

文章编号 1004-924X(2024)07-1075-12

多尺度增强特征融合的钢表面缺陷目标检测

林珊玲^{1,2}, 彭雪玲^{1,2}, 王 栋^{1,2}, 林志贤^{1,2,3}, 林坚普^{1,2*}, 郭太良^{2,3}

- (1. 福州大学 先进制造学院, 福建 泉州 362252;
2. 中国福建光电信息科学与技术实验室, 福建 福州 350116;
3. 福州大学 物理与信息工程学院, 福建 福州 350116)

摘要:针对轻量级目标检测算法在钢表面缺陷检测任务中识别精度低的问题,提出一种多尺度增强特征融合的钢表面缺陷目标检测算法。该算法采用提出的自适应加权融合模块为不同层级特征自适应计算融合权重,将深层语义与浅层细节进行加权融合,使得浅层特征在不丢失细节信息的同时获得丰富的深层语义。利用提出的空间特征增强模块从3个独立方向强化融合特征,通过引出残差旁路增强网络结构的稳定性,使卷积过程能够挖掘到更多的关键信息。根据先验框与真实框的整体交并程度为模型选择更为合适的训练样本。实验结果表明,该算法的检测精度达到80.47%,相比原始算法提升6.81%。该算法的参数量为2.36 M,计算量为952.67 MFLOPs,能快速且高精度检测钢材表面的缺陷信息,具有较高的应用价值。

关键词:缺陷检测;单发多框检测器;增强特征融合;自适应加权融合;空间特征增强

中图分类号:TP394.1;TH691.9 **文献标识码:**A **doi:**10.37188/OPE.20243207.1075

Object detection of steel surface defect based on multi-scale enhanced feature fusion

LIN Shanling^{1,2}, PENG Xueling^{1,2}, WANG Dong^{1,2}, LIN Zhixian^{1,2,3},
LIN Jianpu^{1,2*}, GUO Tailiang^{2,3}

- (1. School of Advanced Manufacturing, Fuzhou University, Quanzhou 362252, China;
2. China Fujian Photoelectric Information Science and Technology Innovation Laboratory,
Fuzhou 350116, China;
3. School of Physics and Information Engineering, Fuzhou University, Fuzhou 350116, China)
* Corresponding author, E-mail: ljp@fzu.edu.cn

Abstract: To address the issue of low recognition accuracy in lightweight algorithms for steel surface defect detection, this paper introduces a Multi-scale Enhanced Feature Fusion (EFF) technique. Initially, an Adaptive Weighted Fusion (AWF) module calculates fusion weights adaptively for different feature levels. This allows shallow features to enrich with deep semantics without compromising detail. Subsequently, the Spatial Feature Enhancement (SFE) module boosts the fused features from three distinct directions and

收稿日期:2023-10-24;修订日期:2023-12-01.

基金项目:国家重点研发计划资助项目(No. 2021YFB3600603, No. 2022YFB3603705);福建省自然科学基金资助项目(No. 2020J01468);国家自然科学基金青年基金资助项目(No. 62101132);国家重点研发计划资助(2023YFB3609400)

improves network stability by integrating residual pathways, enabling the convolution process to extract more critical information. The model then selects better training samples based on the overlap between the prior box and the ground truth. Experimental outcomes show that the proposed method achieves a detection accuracy of 80.47%, marking a 6.81% increase over the baseline algorithm. Moreover, with 2.36 M parameters and 952.67 MFLOPs, this algorithm efficiently and accurately identifies steel surface defects, demonstrating significant practical utility.

Key words: defect detection; Single Shot multibox Detector (SSD); enhanced feature fusion; adaptive weighted fusion; spatial feature enhancement

1 引言

钢坯在轧制阶段由于表面受热不均、轧辊夹杂碎屑等原因,部分区域存在孔洞、冷热裂纹等缺陷,影响铸造成品的抗拉强度、硬度和韧性。早期的钢表面缺陷主要依靠人工查验^[1],通过主观视觉判定钢材损伤位置及受损程度,效率不高并且经常发生漏检。频闪光检测技术虽然能够从一定程度上提高钢表面缺陷的识别准确率,但大量重复性工作导致其检测速度较慢,无法满足工业检测需求^[2]。

近年来,卷积神经网络(Convolutional Neural Network, CNN)发展迅速,基于CNN^[3]的目标检测算法的识别精度不断提高。现阶段的目标检测算法大致分为两类,一类是基于区域候选提议(Region Proposal Network, RPN)的算法,如R-CNN(Region-based CNN)^[4], Faster R-CNN^[5]等,通过生成候选框提升检测精度,需要消耗大量检测时间。另一类是基于回归的算法,如YOLO(You Only Look Once)^[6], SSD(Single Shot multibox Detector)^[7],在回归目标位置的同时直接预测类别。其中,SSD算法采用VGG-16^[8]作为骨干网络进行多尺度分治检测,模型体积与计算量较大。为了降低卷积网络的复杂度,Sandler等^[9]提出MobileNetV2,采用深度可分离卷积进行特征提取。在增强特征表达方面,Lin等设计了特征金字塔(Feature Pyramid Network, FPN)^[10]结构,通过横向连接与自顶向下融合方式,加强了深层语义与浅层细节的交互。Li等受到FPN结构的启发提出FSSD(Feature Fusion Single Shot multibox Detector)^[11],将深层特征上采

样后直接与浅层特征拼接融合,为浅层特征融入了深层语义表示。Hu等^[12]则提出SE(Squeeze and Excitation)注意力机制,通过通道方向的特征挤压与激励,使模型更加关注图像细节。与此同时,许多专家学者开展了钢表面缺陷目标检测算法的相关研究,特征融合结构、注意力机制等成为重要的算法优化手段。He等^[13]提出一种基于深度学习的钢板缺陷检测系统,通过组合池化层、普通卷积、反卷积形成全段多尺度特征融合结构,促进了杂质、刮痕等钢表面缺陷的细粒度表达与深层语义融合。Tang等^[14]提出一种带钢表面缺陷检测方法,采用带注意力机制的残差模块结合多尺度池化结构,解决了由于网络感受野过于单一导致的无法感知补丁、结疤等缺陷尺度变化问题。此外,Guo等^[15]提出一种检测钢表面缺陷的改进算法,结合视觉Transformer检测模块辅助全局特征提取,增强了算法对于裂纹、点蚀等覆盖范围较广的缺陷目标的识别能力。虽然上述方法都能提升钢表面缺陷的检测效果,但识别精度仍具有提升空间,并且网络模型的计算成本较高,实际应用于工业检测设备具有一定难度。

针对钢表面缺陷目标检测算法的识别准确率不高、模型计算量大的问题,本文在SSD-MV2算法的基础上提出一种多尺度增强特征融合的钢表面缺陷目标检测算法(Enhanced FSSD with MobileNetV2, EFSSD-MV2)。首先,设计了一种自适应加权融合模块,采用自适应权重指导不同尺度层级的特征拼接融合。其次,设计了一种空间特征增强模块,进一步加强融合特征的有效信息描述。接着,改进了训练样本选择方法,使

模型根据先验框与真实框的整体交并趋势动态调整训练样本数量,提高损失函数计算的准确性。

2 原理

EFSSD-MV2算法以SSD300算法^[16]框架为基础,采用更加轻量的MobileNetV2代替VGG16作为特征提取骨干网络,并移除它在Conv18层后的分类输出层,算法网络结构如图1所示。为了将深层语义表达快速传导至浅层,选择其中的Conv7,Conv14,Conv18层进行类似FSSD中的多尺度特征融合,并通过增强特征融合(Enhanced Feature Fusion, EFF)单元提升融合效果,包含自适应加权融合(Adaptive Weight-

ed Fusion, AWF)模块与空间特征增强模块(Spatial Feature Enhancement, SFE)。AWF模块用于计算不同层级特征的自适应融合权重,而SFE模块则用于进一步提高融合特征的空间特性。之后,利用附加特征提取网络生成6个不同尺度的特征图,将其特征通道分布由原始的512,1 024,512,256,256,256调整为256,128,128,64,64,32,并采用深度可分离卷积代替普通卷积作为分类与回归预测器,目的是降低网络后端的计算成本。改进的训练样本选择方法应用于模型训练阶段,通过将交并比(Intersection Over Union, IOU)阈值关联先验框与真实框的整体交并程度,使模型获得更为合适的训练样本。最后,采用非极大抑制算法筛选目标框。

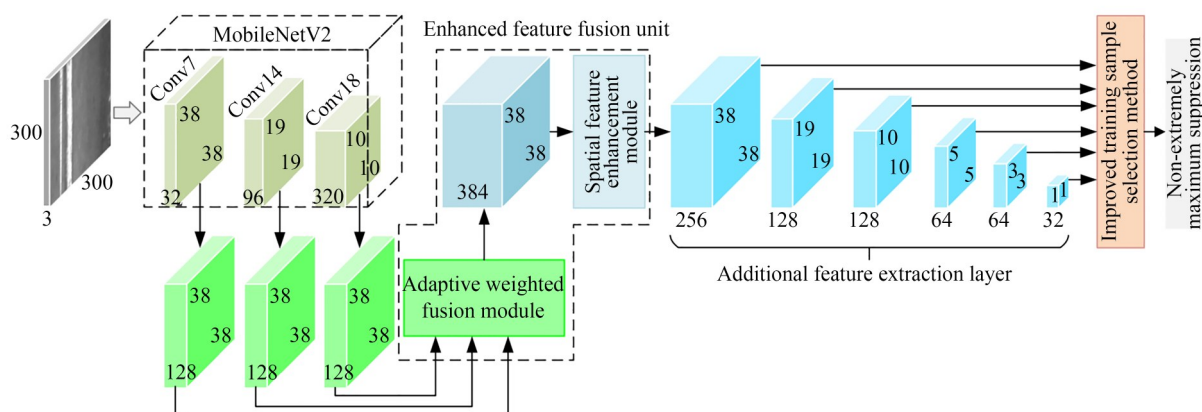


图1 EFSSD-MV2算法的网络结构

Fig. 1 Network structure of EFSSD-MV2 algorithm

2.1 自适应加权融合模块

基础特征融合结构一次性聚合多个不同尺度的特征图^[17]。大尺度特征图包含众多几何细节,而小尺度特征图则包含更为强大的语义表达。这种融合方式的本质是将深层特征直接传递至浅层,使浅层特征在不丢失细节信息的同时获得丰富的深层语义。然而,它在融合前必须规范特征图尺度与通道,弱化了不同层级特征之间的相对差异,抑制了融合特征描述检测目标特异性的能力。加权特征融合是指通过计算每个特征图对检测目标的贡献程度,使得不同尺度的特征信息得到充分融合。为了弥补基础特征融合结构的不足,设计了AWF模块,其网络结构和算

法流程框图如图2(a)和2(b)所示。

将AWF模块的输入特征记为 $X_n \in \mathbb{R}^{H_n \times W_n \times C_n}$, $n \in \{1, 2, 3\}$ 表示特征层顺序; $H_n \times W_n$ 表示输入特征的尺寸,本文算法中依次设置为 38×38 , 19×19 , 10×10 ; C_n 表示通道数,分别为32, 96, 320。原始特征融合分支采用 1×1 的卷积核将所有输入特征的通道数统一为 C , C 设置为128;同时,将输入特征 X_2 与 X_3 的尺寸分别通过双线性插值上采样为 $H_1 \times W_1$,输入特征在规范后记为 $\bar{X}_n \in \mathbb{R}^{H_1 \times W_1 \times C}$ 。AWF模块分支首先采用3个全局平均池化(Global Average Pooling, GAP),分别获取输入特征在通道方向上的全局表示,并按照由浅至深的层级顺序合并得到全局融合特征 $Z \in$

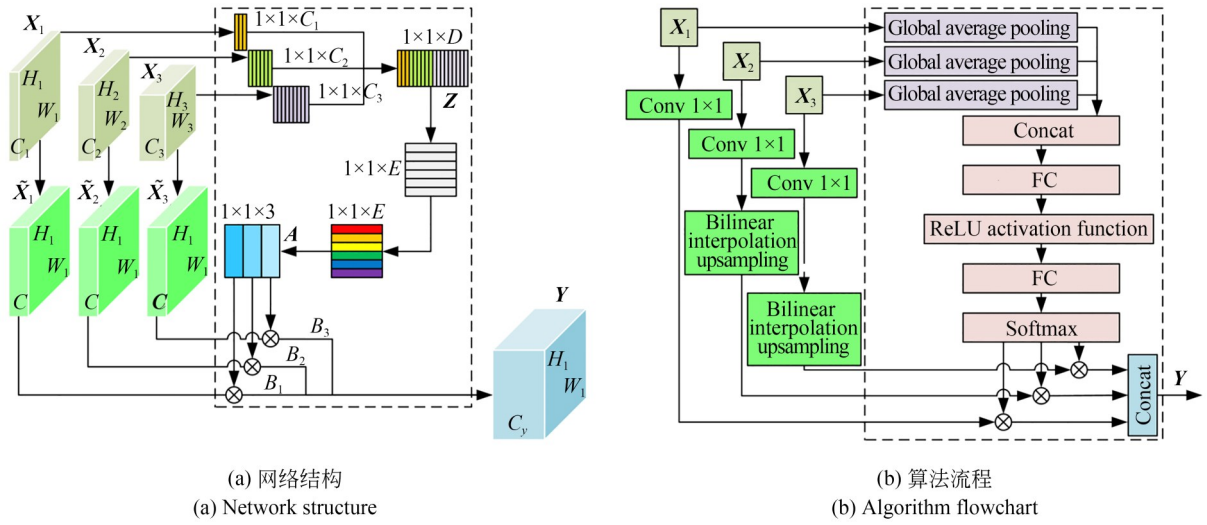


图2 自适应加权融合模块

Fig. 2 Adaptive Weighted Fusion(AWF) module

$\mathbf{R}^{1 \times 1 \times D}$, 其表达式为:

$$Z = \bigoplus_{n=1}^3 z_n = \bigoplus_{n=1}^3 \left\{ \frac{1}{H_n \times W_n} \sum_{i=1}^{H_n} \sum_{j=1}^{W_n} X_n(i, j) \right\}, \quad (1)$$

其中 D 为 C_n 之和, 其值为 448; $\bigoplus$ 表示 Concat 特征拼接, z_n 表示第 n 层输入的全局特征, (i, j) 表示特征空间坐标。采用 2 个全连接层 (Fully Connected, FC) 与 1 个 ReLU 激活函数组合形成瓶颈层, 用于学习特征在通道方向上的全局非线性关系, 从而自适应地生成一组映射特征 $A \in \mathbf{R}^{1 \times 1 \times 3}$, 其表达式为:

$$A = W_2(\text{ReLU}(W_1(Z))), \quad (2)$$

其中: $W_1 \in \mathbf{R}^{E \times D}$ 表示全局特征通过第一个 FC 层后降至 E 个通道的参数, 为了使前后 2 次通道降维更加平缓, E 设置为 32; $W_2 \in \mathbf{R}^{3 \times E}$ 表示特征通过第二个 FC 层后降至 3 个通道的参数。为了更加准确地描述映射特征之间的相对性, 采用 Softmax 函数对 A 的特征值进行归一化指数约束, 该过程可表示为:

$$b_n = \frac{\exp(a_n)}{\sum_{n=1}^3 \exp(a_n)}, \quad (3)$$

其中: a_n 表示 A 中的第 n 个特征值, 对应的权重值为 b_n , 代表特征 $\bar{X}_n$ 对融合结果的影响程度。最后, 将权重值分别与特征层对应相乘, 输出加权融合特征 $Y \in \mathbf{R}^{H_1 \times W_1 \times C_y}$, 其表达式为:

$$Y = \bigoplus_{n=1}^3 B_n = \bigoplus_{n=1}^3 \{b_n \times \bar{X}_n\}, \quad (4)$$

其中: C_y 为 C 的 3 倍, 其值为 384; B_n 表示第 n 层规范化输入特征经过加权后的结果。AWF 模块以特征融合分支为基础, 按照输入特征的全局相对关系, 在通道方向上以极少的参数量分配融合权重。

2.2 空间特征增强模块

为了进一步提取融合特征中包含的上下文先验知识, 同时消除由于输入特征尺度差异而产生的混叠效应, 基础特征融合结构对融合特征应用一个 3×3 的卷积核, 通道数降至原始特征的三分之二。该方法虽然在一定程度上能提高特征的全局表示能力, 但通道数减少会产生特征丢失, 且过于单薄的特征联系会影响附加特征提取网络对于检测目标的分类与定位。因此, 本文结合空间注意力机制 (Spatial Attention Mechanism, SAM) 设计了如图 3 所示的 SFE 模块, 从 3 个不同方向强化融合特征的空间表达, 通过引出残差旁路增强网络结构的稳定性, 使卷积过程能够挖掘到更多的关键信息。

虽然 3×3 的卷积核在提取特征时能够扩大感受野, 但其处理大尺度特征层的计算成本较高, 因此, SFE 模块中所有卷积操作均采用 1×1 的卷积核。SFE 模块首先对加权融合特征 Y 采用一组并行层进行类似 SAM 中的挤压与激励操

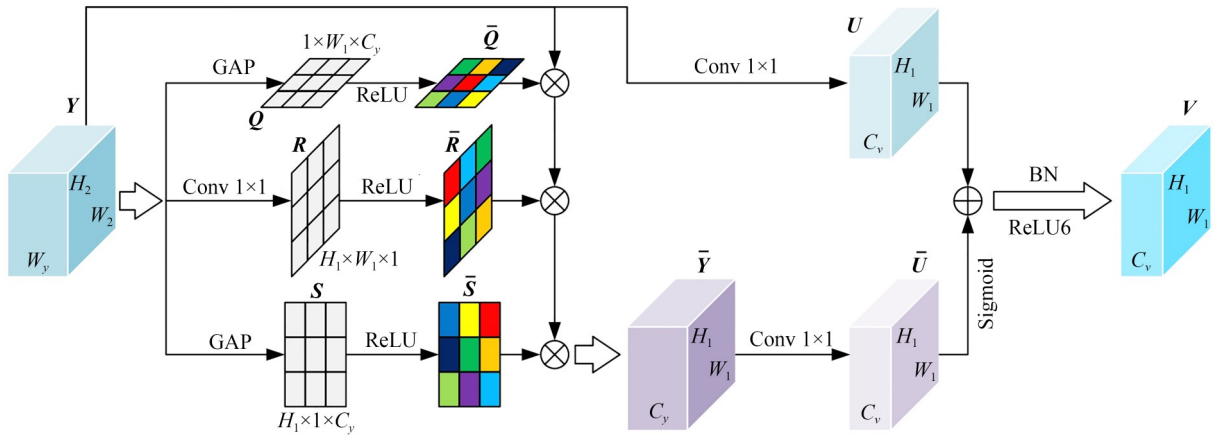


图3 空间特征增强模块

Fig. 3 Spatial Feature Enhancement(SFE) module

作。通过2个并行的GAP编码生成全局特征序列,同时采用1个卷积层提取通道方向上的依赖关系,从而得到融合特征在不同空间取向的独立表示。输出特征的表达式为:

$$\begin{cases} F_q: Y \rightarrow Q \in \mathbb{R}^{1 \times W_1 \times C_y} \\ F_r: Y \rightarrow R \in \mathbb{R}^{H_1 \times W_1 \times 1} \\ F_s: Y \rightarrow S \in \mathbb{R}^{H_1 \times 1 \times C_y} \end{cases} \quad (5)$$

对全局特征 Q, R, S 应用 ReLU 激活函数,目的是增强特征描述的多样性。该过程输出的表达式为:

$$\begin{aligned} \bar{Q} &= \text{ReLU}(Q) \\ \bar{R} &= \text{ReLU}(R), \\ \bar{S} &= \text{ReLU}(S) \end{aligned} \quad (6)$$

其中:输出映射矩阵 $\bar{Q} \in \mathbb{R}^{1 \times W_1 \times C_y}$, $\bar{R} \in \mathbb{R}^{H_1 \times W_1 \times 1}$, $\bar{S} \in \mathbb{R}^{H_1 \times 1 \times C_y}$ 。将映射矩阵与融合特征 Y 按顺序进行逐级矩阵乘法,从3个方向快速强化融合特征的有效空间信息,弱化没有实际表达意义的冗余特征。输出的空间增强特征 $\bar{Y} \in \mathbb{R}^{H_1 \times W_1 \times C_y}$ 中的特征矩阵具有显著的差异性,其表达式为:

$$\bar{Y} = (\bar{S} \cdot (\bar{R} \cdot (\bar{Q} \cdot Y))). \quad (7)$$

为了使 SFE 模块的输出特征通道数与附加特征提取网络的输入通道数保持一致,同时最大程度地保留原始融合特征信息,对融合特征 Y 与空间增强特征 $\bar{Y}$ 分别进行卷积特征提取,输出特征的表达式为:

$$\begin{cases} F_u: Y \rightarrow U \in \mathbb{R}^{H_1 \times W_1 \times C_u} \\ F_v: Y \rightarrow \bar{U} \in \mathbb{R}^{H_1 \times W_1 \times C_v} \end{cases} \quad (8)$$

其中 U 和 $\bar{U}$ 为输出特征,并且 U 的通道数 C_u 小于

$\bar{U}$ 通道数 C_y, C_v 设置为 256。最后,将二者进行类似残差结构中的矩阵叠加,输出特征 $V \in \mathbb{R}^{H_1 \times W_1 \times C_v}$ 的表达式为:

$$V = \text{ReLU6}(\text{BN}(U + \sigma(\bar{U}))), \quad (9)$$

其中: σ 表示 Sigmoid 激活函数,将其应用于特征 $\bar{U}$ 可得到具有约束特性的空间增强系数;批量归一化函数 (Batch Normalization, BN) 用于增强网络模型的稳定性,使模型训练不易发散;ReLU6 激活函数用来进行全局非线性化,相比 ReLU 激活函数能够抑制无效特征,提高特征表达信息的准确性。

2.3 改进训练样本选择方法

在网络模型的训练阶段,SSD 算法按照交并比为每个先验框匹配一个真实框,若该值小于预设阈值则将此先验框判别为负样本。虽然该方法能够简单快速地区分先验框的正负属性,但6个不同尺度的检测层级先验框匹配程度并不相同,并且固定阈值的选取依赖于主观经验,训练阶段无法进行自适应调整。为了缓解这一问题,本文结合自适应阈值思想改进 SSD 算法的训练样本选择方法 (Adaptive Training Sample Selection, ATSS),改进方法具有更加强大的样本包容性。

将改进方法记为改进训练样本选择 (Improved Training Sample Selection, ITSS) 方法,它应用于每个检测层级的先验框与真实框的匹配过程,主要包含正样本选择与阈值计算两个部分。假定在第 l 个尺度检测层级的第 r 个先验框为 P_r ,其中心点坐标为 $C_l(x_1, y_1)$,第 s 个真实框 G_s

的中心点坐标为 $C_2(x_2, y_2)$, 则 C_1 与 C_2 的欧式距离 L_{rs} 可表示为:

$$L_{rs} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \quad (10)$$

对于每个尺度检测层级, 分别为每个真实框筛选前 k 个欧式距离 L 最小的先验框作为候选正样本, 通过超参数 k 可调整纳入候选正样本的先验框数量, 进而控制先验框与真实框的匹配程度。

在 ATSS 方法中, 通过遍历每个真实框所匹配的候选正样本来计算二者的 IOU 值, 并为每个真实框构建一个 IOU 集合; 而 ITSS 方法则更加关注交并比的总体关系, 将前述集合统一归并为 $I = \{I_1, I_2, \dots, I_m\}$, 总数为 m 。为集合 I 计算基本统计学特性可得到自适应 IOU 阈值 T , 其数学表达式为:

$$\begin{cases} M = \frac{\sum_{h=1}^m I_h}{m} \\ S = \sqrt{\frac{\sum_{h=1}^m (I_h - M)^2}{m - 1}} \\ T = M + S \end{cases}, \quad (11)$$

其中: 集合 I 的均值 M 代表候选正样本与真实框的平均交并程度, 而 I 的标准差 S 则用于描述候选正样本与真实框的交并比差异。在实际正负样本判定过程中, 若第 q 个先验框 P_q 与其匹配真实框的 IOU 值 $J_q \leq T$, 则判定 P_q 为负样本。通过 ITSS 方法, 模型能够综合不同检测层级的先验框与真实框的交并程度, 根据不同训练阶段的样本匹配程度动态调整 IOU 阈值, 使损失函数在计算过程中获得更加优质的正负样本。

3 实验结果分析

3.1 数据集与参数设置

为了检验 EFSSD-MV2 对于钢表面缺陷目标的检测性能, 实验采用东北大学公开发布的 NEU-DET 钢材缺陷公共数据集作为训练样本, 该数据集以 PASCAL VOC 数据集格式为标准, 包含 6 类常见的钢表面缺陷图像与标定数据, 分别为裂纹、补丁、杂质、点蚀、结疤和刮痕。其中,

裂纹为钢表面产生的开裂现象; 补丁为金属的某一部分因具有特定特征而与其部分区分开的区域; 杂质为金属表面缺陷中的典型缺陷, 有些夹杂物比较松散, 易脱落, 有的则压入板中; 点蚀是一种腐蚀形式, 集中在钢金属表面的一个非常小的范围内, 并渗入金属内部, 通常直径小但深度较大; 结疤是轧制过程中轧机鳞片被卷入金属中形成的; 刮痕是钢表面磨损的痕迹。在保证实验公平性的前提下, 本文去除了 4 张仅含有难标签的数据样本, 因此, 数据集共包含 1 796 张钢表面缺陷样本图片, 训练集与测试集的比例为 8:2。

所有实验均在戴尔 Precision 7920 高性能图形工作站上完成, 其主要配置为 Intel Xeon Silver 4210R 2.40 GHz CPU 以及 NVIDIA Quadro RTX 5000 GPU, 软件环境为 Python 3.8 与 Pytorch 1.11 深度学习框架, 配备 CUDA 11.6 图形加速平台。网络输入尺寸设为 300×300 , 先验框尺度保持在 (30, 60), (60, 111), (111, 162), (162, 213), (213, 264), (264, 315), 数量分别为 4, 6, 6, 6, 4, 4; 批量大小为 32, 最大训练轮数为 20 000 轮, 推理置信度阈值为 0.01, 检测准确率阈值为 0.7; 采用原始 SSD 算法的 SGD (Stochastic Gradient Descent) 梯度下降法优化训练梯度, 动量参数为 0.9, 权重衰减参数为 0.000 01; 为了使模型在训练过程中达到最优结果, 在学习率调整策略上采用余弦退火学习率策略, 使得网络模型能够逃离局部稳态, 重新寻找更为合适的极值点。学习率 α 的数学表达式为:

$$\alpha = \alpha_{\min} + \frac{1}{2} (\alpha_{\max} + \alpha_{\min}) \left(1 + \cos \left(\frac{T_{\text{cur}}}{T_{\text{max}}} \pi \right) \right), \quad (12)$$

其中: $\alpha_{\min}$ 表示学习率的最小值, 设置为 0; $\alpha_{\max}$ 表示学习率的最大值, 设置为 0.01; T_{cur} 表示当前运行步数, 随实验过程迭代更新; T_{max} 表示单个退火过程所需要的步数, 设置为 2。

3.2 评价指标与消融实验

为了验证所提出方法的性能, 采用所有类别精度均值 (Average Precision, AP), 平均精度均值 (mean Average Precision, mAP) 指标验证多

类别的检测精度,其计算公式如下:

$$P = \frac{TP}{TP + FP}, \tag{13}$$

$$R = \frac{TP}{TP + FN}, \tag{14}$$

$$AP = \int_0^1 P(R) dR, \tag{15}$$

$$mAP = \frac{\sum_{i=1}^K AP_i}{K}, \tag{16}$$

其中:TP表示真正例,FN表示假反例,FP表示假正例,TN表示真反例,K为类别数目。

本文主要采用目标检测任务中常用的平均精度均值指标衡量算法精度,通过帧率(frame per second, fps)、参数量以及计算量等指标量化模型的实时检测性能。以算法在每个类别上的召回率(Recall)为横轴坐标,准确率(Precision)为纵轴坐标绘制 Precision-Recall(*P-R*)曲线,通过计算 *P-R* 曲线与坐标轴围成的区域面积即可得到类别平均精度(Average Precision, AP)。将 SSD-MV2 算法、FSSD-MV2 算法以及 EFSSD-MV2 算法分别在 NEU-DET 数据集上进行实验,3 种算法的 *P-R* 曲线对比如图 4 所示。从图中可以看出,FSSD-MV2 算法的检测精度高于 SSD-MV2 算法,但由于仅采用基础特征融合结构,算法准确率的增加幅度有限;而 EFSSD-MV2 算法在召回率大于 0.2 之后的准确率则明显高于其他两种算法,这表明该算法对钢表面缺陷的检测精度相比改进前显著提升。

为了验证 EFSSD-MV2 算法与提出方法的有效性,本文以 SSD-MV2 算法模型为基础进行消融实验,结果如表 1 所示。其中,mAP_{ITSS} 表示采用 ITSS 方法后的模型平均精度均值,超参数

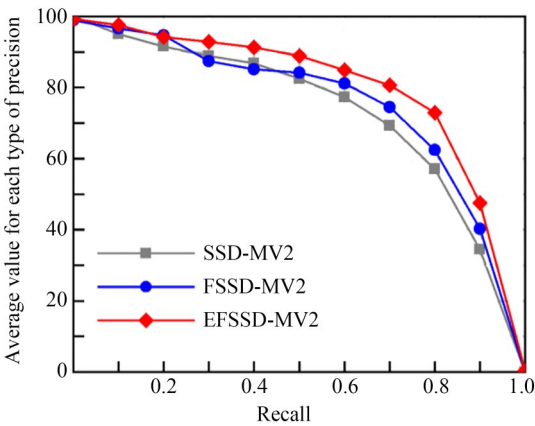


图 4 算法改进前后的检测精度对比

Fig. 4 Comparison of algorithm detection accuracy before and after improvement

$k=30$ 。结果表明,EFSSD-MV2 算法的 mAP 达到了 80.47%,相比 SSD-MV2 算法提升了 6.81%;此外,算法模型的参数量为 2.36 M,计算量为 952.67 MFLOPs,整体参数量与计算成本较低。在传统 SSD 算法的训练样本选择方法下,整体网络结构修改使算法的 mAP 提升了 4.19%;其中,基础特征融合结构使算法的 mAP 提升了 2.45%,计算量增加了 239.48 MFLOPs;EFF 单元使算法的 mAP 提升 1.74%,计算量增加了 144.51 MFLOPs。由此可知,EFF 单元能够有效提高算法的检测精度,计算成本比基础特征融合结构更低。以特征融合结构为基础,单独添加 AWF 模块后算法的 mAP 提升了 1.22%,模型参数量仅增加 0.02 M;单独增加 SFE 模块后算法的 mAP 提升了 1.07%,模型参数量增加 0.1 M。在 SSD-MV2 算法的基础上,将训练样本选择方法更改为 ITSS 方法,算法 mAP 提升了 3.75%。由此可知,AWF 模块以极小的计算成本提升了检测

表 1 EFSSD-MV2 算法的消融实验 (1 M=10⁶)

Tab. 1 Ablation experiments for EFSSD-MV2 algorithm(1 M=10⁶)

方 法	mAP/%	mAPITSS/%	参数量/M	计算量(MFLOPs)
SSD-MV2	73.66	77.41	1.92	568.68
FSSD-MV2	76.11	78.78	2.24	808.16
FSSD-MV2AWF	77.33	79.61	2.26	808.66
FSSD-MV2SFE	77.18	79.98	2.34	952.54
FSSD-MV2EFF	77.85	80.47	2.36	952.67

精度,其原因在于通过在通道方向上的全局特征计算自适应权重;采用SFE模块虽然略微增加计算开支,但也提高了检测精度;通过改进训练样本选择方法,算法的检测精度取得了显著提升。整体而言,本文提出的方法对于SSD-MV2算法的检测精度提升是十分有效的。

3.3 检测算法对比

为了进一步评价EFSSD-MV2算法对于钢表面缺陷目标的检测性能,本文选取目标检测领域10种主流的轻量级检测算法进行比较。实验过程中,所有算法均在前述数据集上训练,保持原始训练参数不变,对比实验结果如表2所示。结果表明,EFSSD-MV2算法在钢表面缺陷检测任务上的mAP指标超过了其他检测算法,对裂纹、点蚀缺陷的检测精度均为最高,说明算法对于检测难度较大的细粒度钢表面缺陷目标的识别具有明显优势。EFSSD-MV2算法相比输入图像尺寸相近的SSDLite算法和NanoDet算法,mAP分别高出6.44%和30.19%,体积分别减少了15.71 MB和1.71 MB,表明算法能够利用较

少的图像输入资源取得较好的检测效果。同时,EFSSD-MV2算法相比超轻量级目标检测算法FastestDet的mAP高出9.05%,相比新型轻量级目标检测算法MutualGuide高出3.75%;相比YOLO系列的经典轻量级检测算法YOLOv3-tiny, YOLOv4-tiny 分别高出12.63%和11.98%,体积则减少了23.87 MB和13.21 MB;相比同系列主流算法YOLOv5n以及YOLOv6n算法,mAP分别高出4.75%和4.92%;相比先进轻量级检测算法YOLOv7-tiny和YOLOX-nano,mAP分别高出5.6%和3.69%。EFSSD-MV2算法在检测精度上超越了同等级别下的YOLO系列算法,而模型体积除了YOLOX-nano之外均相比更小,表明该算法以较低的计算成本,取得了相对较高的钢表面缺陷检测精度。在实时检测速度方面,虽然EFSSD-MV2算法的fps指标处于中间水平,但66.4 frame/s的检测速度能够满足钢铁行业的日常检测需求。据此可以得出,EFSSD-MV2相较于主流轻量级目标检测算法,无论在对钢表面缺陷目标的检测精度还是在模型轻量化程度上都具有明显的优势。

表 2 与其他轻量级目标检测算法的实验结果对比

Tab. 2 Comparison of experimental results with other lightweight target detection algorithms

方 法	AP/%						mAP /%	模型大小 /MB	fps/ (frame•s ⁻¹)
	裂纹	补丁	杂质	点蚀	结疤	刮痕			
YOLOv3-tiny ^[18]	36.09	88.91	73.11	65.41	59.77	83.76	67.84	33.14	60.7
YOLOv4-tiny ^[19]	36.46	87.67	67.37	79.48	58.26	81.67	68.49	22.48	64.1
YOLOv5n ^[20]	38.40	90.70	80.90	85.50	65.60	93.20	75.72	14.25	91.8
YOLOv6n ^[21]	39.30	92.60	79.50	85.20	60.90	95.80	75.55	9.29	86.8
YOLOv7-tiny ^[22]	30.00	91.60	83.80	85.30	65.60	92.90	74.87	11.75	85.0
YOLOX-nano ^[23]	43.32	92.14	79.65	83.10	66.28	96.18	76.78	7.23	54.8
SSDLite ^[24]	42.51	92.93	74.85	84.84	68.22	80.85	74.03	24.98	68.0
NanoDet ^[25]	3.00	87.40	54.40	68.60	30.00	58.30	50.28	10.98	74.8
FastestDet ^[26]	39.70	86.80	64.10	79.40	75.40	83.10	71.42	1.08	97.7
MutualGuide ^[27]	51.03	86.72	81.73	84.78	66.61	89.47	76.72	10.96	68.6
EFSSD-MV2	55.13	91.94	79.54	88.24	74.57	93.42	80.47	9.27	66.4

3.4 超参数讨论

对于ITSS样本选择方法,超参数 k 的取值直接决定候选正样本的数量,需要通过实验验证其取值的合理性。这里以EFSSD-MV2算法的网络结构为基础,分别比较ITSS及ATSS方法中 k 对

算法mAP指标的影响,实验中 k 取值为5~50,间距为5。表3为实验结果,其中,mAP_{ATSS}表示采用ATSS方法后模型的平均精度均值。结果表明,超参数 k 会影响算法的检测精度,当超参数 k 为30时,采用上述两种方法的算法mAP指标分别达到

表 3 超参数 k 对 EFSSD-MV2 算法 mAP 指标的影响
Tab.3 Effect of hyperparameter k on mAP metrics of EFSSD-MV2 algorithm (%)

k	mAP _{ATSS}	mAP _{ITSS}
5	76.55	78.44
10	79.10	79.38
15	79.53	79.44
20	79.04	79.73
25	79.04	80.03
30	79.81	80.47
35	79.70	79.98
40	79.06	79.33
45	78.72	78.26
50	78.90	78.18

极大值。从不同 k 取值下的 mAP 变化趋势可以得出,当 k 取值较小时,由于正样本数量不足,样本选择方法的 mAP 提升效果不明显;当 k 取值较大时,由于正样本数量过多而导致方法的 mAP 提升效果回落;采用 ITSS 方法后算法的 mAP 的整体变化幅度相对较小,原因在于其计算过程考虑了先验框与真实框的整体 IOU 特性,样本筛选过程更加符合自然分布规律,模型经过训练后对于不同的钢表面缺陷拥有更强的泛化检测能力。

3.5 测试结果对比

本文通过 Grad-CAM (Gradient-weighted Class Activation Mapping) 绘制附加特征提取网络的权重可视化效果图,分别对 SSD-MV2 算法、FSSD-MV2 算法以及 EFSSD-MV2 算法进行实验,并为每个类别随机抽取 1 组实验结果,3 种算法的类激活映射输出对比如图 5 所示。从图中可以看出,SSD-MV2 算法各个缺陷类别的热力图分布都较为宽泛,对于裂纹、补丁以及杂质缺陷的特征敏感性较差;FSSD-MV2 算法虽然对于所有缺陷特征表达都具有识别能力,但其特征感知范围仍然存在局限;相比于其他两种算法,EFSSD-MV2 算法则具有更加完整全面的钢表面缺陷捕获能力,主要原因在于 EFF 单元中合理运用了特征融合与注意力机制,网络模型的特征提取更加符合人类视觉。

为了直观展示钢表面缺陷目标检测算法的有效性,本文分别对 3 种算法进行了样例测试,检测效果对比如图 6 所示。可以看出,SSD-MV2 算法对于所有缺陷目标均存在不同程度的漏检问题,检测效果不佳;FSSD-MV2 算法提高了裂纹、补丁以及杂质缺陷的检测质量,但对于刮痕缺陷的检测精度依然较低;而 EFSSD-MV2

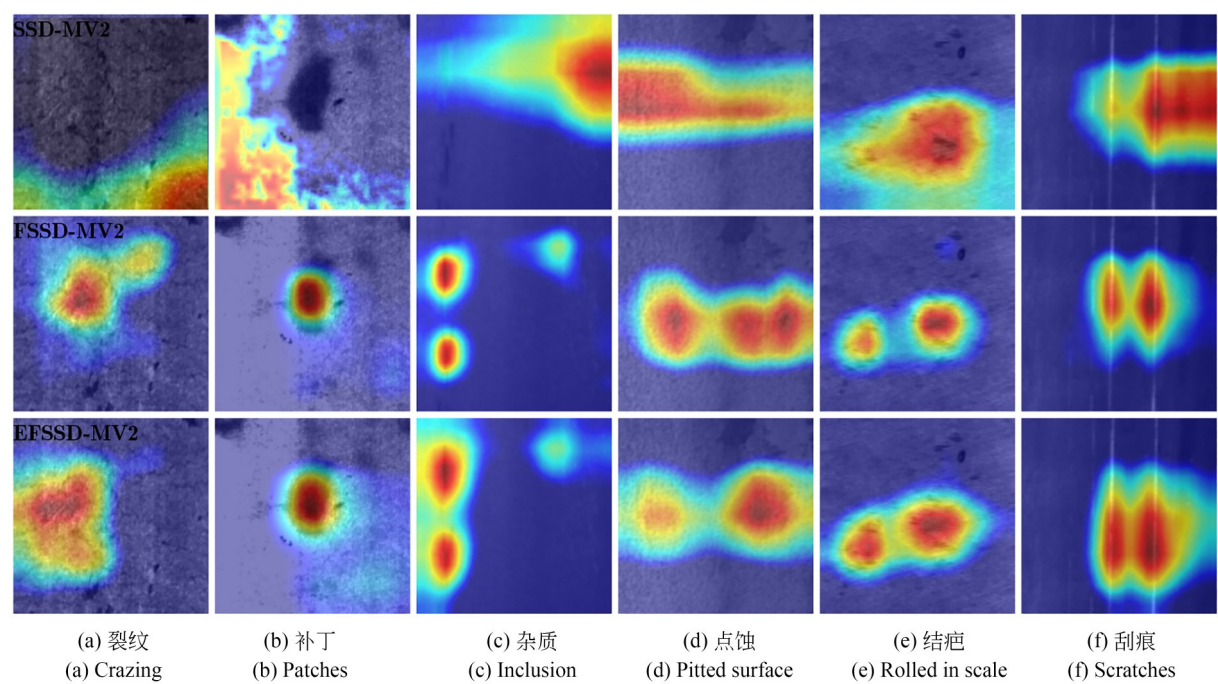


图 5 算法改进前后的类激活映射输出对比

Fig. 5 Comparison of class activation mapping output before and after algorithm improvement

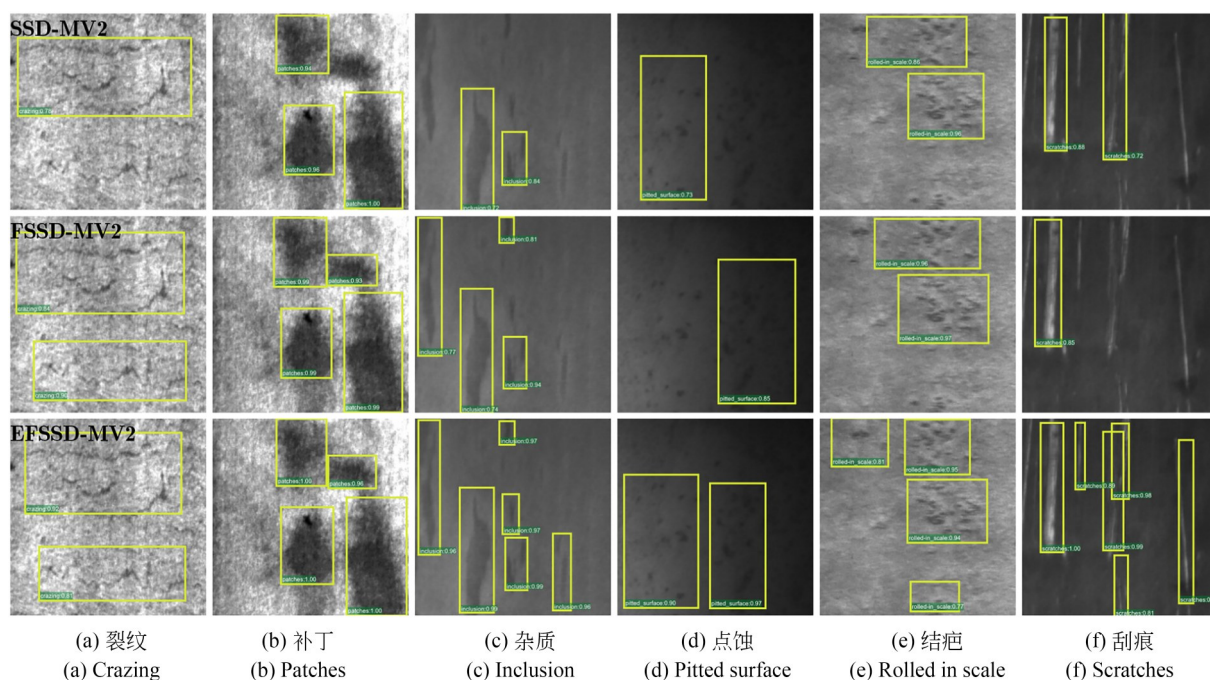


图6 算法改进前后样例的检测效果对比

Fig. 6 Comparison of sample detection results before and after algorithm improvement

算法能够较为准确地检测出6种常见的钢表面缺陷,相比于其他两种算法明显改善了缺陷漏检问题,对不同尺度的缺陷定位更加精准,如样例图中杂质与刮痕缺陷的算法测试结果对比;同时,该算法能够检测出更为细致的缺陷目标,如算法对小目标疤痕的有效定位。

4 结 论

本文以SSD检测模型为框架,提出一种轻量级钢表面缺陷目标检测算法EFSSD-MV2。首先,在基础特征融合结构之后加入EFF单元,其

中AWF模块为不同层级的特征自适应分配融合权重,SFE模块则重点关注融合特征在3个方向的全局表达。而后,通过将先验框与真实框的整体交并特性应用于IOU阈值的自适应计算,使模型在训练阶段能够获得更为合适的数据样本。最后,本文在公共数据集NEU-DET上进行了算法对比实验,并对影响实验结果的超参数进行分析。结果表明,EFSSD-MV2算法具有良好的钢表面缺陷检测性能,所提出的改进方法对算法检测精度的提升是十分有效的。下一步将开展高效特征融合结构的研究,从而以更少的计算开销提升算法的检测精度。

参考文献:

- [1] 李原,李燕君,刘进超,等. 基于改进Res-UNet网络的钢铁表面缺陷图像分割研究[J]. 电子与信息学报, 2022, 44(5): 1513-1520.
LI Y, LI Y J, LIU J CH, *et al.* Research on segmentation of steel surface defect images based on improved res-UNet network[J]. *Journal of Electronics & Information Technology*, 2022, 44(5): 1513-1520. (in Chinese)
- [2] LI Z G, WEI X M, HASSABALLAH M, *et al.* A

- deep learning model for steel surface defect detection [J]. *Complex & Intelligent Systems*, 2024, 10(1): 885-897.
- [3] SALEHI A W, KHAN S, GUPTA G, *et al.* A study of CNN and transfer learning in medical imaging: advantages, challenges, future scope[J]. *Sustainability*, 2023, 15(7): 5930-5958.
- [4] WANG Z, LI P, CUI Y C, *et al.* Automatic detection of individual trees in forests based on airborne LiDAR data with a tree region-based convolutional neural network (RCNN) [J]. *Remote Sensing*,

- 2023, 15(4): 1024-1044.
- [5] 邓姗姗, 黄慧, 马燕. 基于改进Faster R-CNN的小目标检测算法[J]. 计算机工程与科学, 2023, 45(5): 869-877.
- DENG SH SH, HUANG H, MA Y. A small object detection algorithm based on improved Faster R-CNN[J]. *Computer Engineering & Science*, 2023, 45(5): 869-877. (in Chinese)
- [6] 邓亚平, 李迎江. YOLO算法及其在自动驾驶场景中目标检测研究综述[J]. 计算机应用, 2023: 1-12.
- DENG Y P, LI Y J. A review of YOLO algorithm and research on target detection in autonomous driving scenarios [J]. *computer application*, 2023: 1-12. (in Chinese)
- [7] 汪宋, 费树岷. SSD(Single Shot MultiBox Detector)目标检测算法的研究与改进[J]. 工业控制计算机, 2019, 32(4): 103-105.
- WANG S, FEI SH M. Research and improvement of SSD(single shot MultiBox detector) target detection algorithm [J]. *Industrial Control Computer*, 2019, 32(4): 103-105. (in Chinese)
- [8] MINARNO A E, BAGAS S Y, YUDA M, *et al.* Convolutional Neural Network featuring VGG-16 Model for Glioma Classification[J]. *JOIV: International Journal on Informatics Visualization*, 2022, 6(3): 660-666.
- [9] SANDLER M, HOWARD A, ZHU M L, *et al.* MobileNetV2: inverted residuals and linear bottlenecks [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018. Salt Lake City, UT. IEEE, 2018: 4510-4520.
- [10] DENG C F, WANG M M, LIU L, *et al.* Extended feature pyramid network for small object detection [J]. *IEEE Transactions on Multimedia*, 2022, 24: 1968-1979.
- [11] 王国东, 唐金亮, 陈特欢, 等. 改进FSSD的差速器壳体表面缺陷实时检测算法[J]. 计算机工程与应用, 2022, 58(22): 262-270.
- WANG G D, TANG J L, CHEN T H, *et al.* Real-time detection algorithm for differential housing surface defects based on improved FSSD[J]. *Computer Engineering and Applications*, 2022, 58(22): 262-270. (in Chinese)
- [12] HU J, SHEN L, ALBANIE S, *et al.* Squeeze-and-excitation networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(8): 2011-2023.
- [13] HE Y, SONG K C, MENG Q G, *et al.* An end-to-end steel surface defect detection approach via fusing multiple hierarchical features [J]. *IEEE Transactions on Instrumentation Measurement*, 2020, 69(4): 1493-1504.
- [14] TANG M, LI Y Y, YAO W, *et al.* A strip steel surface defect detection method based on attention mechanism and multi-scale maxpooling [J]. *Measurement Science and Technology*, 2021, 32(11): 115401-115412.
- [15] GUO Z X, WANG C S, YANG G, *et al.* MSFT-YOLO: improved YOLOv5 based on transformer for detecting defects of steel surface [J]. *Sensors*, 2022, 22(9): 3467-3482.
- [16] LIU W, ANGUELOV D, ERHAN D, *et al.* SSD: single shot MultiBox detector[M]. *Computer Vision-ECCV 2016*. Cham: Springer International Publishing, 2016: 21-37.
- [17] 温剑, 邵剑飞, 刘杰, 等. 多维注意力机制与选择性特征融合的图像超分辨率重建[J]. 光学精密工程, 2023, 31(17): 2584-2597.
- WEN J, SHAO J F, LIU J, *et al.* Multidimensional attention mechanism and selective feature fusion for image super-resolution reconstruction [J]. *Opt. Precision Eng.*, 2023, 31(17): 2584-2597. (in Chinese)
- [18] GONG H, LI H, XU K, *et al.* Object detection based on improved YOLOv3-tiny [C]. 2019 *Chinese Automation Congress (CAC)*. November 22-24, 2019. Hangzhou, China. IEEE, 2019: 3240-3245.
- [19] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Scaled-YOLOv4: scaling cross stage partial network [C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 20-25, 2021. Nashville, TN, USA. IEEE, 2021: 13029-13038.
- [20] DING Z J, LIU C, LI D W, *et al.* Deep-sea biological detection method based on lightweight YOLOv5n[J]. *Sensors*, 2023, 23(20): 8600-8615.
- [21] LI Y, ZHANG X. Object detection for UAV images based on improved yolov6[J]. *IAENG International Journal of Computer Science*, 2023, 50(2): IJCS_50_2_45.
- [22] 齐向明, 董旭. 改进Yolov7-tiny的钢材表面缺陷检测算法[J]. 计算机工程与应用, 2023, 59

- (12): 176-183.
- QI X M, DONG X. Improved Yolov7-tiny algorithm for steel surface defect detection[J]. *Computer Engineering and Applications*, 2023, 59(12): 176-183. (in Chinese)
- [23] ZHENG Z Z, HU Y H, QIAO Y C, *et al.* Real-time detection of winter jujubes based on improved YOLOX-nano network [J]. *Remote Sensing*, 2022, 14(19): 4833-4852.
- [24] AN Y L, TANG J L, LI Y F. *A MobileNet SSD Lite Model with Improved FPN for Forest Fire Detection* [M]. Image and Graphics Technologies and Applications. Singapore: Springer Nature Singapore, 2022: 267-276.
- [25] MIAO D, WANG Y M, YANG L, *et al.* Foreign object detection method of conveyor belt based on improved nanodet[J]. *IEEE Access*, 2023, 11: 23046-23052.
- [26] ZHANG D P, YU X C, YANG L, *et al.* Data-augmented deep learning models for abnormal road manhole cover detection [J]. *Sensors*, 2023, 23(5): 2676-2694.
- [27] TAI X X, LI M J, XIANG M, *et al.* A mutual guide framework for training hyperspectral image classifiers with small data [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-17.

作者简介:



林珊玲(1991—),女,福建泉州人,博士,硕士生导师,2020年于福州大学获得博士学位,主要从事显示驱动、图像处理等方面的研究。E-mail: sl-lin@fzu.edu.cn

通讯作者:



林坚普(1989—),男,福建泉州人,博士,讲师,副教授,2019年于福州大学获得博士学位,主要从事图像处理、深度学习、液晶透镜器件、裸眼3D显示技术等方面的研究。E-mail: ljp@fzu.edu.cn