

基于视觉大模型的激光粉末床熔融铺粉缺陷检测

谭鲲鹏¹, 唐甲锋¹, 赵志斌^{1*}, 王晨希¹, 张兴武¹, 何卫锋², 陈雪峰¹¹西安交通大学机械工程学院航空发动机研究所, 陕西 西安 710049;²空军工程大学航空动力系统与等离子体技术全国重点实验室, 陕西 西安 710038

摘要 激光粉末床熔融增材制造面临质量稳定一致性的挑战, 铺粉质量是影响成形件质量的重要因素。近年来, 计算机视觉在铺粉缺陷监测中的应用表现突出, 但其性能却受到标注数据数量不足的限制。针对这一问题, 笔者设计了基于视觉大模型分割一切模型(SAM)的铺粉缺陷分割模型(PSAM)。针对 SAM 预训练参数的知识迁移问题, 引入 Adapter 模块实现参数微调; 针对铺粉分割任务中类别信息的需求, 改进了 SAM 中的掩码解码器; 针对工业场景中人工提示难的问题, 提出了自动提示生成器, 实现了视觉提示的自动生成。在训练样本数量仅为 50 的情况下, PSAM 表现出了良好的分割性能, 平均交并比(mIoU)可达到 65.02%, 相较于 Deeplab v3 和 U-Net 分别提升了 8.52 个百分点和 5.31 个百分点。本研究展示了视觉大模型在增材过程监控中的应用价值和潜力。

关键词 激光技术; 激光粉末床熔融; 过程监测; 视觉大模型; 缺陷检测

中图分类号 TP277 文献标志码 A

DOI: 10.3788/CJL240430

1 引言

激光粉末床熔融(LPBF)技术作为目前最为先进的金属增材制造技术, 已被广泛应用于航空航天、医疗健康等领域关键金属零部件的制造^[1-3]。然而, 由于 LPBF 成形过程涉及多种因素的耦合作用, 确保 LPBF 成形零部件的质量稳定一致性是目前面临的重大挑战^[4]。铺粉质量是 LPBF 过程监测中的重要宏观表征, 铺粉缺陷会直接影响当前层的成形质量, 并影响后续成形过程, 为成形件埋下质量隐患^[5]。对铺粉过程中的缺陷进行及时而准确的检测和评估, 有助于及时发现铺粉过程中的异常状况, 并为成形工艺的优化提供参考, 进而提升零件质量^[6-7]。

近年来, 高精度、高自动化的铺粉缺陷监测技术多集中于计算机视觉领域, 比如图像分割技术, 其在缺陷监测中表现优异^[8-10]。美国橡树岭国家实验室的 Scime 等^[11]系统地解释了铺粉缺陷的分类及形成机制, 并使用基于 U-Net 的深度神经网络实现了对铺粉图像的区域分割。然而, 训练这样的分割网络往往需要大量的标注数据。在工业领域中, 图像数据标注需要耗费大量的人力, 并且需要专业人员全程参与, 形成具有丰富标注数据的数据集耗时长、难度大, 这也成为了开发工业领域深度神经网络的障碍^[12-13]。

如今, 以 GPT-4^[14]为代表的大模型正在人工智能领域引领一场巨大的革命。大模型所具有的卓越的

泛化能力, 使其能够通过学习少量额外数据实现学习任务的迁移, 从而使得其在工业领域的部署成为可能^[15]。视觉大模型中的分割一切模型(SAM)^[16]使用 SA-1B 数据集中超过 1000 万的图像和 10 亿的标注进行预训练, 包含极其丰富的图像知识, 能够在不同的下游任务中取得良好表现, 目前已有将 SAM 应用于工业检测的案例。Hu 等^[17]使用热成像缺陷数据集微调 SAM, 开发了用于缺陷检测的大模型 DefectSAM。Chen 等^[18]通过对特定参数进行训练, 将纺织品缺陷专业知识注入 SAM, 实现了基于 SAM 的纺织品缺陷分割。SAM 在工业缺陷检测任务中表现出的优异性能使得 SAM 迁移至铺粉缺陷分割任务成为可能。

然而, 将 SAM 直接应用至铺粉缺陷分割面临诸多问题^[19]。首先, SAM 对图像区域的分割结果不包含类别信息, 分割结果无法与类别相关联^[20], 而在铺粉缺陷图像分割中需要明确缺陷的类型, 以便进行后续的针对性调控。其次, SAM 的训练数据为自然图像, 其特征与铺粉图像相去甚远, 特征分布的较大差异将导致 SAM 在铺粉图像上的分割性能不稳定^[17]。最后, 也是最为重要的一点, SAM 依赖视觉提示(点、边界框、掩码)进行交叉注意力预测, 但 LPBF 成形过程周期长而且会持续不断地产生铺粉图像, 逐张对图像进行人工提示需要大量的人力, 不契合过程监测自动化的目标, 而且包含类别信息的分割任务所需的提示类

收稿日期: 2024-01-02; 修回日期: 2024-03-05; 录用日期: 2024-03-06; 网络首发日期: 2024-03-13

基金项目: 国家重点研发计划(2022YFB4600800)

通信作者: *zhaozhibin@xjtu.edu.cn

型目前尚不明确^[21]。

针对上述问题,笔者设计了铺粉缺陷分割一切模型(PSAM),它是一种能够自动提示的基于SAM的LPBF铺粉缺陷分割方法。针对SAM预训练参数中不包含铺粉图像相关知识的问题,引入了Adapter模块,以实现自然图像向铺粉图像的知识迁移;针对铺粉分割任务中对缺陷进行类别区分的需求,改进了SAM中的掩码解码器,实现了包含类别信息的掩码输出;针对工业场景中人工提示难的问题,引入了自动提示生成器,实现了视觉提示的自动生成。实验结果证明,PSAM能够在少量训练数据的条件下实现良好的缺陷分割性能,分割性能优于常见的分割模型。

2 基本原理

2.1 PSAM 概述

PSAM是以SAM为主干搭建的模型。SAM是一种可提示分割的基础模型,由图像编码器、提示编码器和掩码解码器构成,如图1所示。SAM的工作原理如下:首先,将图像输入图像编码器,图像编码器是使用掩码自编码器(MAE)^[22]预训练的视觉自注意力模型(ViT)^[23],输入图像在其中被高效编码为对应的图像

嵌入;随后,提示编码器将对人工设定的稀疏提示(如点、边界框、文本)和密集提示(掩码)进行编码,生成包含注意力信息的提示嵌入;最后,掩码解码器将图像嵌入、提示嵌入和输出令牌作为输入,通过自注意力、双向交叉注意力进行图像嵌入和提示嵌入之间的交互,并随后通过上采样、多层感知机(MLP)将注意力运算结果进行映射,输出对应提示区域对象分割的二值掩码。

对于一张尺寸为 $H \times W$ 、通道数为 C 的铺粉图像 x ,铺粉图像分割的目标是训练一个可对图像中的每个像素的分类进行预测,并可生成一张尺寸同为 $H \times W$ 的分割图的模型。分割图中的每一个像素属于一个预定义的分类列表 $Y = \{y_0, y_1, \dots, y_k\}$,其中, y_0 代表背景, y_i 代表铺粉图像中不同的区域或缺陷的分类, $i \in \{1, \dots, k\}$ 。在训练过程中经历数次迭代之后,模型能够准确识别并分割出图像中的缺陷。虽然SAM具备极其强大的图像特征提取能力,但由于SAM缺乏铺粉图像相关知识并且模型输出不包含类别信息,直接在铺粉缺陷分割任务上应用SAM不能取得预期的效果,如图2所示。针对铺粉缺陷分割任务,笔者设计了PSAM,其整体结构如图3所示。

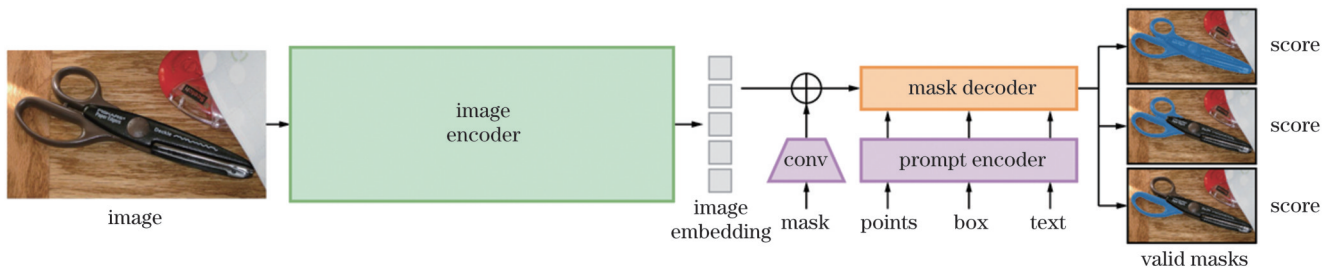


图1 SAM的结构^[16]
Fig. 1 Structure of SAM^[16]

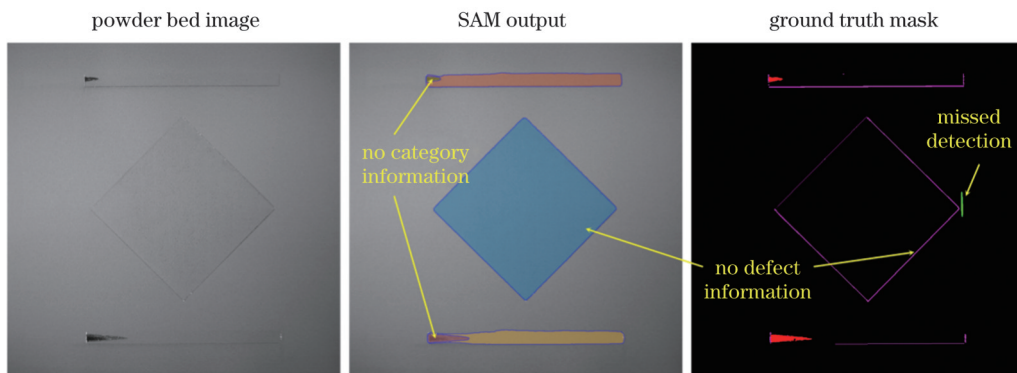


图2 SAM的局限性
Fig. 2 Limitation of SAM

2.2 图像编码器微调

尽管SAM图像编码器可以对输入图像进行高效编码,但其训练主要是针对自然图像进行的,训练数据中并不包含铺粉图像,而这两种图像在颜色、纹理等分布特征上差异显著,这种差异会直接影响SAM图像

编码器对铺粉图像的特征提取^[20]。因此,SAM原有的图像编码器不能直接应用于铺粉缺陷分割任务。此外,虽然重新训练SAM能够实现其在铺粉缺陷分割中的应用,但所需数据量和运算成本在铺粉质量监测问题中是难以实现的。因此,如何使用少量数据实现

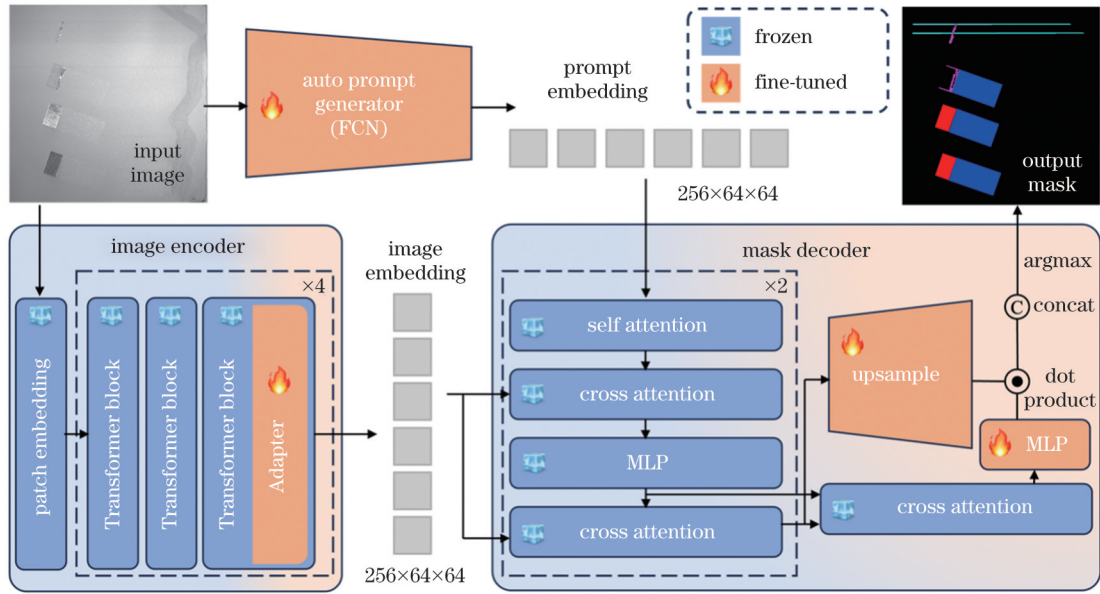


图 3 PSAM 整体结构图

Fig. 3 Overall structure of PSAM

模型知识向铺粉图像的迁移是将 SAM 应用于铺粉缺陷分割任务的关键。

Adaption^[24] 是一种被广泛应用于自然语言处理 (NLP) 的高效的参数微调技术, 针对特定用途对预训练模型进行微调。其主要思想是在模型中插入几个 Adapter^[25] 模块, 冻结其他所有预训练参数, 训练数据只用来更新 Adapter 参数。通过使用 Adapter 模块, 可以将图像编码器提取到的特征映射到对应于铺粉图像的特征空间, 实现针对铺粉缺陷分割的参数微调。PSAM 中的图像编码器使用 ViT-B 模型, 其中包含 12 个 Transformer 模块。为了降低少样本情况下的训练难度, 同时充分发挥 Adapter 模块在模型预训练中的调整作用, 在 PSAM 图像编码器的第 3、6、9、12 个 Transformer 模块中插入 Adapter 模块, 如图 4 所示。Adapter 模块由两部分组成, 其中: 第一部分主要由线性层构成, 通过对通道数进行变换实现特征维度层面的调整; 第二部分主要由卷积层构成, 通过对特征图尺寸进行压缩和还原实现空间尺度层面的调整。为了实现对特征的高效调整, Adapter 模块插入的位置位于 Transformer 模块中的多头注意力层之后。

2.3 类别掩码解码器

SAM 中的掩码解码器可以分为注意力层和分割头两部分, 其中分割头可以设置为对一张图像输出多个掩码。但是, 这些分割图不包含类别信息, 只是对应于同一提示对目标进行不同尺度的分割, 即歧义分割^[20]。为了实现不同铺粉缺陷类型的区分, PSAM 改进了原有的掩码解码器, 改进后的掩码解码器可以实现类别掩码的输出, 如图 5 所示。设置掩码解码器, 为一张输入图片输出对应类别数的 $k + 1$ 张分割图, 每一

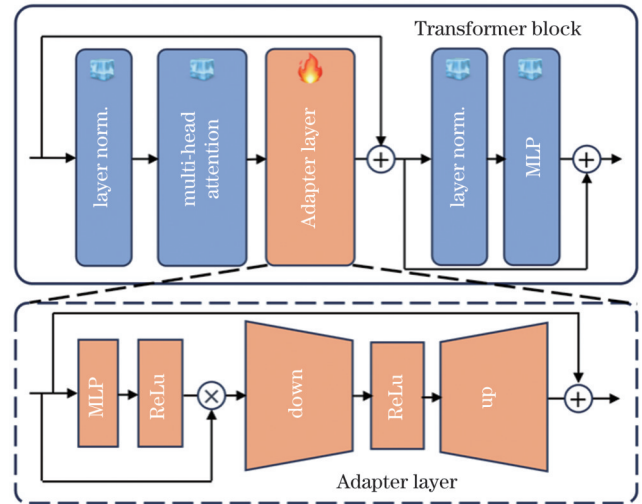


图 4 Adapter 模块的位置和结构

Fig. 4 Position and structure of Adapter layer

张分割图对应着包括背景在内的同一类别区域的分割预测值。

对于某个像素, 其最终的分类结果为 $k + 1$ 张分割图中预测值最高的分割图对应的类别。其计算过程为

$$y^{H \times W} =$$

$$\operatorname{argmax} \left\{ \operatorname{Softmax} \left[\operatorname{concat} \left(L_i^{H \times W} \right) \right] \right\}, i \in \{0, \dots, k\}, (1)$$

式中: $L_i^{H \times W}$ 代表模型对第 i 个类别的输出分割图; $y^{H \times W}$ 代表最终的逐像素分类的分割图; $\operatorname{concat}(\cdot)$ 代表将数据按照类别维度合并; $\operatorname{Softmax}(\cdot)$ 是将输出值转换成范围在 $[0, 1]$ 并且和为 1 的概率分布函数; $\operatorname{argmax}(\cdot)$ 代表取最大值索引的函数, 对应该像素的类别序号。为了能够充分利用原结构的分割能力同时实现类别信息的输出, 在掩码解码器训练过程中冻结注意力层参数, 对分割头的参数进行微调。

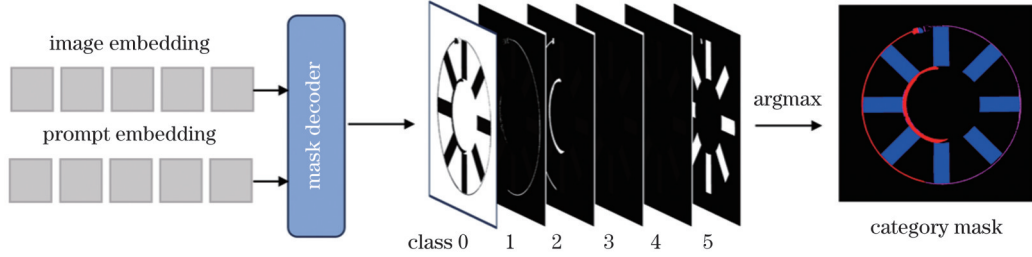


图 5 类别掩码生成过程

Fig. 5 Generating process of category mask

2.4 自动提示生成器

SAM 的卓越性能得益于视觉提示的引入^[26]。在 SAM 中,人工设置的视觉提示通过提示编码器生成提示嵌入,提示嵌入在掩码解码器中与图像嵌入进行双向交叉注意力映射,生成预测掩码。由于 LPBF 成形周期长达数日、逐层铺粉图像数目多达数千张并且单张图像缺陷类型和尺度不一,难以在成形过程中实现人工逐张对图像进行提示。同时,SAM 的提示具有明确的位置信息,但缺乏类别信息。而对于铺粉缺陷分割任务,由于同一张分割图中可能存在多种类别并且同一类别区域之间可能并不连续,SAM 原有的提示类型无法直接应用。针对上述问题,笔者提出了一个自动提示生成器,它通过对输入图片进行特征提取来学习可以直接被掩码生成器利用的提示嵌入。

自动提示生成器的结构如图 6 所示。自动提示

生成器是一个使用残差连接的全卷积网络(FCN)。输入图像首先通过由 7×7 卷积核大小的卷积层和最大池化层组成的预处理层,将通道数扩展到 64,并初步调整特征图的尺寸;之后经过三个残差神经网络模块。每个模块均由主支路和残差支路组成,其中:主支路由两个卷积层组成,每个卷积层后面都跟着批量归一化层(BN)和 ReLu 激活函数;残差支路通过卷积操作之后加到主支路形成残差结构。残差结构允许梯度在网络层之间直接传递,能够有效缓解梯度消失问题,使网络更易优化和收敛。同时,残差结构能够将不同层级的信息直接传递到更深层网络,保留了多层级的信息,对图像特征具有良好的表示能力^[27]。在残差神经网络模块中,图像特征通道数逐步增加,特征图尺寸逐步减小,最后输出形状为 $256 \times 64 \times 64$ 的提示嵌入。

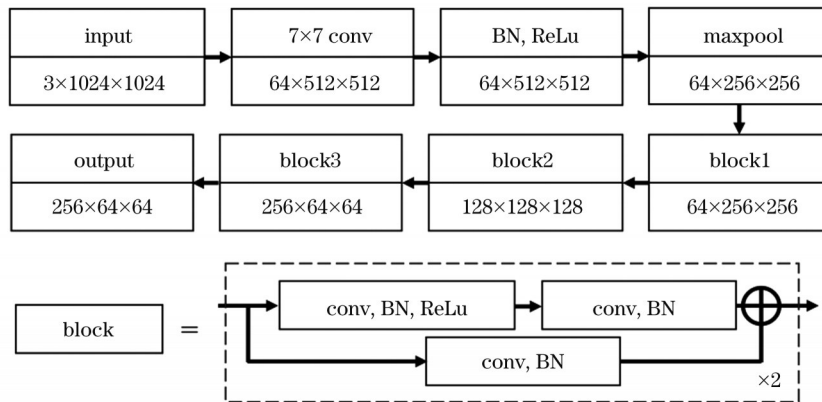


图 6 自动提示生成器结构

Fig. 6 Structure of auto-prompt generator

2.5 损失函数

铺粉图像具有缺陷尺寸跨度大、像素数量极不平衡的特点。在 PSAM 训练过程中,模型输出 $f(x)$ 与标签 y 之间的损失函数 L 表示为

$$L = \lambda_1 L_{CE}[f(x), y] + \lambda_2 L_F[f(x), y] + \lambda_3 L_D[f(x), y], \quad (2)$$

式中: $\lambda_1, \lambda_2, \lambda_3$ 是用来调节各部分权重的参数; $L_{CE}(\cdot)$ 代表交叉熵损失^[28]; $L_F(\cdot)$ 代表焦点损失^[29]; $L_D(\cdot)$ 代表 Dice 损失^[30]。

交叉熵损失是一种常用于分类任务的损失函数。

对于输入图片的每个像素,模型输出一个概率分布,表示该像素属于每个类别的概率。交叉熵损失通过比较模型的预测概率分布与真实标签之间的距离来衡量损失。交叉熵损失的计算公式为

$$L_{CE}[f(x), y] = - \sum_{i=1}^N y_i \log_2(p_i), \quad (3)$$

式中: N 代表类别数目; y_i 代表属于第 i 类的标签; p_i 代表模型对第 i 类的预测概率。

焦点损失通过引入焦点因子来降低易分类样本的权重、增加难分类样本的权重,这有助于模型更关注那些在训练中容易被忽视的少数类别,从而减少类别不

平衡对模型训练过程的影响。焦点损失的计算公式为

$$L_F[f(x), y] = -\sum_{i=1}^N \alpha (1 - p_i)^\gamma \cdot y_i \log_2(p_i), \quad (4)$$

式中: α 是焦点因子; γ 是焦点损失的调节参数。

Dice 损失通过度量模型预测结果与真实标签的相似性来衡量损失, 它的计算过程包含了对两个区域相交部分大小的计算。由于相交部分的计算对每个类别都是平等的, 因此, Dice 损失在本质上对类别不平衡具有一定的鲁棒性。相比于交叉熵损失, Dice 损失更加注重模型在预测分割边界上的性能。Dice 损失的计算公式为

$$L_D[f(x), y] = \sum_{i=1}^N \left(1 - \frac{2p_i y_i + \epsilon}{p_i^2 + y_i^2 + \epsilon} \right), \quad (5)$$

式中: ϵ 是一个极小的常数, 用于防止分母为零的情况出现。

3 实验分析与讨论

3.1 铺粉缺陷数据集

本次实验使用的数据是从装备有粉末床监测系统的 LPBF 设备中使用离轴工业相机收集的多个零件成形过程中逐层粉末铺展图像。原始图片的分辨率从

1400 pixel \times 1400 pixel 到 4300 pixel \times 4300 pixel 不等。这些图像首先通过透视变换转换到前向视角, 之后通过图像裁切删除粉末床成形区域之外的多余部分, 最后降采样到 1024 pixel \times 1024 pixel。在这些图像中共挑选了 406 张包含铺粉缺陷的图像作为铺粉缺陷数据集, 并设置了背景(background)、翘曲(super-elevation)、铺粉不全(incompletion)、振纹(hopping)、拖拽(streaking)和点阵(lattice) 6 种类别。翘曲一般出现在成形轮廓边缘, 是成形过程中温度快速变化产生的局部热应力引起的, 表现为细长的高于粉末床平面的区域, 如图 7(a) 所示。铺粉不全通常是由供粉不足或粉末流动性差导致粉末铺展后不能完全覆盖已熔融区域造成的, 表现为较大面积的零件裸露, 如图 7(b) 所示。振纹表现为与粉末刮刀运动方向垂直的直线, 通常是由粉末刮刀的振动或者粉末刮刀与已熔融区域碰撞造成的, 如图 7(c) 所示。拖拽表现为与粉末刮刀运动方向平行的直线, 通常是由粉末刮刀拖动粉末床上的杂质造成的, 如图 7(d) 所示。点阵是由零件支撑引起的一种特殊形状, 其本身不是一种缺陷。为了将其与翘曲区分, 将其单独列为一类, 如图 7(e) 所示。对这些图像进行逐像素人工标注, 每张图像制作了对应的分割掩码, 掩码中每个像素的数值为图像中该像素对应的类别编号, 标注示例如图 7(f) 所示。

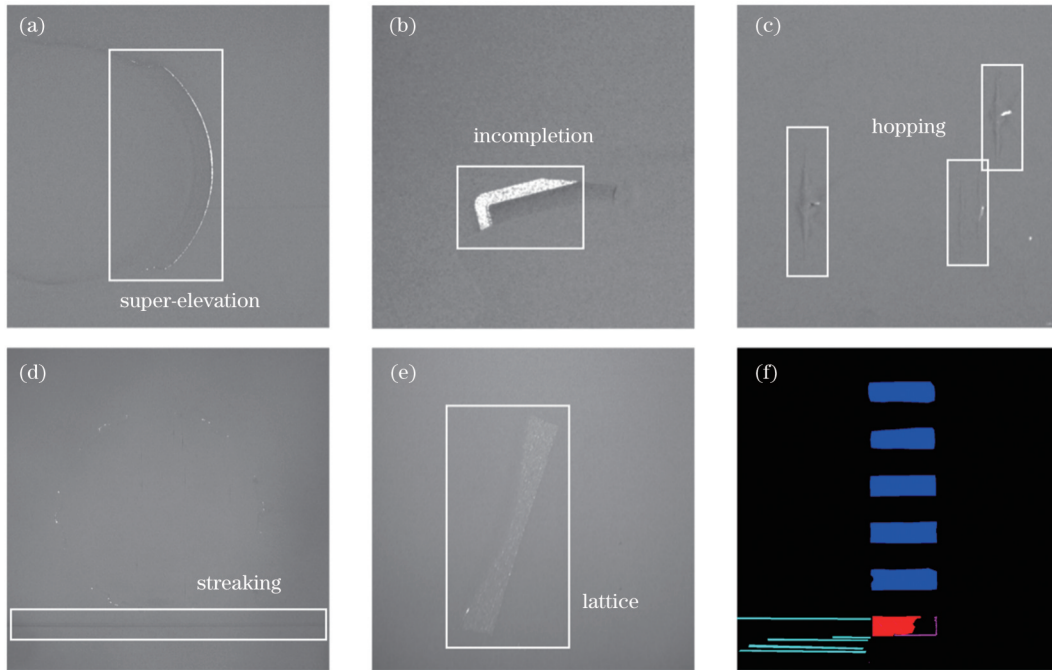


图 7 铺粉缺陷和标注示例。(a)翘曲;(b)铺粉不全;(c)振纹;(d)拖拽;(e)点阵;(f)逐像素标注示例

Fig. 7 Powder-spreading defects and annotation example. (a) Super-elevation; (b) incompletion; (c) hopping; (d) streaking; (e) lattice; (f) pixel-wise annotation example

在本研究中, 为了探究模型在少量数据上的训练效果, 仅使用 50 张图像作为训练集。使用另外的 50 张图像作为验证集, 用于评价模型在训练过程中的性能变化, 并作为保存性能最佳模型的标准。剩余的 306

张图像作为测试集, 用于评价模型的最终性能。

3.2 实验设置和评价指标

实验程序均基于 Python 3.10.0 和 Pytorch 2.1.1 开发, 图形加速环境为 CUDA 11.8, 训练和测试过程在 4

张图形显示卡(Nvidia Tesla V100-SXM3-32 GB)上进行。训练过程中进行 2000 次迭代(若 400 次迭代内没有提升则提前停止训练),初始学习率设置为 0.0001,采用 AdamW 优化器和余弦退火学习率调整策略,批次大小为 8。损失函数中的权重分别设置为 $\lambda_1 = 1$ 、 $\lambda_2 = 2$ 和 $\lambda_3 = 1$ 。焦点损失中的焦点因子 $\alpha = 1$,调节参数 $\gamma = 2$ 。Dice 损失中的常数 $\epsilon = 0.0001$ 。对于每组实验,重复 5 次,以消除随机性对实验结果的影响,取 5 次实验的均值与标准差作为实验的最终结果。

将平均交并比(mIoU,在公式中记为 ξ_{mIoU})作为模型预测结果的评价指标,mIoU 是交并比(IoU)在各个类别上的平均值。IoU(在公式中记为 ξ_{IoU})是图像分割任务中经常使用的评价指标,它度量了模型预测结果与真实值之间的对齐程度,由预测掩码与真实标签的

交集除以并集得到。IoU 和 mIoU 的计算公式分别为

$$\xi_{\text{IoU}} = \frac{q_{\text{TP}}}{q_{\text{TP}} + q_{\text{FP}} + q_{\text{FN}}}, \quad (6)$$

$$\xi_{\text{mIoU}} = \frac{1}{N} \sum_{i=1}^N \xi_{\text{IoU}, i}, \quad (7)$$

式中: q_{TP} 代表模型正确地预测为正例的数量; q_{FP} 代表模型错误地预测为正例的数量; q_{FN} 代表模型错误地预测为负例的数量; N 代表类别数目。

3.3 对比实验

为了验证所设计的 PSAM 整体性能的优越性,选择 Deeplab v3^[31]和 U-Net^[32]这两种在图像分割领域表现突出的网络,在相同的实验设置下进行训练,比较这两种网络和 PSAM 的铺粉缺陷分割性能。实验结果如表 1 所示。

表 1 PSAM 与不同分割网络的性能对比

Table 1 Performance comparison of PSAM with different segmentation networks

Network	mIoU / %	IoU for per category / %					
		Background	Super-elevation	Incompletion	Hopping	Streaking	Lattice
Deeplab v3	56.50±0.23	97.85±0.75	36.74±2.40	49.38±0.89	32.05±2.81	44.71±2.61	78.26±1.48
U-Net	59.71±0.86	98.61±0.08	49.38±0.61	60.06±0.01	30.36±1.72	42.56±2.20	77.24±1.77
PSAM	65.02±0.40	98.66±0.03	50.10±0.65	69.97±1.51	42.51±1.61	51.26±1.14	77.65±1.24

从表 1 中可以看出,由于用于训练的铺粉图像数量极为有限,Deeplab v3 和 U-Net 在测试集上的分割结果不佳,对于翘曲、振纹、拖拽这类所占像素数少、相对极不平衡的类别,IoU 在 50% 以下。这是因为在训练数据多样性不足的情况下,这两种模型仅能学习到少数样本的特定特征,无法获得更丰富、更复杂的特征表示能力。而 SAM 本身具有较强的图像特征提取能力,不需要再使用大量数据训练复杂的网络,因而在最终的分割性能上优于其他模型。相较于 Deeplab v3 和 U-Net,PSAM 的 mIoU 分别提升 8.52 个百分点和 5.31 个百分点。

不同分割网络的可视化结果如图 8 所示。从图 8 中可以看出,在少量训练样本的情况下,PSAM 在分割效果上优于其他两种网络,分割结果更加接近真实标签。对于轮廓形状复杂且难以区分的翘曲、铺粉不全、点阵类别,PSAM 的分割结果相较于其他两种网络具有更加精细的轮廓。由于用于训练的数据不够充分,Deeplab v3 和 U-Net 难以学习、区分这些类别的深层特征信息,分割结果中常常会出现类别之间的混淆,如图中虚线框所示。对于本身所占像素较少且在铺粉图像中不明显的拖拽、振纹类别,PSAM 的分割结果也较好,而 Deeplab v3 和 U-Net 难以进行有效的特征学习,常常会发生误检、漏检的情况,如图中实线框和点划线框所示。

3.4 消融实验

为了验证所设计的 PSAM 中的各种改进方法的

有效性,通过消融实验来比较增减不同方法所带来的铺粉缺陷分割性能的差异。为了验证 Adapter 模块的有效性,设计了不插入 Adapter 模块的模型进行对比。在这种实验设置下,图像编码器的参数将不进行更新。为了验证自动提示的有效性,设计了禁用自动提示生成器的模型进行对比。在这种实验设置下,将一个数值全为 1 的 32×32 矩阵作为密集提示输入,使用 SAM 原有的提示编码器进行编码,提示编码器的参数未被冻结,而是随训练过程进行更新。以上实验的结果如表 2 所示。

由表 2 中的第 1 组结果可以看出,在不添加 Adapter 和自动提示模块的情况下,模型的分割性能低于 Deeplab v3 和 U-Net。这是因为在不对图像编码器参数进行调整的情况下,图像编码器不具备对铺粉图像相关知识的理解,从而使得其对铺粉图像特征的提取变得困难,并且由于没有视觉提示的指导,掩码解码器无法实现高效的交叉注意力。由表 2 中的第 2 组结果可以看出,虽然添加自动提示对模型分割能力有一定程度的提升,但仍不能弥补图像编码器在特征提取能力方面的不足。由表 2 中的第 3 组结果可以看出,通过引入 Adapter 对图像编码器参数进行微调,成功实现了特征提取能力的迁移,从而提高了分割性能,这也展示了参数微调在使大模型适应特定下游任务中的关键作用。由表 2 中的第 4 组结果可以看出,引入 Adapter 和自动提示的模型实现了显著的分割性能提升。

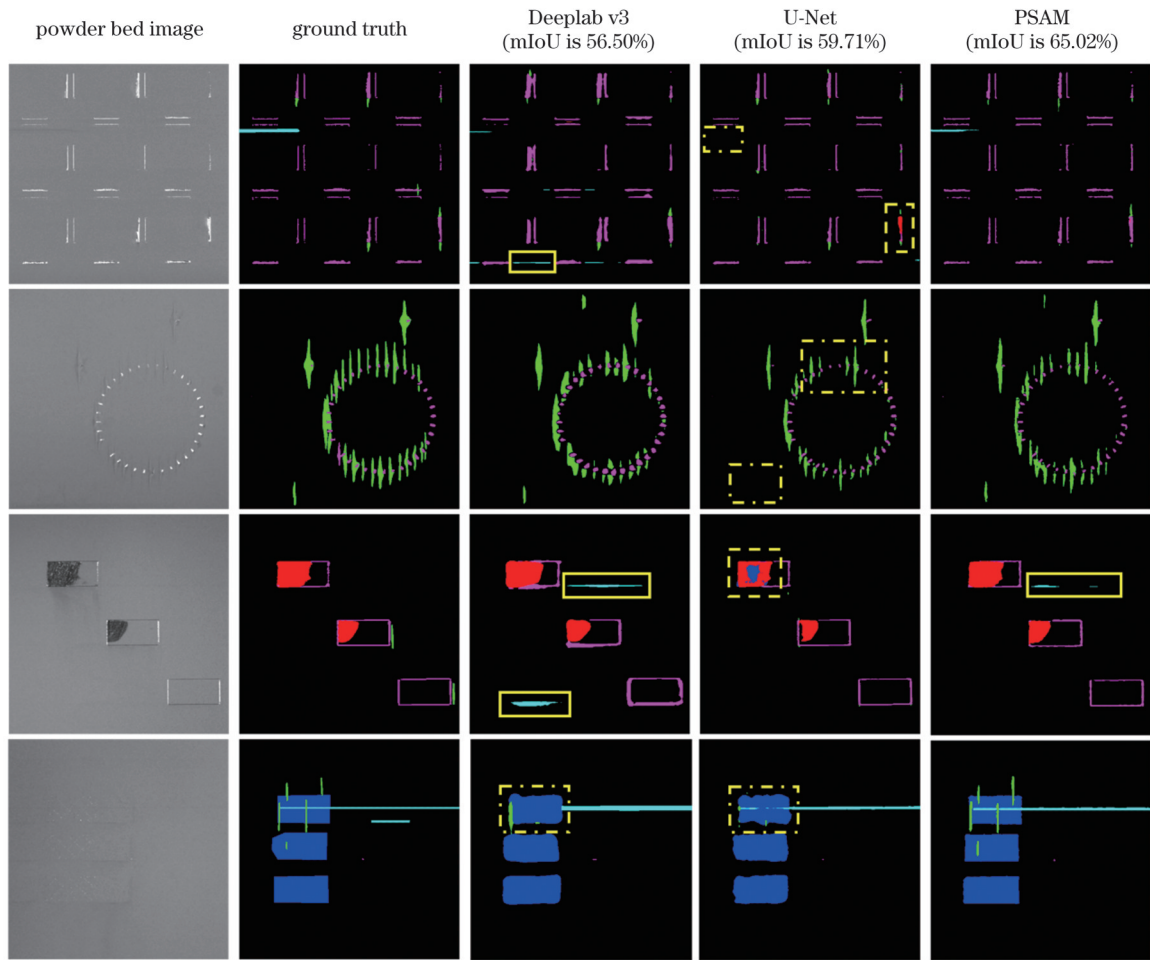


图8 不同网络的分割结果对比

Fig. 8 Comparison of segmentation results from different networks

表2 消融实验结果

Table 2 Ablation experimental results

Group	Adapter	Auto-prompt	mIoU / %	Improvement
1			53.69±0.28	
2		✓	57.23±0.75	+3.54 percentage points
3	✓		58.79±1.23	+5.10 percentage points
4	✓	✓	65.02±0.40	+11.33 percentage points

4 结 论

笔者针对 LPBF 铺粉缺陷识别中存在的训练数据不充分的问题,结合视觉大模型 SAM,设计了用于 LPBF 铺粉缺陷分割的模型 PSAM。首先,针对 SAM 预训练数据中缺少铺粉图像的问题,通过引入 Adapter 模块,实现了图像编码器特征提取能力从自然图像到铺粉图像的迁移。其次,针对 SAM 输出缺乏类别信息的问题,修改了 SAM 中的掩码解码器,使其具备输出类别掩码的能力。最后,针对工业场景中难以进行人工提示的问题,添加了一个自动提示生成器,通过对输入图像进行编码,生成可以被掩

码解码器使用的视觉提示嵌入。本研究所得主要结论如下:

1) 模型 PSAM 能够实现 LPBF 铺粉缺陷的分割,并且在使用少量铺粉图像对模型进行训练的情况下,PSAM 仍能表现出良好的分割性能, mIoU 可达到 65.02%, 相较于 Deeplab v3 和 U-Net 分别提升了 8.52 个百分点和 5.31 个百分点。

2) 所提出的改进方法在 LPBF 铺粉缺陷分割任务中均对模型性能有提升作用。相较于 SAM, PSAM 在 mIoU 上提升了 11.33 个百分点。

3) 所提方法成功地将视觉大模型 SAM 应用至 LPBF 过程监测中,为探索视觉大模型在 LPBF 增材制造过程监测中的应用价值和潜力提供了新途径。

虽然笔者已经实现了在铺粉缺陷分割中应用视觉大模型,但仍未完全开发视觉大模型的应用潜力。大模型所具备的卓越的特征提取能力、零样本泛化能力、多模态知识融合能力,可以为增材过程监控提供全新的解决思路和解决办法,大模型在增材过程监控中的应用仍需要在未来的工作中进一步探究。

参 考 文 献

- [1] 顾冬冬, 张红梅, 陈洪宇, 等. 航空航天高性能金属材料构件激光增材制造[J]. 中国激光, 2020, 47(5): 0500002.
Gu D D, Zhang H M, Chen H Y, et al. Laser additive manufacturing of high-performance metallic aerospace components [J]. Chinese Journal of Lasers, 2020, 47(5): 0500002.
- [2] Kumar M B, Sathiya P. Methods and materials for additive manufacturing: a critical review on advancements and challenges[J]. Thin-Walled Structures, 2021, 159: 107228.
- [3] Chowdhury S, Yadaiah N, Prakash C, et al. Laser powder bed fusion: a state-of-the-art review of the technology, materials, properties & defects, and numerical modelling[J]. Journal of Materials Research and Technology, 2022, 20: 2109-2172.
- [4] 赵志斌, 王晨希, 张兴武, 等. 激光粉末床熔融增材制造过程智能监控研究进展与挑战[J]. 机械工程学报, 2023, 59(19): 253-276.
Zhao Z B, Wang C X, Zhang X W, et al. Research progress and challenges in process intelligent monitoring of laser powder bed fusion additive manufacturing[J]. Journal of Mechanical Engineering, 2023, 59(19): 253-276.
- [5] McCann R, Obeidi M A, Hughes C, et al. *In-situ* sensing, process monitoring and machine control in laser powder bed fusion: a review[J]. Additive Manufacturing, 2021, 45: 102058.
- [6] Taherkhani K, Ero O, Liravi F, et al. On the application of *in situ* monitoring systems and machine learning algorithms for developing quality assurance platforms in laser powder bed fusion: a review[J]. Journal of Manufacturing Processes, 2023, 99: 848-897.
- [7] 姚讯杰, 王佳玮, 杨雁程, 等. 金属构件激光增材制造缺陷产生机理及控制机制探究[J]. 中国激光, 2022, 49(14): 1402802.
Yao X J, Wang J W, Yang Y C, et al. Review on defect formation mechanisms and control methods of metallic components during laser additive manufacturing[J]. Chinese Journal of Lasers, 2022, 49(14): 1402802.
- [8] Zhang Y J, Yan W T. Applications of machine learning in metal powder-bed fusion in-process monitoring and control: status and challenges[J]. Journal of Intelligent Manufacturing, 2023, 34(6): 2557-2580.
- [9] Sarkon G K, Safaei B, Kenevisi M S, et al. State-of-the-art review of machine learning applications in additive manufacturing: from design to manufacturing and property control[J]. Archives of Computational Methods in Engineering, 2022, 29(7): 5663-5721.
- [10] 苏金龙, 陈乐群, 谭超林, 等. 基于机器学习的增材制造过程优化与新材料研发进展[J]. 中国激光, 2022, 49(14): 1402101.
Su J L, Chen L Q, Tan C L, et al. Progress in machine-learning-assisted process optimization and novel material development in additive manufacturing[J]. Chinese Journal of Lasers, 2022, 49(14): 1402101.
- [11] Scime L, Beuth J. A multi-scale convolutional neural network for autonomous anomaly detection and classification in a laser powder bed fusion additive manufacturing process[J]. Additive Manufacturing, 2018, 24: 273-286.
- [12] Zhang Y, Safdar M, Xie J R, et al. A systematic review on data of additive manufacturing for machine learning applications: the data quality, type, preprocessing, and management[J]. Journal of Intelligent Manufacturing, 2023, 34(8): 3305-3340.
- [13] 杨乐森, 周富强. 基于机器视觉的划痕检测技术综述[J]. 激光与光电子学进展, 2022, 59(14): 1415009.
Yang L M, Zhou F Q. Survey of scratch detection technology based on machine vision[J]. Laser & Optoelectronics Progress, 2022, 59(14): 1415009.
- [14] OpenAI, Achiam J, Adler S, et al. GPT-4 technical report[EB/OL]. (2023-12-18)[2023-12-20]. <http://arxiv.org/abs/2303.08774>.
- [15] Moenck K, Wendt A, Prünke P, et al. Industrial segment anything: a case study in aircraft manufacturing, intralogistics, maintenance, repair, and overhaul[EB/OL]. (2023-07-24)[2023-08-30]. <http://arxiv.org/abs/2307.12674>.
- [16] Kirillov A, Mintun E, Ravi N, et al. Segment anything[EB/OL]. (2023-04-05)[2023-12-13]. <http://arxiv.org/abs/2304.02643>.
- [17] Hu B Z, Gao B, Tan C, et al. Segment anything in defect detection[EB/OL]. (2023-11-16)[2023-12-13]. <http://arxiv.org/abs/2311.10245>.
- [18] Chen Z W, Wong W K, Zhong Z F, et al. Effective transfer of pretrained large visual model for fabric defect segmentation via specific knowledge injection[EB/OL]. (2023-06-28)[2023-09-06]. <http://arxiv.org/abs/2306.16186v1>.
- [19] Ji W, Li J J, Bi Q, et al. Segment anything is not always perfect: an investigation of SAM on different real-world applications[EB/OL]. (2023-05-22)[2023-12-07]. <http://arxiv.org/abs/2304.05750>.
- [20] Zhang K D, Liu D. Customized segment anything model for medical image segmentation[EB/OL]. (2023-10-17)[2023-11-29]. <http://arxiv.org/abs/2304.13785>.
- [21] Chen K Y, Liu C Y, Chen H, et al. RSPrompter: learning to prompt for remote sensing instance segmentation based on visual foundation model[EB/OL]. (2023-06-28)[2023-11-29]. <http://arxiv.org/abs/2306.16269>.
- [22] He K M, Chen X L, Xie S N, et al. Masked autoencoders are scalable vision learners[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. New York: IEEE Press, 2022: 15979-15988.
- [23] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale[EB/OL]. (2020-10-22)[2023-12-11]. <https://arxiv.org/abs/2010.11929>.
- [24] Hu J E, Shen Y L, Wallis P, et al. LoRA: low-rank adaptation of large language models[EB/OL]. (2021-10-16)[2023-12-20]. <http://arxiv.org/abs/2106.09685>.
- [25] Wu J D, Ji W, Liu Y P, et al. Medical SAM adapter: adapting segment anything model for medical image segmentation[EB/OL]. (2023-05-13)[2023-09-23]. <http://arxiv.org/abs/2304.12620>.
- [26] Zhang C N, Puspitasari F D, Zheng S, et al. A survey on segment anything model (SAM): vision foundation model meets prompt engineering[EB/OL]. (2023-07-03)[2023-12-01]. <http://arxiv.org/abs/2306.06211>.
- [27] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[EB/OL]. (2015-12-10)[2024-01-24]. <http://arxiv.org/abs/1512.03385>.
- [28] Mannor S, Peleg D, Rubinstein R. The cross entropy method for classification[C]//Proceedings of the 22nd International Conference on Machine Learning, August 7-11, 2005, Bonn, Germany. New York: Association for Computing Machinery, 2005: 561-568.
- [29] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[EB/OL]. (2018-02-07)[2024-01-25]. <http://arxiv.org/abs/1708.02002>.
- [30] Li X Y, Sun X F, Meng Y X, et al. Dice loss for data-imbalanced NLP tasks[EB/OL]. (2020-08-29)[2024-01-25]. <http://arxiv.org/abs/1911.02855v3>.
- [31] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[EB/OL]. (2017-12-05)[2023-12-26]. <http://arxiv.org/abs/1706.05587v3>.
- [32] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI 2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.

Powder-Spreading Defect Detection in Laser Powder Bed Fusion Based on Large Vision Model

Tan Kunpeng¹, Tang Jiafeng¹, Zhao Zhibin^{1*}, Wang Chenxi¹, Zhang Xingwu¹, He Weifeng²,
Chen Xuefeng¹

¹*Institute of Aero-Engine, School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, Shaanxi, China;*

²*National Key Lab of Aerospace Power System and Plasma Technology, Air Force Engineering University, Xi'an 710038, Shaanxi, China*

Abstract

Objective To date, laser powder bed fusion (LPBF) is considered as the most advanced metal additive manufacturing technology. It has been widely adopted for the production of critical metal components in aerospace and healthcare industries. However, realizing quality stability and consistency is challenging because of the coupled effects of various factors during LPBF. The powder-spreading quality is a crucial characteristic of LPBF process monitoring. Defects during powder spreading can introduce defects into the formed components. In recent years, the application of computer vision in powder-spreading defect detection has shown promising results. However, the limited availability of annotated data constrains its performance. Large vision models, such as the segment anything model (SAM), exhibit remarkable generalization capabilities owing to pre-training on an extremely large dataset. This allows its transfer to various downstream tasks with minimal training data. However, owing to the lack of defect knowledge, absence of category information, and dependence on manual prompts, SAM cannot be directly applied to powder-spreading defect segmentation. This study addresses the requirements for powder-spreading defect segmentation by improving SAM, achieving excellent defect segmentation performance with minimal training samples, and exploring the potential application of large vision models in monitoring the additive manufacturing process.

Methods In this study, the powder-spreading defect segment anything model (PSAM), based on SAM, was introduced. The overall structure of PSAM was similar to that of SAM, which consisted of an image encoder, an auto-prompt generator, and a mask decoder. Compared to the original SAM, PSAM incorporated the following improvements: To address the issue of knowledge transfer concerning SAM's pre-trained parameters, four Adapter modules were introduced into the SAM image encoder structure. These Adapter modules enabled efficient adjustment of image feature encoding. They were inserted behind the multi-head attention layer in the transformer module and comprised linear and convolutional layers. To satisfy the requirements for category information in the powder-spreading segmentation task, PSAM utilized an improved mask decoder. This decoder outputted segmentation masks that were equal to the number of categories in a single pass. Each output corresponded to a specific category-segmentation result. These outputs were then integrated to obtain a classification output. To overcome the challenges of manual prompting in industrial settings, an autoprompt generator was designed. This generator is a fully convolutional neural network with residual connections that extracts features from input images and generates prompt embeddings that can be used by a mask decoder. A combination of cross-entropy loss, focal loss, and Dice loss was used as the final loss function. Specifically, mean intersection over union (mIoU) was employed as the evaluation metric.

Results and Discussions This study utilizes an off-axis industrial camera to acquire powder-spreading images during the formation of several components. A subset of these images is selected for pixel-level annotation and is categorized into six classes: background, super-elevation, incomplete, hopping, streaking, and lattice. The images and their corresponding labels are organized into a dataset and are divided according to certain proportions. The model is trained using the training set, and even with only 50 training images, PSAM exhibits excellent segmentation performance. The evaluation of PSAM using the test set yields an mIoU of 65.02%, representing an improvement of 8.51 percentage points over Deeplab v3 and 5.31 percentage points over U-Net (Table 1). The limited amount of data restricts the ability of the model to perform deeper feature learning, thus hindering its capacity for richer feature representation. However, the pretrained SAM possesses strong image-feature extraction capabilities. Therefore, excellent defect segmentation performance can be realized without extensive data training. Ablation experiments are conducted to evaluate the proposed improvements. The results indicate that the introduction of the Adapter module successfully transfers feature representation capabilities, and the automatic prompting effectively guides the mask decoder's output. Compared with the original SAM, PSAM realizes an mIoU improvement of 11.33 percentage points (Table 2).

Conclusions The design of PSAM based on SAM realizes excellent segmentation performance even with a small amount of training data. Firstly, the model achieves transferring image encoder feature extraction capabilities from natural images to powder-spreading images by introducing the Adapter modules. Second, the mask decoder is modified to output the category masks. Finally, by incorporating an auto-prompt generator to encode the input images, automatic generation of visual prompt embeddings is achieved.

Large artificial intelligence (AI) models are rapidly advancing. However, owing to the unique characteristics and complex operating conditions of industrial settings, these models still face challenges in practical applications in industrial scenarios. This study provides a preliminary exploration of the application of large vision models for powder-spreading defect detection in additive manufacturing. However, the full potential of large vision models has yet to be fully realized. The outstanding feature extraction, zero-shot generalization, and multimodal knowledge fusion capabilities of these large models can provide new solutions and approaches for additive manufacturing process monitoring, which are worth further exploration in future research.

Key words laser technique; laser powder bed fusion; process monitoring; large vision model; defect detection