

基于改进 YOLOv4 的电机端盖缺陷检测^①



万卓, 叶明, 刘凯

(南京航空航天大学机电学院, 南京 210016)

通讯作者: 叶明, E-mail: yemingjs@163.com

摘要: 在基于机器视觉实现电机端盖裂纹缺陷检测过程中, 针对复杂背景下目标特征不明显的问题, 使用限制对比度的自适应直方图均衡化的方法加强目标特征. 针对机器视觉系统中训练数据量少且训练图片背景单一导致模型泛化性低的问题, 对比了 Mosaic 和 CutMix 数据增强方法, 并结合多种数据增强策略, 提出了系统的数据集构建方案. 针对使用 YOLOv4 进行单类检测和小目标检测时正负样本不平衡导致检测率低的问题, 提出了自适应多尺度焦点损失+CIoU 损失的加权融合损失函数, 并通过实验得到最优超参数. 最后使用 K-means 算法初始化锚点框, 使模型更适应线状目标的预测. 结果表明, 该方法对于裂纹类别的检测达到了 95.8% 的平均精度 (Average Precision, AP), 相较于改进前提升 9.7%, 单张检测时间 48 ms, 具有一定的工程应用价值.

关键词: 机器视觉; 缺陷检测; YOLOv4; 数据增强; 焦点损失

引用格式: 万卓, 叶明, 刘凯. 基于改进 YOLOv4 的电机端盖缺陷检测. 计算机系统应用, 2021, 30(3): 79-87. <http://www.c-s-a.org.cn/1003-3254/7808.html>

Defect Detection of Motor Cover Based on Improved YOLOv4

WAN Zhuo, YE Ming, LIU Kai

(College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: Adaptive histogram equalization with limited contrast is applied to strengthening the target feature to solve the problem of unclear targets in a complex background during crack detection of motor covers based on machine vision. A systematic dataset construction scheme is proposed by comparing Mosaic and CutMix data augmentation and combining with a variety of data enhancement techniques to address the low generalization of the model induced by the small volume of training data in the machine vision system and single background of training pictures. Besides, a weighted fusion loss function combined with adaptive multi-scale focus loss and CIoU loss is proposed to deal with the low detection rate caused by unbalanced numbers of positive and negative samples in the single class detection and small target detection of YOLOv4, and the optimal hyper parameters are obtained through experiments. Finally, the anchor box is initialized by the K-means algorithm to make the model more suitable for predicting linear targets. Results demonstrate that this method achieves an Average Precision (AP) of 95.8% for detecting crack types, which is 9.7% higher than before, and the single-sheet detection time is 48 ms, presenting the potential for engineering application.

Key words: machine vision; defect detection; YOLOv4; data enhancement; focal loss

在现代化生产过程中, 由于技术工艺水平、操作不当等因素造成的精铸零件表面缺陷是很多重大安全

事故的罪魁祸首, 因此识别和检测精铸零件表面缺陷是一种非常重要的产品质量控制手段.

① 基金项目: 国家自然科学基金 (51405229); 江苏省自然科学基金 (BK20151470)

Foundation item: National Natural Science Foundation of China (51405229); Natural Science Foundation of Jiangsu Province (BK20151470)

收稿时间: 2020-07-06; 修改时间: 2020-08-11; 采用时间: 2020-08-17; csa 在线出版时间: 2021-03-03

根据某企业实际的检测要求,需要在生产线上在线检测电机端盖的侧面裂纹缺陷,如图1所示。电机端盖是电动机的核心零部件,起着支撑传动部分、确定转子与定子的相对位置、保护电机内部和导风的作用,其质量优劣将直接影响电机的可靠运行和工作人员人身安全。该任务的难点体现于:①侧面裂纹目标特征不明显、形态多变且位置不固定,传统的机器视觉算法难以解决;②由于相机视野须覆盖电机端盖整个侧面,导致待检测的裂纹目标相对整个图片很小,属于小目标检测;③在机器视觉系统中,工业相机位置固定,采集的图像的背景单一,不利于深度学习数据集的构建。



(a) 型号一

(b) 型号二

图1 电机端盖实物图

机器视觉系统中运用深度学习目标识别算法,可以解决很多传统视觉算法难以解决的问题。田晓杰等^[1]提出了一种优化的SSD检测方法,用DenseNet网络结构替代VGG网络结构,从而优化了SSD检测算法对小目标的鲁棒性,解决硅片隐裂检测问题,但是检测速度较慢,不能满足实时性。马美荣等^[2]基于RetianNet设计了一个目标检测器,在复杂背景下的检测有极佳的表现,解决了手机主板缺陷检测的问题,缺点是检测器对外界环境变化的适应性一般。目前常用的算法包括Faster R-CNN^[3]、SSD^[4]、YOLO系列^[5-9]等,其中YOLOv4是最优秀的目标检测算法之一,其汇集了之前的众多研究成果,达到了检测精度和效率的最佳平衡,本文基于YOLOv4进行端盖裂纹的检测,并从数据集构建和损失函数两个方面进行一系列改进,使其在端盖数据集上达到更高的检测精度。

1 数据集构建

1.1 图像增强

电机端盖的侧面特征主要包括光亮带、撕裂带和裂纹,如图2所示,我们希望得到的零件能获得100%的光亮带,而实际生产中由于设备缺陷,很难得到完全

的光亮带,断面上通常容易产生撕裂带,但企业的质检标准认为撕裂带不影响零件的使用,只要不出现较深的裂纹,就认为是合格产品,这无疑增加了检测的难度。当裂纹出现在撕裂带的时候,特征极不明显,势必会影响后续的检测过程。因此本文首先通过图像增强的方式强化裂纹目标的特征。

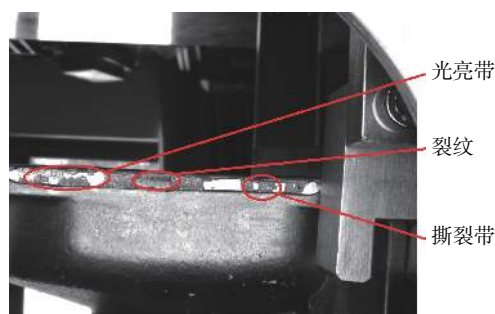
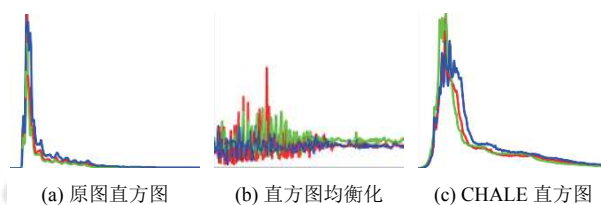


图2 端盖侧面拍摄图

直方图均衡化是常用的图像增强方式,其原理是将图像的直方图按照概率分布做一个简单的稀疏拉伸,如图3(b)所示。但这样的结果会导致变换后的图像虽然对比度得到增强,但是细节纹理部分缺失,同时图片中大片暗背景被加亮很多,而其中少量的实际需要关注的目标对比度反而没有得到改善,如图4(a)。

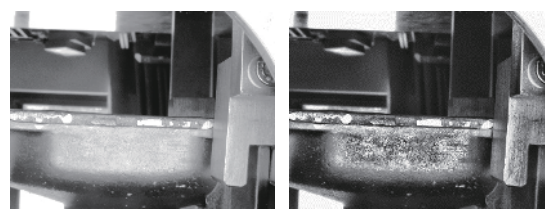


(a) 原图直方图

(b) 直方图均衡化

(c) CHALE 直方图

图3 变换前后直方图



(a) 直方图均衡化

(b) CHALE 变换

图4 变换后结果图

针对这个问题,本文使用一种限制对比度的直方图均衡化(CHALE)算法^[10,11],取得了较好的效果如图4(b)。算法流程如下:

(1) 将原图像划分为 n 个大小为 $x \times y$ (本文中取 13×13)的固定子块;

(2) 对这 n 个固定子块计算灰度直方图 H_i ;

(3) 计算平均分配到每个灰度值的像素个数: $N_{avg} = N_x N_y / N_{gray}$ 其中, N_x 是子区域 x 轴方向的像素数; N_y 是子区域 y 轴方向的像素数, N_{gray} 是子区域中灰度级的数量;

(4) 计算剪切阈值: $N_{res} = N_{clip} N_{avg}$, N_{clip} 设的剪切倍数, 即剪切阈值为平均分配到每个灰度值的像素个数的 N_{clip} 倍;

(5) 使用对应的剪切阈值对每个子块的灰度直方图 H_i 进行剪切, 剪切下来的像素数目重新分配到各直方图的各灰度级中;

(6) 分别对剪切后的每个子区域的灰度直方图进行直方图均衡化;

(7) 获取每个子块中心点的灰度值, 将该点作为参考点, 采用双线性插值的方法对图像中的每个像素进行灰度线性插值, 最终得到的直方图如图 3(c) 所示. 裂纹得到明显区分, 消除了周围背景的干扰, 同时不存在过度增强的问题.

1.2 图像增强

数据增强^[12,13]的目的是扩充数据集并增加输入图像的多样性, 从而使设计的目标检测模型具有更高的鲁棒性和泛化性. 端盖缺陷检测中, 由于没有公开的数据集可以利用, 且数据通过固定位置的 CCD 相机采集, 导致采集的图片背景单一, 通过数据增强辅助构建一个完善的数据集尤为重要.

常用的数据增强方法有 3 类: ① 空间变换: 尺寸变化、随机切片、旋转等; ② 像素变化: 改变饱和度、曝光量等; ③ 信息变化: 随机擦除、添加噪声等. 改变饱和度和曝光量可以模拟工业现场环境的拍摄情况, 如光源亮度衰减、光路失真和外界环境光变化等. 如图 5(b) 和图 5(c) 所示. 对图像增加信噪比为 0.98 的椒盐噪声或高斯噪声, 模拟拍摄过程中产生的噪声, 并弱化高频特征防止模型过拟合, 如图 5(d) 所示.

本文除了使用这两个数据增强方法之外, 对比了 Mosaic 数据增强和 CutMix 数据增强. Mosaic 是一种新的数据增强方法, 该方法在图片上随机选取一个参考点, 过该点作水平和竖直线, 将图片分为 4 个大小不同的部分, 依次填充 4 个训练图片, 如图 6(a) 所示. CutMix 方法是将图片中随机大小的一部分区域裁剪掉, 然后随机填充训练集中的其他数据的区域像素值, 如图 6(b) 所示.

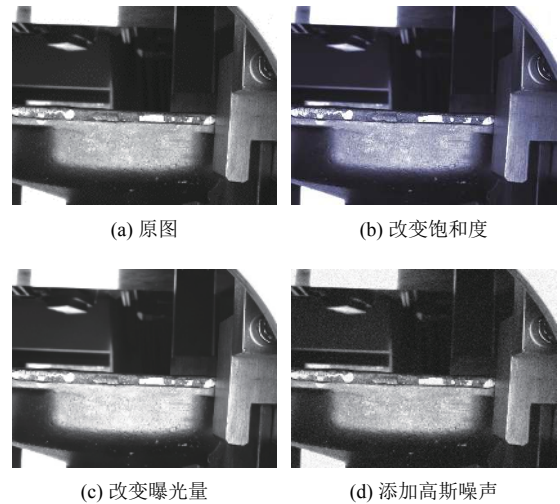


图 5 数据增强效果图

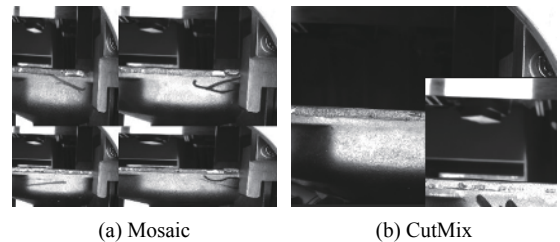


图 6 Mosaic 数据增强和 CutMix 数据增强对比

在第 4 节中, 通过实验对比了 Mosaic 和 CutMix 数据增强在电机端盖数据上的表现, Mosaic 效果略好于 CutMix, 我们分析是由于电机端盖数据集的特殊性造成的, 电机端盖数据是由固定位置相机采集, 图片背景相对固定, 使用 Mosaic 数据增强能带来的更多的背景多样性, 另外, CutMix 可能会分割待检目标, 影响目标完整性.

最终结合多种数据增强方法, 构成了包含 2000 张图片的端盖数据集, 按 6:2:2 的比例划分为训练集、验证集、测试集, 并使用 Colabeler 工具对数据集制作标签.

2 基于改进 YOLOv4 的端盖裂纹识别算法

2.1 YOLOv4

YOLOv4 总体上由骨干网络、颈部网络和头部网络 3 部分组成. 骨干网络采用 CSPDarknet53, 是在 YOLOv3 的 Darknet53 基础上引进了 CSPNet 块思路, 包含 29 个卷积层, 725×725 的感受野^[14,15], 27.6 M 参数, 在速度和检测精度上都有一定的提升. 颈部网络采用了空间金字塔池化 (Spatial Pyramid Pooling, SPP) 模块和路径聚合网络 (Path-Aggregation Neck, PAN).

SPP 用于扩宽模型的感受野, 分离出更重要的上下文信息, 同时不会导致模型推理速度的下降. PAN 是一种比 YOLOv3 中 FPN 网络更好的对不同的骨干层进行

参数聚合的方法, 可以为之后的头部层提供更丰富的信息. 头部网络沿用了 YOLOv3 中的基于锚 (anchor) 的头部. YOLOv4 整体结构图如图 7 所示.

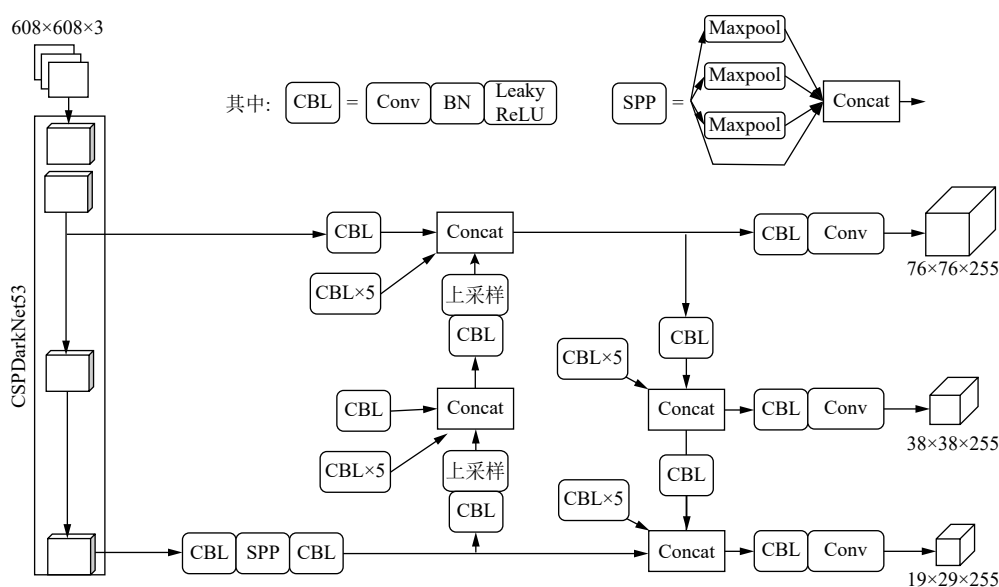


图 7 YOLOv4 结构图

YOLOv4 的主要创新点包括: ① 使用了自对抗训练 (Self-Adversarial Training, SAT) 进行数据增强; ② Mosaic 数据增强; ③ 在跨批次归一化 (Cross Batch Normal, CBN) 的基础上改进得到跨最小批的归一化 (Cross mini-Batch Normal, CmBN) 方法; ④ 改进 PAN, 把通道的连接方法从相加 (Add) 改为拼接 (Concat); ⑤ 修改空间注意力模块 (Spatial Attention Module, SAM), 从逐空间改为逐点.

除此之外, YOLOv4 使用了大量改善模型策略, 包括使用 Mish 激活函数、使用 DropBlock 正则化方法、跨阶段部分连接、改进 SOTA 方法使其更适合单 GPU 训练、使用遗传算法寻找最优超参数等.

2.2 YOLOv4 损失函数计算规则

YOLOv4 中负责计算损失函数和输出结果的 YOLO 层是基于锚 (anchor) 的, YOLO 层在 3 个尺度下进行预测, 每个尺度划分为 $S \times S$ 个网格 (对于 608×608 的输入图片, S 分别为 19、38、76), 每个网格产生 9 个候选框 (anchor box), 每个 anchor box 会经过网络预测得到相应的预测框 (bounding box). 通过人工标注的目标框称为真实框 (ground truth box), 每个 bounding box 和 ground truth box 包括目标的中心坐标 x, y , 宽高

w, h , 置信度, 以及类别 (本文数据集的类别数为 1), 同时每个 box 负责对 3 个 anchor box 进行回归, 因此在一个批次的的一个尺度的计算中, 输出 box 的尺寸为 $(4+1+1) \times 3 = 18$. 前向运算的过程中, 根据 bounding box 和 ground truth box 的差异来计算损失函数.

YOLOv4 的损失函数分为 3 个部分, 置信度损失、定位损失、分类损失. Bounding box 按照一定的规则被划分为正样本、负样本和忽略样本, 对于正样本, 需要计算所有 3 部分损失, 对于负样本, 只需要计算置信度损失, 因为负样本不负责目标, 所以这个 anchor box 产生的 bounding box 的中心坐标、宽高、分类都属于无用信息, 也就不需要计算这部分的损失函数, 但是置信度仍然重要, 因为置信度还可以表示这个 anchor box 产生的 bounding box 中不包含对象, 所以仍然需要计算置信度损失函数. 忽略样本不需要计算损失函数.

YOLOv4 对正负样本和忽略样本的划分规则和 YOLOv3 略有不同, 在 YOLOv3 中, 每个网格对应的 9 个 anchor box 中与 ground truth box 的 IoU (交并比) 最大的那一个为正样本, 负责这个目标的检测, 一个目标只对应一个正样本, 其他的 anchor box 如果和 ground truth box 的 IoU 大于一个阈值 (Ignore_Thresh), 则为忽

略样本, 否则为负样本. 而在 YOLOv4 中, 每个目标可以对应多个正样本, 只要 IoU 大于设定的阈值 (IoU_Thresh, 默认值为 0.213) 就作为正样本, 如果都小于阈值, IoU 最大的为正样本. 这个改进在一定程度上减小了 YOLOv3 中正负样本不平衡的问题, 是 YOLOv4 权衡 COCO 数据集中 80 个类别后的解决策略, 但在本文中的单类检测+小目标检测问题中, 待检测的裂纹目标相对整幅图片很小, 大部分 anchor box 是不包含待检测物体的, 这样会导致负样本的贡献会远大于正样本的贡献, 即没有目标的计算部分贡献会大于有目标的计算部分, 从而网络会倾向于预测单元格不含有物体, 导致平均精度偏低. 下面本文通过改进损失的计算, 来解决这种不平衡的问题.

2.3 改进 YOLOv4 损失函数

YOLOv4 的置信度损失:

$$Loss_{conf} = - \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} [\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - C_i^j)] - \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{noobj} [\hat{C}_i^j \log(C_i^j) + (1 - \hat{C}_i^j) \log(1 - C_i^j)] \quad (1)$$

其中, S 表示网格数, 每个尺度上的特征图划分为 S^2 个网格, B 表示每个网格产生的 anchor box 的个数, I_{ij}^{obj} 和 I_{ij}^{noobj} 表示是否为正负样本, 对于正样本, I_{ij}^{obj} 为 1, I_{ij}^{noobj} 为 0, 对于负样本, I_{ij}^{obj} 为 0, I_{ij}^{noobj} 为 1, \hat{C}_i^j 表示该样本真实置信度, 取值为 0 或 1, C_i^j 表示该样本预测置信度.

考虑到在端盖数据集中出现的正负样本严重不平衡问题, 本文引进焦点损失^[16]的思想, 在原 YOLOv4 置信度损失函数的基础上, 设置样本平衡系数和调制系数. 首先于正样本置信度损失前增加样本平衡系数 α , 于负样本置信度损失前增加系数 $(1-\alpha)$, 当 α 取值在 0.5 到 1 之间时, 负样本的计算权重降低、正样本的计算权重增加, 可以起到平衡正负样本的作用. 由于只有正样本中包含物体, 通过增加正样本的计算权重, 使得模型更倾向于预测单元格中含有目标, 从而能够提高检出率. α 的取值应与正负样本之间的不平衡度成正比, α 越大, 负样本所占权重越低, 平衡效果越明显. 另外, 本文对正样本增加调制系数 $(1-C_i^j)^\gamma$, 对负样本增加调制系数 $(C_i^j)^\gamma$, 该系数可以解决训练样本中过多简

单样本导致训练的结果对困难样本识别率不高的问题. 在本文数据库中, 数据由固定相机采集, 不可避免地会产生大量单一性样本, 这种单一性会导致大量易分的简单样本, 尤其是背景中的简单负样本. 增加调制系数后以及不同的 γ 取值对简单样本和困难样本的计算权重影响如图 8 所示, 其中横坐标 x 为预测置信度, 对于正样本 x 为 C_i^j , 对于负样本 x 为 $1-C_i^j$, 由图可见, 预测置信度 x 接近 1 的简单样本, 增加调制系数后其 loss 减小很大, 而预测置信度 x 较低的困难样本增加系数后其 loss 减少较小, 这样就相当于增加了困难样本的计算权重, 从而聚焦于困难样本的学习, 提高最终的检测精度.

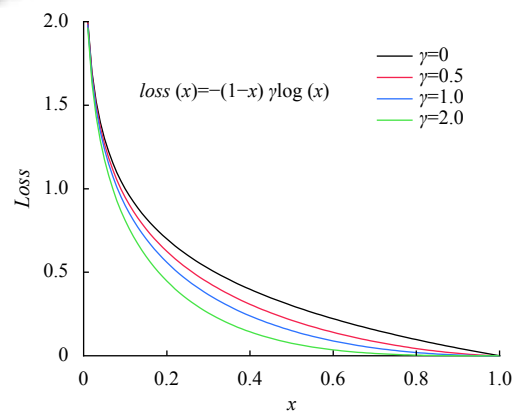


图 8 γ 取值的影响

上述基于焦点损失的改进, 对置信度损失从整体上起到了调优的作用, 并得到了理论上的解释. 但在应用到端盖数据集上, 在多次调参之后, 得到的平均精度的提升依然有限, 如表 1 所示, 焦点损失并不能很好适配于 YOLOv4. 由于 YOLOv4 是基于 3 个尺度的组合预测, 我们分别拆分 19×19 、 38×38 、 76×76 三个尺度, 在计算置信度损失的时候增加正负样本计数器, 在训练过程中统计 80 次迭代中每个尺度下的平均正负样本数, 结果如图 9 所示.

统计发现, 不同尺度下正负样本比不同, 如果使用相同的焦点损失系数 α 和 γ , 对每个尺度下起到的效果并不相同, 因此, 难以得到一组最优的参数值. 由此, 本文提出了更适应 YOLO 的多尺度焦点损失, 以 38×38 尺度为参考点, 对 19×19 和 76×76 的焦点损失系数上增加不同的尺度平衡系数 β , β 的取值根据 3 个尺度下的正负样本比自适应选取, 以此来平衡不同尺度

下不同的正负样本比. 最终, 得到改进后的置信度损失为:

$$Loss_{conf} = -\sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} [\hat{C}_i^j \beta \alpha (1-C_i^j)^\gamma \log(C_i^j) + (1-\hat{C}_i^j) \log(1-C_i^j)] - \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{noobj} [\hat{C}_i^j \log(C_i^j) + (1-\hat{C}_i^j)(1-\beta\alpha)(C_i^j)^\gamma \log(1-C_i^j)] \quad (2)$$

$$\beta = \begin{cases} \frac{N_2}{3N_1}, & S = 76 \\ 1, & S = 38 \\ \frac{3N_2}{N_3}, & S = 19 \end{cases} \quad (3)$$

其中, N_1 、 N_2 、 N_3 为 3 个尺度下的正负样本比. 置信度损失在引进焦点损失策略后, 使用多尺度平衡改进和不改进在不同参数下的表现如表 1 所示. 改进后, 在 $\alpha=0.6$ 和 $\gamma=2$ 的时候取得了最高的 AP 值, 超过了未进行多尺度改进时的最高值. 根据经验 γ 越大, α 参的值要越小一点, 不恰当的参数可能会导致梯度爆炸, AP 值为 0, 表 1 中 $\alpha=0.8$ 和 $\gamma=1$ 时, 改进前后 AP 值均为 0, 即为训练失败.

表 1 改进前后的不同参数情况 AP 对比 (%)

α	γ	改进前 AP	改进后 AP
0.5	0	93.3	93.3
0.8	0	92.3	93.8
0.8	1	0.0	0.0
0.7	1	94.0	94.3
0.6	1	94.3	94.5
0.6	2	94.3	95.8
0.5	2	94.5	95.3
0.4	2	94.1	94.8

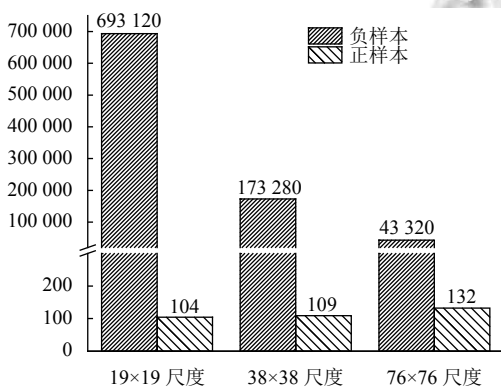


图 9 3 个尺度下正负样本数

YOLOv4 的分类损失计算的是错误分类的损失, 计算公式如下:

$$Loss_{cls} = \sum_{i=0}^{S^2} I_{ij}^{obj} \sum_{c \in \text{classes}} ([\hat{P}_i^j \log(P_i^j) + (1-\hat{P}_i^j) \log(1-P_i^j)]) \quad (4)$$

其中, \hat{P}_i^j 表示真实类别, P_i^j 表示预测类别概率值. 本文中数据集只有一个类别, 因此不对分类损失的计算做过多的优化.

YOLOv4 的定位损失包括中心坐标损失、宽高损失. IoU 损失是目标检测广泛使用的定位损失, 计算公式如下:

$$Loss_{IoU} = 1 - R_{IoU} = 1 - \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \quad (5)$$

其中, B 为预测框, B^{gt} 为真实框. 基于 IoU 损失的改进有很多, Rezatofighi 等^[17] 提出 GIoU 损失, 在 IoU 损失基础上增加了 C 检测框 (C 检测框是包含了检测框和真实框的最小矩形框), 解决了检测框和真实框没有重叠的问题. 但是当真实框和检测框之间出现包含关系的时候, GIoU 和 IoU 效果相同. Zheng 等^[18] 针对这一缺点, 提出 DIOU 损失, 将预测框和真实框都包含进 C 检测框, 但是 DIOU 计算的不是框之间的交并, 而是计算的每个检测框之间的欧氏距离. 本文的改进算法中使用 CIOU 损失^[18] 来计算定位损失, 在 DIOU 损失的基础上增加了宽高比的惩罚项, 这样预测框就会更加的符合真实框. CIOU 损失的计算公式如下:

$$Loss_{CIOU} = Loss_{GIoU} + \alpha v = 1 - R_{IoU} + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (6)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (7)$$

$$\alpha = \frac{v}{(1 - R_{IoU}) + v} \quad (8)$$

其中, b 、 b^{gt} 是 B 、 B^{gt} 的中心点, ρ 为欧氏距离, c 为能同时覆盖 B 和 B^{gt} 的最小包围框的对角线长度, w 和 h 为预测框的宽和高, w^{gt} 和 h^{gt} 为真实框的宽和高.

最终改进后的损失函数由增加自适应多尺度焦点损失的置信度损失、使用 CIOU 的定位损失和默认的分类损失组成, 并通过加权系数 k_1 和 k_2 对 3 部分进行加权融合, 损失函数为:

$$Loss = k_1 (Loss_{conf} + Loss_{cls}) + k_2 Loss_{CIOU} \quad (9)$$

本文中 k_1 和 k_2 的取值分别为 1 和 0.07.

2.4 利用 K-means 算法初始化 anchor box

端盖数据集中的裂纹目标多数为线状, 如果使用

默认的 anchor box 尺寸,可能会导致训练起始阶段由于实际目标尺寸和设定的 anchor box 尺寸差异过大,损失值的下降方向偏离甚至产生梯度爆炸. 在训练之前,使用 K-means 算法对标注出来的目标框做聚类,用聚类的结果来初始化 anchor box 大小,可以有效的避免上述情况,同时可以提高模型的学习效率,加快收敛速度,相当于在训练之前,为模型提供了先验知识.

YOLOv4 网络中对于每个网格的预测需要 9 个候选框,因此类簇个数 k 取 9. K-means 算法对初始值是比较敏感的,对于同样的 k 值,选取的初始类簇中心不同,会影响算法的聚类效果和迭代的次数,经验表明,初始的类簇中心之间的相互距离越远,聚类效果越好^[19],以此为出发点,本文通过以下过程选取初始的类簇中心点:

1) 在标注的数据集中随机选取一个初始框作为初始中心,遍历数据集中所有点 x ,计算它与已选择的类簇中心的距离 $D(x)$;

2) 选取 $D(x)$ 最大的点作为下一个中心点,遍历数据集中剩余点,计算每个点与选取的最近的类簇中心的距离 $D(x)$;

3) 重复步骤 2),直到选出 9 个类簇中心,将这些值作为 K-means 算法的初始类簇中心.

最终通过 K-means 算法对端盖数据集做聚类分析,得到 9 个候选框大小分别为: 21×17 , 41×13 , 35×19 , 42×16 , 54×13 , 48×22 , 73×22 , 97×19 , 90×23 .

3 实验与结果分析

3.1 实验环境

硬件方面,整个电机端盖缺陷检测系统由检测线体、CCD 相机、工业光源、零件工装、光学支架以及工控机组成,如图 10 所示.其中光源选用直径 120 mm 穹顶光源,与相机同轴,可以对相机视野中物体进行均匀照射. CCD 相机安装于光学支架上,搭配 8 mm 镜头,垂直拍摄电机端盖侧面,通过调整光学支架调节相机、光源与零件之间的距离,使得相机视野覆盖整个端盖侧面,且清晰成像.

软件方面,环境配置如下: Intel i5 8300 处理器, 16 GB 内存, NVIDIA GeForce GTX 1060 显卡, OpenCV 版本为 3.3.0,使用 Cuda8.0 和 Cudnn5.1 对训练过程进行加速.

网络参数配置如下:迭代次数为 10000,初始学习

率为 0.002,采用逐步递减策略^[20],训练到一定次数后分两次减小为原学习率的 10%. 动量为 0.95,权重衰减为 0.005.



图 10 电机端盖检测系统构成

3.2 实验结果对比分析

使用不同改进策略后,进行多组对比实验,得到的检测结果如表 2 所示.其中,√表示在该组实验中使用了对应策略, FCL 表示使用焦点损失+CIoU 的加权融合的损失函数, MFCL 表示在 FCL 基础上增加自适应多尺度焦点损失的改进.

表 2 使用不同改进策略检测结果

实验	CLAHE	CutMix	Mosaic	K-means	FCL	MFCL	AP(%)
1							87.3
2	√						88.8
3	√	√					92.8
4	√		√				93.0
5	√		√	√			93.3
6	√		√	√	√		94.5
7	√		√	√		√	95.8

应用本文的改进策略后,平均检测精度均获得了一定程度的提高.使用 CutMix 数据增强或者 Mosaic 数据增强后获得了最大的提升,可见,深度学习是一种数据驱动的方法,对于没有公开数据集可以使用的工业场景,使用合适的数据增强方法构架一个完善的数据集是成功检测的重要环节.

使用自适应多尺度焦点损失+CIoU 的加权融合的损失函数后, AP 值提升了 2.7%,改进损失函数前后的损失值曲线如图 11 所示.实验证明,这样的改进,对于本文中的小目标+单类检测的场景十分有效.

最终的识别结果如图 12 所示,检测的 AP 值为 95.8%,单张图片检测时间为 48 ms,满足工程应用的精度和速度要求.

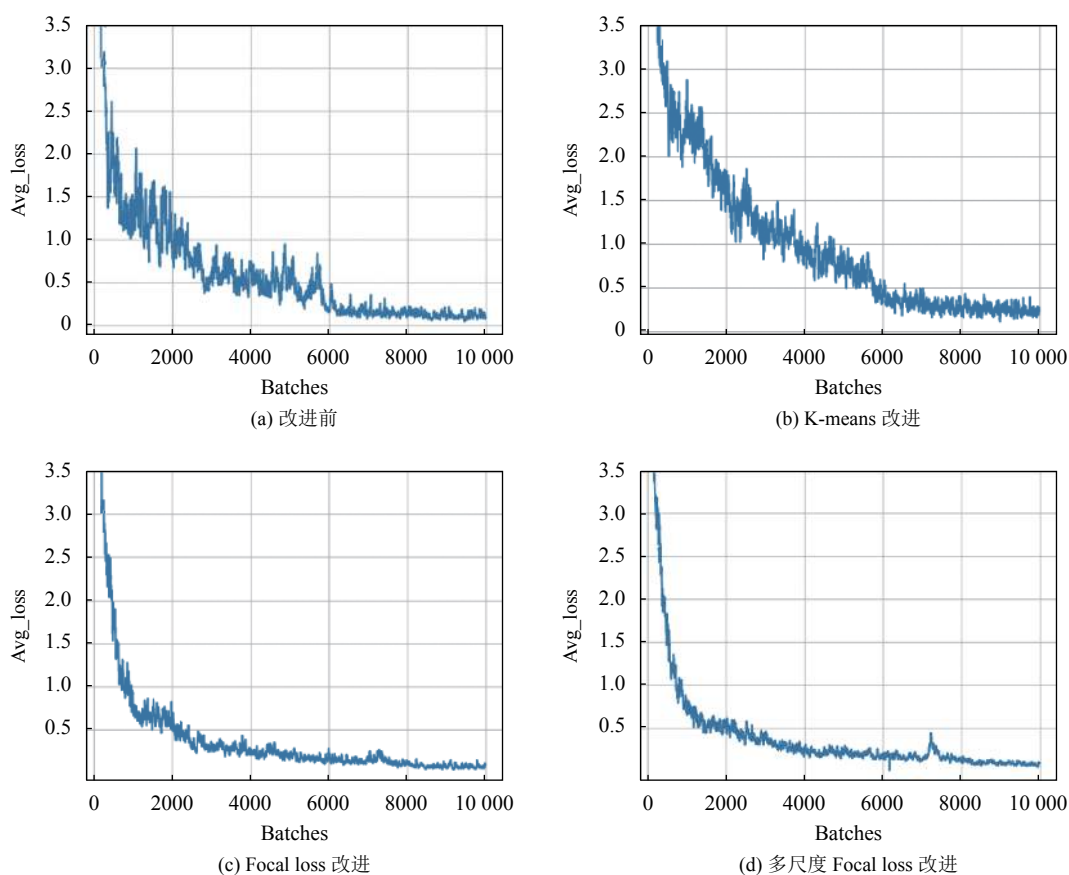


图 11 损失值曲线



图 12 识别结果图

使用本文中数据增强后的数据集,在不同的算法模型下的检测效果如表 3 所示.从实验结果可以看出,使用 two-stage 网络结构的 Faster-RCNN 算法检测速度较慢,其通过 RPN 网络推荐候选框,这一过程耗时较长,难以满足工业场景应用中的实时性要求.而 one-stage 算法中,结合了多策略的 YOLOv4 算法达到了检测精度和速度的最佳平衡,检测精度和速度都优于 SSD 算法,而同样使用焦点损失思想的 RetineNet 虽然实现了较高的检测精度,但代价是增加了检测时间.本文通过优化损失函数和优化 anchor box 初始值等改进,将 YOLOv4 对于小目标的检测精度进一步突破,同时几乎没有增加检测时间,在同类型场景下的具有一定的工程应用价值.

4 结语

本文基于改进 YOLOv4 算法实现了一个电机端盖裂纹缺陷检测系统,首先针对目标特征不明显、训练

数据少且背景单一的问题,从图像增强和数据增强两个方面系统地构建端盖数据集。之后,本文介绍了YOLOv4网络框架并着重解释了损失函数的计算规则,同时创造性地提出了自适应多尺度焦点损失+CIoU损失的加权融合的改进损失函数,并从理论和实验两个方面证明了改进损失函数的有效性。另外,在训练之前,针对裂纹目标多为线状的特点,使用K-means算法初始化anchor box,加快了收敛速度并提高了平均精度。

最终实验结果表明,使用改进策略和改进算法后取得了95.8%的AP值,相较于改进前提升9.7%。不足之处在于改进算法引入了更多的参数,参数的调优需要耗费一定的时间。

表3 不同网络模型的检测效果对比

算法模型	AP(%)	检测时间(ms)
Faster-RCNN	92.3	135
SSD	85.3	58
RetineNet	92.1	98
YOLOv4	93.0	45
本文算法	95.8	48

参考文献

- 田晓杰,程耀瑜,常国立. 基于深度学习优化SSD算法的硅片隐裂检测识别. 机床与液压, 2019, 47(1): 36–40, 60. [doi: 10.3969/j.issn.1001-3881.2019.01.009]
- 马美荣,李东喜. 基于RetinaNet的手机主板缺陷检测研究. 计算机工程与科学, 2020, 42(4): 673–682. [doi: 10.3969/j.issn.1007-130X.2020.04.014]
- 郭梦浩,徐红伟. 基于Faster RCNN的红外热图像热斑缺陷检测研究. 计算机系统应用, 2019, 28(11): 265–270. [doi: 10.15888/j.cnki.csa.007137]
- 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]
- Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multiBox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands, 2016. 21–37.
- Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA, 2016. 779–788.
- Redmon J, Farhadi A. YOLO9000: Better, faster, stronger. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA, 2017. 6517–6525.
- Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv: 1804.02767, 2018.
- Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. arXiv: 2004.10934v1, 2020.
- 张鲁筠,韩锴,徐永贵,等. 基于插值和对比度受限算法的局部直方图均衡. 潍坊学院学报, 2019, 19(2): 1–5, 15. [doi: 10.3969/j.issn.1671-4288.2019.02.001]
- Subramani B, Veluchamy M. Fuzzy gray level difference histogram equalization for medical image enhancement. Journal of Medical Systems, 2020, 44(6): 103. [doi: 10.1007/s10916-020-01568-9]
- Chawla NV, Bowyer KW, Hall LO, et al. SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 2002, 16: 321–357. [doi: 10.1613/jair.953]
- Zhang HY, Cisse M, Dauphin YN, et al. Mixup: Beyond empirical risk minimization. arXiv: 1710.09412, 2017.
- Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, USA, 2012. 1097–1105.
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, 2014.
- Lin TY, Goyal P, Girshick R, et al. Focal loss for dense object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318–327. [doi: 10.1109/TPAMI.2018.2858826]
- Rezatofghi H, Tsoi N, Gwak JY, et al. Generalized intersection over union: A metric and a loss for bounding box regression. Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA, 2019. 658–666.
- Zheng ZH, Wang P, Liu W, et al. Distance-IoU Loss: Faster and better learning for bounding box regression. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 34(7): 12993–13000.
- 张素洁,赵怀慈. 最优聚类个数和初始聚类中心点选取算法研究. 计算机应用研究, 2017, 34(6): 1617–1620. [doi: 10.3969/j.issn.1001-3695.2017.06.004]
- 林俊,党伟超,潘理虎,等. 基于YOLO的安全帽检测方法. 计算机系统应用, 2019, 28(9): 174–179. [doi: 10.15888/j.cnki.csa.007065]