

DOI: 10.16078/j.tribology.2022062

基于改进YOLOv4的扫描电镜 磨粒图像智能识别

王雨薇¹, 陈果^{2*}, 何超¹, 郝腾飞³, 马佳丽¹

(1. 南京航空航天大学 民航学院, 江苏 南京 210016;

2. 南京航空航天大学 通用航空与飞行学院, 江苏 常州 213300;

3. 南京工程学院 汽车与轨道交通学院, 江苏 南京 211167)

摘要: 磨损颗粒分析是设备磨损故障诊断和预测的有效手段, 为了提高磨粒检测的自动化和智能化程度, 提出一种基于改进YOLOv4的目标检测算法, 并应用于航空发动机扫描电镜磨粒图像识别. 首先, 新算法采用VoVNetv2-39替换YOLOv4原主干网络CSPDarknet53, 并引入BiFPN特征金字塔结构与新主干相连, 同时调整模型中所有3×3标准卷积为深度可分离卷积, 以加强多层次特征融合, 构造轻量级网络; 其次, 利用迁移学习解决扫描电镜磨粒图像数量较少的问题, 并通过冻结训练加速模型训练过程; 最后, 应用实际发动机扫描电镜磨粒图像验证, 结果表明: 新算法相较于原YOLOv4网络, 在保证精度的前提下, 网络参数量大幅降低, 识别速度提升51.1%, 满足实际扫描电镜磨粒图像快速、简洁和高精度的检测需求, 具备潜在的工程应用价值.

关键词: 航空发动机; 磨损颗粒; 扫描电镜; 深度学习; YOLOv4; 图像识别

中图分类号: TH117.1

文献标志码: A

文章编号: 1004-0595(2023)07-0809-12

Intelligent Recognition of Wear Particle Images in Scanning Electron Microscope Based on Improved YOLOv4

WANG Yuwei¹, CHEN Guo^{2*}, HE Chao¹, HAO Tengfei³, MA Jiali¹

(1. College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Jiangsu Nanjing 210016, China

2. College of General Aviation and Flight, Nanjing University of Aeronautics and Astronautics,
Jiangsu Changzhou 213300, China

3. School of Automotive and Rail Transit, Nanjing Institute of Technology, Jiangsu Nanjing 211167, China)

Abstract: Wear particle analysis is an effective method for equipment wear fault diagnosis and prediction. In order to improve the automation and intelligence degree of wear particle detection, a target detection algorithm based on improved YOLOv4 was proposed to automatically extract and identify target particles from wear particle images with complex backgrounds. It overcame the error caused by the traditional method of image segmentation in the face of multi-wear particle images, and was applied to wear particle image recognition in aeroengine Scanning Electron Microscope (SEM). Firstly, VoVNetv2-39 was used to replace YOLOv4's original backbone network CSPDarknet53. Its unique improved OSA module guaranteed the gradient and direction propagation without interference, and maintained channel information while owning deeper network depth, thus improved the network performance. Secondly, BiFPN feature pyramid structure was introduced to connect with the new backbone. BiFPN could not only completed the

Received 13 April 2022, revised 16 May 2022, accepted 17 May 2022, available online 1 June 2022.

*Corresponding author. E-mail: cgnuaacca@163.com, Tel: +86-13851875041.

This project was supported by National Science and Technology Major Project (J2019-IV-004-0071) and the Project of AECC Commercial Aircraft Engine Co, Ltd.

国家科技重大专项(J2019-IV-004-0071)和中国航发商用航空发动机有限责任公司项目资助.

feature extraction from top to bottom, but also achieved the weighted fusion of features with different resolutions from bottom to top. At the same time, BiFPN increased the horizontal connection between the input and output of the same level, enriching the semantic information of the feature map. Finally, all 3×3 standard convolutions in the model were adjusted by depthwise separable convolution, in order to build a lightweight network.

The improved YOLOv4's network architecture for image target detection consist of three parts. The first part was the new backbone feature extraction network—VoVNetv2-39. It preliminarily extracted the feature of wear particle images in SEM with the input size of 416×416×3. Three initial effective feature layers were obtained by convolution, combination and addition. The second part was the strengthen feature extraction network—SPP and BiFPN. The first feature layer was passed into BiFPN after OSA Module Stage3 and a convolution block. The second feature layer was passed into BiFPN after OSA Module Stage4 and a convolution block. The third feature layer was passed into Spatial Pyramid Pooling (SPP) after OSA module Stage5 and three convolution blocks. SPP used pooling cores with different scales of 13×13, 9×9, 5×5 and 1×1 for maximum pooling and then merged. The merged results were passed into BiFPN which realized feature fusion of the three initial effective feature layers to extract more effective features. The third part was the prediction network-YOLO Head. The feature map was divided into three kinds of grids with different numbers of 52×52, 26×26 and 13×13. The prior frames of different sizes were generated on each grid. The classification, confidence and coordinates of the wear particles in SEM images were predicted by the prior frames, and then the prediction results were output.

In the experimental stage, firstly, wear particle images in SEM with typical wear characteristics were selected. The data set was expanded by image data enhancement method. The training set, the validation set and the test set were divided to form the self-built SEM wear particle image set. Secondly, the size of all images was unified as 416×416×3, and then these images were input into the improved YOLOv4's network. Thirdly, the transfer learning was used to solve the problem of the small number of wear particle images in SEM. The ImageNet data set was used to pre-train the new model, and the freezing idea was used to accelerate the model training process and obtain the initial weight and deviation of the model. The training parameters were set, and the self-built wear particle image set in SEM was used to start freezing training. In this period, only some parameters of the network were fine-tuned. The unfreezing preparing training began after the freezing training. In this period, the weight and deviation of the model could be updated. Then, the output mAP values were compared to obtain the optimal model. Finally, the actual wear particle images in aeroengine SEM were used to verify. The results showed that, compared with the original YOLOv4 network, the number of the new network's parameters was significantly reduced and the recognition speed was improved by 51.1% on the premise of ensuring accuracy. It could meet the detection requirements of fast, simple and high precision of actual SEM wear particle images, and had the potential engineering application value.

Key words: aeroengine; wear particle; Scanning Electron Microscope (SEM); deep learning; YOLOv4; image recognition

航空发动机各部件在运行过程中通过相互作用产生摩擦, 严重时会造成磨损. 磨损产生的颗粒往往携带关于磨损程度与磨损机理的有用信息. 因此, 磨损颗粒分析是航空发动机磨损状态监测和失效诊断中一种十分有效的手段.

早期磨粒分析技术^[1-3]基于分割后的磨粒图像提取特征, 依据传统模式识别方法确定磨粒类型, 不仅耗时且处理过程复杂, 分析结果带有一定的主观偏差. 近年来, 随着深度学习的兴起, 图像识别领域获得迅猛发展. 将深度学习技术引入磨粒图像识别, 使得磨粒特征自动提取变得可能. Wang等^[4]提出融合按误差反向传播训练的多层前馈神经网络(Back propagation, BP)与卷积神经网络的两级分类模型, 对铁谱分析得

到的疲劳磨粒与严重滑动磨粒的表面细节进行像素级分析, 大幅提高了这两类磨粒的识别精度. Peng等^[5]利用迁移学习与支持向量机提出1种混合卷积神经网络, 识别性能良好, 为铁谱磨粒图像分析提供了新技术手段. Fan等^[6]针对样本相似性设计出虚拟铁谱磨粒图像, 结合迁移学习与卷积神经网络提出1种磨粒图像智能识别方法, 有效解决训练集小样本量问题. Zhang等^[7]提出1个基于类中心向量和距离对比的新模型, 该模型不仅能够识别训练集中出现过的磨粒类别, 也能够识别训练集中未出现过的磨粒新类别而无需重构和再训练, 真正实现模型泛化. Fan等^[8]通过融合传统人工特征提取法和卷积神经网络自动特征提取法提取到的特征构建分类器, 提出1个名为FFWR-Net

的新型磨粒识别网络,使得所提取到的特征更具代表性和全面性,提高了分类精度.赵春华等^[9]考虑到设备实际有效故障数据缺乏和数据标记丢失的特点,优化损失函数,基于迁移学习提出1种新的卷积神经网络集成模型,选用支持向量机作为分类器,试验结果表明,该模型特征表达能力强,识别效果好.

上述研究从不同角度为设备磨损状态检测提供了新技术,大大推动了磨粒智能识别的发展,但是其中大多数模型不具备目标检测能力,本质上只是1种“图像分类器”^[10].因此,本文中针对航空发动机扫描电镜磨粒图像提出1种基于改进YOLOv4的目标检测算法,以期从具有复杂背景的磨粒图像中自动提取目标颗粒并加以识别,克服了传统识别方法在面对多磨粒图像时需要事先进行图像分割所造成的识别误差,新算法采用VoVNetv2^[11]替换YOLOv4中原主干特征提取网络CSPDarknet53,并引入BiFPN^[12]加权双向特征金字塔结构与新主干特征提取网络相连,以增进多层次特征融合,丰富各级特征图的语义信息,同时调整模型中所有3×3普通卷积为深度可分离卷积,大幅降低网络的参数数量和运算成本.

现有的磨损颗粒图像分析技术中,铁谱技术常用于磨损颗粒的定性和定量分析,但是铁谱样本制备过程复杂耗时,磨粒图像质量较难控制,相比之下,扫描电镜磨粒图像则更为清晰,允许微米级视野范围内的有效观察,并能够产生三维立体图像用于分析微观形貌特征.扫描电镜磨粒图像是扫描电子显微镜(Scanning electron microscope, SEM)利用高能聚焦电子束在油样表面逐点扫描成像得到,扫描电镜磨粒分析法是通过电子与物质相互作用产生的信号对油样中磨粒成分、形貌及结构等进行观察与分析.扫描电镜主要由电子光学系统、信号探测系统和真空系统3部分组成,常与进行材料微区成分元素种类与含量分析的能谱仪组成密不可分的整体^[13],应用到前沿科学研究.目前,扫描电镜已广泛应用于航空发动机磨粒分析领域,通过磨粒识别结果反推磨粒产生原因与进行磨损机理表征,获知设备处于正常或异常磨损状态,成为了1种重要检测手段^[14].本文中以提升扫描电镜磨粒图像自动化以及智能化分析水平为研究目的,进一步拓展了扫描电镜技术的应用领域,提升其故障诊断的精度和效率.

最后本文中应用实际航空发动机扫描电镜磨粒图像进行验证,证明新算法可以在高效率与高精度之间达到平衡.

1 扫描电镜图像磨粒智能识别新方法

1.1 基于深度学习的图像目标检测

新兴深度学习方法已开始应用于图像目标检测,其发展可分为以R-CNN (Regions with CNN features)家族为代表的两阶段(Two-Stage)目标检测算法^[15-18]和以YOLO (You only look once)及其衍生算法为代表的单阶段(One-Stage)目标检测算法^[19-22]. Two-Stage目标检测算法第1阶段主要是寻找目标位置并得到候选区域,第2阶段才聚焦于分类候选区域并得到1个更加准确的目标位置,“两步走”战略虽然检测精度高,但也导致算法模型大,占用资源多,检测速度慢,不满足实时要求. One-Stage目标检测算法中YOLO仅通过1个主干网络中便实现候选框定位、特征提取和目标分类等功能,极大地降低模型大小,加快检测速度.随后的衍生算法不断改进,进一步提高检测速度与精度.其中,YOLOv4具有更优的检测性能和更快的检测速度,基本能够满足终端设备性能和速度要求,更适合应用到终端中.

根据扫描电镜图像中磨粒的形状、尺寸和纹理等特征可以确定磨粒类型,不同的磨粒类型可以代表设备不同的磨损机理,因此,本文中采用One-Stage目标检测方法进行扫描电镜磨粒图像目标提取和识别,对YOLOv4网络进行改进,深度强化其特征提取能力,使得模型在训练过程中不断自主学习磨粒尺寸和形貌等特征参数,在满足足够精度前提下,实现磨粒图像的快速智能识别.

1.2 图像目标检测的改进YOLOv4网络

用于图像目标检测的改进YOLOv4网络结构如图1所示,包含3个部分,第1部分为替换掉CSPDarknet53的新主干特征提取网络VoVNetv2-39,对输入尺寸是416×416×3的扫描电镜磨粒图像进行初步特征提取,通过卷积、合并和相加等操作获得3个初步的有效特征层;第2部分为加强特征提取网络SPP和BiFPN,第1个特征层在经过OSA module Stage3与1次卷积块后传入BiFPN,第2个特征层在经过OSA module Stage4与1次卷积块后传入BiFPN,第3个特征层在经过OSA module Stage5与3次卷积块后,传入空间池化金字塔(Spatial pyramid pooling, SPP)^[23],SPP利用13×13、9×9、5×5和1×1等4个不同尺度的池化核进行最大池化处理再合并,合并后的结果传入BiFPN,BiFPN不仅能够完成自上而下的特征提取,还能够实现自下而上、不同分辨率特征的加权融合,同时增加了同层级输

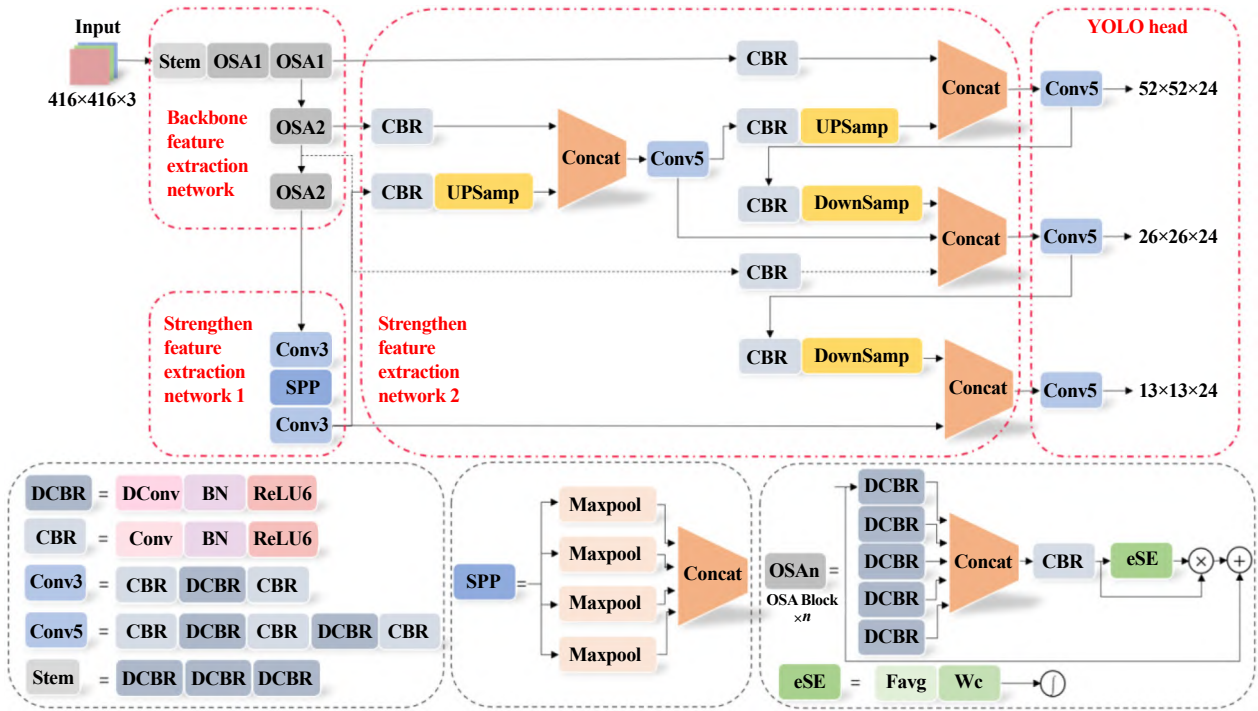


Fig. 1 Improved YOLOv4 network structure

图1 改进YOLOv4网络结构

入、输出之间的横向连接,丰富了特征图的语义信息,该部分实现对3个初步有效特征层的特征融合,提取更有效特征;第3部分为预测网络YOLO Head,将特征图划分为3种不同数量的网格,分别为 52×52 、 26×26 和 13×13 ,每个网格上生成不同尺寸的先验框,通过先验框预测扫描电镜图像中磨粒的类别、置信度与坐标,输出预测结果.其中,网络中所有的 3×3 普通卷积均调整为深度可分离卷积,用于打造轻量级网络.

1.2.1 基于VoVNet的主干特征提取网络

(1) OSA模块

密集卷积网络(Dense convolutional network, DenseNet)^[24]保存并积累具有不同感受野的特征图,区别于每层之间仅1个连接的传统L层卷积网络, DenseNet中共有 $L(L+1)/2$ 个直接连接. DenseNet核心即Dense模块结构如图2所示,其中的每一层都聚合前面所有层的特征.因此, DenseNet在目标检测方面具备更多的特征表示.

但是DenseNet过重的密集连接导致中间层特征的重复学习,最终造成特征冗余. Lee等^[25]以此为优化方向,在Dense模块基础上提出了一次性聚合(One-shot aggregation, OSA)模块,只在最后的特征映射中一次聚合前面所有特征,如图3所示.基于OSA模块搭建骨干网络VoVNet,以包含较多OSA模块的VoVNet-39网

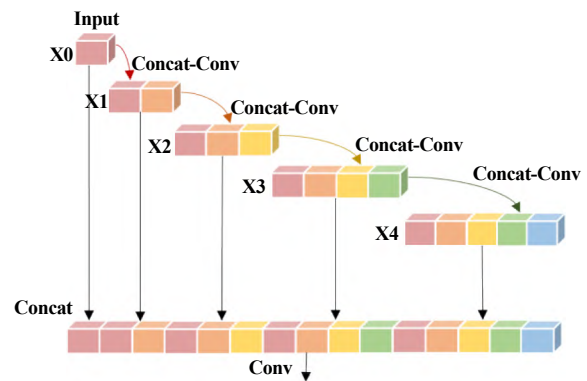


Fig. 2 Dense block structure

图2 Dense模块结构

络为例,输入尺寸为 $416 \times 416 \times 3$ 的图像, VoVNet-39网络详细结构及对应各模块输出列于表1中.不难发现, VoVNet-39由包含3个卷积层的1阶段Stem模块和4阶段OSA模块组成,每个阶段通过步长为2的 3×3 最大池化层下采样连接. VoVNet-39在第2、第3阶段分别部署1个OSA模块,在第4、第5阶段分别部署2个OSA模块,每个OSA模块中都含有5个 3×3 卷积和1个 1×1 卷积.利用OSA模块解决密集连接引起的输入通道数量线性增加问题,基于OSA模块构建的VoVNet网络不仅充分保留了DenseNet的多样化特征表示优势,还降低了内存访问成本,使GPU计算更高效.

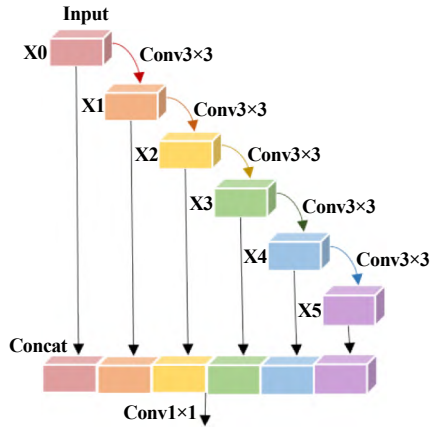


Fig. 3 OSA module structure

图 3 OSA模块结构

(2) 改进OSA模块

改进OSA模块直接将输入加到输出上, 形成恒等映射, 保障梯度和方向传播不受干扰, 进而拥有更深的网络深度; 同时在最后的特征层中引入有效压缩和机理 (effective Squeeze-Excitation, eSE) 模块增强特征, 原SE模块包含两个全连接层和1个Sigmoid函数, 以重新调整输入特征图并突出有用通道, 但也会导致通道信息丢失, eSE模块仅使用1个全连接层, 维护通道信息, 提高网络性能.

eSE模块定义为

$$X_{refine} = \sigma(W_c(F_{avg}(X_{div}))) \otimes X_{div} \quad (1)$$

式中: σ 为Sigmoid激活函数; $X_{div} \in R^{C \times W \times H}$ 是改进OSA模块中 1×1 卷积计算得到的多样化特征图 ($C \times W \times H$ 为输入eSE模块的特征大小); F_{avg} 代表全局平均池化; W_c 代表全连接; \otimes 代表元素相乘; 最后, X_{div} 通过残差连接添加到精细化特征图 X_{refine} 中. 改进OSA模块结构如图4所示, VoVNetv2-39网络即在原VoVNet-39网络结构基础上改进OSA模块得到.

1.2.2 基于BiFPN的加强特征提取网络

特征金字塔网络 (Feature Pyramid Network, FPN)^[26] 是识别不同尺寸目标的基本组成部分, 但其传统的自上而下结构不可避免会受到单向信息流限制. 为了解决这个问题, YOLOv4中采用的路径聚合网络 (Path Aggregation Network, PANet)^[27] 增加了1条自下而上的聚合路径, 精度更高, 但也导致参数量和计算量加大. 因此, 本文中采用的双向特征金字塔网络 (Bidirectional Feature Pyramid Network, BiFPN) 在PANet结构基础上, 首先删去了只有1条输入边的节点, 因为这些节点对融合不同特征的特征网络贡献较小, 由此产生了1个简化的双向网络; 其次在属于同一层级的输入、

表 1 VoVNet-39网络详细结构及对应各模块输出
Table 1 Detailed structure of VoVNet-39 and outputs of corresponding modules

Times of repetition	Type	Filters	Size	Output
Stem stage 1				
1	Convolutional	64	3×3	-
	Convolutional	64	3×3	-
	Convolutional	128	3×3	208×208
OSA module stage 2				
1	Convolutional	128	3×3	-
	Convolutional	128	3×3	-
	Convolutional	128	3×3	-
	Convolutional	128	3×3	-
	Convolutional	128	3×3	-
	Concat	-	-	-
Convolutional	256	1×1	104×104	-
OSA module stage 3				
1	Convolutional	160	3×3	-
	Convolutional	160	3×3	-
	Convolutional	160	3×3	-
	Convolutional	160	3×3	-
	Convolutional	160	3×3	-
	Concat	-	-	-
Convolutional	512	1×1	52×52	-
OSA module stage 4				
2	Convolutional	192	3×3	-
	Convolutional	192	3×3	-
	Convolutional	192	3×3	-
	Convolutional	192	3×3	-
	Convolutional	192	3×3	-
	Concat	-	-	-
Convolutional	768	1×1	26×26	-
OSA module stage 5				
2	Convolutional	224	3×3	-
	Convolutional	224	3×3	-
	Convolutional	224	3×3	-
	Convolutional	224	3×3	-
	Convolutional	224	3×3	-
	Concat	-	-	-
Convolutional	1 024	1×1	13×13	-

输出节点间增加横向连接, 从而能够在不增加成本的情况下融合更多特征; 最后利用快速规范化融合为不同分辨率的特征增加权重, 而不是统一调整到相同分辨率再相加, 打破重要性平衡, 让网络能够学习到不同分辨率的特征. BiFPN结构如图5所示.

1.2.3 基于深度可分离卷积的网络参数简化

普通卷积利用权重矩阵实现通道维和空间维的联合映射, 代价是高计算复杂度、高内存开销和大量权重系数. Kaiser等^[28]提出的深度可分离卷积 (Depthwise Separable Convolutions, DConv) 实施“两步走”战略, 分别对通道维和空间维上的特征进行映射并将结果组合, 保留普通卷积表征学习能力的同时, 减少参数

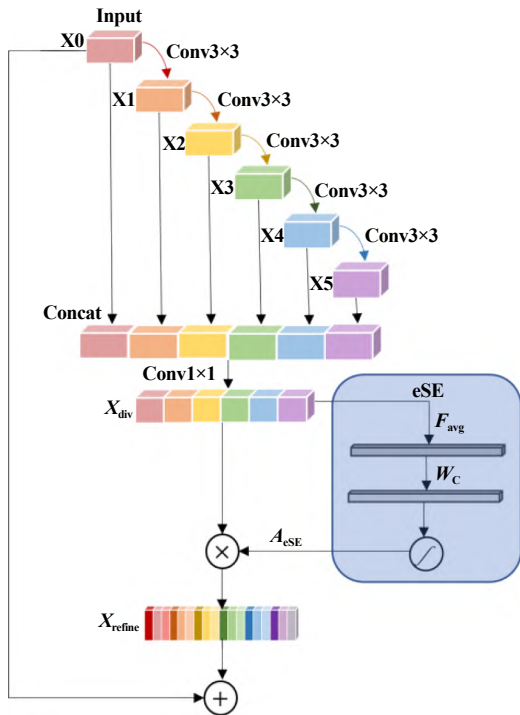


Fig. 4 Improved OSA module structure
图 4 改进OSA模块结构

数量, 提高运算效率. 深度可分离卷积操作如图6所示.

(1)逐通道卷积(Depthwise Convolution, DWConv)

假设输入图像尺寸为 $D_F \times D_F \times M$ (D_F 为输入图像的长和宽, M 为输入图像的通道数), 输出通道数为 N , DWConv为每个通道分配1个卷积核, 在二维平面内进行卷积运算, 生成 M 个特征图. 由于对每个通道进行独立卷积运算, 未能有效利用不同通道在相同空间位置上的特征信息, 且生成的特征图数量因与输入图像通道数相同无法进行扩展, 所以需要PWConv对DWConv得到的特征图进行组合.

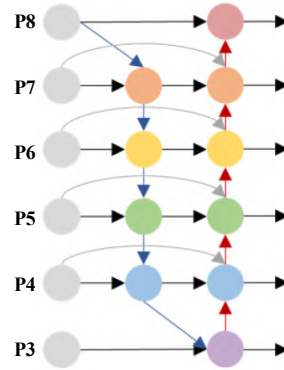


Fig. 5 BiFPN structure
图 5 BiFPN结构

(2)逐点卷积(Pointwise Convolution, PWConv)

PWConv运算与普通卷积运算十分相似, 其卷积核尺寸为 $1 \times 1 \times M \times N$, 对DWConv得到的特征图进行深度方向上加权, 得到扩展后的特征图.

下一节将介绍在迁移学习基础上利用改进YOLOv4网络实现的磨粒识别试验, 以证实本文中所提出的新型模型更符合工业中实时、简洁的目标检测需求.

2 扫描电镜磨粒识别及试验验证

2.1 扫描电镜磨粒识别流程

扫描电镜磨粒识别流程如图7所示. 首先, 选择具有典型磨损特征的扫描电镜磨粒图片, 采用图片数据增强法进行数据集扩充, 并划分训练集、验证集和测试集, 形成自建的扫描电镜磨粒图像集; 其次, 统一所有图片尺寸为 $416 \times 416 \times 3$, 输入改进YOLOv4网络, 根据迁移学习思想利用ImageNet数据集对新模型进行预训练, 获取模型初始化权重和偏差; 接着, 设置训练参数, 利用自建的扫描电镜磨粒图像集开始冻结训

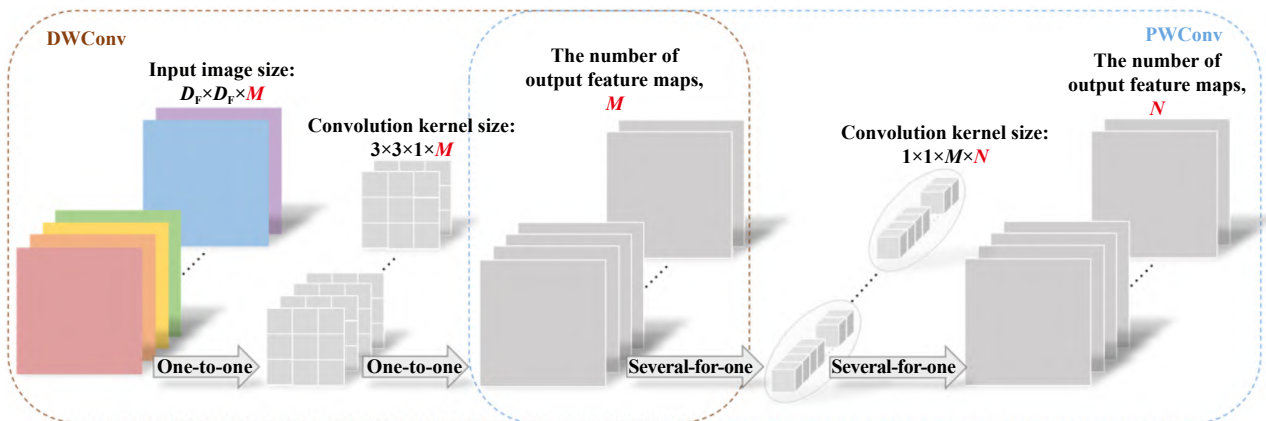


Fig. 6 Depthwise separable convolution operation
图 6 深度可分离卷积操作

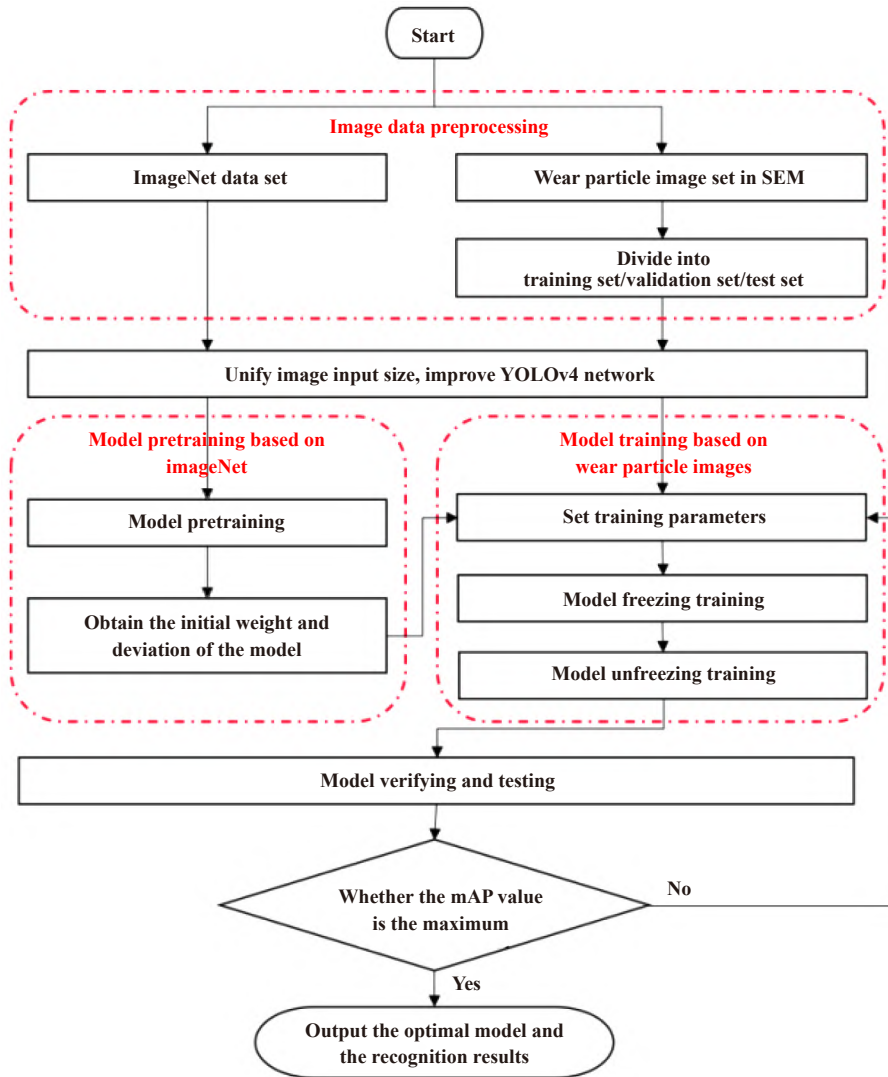


Fig. 7 The process of wear particle intelligent recognition
图 7 磨粒智能识别流程

练,仅微调网络部分参数,冻结训练结束开始解冻训练,模型权重和偏差得以更新;最后,模型验证与测试,比较输出的mAP值并得到最优模型与识别结果。

2.2 扫描电镜磨粒识别关键技术

2.2.1 图像数据预处理

将分析航空发动机扫描电镜磨粒图像问题看作是1个基于图像的目标检测问题.本文中主要进行常见类型的异常磨粒识别研究,通过以往总结得到的专家经验,依据不同磨粒对应的不同典型特征,人工从中国航发商用发动机公司滑油系统中选取684张扫描电镜磨粒图片,包含制造屑、摩擦屑和疲劳磨损屑3种,采用旋转、缩放、随机剪切、亮度调整和噪声添加等图像数据增强方法,将1张图片拓展成5张,形成了含有3 420张图片的磨粒图像集,并按训练集:验证

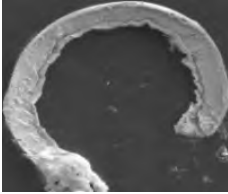
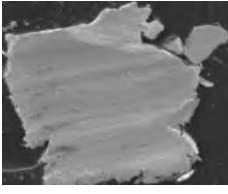
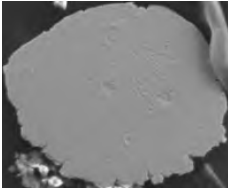
集:测试集=6:2:2的比例划分.其中,旋转以图像中心为旋转中心进行随机角度旋转;缩放按照一定比例缩小或放大图像;随机剪切是随机定义感兴趣区域进行截取并调整为原始图像尺寸;亮度调整为更改图像亮度;噪声添加则是在原始图片基础上随机叠加噪声,一定程度上影响像素平滑.上述操作都是用于获得更多形态的图像数据,并不破坏原图像中的磨粒典型特征.对应磨粒图像及其特征总结列于表2中。

2.2.2 基于ImageNet数据集的网络预训练

由于扫描电镜磨粒图像数量有限,即使是扩充后的磨粒图像集也不足以用来训练深度学习模型.本文作者利用迁移学习来初始化模型权重,以减少对样本数量的依赖。

迁移学习旨在迁移现有知识去解决仅含有少量

表2 磨粒图像及其特征总结
Table 2 Wear particle images and corresponding characteristic summary

Wear particles' types	Wear particle images	Wear particles' characteristics	Cause	Wear mechanism characterization	The number of images	The number of expanded images
Machining particles		Strip; spiral shape, with car marks; shaving shape; filaments; the surface has relatively regular processing marks	All kinds of machine chip residue	The burr of the workpiece is peeled off in the test. The workpiece cleaning is not thorough, causing that the burr drop from the hole in the processing. Always find in the early running period (within 50 hours)	114	570
Wear particles		Sheet or filaments; the surface has irregular friction, cutting and extrusion marks	Abrasive wear	The friction surface is caused by the hard raised profile of the friction pair or the hard particles entering from outside. Always find in the early running period (within 50 hours)	258	1 290
Fatigue particles		The contour is irregular and flaky; the surface is flat and shiny	Fatigue wear	The surface of friction pair deforms under the repeated action of alternating stress, resulting in accumulated damage of friction material and crack propagation, and finally leading to wear particles falling off. Always find in the certain running period (within 50 hours)	312	1 560

标签样本的目标领域学习问题. 基本原理是利用大规模图像集(本文中选用包含有120万个样本的ImageNet图像集)训练模型, 获得模型的初始权重和偏差, 然后通过扫描电镜磨粒图像集重复训练该模型以更新权重和偏差.

2.2.3 基于磨粒图像数据集的网络训练

本文中在迁移学习基础上, 引入了冻结训练的思想, 在利用磨粒图像重训练模型阶段, 考虑到所有目标检测模型, 其神经网络主干部分提取到的特征是通用的, 将主干冻结起来训练可以加快训练效率, 所以对于含有800轮次的模型总训练周期, 设前25%为冻结阶段, 特征提取网络不发生改变, 占用显存较小, 仅对网络进行微调; 后75%为解冻阶段, 特征提取网络发生改变, 占用显存较大, 网络所有参数均受影响.

2.2.4 损失函数

本文中选用的损失函数loss由3部分组成, 分别为坐标损失 $\text{loss}_{\text{coord}}$, 置信度损失 $\text{loss}_{\text{conf}}$ 和类别损失 $\text{loss}_{\text{class}}$.

$$\text{loss} = \text{loss}_{\text{coord}} + \text{loss}_{\text{conf}} + \text{loss}_{\text{class}} \quad (2)$$

式中:

$$\text{loss}_{\text{coord}} = \sum_{i=0}^{s^2} \sum_{j=0}^B I_{i,j}^{\text{obj}} \left[1 - \text{IOU} + \frac{\rho^2(b, b^{\text{gt}})}{c^2} + \frac{16 \left(\arctan \frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w}{h} \right)^4}{\pi^4 \left(1 - \text{IOU} + \frac{4}{\pi^2} \left(\arctan \frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w}{h} \right)^2 \right)^2} \right] \quad (3)$$

从公式3中可以看出, 坐标损失采用的是CIOU, s^2 为网格数量; B 为事先设定的先验框数量; 若第 i 个网格的第 j 个先验框中有目标, 则 $I_{i,j}^{\text{obj}}$ 为1, 反之则为0; IOU为预测框和真实框的交并比; b 和 b^{gt} 分别为预测框和真实框的中心位置; $\rho(b, b^{\text{gt}})$ 为两个中心位置的欧几里得距离; c 代表能够同时包含预测框和真实框的最小矩形的对角线长度; w^{gt} 和 h^{gt} , w 和 h 分别为真实框和预测框的宽与高.

$$\begin{aligned} \text{loss}_{\text{conf}} = & \sum_{i=0}^{s^2} \sum_{j=0}^B I_{i,j}^{\text{obj}} \left[-\widehat{C}_i^j \log(C_i^j) - (1 - \widehat{C}_i^j) \log(1 - C_i^j) \right] + \\ & \lambda_{\text{noobj}} \sum_{i=0}^{s^2} \sum_{j=0}^B I_{i,j}^{\text{noobj}} \left[-\widehat{C}_i^j \log(C_i^j) - \right. \\ & \left. (1 - \widehat{C}_i^j) \log(1 - C_i^j) \right] \end{aligned} \quad (4)$$

式中: \widehat{C}_i^j 与 C_i^j 分别代表真实框与预测框的置信度; λ_{noobj} 为人为设定的调整参数, 通常选择不超过1的正数; 若第 i 个网格的第 j 个先验框中没有目标, 则 $I_{i,j}^{\text{noobj}}$ 为1, 反之则为0, $I_{i,j}^{\text{obj}}$ 和 $I_{i,j}^{\text{noobj}}$ 之和始终为1.

$$\text{loss}_{\text{class}} = \sum_{i=0}^{s^2} I_{i,j}^{\text{obj}} \sum_{c \in \text{classes}} \left[-\widehat{P}_i^j \log(P_i^j) - (1 - \widehat{P}_i^j) \log(1 - P_i^j) \right] \quad (5)$$

式中: \widehat{P}_i^j 与 P_i^j 分别代表真实框与预测框的类别概率.

2.2.5 模型评价指标

模型性能评估通常从精度与速度两个角度出发.

因为磨粒识别本质上也是1种多分类问题, 每一类的准确率指标为平均精度(Average precision, AP), 不能够宏观评价模型精度, 所以引入所有类别的平均精度(Mean average precision, mAP).

$$P = \frac{T_p}{T_p + F_p} \quad (6)$$

$$R = \frac{T_p}{T_p + F_N} \quad (7)$$

$$\text{AP} = \int_0^1 P(R) dR \quad (8)$$

$$\text{mAP} = \frac{\sum_c \text{AP}}{c} \quad (9)$$

式中: P (Precision)为准确率, 也叫查准率, 用于表示某一类别中所有检测出的正样本是否实际都为正样本; R (Recall)为召回率, 也叫查全率, 用于表示某一类别中所有实际正样本是否均被检测出来; T_p (True positives)为此类中被正确划分为正样本的个数; F_p (False positives)为此类中被错误划分为正样本的个数; F_N (False negatives)为此类中被错误划分为负样本的个数, 即漏检的正样本个数; c 为类别数. 对于每一类, 可以规定1个置信度阈值(通常是0.5)来判断是否成正例. 显然, mAP值越大, 模型精度越高.

模型速度评价指标选用FPS (Frames per second), 表示画面每秒传输帧数, 公式如下:

$$\text{FPS} = \frac{1}{t_{\text{average}}} \quad (10)$$

式中: t_{average} 为平均检测1张图片所需的时间. FPS越大, 则模型推理速度越快.

2.3 试验验证结果分析

本文中根据中国航发商用发动机公司提供的磨粒图像, 采用图像数据增强法扩充小样本数据集, 建立包含3类磨损的扫描电镜磨粒图像集, 并按训练集:验证集:测试集=6:2:2的比例划分, 其中训练集、验证集和测试集分别含有2 052, 684和684个样本, 共计3 420个样本. 使用数据标记软件labelimg对每张图片目标区域进行画框标记和种类标注.

本文中基于Pytorch-1.2.0深度学习框架, 利用Python 3.6编程语言搭建新型目标检测模型, 并在硬件NVIDIA 2080TI上实现模型的训练与测试. 模型总训练轮次为800, 前200轮次处于冻结阶段, 学习率设为0.001, 训练批次(batch_size)设为8; 后600轮次处于解冻阶段, 学习率设为0.000 1, 训练批次设为4. 预设9个不同尺寸的先验框, 根据不同特征层的感受野大小进行分配, 具体分配情况列于表3中. 最后预测网络YOLO Head输出大小分别为(13, 13, 24)、(26, 26, 24)和(52, 52, 24), 如图1所示, 即YOLOv4的3个特征层分别将整个图像分为13×13、26×26和52×52的网格, 每个网格左上角的点负责对该网格区域的检测, 最后一维24可分解为3×(4+1+3), 第1个3代表三类预设框, 第2个3代表磨粒类别预测结果, 4代表预测框中心与预测框长宽的调整参数, 1为判断该种类的置信度.

表3 不同尺寸先验框分配情况
Table 3 Distribution of prior frames of different sizes

Feature maps' size	Receptive field	Prior frames' size
13×13	Large	(142×110), (192×243), (459×401)
26×26	Medium	(36×75), (76×55), (72×146)
52×52	Small	(12×16), (19×36), (40×28)

将模型训练与验证过程中损失函数值的变化过程可视化, 如图8所示. 从图8中可以看出, 前200轮处于冻结阶段, 初始学习率为0.001, 模型预训练得到的初始权重被保护, 仅部分网络参数在进行微调, 损失函数值在150轮次左右收敛于3; 后600轮次处于解冻阶段, 初始学习率为0.000 1, 训练过程中如果连续2个轮次损失值没有明显下降, 采用学习率变为原来一半的策略对学习率进行衰减, 所以冻结阶段后50个左右的轮次损失值没有明显变化, 学习率衰减到0.000 1以下, 解冻开始, 学习率突然增加, 损失值也随之突然增加, 权值更新, 损失值再快速下降, 到400轮左右损失

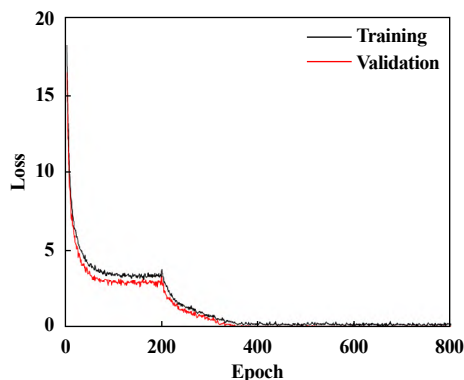


Fig. 8 Loss value during model training and validation

图8 模型训练与验证过程中的损失值

值收敛于0.1.

图9所示为某一疲劳磨粒特征识别过程,3种不同尺寸网格分别输出3张特征热力图,共9张.第1列特征图代表置信度(score),突出显示置信度超过设定阈值的对象,但不进行分类;第2列特征图代表类(class),突出显示需要判别的类的区域;第3列特征图是进行置信度矩阵与类别矩阵的加权,选定的目标对象恰好也是设定的所属类.13×13网格感受野较大,主要进行大目标识别,识别出了图中所有的疲劳屑;26×26网格感受野中等,主要进行一般大小的目标识别,所以只识别出了尺寸较小的2个疲劳磨粒;52×52网格感受野较小,主要进行小目标识别,该图像中没有小尺寸磨粒,因此未能识别出4个疲劳磨粒中的任一个.

为了证明所提网络的先进性和有效性,设定原YOLOv4网络为基线模型,对调整中不同阶段的检测网络进行消融试验,测试结果对比列于表4中.

部分磨粒图像检测结果如图10所示,方框框出的为磨损颗粒,方框上同时标有磨粒种类及对应置信度.

从表4可以看出,采用VoVNetv2-39替换CSPDarknet53作为YOLOv4的主干特征提取网络时,模型mAP值相近,参数量下降,推理速度提升至34.82/帧.更改特征金字塔网络结构为BiFPN时,虽然模型推理速度有所下降,但是检测精度高于原YOLOv4网络,说明高效的双向跨尺度连接能够充分利用深层与浅层特征优势,增进多层次特征融合,很好地弥补了轻量级网络替换造成的精度损失.将网络中的所有3×3普通卷积均调整为深度可分离卷积时,网络大小直接由231.1 MB⁻¹降为120.1 MB⁻¹,推理速度相较于YOLOv4提升了51.1%,平均每张图片的检测时间节省了0.011 05 s,从图10可以看出模型对于重叠磨粒的检测效果也较好,如图10(j), (n)和(o)所示,模型已识别出所有堵塞、拥挤的磨屑,有效解决边界框重叠被抑制的问题,从而减少漏检的可能性,说明该模型同时还能保持较高精度,符合快速、简洁和高精度的检测要求,可应用于工业实践中.

3 结论

本文中主要结论总结如下:

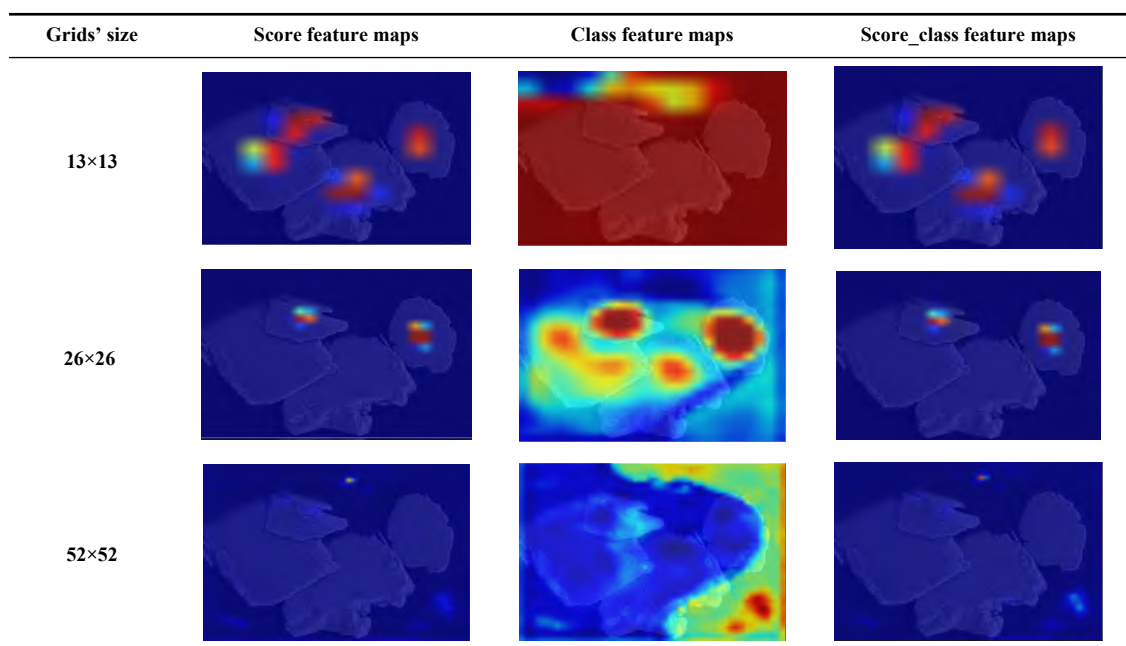


Fig. 9 Feature recognition process of one fatigue wear particle

图9 某一疲劳磨粒的特征识别过程

表4 消融试验结果
Table 4 Ablation experiments' results

Models	AP/%			mAP/%	Networks' size/MB	FPS/frame	Average detection time per image/s
	Machining particles	Wear particles	Fatigue particles				
YOLOv4	95.93	87.96	95.13	93.01	248.62	30.64	0.032 64
YOLOv4-VoVNetv2-39	96.17	85.65	95.32	92.38	217.97	34.82	0.028 72
YOLOv4-VoVNetv2-39-BiFPN	96.94	89.96	96.92	94.61	231.10	32.71	0.030 57
YOLOv4-VoVNetv2-39-BiFPN-DConv	95.28	89.20	93.45	92.64	120.11	46.31	0.021 59

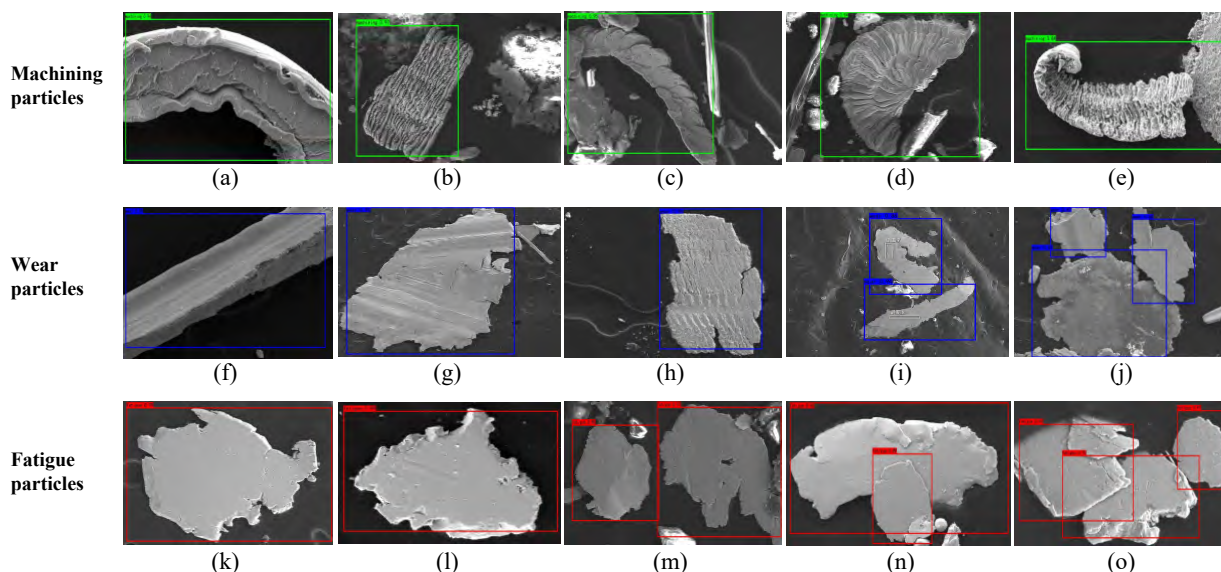


Fig. 10 Parts of wear particle images' detection results based on YOLOv4-VoVNetv2-39-BiFPN-DConv

图10 基于YOLOv4-VoVNetv2-39-BiFPN-DConv得到的部分磨粒图像检测结果

a. 针对现有磨粒图像智能识别算法识别精度往往受主观因素影响很大,且识别过程的自动化和智能化水平很低的问题,提出1种基于改进YOLOv4的目标检测算法,并应用于航空发动机扫描电镜磨粒图像目标检测,实现磨粒的自动快速识别。

b. 与原始YOLOv4网络相比,本文中所提出的新模型在3类磨粒组成的扫描电镜图像数据集上具有最佳的检测结果,即,在保证精度的前提下,网络参数量得到大幅降低,推理速度提升了51.1%,平均每张图片的检测时间节省了0.011 05s,同时对于重叠磨粒也有很好的检测效果。充分表明该模型在精度与速度之间保持了良好的平衡,满足实际扫描电镜磨粒图像快速、简洁和高精度的检测需求,具备很好的潜在工程应用价值。

参考文献

- [1] Chen Guo, Li Ai. Introduction to aircraft detection and diagnosis technology[M]. Beijing: Aviation industry press, 2012 (in Chinese) [陈果, 李爱. 航空器检测与诊断技术导论[M]. 北京: 航空工业出版社, 2012].
- [2] Chen Guo, Zuo Hongfu. Classing parameter's study of lubricating oil metal debris[J]. Acta Aeronautica et Astronautica Sinica, 2002, 23(3): 279-281 (in Chinese). [陈果, 左洪福. 润滑油金属磨粒的分类参数研究[J]. 航空学报, 2002, 23(3): 279-281]. doi: 10.3321/j.issn:1000-6893.2002.03.023.
- [3] Chen Guo, Zuo Hongfu. Two new techniques of image thresholding[J]. Pattern Recognition and Artificial Intelligence, 2002, 15(4): 468-473 (in Chinese). [陈果, 左洪福. 图像阈值分割的两种新技术[J]. 模式识别与人工智能, 2002, 15(4): 468-473]. doi: 10.3969/j.issn.1003-6059.2002.04.016.
- [4] Wang S, Wu T H, Shao T, et al. Integrated model of BP neural network and CNN algorithm for automatic wear debris classification[J]. Wear, 2019, 426-427: 1761-1770.
- [5] Peng Yeping, Cai Junhao, Wu Tonghai, et al. A hybrid convolutional neural network for intelligent wear particle classification[J]. Tribology International, 2019, 138: 166-173. doi: 10.1016/j.triboint.2019.05.029.
- [6] Fan Hongwei, Gao Shuoqi, Zhang Xuhui, et al. Intelligent recognition of ferrographic images combining optimal CNN with transfer learning introducing virtual images[J]. IEEE Access, 2020,

- 8: 137074–137093. doi: [10.1109/ACCESS.2020.3011728](https://doi.org/10.1109/ACCESS.2020.3011728).
- [7] Zhang Taohong, Hu Junnan, Fan Suli, et al. CDCNN: a model based on class center vectors and distance comparison for wear particle recognition[J]. *IEEE Access*, 2020, 8: 113262–113270. doi: [10.1109/ACCESS.2020.3002961](https://doi.org/10.1109/ACCESS.2020.3002961).
- [8] Fan Suli, Zhang Taohong, Guo Xuxu, et al. FFWR-Net: a feature fusion wear particle recognition network for wear particle classification[J]. *Journal of Mechanical Science and Technology*, 2021, 35(4): 1699–1710. doi: [10.1007/s12206-021-0333-6](https://doi.org/10.1007/s12206-021-0333-6).
- [9] Zhao Chunhua, Li Qian, Hu Hengxing, et al. Research of abrasive particle identification based on transfer learning neural network with a new joint loss function[J]. *Lubrication Engineering*, 2021, 46(4): 26–31 (in Chinese). [赵春华, 李谦, 胡恒星, 等. 一种新联合损失函数优化的迁移学习神经网络磨粒识别研究[J]. *润滑与密封*, 2021, 46(4): 26–31]. doi: [10.3969/j.issn.0254-0150.2021.04.005](https://doi.org/10.3969/j.issn.0254-0150.2021.04.005).
- [10] Guan Haojian, He Shizhong, Li Qiuqiu, et al. A review of convolutional neural networks in equipment wear particle recognition[J]. *Tribology*, 2022, 42(2): 426–445 (in Chinese). [关浩坚, 贺石中, 李秋秋, 等. 卷积神经网络在装备磨损颗粒识别中的研究综述[J]. *摩擦学学报*, 2022, 42(2): 426–445]. doi: [10.16078/j.tribology.2021025](https://doi.org/10.16078/j.tribology.2021025).
- [11] Lee Y, Park J. CenterMask: real-time anchor-free instance segmentation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020. doi: [10.48550/arXiv.1911.06667](https://doi.org/10.48550/arXiv.1911.06667).
- [12] Tan M X, Pang R M, Le Q V. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.
- [13] Yan Yuan. Analysis of defects in the production process of color filter and its equipment application[J]. *Science & Technology Information*, 2018, 16(23): 110–111 (in Chinese). [颜媛. 浅谈彩色滤光片生产制程不良解析及设备应用[J]. *科技资讯*, 2018, 16(23): 110–111]. doi: [10.16661/j.cnki.1672-3791.2018.23.110](https://doi.org/10.16661/j.cnki.1672-3791.2018.23.110).
- [14] Fan Hongwei, Hu Deshun, Gao Shuoqi, et al. An intelligent recognition method for ferrographic images based on deep belief network[P]. CN112287813A, 2021 (in Chinese) [樊红卫, 胡德顺, 高烁琪, 等. 一种基于深度信念网络的铁谱图像智能识别方法[P]. CN112287813A, 2021]. doi: [10.3969/j.issn.0254-0150.2021.07.003](https://doi.org/10.3969/j.issn.0254-0150.2021.07.003).
- [15] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2014: 580–587. doi: [10.1109/CVPR.2014.81](https://doi.org/10.1109/CVPR.2014.81).
- [16] Girshick R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision. IEEE, 2015: 1440–1448.
- [17] Ren Shaoqing, He Kaiming, 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. doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
- [18] He Kaiming, Gkioxari G, Dollár P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision. IEEE, 2017: 2980–2988.
- [19] 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. IEEE, 2016: 779–788.
- [20] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 6517–6525.
- [21] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. 2018, arXiv: 1804.02767 [cs. CV].
- [22] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[J]. 2020, arXiv: 2004.10934 [cs. CV]. doi: [10.48550/arXiv.2004.10934](https://doi.org/10.48550/arXiv.2004.10934).
- [23] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904–1916. doi: [10.1109/TPAMI.2015.2389824](https://doi.org/10.1109/TPAMI.2015.2389824).
- [24] Huang Gao, Liu Zhuang, van der Maaten L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 2261–2269.
- [25] Lee Y, Hwang J W, Lee S, et al. An energy and GPU-computation efficient backbone network for real-time object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2019: 752–760.
- [26] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[J]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 936–944. doi: [10.1109/CVPR.2017.106](https://doi.org/10.1109/CVPR.2017.106).
- [27] Liu Shu, Qi Lu, Qin Haifang, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 8759–8768. doi: [10.1109/CVPR.2018.00913](https://doi.org/10.1109/CVPR.2018.00913).
- [28] Kaiser L, Gomez A N, Chollet F. Depthwise separable convolutions for neural machine translation[J]. 2017, arXiv: 1706.03059 [cs. CL]. doi: [10.48550/arXiv.1706.03059](https://doi.org/10.48550/arXiv.1706.03059).