DOI：10.20271/j.cnki.1673-9833.2026.5011

EDMS：一种改进 YOLO11s 的钢铁表面缺陷检测算法

吴佳雨，张龙信，蒲忻怡，高　健

（湖南工业大学 计算机与人工智能学院，湖南 株洲 412007）

摘　要：针对传统钢铁表面缺陷检测模型检测精度较低的问题，提出了一种高效的钢铁表面缺陷检测模型（efficient detection model for steel-surfaces，EDMS）。首先，设计了一种多尺度特征提取网络（multi-scale feature extraction network，MFEN），其使用一种新颖的下采样模块，以在卷积下采样过程中保留更多的细节特征；同时引入高效的多尺度注意力机制，以增强模型对重要信息的捕获能力。其次，构建了多尺度特征融合网络，以实现跨层级特征信息的有效融合。此外，采用跨阶段部分融合的特征提取模块，以进一步提升模型的检测精度。在东北大学表面缺陷检测数据集上的实验结果表明，EDMS 的平均精度均值 *mAP* 达 79.07%，较基准模型 YOLO11s 的提高了 5.50%，与已有主流目标检测模型相比，其 *mAP* 值最高。

关键词：缺陷检测；卷积下采样；注意力机制；特征融合

中图分类号：TP317.4　　**文献标志码**：A　　**文章编号**：1673-9833(2026)05-0082-09

引文格式：吴佳雨，张龙信，蒲忻怡，等 . EDMS：一种改进 YOLO11s 的钢铁表面缺陷检测算法 [J]. 湖南工业大学学报，2026，40(5)：82-90.

EDMS：An Improved Steel Surface Defect Detection Algorithm for YOLO11s

WU Jiayu，ZHANG Longxin，PU Xinyi，GAO Jian

（School of Computer Science and Artificial Intelligence，Hunan University of Technology，Zhuzhou Hunan 412007，China）

Abstract：In view of the flaw of low accuracy in detecting surface defects of traditional steel, an efficient detection model for steel-surfaces (EDMS) has been proposed. Firstly, a multi-scale feature extraction network (MFEN) is designed, with a novel down-sampling module used to preserve more detailed features during the convolutional down-sampling process. Meanwhile, a high-efficient multi-scale attention mechanism is introduced for an enhancement of the model's ability to capture important information. Secondly, a multi-scale feature fusion network is constructed to achieve effective fusion of feature information across levels. Moreover, a cross-stage partial fusion feature extraction module is used to further improve the detection accuracy of the model. The experimental results on the Northeastern University surface defect detection dataset show that the mean average precision (mAP) of EDMS is 79.07%, which is 5.50% higher than that of the benchmark model YOLO11s, with its *mAP* value the highest compared with the existing mainstream object detection models.

Keywords：defect detection；convolutional downsampling；attention mechanism；feature fusion

收稿日期：2025-10-13

基金项目：湖南省自然科学基金资助项目（2023JJ50204）；湖南省教育厅科学研究基金资助项目（25A0422）；湖南省大学生创新训练基金资助项目（S202511535023）

作者简介：吴佳雨，男，湖南工业大学硕士生，主要研究方向为高性能计算，E-mail：jiayuwuchi@foxmail.com

通信作者：张龙信，男，湖南工业大学副教授，博士，硕士生导师，主要研究方向为深度学习与高性能计算，E-mail：longxin.zhang@163.com

1　研究背景

传统的钢铁表面缺陷检测通常依赖于人工肉眼检查，这种方法在实际应用中存在检测速度慢、人工成本高、检测精度低等诸多问题。近年来，深度学习技术迅速发展，被广泛应用于众多领域，并取得了显著成效[1]。相较于传统的检测方法，基于深度学习的目标检测方法展现出更强的鲁棒性、更快的处理速度，以及更精准的目标定位能力[1]。根据不同的处理步骤，可将基于深度学习的检测方法分为以 Faster R-CNN（faster region-based convolutional neural network[2]）为代表的两阶段检测方法和以 YOLO（You Only Look Once[3]）为代表的一阶段检测方法。与两阶段检测方法不同，一阶段检测方法直接通过神经网络预测目标的边界框，省略了生成候选区域这一步骤，减少了中间环节，从而显著提升了推理速度，可满足工业生产中钢铁表面缺陷实时性检测的要求。

Chen H. 等[4]通过引入基于卷积块注意力机制的下采样和上采样路径，提出了一种基于多尺度特征融合和注意力机制的 YOLOv5 模型。该模型能够更好地捕捉带钢表面的纹理信息，提高对复杂背景下中大型目标的检测能力。然而，由于该模型对于小型目标的检测性能欠佳，导致其整体检测精度偏低。Zhao X. Y. 等[5]设计了一种利用注意力机制的多特征图交互金字塔网络，提高了模型对于钢铁表面缺陷的自适应表征能力，强化了对于小型目标的检测能力，但对于复杂背景下的目标和大型目标的检测性能仍有待提高。Li Z. G. 等[6]提出了一种高效的特征融合模块，通过将骨干网络的特征添加至颈部网络以优化特征融合，增强了模型对重要特征的提取能力，在保持少量参数的同时实现了针对钢铁表面缺陷较为精确的检测。文献 [7] 通过在 YOLOv8 主干网络中引入轻量级的 MobileViTv2 网络，在确保模型对钢铁表面缺陷检测的精度不下降的前提下，实现了模型的轻量化。然而，尽管这些钢铁表面缺陷检测方法在模型轻量级优化方面取得了一定的进展，但在检测精度方面还存在着较大的提升空间。

2024 年 9 月 Ultralytics 团队发布了 YOLO11[8]。根据模型复杂度和参数数量可将 YOLO11 划分为 5 个版本，分别是 YOLO11-x、l、m、s、n。其中，YOLO11s 凭借轻量化架构，在实现快速处理的同时保障了较高的检测精度。在 YOLO11s 的基础上进行改进的模型，能够更好地满足钢铁表面缺陷检测中的精度要求。为此，针对当前钢铁表面缺陷检测模型对于多尺度缺陷检测效果不佳和整体检测精度不高等问题，本研究对 YOLO11s 模型进行了优化，设计了一种高效的钢铁表面缺陷检测模型（efficient detection model for steel-surfaces，EDMS）。

2　EDMS 模型建立

EDMS 整体框架结构见图 1，主要分为多尺度特征提取网络（multi-scale feature extraction network，MFEN）、多尺度特征融合网络（multi-scale feature fusion network，MFFN）和检测头（detect head）3 部分。

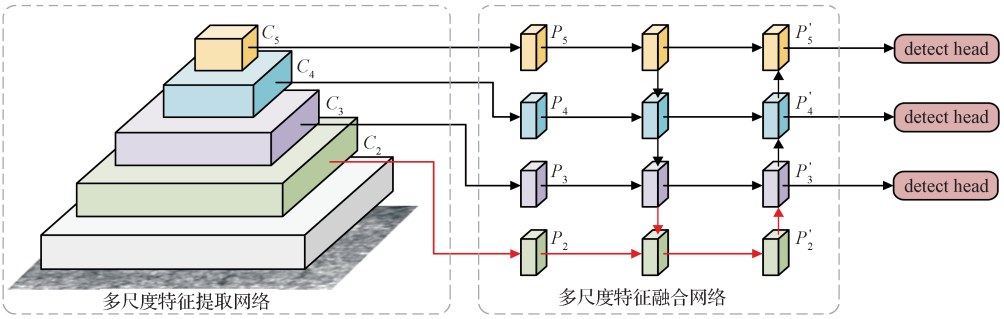


图 1　EDMS 整体框架结构图

Fig. 1　EDMS overall framework diagram

对于输入的待检测钢铁表面缺陷图片，EDMS 首先通过 MFEN 提取出不同尺度的特征图 C_2、C_3、C_4、C_5，其中特征图 $C_i=P_i$（$i \in \mathbb{R}$）。然后，通过 MFFN 融合来自 4 个不同尺度的特征图 P_2、P_3、P_4、P_5，得到聚合了浅层位置信息和深层语义信息的特征图 P'_2、P'_3、P'_4、P'_5。最后，利用检测头对特征图 P'_3、P'_4、P'_5 进行缺陷的分类和回归预测。

MFEN 中，EDMS 设计了深度－空间－注意力下采样模块（depthwise-space-attention downsampling，DSA-Down），在减少模型参数量的同时，有效提升其检测精度。随后，引入高效的多尺度注意力机制（efficient multi-scale attention，EMA[9]），增强模型对于特征图中关键信息的捕获能力。在 MFEN 中，EDMS 优化了 YOLO11s 的特征融合网络结构，如图 1 中红色箭头所示。这能够有效提升模型对于多尺度特征的提取能力。此外，EDMS 还设计了跨阶段部分

融合的特征提取模块（feature extraction module with cross-stage partial fusion，FEMCSPF），以实现模型压缩，在保持高精度检测的同时进一步缩减模型参数。

2.1 MFEN 结构

MFEN 的结构如图 2 所示，主要由卷积－批归一化－SiLU 激活函数 3×3 模块（convolution-batch normalization-sigmoid linear unit 3×3 module，CBS 3×3）、C3k2 模块[8]、快速空间金字塔池化模块（spatial pyramid pooling fast，SPPF）、跨阶段部分连接金字塔压缩注意力模块（cross stage partial with pyramid squeeze attention，C2PSA）、DSA-Down 下采样模块和 EMA 注意力机制构成。

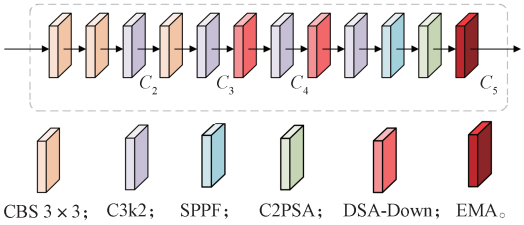


图 2 MFEN 结构示意图

Fig. 2 MFEN structural diagram

在 MFEN 中，CBS 3×3 模块和 DSA-Down 模块用于对特征图的下采样。C3k2 模块是对包含 3 个卷积层的跨阶段部分连接瓶颈模块（CSP bottleneck with 3 convolutions，C3）的改进，其既保持了原有 C3 模块的轻量级特点，也增强了网络对于多尺度特征的提取能力。SPPF 模块对传统特征金字塔模块进行了改良，使用多个小尺寸的 5×5 池化核代替原本的大尺寸池化核，这在降低模型参数量和计算量的同时，极大地提高了模型在应对多尺度目标任务时的检测能力。C2PSA 模块的核心思想是通过并行双分支结构，分别处理通道级特征和空间级特征，通过多尺度特征融合实现特征的多维强化，在提升模型对特征图关键语义信息捕捉能力的同时，增强特征提取模块在复杂环境下的抗干扰性能。

2.1.1 DSA-Down 模块

DSA-Down 下采样模块主要由空间到深度的卷积（space-to-depth convolution，SPDConv[10]）、卷积块注意力模块（convolutional block attention module，CBAM[11]）和深度卷积（depthwise convolution，DWConv[12]）3 部分构成，其结构如图 3 所示，图中⊕为矩阵逐元素相加。

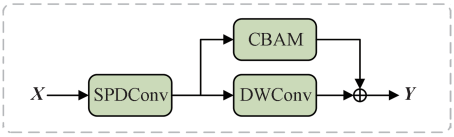


图 3 DSA-Down 模块结构示意图

Fig. 3 DSA-Down module structure diagram

如图 3 所示，首先，DSA-Down 利用 SPDConv 将特征图的每个空间维度降低到通道维度，在减小特征图空间尺寸的情况下保留更多的细粒度信息，同时减少基本卷积操作带来的计算开销。然后，使用 CBAM 模块，利用空间注意力机制和通道注意力机制增强模型对于关键特征的辨识能力，帮助模型聚焦特征图中的关键信息以提高模型的准确性。同时使用 DWConv 对下采样后的特征图做进一步处理，在保持高效提取特征图信息的同时，有效减少模型的参数量和计算量。最后，将经过 CBAM 注意力机制增强后的特征图和经过 DWConv 卷积操作后的特征图做矩阵逐元素相加得到输出特征图。

2.1.2 EMA 注意力机制

EMA 注意力机制能够在处理特征图信息时自动选择重要信息，同时过滤掉冗余信息。该机制能提升模型对关键信息的敏感度，从而增强其在复杂任务上的性能。EMA 注意力机制的基本结构如图 4 所示，图中⊗表示矩阵乘法。

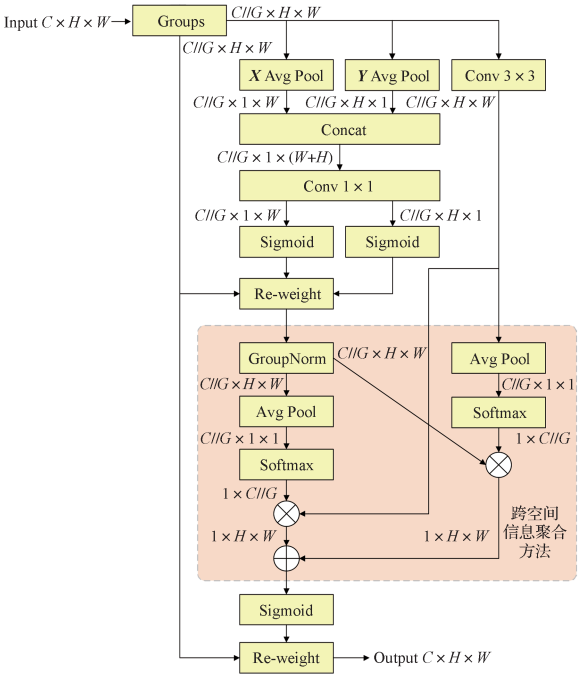


图 4 EMA 注意力机制结构示意图

Fig. 4 EMA attention structure diagram

对于输入特征图 $\boldsymbol{X}\in\mathbb{R}^{C\times H\times W}$，EMA 注意力机制将其划分为 $G(G\ll C)$ 个跨通道的子特征，其中 C、H 和 W 分别为输入特征图的通道数、高和宽。此模块对 G 组子特征进行了批次维度的变换，将输入特征图的形状重新定义为 $\boldsymbol{X}=[X_0,X_1,\cdots,X_{G-1}]$，$X_i\in\mathbb{R}^{C//G\times H\times W}$。

EMA 注意力机制采用 3 条平行的分支来提取分组特征的注意力权重，其中两条平行分支使用 1×1

卷积，另一条则使用 3×3 卷积。整体而言，在 1×1 支路中，两个一维平均池化操作分别应用于编码沿水平维度和沿垂直维度两个方向的空间注意力；而在 3×3 支路中，只有一个 3×3 的卷积用于捕获局部空间上下文信息。一方面，在 1×1 支路中，EMA 注意力机制首先将原始输入特征分别沿水平维度与垂直维度执行一维全局平均池化，得到两条互补的方向编码。然后，将这两条一维向量外积重构为二维注意力图，以此在完整空间维度上整合水平与垂直两个方向的上下文信息，从而捕捉不同位置之间的长程依赖关系。接着，通过逐像素乘法重新校准原始输入特征，使得来自浅层的细节特征能够得到充分保留和强化。另一方面，原始输入特征通过 3×3 支路中的 3×3 卷积提取局部空间信息，从而增强模型对局部空间关系的表征能力。

此外，EMA 中还采用了不同空间维度特征聚合的跨空间信息聚合方法。跨空间信息聚合方法依旧使用了两条分支，一条是 1×1 的分支，另一条为 3×3 的分支。在 1×1 的分支中，使用二维全局平均池化操作以编码其中的全局信息，可以表示为

$$z=\frac{1}{H\times W}\sum_{j}^{H}\sum_{i}^{W}x(i,j), \tag{1}$$

式中：$\boldsymbol{x}\left(\boldsymbol{x}\in\mathbb{R}^{C//G\times H\times W}\right)$为经分组归一化后的输出特征；$z$为通过二维全局平均池化操作后的输出特征图。

为了提高所得结果的有效性，EMA 使用非线性函数 Softmax 拟合线性变换，将 1×1 分支链路中经过 Softmax 处理后的结果与 3×3 分支链路中经过 3×3 卷积操作后的输入特征做矩阵乘法运算，得到了第一个空间注意力权重值。然后，利用类似方法，在 3×3 分支链路中推导出第二个空间注意力权重值。最后，将两个空间注意力权重值聚合，通过 Sigmoid 函数对其进行平滑处理后，与原始特征图相乘得到经过 EMA 注意力机制增强后的输出特征。

2.2　MFFN 架构

MFFN 的核心功能是对 MFEN 中提取出的特征图进行跨尺度融合与增强，其整体结构见图 5。

在卷积神经网络中，随着下采样的不断进行，特征图的分辨率逐渐降低，导致小目标的精确定位信息在深层网络中逐渐丢失。为解决这一问题，EDMS 模型对 YOLO11s 特征融合网络结构进行了改进。首先，通过优化特征金字塔结构增强模型对于细节特征的捕获能力，以有效缓解下采样过程中的信息丢失。其次，在深层网络中引入 FEMCSPF 模块，其采用跨阶段部分连接策略，能同时聚焦局部细节特征和全局上下文信息，使得模型在关注目标局部信息的同时，能高效整合来自上下文区域的多尺度特征，提高模型对于复杂背景中不同尺度目标的感知能力。

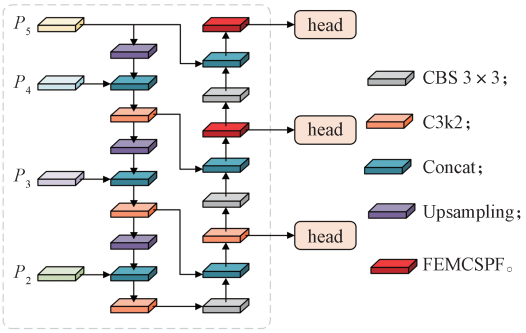


图 5　MFFN 结构示意图

Fig. 5　MFFN structure diagram

2.2.1　对于特征融合网络的改进

优化的特征融合网络（optimized feature fusion network，OFFN）结构如图 6 所示。OFFN 是基于 YOLO11s 特征融合网络进行的改进，在原有 3 层路径聚合特征金字塔网络（path aggregation network with feature pyramid network，PAN-FPN）结构的基础上增加了一条专门用于提取浅层特征的分支（如图中蓝色箭头所示）。此分支的加入使得模型能同时捕获来自更多不同尺度上的特征，提高对小目标的敏感程度，减少多尺度缺陷检测任务中的目标信息丢失。

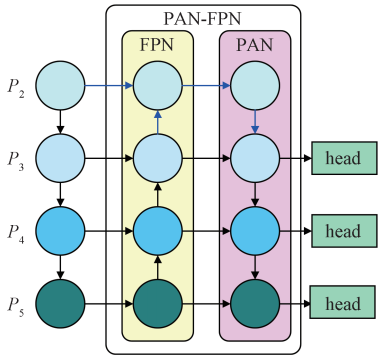


图 6　OFFN 结构示意图

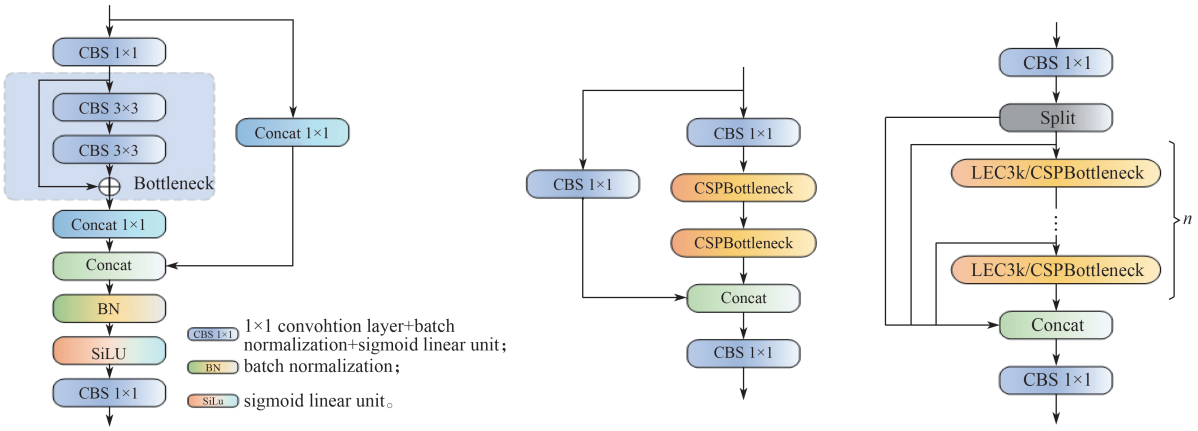
Fig. 6　OFFN structure diagram

OFFN 集成了双向特征融合机制，包含两条特征融合路径：一条为自顶向下的信息传递路径，另一条为自底向上的信息传递路径。在自顶向下的路径中，OFFN 利用特征金字塔网络（feature pyramid network，FPN）将通过 MFEN 网络卷积下采样得到的 4 个不同尺度的特征图进行特征上采样融合，将深层特征图的强语义信息向下传递，与浅层特征图相融合，以增强浅层特征图的语义表达能力。自底向上的路径中，OFFN 使用路径聚合网络（path aggregation network，PAN）对通过上采样融合后的

特征图进行特征下采样融合，将浅层特征图的强定位信息向上传递，与深层特征图相融合，以增强深层特征图的定位能力。通过使用双向路径聚合方法，OFFN 实现了更高效的多尺度特征融合，使得模型能够充分捕获来自不同尺度的特征信息，并通过跨层级信息交互方式强化语义信息与定位信息的互补性，以提高模型在复杂场景下对小目标及遮挡目标的检测性能。

2.2.2 FEMCSPF 模块

在深层网络中，重复运算和信息冗余将导致计算效率低下和模型性能受限。为解决这一问题，本文基于跨阶段局部网络 [13] 中提出的特征分割与部分连接思想，引入了跨阶段部分瓶颈模块（cross-stage partial bottleneck，CSPBottleneck），其结构见图 7a。



a）CSPBottleneck b）LEC3k c）FEMCSPF

图 7 CSPBottleneck、LEC3k 和 FEMCSPF 模块结构示意图

Fig. 7 Schematic diagram of CSPBottleneck, LEC3k and FEMCSPF module structure

CSPBottleneck 模块利用分层提取机制将输入特征分为两部分，第一部分通过 Bottleneck 模块提取深层语义特征，第二部分则跳过了 Bottleneck 模块，仅采用 1 × 1 卷积保留浅层特征。然后，将经过第一部分和第二部分处理的特征图在通道维度上进行拼接，通过卷积 − 批归一化 −SiLU 激活函数 1 × 1 模块（convolution-batch normalization-SiLU 1 × 1 module，CBS 1 × 1）对拼接后的特征图进行特征融合。

CSPBottleneck 模块通过特征分流操作，能够有效减少模型在深层网络中的大量重复运算，提高模型整体运算效率。同时，通过特征图拼接操作，模型既保留了原有浅层特征的定位信息，又兼顾了深层特征的语义信息，可有效降低在深层网络中因卷积操作带来的信息流失，提高模型对小目标和复杂背景中大型目标的检测效果。

轻量且高效的跨阶段部分可以定制卷积模块（lightweight and effective cross stage partial bottleneck with 3 convolutions and customizable kernel size，LEC3k），其结构如图 7b 所示。在 LEC3k 中，输入特征图也被分为两部分，一部分通过 CBS 1 × 1 模块进行卷积操作直接传递，保留浅层特征；另一部分则通过两个 CSPBottleneck 模块处理深层特征。然后，将两部分特征进行拼接，并利用 CBS 1 × 1 模块进行融合。这种设计方式不仅能实现模型轻量级优化，还使模型能高效捕捉来自深层次网络中的特征信息。

本文借助 CSPBottleneck 和 LEC3k 模块具备低计算量以及高效推理速度两方面优势，设计了 FEMCSPF 模块，如图 7c 所示，其中 $n \in \mathbb{Z}^+$。该模块通过并行卷积分支设计，将输入特征划分为两部分，一部分直接传递以保留浅层特征，另一部分通过可选择的判别参数 $n \in \mathbb{Z}^+$ 自由地选择多个 LEC3k 或 CSPBottleneck 模块以处理深层特征，最终通过拼接和 CBS 1 × 1 模块融合来自两部分的特征。这种设计能够增强不同层次间的信息交互能力，提高深层次特征提取任务中特征提取的效率，进一步提升模型对于复杂环境中多尺度目标的检测能力。

3 实验结果与分析

3.1 实验设置

本次实验使用的操作系统为 Ubuntu 20.04.5，CUDA 11.8，PyTorch 1.11.0，Python 3.8.20， 使用的 CPU 为 Intel(R) Xeon(R) Platinum 8474C，GPU 为 RTX 4090D，显存为 24 GB，运行内存为 80 GB。模型训练过程的超参数设置 [1] 如表 1 所示。

表 1 超参数设置

Table 1 Hyperparameter setting

超参数	取值	超参数	取值
Input size	640	Optimizer	SGD
Epoch	200	Initial learning rate	0.01
Batch size	32	Weight decay	0.000 5

Input size 设为 640，既符合模型分辨率要求，又能在计算资源消耗与保留图像特征之间达到良好平衡。Epoch 选取 200，是因多次实验表明此值下模型性能最佳，增大则容易过拟合，减小则会导致学习不充分。Batch size 设为 32，是因经实验验证该数值不仅能保证训练稳定，而且兼顾模型性能。Optimizer 选 SGD，是因其随机性可提升模型泛化能力，且实验表明在当前任务中应用 SGD 效果更佳。Initial learning rate 设为 0.01，参考了 SGD 常用范围，并且通过小规模实验验证，发现该数值能够使模型在训练初期保持稳定，同时实现损失函数的快速下降。Weight decay 取 0.000 5，这是经实验验证此值能够有效防止过拟合，平衡模型的复杂度与泛化能力。

本文中所使用的钢铁表面缺陷图像数据来源于东北大学表面缺陷检测数据集（Northeastern University surface defect detection dataset，NEU-DET）[14]。此数据集包含 6 类常见的热轧钢带表面缺陷，包括裂纹（crazing）、夹杂物（inclusion）、斑块（patches）、表面点蚀（pitted_surface）、轧制氧化皮（rolled_in_scale）和划痕（scratches）。每类缺陷分别含有 300 张分辨率为 200 × 200 像素的灰度图片，共计 1 800 张灰度图片。实验对该数据集进行随机划分，以 8:1:1 的比例划分为训练集、测试集和验证集。划分完成后，训练集包含 1 440 张图片，测试集和验证集各包含 180 张图片。本文在该数据集下进行实验，以验证 EDMS 模型的有效性。

3.2　评价指标

本文的主要目标是提高模型对钢铁表面缺陷的检测效果，因此在实验中使用精确率（precision）*P*、召回率（recall）*R*、平均精确率（average precision）*AP*、每秒帧数（frames per second）*FPS*、平均精度均值（mean average precision）*mAP* 和参数量（parameters）*Params* 作为模型性能的评价指标。

精确率表示模型预测为正样本的所有样本中，实际正样本的比例，它体现了模型对负样本的辨别能力。其计算式如下：

$$P = TP/(TP + FP), \tag{2}$$

式中：*TP*（true positives）为实际为正样本且被模型预测为正样本的样本数；*FP*（false positives）为实际为负样本但被模型预测为正样本的样本数。

召回率为模型正确识别出的正样本占所有实际正类样本的比例，它能够体现模型对正样本的识别能力。其计算式如下：

$$R = TP/(TP + FN), \tag{3}$$

式中：*FN*（false negative）为实际为正样本但被模型预测为负样本的样本数。

在目标检测模型中，平均精确率 *AP* 是衡量模型在特定类别上检测性能的核心指标，它通过精确率与召回率的动态关系 *P-R* 曲线（precision-recall curve）来衡量模型在这个类别上的检测性能。一般而言，*AP* 值越高，模型在此类数据集上的检测效果越好，其计算式如下：

$$AP = \int_0^1 P(r)\mathrm{d}r, \tag{4}$$

式中，*P*(*r*) 为特定召回率 *r* 下的精确率。

平均精度均值为每个类别平均精确率的算数平均值，是目标检测任务中评估模型综合性能的核心指标，其表达式为

$$mAP = \frac{1}{C}\sum_{i=1}^{C} AP_i, \tag{5}$$

式中：*C* 为 NEU-DET 数据集中缺陷类别的总数；AP_i 为第 *i* 类缺陷的平均精确率。

FPS 用于衡量模型每秒可处理的图像帧数，数值越高代表模型的检测速度越快。

参数量 *Params* 决定了模型在存储和运行时所需的内存空间。一般而言，参数量越小，模型越轻巧。

3.3　消融实验

为检验 EDMS 的有效性，本文在相同设备及实验环境中开展了消融实验。具体而言，针对每一个改进点都单独进行了实验，以验证各个改进点的有效性，结果见表 2。

表 2　EDMS 在 NEU-DET 数据集上的消融实验结果

Table 2　EDMS ablation experiment results on the NEU-DET dataset

模型	*P*/%	*R*/%	*mAP*/%	*FPS*/(帧·s^{-1})	*Params*/10^6
Baseline	74.08	70.16	74.95	98.26	9.42
+DSA-Down	78.21	70.15	76.68	92.33	8.81
+EMA	80.17	72.69	78.44	93.64	9.59
+OFFN	76.33	72.02	76.71	91.99	9.57
+FEMCSPF	79.95	70.82	77.48	95.74	9.01
EDMS	80.31	73.43	79.07	85.39	8.60

表 2 中，基准模型（Baseline）为 YOLO11s 基础模型，其 *P* 值为 74.08%，*R* 值为 70.16%，*mAP* 值为 74.95%，*Params* 值为 9.42×10^6，*FPS* 为 98.26 帧·s^{-1}。在 Baseline 中使用 DSA-Down 模块后，相较于基准模型，改进后模型的 *P* 值增长了 5.58%，*mAP* 值提高了 2.31%，*Params* 值下降了 0.61×10^6。这表明 DSA-Down 在对特征图进行下采样的过程中，相较于普通的卷积下采样，能够在增强特征图信息提取能力的同时，降低模型的参数量。EMA 注意力机制可以高效提取并充分利用关键特征，促使模型在

检测任务中能够更精确地完成对缺陷的分类与定位。在 Baseline 中加入 EMA 注意力机制后，相较于基准模型，改进后的模型 *P* 值上升了 8.22%，*R* 值增长了 3.61%，*mAP* 值提高了 4.66%。使用 OFFN 后，相较于基准模型，改进后的模型 *P* 值上升了 3.04%，*R* 值增长了 2.65%，*mAP* 值提高了 2.35%。不难发现 OFFN 实现了更为充分的多尺度特征融合，使模型能够更加有效地捕获来自不同尺度的特征信息，提高了模型在复杂场景下对多尺度目标的检测性能。当在 Baseline 中使用 FEMCSPF 模块后，相较于基准模型，改进后的模型 *P* 值上升了 7.92%，*R* 值增长了 0.94%，*mAP* 值提高了 3.38%，*Params* 值下降了 0.41×10^6。这表明，FEMCSPF 能够利用少量的参数获取大量的特征信息，在降低模型参数量的同时，提高模型对于特征图中关键信息的学习能力。本文提出的 EDMS 的 *P*、*R*、*mAP*、*Params* 和 *FPS* 值分别为 80.31%、73.43%、79.07%、8.60×10^6 和 85.39 帧 /s。相较于基准模型，其 *P* 值上升了 8.41%，*R* 值增长了 4.66%，*mAP* 值提高了 5.50%，*Params* 值下降了 0.82×10^6，*FPS* 值降低了 12.87 帧 /s。可以看出，相较于原始的 YOLO11s，EDMS 虽然在检测速度方面有所下降，但依旧保持着不错的检测速度，并且在减少模型参数量的同时，有效提升了检测精度。

3.4 对比实验

为了评估 EDMS 模型的性能，本文在 NEU-DET 数据集上，与目前较为先进的目标检测模型进行了对比实验，结果如表 3 所示。

表 3 不同检测模型在 NEU-DET 数据集上的检测结果

Table 3 Detection results of different detection models on the NEU-DET dataset

模型	*P*/%	*R*/%	*mAP*/%	*FPS*/(帧·s^{-1})	*Params*/10^6
SSD[15]	80.10	54.08	74.00	37.16	26.29
RT-DETR-l[16]	73.62	63.86	69.69	56.86	32.00
YOLOv8s	75.58	70.78	74.45	113.47	11.13
YOLOv9s	71.41	71.01	75.20	72.91	7.17
YOLOv10s	77.35	57.79	68.74	93.80	8.04
EDMS	80.31	73.43	79.07	85.39	8.60

由表 3 可知，在 NEU-DET 数据集上，EDMS 在 *P* 值、*R* 值、*mAP* 值方面均达到最优，高于目前较为先进的目标检测模型。具体而言，EDMS 较目前较为先进的目标检测模型的 *P* 值平均上升了 6.21%，*R* 值平均增长了 15.63%，*mAP* 值平均提高了 9.19%，*FPS* 值平均增加了 14.10%。这是由于钢铁表面图像通常为特征提取难度较大的灰度图，而 EDMS 在特征融合网络中考虑了更多不同尺度的特征信息，使得模型在钢铁表面缺陷数据集上的综合检测性能优于其他目标检测模型。相比于目前较为先进的目标检测模型的平均 *Params*，EDMS 的 *Params* 下降了 8.33×10^6。这背后的原因为 EDMS 在主干网络中采用了参数量更少的 DSA-Down 模块，并且在特征融合网络中，将原有深层网络中的 C3k2 模块替换为参数量更少的 FEMCSPF 模块。通过这些改进，EDMS 能够在提高检测精度的同时，有效降低模型整体参数量。

为了进一步评估 EDMS 模型对不同种类缺陷的检测性能，本文对比了 EDMS 与目前较为先进的目标检测模型在不同缺陷上的检测效果，如表 4 所示。

表 4 不同模型在 NEU-DET 数据集上对各类缺陷的检测结果

Table 4 Detection results of different detection models for various types of defects on the NEU-DET dataset %

模型	AP_{cr}	AP_{in}	AP_{pa}	AP_{pi}	AP_{ro}	AP_{sc}
SSD	50.79	79.32	84.82	81.70	73.09	74.30
RT-DETR-l	31.14	75.01	88.57	67.93	63.99	91.48
YOLOv8s	48.21	80.94	89.63	75.27	58.93	93.70
YOLOv9s	44.14	88.19	88.52	71.98	62.73	95.66
YOLOv10s	48.31	74.59	83.64	60.42	57.32	88.16
YOLO11s	47.43	84.03	87.65	73.53	62.80	94.27
EDMS	56.21	87.88	87.35	76.39	71.19	95.40

在表 4 中，第 2~7 列中的 AP_{cr}、AP_{in}、AP_{pa}、AP_{pi}、AP_{ro} 和 AP_{sc} 分别为对钢铁表面裂纹、夹杂物、斑块、表面点蚀、轧制氧化皮和划痕类缺陷检测的 *AP* 值。不难发现，EDMS 在 AP_{cr}、AP_{in}、AP_{pa}、AP_{pi}、AP_{ro} 和 AP_{sc} 这 6 项指标上分别达到了 56.21%, 87.88%, 87.35%, 76.39%, 71.19%, 95.40%，相较于目前较为先进的目标检测模型的平均值，分别提高了 24.90%, 9.38%, 0.24%, 6.39%, 12.74%, 6.48%。这表明 EDMS 具有更强的全局感知能力，能够有效提高对于钢铁表面不同种类缺陷分类和定位的准确性。

为了更直观地体现出 EDMS 模型的改进效果，本文采集了不同模型在 NEU-DET 数据集上检测的部分效果图片，如图 8 所示。观察图 8 可知，在裂纹类缺陷检测中，SSD、YOLOv9s 和 YOLOv11s 这几种模型普遍存在漏检问题。RT-DETR-l 不仅存在漏检问题，还会出现误检状况。YOLOv8s 同时存在漏检和重复检测两方面的问题，而 YOLOv10s 模型则出现了误检的现象。在夹杂物类缺陷检测方面，RT-DETR-l 和 YOLOv8s 均存在误检问题。在轧制氧化皮类缺陷检测中，RT-DETR-l、YOLOv8s 和 YOLOv9s 同样存在不同程度的误检问题。此外，SSD 在检测划痕类缺陷时还容易出现漏检的情况。相比之下，EDMS 展现出显著优势，不仅能够实现精准的缺陷分类与定位，避免漏检和重复检测现象，还能保持较高的预测置信度。

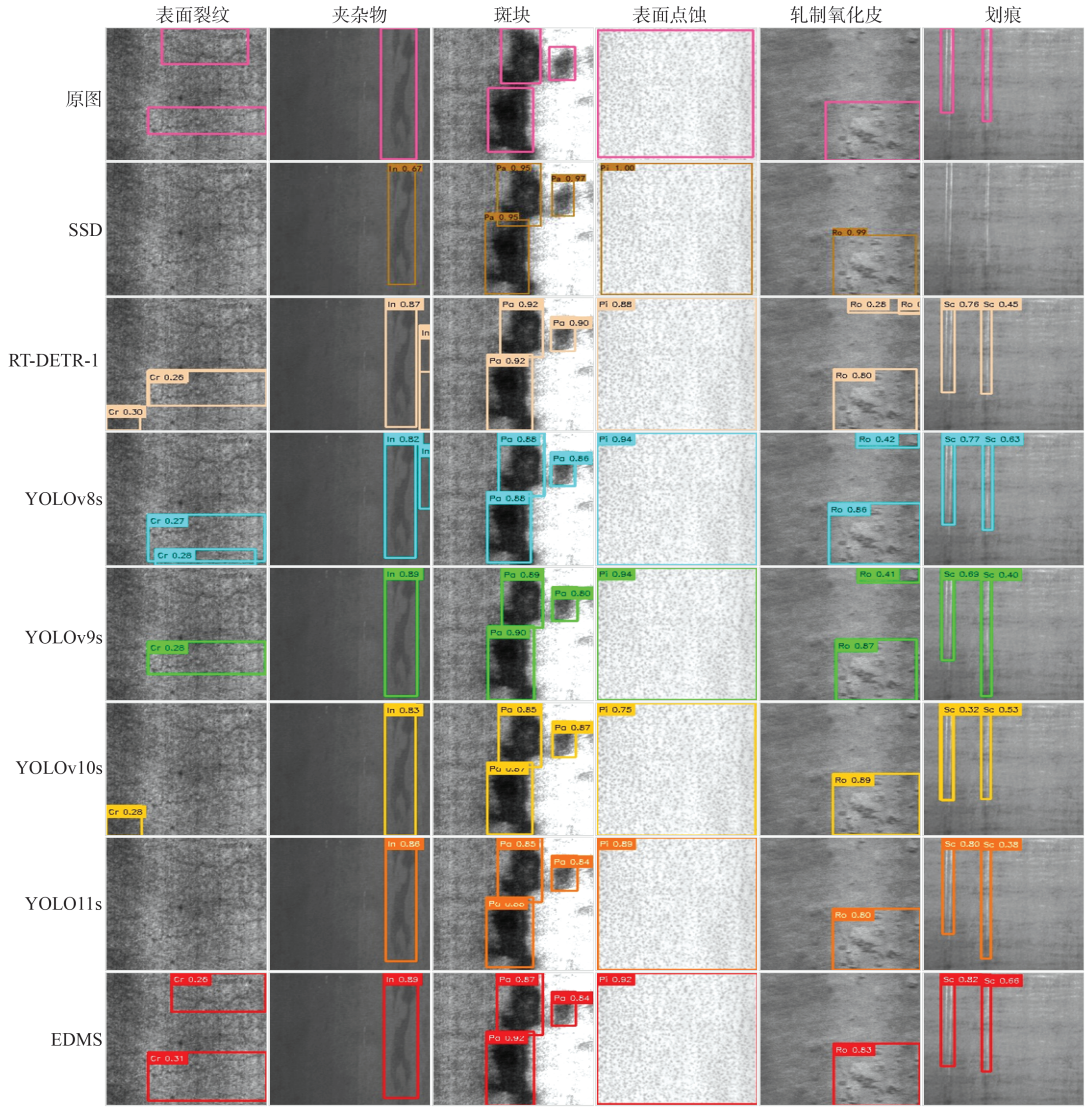


图 8　不同模型在 NEU-DET 数据集上的检测效果图

Fig. 8　Detection effect diagram of different models on the NEU-DET dataset

3.5　鲁棒性实验

为了测试 EDMS 的抗干扰能力，本文通过在 NEU-DET 数据集中分别添加高斯噪声（Gaussian noise，GN）和椒盐噪声（salt and pepper noise，SPN）的方式以制作鲁棒性测试数据集。鲁棒性测试实验主要使用的检测模型为 YOLO11s 和 EDMS，实验结果如表 5 所示。比较表中实验结果，不难发现高斯噪声和椒盐噪声对图片分辨率的影响较大，导致 YOLO11s 和 EDMS 模型的整体检测精度都有所降低，不过 EDMS 仍保持了不错的检测精度。总体而言，在高斯噪声和椒盐噪声的影响下，EDMS 相较于 YOLO11s 的 *mAP* 值分别提高了 2.60% 和 3.37%。具体来说，在高斯噪声实验中，EDMS 在裂纹、夹杂物、斑块、轧制氧化皮和划痕类缺陷上的检测精度均优于 YOLO11s 的，分别提高了 13.08%, 2.20%, 2.28%, 4.06%, 1.52%。在椒盐噪声实验中，EDMS 的 AP_{cr} 和 AP_{pa} 较 YOLO11s 的均有较大幅度的提高，分别提高了 28.97%, 9.02%。尽管在两种噪声的干扰下，EDMS 的检测精度均有所下降，但整体上仍保持着不错的检测性能，这表明其具备较强的抗干扰能力。

表 5　YOLO11s 和 EDMS 在鲁棒性测试数据集上的实验结果

Table 5　Experimental results of YOLO11s and EDMS on robust test datasets　　%

指标	YOLO11s			EDMS		
	None	GN	SPN	None	GN	SPN
mAP	74.95	72.60	72.95	79.07	74.49	75.41
AP_{cr}	47.43	40.61	45.33	56.21	45.92	58.46
AP_{in}	84.03	77.88	81.72	87.88	79.59	77.10
AP_{pa}	87.65	87.66	83.96	87.35	89.66	91.53
AP_{pi}	73.53	77.89	72.57	76.39	76.34	74.40
AP_{ro}	62.80	60.10	61.44	71.19	62.54	56.63
AP_{sc}	94.27	91.48	92.69	95.40	92.87	94.35

4 结语

为解决现有钢铁表面缺陷检测模型检测精度较低的问题，本文提出了一种新颖的 EDMS 检测模型。首先，通过引入 DSA-Down 下采样模块，在保持轻量级的同时提高了模型的检测精度。其次，引入 EMA 注意力机制和 OFFN 网络，使得特征融合网络中不同层次的信息能够得到充分有效的融合。并且通过使用 FEMCSPF 模块强化了模型对于复杂背景中多尺度目标的感知能力。在 NEU-DET 数据集上的实验结果表明，EDMS 的 *mAP* 值达 79.07%，与基准模型 YOLO11s 相比提高了 5.50%。与其他较先进的目标检测模型相比，EDMS 的 *mAP* 值表现最优。此外，本文还对 NEU-DET 数据集中的钢铁表面缺陷图像分别添加了高斯噪声和椒盐噪声，以构建模拟不同噪声干扰下的钢铁表面缺陷图像数据集，并分别验证了 EDMS 在不同噪声干扰下对钢铁表面缺陷检测的鲁棒性。

EDMS 虽然在钢铁表面缺陷检测领域拥有不错的检测性能，但在模型体量方面仍有进一步优化的空间。后续将继续深入研究，采用结构化剪枝等方法，致力于设计轻量化效果更佳的模型，并在保证模型性能稳定的前提下，进一步提升其检测精度。

参考文献：

[1] XIE W N，MA W F，SUN X Y. An Efficient Re-Parameterization Feature Pyramid Network on YOLOv8 to the Detection of Steel Surface Defect[J]. Neurocomputing，2025，614：128775.

[2] REN S Q，HE K M，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，2016，39(6)：1137−1149.

[3] 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 (CVPR). Las Vegas：IEEE，2016：779−788.

[4] CHEN H，QIU J L，GAO D P，et al. Research on Surface Defect Detection Model of Steel Strip Based on MFFA-YOLOv5[J]. IET Image Processing，2024，18(8)：2105−2113.

[5] ZHAO X Y，ZHAO J D，HE Z X. A Multiple Feature-Maps Interaction Pyramid Network for Defect Detection of Steel Surface[J]. Measurement Science and Technology，2023，34(5)：055401.

[6] 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.

[7] LÜ Z L，ZHAO Z Q，XIA K W，et al. Steel Surface Defect Detection Based on Mobile ViTv2 and YOLOv8[J]. The Journal of Supercomputing，2024，80(13)：18919−18941.

[8] KHANAM R，HUSSAIN M. YOLOv11：An Overview of the Key Architectural Enhancements[EB/OL]. [2025−04−04]. 2024: arXiv: 2410.17725. https://arxiv. org/abs/2410.17725.

[9] OUYANG D L，HE S，ZHANG G Z，et al. Efficient Multi-Scale Attention Module with Cross-Spatial Learning[C]//ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Rhodes Island：IEEE，2023：1−5.

[10] SUNKARA R，LUO T. No More Strided Convolutions or Pooling：A New CNN Building Block for Low-Resolution Images and Small Objects[C]//Machine Learning and Knowledge Discovery in Databases. Cham：Springer，2022：443−459.

[11] WOO S，PARK J，LEE J Y，et al. CBAM：Convolutional Block Attention Module[C]//Computer Vision-ECCV 2018. Cham：Springer，2018：3−19.

[12] CHOLLET F. Xception：Deep Learning with Depthwise Separable Convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu：IEEE，2017：1251−1258.

[13] WANG C Y，LIAO H Y M，WU Y H，et al. CSPNet：A New Backbone that Can Enhance Learning Capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle：IEEE，2020：390−391.

[14] BAO Y Q，SONG K C，LIU J，et al. Triplet-Graph Reasoning Network for Few-Shot Metal Generic Surface Defect Segmentation[J]. IEEE Transactions on Instrumentation and Measurement，2021，70：5011111.

[15] LIU W，ANGUELOV D，ERHAN D，et al. SSD：Single Shot MultiBox Detector[C]//Computer Vision–ECCV 2016. Cham：Springer，2016：21−37.

[16] ZHAO Y A，LÜ W Y，XU S L，et al. DETRs Beat YOLOs on Real-Time Object Detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle：IEEE，2024：16965−16974.

（责任编辑：廖友媛）