

doi: 10.3969/j.issn.1005-7854.2021.01.014

基于改进 YOLOv3 的锌阴极板残留物图像识别方法

鲁恒润 李 强 杨文旺

(矿冶科技集团 北矿机电科技有限责任公司, 北京 100160)

摘 要: 为识别剥锌机锌阴极板的残留物, 研究了一种基于深度学习的阴极板残留物图像识别方法。在 YOLOv3 单阶段图像检测模型的基础上, 将原始网络结构 Darknet53 替换成具有深度可分离卷积的 Xception 网络结构来提升检测模型的特征提取能力, 通过改进候选框生成策略来减轻训练过程中正负样本不平衡。实验从训练过程损失曲线以及识别准确率分别对比了改进结构与原始结构的识别效果。结果表明, 基于改进 YOLOv3 的方法提升了检测模型的泛化能力, 对阴极板残留物的识别准确度高达 95%, 较原始的 YOLOv3 模型提升了 3%。

关键词: 深度学习; 残留物识别; 阴极板; YOLOv3

中图分类号: TH16 **文献标志码:** A **文章编号:** 1005-7854(2021)01-0074-06

Image recognition method of zinc cathode plate residue based on improved YOLOv3

LU Heng-run LI Qiang YANG Wen-wang

(BGRIMM Machinery and Automation Technology Co. Ltd., BGRIMM Technology Group, Beijing 100160, China)

Abstract: In order to identify the residue of zinc cathode plate of zinc stripping machine, an image recognition method based on deep learning was studied. Based on the YOLOv3 single-stage image detection model, the original network structure Darknet53 is replaced by the Xception network structure with deep separable convolution to improve the feature extraction ability of the detection model, and by improving the candidate box generation strategy to reduce the imbalance between positive and negative samples. The experiment compares the recognition effect of the improved structure and the original structure from the loss curve of training process and the recognition accuracy. The results show that the method based on improved YOLOv3 improves the generalization ability of the detection model, the recognition accuracy of cathode plate residue is up to 95%, 3% higher than the original YOLOv3 model

Key words: deep learning; residue identification; cathode plate; YOLOv3

阴极板的锌片脱离是锌电解生产工艺中的重要步骤之一, 自动剥锌机通过剥刀可实现锌片自动剥离, 但由于剥锌机在剥离锌片过程中会产生锌片剥离不彻底导致极板残留锌片。传统方法需要工人使用工具进行二次剥离, 人工剥离效率低、劳动强度大。随着机器视觉技术的快速发

展, 使用视觉技术识别锌阴极板残留具有非接触、实时性好的特点, 可以有效提高剥锌机的剥锌效率, 提升装备智能化水平。

在图像识别领域, 传统的识别方法大都采用基于特征的方式^[1-2], 通过提取目标识别物的纹理、颜色、形状等信息进行识别, 该方法依赖于人工设计的特征, 当目标物的颜色、光照等发生变化时, 识别准确度、稳定性无法保证。另外, 基于模板匹配的方法, 根据图像之间的相似度进行识别, 同样容易受到外界干扰(光照、拍摄角

收稿日期: 2020-06-23

基金项目: 北京市科技计划项目(Z171100000817011)

第一作者: 鲁恒润, 硕士, 主要从事计算机视觉、深度学习方面的研究。E-mail: l_hengrun@bgrimm.com

度等)造成的影响。

近年来,深度学习技术在图像识别等领域快速发展,基于深度学习的图像识别方法通过多层卷积神经网络自动提取图像的特征,具有较高的鲁棒性和识别准确度^[3-13]。本文以算法 YOLOv3^[12]为基础,通过改进算法的主干网络以及候选框生成策略提升了模型的泛化能力,提高锌阴极板的残留物识别的准确度,实现故障锌阴极板的在线识别。

1 YOLOv3 算法原理

YOLO(You only look once)系列算法是单阶段目标检测器中经典算法之一,由 Joseph Redmon 等在 2016 年首次提出。YOLOv3 是其第三代版本,在 YOLOv1^[10]和 YOLOv2^[11]算法的基础上,结合 FPN 特征金字塔^[13]和 Faster-rcnn^[8]的 Anchor 机制,显著提升了单阶段检测器的检测精度。

1.1 边界框预测

YOLOv3 使用 k-means 聚类方法为每个尺度各生成 3 个锚点框,网络预测边界框的 4 个坐标值,计算方法如公式 1~4。

$$b_x = \sigma(t_x) + c_x \quad (1)$$

$$b_y = \sigma(t_y) + c_y \quad (2)$$

$$b_w = p_w e^{t_w} \quad (3)$$

$$b_h = p_h e^{t_h} \quad (4)$$

式中, t_x 、 t_y 、 t_w 、 t_h 为网络的预测输出; (c_x, c_y) 为相对于图像左上角的偏移值; p_w 、 p_h 为锚点框的宽度和高度; b_x 、 b_y 、 b_w 、 b_h 预测值转化为实际输出坐标值, $\sigma(\cdot)$ 为 Sigmoid 函数。

1.2 多尺度预测

为解决小目标检测较差问题,YOLOv3 采取了 FPN 特征金字塔网络的思想^[12-14],输出三种不同尺度的特征图并将不同尺度特征图相互融合做预测。YOLOv3 使用了 52×52 、 26×26 和 13×13 三种尺度的输出作为预测,三种不同尺度的特征图分别从网络的不同层输出且相互融合,包含了更多的语义信息,大尺度的特征图语义信息丰富,适合检测小目标物体,通过多尺度预测提升了对小目标检测的效果。

1.3 主干网络

YOLOv3 使用 darknet53 作为主干网络,

darknet53 具有 52 个卷积层和 1 个全连接层做成,交替使用 3×3 和 1×1 的卷积核,采用残差模块解决网络层数加深梯度消失或梯度爆炸导致的训练退化问题。YOLOv3 的 3 个尺度输出分别从主干网络的 32 倍下采样、16 倍下采样和 8 倍下采样层获得,尺度特征图之间通过 1×1 的卷积和 2 倍上采样操作分别进行尺度融合。

1.4 损失函数

YOLOv3 的损失函数由 3 部分组成,分别是坐标损失、置信度损失和类别损失。其中坐标损失采用平方和误差进行评价,置信度和类别概率使用二值交叉熵进行评价。

2 算法改进

2.1 主干网络结构

网络的输入由 $256 \times 256 \times 3$ 修改为 $512 \times 512 \times 3$ 分辨率提升了一倍,有利于保留更多的有用信息。使用改进的深度可分离卷积网络 Xception^[14]代替原有的 Darknet53 主干网络。如图 1 所示,Xception 网络由输入头、中间体和输出体三部分组成,网络结合了残差和宽度拓宽思想,在 Inception 网络的基础上将普通的卷积替换成可分离卷积。三个预测图分别从 Xception 网络模块 Block4、Block13 和 Block14 的可分离卷积层输出,大小分别为 $64 \times 64 \times 728$ 、 $32 \times 32 \times 1024$ 和 $16 \times 16 \times 2048$ 。与 YOLOv3 的结构相同,尺度特征图之间通过 1×1 的卷积和 2 倍上采样操作分别进行尺度融合操作。

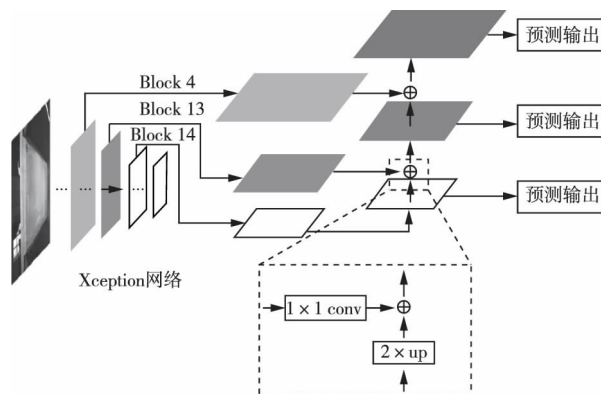


图 1 YOLOv3-Xception 网络结构

Fig 1 Network structure of YOLOv3-Xception

2.2 正负样本不平衡处理

YOLOv3 单阶段目标检测在候选窗口的生成

上采用锚点机制，当正样本较少情况下会产生大量的负样本，大量负样本会导致网络过度关注负样本，降低模型的识别效果(增加漏检率)。在阴极板残留物识别属于单类目标识别且绝大多数故障板仅有1块或者2块呈块状分布的残留区域。实验中发现这种候选框生成机制会导致严重的正负样本失衡问题，正负样本比高达1:318。

通过观察阴极板残留，发现阴极板残留物所在区域主要分布在极板的两侧。为了进一步统计阴极板残留的分布区域规律，针对残留物目标统计训练集中所有图片目标中心所落区域的散点如图2所示(黑色点为训练集中标记的残留物区域的矩形框的中心点位置，(x,y)中心点在原图中的像素点位置)。

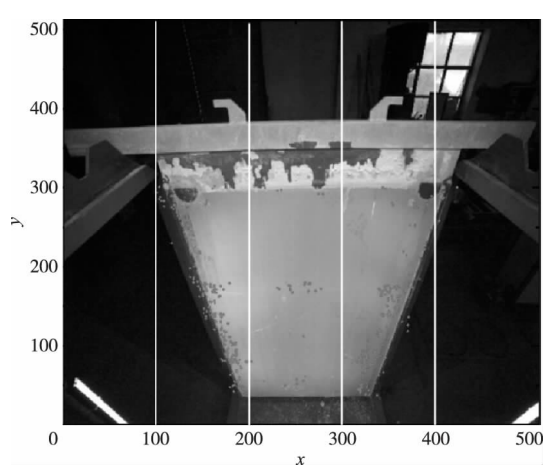


图2 目标中心所落区域分布
Fig.2 Target center distribution

从图2中可以看出，阴极板残留物主要分布在极板两侧，中间区域绝大部分为负样本区域。如图3所示，以横坐标x将图片划分为5个区域，分别统计5个区域内落点占比。

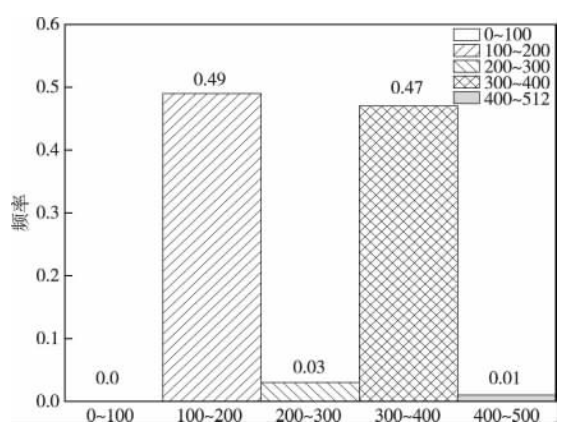


图3 目标中心所落区域分布频率
Fig.3 Distribution frequency of target center

在训练过程对于预测的坐标值(b_x, b_y, b_w, b_h)当迭代次数小于等于N时，按照原有Anchor机制生成候选框。当迭代次数大于N且生成框的横坐标位于0~100、200~300或者400~500区域时，预测Anchor坐标值按照指数进行衰减，如公式5所示。

$$\begin{cases} (b_x, b_y, b_w, b_h) & \text{IF epoch} \leq N \\ (b_x, b_y, b_w, b_h) e^{-(\text{epoch}-N)/2} & \text{ELSE} \end{cases} \quad (5)$$

式中，epoch为迭代次数，N为可调超参数；(b_x, b_y, b_w, b_h)为预测坐标值。在训练初期正负样本对网络的训练影响不大，当训练到一定次数后，由于正负样本的失衡会导致网络对负样本倾斜，此时需要减低负样本的权重值，使网络增加对正样本关注。根据阴极板残留区域的分布规律可知，残留主要分布在阴极板的两侧，因此可对除两侧之外的其他区域(负样本出现概率大的区域)的预测坐标值进行衰减。

3 实验结果与分析

3.1 参数配置

使用Python3.5编程语言、Tensorflow1.14深度学习框架，结合Intel(R)-CPU-E5-2620-V4处理器，在Ubuntu 16.04操作系统下进行实验。使用Nvidia Titan RTX显卡、CUDA10.0和cuDNN7.6调用GPU进行加速。迭代次数N取1000，使用Adam梯度下降，学习率取0.001，batch大小为4。

3.2 实验数据集分布

如表1所示，数据集分为训练数据集和测试数据集，共486幅图片数据。其中有残留图片384幅，从中随机抽取318幅作为训练数据集，66幅作为测试数据集，无残留数据102幅，从中随机抽取29幅作为训练数据集，73幅作为测试数据集。

表1 训练和测试集分布			
Table 1 Distribution of training and test sets			
类别	数量/幅	大小	
训练	有残留	318	512×512
	无残留	29	512×512
测试	有残留	66	2592×2048
	无残留	73	2592×2048

3.3 实验结果分析

表2所示为3种不同模型的识别结果，

Model1 使用的是原始 YOLOv3-darknet53 模型，Model2 使用的原始 YOLOv3 中将 Darknet 主干网络替换为 Xception 网络的 YOLOv3-Xception 模型，Model3 使用的改进的 YOLOv3-Xception 模型(替换主干网络为 Xception 网络并调整生成

候选框策略)。从表 2 可知，相比原始的 YOLOv3-darknet53 模型，改进的 YOLOv3 模型在有残留识别率和无残留识别率上均有较大提升。测试识别效果显示：Model 3 > Model 2 > Model 1。

表 2 识别结果

Table 2 The results of recognition

序号	使用模型	迭代次数	有残留识别率	无残留识别率	综合识别率
Model 1	YOLOv3-Darknet53	1 000	0.910	0.930	0.920
Model 2	YOLOv3-Xception	1 000	0.930	0.950	0.940
Model 3	Ours	1 000	0.950	0.960	0.955

为进一步分析原因，做出了 3 种模型训练过程的损失曲线图。如图 4 所示，Model 1 训练集损失曲线随着迭代次数增加逐步下降并趋于平缓，但交叉验证集损失曲线振荡下降且曲线始终位于训练集损失曲线上方，表明模型处于欠拟合状态，学习效果较差。

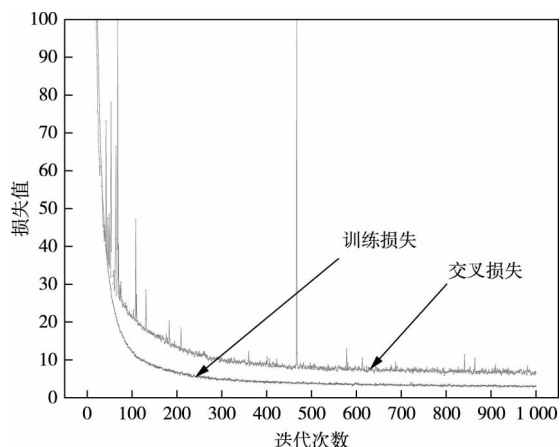


图 4 YOLOv3-Darknet53 模型损失曲线

Fig 4 Loss curves of YOLOv3-Darknet53 model

如图 5 所示，Model 2 训练集损失曲线和交叉验证集损失曲线随着迭代次数增加逐步下降并趋于平缓，交叉验证集损失曲线略微位于训练集损失曲线上方，说明相对于原始的 YOLOv3-Darknet53 主干网络，Xception 网络提升了模型的拟合能力。

如图 6 所示，Model 3 训练集损失曲线和交叉验证集损失曲线随着迭代次数增加逐步下降并趋于平缓，且 2 个曲线几乎重合，模型的拟合能力进一步得到了提升，主要原因是通过改进候选框生成策略，解决了模型正负样本失衡问题，使得模型对正负样本的学习能力得到了提升。

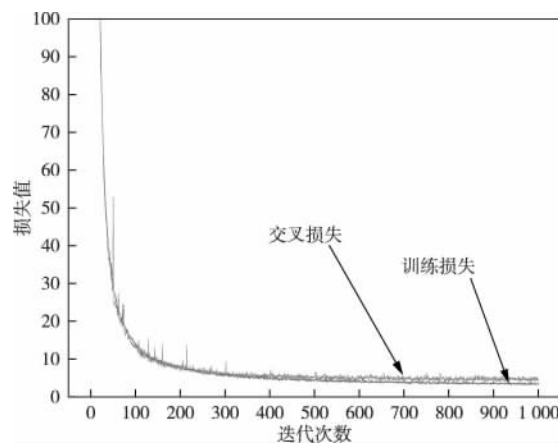


图 5 YOLOv3-Xception 模型损失曲线

Fig 5 Loss curves of YOLOv3-Xception model

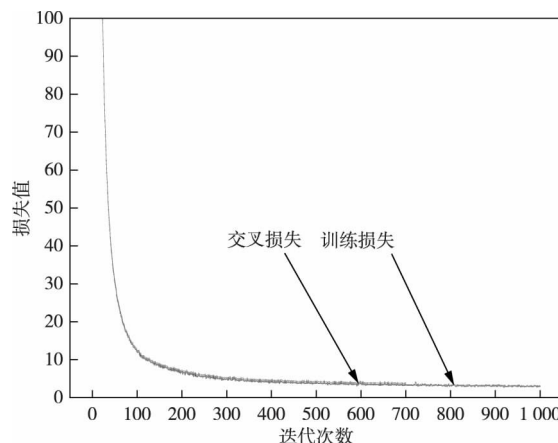


图 6 改进的 YOLOv3-Xception 模型损失曲线

Fig 6 Loss curves of improved YOLOv3-Xception model

训练过程损失曲线进一步证明了通过主干网络和候选框生成策略的调整，检测模型的泛化能力得到了明显提升。如图 7 所示，列举了常见的几种锌极板残留情况以及使用本文改进算法的识别结果。

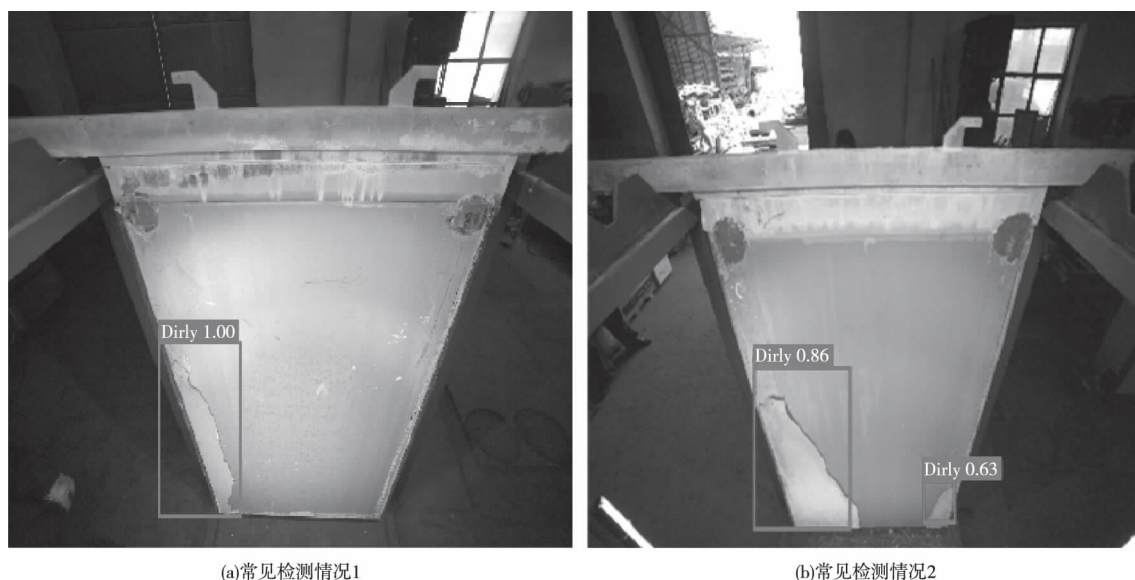


图 7 实际检测效果

Fig. 7 Effect of actual detection

4 结论

1)针对原有网络特征提取能力不足,使用Xception网络调整了原有Darknet53主干网络结构,提升了模型的特征提取能力。

2)通过训练过程预测值指数衰减改进候选框生成策略,有效减低正负样本不平衡导致学习向负样本偏移问题,有利于降低识别的漏检率。

参考文献

- [1] 张泽苗,霍欢,赵逢禹. 深层卷积神经网络的目标检测算法综述[J]. 小型微型计算机系统, 2019, 40(9): 1825.
ZHANG Z M, HUO H, ZHAO F Y. Survey of object detection algorithm based on deep convolutional neural networks[J]. Journal of Small Microcomputer Systems, 2019, 40(9): 1825-1831.
- [2] 鲁恒润,王卫东,徐志强,等. 基于机器视觉的煤矸特征提取与分类研究[J]. 煤炭工程, 2018, 50(8): 137-140.
LU H R, WANG W D, XU Z Q, et al. Extraction and classification of coal and gangue image features based on machine vision [J]. Coal Engineering, 2008, 50(8): 137-140.
- [3] 王萍萍,仇润鹤. 基于YOLOv3的车辆多目标检测[J]. 科技与创新, 2020(3): 68-70.
WANG P P, QIU R H. Vehicle multi-target

detection based on YOLOv3 [J]. Science and Technology & Innovation, 2020(3): 68-70.

- [4] 舒军,吴柯. 基于改进YOLOv3的航拍目标实时检测方法[J]. 湖北工业大学学报, 2020, 35(1): 21-24.
SHU J, WU K. Real-time aerial target detection method based on improved YOLOv3[J]. Journal of hubei university of technology, 2005, 35(1): 21-24.
- [5] 陈正斌,叶东毅,朱彩霞,等. 基于改进YOLOv3的目标识别方法[J]. 计算机系统应用, 2020, 29(1): 49-58.
CHEN Z B, YE D Y, ZHU C X, et al. Object recognition method based on improved YOLOv3[J]. Computer System Application, 2020, 29 (1): 49-58.
- [6] 李成,汪小凯,章振原,等. 基于深度残差网络的有色金属破碎料视觉识别方法[J]. 有色金属工程, 2019, 9(8): 127-134.
LI C, WANG X K, ZHANG Z Y, et al. Visual recognition method for nonferrous metal crushed aggregates based on deep residual network [J]. Nonferrous Metal Engineering, 2019, 9(8): 127-134.
- [7] 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, Columbus, USA: IEEE, 2014: 580-587.

- [8] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.
- [9] LIU W, ANGUELOV D, ERHAN D, SZEGEDY C, et al. SSD: Single shot multi-box detector [C] // Proceedings of the 2016 European Conference on Computer Vision, Berlin: Springer-Verlag, 2016: 21-37.
- [10] REDMON J, DIVVALA S K, FARHADI A, et al. You only look once: Unified, real-time object detection [C] // Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA: IEEE, 2016: 778-779.
- [11] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C] // Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA: IEEE, 2017: 6517-6525.
- [12] REDMON J, FARHADI A. YOLOv3: An incremental improvement [C] // Proceedings of the Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 1-4.
- [13] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C] // Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA: IEEE, 2017: 2117-2125.
- [14] CHOLLET F. Xception: Deep Learning with Depthwise Separable Convolutions [C] // Proceedings of the IEEE conference on computer vision and pattern recognition. Honolulu, Hawaii: IEEE, 2017: 1251-1258.