

融合注意力机制的3D打印缺陷检测算法

张俊杰^{1,2}, 沈震^{1,3}, 方启航^{1,2}, 董西松^{1,3}, 王迪⁴, 熊刚^{1,3}

(1. 中国科学院自动化研究所多模态人工智能系统全国重点实验室, 北京 100190; 2. 中国科学院大学人工智能学院, 北京 100049; 3. 中国科学院自动化研究所北京市智能化技术与系统工程技术研究中心, 北京 100190; 4. 华南理工大学机械与汽车工程学院, 广东 广州 510641)

摘要: 近年来, 3D 打印技术在越来越多的行业中发挥着重要作用, 但是, 其作为一种新技术, 相对于传统制造, 在打印过程中出现缺陷的情形较多, 这些缺陷会显著影响最终产品的性能。鉴于 3D 打印部件通常具有复杂和高度优化的几何形状, 传统的检测技术难以满足其对精确性和效率的要求。为解决这一难题, 提出了一种基于改进版 YOLOv5 的 3D 打印缺陷检测算法。该算法对 YOLOv5 模型进行了深入改进, 通过替换损失函数和引入注意力机制, 成功实现了模型的轻量化。新设计的检测系统具有参数规模较小、推理速度快、检测精度高和鲁棒性较好的特点。与原始的 YOLOv5s 模型相比, 改进后的轻量化模型在检测精度上达到了 94.2%, 同时在参数规模和数据大小上几乎减半。这一进步不仅提升了检测效率, 还为 3D 打印缺陷检测与故障诊断提供了一种有效的技术手段。

关键词: 3D 打印; 缺陷检测; 深度学习; YOLOv5; 注意力机制

中图分类号: TP391.41

文献标志码: A

doi: 10.11959/j.issn.2096-3750.2025.00464

A 3D printing defect detection algorithm incorporating attention mechanism

ZHANG Junjie^{1,2}, SHEN Zhen^{1,3}, FANG Qihang^{1,2}, DONG Xisong^{1,3}, WANG Di⁴, XIONG Gang^{1,3}

1. National Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

2. School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China

3. Beijing Engineering Research Center of Intelligent Technology and Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

4. School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510641, China

Abstract: In recent years, 3D printing technology has played an increasingly important role in a growing number of industries. However, as a relatively new technology, it tends to exhibit more defects during the printing process compared to traditional manufacturing methods. These defects can significantly impact the performance of the final product. Given that 3D printed parts typically have complex and highly optimized geometric shapes, traditional detection technologies struggle to meet the demands for precision and efficiency. To address this challenge, this paper introduces a 3D printing defect detection algorithm based on an improved version of YOLOv5. The algorithm makes extensive refinements to the YOLOv5 model, achieving model lightweighting by replacing the loss function and introducing an attention mechanism. The newly designed detection system is characterized by a smaller parameter scale, rapid inference speed, high detection accuracy, and strong robustness. Compared to the original YOLOv5s model, the improved lightweight model has achieved a detection accuracy of 94.2%, and has nearly halved the parameter scale. This advancement not only enhances detection efficiency but also provides an effective technical solution for 3D printing defect detection and fault diagnosis.

收稿日期: 2024-11-15; 修回日期: 2025-01-08

通信作者: 董西松, xisong.dong@ia.ac.cn

基金项目: 国家自然科学基金资助项目 (No. 92267103, No. 92360307, No. 62461160259)

Foundation Items: The National Natural Science Foundation of China (No. 92267103, No. 92360307, No. 62461160259)

Key words: 3D printing, defect detection, deep learning, YOLOv5, attention

0 引言

3D打印技术作为一种创新的增材制造方式，正在引发一场工业变革。它通过逐层堆叠材料来创建三维物体，已经在制造业、医疗、建筑、教育等多个领域得到广泛应用。3D打印不仅加快了产品从设计到原型的进程，还通过按需定制生产，显著提高了生产效率和降低了成本，推动了产业创新。

尽管3D打印技术具有巨大潜力，但其制造过程中的复杂性也带来了挑战。使用不当的材料、不适宜的打印参数设置、打印机的故障或环境因素，都可能导致打印过程中出现孔洞、裂纹等缺陷，影响成品的质量和性能。

因此，检测3D打印过程中的实时情况尤为重要，及时发现缺陷并立即报警，能很大程度上减轻材料的浪费，并可以为分析缺陷形成的原因提供实时性材料。对打印过程中的缺陷进行检测的研究对本行业的发展和提高成品性能具有重要意义。目前，在缺陷检测方向的主要研究方向见表1。

在现代工业检测领域，工业互联网技术、计算机视觉和机器学习已成为提升检测精度和效率的关键技术。通过集成预测特征层、改进聚类算法、上下文增强模块和优化的相似度量方法^[1-3]，S-YOLOv5s检测精度实现了近5%的提升。深度卷积和编码技术在故障预测与诊断中的应用^[4]，为信号采集和云端故障检测提供了新的解决方案。文献[2-5]提到的各种深度学习模型和方法，显著提高了从轮辋焊缝

到PCB缺陷的检测能力。此外，机器学习技术^[4]通过分析工业互联网数据，有效提升了电网负荷管理和PCB缺陷检测的精度。这些研究成果不仅展示了跨学科技术融合潜力，也为3D打印等新兴制造技术的缺陷检测提供了宝贵的参考和启示。

尽管目标检测技术在多个工业领域取得了显著进展，如在金属材料表面缺陷检测、印制电路板(PCB, printed circuit board)瑕疵识别以及手工艺品质量评估中展现出较高的准确性和效率，但在3D打印这一新兴且快速发展的领域，现有技术仍面临诸多挑战。在工业互联网^[5]中，主要针对传统制造业产品表面缺陷进行检测，此类缺陷的特点是多样性较低，形状样式较为单一，对模型的泛化能力和鲁棒性要求不高。然而，3D打印过程中可能出现的缺陷类型多样，包括但不限于孔洞、裂纹、层间剥离等，这些缺陷的检测需要对模型的几何形状和打印过程有深刻的理解。图1展示了3D打印过程中可能出现的4中情况，图1(b)与图1(d)在形状、面积和周围纹理等方面差异巨大，对模型的抽象特征提取能力和鲁棒性提出了更高的要求。现有的目标检测技术多针对特定目标和环境优化，缺乏对3D打印特有缺陷的泛化能力，难以实现精准识别。

为了提升3D打印缺陷检测的精度和效率，本文采用了以下几种关键技术。

(1) 替换原始的交并比(IoU, intersection over union)损失函数。传统的IoU损失函数在处理边界框回归时存在局限性，尤其是在检测框与真实框没

表1 在缺陷检测方向的主要研究方向

研究方向	核心技术
工业互联网技术	文献[1]通过预测特征层、K-means++聚类和相似度量改进工业互联网技术，提高S-YOLOv5s检测精度4.92% 文献[2]结合工业互联网和计算机视觉，提升轮辋焊缝检测效率和精度 文献[3]应用深度卷积和编码于故障预测和诊断，实现信号采集和云端故障检测
计算机视觉	文献[4]结合工业互联网和深度学习，提出改进的目标检测模型 文献[5]提出视觉显著性和阈值融合的缺陷分割方法 文献[6]自动检测和分类刀具表面缺陷 文献[7]结合卷积神经网络和支持向量机，提出纺织品深层特征提取和缺陷分类方法 文献[8]基于TensorFlow的更快速区域卷积神经网络算法，用于铆钉缺陷检测 文献[9]提出改进网络结构和优化损失函数的缺陷检测算法，提高检测精度 文献[10]使用ResNet50和特征金字塔网络检测印刷电路板(PCB, printed circuit board)小缺陷 文献[11]提出多注意力机制融合多缺陷检测方法 文献[12]提出基于自适应注意力融合的AC-YOLO网络，用于绝缘子多缺陷检测
机器学习	文献[13]利用工业互联网数据，建立人工智能模型，提高电网负荷管理和PCB缺陷检测精度

有重叠的情况下。为此，本文引入了多种改进的IoU损失函数，包括完全交并比（CIoU, complete intersection over union）、距离交并比（DIOU, distance intersection over union）、增强交并比（EIoU, enhanced intersection over union）和平方交并比（SIoU, square intersection over union）。这些改进的损失函数不仅能够更精确地度量检测框与真实框之间的关系，还能加速模型的收敛，提升检测精度。实验结果显示，CIoU在本文数据集中表现最佳，使得平均精度（mAP, mean average precision）提高了近1%。

(2) 引入注意力机制。3D打印缺陷的多样性和复杂性要求模型能够有效关注图像中的关键区域。为此，本文在改进的YOLOv5模型中引入了3种注意力机制：具有自适应加权多输入功能的简单注意力模块（SimAM, simple attention module with adaptively-weighted multi-inputs）机制、压缩激励注意力（SE-Attention, squeeze-and-excitation attention）机制和混洗注意力（Shuffle-Attention）机制。这些注意力机制通过增强模型对重要特征的关注，提升了模型的泛化能力和检测性能。特别是SimAM，由于其无须额外参数且易于集成到现有卷积神经网络中，表现出色，成为本文的最佳选择。

(3) 模型轻量化。为了满足实际应用中的速度和模型大小需求，本文采用了Ghost Net来实现模型轻量化。传统的YOLOv5模型结构较为复杂，导致推理过程中的计算量较大。通过将主干网络和头部网络替换为Ghost Net模块，本文显著减少了模型的参数规模和计算量，同时保持了高检测精度。具体来说，改进后的模型参数规模从 7.6×10^6 减少到 3.6×10^6 ，数据大小从14.4 MB减少到7.7 MB，推理速度也得到了显著提升。

通过上述方法的结合，本文构建了一种高效、轻量化的3D打印缺陷检测系统，能够在保证高检测精度的同时，满足实际应用中的实时性和资源限制要求。

1 方法

1.1 数据预处理

本文使用了中国科学院自动化研究所实验室3D打印机的数据集，包含故障、严重错误和损毁、正常4种情况（种类界定标准及数量见表2）。

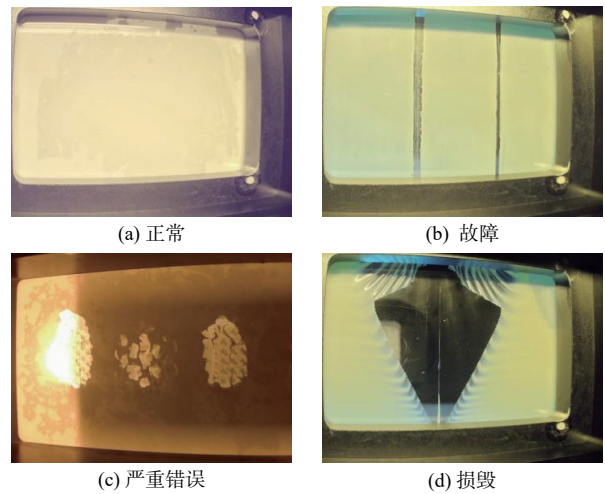


图1 3D打印过程中可能出现的4中情况

表2 种类界定标准及数量

图片种类	界定标准	处理方法	数量
故障	出现1~2道不显眼划痕	目标检测	190
严重错误	连续5张以上划痕位置相同	报警暂停	94
损毁	离型膜脱落或大面积空洞	报警暂停	9
正常	无异常	正常打印	260

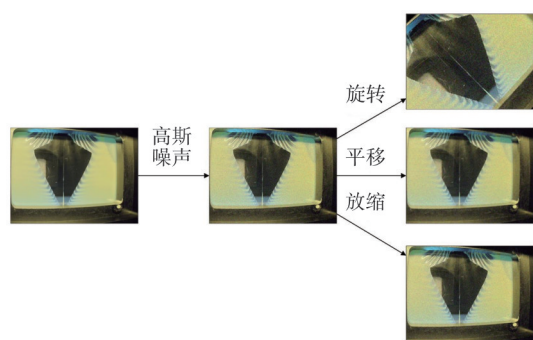
数据增强是指在机器学习和深度学习中通过对原始数据进行一系列变换来生成新的训练样本的技术。数据增强采用了颜色变换，向数据集中加入随机高斯噪声以提高模型鲁棒性（数据集增强前后对比如图2所示）。这种方式不仅增加了数据集的多样性，还增强了模型在不同光照条件和环境下的泛化能力。

设像素点在图片坐标轴上的坐标为 $P(x_0, y_0)$ ，与原点 O 的距离为 r ，连线与 X 轴的夹角为 α ，使其绕轴旋转角 β 后的新坐标为 $P(x, y)$ ，对图像进行旋转操作的数学原理如下

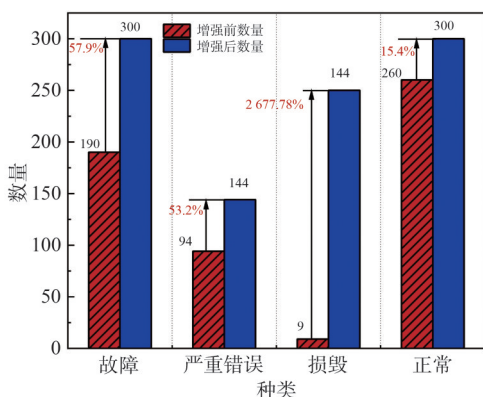
$$\begin{cases} x_0 = r \cdot \cos \alpha \\ y_0 = r \cdot \sin \alpha \\ x = \cos(\alpha - \beta) \\ y = \sin(\alpha - \beta) \end{cases} \quad (1)$$

$$\begin{bmatrix} 1 \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} \cos \beta & \sin \beta & 0 \\ -\sin \beta & \cos \beta & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} \quad (2)$$

通过几何变换的处理方法在数据增强中应用最简单也最多，但是此方法本质上并没有改变图像内容，容易造成数据样本不丰富、鲁棒性下降等影响。本文在几何变换的基础上采用了颜色变换进行



(a) 数据增强方法



(b) 数量对比

图2 数据集增强前后对比

数据增强，向数据集中加入随机高斯噪声，一定程度上可以提高模型鲁棒性。

1.2 网络架构

1.2.1 C3 模块

YOLOv5中的C3模块示意图如图3所示，它通过跨阶段部分网络增强了深度学习神经网络的感受能力和特征表征能力。C3模块是YOLOv5中一个核心的特征提取结构。它由两个主要分支构成：一是包含多个Bottleneck模块的主分支，用于深入提

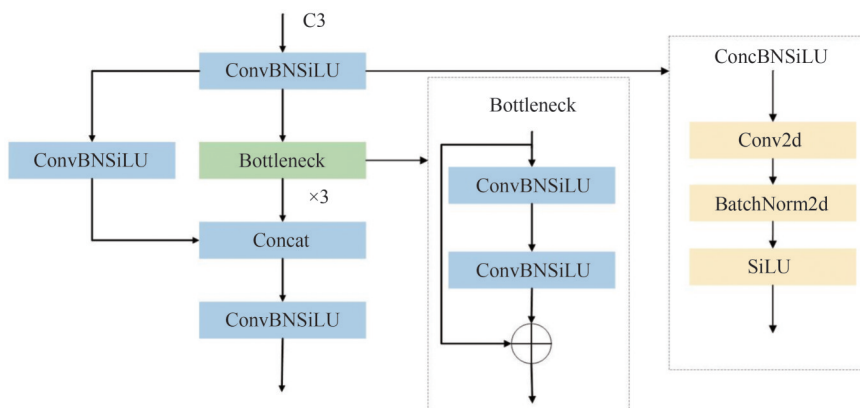


图3 C3模块示意图

取特征；二是简单的快捷分支，用于保留原始信息。两个分支的输出通过Concat（拼接）操作进行融合，随后再经过一个ConvBNSiLU模块（包含Conv2d、BatchNorm2d和SiLU激活函数）。这种设计结合了密集特征学习和残差连接，有效地减少了计算量，同时增强了网络的特征表示能力。

1.2.2 SPPF 模块

快速空间金字塔池化（SPPF, spatial pyramid pooling-faster）模块是一种高效的多尺度特征融合技术。它通过串联的Maxpooling操作，将输入的特征图在不同尺度下进行池化，从而生成多尺度的特征。这些特征随后与原始特征一起被拼接起来，形成一个固定长度的特征表示。这种设计比传统的空间金字塔池化更快，能够高效地处理不同尺寸的输入图像，为后续的全连接层提供鲁棒的特征。SPPF模块示意图如图4所示。

1.3 替换IoU损失函数

IoU是一种用于量化测定指定数据集中对目标物体检测精度的函数，损失函数用于边界盒预测，它将预测盒的4个边界作为一个整体单位进行回归。一般情况下，凡是输出中存在一个预测框的任务都可以使用该种函数^[13]。

广义交并比（GIoU, generalized intersection over union）相较于IoU能在更广义的层面上进行计算，其创新点在于通过引入一个广义的版本作为一个新的损失和一个新的度量标准来解决IoU的弱点，巧妙地解决了检测框和真实框没有重叠的问题^[14]。

GIoU判断原理如图5所示，预测框的面积记为A，真实框的面积记为B，最小外接框的面积记为C，计算最小外接框中除去预测框与真实框的面积再除以最小外接框的面积，这个比值即GIoU

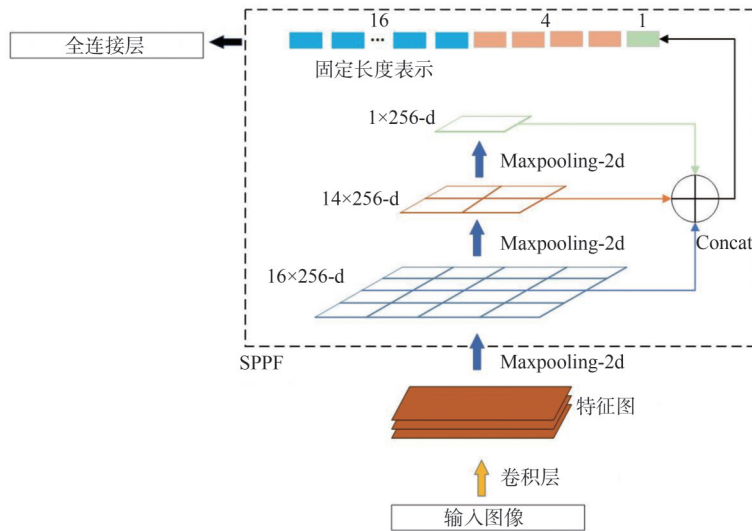


图4 SPPF模块示意图

$$L_{GIoU} = 1 - IoU + \frac{|C - (A \cup B)|}{|C|} \quad (3)$$

其中，GIoU损失函数是对传统IoU损失的有效改进。它在衡量预测框与真实框重叠程度的同时，引入了最小外接框的概念来解决两框无交集时IoU损失无法优化的缺陷。具体来说，GIoU损失不仅包含常规的1-IoU项，还添加了最小外接框中不包含预测框和真实框的区域的面积。当两框相距较远时，这一项会增大损失值，促使预测框向真实框移动，从而显著提升了边界框回归的稳定性和准确性。

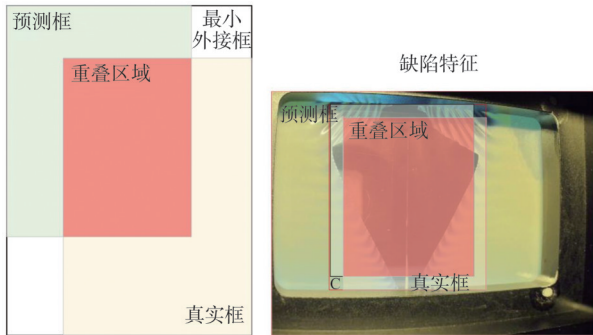


图5 GIoU判断原理

DIoU考虑GIoU的缺点，其计算的不再只是IoU的面积，而是增加了针对预测框与真实框之间的欧氏距离的惩罚项。相较于GIoU，DIoU通过结合预测框和真实框之间的归一化距离，使真实框的回归更加稳定。DIoU在处理中心点偏移较大的缺陷时表现出色，如当缺陷出现在图像边缘或角落时，DIoU能够更快地收敛，提高检测速度^[15]。DIoU计算原理如图6所示。

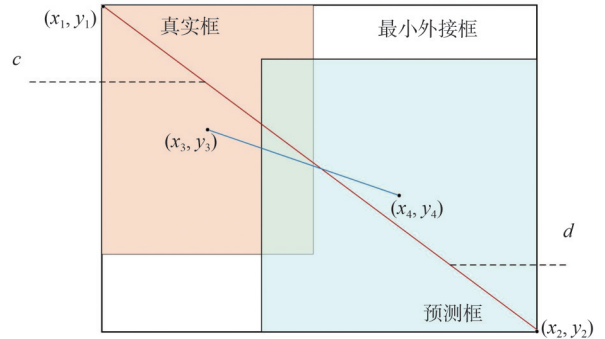


图6 DIoU计算原理

DIoU比GIoU更加符合真实框的回归机制，更加符合现实世界中生物对目标的位置的捕捉过程，使真实框的回归变得更加稳定

$$DIoU = \frac{\rho(A_{ctr}, B_{ctr})^2}{c^2} \quad (4)$$

其中， ρ 为两点之间的欧氏距离， $\rho(A_{ctr}, B_{ctr})^2$ 为中心点距离平方， c 为最小外接框对角线长度。

DIoU考虑了重叠面积和中心点的距离，当真实框包含预测框时，其直接度量的是两个框的距离，因此相较于GIoU，DIoU更具有收敛性。

CIoU在DIoU的基础上额外增加了预测框的尺度的损失，通过提高高和宽的损失，促进了真实框的稳定性进一步提高^[15]，避免了像IoU和GIoU中出现的发散问题。CIoU特别适用于处理那些长宽比差异较大的缺陷，如细长的裂缝或不规则的孔洞。CIoU能够敏锐地识别这些差异，从而提高检测精度

$$L_{CIoU} = 1 - IOU(A, B) + \frac{\rho(A_{ctr}, B_{ctr})^2}{c^2} + av \quad (5)$$

$$v = \frac{4}{\pi^2} \times (\tan^{-1}(\frac{w^{gt}}{h^{gt}}) - \tan^{-1}(\frac{w}{h}))^2 \quad (6)$$

$$\alpha = \frac{v}{(1 - \text{IoU}) + v} \quad (7)$$

其中, w^{gt} 、 h^{gt} 分别为真实框的宽和高, w 、 h 分别为预测框的宽和高, av 为动态加权的宽高比惩罚项, a 是一个权重系数, v 用于衡量宽高比一致性。会根据当前 IoU 动态调整, 使得模型在训练初期更注重宽高比的对齐。CIoU 损失函数是 IoU 损失的改进版, 它综合考量了目标检测框的重叠面积、中心点距离和宽高比 3 个关键因素。这种设计使得 CIoU 损失函数比传统 IoU 更全面, 有助于模型实现更精准、更稳定的边界框回归。

EIoU 主要包括 3 个部分, IoU 损失 L_{IoU} 、距离损失 L_{dis} 、横框比损失 L_{asp} , 其相对于 CIoU 的创新点在于引入了横框比损失函数, 该损失函数直接量化地描述了预测框与真实框的差异, 相较于 CIoU, EIoU 的收敛速度与定位效果更胜一筹^[6]。对于 3D 打印缺陷检测, EIoU 特别适合处理那些形状不规则且角度复杂的缺陷, 如裂纹或分层剥离。EIoU 能够快速调整预测框的角度和位置, 使其更接近真实框, 从而提高定位精度

$$L_{\text{EIoU}} = L_{\text{IoU}} + L_{\text{dis}} + L_{\text{asp}} = 1 - \text{IoU} + \frac{\rho(A_{\text{ctr}}, B_{\text{ctr}})^2}{c^2} + \frac{\rho(w, w^{gt})^2}{c_w^2} + \frac{\rho(h, h^{gt})^2}{c_h^2} \quad (8)$$

其中, c_w 和 c_h 分别为最小外接框的宽和高。

SIoU 考虑期望回归之间向量的角度, 重新定义角度惩罚度量, 它可以使预测框快速漂移到最近的轴, 随后则只需要回归一个坐标 (X 或 Y), 这有效地减少了自由度的总数^[7]。SIoU 判断原理如图 7 所示。

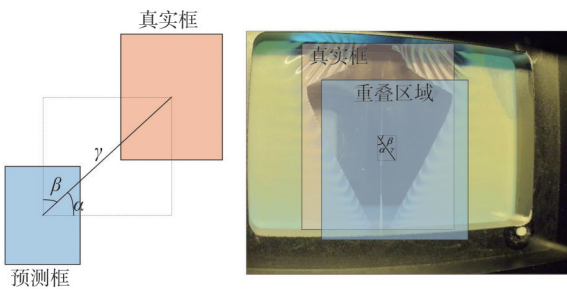


图7 SIoU判断原理

当预测框与真实框在 x 轴或者 y 轴上中心点对齐时, 即图 7 中 α 角为 0 或者 β 角为 0 时, 其角度损失函数为

$$\Delta = 1 - 2 \times \sin(\sin^{-1}\theta - \frac{\pi}{4}) \quad (9)$$

如果返回值为 0, 则代表预测框与真实框的角度损失为 0, 在角度维度高度重合, 为预测框进一步逼近真实框提供了新的依据标准。对于 3D 打印缺陷检测, SIoU 特别适用于处理那些角度变化较大的缺陷, 如旋转的裂缝或倾斜的孔洞。SIoU 能够快速调整预测框的角度, 使其更符合实际缺陷的形态。

1.4 注意力机制的引进

注意力机制允许模型在不同尺寸和类型的缺陷上进行有效检测, 提高了模型对新数据的泛化能力, 可以帮助模型更加关注图像中的关键区域。本文通过在改进损失函数的 YOLOv5s 模型的基础上再引入 SimAM、SE-Attention、Shuffle-Attention 注意力机制, 并设置对照试验测定本数据集的最佳匹配的注意力机制。

1.4.1 引入 SimAM 机制

SimAM 是一种用于卷积神经网络, 虽然概念比较简单却十分有效的注意力机制模块, 其通过优化一个能量函数来发现每个神经元的重要性, 可推导出一种能量函数的封闭解, 且此解可以通过不到 10 行代码实现^[8]。其一大创新点在于不用向原始网络模块中添加通道数, 也就是无须匹配原网络中的通道数, 可以自由地加入网络中任何一层, 避免了在调整结构上花费大量精力。SimAM 特别适用于处理那些局部特征明显的缺陷, 如细小的孔洞或表面划痕, 它能够快速捕捉到这些局部特征, 从而提高检测精度。

为了寻找最简单神经元, 度量神经元之间的线性可分性, SimAM 中定义了如下能量函数

$$e_t(w_t, b_t, y, x_t) = (y_t - \hat{t})^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_0 - \hat{x}_i)^2 \quad (10)$$

其中, w_t 、 b_t 分别为线性变换的权重和偏置, y 为当前目标神经元的值, x_t 为同一特征图中, 除 y 以外的其他神经元; $M=H \times W$ 为一个通道内所有神经元的个数; $\hat{t} = w_t t + b_t$, $\hat{x}_i = w_t x_i + b_t$ 是 t 和 x_i 的线性变换, t 是目标神经元。

式(10)可化简为

$$\tilde{X} = \text{Sigmoid}(\frac{1}{E}) \otimes X \quad (11)$$

其中, Sigmoid 作为激活函数, 将能量 E 的倒数激活后与 X 做向量乘积。

在原网络架构头部的第 10 层与第 11 层、第 12 层与第 13 层、第 21 层与 22 层之间加入 SimAM 模块，无须输入接收通道数。插入层级示意图如图 8 所示。

1.4.2 引入 SE-Attention 机制

SE 模块并非一个完整的网络，而是一个子结构，一般用于嵌套在其他分类模块中，其核心思想是根据损失函数的趋势去设置一个使有效的特征图的权重更大，无效或者效果很小的特征图的权重更小的权重分布去训练模型^[19]。其操作为将输入的特征图进行一次标准的卷积操作，进行全局平均池化，将前两部所得特征图使用 ReLU 函数激活，乘

以线性层后通过 Sigmoid 函数转化为 0-1 分布值，SE-Attention 特别适用于处理那些全局特征较为复杂的缺陷，如大面积的裂纹或分层剥离。SE-Attention 能够有效地增强模型对全局特征的理解，从而提高检测的鲁棒性。SE-Attention 机制示意如图 9 所示。

$$F_{tr}: X \rightarrow U, X \in \mathbf{R}^{W' \times H' \times C'}, U \in \mathbf{R}^{W \times H \times C} \quad (12)$$

$$u_c = v_c X = \sum_{s=1}^{C'} v_c^s \times x^s \quad (13)$$

$$z_c = F_{sq}(u_c) = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H u_c(i, j) \quad (14)$$

$$s = \text{sigmoid}(W_2 \text{ReLU}(W_1 z)) \quad (15)$$

$$\tilde{x} = F_{scale}(u_c, s_c) = s_c \times u_c \quad (16)$$

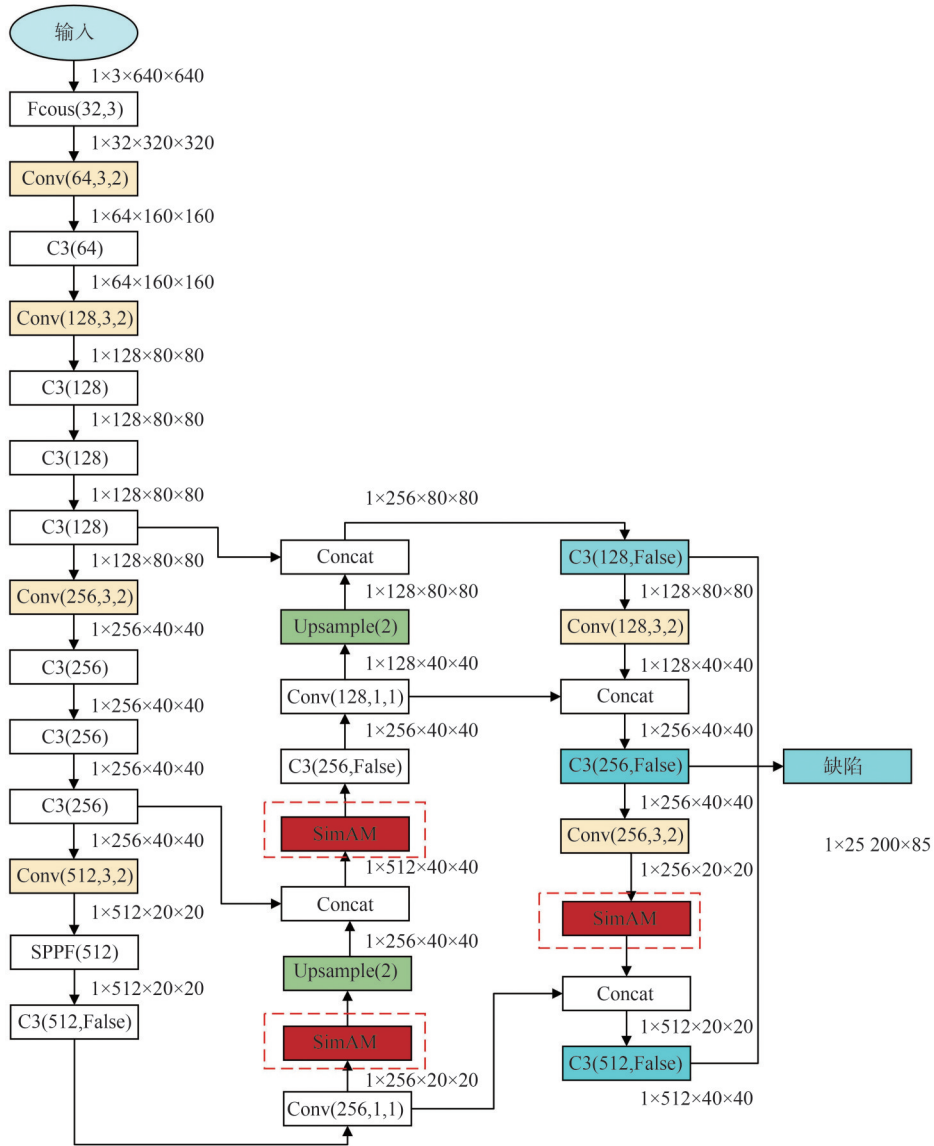


图 8 插入层级示意图

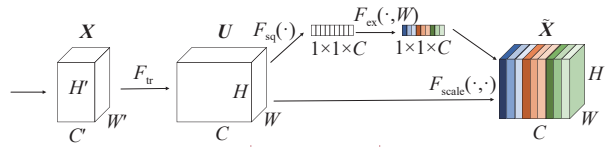


图9 SE-Attention 机制示意图

其中, X 、 U 分别为输入特征图和输出特征图, u_c 为 x 的第 c 个通道, W_1 、 W_2 分别为第一个、第二个全连接层的权重矩阵, s_c 为向量 s 中第 c 个通道对应的标量权重, \hat{x}_c 为重标定后的第 c 个通道。

文献[20]引入了 Shuffle-Attention 机制, 它融合了通道和空间注意力机制, 以增强特征表示, 使用全局平均池处理输入信息, 生成通道统计数据, 然后使用一个门控机制配合 Sigmoid 激活函数, 生成紧凑特征, 最后使空间维度与通道维度的注意力机制互补, 适用于处理那些同时具有局部和全局特征的缺陷, 如复杂的几何形状或多重缺陷叠加。Shuffle-Attention 能够同时捕捉到局部和全局特征, 从而提高模型的综合检测能力

$$s = F_{gp}(X_{k1}) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_{k1}(i, j) \quad (17)$$

$$X'_{k1} = \sigma(F_c(s)) \cdot X_{k1} = \sigma(W_1 s + b_1) \cdot X_{k1} \quad (18)$$

$$X'_{k2} = \sigma(W_2 \cdot GN(X_{k2}) + b_2) \cdot X_{k2} \quad (19)$$

其中, 式(17)首先通过全局平均池化提取特征图的全局上下文信息 s 。接着, 式(18)利用该上下文信息, 通过一个全连接层和 Sigmoid 激活函数生成通道注意力权重, 并将其与原始特征图 X_{k1} 逐元素相乘, 从而实现通道特征的动态加权。最后, 式(19)采用类似的机制, 但基于全局归一化方法来计算权重, 进一步增强特征图 X_{k2} 的表示能力。

1.5 模型轻量化

传统的 YOLOv5 网络架构中, 主干网络使用的是 C3 网络架构, 其由 3 个卷积层、3 个瓶颈层与一个拼接组成, 结构较为复杂, 导致模型在推理过程中的计算量大幅度增加, 为了解决这个问题, 本文更换 YOLOv5 的主干网络与头部网络, 引入 Ghost Net 网络架构, 作为主干和头部网络的基础组成模块, 以达到精简模型结构和减少参数数量和加快推理速度的目的。

GhostBottleneck 由 Ghost 模块堆叠而成, Ghost 是一种即插即用的模块, 旨在使用更少的参数来生成更多的特征图, 从而达到减小参数规模的目

的^[21]。Ghost Net 提出了一种新的 Ghost 模块, 旨在使用一种廉价的线性变换生成更多的特征图。

对输入的特征图进行简单的卷积操作。第一步为生成 $m=n/s$ 个本征特征图, Y' 为输出本征特征图, f' 为所用滤波器, 生成 n 个 Ghost 特征图; 第二步为通过生成的 m 个本征特征图, 对其每一个通道都进行廉价线性运算再生成 $(s-1)$ 个 Ghost 特征图, 共计生成 $m(s-1)$ 个特征图, 将第一步生成的 m 个本征特征图与第二步生成的 $m(s-1)$ 个 Ghost 特征图进行拼接, 得到拼接的 $m \times s$ 个特征图作为最终输出, Ghost 生成特征图如图 10 所示。

$$Y' = X \times f' \quad (20)$$

$$y'_j = \phi_j^i(y'_i), \forall i = 1, 2, \dots, m, j = 1, 2, \dots, s, \quad (21)$$

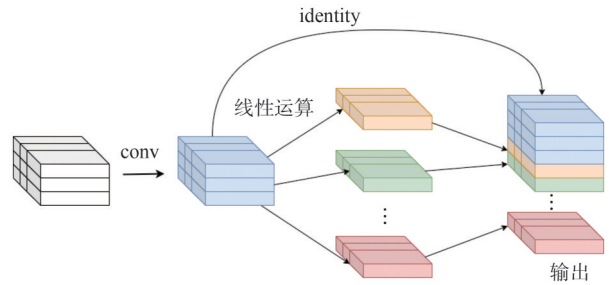


图10 Ghost生成特征图

Ghost 模块主要的运算在一次等价映射与 $m(s-1)$ 次廉价的线性变换。假设要输出 n 个特征图, 那么就需要进 $n-m=m \times (s-1) = n/s \times (s-1)$ 次线性变换, 加上原本 m 个本征特征图, 假设每次线性变换的 kernels 为 $d \times d$, 卷积操作的核为 $k \times k$, d 为超参数, 可得出 Ghost 模块与标准卷积的计算比为

$$r_s = \frac{n \cdot h' \cdot w' \cdot c \cdot k \cdot k}{\frac{n}{s} \cdot h' \cdot w' \cdot c \cdot k \cdot k + \frac{n}{s} \cdot (s-1) \cdot h' \cdot w' \cdot d \cdot d} \approx \frac{s \cdot c}{s + c - 1} \approx s \quad (22)$$

因为 c 的值一般远大于 s , 所以可以近似为 s , 换言之, 在第一步生成本征特征图时, 本征特征图生成得越少, 参数规模削减的效果越好。

GhostBottle 如图 11 所示。当步幅 stride=2 时, 加入的 DWConv 的作用是对特征图的宽和高进行压缩, 缩小到原来的 1/2, 残差部分也会加入类似下采样的操作, 便于残差相加操作时大小可以相匹配。

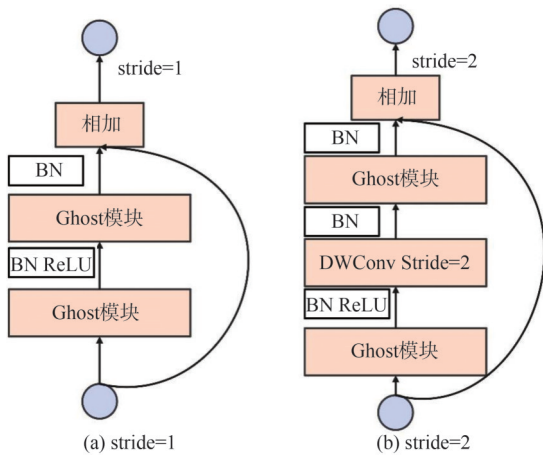


图11 GhostBottle

2 实验结果

2.1 替换IoU损失函数

本文引入了5种改进的IoU损失函数，并在无预训练权重的情况下进行了多次实验，测定其mAP等参数。

原始对照组在同样在无预训练权重的情况下训练，初始模型为YOLOv5s轻量级模型。训练集与验证集的比值为2:8，经过预处理的数据集图片被设置为480×480大小输入模型，初始学习率设为0.001，权重衰减为0.0005，训练轮次为100，批处理大小为16。YOLOv5s模型无预训练权重训练结果见表3。

表3 YOLOv5s模型无预训练权重训练结果

mAP(IoU为0.5)	mAP(IoU为0.5~0.95)	精准率	召回率
0.933	0.719	0.920 2	0.890 6

引入5种损失函数，所有训练条件与对照组完全相同，取3次训练结果取平均值作为实验组，与对照组进行对比，结果见表4。

表4 改进IoU实验结果

损失函数	mAP(IoU为0.5)	mAP(IoU为0.5~0.95)	精准率	召回率
GIoU	0.937	0.751	0.925	0.912
DIoU	0.939	0.751	0.927	0.913
CoU	0.941	0.754	0.929	0.930
SIoU	0.929	0.749	0.929	0.889
EIoU	0.940	0.753	0.932	0.907

实验结果显示，CIoU在本文数据集中表现最佳，说明CIoU的优势在于它能够敏锐地识别3D打印缺陷中长宽比差异较大的情况，如细长的裂缝或不规则的孔洞，从而提高检测精度。

CIoU为最适配本文所用数据集的损失函数。

根据其原理推测是数据集中的故障与严重错误虽然整体相似，但是故障的长宽比远大于严重错误，CIoU损失函数敏锐地识别了这一差别，使故障与严重错误的mAP提高，导致整体mAP上升近1%，平均mAP上升近4%。

2.2 引入注意力机制

在引入CIoU损失函数为前提的情况下，加入SimAM、SE-Attention、Shuffle-Attention注意力机制，在与原始对照组相同条件下进行训练选定最佳匹配的注意力机制。

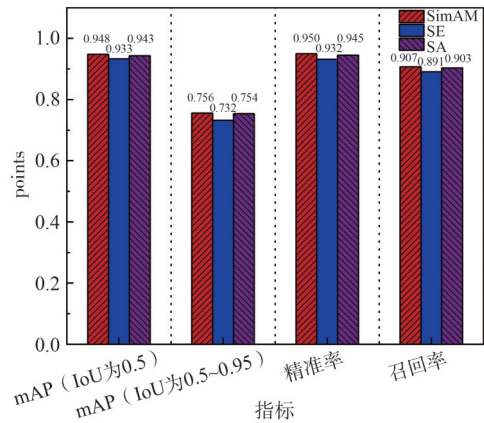


图12 3种注意力机制对比

实验结果显示，SimAM表现最好，推测其原因是没有通道数限制，且不需要额外参数，能够轻松地集成到现有的卷积神经网络中。SimAM通过优化能量函数，帮助模型更专注于图像中的关键区域，从而提高了检测精度。

SE-Attention通过动态调整特征图的权重，增强了模型对全局特征的理解，特别适用于处理那些全局特征较为复杂的缺陷。Shuffle-Attention则通过融合通道和空间注意力机制，提升了模型对局部和全局特征的综合理解，特别适用于处理那些同时具有局部和全局特征的缺陷。

2.3 轻量化

在改进YOLOv5模型的基础上，将原有的主干网络和头部改为由Ghost模块，Conv改为Ghost Conv，C3改为Ghost C3，其他层结构不变。为确定模型改进部分（IoU与注意力机制）对精度的提升所做的贡献，现测定原始模型与改进模型经轻量化后的模型性能，其他实验条件与第2节中的实验条件相同，改进前后模型对比见表5。

从表5可以得出，精度相较于改进模型略有下

表5 改进前后模型对比

阶段	mAP		精准率	召回率	参数规模	GFLOPs	数据大小 /MB
	(IoU为0.5)	(IoU为0.5~0.95)					
改进前	0.933	0.719	0.920	0.890	7.12×10^6	16.0	14.4
改进后	0.942	0.745	0.949	0.896	3.68×10^6	8.0	7.7

降，但是参数规模大幅度下降，推理速度显著提升，由此可见，网络是成功的，以丢失细微检测精度的代价实现了参数规模等大幅度下降。

综上所述，与ResNet等当前流行检测算法相比，在检测精度基本一致的情况下，本文模型大小实现了近半削减。

3 结束语

本文提出了一种基于改进YOLOv5算法的3D打印缺陷检测技术，使用中国科学院自动化研究所收集的数据集，通过优化IoU损失函数并引入注意力机制，显著提升了检测精度。针对特殊环境下对速度和模型大小的需求，采用Ghost Net实现模型轻量化，减少了参数规模，加快了推理速度，同时保持了高准确率。

本文研究内容包括数据增强、损失函数改进、注意力机制和模型轻量化。首先，通过数据增广扩充了初始数据集，解决了类别不平衡问题；其次，对比多种IoU损失函数后，选择了CIoU，并引入了SimAM以增强特征识别；最后，进行模型轻量化，可实现上传至嵌入式设备等。

实验结果显示，改进后的算法不仅保持了高识别准确率，还在运算速度和鲁棒性方面取得了显著提升。这一成就归功于算法的优化，使得系统在面对复杂和多变的数据时，能够更加迅速和稳定地进行处理。此外，通过对模型进行有效的轻量化处理，确保了系统在保持高性能的同时，也易于部署和使用。这些成果为3D打印领域的缺陷检测技术带来了创新。

尽管目前的系统在数据集的多样性以及实际应用测试方面还有待进一步扩展，但笔者已经制定了明确的未来研究方向，计划通过扩大数据集的范围，引入先进的图像预处理技术，并开发实时监测3D打印过程的算法以克服这些局限性。

参考文献：

[1] 喻朝新, 郭松涛, 郭佳哲. 基于工业互联技术的胶料质量缺陷检测模型构建分析[J]. 粘接, 2023, 50(6): 19-22.
YU C X, GUO S T, GUO J Z. Research on rubber defect identifi-

cation and detection based on industrial interconnection technology[J]. Adhesion, 2023, 50(6): 19-22.

- [2] 袁朕鑫. 工业互联网中边缘端轮辋焊缝检测系统设计与实现[D]. 济南: 济南大学, 2023.
YUAN Z X. Design and implementation of edge end rim weld detection system in industrial Internet[D]. Jinan: University of Jinan, 2023.
- [3] 袁慧苗. 基于工业互联网的旋转机械系统状态检测与故障诊断研究[D]. 济南: 齐鲁工业大学, 2023.
YUAN H M. Research on condition detection and fault diagnosis of rotating machinery system based on industrial Internet. [D]. Jinan: Qilu University of Technology, 2023.
- [4] 佟星. 工业互联网场景下基于人工智能的图像和数据分析研究[D]. 广州: 广东技术师范大学, 2021.
TONG X. Image and data analysis based on artificial intelligence in industrial Internet. [D]. Guangzhou: Guangdong University of Technology and Education, 2021.
- [5] 陈立君. 基于流形支持向量机的木材表面缺陷识别方法的研究[D]. 哈尔滨: 东北林业大学, 2015.
CHEN L J. The research for recognition method of wood surface defect based on svm combined manifold. [D]. Harbin: Northeast Forestry University, 2015.
- [6] 刘浩, 陈再良, 王善翔. 基于改进YOLOv3算法的刀具表面缺陷检测[J]. 组合机床与自动化加工技术, 2021(11): 87-90.
LIU H, CHEN Z L, WANG S X. Tool surface defect detection based on improved YOLOv3 algorithm[J]. Modular Machine Tool & Automatic Manufacturing Technique, 2021(11): 87-90.
- [7] QIU J H, HU Y H, CUI J R, et al. Textile defect classification based on convolutional neural network and SVM[J]. AATCC Journal of Research, 2021, 8(1_suppl): 75-81.
- [8] 舒琪. 基于机器视觉的铆钉筛选系统设计[D]. 成都: 西南交通大学, 2022.
SHU Q. Design of a rivet screening system based on machine vision[D]. Chengdu: Southwest Jiaotong University, 2022.
- [9] YU Z L, LEI Y Q, SHEN F, et al. Application of improved YOLOv5 algorithm in lightweight transmission line small target defect detection[J]. Electronics, 2024, 13(2):
- [10] HU B, WANG J H. Detection of PCB surface defects with improved faster-RCNN and feature pyramid network[J]. IEEE Access, 2020, 8: 108335-108345.
- [11] HU Y M, WEN B, YE Y S, et al. Multi-defect detection network for high-voltage insulators based on adaptive multi-attention fusion[J]. Applied Sciences, 2023, 13(24): 13351.
- [12] YU J, JIANG Y, WANG Z, et al. UnitBox: an advanced object detection network[J]. arXiv preprint, 2016, arXiv: 1608.01471.
- [13] 赵英伟. 面向工业场景的机器学习算法应用研究[D]. 成都: 电子科技大学, 2023.
ZHAO Y W. Machine learning algorithms for industrial scenarios applied research. [D]. Chengdu: University of Electronic Science and Technology of China, 2023.

[14] REZATOFIGHI H, TSOI N, GWAK J, et al. Generalized intersection over union: a metric and a loss for bounding box regression[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 658-666.

[15] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

[16] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.

[17] GEVORGYAN Z. SIoU loss: more powerful learning for bounding box regression[J]. arXiv preprint,2022,arXiv:2205.12740.

[18] YANG L X, ZHANG R Y, LI L D, et al. SimAM: a simple, parameter-free attention module for convolutional neural networks[C]//Proceedings of the International Conference on Machine Learning, 2021.

[19] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7132-7141.

[20] ZHANG Q L, YANG Y B. SA-net: shuffle attention for deep convolutional neural networks[C]//Proceedings of the ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2021: 2235-2239.

[21] HAN K, WANG Y H, TIAN Q, et al. GhostNet: more features from cheap operations[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 1577-1586.



沈震(1982-), 男, 博士, 中国科学院自动化研究所多模态人工智能系统全国重点实验室研究员, 主要研究方向为智能制造、无人系统。



方启航(1997-), 男, 博士, 中国科学院自动化研究所多模态人工智能系统全国重点实验室博士后, 主要研究方向为智能制造。



董西松(1978-), 男, 博士, 中国科学院自动化研究所多模态人工智能系统全国重点实验室副研究员/副教授, 主要研究方向为复杂系统的建模与分析、社会制造。



王迪(1986-), 男, 博士, 华南理工大学机械与汽车工程学院教授, 主要研究方向为金属增材制造、装备金属增材制造、工艺金属增材制造、医学和工业应用激光加工。

[作者简介]



张俊杰(2002-), 男, 中国科学院自动化研究所多模态人工智能系统全国重点实验室硕士生, 主要研究方向为智能制造。



熊刚(1969-), 男, 博士, 中国科学院自动化研究所多模态人工智能系统全国重点实验室研究员, 主要研究方向为并行控制与管理、复杂系统的建模与优化、云计算和大数据、智能制造以及智能交通系统。