https://doi.org/10.12442/j.issn.1002-185X.20250360

基于联合注意力机制的GH4169合金晶粒分割算法研究

邓博文，李 科，张馨元，赵燕茹，史蒲英，王 虎，高慧贤，杨 超
（西部超导材料科技股份有限公司，陕西 西安 710018）

摘 要：针对高温合金晶粒形貌复杂、尺寸差异大、边界模糊且晶粒紧密连接导致分割难题，本研究提出一种融合通道与空间信息的联合抑制注意力机制的晶粒度分割网络。该网络结合Swin Transformer的全局建模与深度卷积神经网络（convolutional neural network，CNN）的局部细节恢复能力，并在解码器中引入融合通道与空间信息的联合抑制注意力机制，有效抑制噪声和纹理干扰，提升特征筛选与泛化能力，同时强化浅层与深层特征融合，显著改善晶粒边界连贯性。结果表明，所提算法在自制金相组织数据集上IoU、F1-score分别可达到67.34%和78.62%，各项指标均优于主流高温合金中的晶粒分割算法。

关键词：高温合金；金相图像分割；注意力机制；CNN
中图法分类号：TG132.3^{+}3 **文献标识码：**A **文章编号：**1002-185X（2026）10-2565-08

1 引 言

精确的晶粒结构分布信息对于高温合金材料的性能评估、工艺优化以及服役寿命预测等研究具有重要意义[1]。显微图像中晶粒的语义分割是材料微观组织智能分析的重要组成部分，其主要任务是从高分辨率金相图像中提取晶粒相关特征，实现对图像中所有像素的语义分类，从而完成对晶粒区域的精确分割与描述。

然而，高温合金中的晶粒结构常常呈现形貌复杂、尺寸差异大、边界模糊及晶粒连接紧密等特性，显著增加了分割任务的难度。高温合金在制备过程中，受成分、温度、冷却速率等多种因素影响，晶粒形态变化多样，从细小的等轴晶到长条状的柱状晶均有可能出现，这种复杂形貌使得通用的分割算法难以准确识别每个晶粒。晶粒尺寸方面，同一图像中晶粒尺寸差异可达数倍甚至数十倍，小晶粒容易在分割中被遗漏或误判，大晶粒的细节特征也可能因分辨率不足而丢失。同时，由于晶粒生长过程中相互挤压，边界部分模糊不清，紧密相连的晶粒更是难以区分。尤其是在存在孪晶、亚晶界以及第二相颗粒等微观组织时，晶界位置往往难以准确判定。孪晶结构与正常晶界的灰度和纹理特征相似，仅存在取向差异，这对基于视觉特征的分割算法构成巨大挑战；亚晶界通常较细且对比度低，在图像中难以清晰呈现；第二相颗粒的存在可能干扰分割模型对晶界的判断，其位置和形态变化也增加了分割的复杂性。此外，因样品制备过程中的腐蚀不均匀、光照条件变化、成像对比度低以及背景夹杂等因素，也进一步干扰了晶粒边界的识别与提取，给高精度自动分割带来了严峻挑战。蚀刻不均会导致同一晶粒内部灰度差异过大，光照条件变化使图像局部亮度不一，成像对比度低模糊了晶界细节，而背景夹杂的存在则可能被误判为晶粒或晶界，这些问题都极大地影响了分割的准确性和可靠性。

随着高通量显微成像、图像识别与深度学习等技术的快速发展，高分辨率材料微观组织图像的语义分割问题受到了广泛关注。针对该问题，众多研究者提出了多种基于传统图像处理和深度学习的分割方法[2–4]，以期实现对晶粒、析出相及界面结构等微观组织的精准识别。具体而言，传统方法主要依赖边缘检测、阈值分割和形态学操作等手段，尽管在特定条件下具有一定效果，尤其对人工先验所得构造特征提取模型依赖严重，进而对噪声、边界模糊和形貌复杂性敏感，导致方法的泛化能力有限。

近年来，深度卷积神经网络（convolutional neural network, CNN）[5]、注意力机制[6]以及Vision Transformer[7]等新型架构的引入，显著推动了材料显微组织图像分割技术的发展。特别是在语义分割任务中，诸如U-Net[8]及其多种变体、DeepLab系列[9]、Mask R-CNN[10]等网络模型，结合图像增强策略与边界感知机制，被广泛应用于各类复杂微观结构的识别任务，并在多个实际应用场景中取得了良好的分割性能与推广效果。如Munoz-Rodenas等人[11]基于DeepLabV3+、U-Net和SegNet等典型分割网络对钢铁材料的显微图像进行了对比研究，并分别引

收稿日期：2025-10-08
基金项目：陕西省重点研发计划（2025CY-YBXM-237）
作者简介：邓博文，男，1997年生，硕士，西部超导材料科技股份有限公司，陕西 西安 710018，电话：029-86514511，E-mail：dbw@c-wst.com

入预训练的 ResNet50、ResNet18 和 MobileNetv2 作为特征提取器，用于低碳钢退火试样中珠光体与铁素体相的自动分割任务。验证了深度语义分割网络在钢铁组织识别中的有效性；Pokharel 等人[12]结合 DeepLabV3Plus 等模型优化多尺度孔隙识别，有效解决不规则结构与杂质干扰下的自动化表征难题；Hirabayashi 等人[13]将神经网络的语义分割技术应用于多晶陶瓷三维成像的二次电子图像，利用全卷积网络和 U-Net 等深度学习模型，借助改进深度准确性的训练图像，成功在数分钟内重建出单个体素分辨率 20 nm 的千兆级体素 3D 显微结构；Zhou 等人[14]针对高温合金多相遮挡致显微结构分割难的问题，提出 MPU-Net 方法，通过端到端同步检测晶界与析出物，并设计两种后处理策略优化边界检测，经 OM 和 FESEM 图像验证，该方法性能优于现有方案，实现高效自动化显微结构分析；Sun 等人[15]针对金相组织图像标注数据稀缺与复杂显微结构分割难题，设计一种自适应并行注意力模块（adaptive parallel attention module, APAM）和融入多尺度扩张卷积的多头自适应注意力模块，结合逐通道注意力与 LSTM 建模通道关系，在 MetalDAM 和 UHCS 公开数据集上达到了较好的效果；Duan 等人[16]提出基于改进通道注意力机制的对抗生成网络（generative adversarial network，GAN）方法，嵌入通道注意力模块并结合全局上下文机制，用 Focal Loss 优化损失函数，有效缓解了原始金相图像噪声多、边界缺失等导致晶粒提取难的问题。

高温合金微观组织图像中，晶粒结构呈现出显著的非均质性，表现为尺度差异大、边界模糊、纹理复杂以及背景噪声干扰强烈等特征。为提升对复杂结构的识别能力，当前图像分割网络多采用多尺度特征融合策略，将浅层的细节信息与深层的语义信息进行结合。然而，在微观组织图像这一纹理细腻、噪声丰富的场景下，浅层特征中往往包含大量与晶粒无关的背景纹理或伪边界响应，若在融合阶段直接与深层语义特征连接，极易引入冗余信息和误导性响应，进而导致模型产生特征混淆的风险。为缓解这一问题，研究者们逐步引入注意力机制来增强网络的特征选择能力。其中空间注意力机制可有效地增强网络在空间上对于长距离依赖关系与上下文信息的能力。而通道注意力机制可以引导网络自适应地对目标通道进行选择，进而可增强网络对待分割目标有效特征的关注能力，二者结合在一定程度上提升了模型在结构复杂图像中的分割性能。典型代表如 SENet、CBAM、DANet 等已在金属微观组织图像等边界模糊场景中取得良好效果。尽管上述基于空间注意力以及通道注意力的方法在一定程度上改善了模型的特征提取能力，当前多数主流注意力机制核心关注于“增强有效特征”，相对忽略对“抑制干扰信息”的显式建模能力，特别是在多尺度特征融合过程中，对于浅层引入的冗余纹理响应缺乏有效控制，仍可能影响整体分割质量。

针对上述痛点，本研究提出一种面向高温合金 GH4169 晶粒语义分割任务的高性能分割网络—SwinGrainNet。该网络采用 Transformer 与 CNN 相结合的编解码结构，充分融合 Swin Transformer 在全局上下文建模方面的优势与 CNN 在局部特征提取和空间细节恢复方面的能力，提升模型对复杂晶粒结构的特征挖掘能力。特别地，在解码器阶段引入通道-空间联合抑制注意力机制（channel-spatial dropout attention module，CSDA），通过融合通道注意力与空间注意力，有效增强浅层与深层特征融合的选择性和判别性。同时，设计抑制融合策略，显式建模“信息增强+噪声抑制”的融合过程，有效过滤浅层特征中干扰性的背景纹理与伪边界响应，缓解浅层噪声对深层语义表达的干扰问题，从而提升边界建模的鲁棒性与晶粒分割的准确性。

2　实　验

为了解决上述问题，本文提出一种基于 Transformer 与 CNN 编解码结构的高性能分割网络（SwinGrainNet）：SwinGrainNet 的编码器基于 Swin Transformer 架构，解码器基于 ConvNext 架构。编码器主要用于提取高温合金金相组织图像中晶粒的纹理、区域边界等浅层特征以及类别、分布等深层语义特征。解码器主要用于对特征图进行分辨率恢复，采用渐进式上采样对已提取特征图进行分辨率恢复。SwinGrainNet 网络总体结构如图 1 所示。

2.1　编解码器骨干网络

（1）基于 Swin Transformer 的编码器网络

网络中基于 Swin Transformer 的特征提取模块主要由两个阶段组成。第 1 阶段通 Patch Partition 和 Patch Embedding 操作提取图像的低级局部结构特征。Patch Partition 将输入图像划分为固定尺寸的图像块（如 4×4），随后使用卷积核大小为 4×4、步长为 4、填充为 0 的卷积层对图像块进行线性嵌入（Patch Embedding），输出维度为 96 的初始特征表示。该阶段接一个 LayerNorm 层用于稳定特征分布，提升网络训练的收敛性能。

第 2 阶段为层级化的 Swin Transformer 编码模块，由 4 个连续阶段（Stage 1 至 Stage 4）组成，每个阶段均包含多个 Swin Transformer Block。具体地，Stage 1 至 Stage 4 分别包含 2、2、6、2 个 Block，Block 内部由改进的窗口多头自注意力模块与跨窗口注意力模块组成，以建模图像中的空间上下文与长程依赖关系。在每两个阶段之间插入 Patch Merging 模块，以步长为 2 的窗口进行空间降采样，同时将通道数扩大 2 倍，以保留更丰富的语义信息，进而能够有效提取金相图像中的局部纹理、边界轮廓及深层语义特征。其参数如表 1 所示。

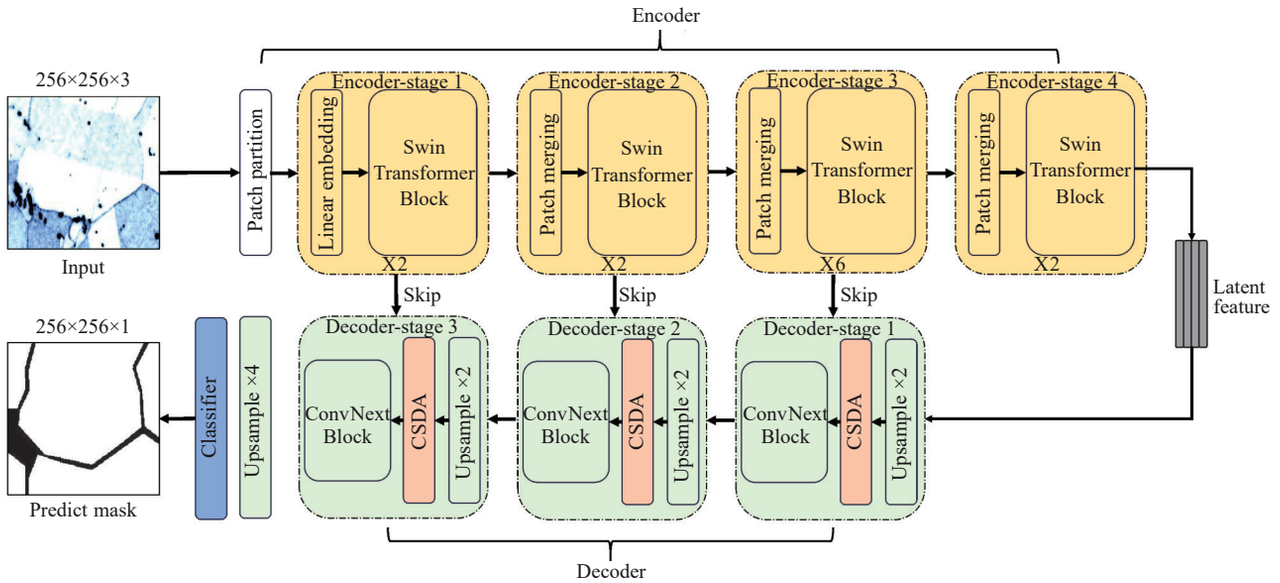


图1　SwinGrainNet网络整体结构图

Fig.1　Schematic of SwinGrainNet

表1　编码器网络参数表

Table 1　Encoder module configuration

Stage	Input size (height×width)	Number of block	Number of channels	Downsampling method
Patch embedding	256×256	-	96	Conv (4×4, stride=4, padding=0)
Stage 1	64×64	2	96	-
Stage 2	32×32	2	192	Patch merging (stride =2)
Stage 3	16×16	6	384	Patch merging (stride =2)
Stage 4	8×8	2	768	Patch merging (stride =2)

（2）基于ConvNeXt的解码器网络

网络中基于ConvNeXt的解码器模块旨在逐步恢复空间分辨率并融合编码器提取的多尺度特征，从而实现更高精度、更具结构完整性的金相图像边界分割输出。该模块包含4个解码阶段，分别对应于编码器中4个不同分辨率层级（Stage 1 至 Stage 4）的输出特征。在解码过程中，首先将编码器末端输出的深层特征图通过双线性插值进行上采样，并与上一层编码器输出进行通道拼接，以融合语义与空间信息。拼接后的特征经过通道-空间协同的注意力抑制模块（CSDA）处理，能够突出图像中关键结构区域并有效抑制背景噪声。

在完成多尺度特征融合后，解码器引入轻量化的ConvNeXt Block以进一步对融合特征进行深度重建与表达增强。该模块通过深度可分离卷积与逐点卷积的组合实现对局部空间上下文的高效建模，同时配合LayerNorm归一化层与GELU激活函数增强非线性表达能力与特征选择性。此外，在通道维度与空间维度中引入Dropout正则化策略，有效缓解过拟合风险并提升模型泛化性能。解码器整体沿用渐进式上采样的解码机制，逐层恢复空间分辨率并整合多层语义特征，最终生成具有精细边界结构的分割预测图。在保持较高计算效率的同时，解码器充分发挥远程上下文感知与局部边缘建模的优势，显著增强对金相图像中晶粒边界与细节纹理的辨识能力。其具体参数配置如表2所示。

2.2　通道-空间协同的注意力抑制模块（CSDA）

在解码器的特征融合阶段，本研究设计了一种融合通道与空间信息的联合抑制注意力机制，用以实现特征筛选与泛化能力的提升。该模块通过动态调节特征图中

表2　解码器网络参数表

Table 2　Decoder module configuration

Stage	Input size (height×width)	Number of channels	Upsampling method
1	16×16	512	Upsample (scale=2)
2	32×32	256	Upsample (scale=2)
3	64×64	128	Upsample (scale=2)
4	256×256	2	Upsample (scale=4)

不同空间区域及通道维度的关注权重，有效抑制噪声和纹理突变等干扰信息，强化模型对目标晶粒区域的聚焦与判别能力。同时，结合抑制融合机制，模块进一步增强了模型的泛化性能，有效防止过拟合的发生。此外，该模块通过强化浅层与深层特征的融合与交互，显著提升晶粒边界的连贯性与完整性，实现了多尺度信息的高效补充与利用。令$F_{\mathrm{in}} \in R^{H_{\mathrm{f}} \times W_{\mathrm{f}} \times C_{\mathrm{f}}}$表示输入 CSDA 的特征图，$F_{\mathrm{out}} \in R^{H_{\mathrm{f}} \times W_{\mathrm{f}} \times C_{\mathrm{q}}}$表示 CSDA 输出特征图，其中，$H_{\mathrm{f}}$，$W_{\mathrm{f}}$分别为输入 CSDA 的特征图的高、宽，$C_{\mathrm{f}}$，$C_{\mathrm{q}}$分别代表输入输出特征图通道数。CSDA 模块的流程描述如下：

首先，将输入的特征图F_{in}输入通道注意力模块（channel attention module，CAM）中。该模块旨在对不同通道的重要性进行建模，从而增强具有判别性的特征并抑制冗余信息。具体地，通道注意力模块首先对输入特征图在空间维度上分别进行全局平均池化（global average pooling）和全局最大池化（global max pooling），以不同的特征压缩方式，获得特征图不同的通道描述向量。将这两个向量分别输入共享权重的卷积层进行特征变换，随后，将两个输出向量通过逐元素相加的方式融合，并通过 Sigmoid 激活函数生成归一化的通道注意力权重。最后将权重向量与输入特征图进行逐元素乘法，以实现对不同通道的重要性加权调制。其数学表达如式（1）至式（3）所示。

$$F_{\mathrm{avg}}^{\mathrm{c}} = f_{\mathrm{conv}}\{ f_{\mathrm{leakyReLU}}[f_{\mathrm{conv}}(f_{\mathrm{s_avgpool}}(F_{\mathrm{in}}))]\} \tag{1}$$

$$F_{\mathrm{max}}^{\mathrm{c}} = f_{\mathrm{conv}}\{ f_{\mathrm{leakyReLU}}[(f_{\mathrm{conv}}(f_{\mathrm{s_maxpool}}(F_{\mathrm{in}}))]\} \tag{2}$$

$$F_{\mathrm{cam}} = f_{\mathrm{sigmoid}}(F_{\mathrm{avg}} + F_{\mathrm{max}}) \otimes F_{\mathrm{in}} \tag{3}$$

其中，$F_{\mathrm{avg}}^{\mathrm{c}} \in R^{1 \times 1 \times C_{\mathrm{f}}}$代表空间自适应平均池化分支输出特征图；$F_{\mathrm{max}}^{\mathrm{c}} \in R^{1 \times 1 \times C_{\mathrm{f}}}$代表空间自适应最大池化分支输出特征图；$F_{\mathrm{cam}} \in R^{H_{\mathrm{f}} \times W_{\mathrm{f}} \times C_{\mathrm{f}}}$代表通道注意力模块输出特征图；$f_{\mathrm{conv}}$代表卷积操作；$f_{\mathrm{leakyReLU}}$代表带泄露线性整流单元操作；$f_{\mathrm{s_avgpool}}$代表输出空间大小为 1 的空间自适应平均池化操作；$f_{\mathrm{s_maxpool}}$代表输出空间大小为 1 的空间自适应最大池化操作。

其次，将输入的特征图F_{in}输入空间注意力模块（spatial attention module，SAM）。SAM 旨在通过捕捉特征图中显著区域的空间分布模式，引导模型关注关键的空间区域。具体地，空间注意力模块首先对输入特征图沿通道维度分别执行最大池化和平均池化操作，后将这两个空间图在通道维度进行拼接，输入卷积核大小为 7 × 7 的卷积层，进而对特征图局部区域的显著性空间关系进行捕获。最终，通过 Sigmoid 激活函数输出空间注意力图，并与输入特征图按空间维度进行逐元素相乘，实现空间维度上的特征重加权。其数学表达如式（4）至式（6）所示：

$$F_{\mathrm{avg}}^{\mathrm{s}} = f_{\mathrm{c_avgpool}}(F_{\mathrm{in}}) \tag{4}$$

$$F_{\mathrm{max}}^{\mathrm{s}} = f_{\mathrm{c_maxpool}}(F_{\mathrm{in}}) \tag{5}$$

$$F_{\mathrm{sam}} = f_{\mathrm{sigmoid}}\{ f_{\mathrm{conv}}[f_{\mathrm{concat}}(F_{\mathrm{avg}}^{\mathrm{s}}, F_{\mathrm{max}}^{\mathrm{s}})]\} \otimes F_{\mathrm{in}} \tag{6}$$

其中，$F_{\mathrm{avg}}^{\mathrm{s}} \in R^{H_{\mathrm{f}} \times W_{\mathrm{f}} \times 1}$代表通道自适应平均池化分支输出特征图；$F_{\mathrm{max}}^{\mathrm{s}} \in R^{H_{\mathrm{f}} \times W_{\mathrm{f}} \times 1}$代表通道自适应最大池化分支输出特征图；$F_{\mathrm{cam}} \in R^{H_{\mathrm{f}} \times W_{\mathrm{f}} \times C_{\mathrm{f}}}$代表空间注意力模块输出特征图；$f_{\mathrm{c_avgpool}}$代表输出通道数量为 1 的通道自适应平均池化操作；$f_{\mathrm{s_maxpool}}$代表输出通道数量为 1 的通道自适应最大池化操作。

为了进一步提升模型的泛化能力与鲁棒性，本模块引入了抑制融合策略，即在通道注意力与空间注意力输出之间添加随机失活机制。该机制在训练阶段随机抑制部分通道或空间响应，从而有效缓解过拟合问题。最终的输出特征图可表示为$F_{\mathrm{out}}^{\mathrm{m}} \in R^{H_{\mathrm{f}} \times W_{\mathrm{f}} \times 2C_{\mathrm{f}}}$，其数学表达如式（7）所示：

$$F_{\mathrm{out}}^{\mathrm{m}} = f_{\mathrm{concat}}[f_{\mathrm{dropout}}(F_{\mathrm{sam}}), f_{\mathrm{dropout_2d}}(F_{\mathrm{cam}})] \tag{7}$$

在最终的特征融合模块（feature fusion module, FFM）中，为增强特征的非线性表达能力，首先对式(7)得到的中间特征图$F_{\mathrm{out}}^{\mathrm{m}}$进行通道维度扩张至 2 倍。随后，引入通道重排机制（channel shuffle），通过通道分组与交叉打乱，促进跨通道信息交互。后再经 1 × 1 卷积将特征图还原至目标通道数。为提升梯度传递的稳定性与特征复用能力，引入残差连接，最终实现特征表达能力增强。其数学表达如式（8）所示：

$$F_{\mathrm{ffm}} = f_{\mathrm{conv}}\{ f_{\mathrm{conv}}[f_{\mathrm{c_shuffle}}(f_{\mathrm{conv}}(F_{\mathrm{out}}^{\mathrm{m}}))] + F_{\mathrm{out}}^{\mathrm{m}}\} \tag{8}$$

CSDA 模块整体结构如图 2 所示。

2.3　实验设置

实验平台工作站搭载 Intel Xeon E5 2650 处理器，376 GB 内存，3 个 NVIDIA L4 24G 显卡；深度学习框架采用 Pytorch-1.8，以及 NVIDIA 公司 CUDA11.2 的 GPU 运行平台以及 cuDNN8.0 深度学习 GPU 加速库。实验数据集为分辨率 2560 × 2048 的 GH4169 金相组织图像，裁剪至 256 × 256 分辨率后共包含 800 张局部组织图像，按照 7:2:1 进行数据集划分，其中训练集 560 张、验证集 160 张、测试集 80 张。SwinGrainNet 网络训练过程中，输入裁剪后分辨率 256 × 256 的金相组织图像，使用随机水平翻转与归一化操作策略实现对金相组织图像数据的增强。实验中，网络设置训练参数批大小为 12、学习轮数为 151、初始学习率为 1×10^{-3}，使用 AdamW 优化器进行优化。网络训练过程中，使用余弦学习率调整策略，调整周期为 20 轮，最小学习率设置为 1×10^{-6}。

3　结果与分析

3.1　实验评价指标

为了验证提出算法的有效性，采用交并比(intersection over union，IoU)和 F1 分数（F1-score）两个评估指标进行算法有效性对比评价，各指标定义公式表示

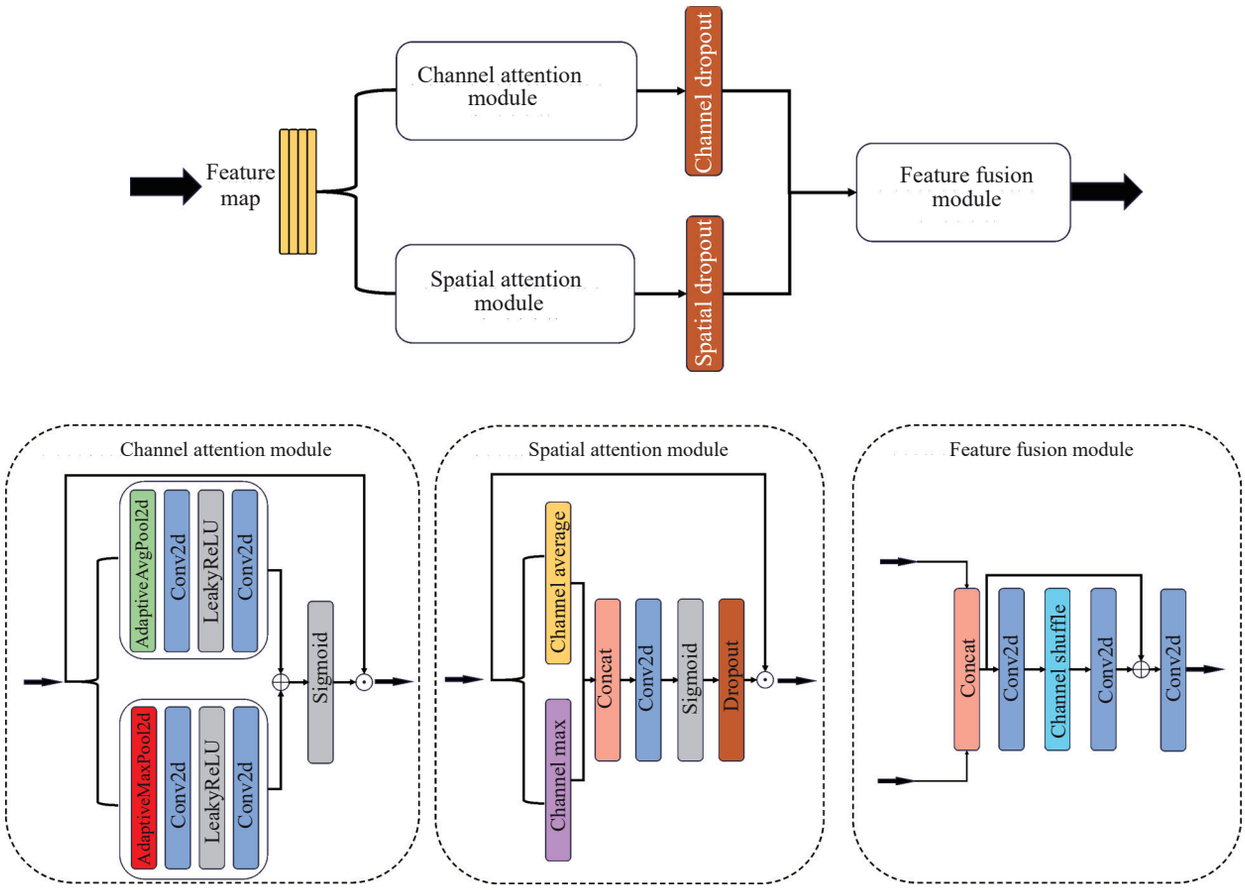


图2 CSDA模块整体结构图

Fig.2 Schematic of CSDA

如式(9)与式(10)：

$$\mathrm{IoU}=\frac{TP}{TP+FN+FP} \tag{9}$$

$$\text{F1-score}=\frac{2\cdot\text{precision}\cdot\text{recall}}{\text{precision}+\text{recall}} \tag{10}$$

式中，TP为金相组织图像分割结果中晶粒类别分割正确的像素点总数；FP为金相组织图像分割结果中晶粒分割错误的像素点总数；FN为金相组织图像分割结果中背景类别分割错误的像素点总数。

3.2　消融实验

为了验证所提出的CSDA模块的有效性，在自制金相图像数据集上开展了消融实验。消融实验主要从分割结果的IoU、F1-score等指标进行对比，各变种模型的相关参数设置和训练策略均相同，以文献[8]的语义分割模型作为基线模型(Baseline)。表3给出网络各变种的配置与说明，消融实验对比结果如表4所示。

由表4可知，Baseline + CAM + FFM比Baseline分别提高2.02%(IoU)和0.86%(F1-Score)；Baseline + SAM + FFM比Baseline分别提高1.78%(IoU)和0.78%(F1-Score)；Baseline+CAM+ SAM +FFM比Baseline分别提高2.98%(IoU)和2.09%(F1-Score)；Baseline+CAM+SAM+FFM+dropout比Baseline分别提高3.26%(IoU)和2.39%(F1-Score)。分析表4可知，对比Baseline，引入通

表3 SwinGrainNet模型变种结构表

Table 3 Variant structure of SwinGrainNet

Network variant	Detail
Baseline	Baseline model Swin Transformer+ConvNeXT
Baseline+CAM+FFM	Channel attention module+Feature fusion module
Baseline+SAM+FFM	Spatial attention module+Feature fusion module
Baseline+CAM+SAM+FFM	Channel attention module+Spatial attention module+Feature fusion module
Baseline+CAM+SAM+FFM+Dropout	Channel attention module+Spatial attention module+Feature fusion module +Dropout

道注意力模块、空间注意力模块、特征融合模块以及特征融合抑制机制后模型性能有较为明显的提升，证明所提出的模块能显著提升模型对待分割晶粒的区域与边界信息的关注能力。

3.3　对比实验

本小节为验证所提SwinGrainNet有效性，与几个主流先进语义分割网络在金相数据集上进行了晶粒分割对比实验。对比方法包括FCN[17]、Unet[8]、SegNet[18]、DeeplabV3[9]、PSPNet[19]以及Swin-Unet[20]。

图3为各类对比方法在金相数据集部分分割结果，从左到右，图中第1列Input为输入测试的金相图晶粒图像，第2列Ground Truth为真实标签图像。第3列到第8列分别为FCN、Unet、SegNet、DeeplabV3、PSPNet、Swin-Unet算法的语义分割结果，第9列Ours为本文所提SwinGrainNet的语义分割结果。

（1）定性分析。图3a~3g分别展示了在晶粒组织中存在第二相颗粒干扰、腐蚀不均匀、孪晶及晶粒边缘特征不明显的复杂情况下所提出的SwinGrainNet与其余先进算法的分割效果比较。总体来看，提出的算法SwinGrainNet相较于FCN、Unet、Swin-Unet等算法能更

表4 消融实验结果

Table 4 Ablation experimental results

Method	IoU/%	F1-score/%
Baseline	65.01	76.78
Baseline+CAM+FFM	66.32	77.44
Baseline+SAM+FFM	66.17	77.38
Baseline+CAM+SAM+FFM	66.95	78.39
Baseline+CAM+SAM+FFM+dropout	67.13	78.62

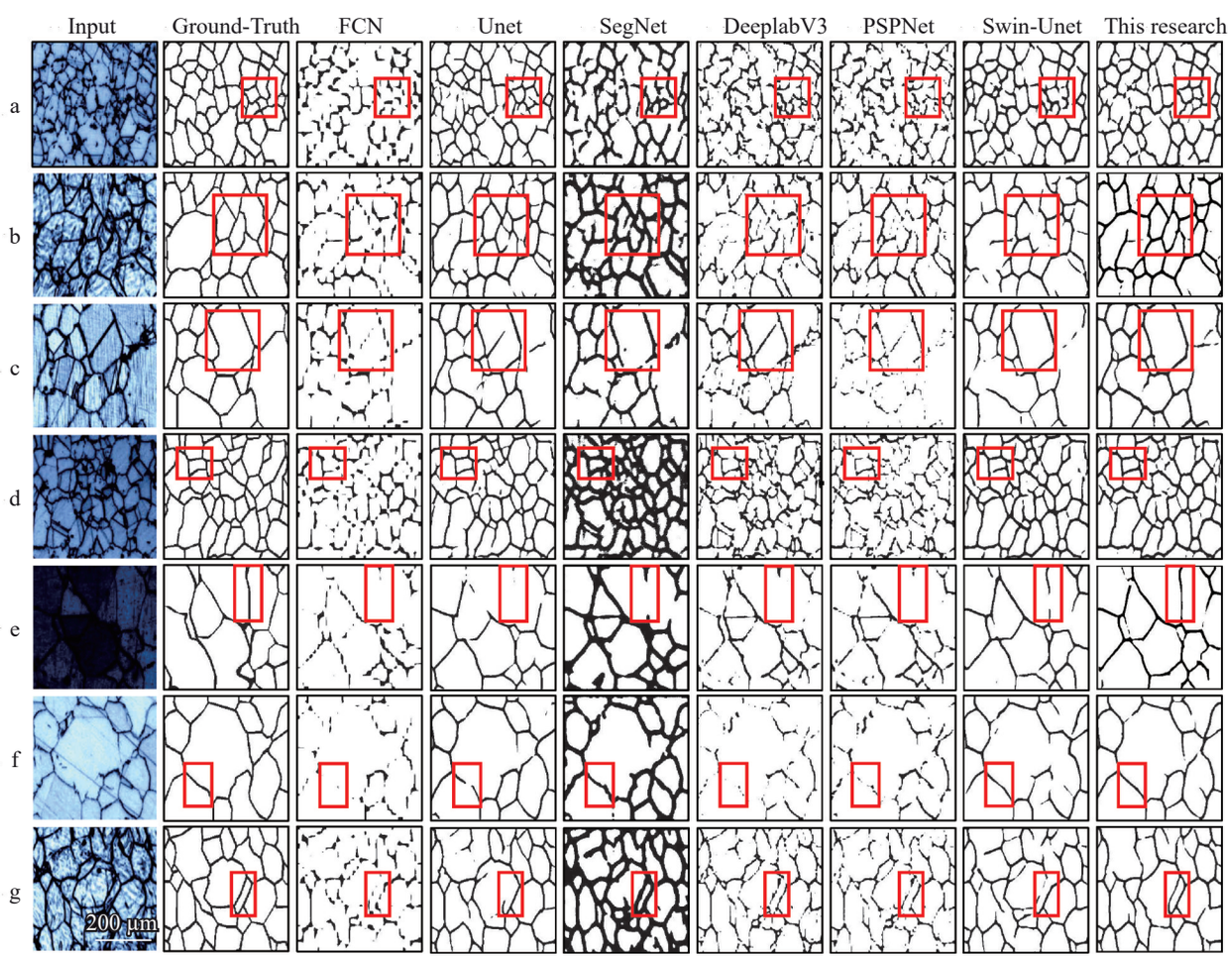


图3 金相图像数据集定性对比实验结果

Fig.3 Qualitative comparison results of metallographic image dataset: (a)–(g) OM images of different areas for the sample

表5 金相数据集定量对比实验结果

Table 5 Quantitative comparison on metallographic image datasets

Method	IoU/%	F1-score/%
FCN	60.47	72.37
UNet	64.87	76.44
SegNet	60.27	72.25
DeepLabV3	59.34	71.18
PSPNet	59.98	71.95
SwinUnet	66.58	77.89
SwinGrainNet	67.34 ↑	78.62 ↑

好地将金相组织图像中晶粒的空间细节信息以及语义分析，使得SwinGrainNet能较好地缓解晶粒边界不清晰及受到孪晶、阴影等干扰信息影响所导致的分割不精确问题；对于受到阴影、光线影响较大的情况，如图3a、3d所示，对比FCN、Unet、Swin-Unet等算法，所提出算法能精确的关注到不显著晶粒的边缘与区域形状，从而实现对受到阴影、光线影响较大晶粒边界的精确分割；如图3b、3c、3g所示，金相组织图像中的真实边界受到腐蚀不均匀及孪晶等问题的影响，相较于FCN、Unet、Swin-Unet等算法，所提算法可有效地注意到各类腐蚀不均匀及孪晶影响下的准确晶粒边界，从而实现腐蚀不均匀及孪晶等复杂因素影响下的晶粒边界分割；如图3e、3f所示，在过曝或欠曝等极端成像条件下，相较于FCN、Unet、SegNet、PSPNet等网络本研究所提出的SwinGrainNet依能鲁棒的提取关键边界特征，分割性能较优。综上所述，从金相数据集的对比实验效果可以看出，提出的基于通道-空间协同的注意力抑制模块的金相语义分割方法可有效对复杂场景下的晶粒边界进行提取。

（2）定量分析。不同算法在金相数据集上分割结果的定量分析表如表5所示，定量分析所采用的指标主要为交并比(IoU)和F1分数(F1-score)。由表5可知，提出的SwinGrainNet模型在自制金相数据集上IoU为67.34%，F1-score为78.62%。实验结果表明，所提出的SwinGrainNet相较于其他对比方法在各项指标上均取得了显著提升。在IoU指标上，SwinGrainNet相较于FCN、UNet、SegNet、DeepLabV3和PSPNet分别提升了6.87%、2.47%、7.07%、8.00%和7.36%；在F1-score指标上，分别提升了6.25%、2.18%、6.37%、7.44%和6.67%。此外，与基于Transformer架构的Swin-Unet相比，SwinGrainNet在IoU和F1-score上仍分别提升了0.76%和0.73%。

4 结 论

1）构建了以Swin Transformer为编码器、ConvNeXt为解码器的骨干网络。利用Swin Transformer的全局建模能力提取深层语义特征，结合ConvNeXt的局部特征提取优势进行渐进式分辨率恢复，有效解决了传统方法难以兼顾大尺度结构信息与小尺度细节特征的问题。

2）设计了通道-空间协同的注意力抑制模块（CSDA）。该模块融合通道注意力、空间注意力及抑制融合机制，通过动态调节特征权重增强关键区域聚焦能力，同时利用随机失活策略降低过拟合风险，显著提升了特征融合效果与模型泛化能力。

3）在自建金相图像数据集上的定量与定性实验表明，SwinGrainNet的核心性能指标优于多种先进对比算法。其中，IoU达到67.34%，F1-score达到78.62%，相较于Swin-Unet分别提升了0.76%和0.73%；且在孪晶干扰、腐蚀不均匀及极端成像条件下，仍能实现鲁棒且精确的晶粒边界分割。

参考文献　　References

[1] Cao Hui(曹 卉), Xu Hanzong(徐汉宗), Li Haipeng(李海鹏) *et al*. *Rare Metal Materials and Engineering*(稀有金属材料与工程)[J], 2025, 54(3): 569

[2] Alrfou K, Zhao T, Kordijazi A. *Integrating Materials and Manufacturing Innovation*[J], 2024, 13(3): 703

[3] Tang K K, Zhang P, Zhao Y D *et al*. *Engineering Fracture Mechanics*[J], 2024, 303: 110149

[4] Yan H L, Liang R, Zhang Y *et al*. *Materials Characterization*[J], 2025, 225: 115150

[5] Lecun Y, Bottou L, Bengio Y *et al*. *Proceedings of the IEEE*[J], 1998, 86(11): 2278

[6] Vaswani A, Shazeer N, Parmar N *et al*. *Advances in Neural Information Processing Systems*[J], 2017, 30: 5998

[7] Liu Z, Lin Y T, Cao Y *et al*. *Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows*[C]. Montreal: IEEE, 2021: 10012

[8] Ronneberger O, Fischer P, Brox T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*[C]. Munich: Springer, 2015: 234

[9] Chen L C, Zhu Y K, Papandreou G *et al*. *Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*[C]. Munich: Springer, 2018: 801

[10] He K M, Gkioxari G, Dollár P *et al*. *Mask R-CNN*[C]. Venice: IEEE, 2017: 2961

[11] Munoz-Rodenas J, Garcia-Sevilla F, Miguel-Eguia V *et al*. *Applied Sciences*[J], 2024, 14(6): 2297

[12] Pokharel B, Pandey D S, Sapkota A *et al*. *IEEE Access*[J], 2024, 12(50217–50243): 5

[13] Hirabayashi Y, Iga H, Ogawa H *et al*. *npj Computational Materials*[J], 2024, 10(1): 46

[14] Zhou P, Zhang X Y, Shen X J *et al*. *Computational Materials Science*[J], 2024, 242: 113063

[15] Sun Y L, Huang X Y. *Expert Systems with Applications*[J], 2025, 276: 126925

[16] Duan X Y, Chen Y, Duan X B *et al*. *Materials*[J], 2025, 18(2): 253

[17] Long J, Shelhamer E, Darrell T. *Fully Convolutional Networks for Semantic Segmentation*[C]. Massachusetts: IEEE, 2015: 3431

[18] Badrinarayanan V, Kendall A, Cipolla R. *IEEE Transactions on Pattern Analysis and Machine Intelligence*[J], 2017, 39(12): 2481

[19] Zhao H S, Shi J P, Qi X J *et al. Pyramid Scene Parsing Network*[C]. Honolulu: IEEE, 2017: 2881

[20] Cao H, Wang Y Y, Chen J *et al*. *Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation*[C]. Cham: Springer Nature Switzerland, 2022: 205

Research on Segmentation Algorithm for GH4169 Alloy Grains Based on a Joint Attention Mechanism

Deng Bowen, Li Ke, Zhang Xinyuan, Zhao Yanru, Shi Puying, Wang Hu, Gao Huixian, Yang Chao

(Western Superconducting Technologies Co., Ltd, Xi'an 710018, China)

Abstract: To tackle the segmentation challenges arising from the complex morphology, significant size variations, blurred boundaries, and tight interconnections of superalloy grains, this research presents a grain segmentation network integrated with a joint suppression attention mechanism that fuses channel and spatial information. The proposed network combines the global modeling capacity of Swin Transformer and the local detail restoration capability of a convolutional neural network (CNN), and embeds the aforementioned joint suppression attention mechanism, which integrates channel and spatial information, into the decoder. Results show that this mechanism effectively suppresses noise and texture interference, enhances the abilities of feature screening and generalization, and reinforces the fusion of shallow and deep features, thereby markedly improving the continuity of grain boundaries. Experimental results demonstrate that the proposed algorithm achieves an IoU of 67.34% and an F1-score of 78.62% on the self-constructed metallographic dataset, with all metrics outperforming those of mainstream grain segmentation algorithms for superalloys.

Key words: superalloy; metallographic image segmentation; attention mechanism; CNN

Corresponding author: Yang Chao, Senior Engineer, Western Superconducting Technologies Co., Ltd., Xi'an 710018, P. R. China, E-mail: yangc@c-wst.com