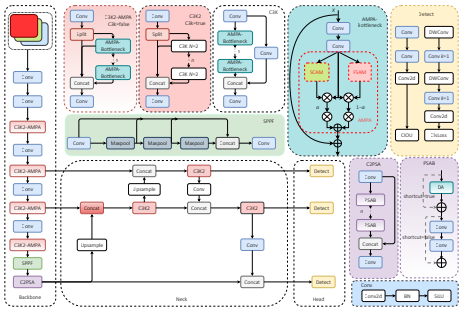
OEE | 光电工程 Opto-Electronic Engineering

Article

2026 年 1 月, 第 53 卷, 第 1 期

基于自适应多重并行注意力机制的微尺度刀具缺陷检测方法



刘子豪[1*], 陈泉达[2], 金友振[1,3], 何伟军[4], 鹿业波[5]

[1]嘉兴大学人工智能学院, 浙江 嘉兴 314001;
[2]浙江理工大学人工智能学院, 浙江 杭州 310018;
[3]天津大学电气自动化与信息工程学院, 天津 300072;
[4]恒锋工具股份有限公司, 浙江 嘉兴 314300;
[5]嘉兴大学机械工程学院, 浙江 嘉兴 314001

摘要: 在精密刀具表面微尺度缺陷检测中, 传统注意力机制多采用单向或串行执行方式, 难以全面捕捉刀具表面缺陷多样化的形态及其复杂的纹理演变模式, 导致模型泛化能力受限。为解决上述问题, 本文提出了一种基于自适应多重并行注意力 (adaptive multi-parallel attention, AMPA) 的精密刀具缺陷检测模型。首先, 构建星通道注意力模块 (star-channel attention module, SCAM), 借助星操作 (star operations) 增强通道间的非线性耦合, 提高对关键通道特征的表达能力。其次, 设计结合频域与空间信息的频域-空间注意力模块 (frequency-spatial attention module, FSAM), 有效捕获图像中的低频、中频和高频特征以及空间结构信息。最后, 通过可学习参数动态调整两个模块的融合权重, 实现特征的自适应融合, 提升检测性能。实验结果表明, 将所提方法集成至主流 YOLOv11 模型后, 在自建的精密刀具缺陷数据集上实现了 90.2% 的 mAP0.5, 较传统 YOLOv11 提升 2.4%, 充分验证了模型在刀具表面微尺度目标缺陷检测中的有效性; 同时, 在公开的 NEU 钢材表面缺陷数据集上达到 76.1% 的 mAP0.5, 较传统 YOLOv11 提升 1.3%, 展示了模型良好的泛化能力。

关键词: 精密刀具; 注意力机制; 缺陷检测

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250173 | CSTR: 32245.14.oee.2026.250173 | 中图分类号: TP391.4 | 文献标识码: A

刘子豪, 陈泉达, 金友振, 等. 基于自适应多重并行注意力机制的微尺度刀具缺陷检测方法 [J]. 光电工程, 2026, **53**(1): 250173

Liu Z H, Chen Q D, Jin Y Z et al. Adaptive multiscale parallel attention mechanism for micro-scale tool defect detection method[J]. *Opto-Electron Eng*, 2026, **53**(1): 250173

1 引 言

随着现代制造业对高精度加工需求的提升, 精密刀具作为加工某些国之重器 (国产 919 大飞机、核工业装备) 的核心器件, 其质量越来越受到重视。刀具表面微尺度缺陷如磨损、裂纹、磕碰等在高强度加工环境下迅速扩展, 损伤部位的缺陷会不断被放大, 影响终端产品性能及安全, 严重制约产品加工制造效率。相比于其他金属零部件, 刀具缺陷表现出更微小的形态和纹理特征, 极大增加了自动化检测的难度, 因此, 研发高效准确的刀具表面微尺度缺陷检测技术成为智能制造的关键。

深度卷积神经网络 (Convolutional neural networks,

收稿日期: 2025-06-18; 修回日期: 2025-10-05; 录用日期: 2025-10-09; 网络出版日期: 2026-01-25

基金项目: 国家自然科学基金资助项目 (62374074); 浙江省“尖兵领雁”研发攻关计划 (2024C04028); 嘉兴市公益性研究计划项目 (SQGY202400009); 校企合作项目 (00523144); 海盐重点研发计划项目 (2024ZD03); 嘉兴大学人才项目 (CD70623008)

*通信作者: 刘子豪, lzh@zjxu.edu.cn

版权所有©2026 中国科学院光电技术研究所

CNN) 因在计算机视觉任务中表现出卓越性能，广泛应用于工业缺陷检测任务中。CNN 通过卷积层与非线性激活函数的叠加，能够有效提取图像的多层次抽象特征，进而生成特征丰富的全局表示[1]。尽管基于 CNN 的目标检测模型在自然场景表现卓越，但面对形态多样且结构复杂的工业缺陷，尤其是精密刀具表面微缺陷时，传统 CNN 往往难以捕捉到关键特征，限制了检测效果和模型的泛化能力[2-3]。近年来，注意力机制逐渐成为提升神经网络性能的重要手段，借鉴人类视觉系统的选择性注意特质，注意力机制通过聚焦重要特征并抑制冗余信息，显著增强了模型的表征能力[4-5]。然而，现有主流注意力方法 (如通道注意力和空间注意力) 在捕捉特征间复杂依赖关系方面仍存在不足[6]，尤其在工业缺陷微尺度、形态及背景的多样性使单一维度注意力难以兼顾所有关键信息，导致检测性能提升受限。注意力机制的本质是通过局部聚焦来提取关键信息，在目标检测中作为特征优化模块，有效增强显著特征的表达[5,7]，但现有策略尚难覆盖工业缺陷的全方位特征表达需求。

基于以上注意力机制中的问题，学者们从通道、空间及混合维度入手提出了多种注意力机制构造思路，尝试提升网络对关键特征的捕捉能力。例如，Hu 等人提出的 SE (squeeze-and-excitation) 模块通过全局信息压缩对通道特征进行重新校准[4]，实现对特征的精准提取；Wang 等设计的 ECA (efficient channel attention) 模块利用一维卷积降低计算复杂度，同时保持性能[8]；Shen 等设计了 GENet 结构，通过上下文建模挖掘空间特征间关系[9]；Wang 等提出 Non-Local 定向机制，利用全局像素相似性捕获长距离依赖，但计算复杂度较高[10]。进一步地，混合注意力机制如 Woo 等提出通过串联通道与空间注意力 (convolutional block attention module, CBAM)，可显著增强模型鲁棒能力[6]；Ji 等在 Faster R-CNN 中引入结合通道与位置注意力的 SPAM (squeeze and position attention mechanism) 模块，提升了复杂背景下小目标检测能力[2]。然而，该方法在串行机制下计算时依赖顺序处理多个模块，易导致信息丢失且计算开销大。类似地，Chen 等提出 DCAM-Net (deformable convolution and attention mechanism network) 网络结构，通过把可变形卷积融入空间注意力模块中，提升缺陷定位能力，但该方法只利用空间注意力，忽略了通道维度的核心特征，限制了对复杂缺陷特征的深度表征性能[11]。Wu 等针对皮革缺陷提出的多尺度融合注意力网络，自适应选择多尺度特征，实现了高效准确识别[12]。

尽管上述文献中提出了较多关于不同注意力机制的研究方法，但将这些方法直接运用于检测刀具表面的外观微小缺陷目标时，会出现以下两个关键问题：

1) 微尺度缺陷难以识别问题。现有注意力机制在处理精密刀具表面微小缺陷的形态多样性时，缺乏对通道特征的有效耦合与表达，导致无法精确识别关键信息。

2) 缺陷纹理多样性问题。在面对精密刀具表面复杂缺陷的多样化纹理特征时，传统方法未能充分利用频域信息来捕捉缺陷的细微变化，影响了检测的准确性。

针对以上问题，本文提出一种自适应多重并行注意力机制 (adaptive multi-parallel attention, AMPA)，该机制并行处理通道、空间及频域三类注意力。首先，通过星操作增强通道间的非线性耦合，提高对关键通道特征的表达能力，有效应对刀具表面缺陷形态的多样性。同时，引入频域-空间协同的注意力模块，整合频域和空间特征，通过对不同尺度的纹理信息加权，提升对复杂缺陷的识别能力，全面捕获刀具图像中的微尺度特征。最后，采用可学习参数动态调整策略，实现多维特征的自适应融合，增强了模型对多尺度、多维特征的捕捉能力，提升精密刀具缺陷检测的整体性能与泛化能力。

2 方　法

2.1 YOLO-AMPA 总体架构

YOLO 模型作为典型的单阶段目标检测算法，其核心思想是将目标检测任务视为回归问题，通过一次前向传播直接预测目标的边界框和类别概率，因而在速度与精度的平衡上表现较好。近年来，YOLO 模型持续优化骨干网络和特征融合模块，不断提升检测性能[13-17]。然而，当把 YOLO 模型应用于检测具有微尺度特征和复杂纹理背景的精密刀具缺陷时，其存在特征表达不足、难以精准捕获刀具缺陷的细微差异等问题。为进一步增强模型在刀具表面缺陷特征的表征能力，本研究将自适应多重并行注意力机制模块 (AMPA) 集成到 YOLOv11 的 C3K2 模块中，采用 AMPA-Bottleneck 模块替代传统的 Bottleneck 结构，以提升特征提取效率并强化特征信息的交互与融合。本文所提的 YOLOv11-AMPA 模型架构如图 1 所示，其中，图 1 通过把 AMPA 模块嵌入 YOLOv11 中 C3K2 模块为基准进行展示。

2.2 AMPA 总体框架

本节提出的 AMPA 是一种可灵活嵌入主流深度卷

积神经网络的通用特征增强模块，其设计初衷是解决传统注意力机制在缺陷检测中存在的局限性，如串行处理导致的信息丢失，单一维度 (通道或空间) 难以兼顾缺陷的多特征 (形态多样性、纹理复杂性) 等问题。以 YOLOv11 为例，AMPA 可嵌入特征提取网络 (Backbone) 的 C3K2 模块中 (如图 1 所示)，通过动态加权关键特征 (如微尺度缺陷的边缘通道、高频纹理) 与抑制冗余信息 (如背景干扰)，实现对刀具表面微小缺陷的精准识别。

AMPA 的核心架构由两个功能互补的子模块构成：星通道注意力模块 (SCAM) 与频域-空间协同注意力模块 (FSAM)。两者采用并行处理机制，从不同维度挖掘缺陷特征：SCAM 聚焦通道维度，通过星操作强化通道间的非线性耦合，捕捉微尺度缺陷的关键通道响应 (如磕碰缺陷的边缘纹理通道)；FSAM 则整合频域与空间维度，利用快速傅里叶变换 (FFT) 提取全频域信息 (低频轮廓、高频细节) 并融合空间结构特征，应对复杂纹理背景下的缺陷定位。这种并行设计能最大限度保留两类特征的独特性—SCAM 与 FSAM 的特征互信息较低 (冗余度小)，总信息熵高于串行结构 (避免后序模块覆盖前序特征)，为高效融合奠定基础。

具体地，给定输入特征图 $\boldsymbol{X}\in\mathbf{R}^{H\times W\times C}$，AMPA 并行生成一维通道注意图 $\boldsymbol{M}_{\rm sc}\in\mathbf{R}^{C\times1\times1}$ 和二维频域空间注意图 $\boldsymbol{M}_{\rm fs}\in\mathbf{R}^{1\times H\times W}$，并引入可学习参数 α 来动态调整两者的融合权重，实现输入特征的有效加权融合，以上过程可用式 (1)~(3) 表示:

$$\boldsymbol{F}_1=\boldsymbol{M}_{\rm sc}(\boldsymbol{X})\otimes\boldsymbol{X}\,, \tag{1}$$

$$\boldsymbol{F}_2=\boldsymbol{M}_{\rm fs}(\boldsymbol{X})\otimes\boldsymbol{X}\,, \tag{2}$$

$$\boldsymbol{F}=\alpha\otimes\boldsymbol{F}_1+(1-\alpha)\otimes\boldsymbol{F}_2\,, \tag{3}$$

式中：$\otimes$ 表示逐元素相乘，通道注意力图 $\boldsymbol{M}_{\rm sc}$ 通过扩展其空间维度 (从 $C\times1\times1$ 到 $C\times H\times W$) 与输入特征图 $\boldsymbol{X}$ 进行逐元素相乘，强化了重要通道的特征响应；频域空间注意力图 $\boldsymbol{M}_{\rm fs}$ 则通过扩展其通道维度 (从 $1\times H\times W$ 到 $C\times H\times W$) 与 $\boldsymbol{X}$ 逐元素相乘，以强调图像中频域和空间上同等重要的区域特征。最终，由可学习参数 α 动态调节两类注意力的融合比例，实现对特征的自适应加权。输出

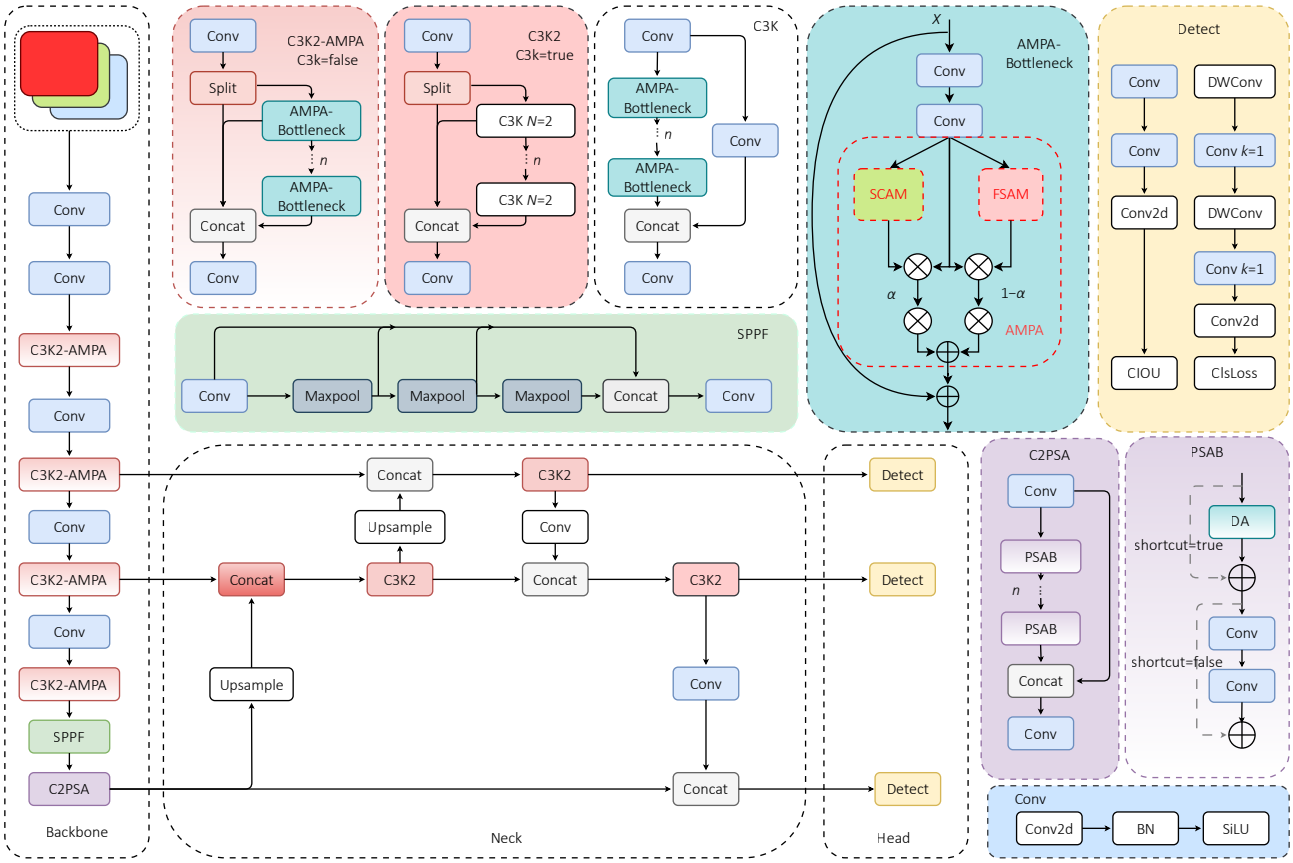


图 1 YOLOv11-AMPA 架构图

Fig. 1 YOLOv11-AMPA architecture diagram

特征图 $\boldsymbol{F}$ 综合了多维度的语义信息，大幅提升模型的特征表达能力和判别力。其中加权系数α为一个可学习标量参数，其初始值设为 0.5。在前向传播时，α直接参与式 (3) 的加权融合；在反向传播时，α与网络其余权重一起接收来自检测损失 (包括 cls-loss 与 box-loss) 的梯度，并通过 SGD 优化器同步更新。整个训练流程无需额外监督。图 2 展示了 AMPA 模块的整体结构图。

2.3 星通道注意力模块 (SCAM)

针对精密刀具表面缺陷微尺度的问题，传统通道注意力机制难以充分挖掘通道间关于微小特征映射的非线性耦合及复杂交互关系，本节所提的 SCAM 模块属于 AMPA 中的一个子模块，用于将传统卷积中的权重矩阵拆分为多组稀疏的子卷积核，通过多路径并行计算实现复杂的交互特征融合，生成丰富的特征表达。同时，融合星操作 (star operation) 的思想[18]，将聚合后的特征信息映射到一个高维的隐式特征空间。基于这种设计，SCAM 能更有效地挖掘通道之间的复杂关系，提升模型对重要特征的表达与区分能力。图 3 展示了星操作的结构图。

SCAM 的具体执行步骤如下:

1) 通道描述向量计算。为了计算通道注意力，对输入特征图$\boldsymbol{X}\in\mathbf{R}^{H\times W\times C}$在空间维度上进行全局平均池化，生成通道描述向量$\boldsymbol{Z}_c$，其形状为$C\times1\times1$，如式 (4)：

$$\boldsymbol{Z}_c=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\boldsymbol{X}_c(i,j)\,, \tag{4}$$

式中：$\boldsymbol{X}_c$ 表示第 c 通道特征图； (i,j) 表示特征图上的空间坐标，其主要作用是整合通道维度的全局信息，为后续的通道间交互提供基础。

2) 特征扩展。通过一个卷积核权重矩阵 $\boldsymbol{W}_1\in\mathbf{R}^{C\times2C}$ 将通道数扩展为原来的两倍，拓展特征表达能力，使模型能够学习更丰富的通道特征。经过激活函数 *Relu6* 处理，输出结果如式 (5) 所示：

$$\boldsymbol{y}_1=Relu6\,(\boldsymbol{W}_1\boldsymbol{z})\,. \tag{5}$$

3) 特征耦合。将扩展后的向量$\boldsymbol{y}_1$分为两部分$\boldsymbol{y}_2$和$\boldsymbol{y}_3$，其中$\boldsymbol{y}_2$通过*Relu6*激活函数进行非线性变换，$\boldsymbol{y}_3$则保持线性，然后通过逐元素相乘的方式将两部分进行耦合。如式 (6)~(7) 所示：

$$\boldsymbol{y}_2=\boldsymbol{y}_1[:,:C]\text{，}\quad\boldsymbol{y}_3=\boldsymbol{y}_1[:,C:]\,, \tag{6}$$

$$\boldsymbol{z}_1=Relu6\,(\boldsymbol{y}_2)\otimes\boldsymbol{y}_3\,, \tag{7}$$

式中：$\boldsymbol{y}_2,\boldsymbol{y}_3\in\mathbf{R}^{B\times C}$。

4) 注意力权重生成。将耦合后的特征 $\boldsymbol{z}_1$ 输入全连接层进行通道组合，生成注意力权重$\boldsymbol{z}_2$，随后通过 Sigmoid 激活函数 σ将权重值归一化至 [0,1]，用于对原始输入特征图 $\boldsymbol{X}$ 进行加权得到输出$\boldsymbol{F}_1$：

$$\boldsymbol{z}_2=\boldsymbol{W}_2\boldsymbol{z}_1\,, \tag{8}$$

$$\boldsymbol{F}_1=\sigma(\boldsymbol{z}_2)*\boldsymbol{X}\,, \tag{9}$$

其中，$\boldsymbol{W}_2\in\mathbf{R}^{C\times C}$为全连接层权重，用于全局特征组合，连接前后层的所有神经元。随着网络训练的不断迭代，模型能够自适应调整通道权重，突出重要特征，弱化无关信息。图 4 展示了完整的 SCAM 结构图。

2.4 频域-空间协同注意力模块 (FSAM)

针对刀具表面纹理的复杂性导致模型无法精准识别的问题，本节提出频域-空间协同注意力模块 (FSAM)，该模块是 AMPA 中的一个子模块，通过融合频域与空间信息解决微尺度缺陷的复杂纹理特征 (如微小裂纹的高频细节与磕碰的低频轮廓) 难以有效提取的问题。微尺度缺陷的频谱呈现显著分化-高频成分 (如缺陷边缘、棱角) 能量占比低但对识别至关重要，低频成分 (如刀具整体轮廓) 能量占比高但可能包含冗余背景信息。FSAM 通过频域变换与动态加权机制，实现对不同频率成分的

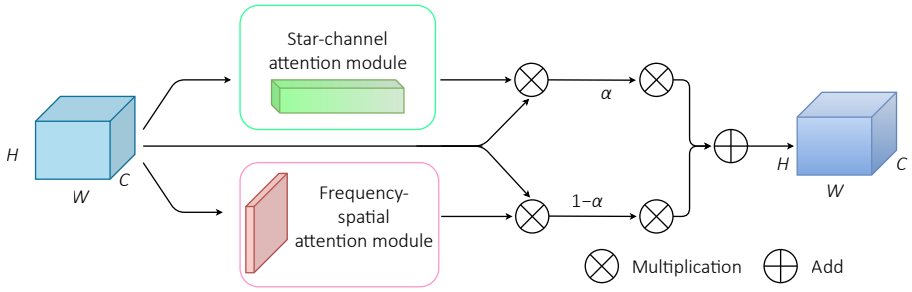


图 2 AMPA 模块的整体结构

Fig. 2 Overall structure of the AMPA module

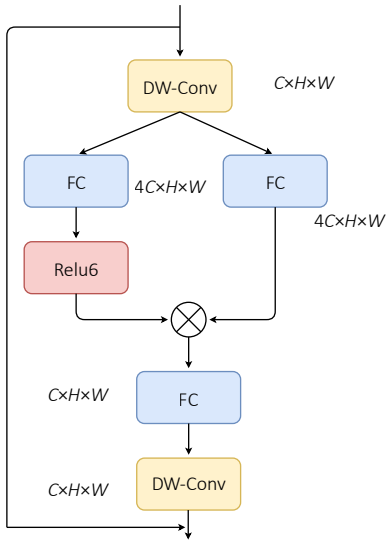


图 3 Star operations 结构图

Fig. 3 Structure diagram of star operations

自适应筛选与增强。首先，对输入特征图进行全局汇聚，分别计算平均池化和最大池化，生成空间注意特征图。然后，采用快速傅里叶变换 (FFT) 将特征图转换至频域，提取低频和高频信号，并通过一个卷积层生成位置编码，对空间特征进行增强，接着对增强后的特征进行通道归一化处理，生成频域注意特征图。最后，将两者通过逐元素相乘进行融合，生成具有增强表达力的注意力图，融合后的特征图被用于更新原始特征图，提升模型在复杂环境下对刀具缺陷的识别能力。图 5 展示了 FSAM 结构图，包含频域分支 (FFT 2D) 以及空间信息整合过程。

FSAM 具体执行步骤如下：

1) 特征聚合。首先，采用两个池化操作 (平均池化和最大池化) 对输入特征图的每个通道进行聚合，生成两个 2D 特征图—$\boldsymbol{F}^{\mathrm{s}}_{\mathrm{avg}} \in \mathbf{R}^{1\times H\times W}$ 和 $\boldsymbol{F}^{\mathrm{s}}_{\mathrm{max}} \in \mathbf{R}^{1\times H\times W}$，其中，每个特征图分别表示基于平均池化和最大池化的空间响

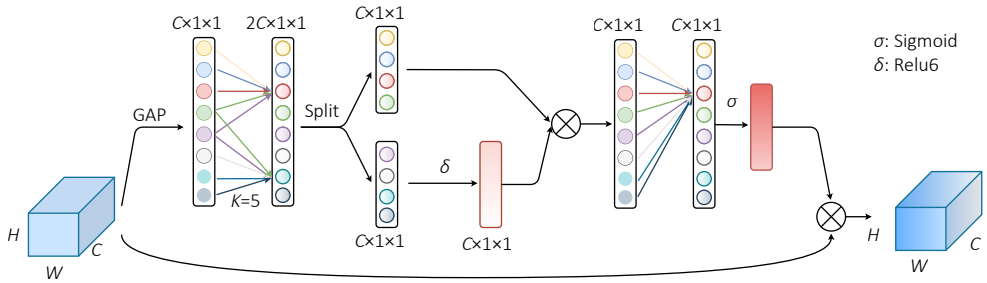


图 4 SCAM 结构图

Fig. 4 SCAM structural diagram

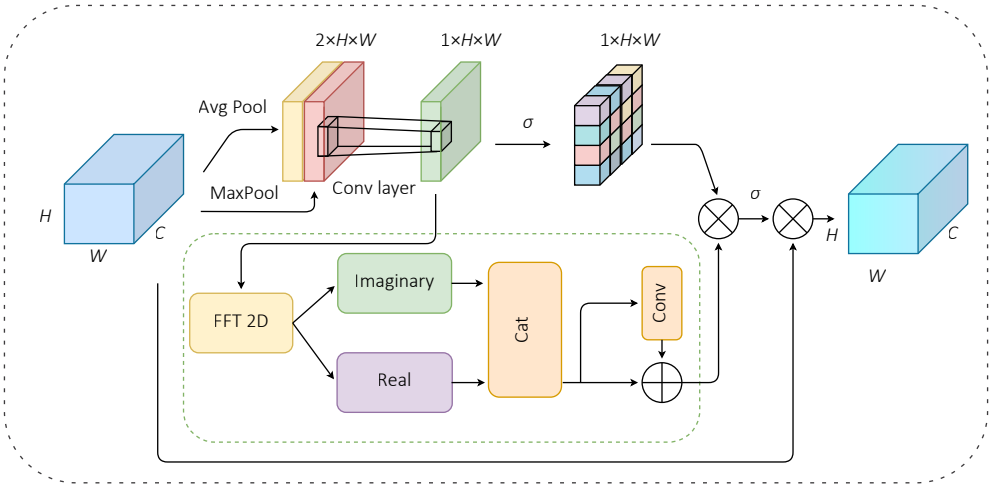


图 5 FSAM 结构图

Fig. 5 FSAM structural diagram

应。然后，将两个特征图在通道维度上拼接，并通过卷积操作融合特征，产生最终的二维空间注意特征图 $\boldsymbol{M}$，以上过程的计算可用式 (10) 表示：

$$\begin{aligned}\boldsymbol{M} &= Conv([AvgPool(\boldsymbol{F}); MaxPool(\boldsymbol{F})]) \\ &= Conv\left(\left[\boldsymbol{F}^{\mathrm{s}}_{\mathrm{avg}}; \boldsymbol{F}^{\mathrm{s}}_{\mathrm{max}}\right]\right) .\end{aligned} \tag{10}$$

2) 空间注意力权重计算。对上述获取的特征图 $\boldsymbol{M}$ 进行 Sigmoid 非线性激活 (σ) 操作，得到空间注意力权重 $\boldsymbol{M}_{\mathrm{s}}$：

$$\boldsymbol{M}_{\mathrm{s}} = \sigma(M(\boldsymbol{F})) . \tag{11}$$

3) 频域特征提取。将输入特征图 $\boldsymbol{X} \in \mathbf{R}^{H\times W\times C}$ 经过快速傅里叶变换转换到频率域，获取复数域频谱特征表示 $\boldsymbol{F}(\boldsymbol{X})$：

$$\boldsymbol{F}(\boldsymbol{X})(u,v) = \sum_{x=0}^{H-1}\sum_{y=0}^{W-1} Y(x,y)\cdot \mathrm{e}^{-\mathrm{j}2\pi(ux/H+vy/W)} , \tag{12}$$

式中：$\boldsymbol{F}(\boldsymbol{X})(u,v)$ 包含了输入特征图在不同频率上的信息；u 和 v 是频率坐标；$Y(x,y)$ 是空间域特征值，x 和 y 是空间坐标；j 是虚数单位；$\mathrm{e}^{-\mathrm{j}2\pi(ux/H+vy/W)}$ 表示复指数函数。

4) 频域位置编码与特征增强。输入特征经过二维 FFT 后得到复数域特征，分别提取其实部 $\boldsymbol{F}_{\mathrm{R}} \in \mathbf{R}^{H\times W\times C}$ 和虚部 $\boldsymbol{F}_{\mathrm{I}} \in \mathbf{R}^{H\times W\times C}$。随后，将两部分在通道维度进行拼接，得到 $\boldsymbol{F}_{\mathrm{C}} \in \mathbf{R}^{H\times W\times 2C}$ 的联合特征。最后，由一个二维 Conv 层对 $\boldsymbol{F}_{\mathrm{C}}$ 进行位置编码，动态学习不同频率成分的重要性，例如，对于微尺度缺陷的高频细节 (如裂纹边缘)，卷积核参数会自动增大对高频区域的响应权重；而对背景相关的低频冗余信号 (如刀具表面正常纹理)，则抑制其响应强度。然后以残差的形式与 $\boldsymbol{F}_{\mathrm{C}}$ 合并，以获得特征 $\boldsymbol{F}^{\mathrm{Res}}_{\mathrm{c}} \in \mathbf{R}^{H\times W\times C}$，以上过程可以用式 (13) 和 (14) 表示：

$$\boldsymbol{F}_{\mathrm{C}} = cat(\boldsymbol{F}_{\mathrm{R}}, \boldsymbol{F}_{\mathrm{I}}) , \tag{13}$$

$$\boldsymbol{F}^{\mathrm{Res}}_{\mathrm{c}} = \boldsymbol{F}_{\mathrm{C}} + Conv(\boldsymbol{F}_{\mathrm{C}}) . \tag{14}$$

5) 频域注意力归一化。为减小不同频率分量在融合时的影响偏差，对频域特征 $\boldsymbol{F}^{\mathrm{Res}}_{\mathrm{c}}$ 按通道维度进行归一化处理，得到归一化频域注意力权重 $\boldsymbol{M}_{\mathrm{f}}$：

$$\boldsymbol{M}_{\mathrm{f}} = \frac{\boldsymbol{F}^{\mathrm{Res}}_{\mathrm{c}}}{\sum_{i=1}^{H}\sum_{j=1}^{W}\boldsymbol{F}^{\mathrm{Res}}_{\mathrm{c}}[\ldots,i,j]} , \tag{15}$$

式中：分母是对每个通道的频谱能量在空间维度上的求和。

6) 空间与频域注意力融合及输出。将归一化后的频域注意力 $\boldsymbol{M}_{\mathrm{f}}$ 与空间注意力 $\boldsymbol{M}_{\mathrm{s}}$ 逐元素相乘，实现多维特征的融合增强，进而加权输入特征图 $\boldsymbol{X}$，得到最终输出结果 $\boldsymbol{F}_2$：

$$\boldsymbol{F}_2 = X * Sigmoid(\boldsymbol{M}_{\mathrm{f}} * \boldsymbol{M}_{\mathrm{s}}) . \tag{16}$$

3 实验结果与分析

3.1 实验环境及参数设置

本文实验基于 Windows 11 的 64 位操作系统，CPU 为 Intel (R) Core (TM) i9-10900X，GPU 型号为 NVIDIA GeForce RTX 4090 (显存为 24 G)，服务器运行内存为 64 GB，固态硬盘 (SSD) 型号为三星 Samsung 870 QVO，硬盘大小为 1 TB，机械硬盘型号为 WDC-WD20EZAZ-22L9GB0。实验基于 Python 语言设计开发，程序实现平台采用 PyCharm，深度学习框架为 Pytorch 2.0.0，CUDA 版本为 11.8。模型中的超参数设置包括：优化器 (optimizer) 为随机梯度下降 (stochastic gradient descent, SGD)，动量 (momentum) 为 0.937，权重衰减 (weight decay) 设定为 0.0005，学习率调度采用余弦退火 (cosine annealing) 从初始值 0.01 开始，训练轮数为 200，批处理大小为 16，数据增强方法采用随机裁剪的 Mosaic 技术，图像大小在训练阶段与测试调整为 640 pixel×640 pixel。

3.2 数据集

本文所构建的精密刀具数据集包含 2185 张含缺陷刀具图像，包含单类磕碰缺陷，此类缺陷检测的难点在于所有要检测的对象均为小目标，平均尺寸约为 20 pixels。数据集中每张图像的分辨率为 1920 pixel×1080 pixel，并在前期搭建的精密刀具机器视觉系统中采集[19]，数据集按 8:1:1 的比例分为训练集、验证集、测试集，其中训练集 1749 张图像，验证集和测试集各 218 张图像。为测试算法在不同应用场景下的泛化性，本文引入东北大学钢材表面缺陷 NEU-DET[20] 数据集，包含六种钢表面缺陷—裂纹、夹杂物、斑点、麻面、轧入氧化皮和划痕，每类缺陷样本数量均为 300，每张图像分辨率为 200 pixel× 200 pixel，总共包含 1800 张图像。图 6(a) 展示了自建数据集中部分样本，图 6(b) 展示了 NEU 数据集的部分样本。NEU-DET 的训练集，验证集和测试集的比率设置为 8:1:1，其中 1440 张图像用于训练，而验证与测试各 180 张图像。两类数据集虽均聚焦工业金属表面缺陷检测，存在一定共性 (如均需从金属纹理

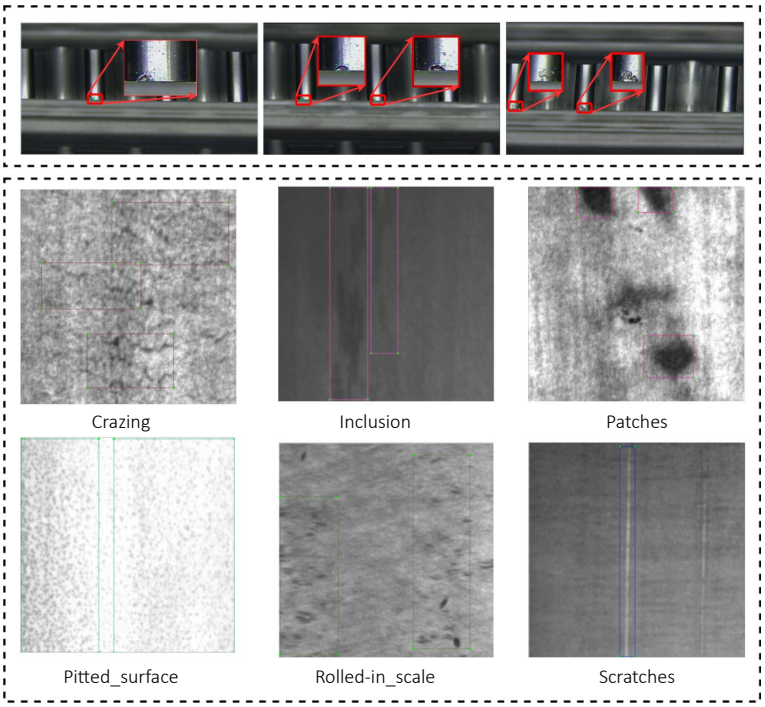


图 6 缺陷样本示意图。(a) 刀具数据集；(b) NEU 数据集

Fig. 6 Illustration of defect samples. (a)Tool dataset; (b) NEU dataset

背景中识别缺陷，依赖对细微特征的捕捉能力)，但在核心特性上存在显著差异，具体如下：从缺陷尺寸看，自建数据集缺陷平均尺寸约 20 pixels，属于典型微尺度目标；NEU 数据集缺陷尺寸相对更大，属于中大尺度目标。从图像属性看，自建数据集图像分辨率为 1920 pixel×1080 pixel，细节信息丰富，采集自光滑的刀具表面 (硬质合金材质)，背景相对单一 (仅含刀具表面区域)；NEU 数据集图像分辨率为 200 pixel×200 pixel，细节压缩明显，采集自粗糙的钢材表面 (热轧板材)，背景包含轧制纹理、氧化皮覆盖等复杂干扰。

3.3 性能平均指标

使用平均准确度 mAP (mean average precision)、推理速度 (frames per second)、GFLOPS (giga floating-point operation per second) 及参数量 (parameters) 四项指标评价本文所提算法的有效性，如式 (17)~(20) 所示：

$$P=\frac{TP}{TP+FP}, \tag{17}$$

$$R=\frac{TP}{TP+FN}, \tag{18}$$

$$mAP=\frac{\sum_{i=1}^{n}\int_{0}^{1}P(R)\mathrm{d}R}{n}, \tag{19}$$

$$FPS=\frac{1}{r}, \tag{20}$$

式中：TP 表示正确识别的缺陷数量；FN 表示未识别的缺陷数量；FP 表示被错误识别的非缺陷数量；P 表示评价模型结果的精确度 (precision) 指标；R 表示评价模型结果的召回率 (recall) 指标；n 表示刀具缺陷的类别数。实验中涉及到的 mAP0.5 表示在 IoU 阈值为 0.5 时的平均精确度 (AP)，IoU 阈值是衡量预测框与真实框重叠程度的指标，取值范围从 0 到 1，当预测框与真实框的 IoU 大于等于 0.5 时，才算作是一次正确检测。GFLOPS 表示每秒 10 亿次的浮点运算数，常作为 GPU 性能参数，1 GFLOPS = 1000 MFLOPS；FPS 表示模型每秒能够处理的图片数；r 表示推理单张图片的时间。

3.4 消融实验

在本节中，采用最新的 YOLOv11s (small 版本) 作为基准模型，在精密刀具数据集上进行消融实验，论证所提算法嵌入到不同深度神经网络模型中的影响。

3.4.1 标准 AMPA 在网络中不同位置的影响

在目标检测的神经网络中，Backbone 主要用于提取图像特征，而 Neck 则用于筛选和优化这些特征以提高

检测性能。本文将 2.2 节提出的可嵌入 AMPA 模块集成到 YOLOv11 的 C3K2 模块中，构成 C3K2-AMPA 模块，分别置于深度神经网络的 Backbone 和 Neck 中，并记录其对目标检测性能的影响。基于此，本实验囊括三种情况：仅在 Backbone 特征提取模块中引入 AMPA 结构 (AMPA-Backbone，将 Backbone 中的 C3K2 模块替换为 C3K2-AMPA)、仅在 Neck 多尺度特征融合和优化模块中引入 AMPA 结构 (AMPA-Neck，将 Neck 中的 C3K2 模块替换为 C3K2-AMPA)，以及同时在两者中同时引入该结构 (AMPA-ALL)。

由表 1 可知，将 AMPA 加在 Backbone 后，mAP0.5 提升幅度最大，达到了 90.2%，较 Baseline 增加 2.4%，因为 Backbone 直接导入原始图像，担负从底层纹理到高层语义的大规模特征提取任务，此时加入注意力机制能显著改善整体特征质量。不过，由于 Backbone 在网络中规模较大，该方法会带来更明显的参数 (从 9.4 M 增加至 10.9 M) 和计算量 (GFLOPS 从 21.7 增至 23.1) 的上升，同时推理速度略有下降 (FPS 从 76 降至 69)。相比之下，将 AMPA 应用于 Neck (AMPA-Neck) 中，检测精度仅提升了 0.6%，表明 Neck 主要起到多尺度特征整合的作用，对特征本身的提炼相对有限，但引入 AMPA 产生的代价也较小 (参数和计算量增幅较小，FPS 稳定在 71 左右)。当在 Backbone 与 Neck 同时引入 AMPA (AMPA-ALL) 时，检测精度增加 1.9%，介于前两者之间，同时模型参数和计算量达到最大的 11.8 M 和 24.0 GFLOPS，速度降至 65 FPS。当 Backbone 已经大幅度提升特征质量后，Neck 再加入同样的注意力模块，常会碰到“冗余”或“性能饱和”的问题，即 Backbone 部分的改进对后端特征已有充分增强，Neck 再加同类模块无法再成倍放大提升。因此，即便 Backbone 与 Neck 同时使用 AMPA，整体性能提升会介于只在 Neck 使用与只在 Backbone 使用之间，难以简单叠加到更高的水平，同时还要付出更多的模型开销与推理时间。

3.4.2 SCAM 与 FSAM 并行与串行对模型的影响

为了进一步探索星通道注意力 (SCAM) 和频域空间注意力 (FSAM) 在神经网络中的最佳执行方式，本实验开展四种组合策略：i. 先应用 SCAM，再应用 FSAM (图 7(a))；ii. 先应用 FSAM，再应用 SCAM (图 7(b))；iii. 同步执行 FSAM 和 SCAM 后相加 (图 7(c))；iv. 同步

表 1 所提 AMPA 结构在神经网络中不同位置的效果

Table 1 The effects of the proposed AMPA structure at different positions in the neural network

Model	mAP0.5/%	Parameters/M	GFLOPS	FPS
Baseline	87.8	9.4	21.7	76
AMPA-Backbone	90.2 (+2.4)	10.9	23.1	69
AMPA-Neck	88.4 (+0.6)	10.3	22.6	71
AMPA-ALL	89.7 (+1.9)	11.8	24.0	65

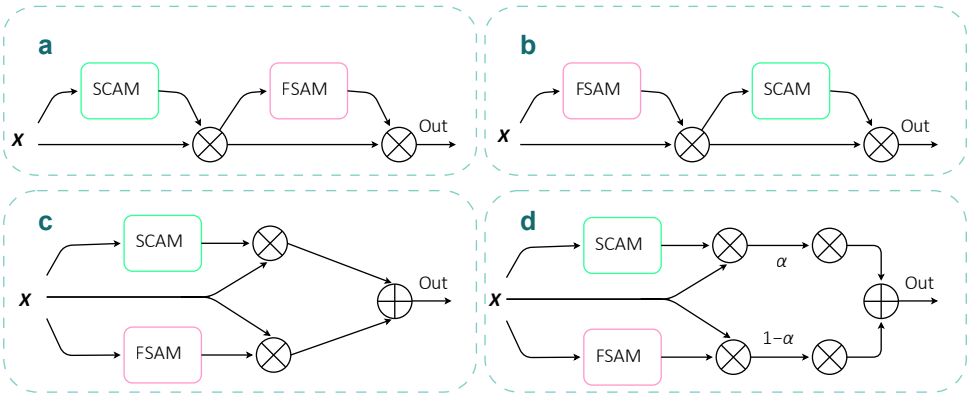


图 7 FSAM 与 SCAM 不同组合形式。(a) 先应用 SCAM，再应用 FSAM(顺序执行)；(b) 先应用 FSAM，再应用 SCAM (顺序执行)；(c) SCAM 与 FSAM 并行执行后直接相加；(d) SCAM 与 FSAM 并行执行，引入可学习权重 α 进行加权融合。

Fig. 7 Different combination forms of FSAM and SCAM. (a) SCAM followed by FSAM (sequential execution); (b) FSAM followed by SCAM (sequential execution); (c) Parallel execution of SCAM and FSAM with direct summation; (d) Parallel execution with learnable weight α for weighted fusion.

执行 FSAM 和 SCAM，并引入一个可学习参数α，使 SCAM 的输出乘以α，对 FSAM 的输出乘以 $(1-\alpha)$，最后加权融合 (图 7(d))。

以上四种融合策略尝试在表 1 表现较好的 AMPA-Backbone 实验基础上进行消融实验，深入分析不同注意力组合方式对模型性能的影响。

由表 2 可知，不同组合方式在精度指标上呈现明显差别。串行结构 (i, ii) 虽能带来性能提升，但两种注意力的作用具备先后顺序，而后者在一定程度上覆盖或重新加权前者的特征，使得两种注意力之间的互补性没有得到最大化发挥，最终仅能获得 0.9% 与 0.7% 的提升。而并行结构 (iii, iv) 中，SCAM 和 FSAM 同时处理同一输入，再进行融合，最大限度保留了两种注意力对特征的不同关注点。尤其是方式 (iv) 通过可学习参数进行自适应权衡，使网络能够根据输入特征和训练过程来自行分配通道注意力和频域空间注意力的相对比例，从而取得最高的检测精度 (增加 2.4%)。因此，方式 (iv) 的加权融合策略是最佳组合方式。

3.4.3 AMPA 在 Bottleneck 中不同位置的性能对比

所提的 AMPA 结构除了可以嵌入在 Backbone 和 Neck 发挥作用外，还可以嵌入在神经网络的 Bottleneck 结构的不同位置探索其对检测精度的贡献，主要包含：1) Bottleneck (图 8(a)) 标准 Bottleneck 结构 (无 AMPA)；2) AMPA-Bottleneck-Standard (图 8(b)) 将 AMPA 置于卷积之后、残差连接之前；3) AMPA-Bottleneck-PRE (图 8(c)) 将 AMPA 置于卷积之前。4) AMPA-Bottleneck-POST (图 8(d)) 将 AMPA 置于残差连接之后；5) AMPA-Bottleneck-PARALLEL (图 8(e)) 将 AMPA 模块与卷积平行后相加。

由表 3 可知，在 Bottleneck 中不同位置插入 AMPA 模块会直接影响 Bottleneck 操作和注意力提取之间的“交互时机”和“交互深度”，导致 mAP0.5 的不同提升效果。根据 AMPA-Bottleneck-PRE 结果，将注意力集中在 Bottleneck 之前，先对输入特征进行注意力加权，这

表 2　AMPA 并行与串行对模型的影响

Table 2　The impact of AMPA parallelism versus serialization on the model

Model	mAP0.5/%	FPS
Baseline	87.8	76
a	88.7 (+0.9)	71
b	88.5 (+0.7)	71
c	89.5 (+1.7)	72
d	90.2 (+2.4)	69

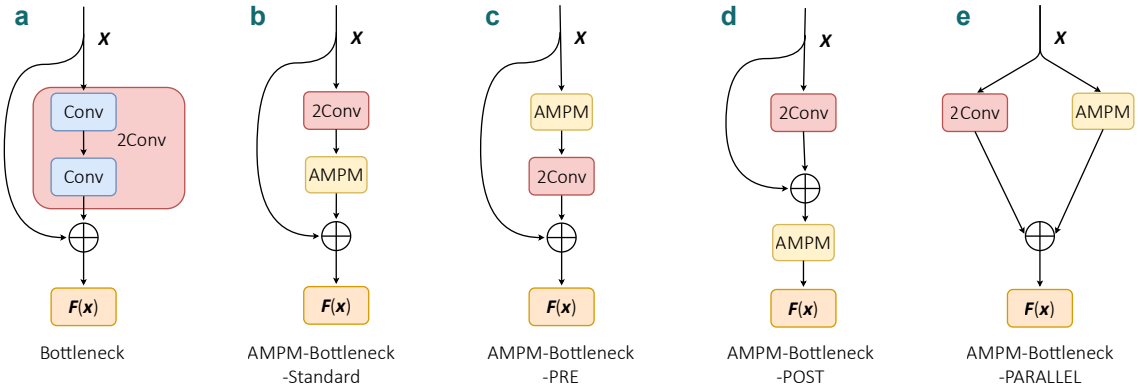


图 8　AMPA 模块在 Bottleneck 不同位置集成设计。(a) 标准 Bottleneck 结构 (无 AMPA)；(b) AMPA 置于卷积之后、残差连接之前；(c) AMPA 置于卷积之前 (AMPA-Bottleneck-PRE)；(d) AMPA 置于残差连接之后 (AMPA-Bottleneck-POST)；(e) AMPA 与卷积并行执行后相加 (AMPA-Bottleneck-PARALLEL)

Fig. 8　Integration design of the AMPA module at different positions within the bottleneck. (a) Standard Bottleneck without AMPA; (b) AMPA placed after convolution and before residual connection; (c) AMPA placed before convolution (AMPA-Bottleneck-PRE); (d) AMPA placed after residual connection (AMPA-Bottleneck-POST); (e) AMPA parallel to convolution with element-wise summation (AMPA-Bottleneck-PARALLEL)

表 3 AMPA 在 Bottleneck 中不同位置的性能对比

Table 3 Performance comparison of AMPA at different positions in the Bottleneck

Model	mAP0.5/%	FPS
Baseline	87.8	76
AMPA-Bottleneck-PRE	89.9 (+2.1)	72
AMPA-Bottleneck-POST	89.3 (+1.5)	71
AMPA-Bottleneck-PARALLEL	87.2 (−0.6)	68
AMPA-Bottleneck-Standard	90.2 (+2.4)	69

样有助于后续卷积更有效地聚焦于被强调的关键信息，从而实现显著的性能提升。根据 AMPA-Bottleneck-POST 结果，在 Bottleneck 最后才对输出特征进行注意力加权；虽然最终输出得到强化，但中间的卷积操作仍然在“未经过注意力筛选”的特征上进行，因而整体提升程度 (增加 1.5%) 不如 PRE 方案。根据 AMPA-Bottleneck-PARALLEL 结果，在并行结构中，“卷积分支”和“AMPA 分支”各自独立处理相同的输入特征，最后再融合处理，由于 AMPA 与卷积未能在同一路径上进行深度融合，导致二者互补性有限，甚至出现微弱下降 (减少 0.6%)。以上实验结果表明，将 AMPA 模块与卷积紧密耦合，使卷积和注意力在同一路径中深度交互、相互补充，在特征提炼和表达上实现最佳协同，因而获得最高的 mAP0.5 提升 (增加 2.4%)。总的来说，AMPA 模块出现在越能与卷积“深度协作”的位置，越能让注意力机制和卷积操作形成互补提升，从而带来更显著的精度提升。

3.4.4 AMPA 与主流注意力机制性能对比

为了进一步评估所提方法的性能，本文将其与现有几种主流注意力机制进行对比，包括 SE、CBAM 和自注意力机制 (self-attention)，它们在网络架构中的位置与 AMPA-Bottleneck-Standard 位置一致。

由表 4 可知，不同注意力机制在同等位置下对模型的提升幅度和计算代价差异明显，主要原因在于各注意力模型对特征的关注维度和方式不同：SE 仅关注通道维度，虽然参数开销较小但提升精度也少 (增加 0.4%)；自注意力 (self-attention) 在全局范围内建立特征依赖关系，理论上能够捕捉更丰富的上下文，但需要更多的计算量，使得精度提升 (增加 1.1%) 与参数增长、速度下降成正比；CBAM 同时处理通道和空间两方面注意力，相比单一路径的 SE，在强调空间信息上有更好的表现，带来 1.3% 的性能提升；AMPA-Bottleneck-Standard 不仅对通道进行再校准，还整合了频域空间注意力，深入挖掘特征在不同维度 (通道、空间、频域) 上的互补性，最终获得最高的 mAP0.5 提升 (增加 2.4%)。因此，不同注意力在通道、空间、频域等维度关注的程度和方式差异，导致特征表达能力与最终检测性能的差别；同时，越是多维度且更加灵活、耦合程度更深的注意力机制，越能实现显著的性能提升。AMPA 模块引入后，Bottleneck 结构能够更有效地捕获关键特征，提升模型在刀具表面缺陷检测任务中的表现和泛化能力。

3.5 不同检测算法实验对比

为了进一步验证本文改进算法的有效性，本文采用 YOLO 系列中的 YOLOv8[14]、YOLOv9[15]、YOLOv10[16] 和 YOLOv11[17] 作为对比模型，这四种模型整体结构相

表 4 AMPA 与其他注意力机制性能对比

Table 4 Performance comparison of AMPA with other attention mechanisms

Model	mAP0.5/%	Parameters/M	GFLOPS	FPS
Baseline	87.8	9.4	21.7	76
SE[4]	88.2 (+0.4)	9.5	21.8	72
Self-Attention[21]	88.9 (+1.1)	12.3	24.2	63
CBAM[6]	89.1 (+1.3)	10.1	22.1	70
AMPA-Bottleneck-Standard	90.2 (+2.4)	10.9	23.1	69

似，特征提取部分都含有 Bottleneck 模块，因此，本文将所提 AMPA 模块分别集成到以上四种模型中，验证所提方法的有效性。

3.5.1 在精密刀具数据集上进行对比实验

本文把所提出的 AMPA 模块分别嵌入 YOLOv8、YOLOv9、YOLOv10、YOLOv11 模型架构中，与传统方法在精密刀具数据集上进行了结果对比。为了更直观地展示 AMPA 模块的有效性，本文对 YOLOv11-AMPA 在刀具数据集上的检测结果进行了可视化，并加入了热力图对比 (图 9(a): YOLOv11；图 9(b): YOLOv11-AMPA)。图 9 展示了其在精密刀具数据集上的检测结果，结果中包含了预测框、缺陷种类、置信度以及热力图，方便对比 AMPA 模块对模型检测能力的影响。

表 5 展示了改进的 YOLO 算法与原始 YOLO 系列算法在精密刀具数据集上的实验结果比较，涵盖了 mAP0.5、FPS、GFLOPS 及参数量四个指标。

结合图 10 和表 5，针对微小尺度缺陷目标，改进后的模型在 mAP0.5 和更综合的 mAP0.5-0.95 指标上均表现出显著提升，充分验证了 AMPA 模块的有效性：在 mAP0.5 指标上，YOLOv8-AMPA 从 88.1% 提升至 90.8% (+2.7%)，YOLOv9-AMPA 从 87.8% 提升至 90.3% (+2.5%)，YOLOv10-AMPA 从 86.4% 提升至 88.5% (+2.1%)，

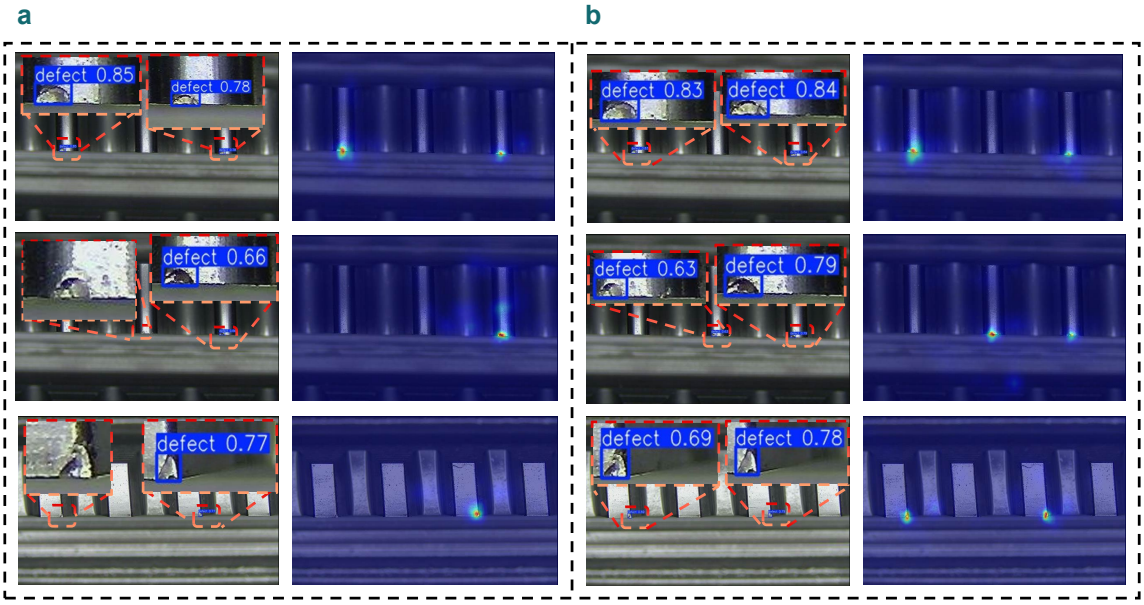


图 9 刀具数据集上的可视化结果。(a) YOLOv11；(b) YOLOv11-AMPA

Fig. 9 Visualization results on the cutting tool dataset. (a) YOLOv11; (b) YOLOv11-AMPA

表 5 本文改进 YOLO 算法与原始 YOLO 系列算法的实验结果比较

Table 5 This paper compares the experimental results of the improved YOLO algorithm with those of the original YOLO series algorithms

Model	mAP0.5/%	mAP0.5-0.95/%	Parameters/M	GFLOPS	FPS
YOLOv8[14]	88.1	48.8	11.1	28.8	72
YOLOv8-AMPA	90.8 (+2.7)	49.9 (+1.1)	12.8	30.3	64
YOLOv9[15]	87.8	47.6	7.3	27.6	70
YOLOv9-AMPA	90.3 (+2.5)	48.9 (+1.3)	9.1	29.1	63
YOLOv10[16]	86.4	47.9	8.0	24.8	83
YOLOv10-AMPA	88.5 (+2.1)	48.9 (+1.0)	9.6	26.8	75
YOLOv11[17]	87.8	48.7	9.4	21.7	76
YOLOv11-AMPA	90.2 (+2.4)	50.0 (+1.3)	10.9	23.1	69

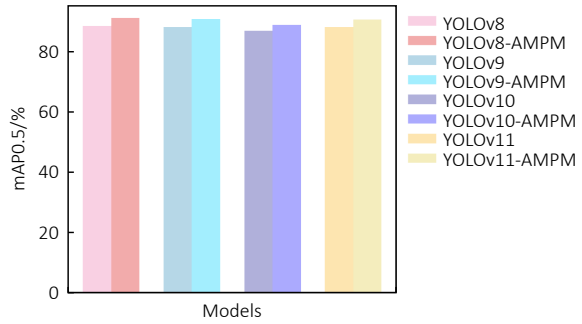


图 10 不同模型在刀具数据集上的精度对比条形图

Fig. 10 Accuracy comparison bar chart of different models on the tool dataset

YOLOv11-AMPA 从 87.8% 提升至 90.2% (+2.4%)，表明 AMPA 模块能有效增强模型对微尺度缺陷的识别能力。更重要的是，在反映不同 IoU 阈值 (0.5~0.95) 综合性能的 mAP0.5-0.95 指标上，改进后的模型同样表现优异：YOLOv11-AMPA 从 48.7% 提升至 50.0% (+1.3%)，YOLOv9-AMPA 从 47.6% 提升至 48.9% (+1.3%)，YOLOv8-AMPA 和 YOLOv10-AMPA 分别提升 1.1% 和 1.0%。这一结果进一步验证了 AMPA 模块的优越性—微尺度缺陷 (平均 20 像素) 的定位精度对 IoU 阈值敏感，在高严格度阈值 (如 *IoU*=0.9) 下，传统模型易因定位偏差导致检测失败，而 AMPA 通过 SCAM 的通道校准 (增强缺陷边缘通道) 和 FSAM 的频域-空间融合 (捕捉微小细节的精准位置)，使模型在不同 IoU 下均能保持稳健性能，尤其在严格阈值下仍能精准定位微小缺陷，这对工业检测中“既避免漏检，又保证定位精度”的需求至关重要。而效率下降的主要原因在于 AMPA 模块引入的星操作 (SCAM 中多路径卷积核的稀疏耦合) 和频域变换 (FSAM 中的 FFT 及频域特征处理)，增加了计算量 (如 YOLOv11-AMPA 参数量从 9.4 M 增至 10.9 M，GFLOPS 从 21.7 增至 23.1)，导致 FPS 从 76 降至 69(下降 7)。但从实际工业场景的性价比来看，该权衡具有合理性：改进后模型的 FPS(69) 仍远高于生产线在线检测的最低实时性要求 (通常≥30 FPS)，完全满足连续检测需求；而 mAP0.5 (+2.4%) 和 mAP0.5-0.95 (+1.3%) 的双重提升，直接降低了微尺度缺陷的漏检率和定位偏差率，其对生产安全性的增益 (避免因缺陷导致的加工事故) 远超过效率小幅下降的影响。此外，表 5 中其他模型 (如 YOLOv8-AMPA、YOLOv9-AMPA) 的结果也显示，mAP0.5 提升与 mAP0.5-0.95 提升的双重增益，与效率损耗 (FPS 下降 6~10) 的配比，在工业缺陷检测的高可靠性需求下具有实际应用价值。虽然改进带来了参数量和计算复杂度的增加，但双重精度指标的显著提升证明了方法的有效性，且在实际应用中仍然能够满足实时检测的需求。

3.5.2 在 NEU 数据集上泛化实验

为了验证所提方法的泛化性，本文在公共 NEU 数据集上进行了实验。与精密刀具数据集不同，NEU 数据集包含六个不同类别，这为模型泛化能力提供了更为复杂的测试环境。图 11 展示了 YOLOv11-AMPA 在 NEU 数据集上可视化检测结果。

在表 6 中，对改进的 YOLO 算法与原始 YOLO 系列算法在 NEU 数据集上的实验结果进行了比较。

由表 6 和图 12 可知，改进后的模型在 mAP0.5 和 mAP0.5-0.95 两项指标上均有提升，充分验证了 AMPA 模块在多类别钢材缺陷检测中的有效性：在 mAP0.5 指标上，YOLOv8-AMPA 从 76.9% 提升至 78.4% (+1.5%)，YOLOv9-AMPA 从 76.7% 提升至 78.1% (+1.4%)，YOLOv10-AMPA,YOLOv11-AMPA 分别提升 1.3%，表明 AMPA 在多类别缺陷检测中能有效增强模型的识别能力。更重要的是，在 mAP0.5-0.95 指标上，改进模型同样表现优异：YOLOv11-AMPA 提升 1.0%，YOLOv8-AMPA，YOLOv9-AMPA 和 YOLOv10-AMPA 分别提升 0.8%，0.7% 和 0.7%。这一结果印证了 AMPA 模块对多形态缺陷的稳健适配性—NEU 数据集中的裂纹 (细长纹理)、划痕 (局部结构) 等缺陷对定位精度敏感，高 IoU 阈值 (如 0.8~0.95) 下传统模型易因边缘定位偏差导致性能下降，而 AMPA 使模型在不同严格度的 IoU 标准下均能保持稳定性能。从效率来看，改进模型的参数量 (如 YOLOv8-AMPA 从 11.1 M 增至 12.8 M) 和 GFLOPS 略有增加，导致 FPS 下降 (如 YOLOv8-AMPA 从 70 降至 61)，但仍满足工业在线检测的实时性需求 (≥30 f/s)。结合 mAP0.5 和 mAP0.5-0.95 的双重提升，说明 AMPA 模块在增加可控计算成本的同时，不仅提高了单一阈值下的识别精度，更增强了模型在多场景、多标准下的综合检测能力，充分体现了其在多类别工业缺陷检测中的应用价值。综上，AMPA 模块通过融合通道、频域与空间特征，既提升了单一阈值下的检测精度，又增强了不同 IoU 标准下的稳健性，有效优化了多类别钢材表面缺陷的检测性能，验证了方法的有效性与实用性。

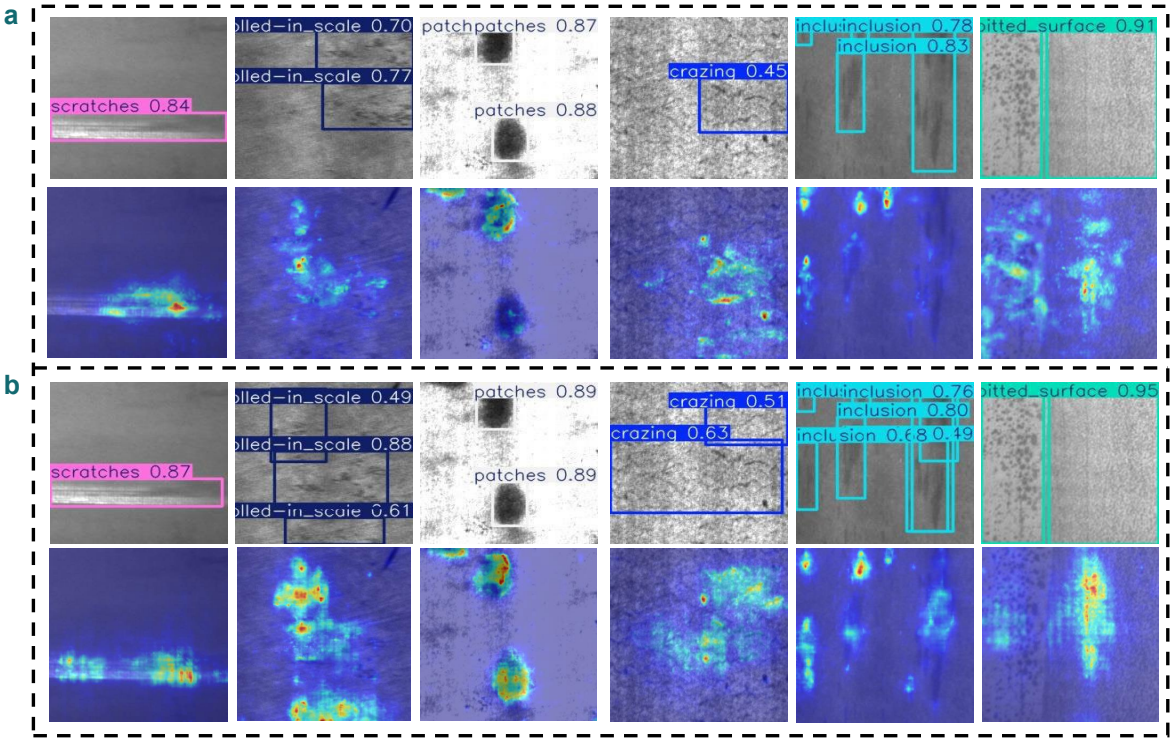


图 11 NEU 数据集上的可视化结果。(a) YOLOv11；(b) YOLOv11-AMPA

Fig. 11 Visualization results on the NEU dataset. (a) YOLOv11; (b) YOLOv11-AMPA

表 6 本文改进 YOLO 算法与原始 YOLO 系列算法的实验结果比较

Table 6 Comparison of experimental results between the improved YOLO algorithm and the original YOLO series algorithms

Model	mAP0.5/%	mAP0.5-0.95/%	Parameters/M	GFLOPS	FPS
YOLOv8[14]	76.9	43.4	11.1	28.8	70
YOLOv8-AMPA	78.4 (+1.5)	44.2 (+0.8)	12.8	30.3	61
YOLOv9[15]	76.7	42.4	7.3	27.6	66
YOLOv9-AMPA	78.1 (+1.4)	43.1 (+0.7)	9.1	29.1	59
YOLOv10[16]	71.5	40.4	8.0	24.8	79
YOLOv10-AMPA	72.8 (+1.3)	41.1 (+0.7)	9.6	26.8	71
YOLOv11[17]	74.8	42.5	9.4	21.7	71
YOLOv11-AMPA	76.1 (+1.3)	43.5 (+1.0)	10.9	23.1	65

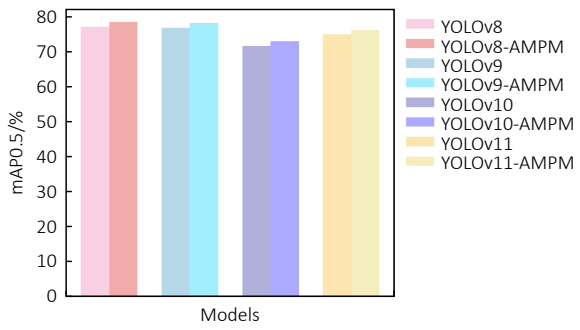


图 12 不同模型在 NEU 数据集上的精度对比条形图

Fig. 12 Comparison of accuracy of different models on the NEU dataset bar chart

4 结 论

本文针对现有注意力机制在精密刀具表面微小缺陷检测中存在的问题，提出了一种自适应多重并行注意力机制 (adaptive multi-parallel attention, AMPA) 模块。首先，将星通道注意力 (star-channel attention module, SCAM) 与频域-空间协同注意力 (frequency-spatial attention module, FSAM) 并行执行，结合可学习的动态权重，实现通道维度与频域-空间特征的自适应融合。为验证方法有效性，本文将 AMPA 模块分别嵌入主流 YOLO 系列模型 (YOLOv8 至 YOLOv11)，并在自建的

精密刀具数据集上进行评估，检测精度最高获得 2.7% 的提升。为了检验算法的泛化能力，进一步在钢材表面缺陷数据集 NEU 上进行实验，获得最高 1.5% 的精度提升。实验结果表明，AMPA 模块不仅显著优化了针对精密刀具的小目标缺陷检测效果，也能有效提升钢材表面多类别复杂缺陷的识别性能，展示了其在不同工业场景下增强特征表达和提升模型性能的广泛适用性。虽然本方法在微尺度单类刀具缺陷检测中表现出有效性，但在复杂工业场景中仍存在不足：对多类微尺度缺陷的适应性有待增强，面对极端光照、模糊等干扰时鲁棒性需提升，且轻量化部署下实时性仍有优化空间。未来将通过引入轻量化模块等设计进一步改进。

作者贡献

刘子豪：实验实施计划和论文撰写的设计；陈泉达：论文的撰写和实验实施；金友振：实验图像样本的采集和公开数据集的收集；何伟军：刀具识别技术方面的指导；鹿业波：对刀具缺陷成因的分析。

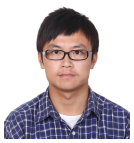
利益冲突

所有作者声明无利益冲突

参考文献

1. Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]//*Proceedings of the 26th International Conference on Neural Information Processing Systems*, 2012: 1097–1105.
2. Ji C, Wang H B, Li H H. Defects detection in weld joints based on visual attention and deep learning[J]. *NDT & E Int*, 2023, **133**: 102764.
3. Yeung C C, Lam K M. Efficient fused-attention model for steel surface defect detection[J]. *IEEE Trans Instrum Meas*, 2022, **71**: 2510011.
4. Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//*Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 7132–7141. https://doi.org/10.1109/CVPR.2018.00745.
5. Wang F, Jiang M Q, Qian C, et al. Residual attention network for image classification[C]//*Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 6450–6458. https://doi.org/10.1109/CVPR.2017.683.
6. Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//*Proceedings of the 15th European Conference on Computer Vision*, 2018: 3–19. https://doi.org/10.1007/978-3-030-01234-2_1.
7. Guo M H, Xu T X, Liu J J, et al. Attention mechanisms in computer vision: a survey[J]. *Comput Visual Media*, 2022, **8**(3): 331−368.
8. Wang Q L, Wu B G, Zhu P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//*Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 11531–11539. https://doi.org/10.1109/CVPR42600.2020.01155.
9. Hu J, Shen L, Albanie S, et al. Gather-excite: exploiting feature context in convolutional neural networks[C]//*Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2018: 9423–9433.
10. Wang X L, Girshick R, Gupta A, et al. Non-local neural networks[C]//*Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 7794–7803. https://doi.org/10.1109/CVPR.2018.00813.
11. Chen H X, Du Y Z, Fu Y Q, et al. DCAM-Net: a rapid detection network for strip steel surface defects based on deformable convolution and attention mechanism[J]. *IEEE Trans Instrum Meas*, 2023, **72**: 5005312.
12. Wu C, Lei S C, Xu H W, et al. Multi-scale fusion attention network for industrial surface defect classification[C]//*2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2024: 594–599. https://doi.org/10.1109/CSCWD61410.2024.10580645.
13. Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//*Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 7464–7475. https://doi.org/10.1109/CVPR52729.2023.00721.
14. Yu X Y, Yan P, Zheng S K, et al. YOLOv8-WTDD: multi-scale defect detection algorithm for wind turbines[J]. *J Supercomput*, 2025, **81**(1): 32.
15. Wang C Y, Yeh I H, Mark Liao H Y. YOLOv9: learning what you want to learn using programmable gradient information[C]//*18th European Conference on Computer Vision*, 2025: 1–21. https://doi.org/10.1007/978-3-031-72751-1_1.
16. Wang A, Chen H, Liu L H, et al. YOLOv10: real-time end-to-end object detection[C]//*Proceedings of the 38th International Conference on Neural Information Processing Systems*, 2025: 3429.
17. Khanam R, Hussain M. YOLOv11: an overview of the key architectural enhancements[Z]. arXiv: 2410.17725, 2024. https://doi.org/10.48550/arXiv.2410.17725.
18. Ma X, Dai X Y, Bai Y, et al. Rewrite the stars[C]//*Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024: 5694–5703. https://doi.org/10.1109/CVPR52733.2024.00544.
19. Chen Q D, Liu Z H, Jin Y Z, et al. Research on surface defect cognition neural network model for precision cutter tool[J]. *J Optoelectron·Laser*, 2025.
陈泉达, 刘子豪, 金友振, 等. 精密刀具表面缺陷认知神经网络模型研究[J]. 光电子•激光, 2025.
20. Song K C, Yan Y H. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects[J]. *Appl Surf Sci*, 2013, **285**: 858−864.
21. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017: 6000–6010.

作者简介



【通信作者】刘子豪 (1988-)，男，副教授，博士，从事工业产品在线无损检测技术研究。

E-mail：lzh@zjxu.edu.cn

Adaptive multiscale parallel attention mechanism for micro-scale tool defect detection method

Liu Zihao[1*], Chen Quanda[2], Jin Youzhen[1,3], He Weijun[4], Lu Yebo[5]

[1] School of Artificial Intelligence, Jiaxing University, Jiaxing, Zhejiang 314001, China;
[2] Zhejiang Sci-Tech University, Hangzhou, Zhejiang 310018, China;
[3] School of Electrical Automation and Information Engineering, Tianjin University, Tianjin 300072, China;
[4] EST Tools Co., Ltd., Jiaxing, Zhejiang 314300, China;
[5] School of Mechanical Engineering, Jiaxing University, Jiaxing, Zhejiang 314001, China

Abstract

Objective Precision cutting tools, as the core components for the construction of national key equipment, have attracted growing attention to their quality. Micro-scale defects on the tool surface, such as cracks and bumps, expand rapidly under high-intensity machining environments. The defects at the damaged parts will be continuously amplified, which affects the performance and safety of end products, also severely restricts the efficiency of product processing and manufacturing. The defects on the surface of cutting tools exhibit much smaller morphological and texture features, and this greatly increases the difficulty of automated detection. Therefore, the development of efficient and accurate micro-scale defect detection technology for tool surfaces has become a key factor in intelligent manufacturing. Although a large number of research methods related to different attention mechanisms have been proposed in the published literature. However, when these methods are directly applied to detect tiny appearance defects on the tool surface, the following two key problems will arise: 1) the problem of difficulty in identifying micro-scale defects. When existing attention mechanisms deal with the morphological diversity of tiny defects on the surface of precision cutting tools, they lack effective coupling and expression of channel features, resulting in the inability to accurately identify key information. 2) The problem of defect texture diversity. When facing the diverse texture features of complex defects on the surface of precision cutting tools, traditional methods fail to make full use of frequency domain information to capture subtle changes in defects, which affects the accuracy of detection.

Methods To address these issues, a precision tool defect detection model is proposed based on adaptive multi-parallel attention (AMPA). The core architecture of AMPA consists of two functionally complementary sub-modules: the star channel attention module (SCAM) and the frequency-spatial collaborative attention module (FSAM). Adopting a parallel processing mechanism, the two modules mine defect features from different dimensions: SCAM focuses on the channel dimension, strengthens the nonlinear coupling between channels through star operations, and captures the key channel responses of micro-scale defects; FSAM integrates the frequency and spatial dimensions, extracts full-frequency domain information using fast Fourier transform (FFT), and fuses spatial structural features to address defect localization under complex textured backgrounds. This parallel design can maximize the preservation of the uniqueness of the two types of features-SCAM and FSAM have low feature mutual information (low redundancy), and the total information entropy is higher than that of serial structures (preventing subsequent modules from overwriting preceding features), laying a foundation for efficient feature fusion. From the aspects of actual processing lists as follows. Initially, a star-channel attention module (SCAM) is constructed, utilizing star pperations to enhance nonlinear coupling between channels, thereby improving the representation ability of key channel features. Secondly, a frequency-spatial attention module (FSAM) is designed to combine frequency domain and spatial information, effectively capturing low-frequency, mid-frequency and high-frequency features as well as spatial structural information in images. Finally, the fusion weights of these two modules are dynamically adjusted through learnable parameters to achieve adaptive feature fusion, enhancing the detection performance.

Results and Discussions The dual improvements of the proposed methods demonstrate that while incurring a controlled computational cost, the AMPA module not only enhances detection accuracy under a single threshold, but also strengthens the model's comprehensive detection capability across multiple scenarios and evaluation criteria, fully embodying its application value in multi-category industrial defect detection. Experimental results demonstrate that integrating the proposed method into mainstream YOLOv11 model can achieve a mAP0.5 of 90.2% on a self-constructed precision tool defect dataset, achieving an improvement of 2.4% compared with the traditional YOLOv11. This result validates the model's effectiveness in detecting micro-scale target defects on tool surfaces. Additionally, on the

Foundation item: National Natural Science Foundation of China (62374074), Zhejiang Provincial "Top-Tier & Leading Talents" R&D Program (2024C04028), Jiaxing Municipal Public Welfare Research Program (SQGY202400009), University-Enterprise Cooperation Project (00523144), Haiyan Key R&D Program (2024ZD03), Talent Project of Jiaxing University (CD70623008)

*Correspondence: Liu Z H, E-mail: lzh@zjxu.edu.cn

publicly available NEU steel surface defect dataset, the proposed method can reach a mAP0.5 of 76.1%, surpassing traditional YOLOv11 by 1.3%, demonstrating the model's robust generalization capability. In summary, by fusing channel, frequency-domain and spatial features, the AMPA module not only improves detection accuracy under a single threshold, but also boosts robustness across different IoU criteria, effectively optimizing the detection performance for multi-category steel surface defects and verifying the effectiveness and practicality of the proposed method.

Conclusions Aiming at the problems of existing attention mechanisms in the detection of micro-defects on the surface of precision cutting tools, an adaptive multi-parallel attention (AMPA) module is proposed in this paper. First, the star-channel attention module (SCAM) and the frequency-spatial attention module (FSAM) are executed in parallel, combined with learnable dynamic weights, to realize the adaptive fusion of channel-dimensional features and frequency-spatial features. To verify the effectiveness of the proposed method, the AMPA module is embedded into mainstream YOLO series models (from YOLOv8 to YOLOv11) respectively, and evaluated on a self-built precision cutting tool dataset, achieving a maximum improvement of 2.7% in detection accuracy. For testing the generalization ability of the algorithm, further experiments are conducted on the NEU steel surface defect dataset, with a maximum accuracy improvement of 1.5% obtained. Experimental results show that the AMPA module not only significantly optimizes the detection performance of small-target defects on precision cutting tools, but also effectively enhances the recognition performance of multi-category complex defects on steel surfaces, demonstrating its wide applicability in enhancing feature representation and improving model performance across different industrial scenarios. Although the proposed method shows effectiveness in the detection of micro-scale single-category tool defects, it still has limitations in complex industrial scenarios: its adaptability to multi-category micro-scale defects needs to be enhanced, its robustness against interferences such as extreme illumination and blurring should be improved, and the real-time performance under lightweight deployment still has room for optimization. Future research will further improve the method by introducing lightweight module designs and other strategies.

Keywords: high-precision tool; attention mechanism; defect detection

Liu Z H, Chen Q D, Jin Y Z et al. Adaptive multiscale parallel attention mechanism for micro-scale tool defect detection method[J]. *Opto-Electron Eng*, 2026, **53**(1): 250173; DOI: 10.12086/oee.2026.250173

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文