测算法作为基础框架进行改进，Xiao [8]等人将 Faster R-CNN 作为主干网络进行小样本目标检测研究。Redmon [9]等人以 YOLOv3 作为主干网络，Yang [10]等人以 SSD 作为主干网络。目前相关研究主要分为两大类，分别是主流方法和基于域适应的方法，主流方法可进一步分为基于度量学习、基于数据增强、基于迁移学习和基于元学习的方法，其中以基于元学习的方法研究最为广泛。两者的区别如下图 3 所示，主流方法将数据划分成基类和新类，图 3 中的基类和新类都是自然景物图像数据，基类包括人、飞机、自行车、汽车、马和羊这六类，新类为摩托车、沙发、电视和牛，它们属于同一个数据领域，但是类别标签不一致，通过在丰富的基类样本上训练，来实现知识新类别的适应；而图 3(a)、图 3(b)中将数据分为源域和目标域，两者的类别标签一致，场景不一致，源域是卡通和素描场景下的马和狗，目标域是自然场景下的马和狗，凭借丰富的源域数据和少数目标域数据来完成域间的知识迁移。Wang [11]等人通过引入源和目标配对机制并提出双平模块，第一次成功实现了域适应小样本目标检测。

一般而言，域适应的目标检测方法都是基于无监督学习的，即利用大量标记好的源域和大量未标记的目标域来训练，以特征对齐的方式在源域中对模型进行预训练，然后对齐到目标域，然而现实中常常存在目标域数据不够的情况，这使得模型性能大打折扣，克服这类问题的一种直观方法是直接合并源域和目标域数据，整合后进行数据增强，但这种盲目的增强方很有可能扩大域差异。本文采用 AcroFOD 网络实现接触网鸟窝检测，它是一种跨域小样本目标检测方法，以 YOLOv5 [12]网络为基本框架进行改进。如图 4 所示，网络主要包含三个步骤，首先通过域感知增强来扩展目标域的数据分布，然后将扩展后的数据和目标域数据同时输入到特征提取器，提取后的特征向量再经定向优化策略来过滤掉不适合减小域偏移的样本，最后再将优化后的数据送入检测器进行训练，训练后的模型能够缓解领域之间的差距所导致的性能下降。



Figure 3. Differences between mainstream few-shot object detection methods and domain-adaptive few-shot object detection methods

图 3. 主流小样本目标检测方法方法与域适应小样本目标检测方法区别

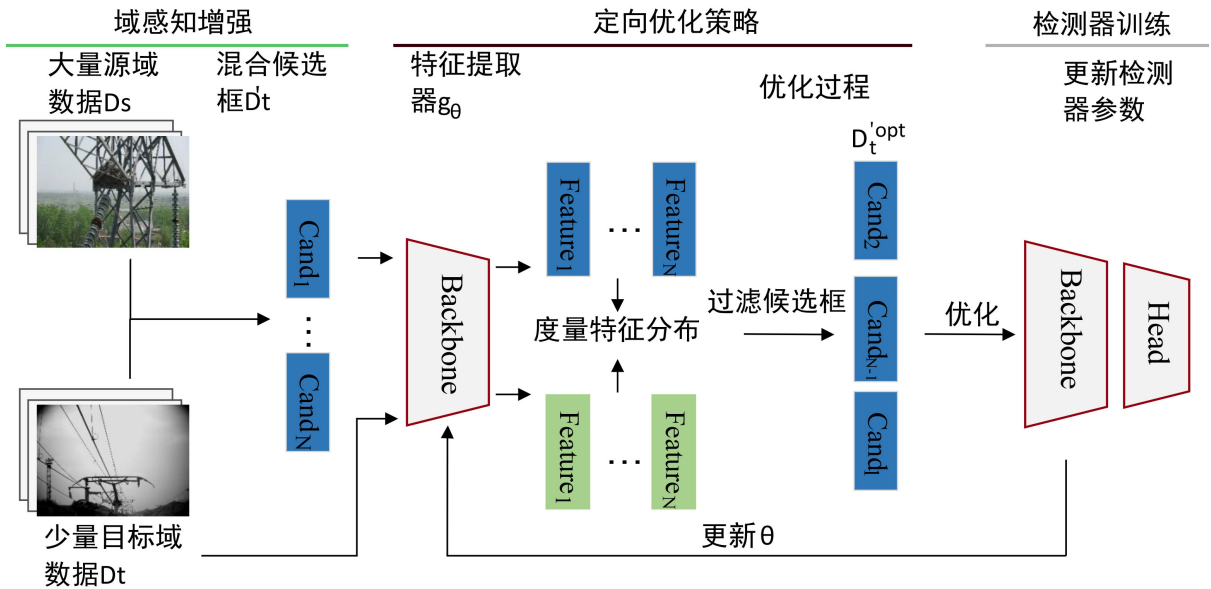


Figure 4. AcroFOD algorithm flowchart
图 4. AcroFOD 算法流程图

3.2.1. 域感知增强

域感知增强的目的是利用目标域数据的前景和背景信息特征，与源域数据进行融合，使源域和目标域图像的融合更加多样化和合理。该算法提出基于 image-level 和基于 box-level 的域感知增强，这两种方式充分利用了目标域数据有限的特征信息，比盲目整合的方式效果更好，具体实现方式如下：

Image-level: 对于一批给定的源域数据 $B_s = \{x_i^s, y_i^s\}$ 和目标域数据 $B_t = \{x_i^t, y_i^t\}$ ，分别从 B_s 和 B_t 中选取 m 和 n 组数据并随机混合到一个单一的图像 x^{aug} 中：

$$x^{aug} = x_0^{aug} + \sum_{i=1}^m \sum_{j=1}^n A_{(i,j)} (\gamma_i x_i^s + \gamma_j x_j^t) \quad (1)$$

其中 x_0^{aug} 代表一个初始化的空白图像， $A_{(i,j)}$ 代表一个为 $\{x_i^s, y_i^s\}$ 配对的变换矩阵， γ_i 和 γ_j 分别是 x_i^s 和 x_j^t 对应的权重值，然后重新计算 x^{aug} 对应的分类标签 y_c^{aug} 和回归标签 y_b^{aug} ：

$$y_b^{aug} = Contact(A_{(i,j)}^T y_{ib}^s, A_{(i,j)}^T y_{jb}^t) \quad (2)$$

$$y_c^{aug} = Contact(\gamma_i y_{ic}^s, \gamma_j y_{jc}^t) \quad (3)$$

其中 y_{ib}^s 和 y_{jb}^t 分别代表源域和目标域感兴趣区域的边界框坐标， y_{ic}^s 和 y_{jc}^t 则表示相应的类别置信度， γ_i 和 γ_j 为对置信度的权值。通过公式(1)~(3)，我们可以以域拼接的形式和域重新分配的方式来生成更多多样化的图像，直观图像如图 5 左半部分所示。

Box-level: 为了更有效的利用标注好的目标信息，可以将目标从背景中脱离，然后放入其他区域中，为了提高增强方式的通用性，这里提出一种带有边界框标签的 box-level 域感知增强方法。在调整好宽度和高度之后，通过交换源域和目标域的边界框 b_s 和 b_t 来组合彼此的特征，公式如下：

$$b_{(p,q)}^{aug} = \beta_{(p,q)} b_{(p,q)}^s + (1 - \beta_{(p,q)}) b_{(p,q)}^t \quad (4)$$

其中 (p, q) 表示边界框 b_s 和 b_t 像素的索引值， β 代表索引值对应的权重。当 $\beta_{(p,q)} = 0$ 时，视为图像间的直接交换，当时，增强方式为混合交换，其中遵循 Beta 分布。为了获取每个像素的尺度信息，进一步提出

一种高斯交换，采用高斯映射的方式来增强图像，此时的定义如下：

$$\beta_{(p,q)} = \exp \left[- \left(\frac{(p-u_x)^2}{\sigma_x^2} + \frac{(q-u_y)^2}{\sigma_y^2} \right) \right] \quad (5)$$

$$\sigma_x = \frac{w}{W} \sqrt{\frac{hw}{2\pi}}, \sigma_y = \frac{h}{H} \sqrt{\frac{hw}{2\pi}} \quad (6)$$

其中 H, W 为图像的高度和宽度， σ_x 和 σ_y 为边界框 b_s 和 b_t 中 x 轴和 y 轴的方差， u_x 和 u_y 是 b_s 和 b_t 对应的平均值。三种图像增强的对比方式如图 5 右半部分所示。

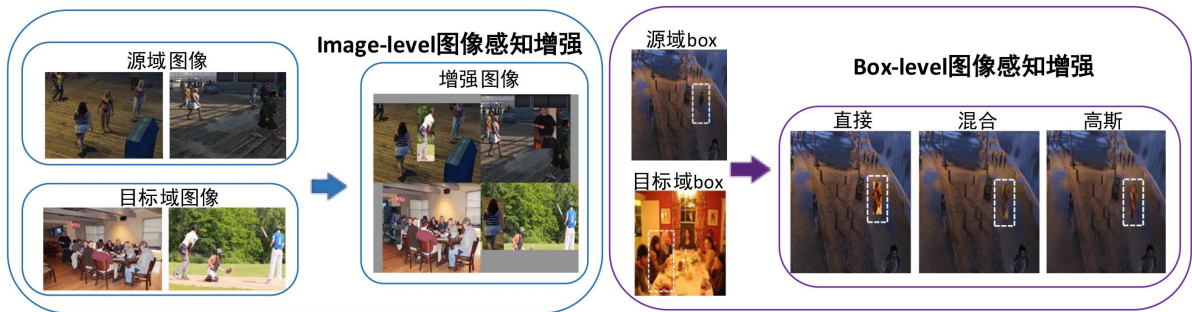


Figure 5. Domain-aware enhanced visualization

图 5. 域感知增强可视化

3.2.2. 定向优化策略

通过域感知增强后，能够生成一系列增强数据 $D'_i = (x_i^{aug}, y_i^{aug})$ ，这些数据和有限的目标域数据 D_t 一起作为特征提取器 g_θ 的输入。假设 D_t 的特征分布为 P_t ， D'_i 的特征分布为 P_{aug} ，我们的目的是希望 P_t 与 P_{aug} 尽可能的接近，这样才能最大限度缓解预间隙导致的模型性能下降，然而 D'_i 中不可避免的会存在一些噪声数据，这些数据与 D_t 的关联性低，会削弱模型的泛化能力，因此算法提出定向优化策略，消除与目标域数据关联性较弱的噪声数据。假设检测器 $f_\theta = (g_\theta, h_\theta)$ 由一组参数 θ 定义且由特征提取器 g_θ 和头部 h_θ 网络组成，特征提取器的增强数据输出 z_i^{aug} 以及目标域数据输出 z_j^t 表示如下：

$$z_i^{aug} = g_\theta(x_i^{aug}), z_j^t = g_\theta(x_j^t) \quad (7)$$

然后根据度量函数 $dist_f$ 对两者的特征表示距离进行排序：

$$d_i^{aug} = dist_f \left(z_i^{aug}, \{z_j^t\}_{j=1}^n \right) \quad (8)$$

为了过滤掉一定数量的噪声样本，选用收缩比 k ($0 < k \leq 1$) 来减少增强的数据量，并定义了一个优化函数 Φ_{opt} 来优化排序，公式如下：

$$D_t^{opt} = \Phi_{opt} \left(D'_i, \{d_i^{aug}\}, k \right) \quad (9)$$

计算后可以选出最优的候选数量并按序列递增，这些数据可以更好地反映目标域的分布 p_t ，经过多次训练迭代后更新 θ ，选择的距离函数 $dist_f$ 用于判断 z_i^{aug} 和 z_j^t 之间的相关性，这里可以选择被广泛使用的最大平均偏差(MMD)和余弦距离(CS)度量函数。定向优化前和定向优化后的适应情况如图 6 所示，最后将优化后的数据投入检测器进行训练，有利于检测器减轻域偏移现象。

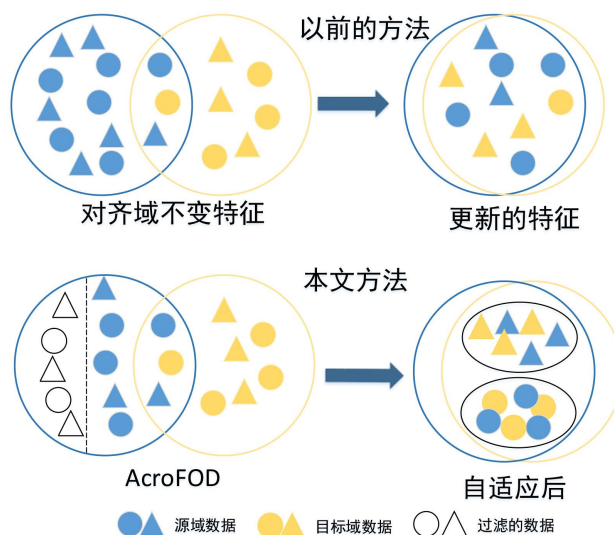


Figure 6. Directional optimization strategy

图 6. 定向优化策略

3.3. 后处理算法

根据巡检视频中已有数据的分析, 其中带有鸟窝的图像可遇不可求, 而带有鸟窝的图像数据中, 目标在整张图像中的占比也比较小, 因此在检测过程中不可避免地会出现误检, 无法投入实际系统的应用。本文结合鸟窝区域的几何特点, 利用 LSD 算法检测直线数量和鸟窝外部长宽特征, 大幅度减低算法误报。其中 LSD 是一种传统直线检测算法, 通过利用检测图像中局部轮廓灰度的剧烈变化, 查找近似矩形区域来获得其中直线的数量, 该方法具有结果准确, 误检可控, 同时不需要调节参数的优点, 算法流程如图 7 所示:

步骤 S1 以高斯下采样的方式对输入的灰度图像区域进行缩放, 消除锯齿效应, 采样率为 0.8。

步骤 S2 计算图像中各点的梯度值和方向, 其中梯度计算公式如下:

$$g_x(x, y) = \frac{i(x+1, y) + i(x+1, y+1) - i(x, y) - i(x, y+1)}{2} \quad (10)$$

$$g_y(x, y) = \frac{i(x, y+1) + i(x+1, y+1) - i(x, y) - i(x+1, y)}{2} \quad (11)$$

$$\text{ang}(x, y) = \arctan\left(\frac{g_x(x, y)}{-g_y(x, y)}\right) \quad (12)$$

其中 $g_x(x, y)$ 和 $g_y(x, y)$ 代表剃度值大小, $\text{ang}(x, y)$ 代表梯度方向, $i(x, y)$ 是灰度图上像素点 (x, y) 处的灰度值。

步骤 S3 将梯度值划分为 1024 个等级进行伪排序, 等级涵盖了梯度从 0~255 的变化范围, 然后将剃度值小于某个种子像素点 p 的点的状态设置为 USED, 其余为 UNUSED。

步骤 S4 取出剩下的排序表中状态为 UNUSED 的点。

步骤 S5 从排序表中选择一个未使用的像素作为种子点用于区域增长, 区域生长算法将用于生成一个 line support region。

步骤 S6 分割出来的直线用一个外接矩形来表示, 计算生长得到的区域同性点密度 D 。步骤 S7 如果 D 小于截断阈值 D_0 , 则截断区域, 转至步骤 S6; 如果 D 大于截断阈值 D_0 , 则计算错误控制函数 NFA。

步骤 S8 如果 NFA 小于设定的错误控制阈值 ϵ ，输出直线段数据，并判断排序表是否还有状态为 UNUSED 的点，如有，则转至步骤 S4，否则结束算法。

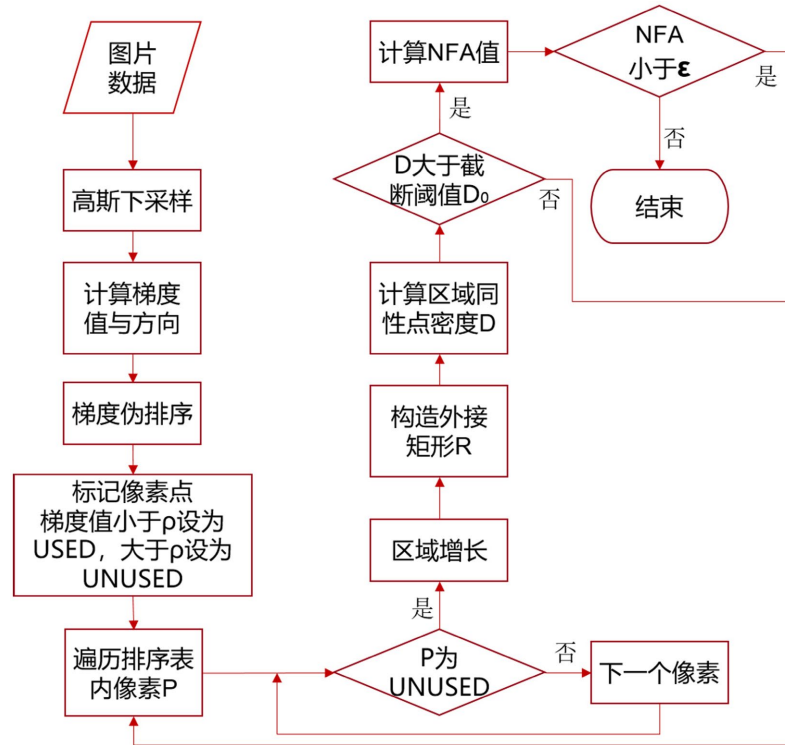


Figure 7. LSD algorithm flowchart
图 7. LSD 算法流程图

首先从外观上来看，鸟窝的目标框是一个长宽比相差不大的矩形，根据这一特点，当模型检测到的目标框高度与宽度有显著差异时，明显是与实际情况不符的，因此本文在后续添加判断：每次检测后根据矩形框的 4 个点坐标计算出对应的长和宽，当两者的差值在 3 倍以上时，认为该区域不是鸟窝。其次，鸟窝是由枯枝、青叶和草根等驻扎而成，形状粗糙，有很多凹凸点，根据其并不平滑的特点，将 LSD 直线检测算法用于检测到的目标框，查找区域中直线个数，而鸟窝区域由于凹凸点多，直线个数也会比较多，因此可以在后续处理中增加判断：对于每次检测后的目标区域进行灰度化处理，使用 LSD 算法检测区域图像的直线段个数，如果直线个数少于设定的阈值 30，则认为该区域不存在鸟窝。

4. 实验

高铁 3C 系统视频中鸟窝样本极度稀缺，可能出现在接触网的横梁、支柱和杆塔这三个地方，数据采集的困难性使得常规的深度学习模型性能受到限制，且鸟窝的目标面积很小，背景区域的干扰很容易影响到后续算法的性能，而面向具体类别的小目标检测算法设计成本高、难度大。因此在对接触网鸟窝进行检测前，先利用 YOLOv8 网络提取出鸟窝的潜在区域，并保存下来，然后将裁剪后的潜在区域图像用于鸟窝的检测，以此来避免鸟窝目标过小的问题，减小大面积背景的干扰。此外，将领域自适应与小样本目标检测应用到鸟窝的检测识别中，输电线路中的鸟窝数据比接触网鸟窝数据相对更易采集，因此选择输电线路领域中大量标注过的鸟窝图像作为源域，目标域为极少数的接触网鸟窝图像，利用 AcroFOD 算法实现鸟窝领域风格的迁移。

4.1. 数据

第一级检测器的数据集由 3C 系统实际运行采集的接触网可见光全景图像组成, 通过抽帧和人工筛选, 共收集到 5097 张图像, 其中训练集和测试集比例为 9:1, 图 8 展示了部分鸟窝驻扎的横梁、杆塔和支柱区域, 采用标注软件对图像中的目标区域横梁、支柱和杆塔进行标注, 标注信息保存在 xml 文件中, 包括了矩形框类别以及左上角和右下角的坐标信息; 第二级检测器的数据集由源域和目标域组成, 选择输电线路领域的鸟窝图像作为源域, 并用标注软件对其进行标记处理, 该场景下鸟窝数据的采集难度相对较低, 利用无人机巡检可以获取一定数量的鸟窝数据, 共收集到 665 张图像, 全部用于训练, 目标域为接触网场景下的鸟窝, 仅标记 6 张图像用于模型训练, 训练数据为整张的原始图像。测试数据中包含鸟窝的图像为 39 张, 不包含鸟窝的接触网图像有 481,038 张, 用于测试模型的误报率, 图 9 显示了源域和目标域数据之间的领域差异。

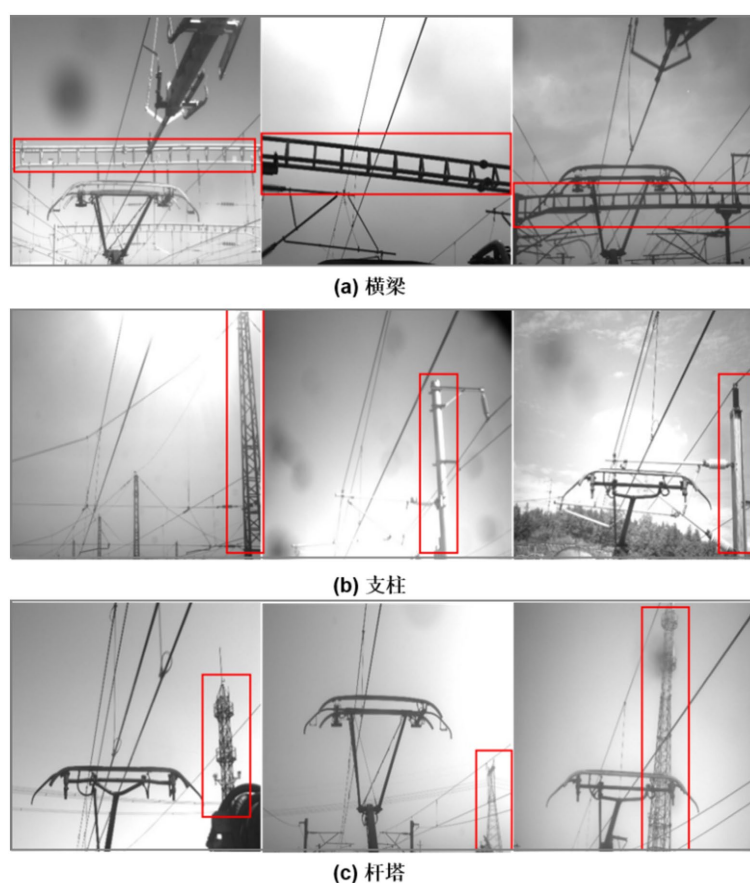


Figure 8. Data display of beams, columns, and towers

图 8. 横梁、支柱、杆塔数据展示

4.2. 设置

服务端使用了 Ubuntu 18.04 操作系统, 算法在基于 Pytorch 1.8.1 的深度学习框架中部署, 使用 Python 3.8 进行网络构建和编程, 显卡采用 NVIDIA RTX 3090 搭配 CUDA11.0。第一级检测器 YOLOv8 网络的实验参数设置如下: 以 YOLOv8n 为基准网络模型, 输入图像尺寸设置为 640×640 , 初始学习率为 0.01, 优化器采用 Adam, 动量因子为 0.937, 权重衰减系数 0.0005, 训练代数 1000 代, 训练批次大小为 32。

第二级检测器 AcroFOD 网络的实验参数设置如下：输入图像尺寸设置为 640×640 ，学习率为 0.01，优化器采用 SGD，动量因子为 0.937，权重衰减系数 0.0005，训练的代数为 500，训练批次大小为 16。



Figure 9. Source domain (left) and target domain (right) images
图 9. 源域(左)和目标域(右)图像

4.3. 结果

4.3.1. 鸟窝潜在区域定位

在鸟窝潜在区域检测任务中，采用精度(Precision)、召回率(Recall)和平均准确度的平均值(Mean Average Precision, mAP)来对模型性能进行综合评估，图 10 为 YOLOv8 的横梁、支柱、杆塔检测结果，训练出来的模型可以准确地将目标区域检测出来。测试集数据为 510 张，最终测试结果如表 1 所示，均值达到了 96%的精度，96%的召回率以及 97%的 mAP 值。



Figure 10. YOLOv8 detection results
图 10. YOLOv8 检测结果

Table 1. YOLOv8 test results
表 1. YOLOv8 测试结果

模型	类别	Precision	Recall	mAP (50)
YOLOv8	平均	0.96	0.96	0.97
	横梁	0.99	0.96	0.99
	杆塔	0.92	0.94	0.95
	支柱	0.96	0.98	0.96

为了说明 YOLOv8 模型的有效性, 在同等实验设置下相继比较了该算法与其他方法: Faster R-CNN [13]、SSD [14]和 YOLOv3 [9]在同一测试集中的性能。实验结果如图 11 和表 2 所示, 其中 Faster R-CNN 为二阶段目标检测方法, SSD、YOLOv3 和 YOLOv8 为一阶段目标检测方法, 由图 11 可见, 本文采用的 YOLOv8 网络在三种指标上均取得了最高精度, 其中 mAP 指标超出 Faster R-CNN 2%, Precision 指标超过 Faster R-CNN 和 SSD2%, 由表 2 可见, YOLOv8 检测速度优于其他同类型算法。可见在对鸟窝潜在区域的检测识别上, YOLOv8 网络模型具有更好的性能。

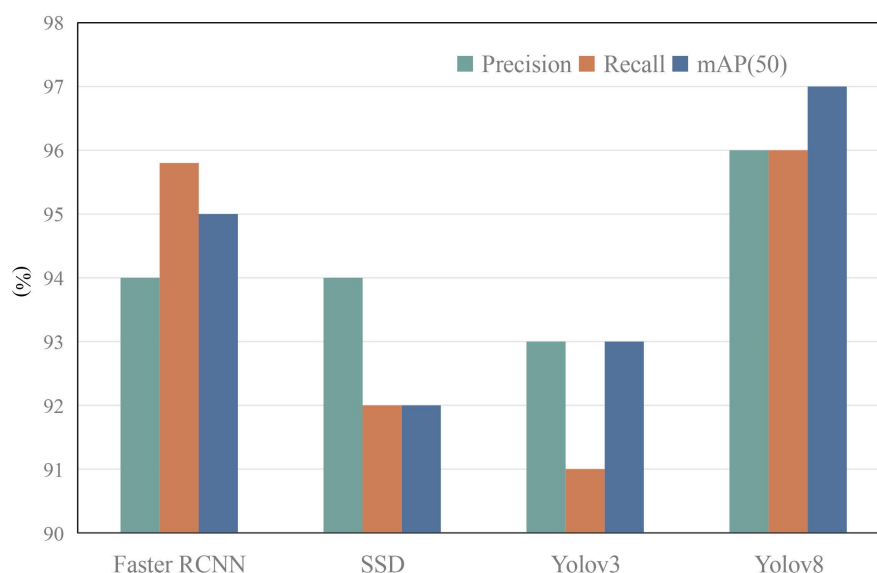


Figure 11. Performance comparison of different algorithms

图 11. 不同算法性能对比

Table 2. Comparison of detection speeds of different algorithms

表 2. 不同算法检测速度对比

算法	时间(ms)
Faster R-CNN	64
SSD	45
YOLOv3	32
YOLOv8	30

4.3.2. 类别迁移与域迁移对比

为了更好的证明接触网场景中基于域适应的小样本目标检测方法的可行性优于主流的小样本目标检测方法, 对 6 种通用主流算法 TFA [15]、FSCE [16]、MetaRCNN [17]、FSDetView [18]、MPSR [19]和 VFA [20]进行了复现, 选择 PASCAL VOC 数据集中 15 个类别作为基类, 5 个类别作为新类, 结果如表 3 所示, 其中 Class Split1 表示新类与基类的一种划分方式, 新类包括鸟、公交车、奶牛、摩托车和沙发, Shot 中的数字表示新类中每个类别的训练数为 5 张, 从表 3 中可以看出, FSDetView 算法的性能最佳。继而将该算法应用至本场景中进行训练, 以 3C 系统中已有大量标注过的受电弓、吊弦、上环线、下环线、中锚线和尼龙套作为基类, 鸟窝为新类, 将训练后的模型用于测试, 部分测试结果可视化如图 12 所示, 可以看出, 对于同一场景模型只能检测出基类, 而无法将基类的知识迁移到新类上, 实现鸟窝的检测识别,

然后将 FSDetView 和 AcroFOD 网络在同一测试集中进行测试，结果如图 13 所示，AcroFOD 方法的检出率远高于 FSDetView。综上所述，对于接触网中的鸟窝，本文所采用的域适应目标检测方法 AcroFOD 要优于基于主流的方法。



Figure 12. FSDetView detection results
图 12. FSDetView 检测结果

Table 3. Reproduction results of mainstream few-shot object detection algorithms
表 3. 主流小样本目标检测算法复现结果

Shot	算法	Class Split1					Mean	
		Bird	Cus	Cow	Motorbike	Sofa	Novel	Base
5	TFA (ICML'2020)	0.244	0.309	0.244	0.304	0.185	0.257	0.444
	FSCE (CVPR'2021)	0.269	0.463	0.219	0.440	0.254	0.406	0.432
	MetaRCNN (ICCV'2019)	0.524	0.638	0.456	0.640	0.341	0.599	0.625
	FSDetView (ECCV'2020)	0.510	0.697	0.547	0.664	0.329	0.625	0.651
	MPSR (ECCV'2020)	0.465	0.545	0.350	0.591	0.289	0.582	0.626
	VFA (AAAI'2023)	0.556	0.676	0.450	0.674	0.379	0.591	0.611

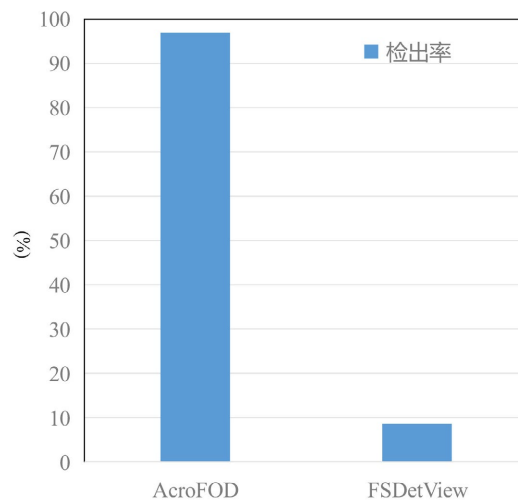


Figure 13. Comparison of the results of the method in this paper with FSDetView
图 13. 本文方法与 FSDetView 结果对比

4.3.3. 消融实验

考虑到投入训练的目标域样本个数是左右 AcroFOD 网络性能的一个重要参数, 这里选择不同数量的标记样本进行实验, 保持其他实验设置不变, 从已有的接触网鸟窝数据中分别抽取 4、5、6、7、8 张图片, 实验结果如表 4 所示, 当样本数从 4 增加到 6 时, 检出率和误报率都有了一定改善, 标记样本的增加对模型性能起到了一定的优化作用, 当样本数从 6 增加到 8 时, 检出率没有变化, 但误报率反而增高了。综上所述, 在域适应小样本目标检测网络 AcroFOD 网络中, 需要小规模标记的目标数据来检测鸟窝, 从精度提升程度和误报情况上综合考虑, 适合投入训练的目标域鸟窝数量为 6 个。

Table 4. Comparison of results with different target domain data volumes

表 4. 不同目标域数据量结果对比

目标域样本数	检出率	误报率
4	87.5%	5.6%
5	90.6%	4.1%
6	96.9%	2.4%
7	96.9%	3.5%
8	96.9%	4.5%

由上述对 AcroFOD 网络结构的介绍可知, 在进行域感知增强时, 基于 imag-level 有域拼接和域重新分配两种方式, 基于 box-level 有直接交换、混合交换和高斯交换 3 种方式, 不同组合类型的域感知增强会对模型性能产生一定影响, 本文对多种拼接方式进行实验, 保持其他实验设置固定不变, 目标域训练个数为 6, 结果如表 5 所示, 实验表明, 当只采用图像级增强中的拼接方式时, 误报率高达 6.8%, 采用组合形式后误报情况有了明显的下降, 图像级与盒级的增强都能够明显的改善模型性能, 其中拼接、重分配和直接交换这 3 种增强方式相结合时, 检出率较高的同时误报率降到了 2.4%, 取得了最优结果。

Table 5. Multi-layer domain-aware enhancement of different combination types

表 5. 不同组合类型的多层域感知增强

	Image-level		Box-level			检出率	误报率
	拼接	重分配	直接交换	混合交换	高斯交换		
AcroFOD	✓					93.8%	6.8%
	✓		✓			87.5%	5.1%
	✓			✓		96.9%	4.4%
	✓				✓	93.8%	4.9%
	✓	✓				87.5%	5.5%
	✓	✓	✓			96.9%	2.4%
	✓	✓		✓		96.9%	4.9%
	✓	✓			✓	93.8%	4.6%

4.3.4. 直接检测与级联检测对比

在接触网鸟窝检测识别任务中, 利用检出率和误报率来对模型性能进行评估, 检出率是指正确预测的样本数与真实样本总数的比值, 误报率是指预测错误的个数与预测总数的比值, 检出率越高的同时误

报率越低，说明模型越可靠。为了更好的说明本文提出的级联检测方法的优越性，先对原始图像进行直接检测，检测结果与级联检测的结果进行对比，部分样本的检测结果如图 14 所示，上半部分为直接检测结果，可以看出，左边图中的鸟窝没有被检测出来，且还有一个误检框位于横梁区域上方，右边图中的鸟窝虽然被检测到，但在横梁区域下方仍然有一个误检框。下半部分为级联方法测试结果，先用第一级检测器检测并截取出横梁区域，对于单独的横梁区域，鸟窝均被成功检测到。第一级检测器起到放大物体的作用，让第二级检测器能够仅在原始图像的部分区域检测目标，待测目标的平均 IOU 值明显大于直接检测的方法，直观来看，提高了精度的同时还降低了误报率。实验结果如表 6 所示，可以看出无论是检出率还是误报率，本文提出的级联的方法都要优于直接用原始图像进行检测。

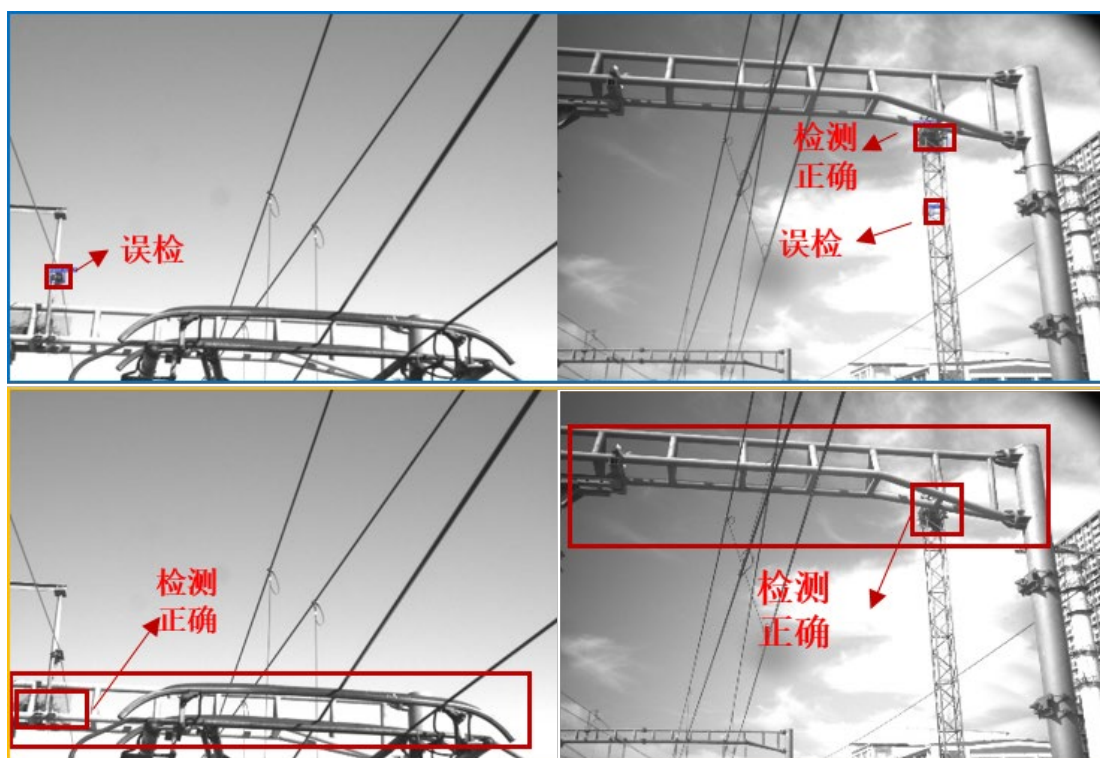


Figure 14. Partial sample detection results

图 14. 部分样本检测结果

Table 6. Comparison of two detection methods

表 6. 两种检测方式对比

	检出率	误报率
直接检测	90.6%	3.8%
级联	96.9%	2.4%

4.3.5. 后处理优化

当目标域训练个数为 6，采取级联检测的方式，且利用拼接、重分配和直接交换的组合类型时，模型能够达到 96.9% 的检出率，但误报率仍高达 2.4%，无法投入实际系统的应用。因此，为了进一步降低误报情况，本文结合鸟窝区域的形状特点，对程序进行了进一步的优化，将后处理中描述的长宽比判断和 LSD 直线检测作为算法的改进部分。

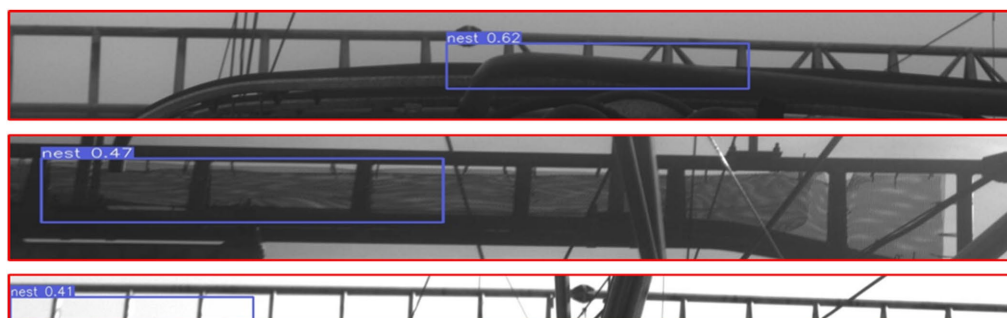


Figure 15. False detection cases with significantly different aspect ratios

图 15. 长宽比相差较大的误检情况

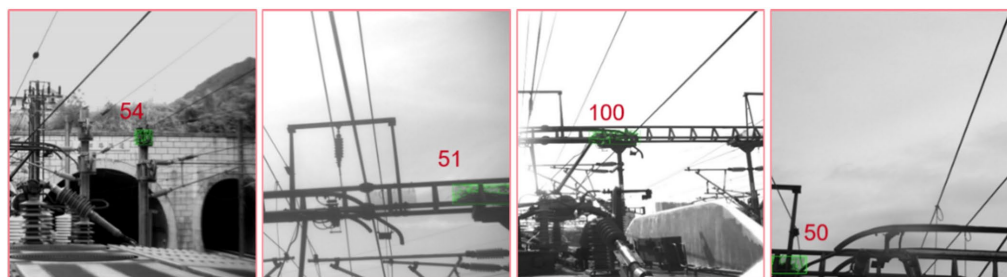


Figure 16. Number of straight line detections in the bird's nest area

图 16. 鸟窝区域的直线检测个数

图 15 显示了部分长宽比相差较大的误检情况，图 16 显示了部分鸟窝区域的直线检测数量情况，为了验证算法优化后的有效性，本文统计了 48 万张测试集的误报情况，结果如表 7 所示，在保持原来检出率的同时，误报率由之前的 2.4% 降到了 0.1%，可见优化后的算法程序能够有效降低误报。

Table 7. Comparison of results before and after algorithm optimization

表 7. 算法优化前后结果对比

	检出率	误报率
优化前	96.9%	2.4%
优化后	96.9%	0.1%

5. 结论

本文针对高速铁路接触网中的鸟窝检测提出了一种级联的解决方法，在第一级检测器中用前沿目标检测网络 YOLOv8，以较高精度定位鸟窝可能出现的横梁、杆塔、支柱区域，大幅缩小目标物检测的搜索范围，避免了部分无关背景的干扰。在第二级检测中只需检测潜在区域中的鸟窝位置，将域适应小样本目标检测方法首次应用在接触网鸟窝场景，标注少量目标域数据就能获得较高精度。为说明 AcroFOD 算法在本场景中的优越性，将 AcroFOD 与主流的小样本目标检测方法 FSDetView 进行对比，该方法得到了 96.9% 的检出率，远远高出主流算法，证明了场景迁移比类别迁移更适用于高速铁路接触网上的鸟窝识别。并在后续对算法进行优化，在保持相同检出率的前提下，误报率由 2.4% 降到了 0.1%，体现了本文方法的可行性。此外轨道交通普遍存在故障样本数量很少的情况，小样本学习方法在鸟窝识别中的有效性证明该方法可移植到弓网异常识别项目和主动防撞等场景中，为少量异常样本训练导致故障识别模型鲁棒性和泛化性差的问题提供一种低成本，高精度的解决方案。

参考文献

- [1] Lin, B. and Lin, R. (2017) An Approach to the High-level Maintenance Planning for EMU Trains Based on Simulated Annealing. arXiv: 1704.02752.
- [2] 段旺旺, 唐鹏, 金炜东, 等. 基于关键区域 HOG 特征的铁路接触网鸟巢检测[J]. 中国铁路, 2015(8): 73-77.
- [3] 徐晶, 韩军, 童志刚, 等. 一种无人机图像的铁塔上鸟巢检测方法[J]. 计算机工程与应用, 2017, 53(6): 231-235.
- [4] 杨沛. 双判别器生成对抗网络及其在接触网鸟巢检测的应用研究[D]: [硕士学位论文]. 成都: 西南交通大学, 2018.
- [5] 王纪武, 罗海保, 鱼鹏飞, 等. 基于 FasterR-CNN 的铁路接触网鸟巢检测[J]. 铁道机车车辆, 2020, 40(2): 78-81, 108.
- [6] Gao, Y., Yang, L., Huang, Y., Xie, S., Li, S. and Zheng, W. (2022) AcroFOD: An Adaptive Method for Cross-Domain Few-Shot Object Detection. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M. and Hassner, T., Eds., *Computer Vision—ECCV 2022*, Springer, 673-690. https://doi.org/10.1007/978-3-031-19827-4_39
- [7] Jocher, G., Chaurasia, A. and Qiu, J. (2023) Ultralytics YOLO (Version 8.0.0) [Computer Software]. <https://github.com/ultralytics/ultralytics>
- [8] Xiao, Z., Zhong, P., Quan, Y., Yin, X. and Xue, W. (2020) Few-Shot Object Detection with Feature Attention Highlight Module in Remote Sensing Images. 2020 *International Conference on Image, Video Processing and Artificial Intelligence*, Shanghai, 22-23 August 2020. <https://doi.org/10.1117/12.2577473>
- [9] Redmon, J. and Farhadi, A. (2018) YOLOv3: An Incremental Improvement. arXiv: 1804.02767.
- [10] Yang, Z., Wang, Y., Chen, X., Liu, J. and Qiao, Y. (2020) Context-Transformer: Tackling Object Confusion for Few-Shot Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, **34**, 12653-12660. <https://doi.org/10.1609/aaai.v34i07.6957>
- [11] Wang, T., Zhang, X., Yuan, L. and Feng, J. (2019) Few-Shot Adaptive Faster R-CNN. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 7166-7175. <https://doi.org/10.1109/cvpr.2019.00734>
- [12] Jocher, G.R., et al. (2021) ultralytics/yolov5: v5.0 - YOLOv5-P6 1280 Models, AWS, Supervise.ly and YouTube Integrations. Zenodo.
- [13] Ren, S., He, K., Girshick, R. and Sun, J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. <https://doi.org/10.1109/tpami.2016.2577031>
- [14] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., et al. (2016) SSD: Single Shot Multibox Detector. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *Computer Vision—ECCV 2016*, Springer, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
- [15] Wang, X., et al. (2020) Frustratingly Simple Few-Shot Object Detection. arXiv: 2003.06957.
- [16] Sun, B., Li, B., Cai, S., Yuan, Y. and Zhang, C. (2021) FSCE: Few-Shot Object Detection via Contrastive Proposal Encoding. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 7348-7358. <https://doi.org/10.1109/cvpr46437.2021.00727>
- [17] Yan, X., Chen, Z., Xu, A., Wang, X., Liang, X. and Lin, L. (2019) Meta R-CNN: Towards General Solver for Instance-Level Low-Shot Learning. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October-2 November 2019, 9576-9585. <https://doi.org/10.1109/iccv.2019.00967>
- [18] Xiao, Y., et al. (2020) Few-Shot Object Detection and Viewpoint Estimation for Objects in the Wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **45**, 3090-3106.
- [19] Wu, J., Liu, S., Huang, D. and Wang, Y. (2020) Multi-Scale Positive Sample Refinement for Few-Shot Object Detection. In: Vedaldi, A., Bischof, H., Brox, T. and Frahm, J.M., Eds., *Computer Vision—ECCV 2020*, Springer, 456-472. https://doi.org/10.1007/978-3-030-58517-4_27
- [20] Han, J., Ren, Y., Ding, J., Yan, K. and Xia, G. (2023) Few-Shot Object Detection via Variational Feature Aggregation. *Proceedings of the AAAI Conference on Artificial Intelligence*, **37**, 755-763. <https://doi.org/10.1609/aaai.v37i1.25153>