

中图分类号:

单位代号: 10280

密 级:

学 号:

上海大学



硕士学位论文

SHANGHAI UNIVERSITY
MASTER'S DISSERTATION

题 目	面向复杂纹理的材料微观 结构分割与识别方法研究
--------	----------------------------

作 者 杨 珅

学科专业 计算机应用技术

导 师 韩越兴

完成日期 2022.03.01

姓 名:

学号:

论文题目: 面向复杂纹理的材料微观结构分割与识别方法研究

上海大学

本论文经答辩委员会全体委员审查, 确认符合上海大学硕士学位论文质量要求。

主任: 何科

委员: 王书 李欣

导 师: 韩越兴

答辩日期: 2022.06.19

姓 名：

学号：

论文题目：面向复杂纹理的材料微观结构分割与识别方法研究

原创性声明

本人声明：所呈交的论文是本人在导师指导下进行的研究工作。除了文中特别加以标注和致谢的地方外，论文中不包含其他人已发表或撰写过的研究成果。参与同一工作的其他同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

签 名：_____ 杨坤 _____ 日 期： 2022.06.19

本论文使用授权说明

本人完全了解上海大学有关保留、使用学位论文的规定，即：学校有权保留论文及送交论文复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容。

（保密的论文在解密后应遵守此规定）

签 名：_____ 杨坤 _____ 导师签名：_____ 韩越兴 _____ 日 期： 2022.06.19

上海大学工学硕士学位论文

面向复杂纹理的材料微观结构分
割与识别方法研究

姓 名：杨 琿

导 师：韩越兴

学科专业：计算机应用技术

上海大学计算机工程与科学学院

2022 年 3 月

A Dissertation Submitted to Shanghai University for the Degree
of Master in Engineering

**Research on Segmentation and
Recognition Method based on
Materials with Complex Textures**

MA Candidate: Shen Yang

Supervisor: Yuexing Han

Major: Computer Application Technology

**School of Computer Engineering and Science, Shanghai
University**

03, 2022

摘 要

新材料的研发是推动材料科学发展的重要动力,而材料基因工程是一门目前材料科学领域的前沿交叉学科,能够缩短新材料的研发周期及降低时间人力成本,对材料图像微观结构分割与识别能够为材料基因工程中的数据库构建提供数据基础。材料图像处理主要是分为基于非学习的图像处理方法和基于学习的图像处理方法。基于非学习的材料图像处理方法存在效率低下、分割结果不准确等缺点;基于学习的材料图像处理包括机器学习和深度学习技术,但材料图像存在的训练样本不足,纹理复杂和各相分布不平衡等问题阻碍机器学习和深度学习的应用;并且由于材料的微观结构形状各异、各相分布不均,因此大多的材料图像处理方法主要针对特定材料图像而设计,很难找到一种高效且适用性广的材料图像分割方法。为了提高材料图像分割的性能,本文通过分析材料图像的特点,结合目前先进的计算机图像处理技术、机器学习技术及深度学习技术,提出两种高效、精确的材料图像分割方法。

首先,针对材料图像训练样本不足和各相分布不平衡的问题,本文提出一种基于超像素与深度学习的材料图像分割方法。该方法抓住材料图像同相像素具有高度的相似性的特点,先用经典的超像素算法(SLIC)算法获取不同数量的超像素;然后提取每个超像素的最大内接矩形块作为训练样本,扩充了样本数量,解决了材料图像中的小样本问题;接着,将矩形块放入改进的 DenseNet 提取特征并训练生成模型,改进的 DenseNet 块通过添加特征增强模块(FE block)和提出的过渡层上采样方法提高了矩形块识别率;最后,使用滑动窗口在大小为 $n*n$ 像素图像获取 $n*n$ 个矩形,模型预测窗口内矩形块类别,表示中心像素点类别,连接同类别像素点,完成分割任务。同时基于 Focal 损失,提出 Precision Focal 损失,使用机器学习度量——精度替代调节因子中的置信度,聚焦难分类的样本,解决了分类识别任务中的数据分布不平衡的问题,优化训练过程。

接着,针对上述方法中矩形块感知区域较小,不含全局信息,不适用各相纹理相似的材料图像的缺陷,本文提出一种基于复杂纹理特征融合的材料图像分割方法,该方法基于 FCN 模型,在下采样阶段,提出级联的特征融合模块融合高层

与低层语义特征；然后使用多尺度学习模块深度挖掘细粒度纹理信息和全局上下文信息；在上采样阶段，提出一种空间与通道混合的注意力机制模块对特征图附加权重，优化资源配置。同时还基于 Dice 损失提出 CE-Dice 损失，结合交叉熵损失和 Dice 损失的优势，训练过程更稳定，解决了语义分割任务中数据分布不平衡的问题，优化训练过程。

本研究通过使用丰富多样的材料图像作为实验数据，设计全面的对比实验和消融实验，实验验证了本研究提出的两种方法在分割、识别微观结构上具有较好的泛化性能与可靠性，节省了大量时间、人力成本，为材料基因工程中的数据库的构建提供可靠的数据基础。

关键词： 材料图像分割；超像素分割；特征融合；注意力机制；多尺度学习

ABSTRACT

Researching and developing new materials is an important driving force for the development of materials science. Materials genetic engineering is a cutting-edge interdisciplinary discipline in the field of materials science, which can shorten the research and development cycle of new materials. Segmenting and identifying microstructure in materials images can provide data basis for designing materials. Material image processing methods consist of non-learning-based image processing methods and learning-based image processing methods. Non-learning-based material image processing methods have disadvantages such as low efficiency and inaccurate segmentation results; learning-based material image processing includes machine learning and deep learning techniques, but the problems of small samples, complex texture and unbalanced distribution of each phase in material images hinder the application of machine learning and deep learning. It is difficult to find an efficient and widely applicable material image segmentation method because most of the material image processing methods are mainly designed for specific material images. In order to improve the performance of material image segmentation, this paper proposes two efficient and accurate material image segmentation methods by analyzing the characteristics of material images, which combines the current advanced computer image processing technology, machine learning technology and deep learning technology.

Firstly, to solve the problem of insufficient training samples and unbalanced distribution of each phase, this paper proposes a material image segmentation method based on superpixel algorithm and deep learning. This method captures the characteristic in material image that pixels in the same phase have high similarity. Firstly, we use the classic superpixel algorithm(SLIC algorithm), to obtain different numbers of superpixels; then extract the largest inscribed rectangular block of each

superpixel as training samples, which solve the problem of small samples. Next, put these rectangular blocks into the improved DenseNet to extract features and recognize them. After training rounds, we use a sliding window with to get $n*n$ rectangular blocks in image with $n*n$ pixels. The generated model after training rounds is used to predict these rectangular blocks' category, which represents the pixel of middle in each rectangular block. Connecting the pixel of the same category, the segmentation task is completed. Also based on Focal loss, we propose Precision Focal loss, which replaces the confidence in the conditioning factor with machine learning metric—precision. Precision focal loss focuses on hard-to-classify samples, solving the problem of unbalanced data distribution in classification recognition tasks and optimising the training process.

Finally, above method is not applicable for phases with similar textures, because the rectangular blocks don't contain global information, this paper proposes a material image segmentation method based on multi-feature fusion. This method is based on fully convolutional network (FCN). In the down-sampling stage, cascading feature fusion blocks are used to fuse high-level semantic features from low-level semantic features. Then, multi-scale learning blocks are used to deeply mine texture information. In the up-sampling stage, the attention mechanism blocks are used to attach weight to feature maps and optimize resource allocation. Meanwhile, this paper proposes CE-Dice loss based on Dice loss, which combines the advantages of cross-entropy loss and Dice loss for a more stable training process, solveing the problem of unbalanced data distribution in semantic segmentation tasks, and optimizing the training process.

By designing comprehensive comparison experiments and ablation experiments on rich and diverse material image data sets, it verifies that the two methods in this paper have better generalization performance and reliability for segmenting and recognizing microstructure. Thus, these two methods save a lot of time and labor cost, and provide a reliable data base for designing materials database in materials genetic engineering.

Keywords: material image processing, superpixel algorithm, feature fusion, attention mechanism, multi-scale learning

目录

第一章 绪论	1
1.1 课题来源	1
1.2 课题背景概述	1
1.3 课题研究的目的与意义	2
1.4 国内外研究现状	4
1.4.1 图像分割的相关研究	4
1.4.2 材料图像微观结构识别与分割相关研究	6
1.5 论文的主要研究内容	8
第二章 相关理论与技术概述	10
2.1 语义分割网络基础知识	10
2.2 特征融合与多尺度学习	14
2.3 注意力机制	17
2.4 语义分割评价指标	20
2.5 本章小结	21
第三章 基于超像素与深度学习的材料图像分割方法	22
3.1 方法概述	22
3.2 超像素算法	23
3.3 矩形块的分类	25
3.4 改进的 DenseNet	26
3.5 Precision Focal 损失	30
3.6 缩短分割时间的方法	31
3.7 实验与讨论	31
3.7.1 实验数据集介绍	31
3.7.2 参数选择	32

3.7.3 分割结果的对比	33
3.7.4 消融实验	39
3.8 本章小结	50
第四章 基于复杂纹理特征融合的材料图像分割方法	52
4.1 方法概述	52
4.2 全卷积神经网络	53
4.3 VGG 模块	54
4.4 级联的特征融合模块	56
4.5 多尺度学习模块	57
4.6 空间与通道混合注意力机制	58
4.7 CE-Dice 损失	59
4.8 实验与讨论	60
4.8.1 实验数据集介绍	60
4.8.2 参数选择	61
4.8.3 分割结果的对比	61
4.8.4 消融实验	66
4.8.5 本章方法在上一章数据集的结果	73
4.9 本章小结	74
第五章 总结与展望	76
5.1 结论	76
5.2 工作展望	77
参考文献	79
作者在攻读硕士学位期间所参与的项目	90
致谢	91

第一章 绪论

1.1 课题来源

本课题来源科技部国家重点研发计划：“材料基因工程关键技术与支撑平台”专项中“材料基因工程专用数据库平台建设与示范应用”项目中课题“陶瓷涂层材料数据自动采集、整合与应用”中的子课题：“材料复杂显微结构高精度图像处理方法研究”(编号: 2018YFB0704400); 科技部国家重点研发项目“材料基因工程关键技术与支撑平台”专项中“国家新材料数据库平台建设关键技术研究”项目中课题：“材料大数据技术研究”(编号: 2020YFB0704500); 上海市自然科学基金面上项目：“小样本环境下物体自适应识别方法研究”(编号: 20ZR1419000); 之江实验室科研攻关项目(编号: 2021PE0AC02) 资助。

1.2 课题背景概述

随着科技与经济的发展，新材料的研发与设计对各行各业的发展具有重大的意义，技术的革命与经济水平的提升依赖于高性能、高规格的先进材料。然而，受制于当前科技水平与经济环境，从新材料的研发与设计至大规模的批量应用于工业化生产耗时耗力，与工业界追求快速的技术革命的目标相矛盾，因此，如何加快新材料的研发与设计是 21 世纪材料科学领域的一项重大挑战。当前，全球高度重视新材料的研发计划，如美国于 2011 年正式颁布“材料基因组计划”(Materials Genome Initiative) [1]，目的是缩短新材料的研发周期及降低时间人力成本。中国于 2012 年启动《材料科学系统工程发展战略研究—中国版材料基因组计划项目》，标志着我国将材料科学上升到国家重大战略。

材料是科技发展的核心基础之一，是社会文明进步和国民经济的骨干支柱，作为当前材料领域的先进技术，材料基因工程是一门交叉多种研究领域的综合性学科，有机融合了材料高效计算设计、先进实验技术与大数据、人工智能等技技术，掀起了新材料研发模式的变革，对于加快研发进度、降低研发成本、推动材料产业高水

平发展具有不可或缺的意义。材料基因技术三大组成要素包括高通量计算、高通量材料实验方法和材料数据库 [2]，其中对微观结构的分割与识别能够提取有效的微观组织数据，从而为材料数据库构造提供数据基础。基于上述的研究流程，新材料的研发周期、人力成本、时间成本得到进一步优化，具有深远的研究意义。

精确的材料图像分割能够为材料基因工程中的材料数据库构造提供数据基础，例如材料微观结构分割与识别能够测定各相的成分与含量，迅速筛选出有利于目标化合物合成的关键微观结构特征，为快速并行模拟各相成分与性能优化提供驱动力；统计分割出的各相面积、周长等性能数据，为实验方法的改进与创新提供数据支撑；用材料微观结构分割与识别技术提取的各相特征、性能数据。以上都是建立材料数据库的重要组成部分，有利于减少材料重复实验和测试。因此，本研究尝试结合数字图像处理技术、机器学习技术、深度学习技术并应用于材料图像微观结构精确的分割与识别。针对不同的特点的材料图像提出相应的分割方法，为材料基因工程技术创新与发展提供数据基础。

1.3 课题研究的目的是与意义

对材料微观结构的分割与识别技术分为基于非学习的处理技术与基于学习的处理技术。非学习的图像处理技术包括传统的图像处理算法，基于学习的图像处理技术包括基于机器学习的图像处理算法和基于深度学习的图像处理算法。材料图像通常具有复杂的纹理，同种材料的微观结构在不同的环境下差异往往较大，各相的分布也通常不均匀，导致基于非学习的处理技术在材料图像上表现一般。同时，许多材料图像存在噪点，噪点由于制备工艺的不同及采集方式的选用而生成的，阻碍了基于非学习的图像处理技术的应用。另外，不同的材料的微观结构差异较大，而非学习的图像处理技术通常针对特定的一种材料图像进行设计，因此，非学习的图像处理技术的泛化性能及适用性范围具有一定的局限性。机器学习与深度学习的飞速发展材料图像的微观结构分割与识别提供新思路，但材料图像数据集存在特有的难点，一是小样本问题，由于制备设备成本、温度等实验环境因素的限制，生成的样本数量通常只有几十张甚至几张，而机器学习技术与深度学习技术需要大量的训

练样本才能有效训练；二是纹理复杂，在制备工艺流程中，材料微观结构可能发生化学反应或重叠融合，导致各相纹理复杂且边界模糊，给精确的人工标注带来挑战；三是数据分布不平衡，待分割与研究的某些相的区域可能分布较少，导致获取的各相的样本数量不平衡，导致损失函数的梯度无法收敛，影响训练效果。因此，找到一种通用、鲁棒性强的针对几种材料图像微观结构识别与分割的统一方法，准确的识别和分割材料图像具有较高的难度，给先进的计算机技术的应用与发展带来一定的挑战，要求设计与分析材料图像的研究者具备材料科学的基本专业领域知识和丰富的实践水平 [3, 4]。

综合考虑上述因素，基于非学习的处理技术存在很多弊端，一是对于复杂纹理的材料图像适用性一般，二是需要具备一定的专业领域知识导致实验结果容易受主观因素的影响，而由于材料图像存在小样本、纹理复杂和数据分布不平衡的问题，阻碍了基于学习的处理技术在材料图像领域的应用。因此，使用先进的计算机技术解决上述的三个难点具有重要意义，一是能够减少非学习图像处理手段对结果的影响，自动化的处理分析数据，能够大幅度的提升结果的可靠性与准确度；二是随着硬件与算法的发展，先进的计算机技术具有高效的处理速度，节省了大量的人力成本、时间成本，推动了新材料研发进程及材料科学的发展，三是上述的三个难点也是机器学习与深度学习领域尚未完全解决的问题，以上三个问题的解决能够推动机器学习与深度学习领域的发展。

本课题研究分析材料图像的特性，结合计算机视觉技术、机器学习技术及深度学习技术，分别提出两种分割与识别材料图像微观结构的方法，为解决了材料图像分割中的存在的小样本、边界模糊和数据分布不平衡的问题提供新思路，满足了精确分割与识别微观结构的任务要求，推动了材料基因工程技术与深度学习技术的发展。

1.4 国内外研究现状

1.4.1 图像分割的相关研究

图像分割是计算机视觉领域的研究核心之一，具有重要的研究价值，同时也存在挑战性。图像分割就是将待分析的区域识别出来，为进一步的分析提供数据基础。目前各个领域因图像分割技术而取得革命性的发展。例如在无人车自动驾驶中，成功的驾驶任务离不开在不断的环境中感知、识别物体和规划执行策略，高精度的图像分割技术可以实时提供有关道路的信息，从而辅助车辆做出正确的决策；在人脸识别中，准确分割面部的各个部位，能够估计性别、表情、年龄和种族等信息，为后续的面部分析提供数据基础；在遥感图像中，精确分割雷达图像的目标、提取遥感云图的不同云系与背景等，有助于即时有效的天气预报，辅助生产活动决策；在医学领域中，准确分割病灶能够加快诊断速度，以及对未知病毒进行推测分析，推动医学的发展；在智慧农业中，机器人利用图像视觉技术提取分割并识别不同农田区域的图像，提高了农业效率和降低生产成本；在材料领域中，对于图像的微观结构定量分析能够揭示材料的宏观性能、微观结构和制备工艺三者关系，为新材料研发提供数据基础。目前图像分割技术可分为非学习的方法和深度学习的方法。

非学习的图像分割方法可分为基于阈值、边缘、区域、聚类和图论等方法。基于阈值的分割方法就是先获取图像的灰度直方图 [5, 6, 7]，然后设定不同的灰度阈值，灰度值在某一个范围内的像素被认为是同类像素，基于阈值的分割方法适用于像素灰度变化规律明显，前景和背景区分度较高的图像，实现简单且算法效率高，但该类方法仅仅考虑了像素灰度值，忽略高层的语义信息、空间分布、边界信息等特征，鲁棒性较差，不适用于复杂纹理图像，常结合其它分割算法使用。基于边缘的分割方法思路是比较某像素与邻域像素的特征差异度，若相差较大，则认为该像素处于边界区域，找到所有的边界点并连接，则可确定边缘轮廓，从而将图像分为不同的区域，该方法常用于边缘检测，目前的边缘检测算法主要有 Sobel 算子 [8]、LoG 算子 [9]、Canny 算子 [10]、Roberts 算子 [11]，这些算法简单快捷，性能相对优良，常用于图像预处理阶段，对边界不明显的图像适用性一般。基于区域的分割方法考虑

了图像的空间信息，根据像素特征的相似性划分像素点，同类像素构成区域，根据区域进行分割的方法主要分为区域生长法 [12] 和分裂合并法 [13]，区域生长法计算简单，但易受噪声影响，导致区域不完整；分裂合并法对复杂纹理的图像效果较好，但计算复杂，边界可能被破坏。基于聚类的分割方法就是选择几个初始像素点，聚集与初始像素点特征相似的其它像素点，反复迭代直至收敛，所有的像素点都被划分到不同的区域，从而完成分割，以 MeanShift[14, 15, 16] 和超像素分割算法 [17, 18] 为代表，广泛应用于姿势估计、目标跟踪及识别、图像匹配等计算机视觉领域，是经典的图像处理手段。基于图论的分割算法就是将分割问题转换为图的求解，设定目标函数与损失函数，并求出最优解，包括 Graphcut[19]、Grabcut[20]、OneCut[21]，然而该类算法计算量通常较大，适用于类内相似度高的图像。总的来说，非学习的图像分割算法种类丰富多样，但往往适用于特征简单、边界清晰的图像，同时特征的选取与设计依赖人工选择，主观因素较大，无法完成复杂场景的高精度分割的任务要求，因此非学习的图像分割算法通常应用于预处理步骤。

自从 FCN(Fully convolutional network)[22] 问世以来，深度学习在图像分割领域飞速发展，卷积神经网络可以充分的挖掘图像语义信息，针对不同的图像分割场景，一系列的基于深度学习的语义分割模型被提出，使得图像分割的应用领域进一步扩大。FCN 生成的特征图还较为粗糙，SegNet[23] 通过添加更多的短连接 (short cut) 使得结果更加精确，同时 SegNet 在池化层复制特征图的索引而不是在编码器中复制特征图，算法的效率得到提升。DeepLab v1,v2[24, 25] 通过使用空洞卷积在保持图像分辨率的同时增大了感受野，DeepLab v3[26] 则设计了空间金字塔池化，整合了上下文多尺度信息，捕获和挖掘高层次的语义信息，进一步细化分割结果。PSPNet[27] 使用金字塔池层集成上下文信息，并关注了局部特征，以使结果更精确。与此同时，基于深度学习的图像分割算法也推动了一些特定领域的飞速发展。在医学图像中，UNet[28],UNet++[29],Attention-UNet[30],CENet[31] 等医学图像分割模型在多个公开的医学图像数据集取得良好的效果。在自动驾驶领域,ICNet[32],ECANet[33], ND-Net[34],ESPNet[35] 等网络在实时道路检测与行人识别中取得高精度的算法效率。总的来说，基于深度学习的图像分割算法应用广泛，是目前的研究热点。

1.4.2 材料图像微观结构识别与分割相关研究

材料图像通常具有复杂的纹理，微观结构的复杂程度也各不相同，目前针对材料图像微观结构识别与分割方法的相关研究可分为两类。

第一类是基于非学习的图像处理算法，即将传统的图像处理算法应用于材料图像，如阈值法、区域生长法、边缘检测法等。Chatterjee[36] 等人使用阈值法分割材料图像，但阈值方法仅在具有高对比度的材料图像中表现良好；Sun 等人 [37] 改进了阈值方法获得的边界，以获得更完整的边界；Anayev 等人 [38] 和 Lopez 等人 [39] 在材料图像中使用形态滤波方法，在简单的材料图像上表现良好；Chen 等人 [40] 和 Liu 等人 [41] 改进分水岭算法以适应材料图像，但容易产生过分割的问题。以 Canny 算法 [42]、Sobel 算法 [43]、Roberts 算法 [44] 为代表的边缘检测算法在识别材料图像各微观结构区域边界已有应用，Liao 等人 [45] 使用方向链码检测材料图像中的边缘；蒋明星等人 [46] 使用多尺度膨胀算法重建材料图像中的边界；邓仕超等人 [47] 将 canny 算法和灰度轮廓相结合，生成材料图像的闭合边缘。Iacoviello 等人 [48] 使用动态轮廓算法检测球墨铸铁图像中的微观结构；Han 等人 [49] 提出了一种基于 Radon 变换和累计概率霍夫变换的方法来分别检测菊池带边缘的直线和线段，为量化地描述菊池带的信息提供数据基础；Zhao 等人 [50] 通过构建马尔可夫随机场对材料图像分割，分割结果一定程度上有提升；丁贤云等人 [51] 利用人工鱼群分割算法分割材料图像，算法具有较快的收敛速度，但只适用于灰度区分明显的材料图像；朱做涛等人 [52] 通过提取铝合金图像的二值特征，设计了对铝合金腐蚀等级评定方法。王宝珠等人 [53] 结合中值滤波、脊检测技术、形态学重建技术和分水岭分割算法解决了晶粒图像中边界不连续、噪声严重的问题，实现晶粒图像的自动分割；秦湘阁等人 [54] 利用三维可视化技术重建晶粒组织模型，实现识别和定量分析晶粒组织；Tarabalka[55] 等人使用分水岭算法分割与分类材料高光谱图像，但该方法对边缘敏感，易受噪声影响，容易产生过分割的问题；Chen[56] 等人采用聚类算法划分材料图像像素实现图像分割，但算法的复杂度较高，分割时间较长。总的来说，基于非学习的材料图像处理算法对数据量没有要求，算法执行效率较高，但存

在如下缺陷，一是如果图像各区域纹理特征存在差异性，易产生过分割或欠分割的问题；二是具有复杂纹理的材料图像往往边界不连续，导致提取虚假边缘，影响分割结果；三是基于非学习的材料图像处理算法通常是针对特定材料图像或任务而设计，使用范围有限。因此基于非学习的图像处理算法只适用于纹理简单的材料图像或应用于图像处理预处理步骤，对于具有复杂纹理的材料图像表现一般。

第二类是基于学习的图像处理算法，包括基于机器学习和基于深度学习的材料图像分割与识别算法。Azimi 等人 [57] 结合 FCN 与投票法分割低碳钢微观结构，与当时最先进的分割技术相比，将精确度提高了近 46% 以上，为钢铁质量评估提供可靠的依据；DeCost 等人 [58] 提出一种深度卷积神经网络分割模型，能够从超高碳钢数据集分割雪明石颗粒、晶界碳化物、球化颗粒基体、无颗粒晶体等微观结构，为后续的定量分析提供可靠的数据基础；Li 等人 [59] 结合卷积神经网络与局部分析法，成功的检测具有高对比度、亮度和磁化的微观组织缺陷；Dai 等人 [60] 提出一种基于特征工程和机器学习的方法来预测高熵合金相的形成，与非学习的图像处理算法相比，该方法在预测高熵合金的相位形成方面可以达到更高的精度；Yang 等人 [61] 利用支持向量分类 (SVC) 方法从巨大的虚拟样本图像中识别候选的过氧化物结构，然后通过 SVR 回归模型预测候选过氧化物的带隙，成功地从 6529 个虚拟样品中筛选出 60 个潜在有用的过氧化物，为发现新的过氧化物提供新思路；Papa 等人 [62] 利用预训练的神经网络模型识别金属和石墨微观结构，减少金相图像预处理和特征工程的构建工作量；Han 等人 [63] 提出了一种混合算法，该算法结合了高斯滤波法、mean-shift 算法、洪泛填充法、改进高斯差分法和聚类算法，准确自动分割具有复杂纹理的金相图像；Wang 等人 [64] 提出一种自动分割重叠的纳米颗粒的技术，首先使用采用均方误差和种子填充算法来消除噪音并提高原始图像的质量，然后使用聚类算法检测分割重叠区域。何维娜等人 [65] 利用多层感知机对球墨铸铁、可锻铸铁和灰铸铁的金相图像进行了分割提取，该方法可应用于金相图像微观结构的分割和量化；Sheng 等人 [66] 针对热电化合物，提出两种机器学习图像分割策略：应用全连接的神经网络批量分割背散射电子图像中的不同相，同时使用 K-Means 算法在没有人标注的情况下，进一步在热电化合物图像中寻找可能被忽

略的化合物；Lai[67] 等人针对材料图像中一定数量的低分辨率相同材质图像，提出了一种结合扩张算子、灰度共现矩阵 (GLCM)、霍夫变换和聚类的方法，在陶瓷数据集上取得 95.75% 的分割精度；Han 等人 [68] 针对晶体结构识别中的小样本问题，对原子空间随机扰动从而生成新的数据样本，从而解决了样本不足及不平衡问题，使用多层感知机对晶体结构分类，该方法给晶体结构的研究带来新的突破，并首次实现对每个晶体结构类型的概率判断；Liu 等人 [69] 开发了一种方便可靠的基于人工智能的认证策略，可以对具有不同清晰度、亮度、旋转度、放大度的不可克隆的花状图案进行快速认证。但是，机器学习模型需要预定义特征，具有一定的主观性，且机器学习的算法的适用范围有限；深度学习模型有效训练需要大量的训练样本，而材料图像由于制备成本、实验环境等因素的影响，获取的材料图像往往较少；且相的构成复杂、边界模糊，标注成本较高；待分割识别相的分布也往往不平衡。因此上述的因素限制了目前机器学习和深度学习算法在材料图像分割任务的应用。

综上所述，虽然近些年关于材料图像微观结构识别与分割的研究越来越多。但非学习的图像处理算法使用范围有限，只适用于对比度明显的简单材料图像；而材料图像中小样本、纹理复杂、数据分布不平衡等问题限制了基于学习的图像处理算法的应用。因此，如何解决材料图像分割存在的难点，准确识别和分割复杂纹理图像的微观结构，并使得分割与识别算法具有较好的鲁棒性、泛化性能是材料图像领域的急需解决的问题。

1.5 论文的主要研究内容

本论文以作者在攻读硕士研究生阶段的成果为基础，为了解决材料图像分割中小样本、纹理复杂和数据分布不平衡的问题，针对不同特点的材料图像，分别研究提出了两种材料图像微观结构识别与分割方法，通过在多种数据集上进行对比实验和消融实验以验证本研究提出方法的有效性。

本论文各章安排如下：

第二章简要介绍本研究涉及的相关理论基础知识，首先介绍了语义分割网络基本知识，然后介绍本研究使用技术的研究现状，主要是特征融合、多尺度学习模块

和注意力机制，最后介绍语义分割评价指标的计算原理。

第三章紧紧围绕材料图像的特性，解决了材料图像中样本不足和相的分布不平衡的情况下，提出一种基于超像素与深度学习的材料图像分割算法，并进行相关的对比实验和消融实验验证方法的有效性。

第四章针对第三章方法存在的缺陷，提出一种基于复杂纹理特征融合的材料图像分割算法，并进行相关对比实验和消融实验验证方法的有效性。

第五章对全文回顾与展望，回顾了本文的主要研究成果、创新点以及不足之处，并提出对未来研究方向的展望。

第二章 相关理论与技术概述

深度学习技术的飞速发展使得计算机视觉领域不断取得跨越式突破，同时也使得材料图像处理领域迈向新台阶，本研究以现有成熟的深度学习技术为理论支撑，结合材料图像的特点，设计了更有效的材料图像分割方法，解决了材料图像存在的难点。因此，本章回顾与本研究相关技术的理论基础和研究现状，为后续章节的方法和理论提供理论依据。

2.1 语义分割网络基础知识

语义分割就是按照人类理解的语义信息，对整张图像所有像素的类别赋予标签，最终区分不同种类的像素点，即实现了像素级别的分类任务。编码-解码网络是语义分割常使用经典框架，编码阶段主要包含卷积层、池化层，解码阶段主要包含归一化层，反卷积层，激活函数层等。

卷积层 (conv) 是深度学习最基本的操作方式，输入图2.1(a) 所示的矩阵 A ，大小为 $a \times a$ 。图2.1(b) 为过滤器 K ，即卷积核，大小为 $b \times b$ ，过滤器 K 的大小又成为感受野， K 的深度 d 与输入矩阵 A 维度相同，过滤器大小为 K 按照设定的步长在矩阵 A 上移动，每次卷积操作后，矩阵维度都在缩小，相比于矩阵中心点，处于边缘的点在卷积中被计算的次数很少，导致边缘信息丢失，为了弥补该缺陷，在矩阵 A 边缘处补 0，保证输出大小不变，边缘点也可以被计算，这种方式称为“padding”。 K 与矩阵 A 的元素进行点积运算，最终输出矩阵 B (图2.1(c))。

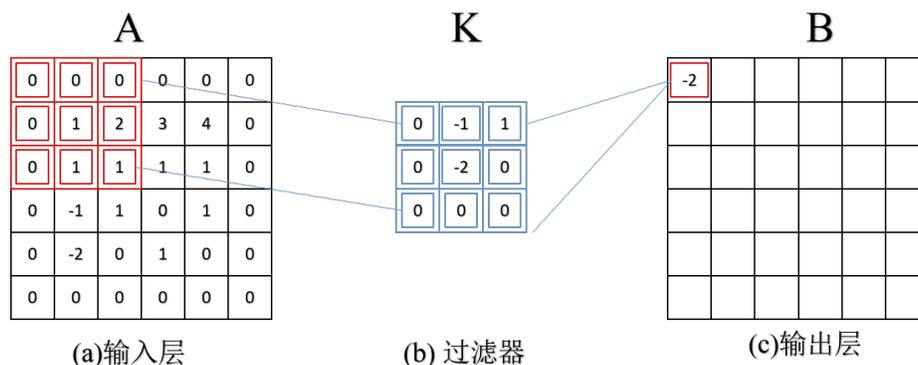


图 2.1: 卷积示意图。

空洞卷积 [24] 是一种常见的卷积方式，向卷积层引入一个称为“扩张率”的新参数，表示卷积核处理数据时，每个核的间距。图2.2显示了扩张率分别为 1、2、4 时的卷积核。没有红点标记的位置填充 0，红色标记位置同正常卷积核。如图2.2(a)所示，对原始特征 f_0 使用扩张率为 1 的空洞卷积生成特征 f_1 ，特征 f_1 相对于 f_0 的感受野为 3×3 ，对整个使用扩张率为 2 的空洞卷积生成特征 f_2 (图2.2(b))，感受野为 7×7 ，即整个深色区域，同理，扩张率为 4 时，生成的特征 f_3 感受野为 15×15 (图2.2(c))。因此空洞卷积具有如下优势：一是扩大了感受野，且不丢失分辨率，对于检测和分割任务中十分有效，大的感受野能够检测分割面积较大的目标，较高的分辨率可以精确定位目标的位置信息。二是提取不同尺度的上下文信息，当设置不同的扩张率时，即感受野不同，获取不同尺度的信息。

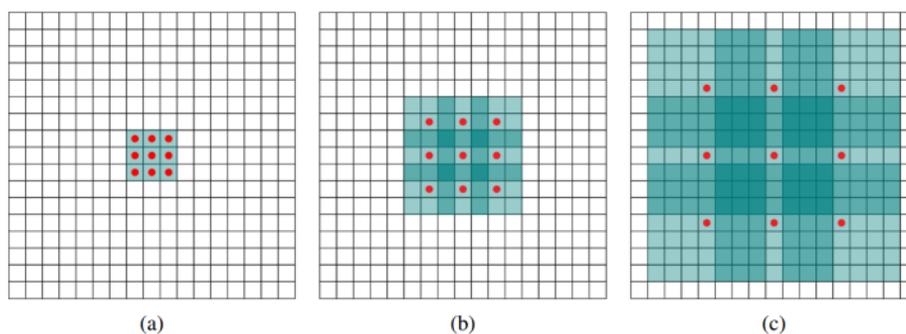


图 2.2: 空洞卷积示意图：(a) 扩张率为 1；(b) 扩张率为 2；(c) 扩张率为 4。

池化层 (pooling) 模仿人的视觉生理现象对输入的对象扩大感受野，抽取高层

语义特征表示图像，池化主要有 3 个作用：(1) 保留关键信息，减少冗余信息，减少计算参数量；(2) 提升模型的尺度不变性、旋转不变性；(3) 避免过拟合。常见的池化操作可分为最大池化、平均池化、全局最大池化、全局平均池化。如图 2.3 所示，最大池化就是保留每个子区域的最大值，其它像素值被舍弃，最大池化的优点在于能保留图像的边缘和纹理信息。平均池化就是求每个子区域的平均值，最大有点在于保留并突出背景信息。全局最大池化则保留整张特征图的最大值，全局平均池化就是求整张特征图的平均值，通常用于替换分类器中的全连接层或密集连接层。

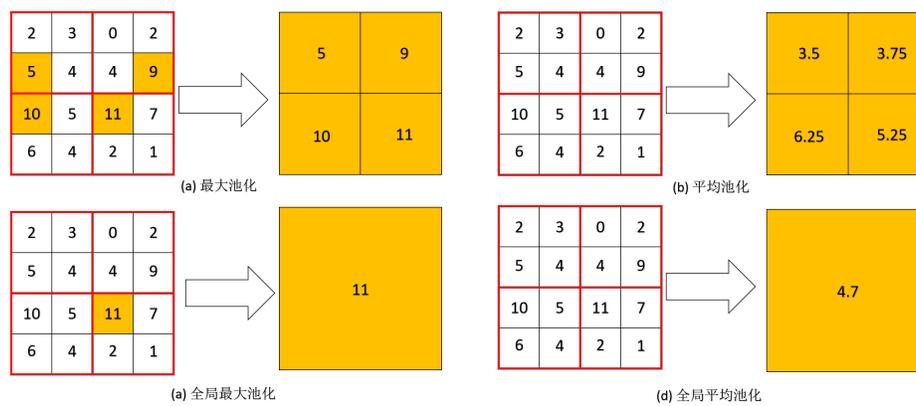


图 2.3: 池化示意图。

神经网络上一层神经元的输出值为下一层神经元的输入值，上层的输出与下层节点输入之间一般呈现函数关系，使用的函数称为激活函数。激活函数使得输出输入不再是线性关系，增强神经网络表达能力，可以几乎逼近任意损失函数。Sigmoid 函数、Tanh 函数和 ReLU 函数是几种经典的激活函数。Sigmoid(图2.4(a)) 函数能够把输入的连续值变换为 0 和 1 之间，因此常用于二分类任务的卷积神经网络。tanh 函数 (图2.4(c)) 解决了 Sigmoid 函数的不是 zero-centered 输出问题。ReLU 函数 (图2.4(b)) 解决了 Sigmoid 和 Tanh 函数的梯度消失问题，且计算速度和收敛速度都快于这两种函数。

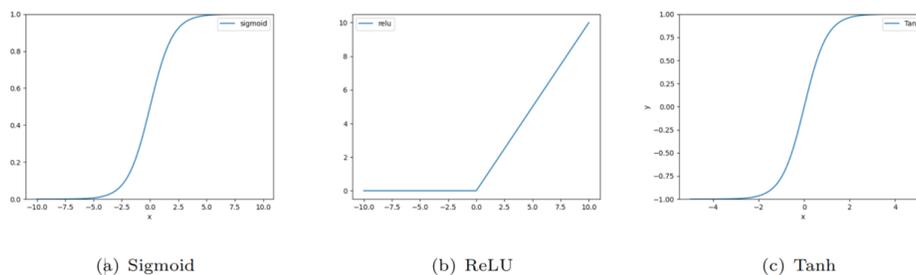


图 2.4: 激活函数。

中间层的输入是前面网络层的输出，因此梯度下降更新参数时，中间每一层的输入分布将会改变，随着神经网络层数的加深，分布改变越明显，该现象称为内部协变量偏移。归一化层统一了每一个神经层的分布，归一化方法可分为批量归一化 (Batch Normalization, BN), 层归一化 (Layer Normalization, LN), 权重归一化 (Weight Normalization, WN), 局部响应归一化。批量归一化将输入参数的每一维都归一到标准正态分布。当神经元的输入分布在神经网络中动态变化，则无法进行批量归一化操作，此时需要使用层归一化对中间层所有神经元归一化。权重归一化实现了对神经网络连接权重归一化，使用再参数化方法，在长度和方向两个维度上，连接权重被分解为两种参数。局部响应归一化常用于图像处理，实现了对邻近的特征图归一化，局部响应归一化常应用于激活函数后，只对邻近的神经元归一化，不减去均值。

在语义分割解码阶段，上采样层对编码阶段的特征图逐步解码，恢复特征图的信息，得到最终的分割结果。上采样方法主要分为插值法 (图2.5(a))、Sub-pixel 卷积 (图2.5(b))、反卷积 (图2.5(c))。最近邻插值法 (nearest-neighbor interpolation) 属于代表性的插值法，该方法将变换后像素的灰度值设置为距它最近的像素的灰度值。

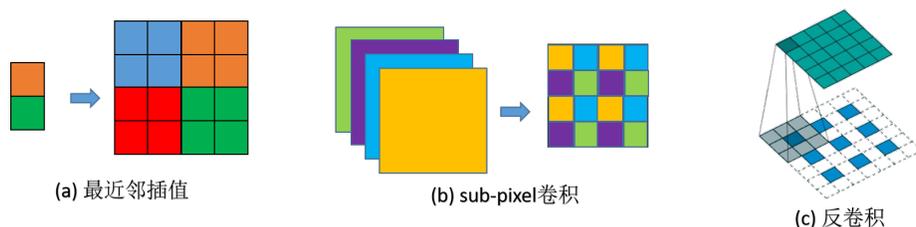


图 2.5: 上采样的方法。

Sub-pixel 卷积 [70] 在不填充 0 的情况下，使用仅有的常规卷积网络进行上采

样,如图2.5(b)所示,假设上采样率为 s , Sub-pixel 卷积先通过常规卷积得到 $s*s$ 张特征图,再将 s 张特征图组合得到宽高都放大了 s 倍的高分辨率的特征图,在 sub-pixel 卷积层后添加卷积层有助于减小计算量,加速学习过程。对于输入的原特征图 x ,步长为 s 的反卷积可分为3个步骤,首先在高 H 和宽 W 的方向上,分别插入 $(s-1)*(H-1)$ 个0和 $(s-1)*(W-1)$ 个0,得到新的特征图 x' ,然后设置新的卷积核大小,最后使用新的卷积核在新的特征图上做常规的卷积,实现了对特征图的上采样。

2.2 特征融合与多尺度学习

按照融合策略的顺序,特征融合可分为早融合与晚融合。早融合就是先融合多层特征,然后网络基于融合的特征做出决策,实现检测与分割任务,称为跳过连接(skip connection),主要是拼接(concat)和叠加(add)操作,图2.6展示了拼接和叠加操作的原理,可以看出拼接操作是合并两张特征图,不进行任何数学运算,增加了通道数,叠加操作是两张特征图相同位置的值相加,生成一张新的特征图,通道数不变。以 UNet, UNet++, UNet+++, ResNet, DenseNet 等网络为代表。晚融合则是在部分融合层上就进行决策,最终综合每个融合层的决策结果,实现检测与分割任务,以 Single Shot MultiBox Detector (SSD)[71], Multi-scale CNN(MS-CNN)[72], Feature Pyramid Network(FPN)[73] 等方法为代表。

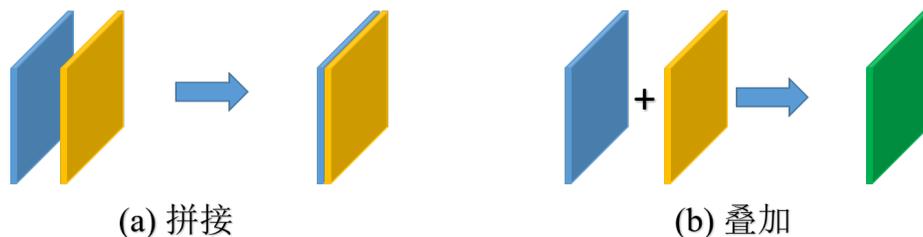


图 2.6: 拼接与叠加操作。

多尺度就是提取同种信号中不同粒度的信息,在不同尺度,可以获取不同的特征,完成不同需求的任务,粒度更小的采样可以捕获更多的细节信息,如高层的语

义特征、空间位置；粒度更大的采样可捕获物体的整体轮廓信息，因此将小粒度采样获取的信息与大粒度的获取的信息融合能够更全面的挖掘特征，优化任务结果。多尺度学习在非深度学习算法和深度学习算法中均有应用。在非深度学习的图像处理领域，多尺度学习可分为两种 [74]：一是构造图像金字塔，即对一组不同分辨率的图像提取特征并融合，增强特征表达能力；二是构造高斯金字塔，以 SIFT 算法 [14]，为代表设置不同的采样因子对图像采样，每个采样层使用不同大小的滤波器，频率从上一层到下一层以因子 2 逐步增加，具有较大频率跨越范围，可以较好的对图像多尺度的描述。在深度学习领域，多尺度学习可分为：(1) 多尺度输入网络，以 MTCNN[75] 为代表，将图像金字塔中的每层的图片输入到网络中，再将每层的结果融合。(2) 多尺度特征融合网络，多尺度特征融合网络分为两种，一种是并行多分支网络，即在不同的感受野下提取特征，大的感受野捕获边界轮廓信息，小的感受野捕获细粒度的高层语义信息，如 Inception[76]，DeepLab[24]，PSPNet[27]，Trident networks[77]。串行的多尺度结构以 FCN、UNet 为代表，通过跳过层实现特征组合。并行的结构能够在同一层级获取不同的感受野特征并融合传递到下一层，串行的结构很好的实现计算量与模型能力之间的平衡，串行结构将不同层级的高级语义特征融合，适用于对边界效果要求较高的语义分割任务。(3) 多尺度特征预测融合网络，以 SSD[71]，FPN[73]，PAN[78] 为代表，即在使用不同尺度的特征输出预测结果，再将结果融合。

Inception 架构的是多尺度学习中常用的网络，早期增强网络鲁棒性的思路就是增加网络深度和宽度，但随着网络深度的增加，参数量也随之提升，一是导致过拟合、计算复杂度高；二是网络越深，梯度容易消失，不利于训练和优化模型，Inception 网络则缓解这两个问题的发生 [79]。如图 Inception V1[79] 基本组成结构有四个成分。 1×1 卷积， 3×3 卷积， 5×5 卷积， 3×3 最大池化，通过多个卷积核提取图像不同尺度的信息，最后进行融合，可以得到图像更好的表征，最后对四个成分运算结果进行通道上组合，Inception V2[80] 两个 3×3 卷积代替 5×5 卷积，可以降低参数量。Inception V3[76] 则将 $n \times n$ 卷积分解为 $1 \times n$ 卷积和 $n \times 1$ 卷积，是降低参数量的同时，通过非对称的卷积拆分，可以处理更多更丰富的空间特征。

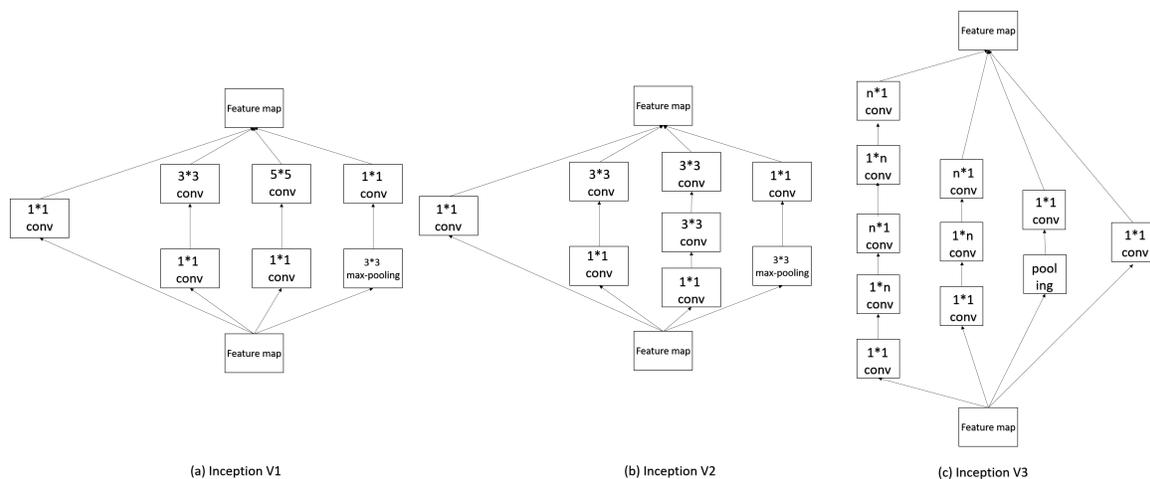


图 2.7: Inception V1、V2、V3 结构。

针对深层的卷积神经网络未充分融入全局信息的缺陷，PSPNet[27] 通过设计一种多尺度学习模块——金字塔池化模块 (图2.8) 有效的提取全局信息。金字塔池化模块包含四同尺度的池化组，大小分别是 1×1 、 2×2 、 3×3 和 6×6 ，对输入的特征图使用四种尺度的池化操作，获得四种不同尺度的子特征图，然后分别使用 1×1 卷积统一通道数，然后，通过双线性插值直接对低维特征图进行上采样，使其与原始特征图尺度相同；最后，将不同层级的特征图拼与原始的特征图拼接，拼接的特征图融合了不同的粒度的特征信息。

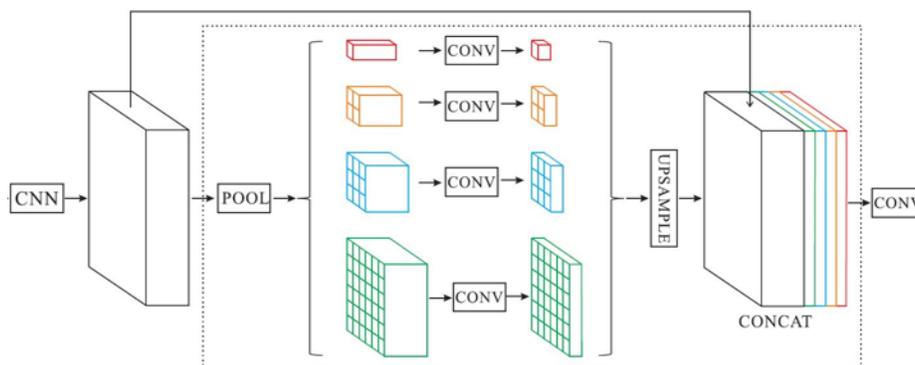


图 2.8: 金字塔池化层 [27]。

2.3 注意力机制

注意力机制来源于认知学的研究，在认知学中，由于人眼处理信息的能力的局限性，人类通常会选择性的关注所有信息的一部分，忽略其它可见的信息，该研究现象即为注意力机制 [81]。注意力机制的核心思想就是处理信息时，决定哪些信息是重要的，按照重要性分配有限的资源。在深度学习中，注意力机制对于神经网络中计算效率的提升、结果的优化起着重要的作用，并在自然语言处理、机器翻译、计算机视觉中有着大量的应用。计算机视觉领域中，注意力机制的核心思想是找到现有数据之间的关联性，并突出重要的特征。按照原理划分 [82]，注意力机制主要分为：硬注意力机制和软注意力机制 [83]。硬注意力机制关注像素级别的关联，通常需要使用强化学习来实现。软注意力机制关注空间域和通道域的依赖关系，主要分为空间注意力机制和通道注意力机制 [84]。

在视觉任务中，空间注意力机制和通道注意力机制是常用的注意力机制技术。特征图中每个区域的贡献度不同，甚至有些区域的特征信息会对干扰最终的结果，空间注意力机制就是寻找图像中最重要的区域并分配较大的权重关注此部分。如图2.9所示，STN(Spatial Transformer) 是一种经典的空间注意力机制 [85]，包含三个模块位置网络模块 (Localisation net)、网格生成模块 (Grid generator) 和上采样器 (Sampler)。输入一张特征图 U ，位置网络对特征图 U 施加 L 操作， L 操作使用公式 (2.1) 对特征图进行仿射变化，包括平移、缩放、旋转等，参数 a 、 b 、 c 、 d 控制缩放和平移，参数 e 和 f 控制平移， (x, y) 表示输入的像素点位置坐标， (x', y') 表示对应输出的位置坐标，最终保留参数 $\theta = \{a, b, c, d, e, f\}$ 。公式 (2.2) 表示，网格生成器则使用 τ 操作使用生成的参数 θ 在空间域上对特征图上每一点 (x_i, y_i) 进行仿射变换；最后，使用上采样器将仿射变换后的特征图还原至输入时的大小，输出特征图 V 。STN 其本质就是对图像的每个区域进行一些变换以获取权重，寻找所有区域中对任务最重要的部分。

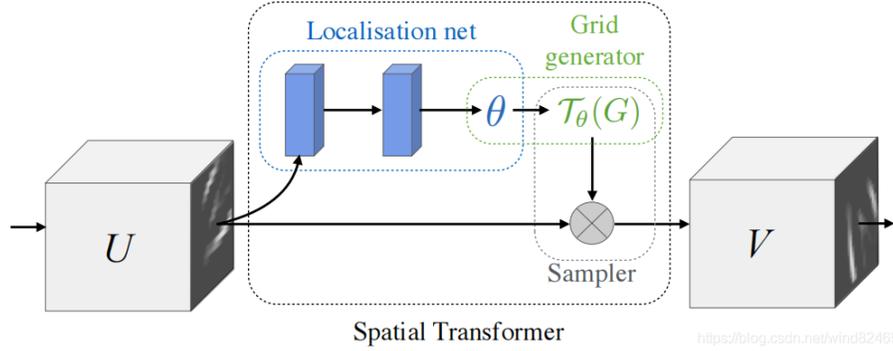


图 2.9: STN 网络结构 [85]。

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix} \quad (2.1)$$

$$\begin{pmatrix} x'_i \\ y'_i \end{pmatrix} = \tau_\theta(G) = A_\theta \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} \quad (2.2)$$

一张 2 维的图像输入到神经网络中，一维是图像的尺度空间，包括图像的长和宽，另一个维度就是通道。卷积神经网络每层都有不同数量的卷积核，一个通道被分配不同的卷积核提取以通道信息，与空间注意力机制相比，通道注意力在于分配各个卷积通道之间的资源，分配粒度上比前者大了一个级别。如图2.10所示，SENet[86]是一种经典的通道注意力机制类网络，对于输入的大小为 $H' \times W' \times C'$ 特征图 X 使用 $F_{tr}(\cdot)$ 操作变换为大小为 $H \times W \times C$ 的特征图 U ， $F_{tr}(\cdot)$ 包含一系列卷积操作，然后使用 $F_{sq}(\cdot)$ 将特征图 U 压缩成 $1 \times 1 \times C$ ，然后使用 $F_{ex}(\cdot, W)$ 每个特征通道生成权重，其中参数 W 被显式地建模特征通道间的相关性；最后，使用 $F_{ex}(\cdot, \cdot)$ 将附加权重后的特征图大小还原至 $H \times W \times C$ 。而后 SKNet[87] 等方法将通道加权的思想和 Inception 中的多分支网络结构进行结合，进一步实现了性能的提升。

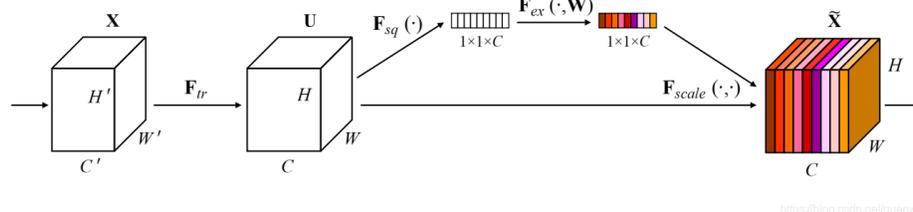


图 2.10: SENet 网络结构 [86]。

空间和通道混合注意力模型，以 CBAM(Convolutional Block Attention Module)[88] 为代表。如图2.11所示，CBAM 包含一个空间注意力模块和一个通道注意力模块。在通道注意力模块中，首先对输入大小为 $H \times W \times C$ 的特征图同时使用最大池化 (MaxPool) 和平均池化 (AvgPool)，生成两个大小为 $1 \times 1 \times C$ 的压缩特征图，再将它们分别送入一个两层的多层感知机 (Shared MLP) 中，最后再将得到的两个特征相加，经过一个 Sigmoid 激活函数得到通道的权重系数 M_c 。在空间注意力模块中，首先对输入大小为 $H \times W \times C$ 的特征图使用最大池化 (MaxPool) 和平均池化 (AvgPool) 得到两个大小为 $H \times W \times 1$ 的特征图，并拼接两个特征图；然后，经过一个 7×7 的卷积层，激活函数为 Sigmoid，最终得到空间域的权重系数 M_s 。通道注意力模块和空间注意力模块可以按照不同的方式使用，即并行或者顺序组合，该研究实验发现顺序组合并且将通道注意力放在前面可以取得更好的效果。总的来说，通道方向的注意力机制捕获了特征的重要性，空间方向的注意力机制捕获了是空间位置的重要性，两种机制的结合可以更全面的对网络捕获的信息处理，优化资源的配置，提升任务效率与结果。

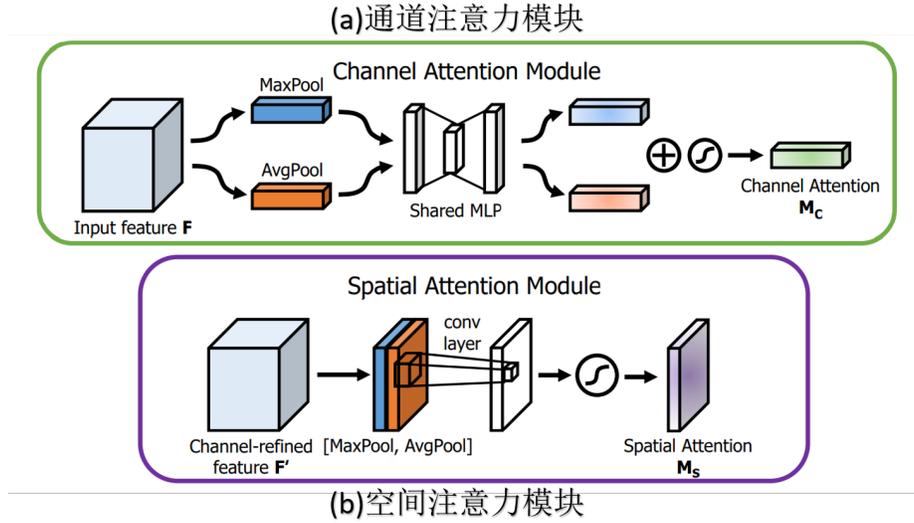


图 2.11: CBAM 网络结构 [88]。

2.4 语义分割评价指标

平均交并比 (mIoU) 和平均 Dice(mDice) 是常用的语义分割结果评估指标, 它能真实的反映预测的分割效果与真实分割结果的差异性, 具体定义如下: 图像中共有 k 类 (包括背景), 真正例 (TP) 表示像素实际类别为正例, 模型预测为正例的个数; 真反例 (TN) 表示像素类别实际为反例, 模型预测为正例的个数; 假反例 (FN) 表示像素实际类别为正例, 模型预测为负例的个数; 假正例 (FP) 表示像素实际类别为负例, 模型预测为负例的个数。第 i 类像素的交并比 (IoU)[89] 的定义为:

$$IoU_i = \frac{TP}{TP + FP + FN} \quad (2.3)$$

平均交并比 (mIoU) 即为每类 IoU 的平均值:

$$mIoU = \frac{\sum_{i=1}^k IoU_i}{k} \quad (2.4)$$

第 i 类像素的 Dice[90] 的定义为:

$$Dice_i = \frac{2TP}{2TP + FP + FN} \quad (2.5)$$

平均 Dice(mDice) 即为每类 Dice 的平均值:

$$mDice = \frac{\sum_{i=1}^k Dice_i}{k} \quad (2.6)$$

2.5 本章小结

本章首先介绍了语义分割网络的基本架构。接着简要回顾了与本论文工作相关的技术: 特征融合和多尺度学习模块、注意力机制模块, 并且详细介绍与本论文相关的模块原理: Inception 架构、金字塔池化模块、空间与通道注意力机制。最后对本论文使用的评价指标进行简要介绍。

第三章 基于超像素与深度学习的材料图像分割方法

由于材料图像通常缺乏足够数量的训练样本，阻碍机器学习技术和深度学习技术在材料图像中的应用。本章抓住材料图像的一个重要特性——同相的像素具有高度的相似性，提出了一种基于超像素和深度学习的材料图像分割方法，为解决材料图像中的小样本问题和数据分布不平衡问题提供新思路。该方法分为三个步骤：SLIC算法 [91] 获得不同数量的超像素；然后提取每个超像素中最大的内接矩形块，将矩形块放入卷积神经网络 (CNN) 中；最后预测输出整张图像每个像素点的类别，可实现对材料图像微观结构识别与分割的任务。本章研究成果已被期刊《Computational Materials Science》录用。

3.1 方法概述

材料基因工程是一门目前材料科学领域的前沿交叉学科，能够缩短新材料的研发周期及降低时间人力成本。而材料基因工程技术三大组成要素包括高通量计算、高通量材料实验方法和材料数据库，其中精确的对材料图像的微观结构分割与识别能够提取关键信息，有助于材料性能的挖掘，为材料数据库提供数据基础。但材料图像中小样本、纹理复杂和数据分布不平衡的问题阻碍了机器学习与深度学习技术的应用。以往的基于机器学习与深度学习的材料图像分割方法有两种思路，人工挑选和机器学习算法根据特征对像素分类 [92]，但挑选和设计的人工特征具有主观性且机器学习模型的泛化能力有限；使用小样本学习技术扩充样本数量，如使用对抗生成网络 (GAN) 和变分自动编码器 (VAE)[93]，但生成的纹理信息与真实纹理信息还是存在差异，影响后续的分割与识别任务。本章分析和研究材料图像的特性——同相的像素具有高度的相似性，因此如果能够学习部分相的像素特征，就能识别和分割整张材料图像。

本章提出了一种基于超像素和深度学习的材料图像分割方法，为解决材料图像中的小样本问题和数据分布不平衡问题提供新思路。图3.1展示了该方法的流程图。具体而言，首先获取矩形块，即采用经典的超像素算法——SLIC 算法 [91] 获得不

同数量的超像素，然后提取每个超像素中最大的内接矩形块。然后识别矩形块，具体是将矩形块放入卷积神经网络（CNN）中，本章选择并改进 DenseNet[94] 作为主干网络对其进行识别，此外，由于材料图像中的相存在分布不均匀与某些相难以区分的问题，选择并改进 Focal 损失以解决数据分布不平衡的问题。最后预测输出整张图像每个像素点的类别，在训练结束后，使用一个步长为 1 大小为 $l \times l$ (l 为奇数) 滑动窗口在大小为 $n \times n$ 整张图像上滑动，为了保证输出每个像素点的类别，将样本的最左侧和最右侧 $((l-1)/2) \times n$ 区域剪切并拼接到左侧和右侧区域，并剪切最顶部和最底部 $((l-1)/2) \times n$ 区域并拼接到顶部和下部区域。在大小为 $n \times n$ 像素的样本上，最终获得 $n \times n$ 个矩形块，模型预测矩形块的类别，表示每个矩形块中间的像素点的类别，连接同类别的像素，即可实现对材料图像微观结构识别与分割的任务。

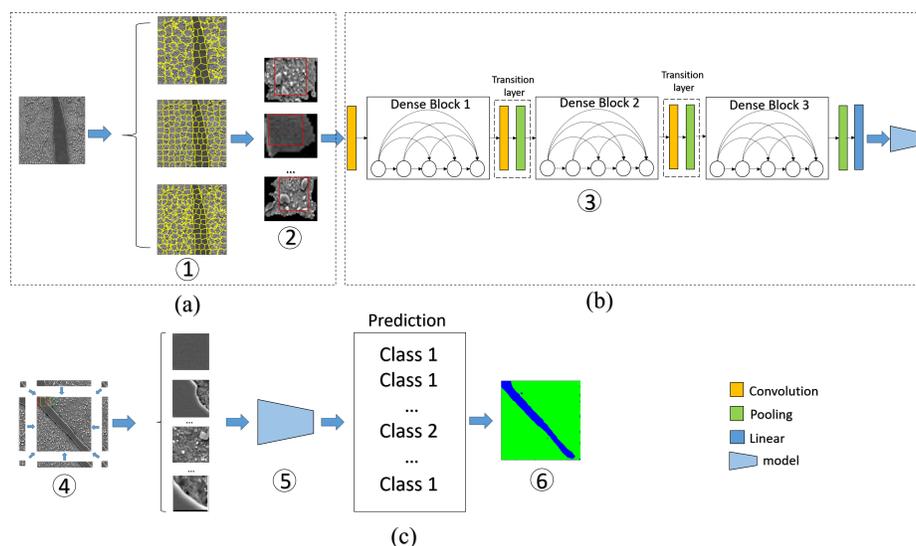


图 3.1: 方法框架。(a) 获取矩形块: ①使用超像素算法获取超像素。②提取每个超像素的最大内接矩形块并标记类别。(b) 识别矩形块: ③将所有最大内接矩形块放入卷积神经网络中提取特征和识别，最终生成训练模型。(c) 分割阶段: ④使用一个步长为 1 的 $l \times l$ 滑动窗口滑动 $n \times n$ 次 (每行滑动 n 次，每列滑动 n 次)。⑤使用生成的模型预测每个矩形块的类别，表示矩形块中心像素点类别。⑥将同一类别的像素连接起来，得到分割结果。

3.2 超像素算法

超像素指在一个邻域内，一组具有相似纹理、颜色、亮度等特征的像素构成的不规则像素块。用少量的超像素替代单个像素点增强特征的表达能力，同时降低算

法的复杂度。在材料图像中，同相的像素及邻域像素具有高度相似性，因此本章提出的方法首先使用 SLIC 算法 [91] 获取矩形块。

SLIC 算法首先将输入的图像每个像素转换为大小为 3 维 Lab 颜色空间和 2 维位置坐标 (x, y) ，因此一个像素由 5 维的特征向量 $[l, a, b, x, y]$ 表示。其中，一个像素在 Lab 彩色空间可表示为 $[l, a, b]$ ， l 表示亮度，值域大小为 0-100， a, b 共同表示色度。相比于 RGB、CMYK 色彩空间，Lab 色彩空间更接近人类生理感知，另外能表示的色域空间更广。SLIC 算法的实现步骤如下：

算法 3.1 SLIC 算法实现步骤

1. 初始化种子点，设定超像素的个数 K ，在有 N 个像素的图片中均匀分配种子点，则超像素的大小为 N/K ，相邻种子点的距离 S 近似为 $\sqrt{N/K}$ ；
2. 在以种子点为中心的 $n * n$ 的邻域范围内重新选择种子点，具体方法为，计算邻域内所有像素点的梯度值，新的种子点为该邻域内梯度最小的像素点；
3. 使用 K-means 算法为每个邻域内的像素赋予类别标签，如图3.2所示，和标准的 K-means 不同的是，SLIC 算法的搜索范围为 $2S * 2S$ ，目的是加快收敛速度；
4. 按公式 (3.1) 求被搜索到像素点 j (由 5 维向量 $[l_j, a_j, b_j, x_j, y_j]$ 表示) 与种子点 i (由 5 维向量 $[l_i, a_i, b_i, x_i, y_i]$ 表示) 的距离 D' ， d_c 代表颜色距离， d_s 代表空间距离， m 为常数，取值范围 $[1, 40]$ 。将在邻域内与种子点距离最近的像素点作为新的聚类中心；

$$\begin{cases} d_c = \sqrt{(l_j - l_i)^2 + (a_j - b_i)^2 + (b_j - b_i)^2} \\ d_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \\ D' = \sqrt{\left(\frac{d_c}{m}\right)^2 + \left(\frac{d_s}{S}\right)^2} \end{cases} \quad (3.1)$$

5. 迭代步骤 1-4，直至聚类中心不再发生变化，算法停止。
-

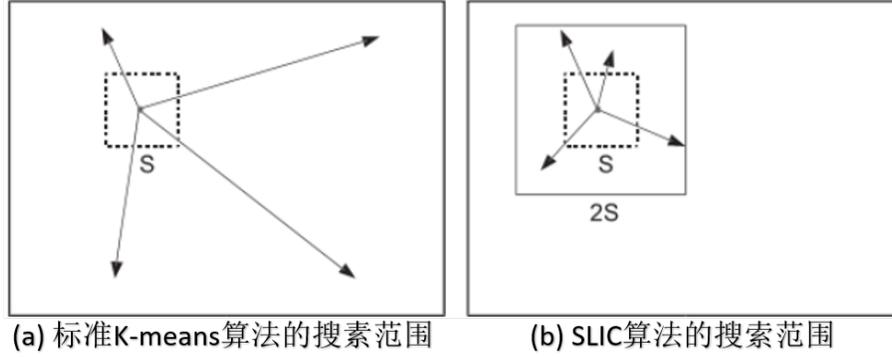


图 3.2: SLIC 的搜索范围与 K-means 算法的搜索范围对比。

算法3.1产生的超像素可能会过小或单个超像素被分解为多个不连续的区域，为了弥补该瑕疵，标记图像中每个超像素为 -1 ，按照 Z 字形走向 (从左到右，从上到下) 扫描每个超像素，若超像素过小或不连续，则与邻近的超像素合并，并将标签设置为与邻近超像素类别，直到遍历完为止。

3.3 矩形块的分类

本章设置超像素的数量分别为 100、150 和 200，然后提取每个超像素区域中最大的内接矩形块，并将其放入改进的 DenseNet 中进行训练，当超像素数为 100 时，改进的 DenseNet 中提取每个超像素中最大内接矩形块 b_i 的特征。最后，将同一类别矩形块的特征连接在一起，形成特征向量集 R_i ，表示相 i 。

$$R_i = D(b_1, b_2, \dots, b_i), \quad (3.2)$$

其中 $D(\cdot)$ 表示 Dense Block。

当超像素的个数设置为 150 和 200 时，将两个尺度下属于同一类别的矩形块的特征表示为 R'_i, R''_i 。将三种类型的特征向量集连接为 R_{f_i} ，

$$R_{f_i} = C(R_i, R'_i, R''_i) \quad (3.3)$$

其中， $C(\cdot)$ 表示连接三种不同尺度的同相特征向量集。使用 *Softmax* 函数识别每

个矩形块，最后生成识别模型 M

$$M = \text{softmax}(R_{f_1}, R_{f_2}, \dots, R_{f_i}). \quad (3.4)$$

训练结束后，使用模型 M 预测每个矩形块 b_j 的类别 u_j ,

$$u_j = M(b_j). \quad (3.5)$$

3.4 改进的 DenseNet

特征是准确识别每个矩形块的关键，为了提高特征提取能力，经典的卷积网络通常思路是加深网络层数或拓宽网络结构，但网络深度达到一定程度时，会产生两个问题，一是误差升高，二是反向传播无法将梯度反馈到前层网络，网络参数无法更新，产生梯度消失和信息丢失的问题，训练效果不佳。已有很多研究提出了解决方法，如 ResNet[95]，其核心思想增加残差映射，跳过本层或多层运算，同时反向传播时，直接将梯度传给上一层。 x 为输入的特征图，经过多次卷积层，输出特征 $F(x)$ ，定义 $H(x) = F(x) + x$ ，公式 (3.6) 表示残差映射，图3.3显示了一个残差映射，ResNet 直接学习 $F(x) = 0$ 在参数量不增加的情况下避免梯度消失问题， $F(\cdot)$ 和 $H(\cdot)$ 表示非线性的转换函数，代表 3 个连续的操作的组合，包括批量归一化 (BN)、ReLU 激活函数、卷积 (conv)，该类学习方式的参数量不会增加。

$$F(x) = H(x) - x \quad (3.6)$$

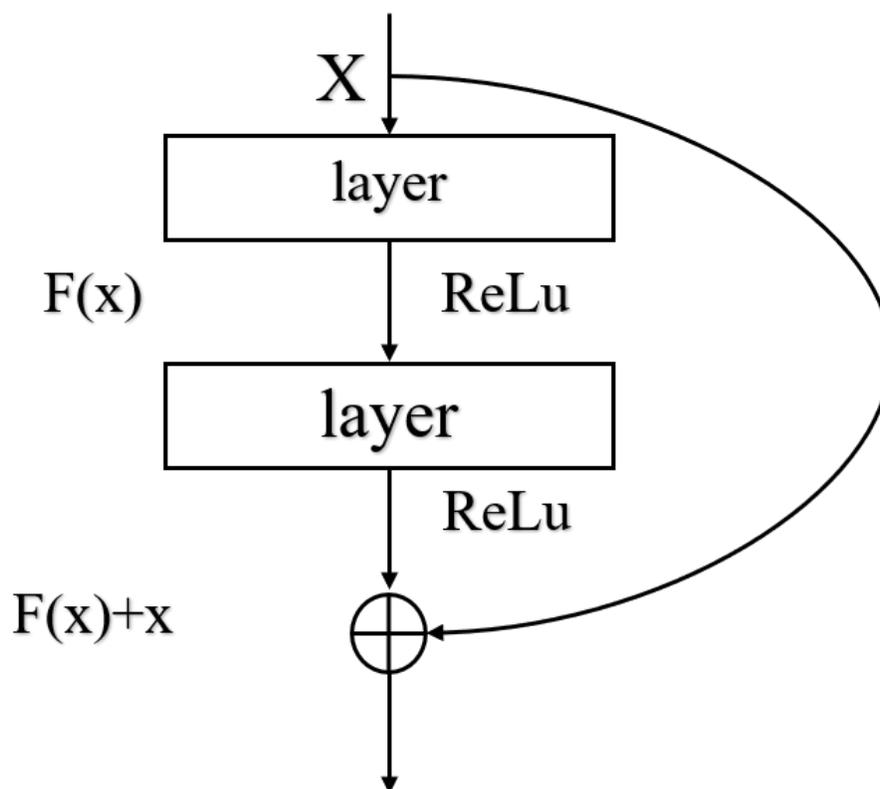


图 3.3: ResNet 的残差结构 [95]。

DenseNet 另辟蹊径，可由公式 (3.7) 表示，在保证每层间的最大信息传输前提下，直接在所有层间建立连接，由于 DenseNet 需要对不同层的特征图进行拼接操作，所以不同层的特征图需要保持相同的大小，使得下采样难以实现，因此 DenseNet 分为多个 Dense Block，DenseNet 可用公式表示为：

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (3.7)$$

其中 x_l 表示网络第 l 层的输出， $[x_0, x_1, x_2, \dots, x_{l-1}]$ 表示将第 $0, 1, 2, \dots, l-1$ 的特征图拼接，

同时 DenseNet 还可以进行优化。同一个 Dense block 内的特征图大小必须保持相同大小，导致无法实现对特征图的下采样，为此不同的 Dense Block 间设置过渡层 (Transition layer) 实现下采样，过渡层包含一层批量归一化层 (BN)， 1×1 卷积和 2×2 平均池化层。在 Dense block 中，每个非线性转换函数 H_l 设置增长

率 (Growth rate) K 控制每个 Dense block 内的特征图数量, 第 i 层的网络输入为 $k_0 + (i - 1) * K$, 其中 k_0 表示网络初始输入的特征图数。Dense block 内不同层的特征图拼接会导致特征图通道数较大, 给网络训练造成负担, 因此使用 1×1 的卷积层, 对特征图通道数降维, 称为瓶颈层 (bottleneck layer), 同时使用使用参数 θ 控制过渡层生成的特征图数量为 $\theta * m$, m 表示一个 Dense block 内的特征图总数, 参数 θ 称为压缩率 (compression factor)。

总之, DenseNet 通过特征复用和设置旁路分支的方式, 在一定程度上缓解了梯度消失的问题, 加强了特征传递和利用特征, 与 ResNet 相比, 训练参数大大减少, 算法效率更高。同时, DenseNet 融合高层与低层的语义特征, 倾向于提供更平滑的决策边界, 因此在训练样本严重不足时 (比如材料图像) 依旧表现良好。因此, 本章选择 DenseNet 作为骨干网络。

纹理信息是精确识别以材料图像为代表的纹理图像的关键, 具有复杂纹理的材料图像含有丰富的语义信息, 网络的主要思路是增加关键层的感受野, 保留关键纹理信息同时减少其它冗余信息对识别矩形块的影响, 以确保决策时不忽略重要信息。池化和空洞卷积是经典的增大感受野方式之一。

本章对 DenseNet 做如下改进, 如图3.4所示, 在特征图添加到下一层之前, 设计了特征增强模块 (Feature Enhance block, FE block), 用于扩大感受野保留关键的纹理信息和边缘信息, 减少其它冗余信息对矩形块识别的干扰。特征增强模块包含三个用于下采样的 3×3 maxpooling 层、一个批量归一化层 (BN), 一个 ReLU 层, 最后使用三个 3×3 的空洞卷积层将特征图恢复至原来大小, 保留全局信息。改进的 DenseNet 第 l 层输出 x_l 可表示为:

$$x_l = H_l([FE(X_0), FE(X_1), FE(X_2), \dots, FE(X_{l-1})]), \quad (3.8)$$

其中 $FE(\cdot)$ 表示在特征图被拼接到下一层之前, 对该特征图施加 FE block。

如图3.5所示, 在过渡层上采样阶段, 本章提出一种基于 Sub-pixel 卷积的上采样方法, 包含一层空洞卷积、一层 sub-pixel 卷积和一个空洞卷积层。使用空洞卷积目的一是保留分辨率的同时扩大感受野, 同时像素的空间位置不变, 二是设置不同的扩展率, 感受野不同, 从而获取多尺度信息, 有利于提取纹理特征, 三是空洞卷

积不需要引入额外参数，进一步降低计算量。反卷积通过在卷积核填充 0 实现上采样，但影响梯度，sub-pixel 卷积则先使用卷积得到 $r * r$ 张特征图，再将 $r * r$ 张特征图组合，实现将特征图放大 r 倍，保持分辨率，减小计算量且具有更强的网络表示能力，且在一层卷积后使用 sub-pixel 卷积可以加快学习过程 [70]。总的来说，改进的 DenseNet 添加的特征增强模块保留关键的纹理信息和边缘信息，本章提出的上采样方法能更好的恢复特征图信息，从而提高了矩形块的识别效果。

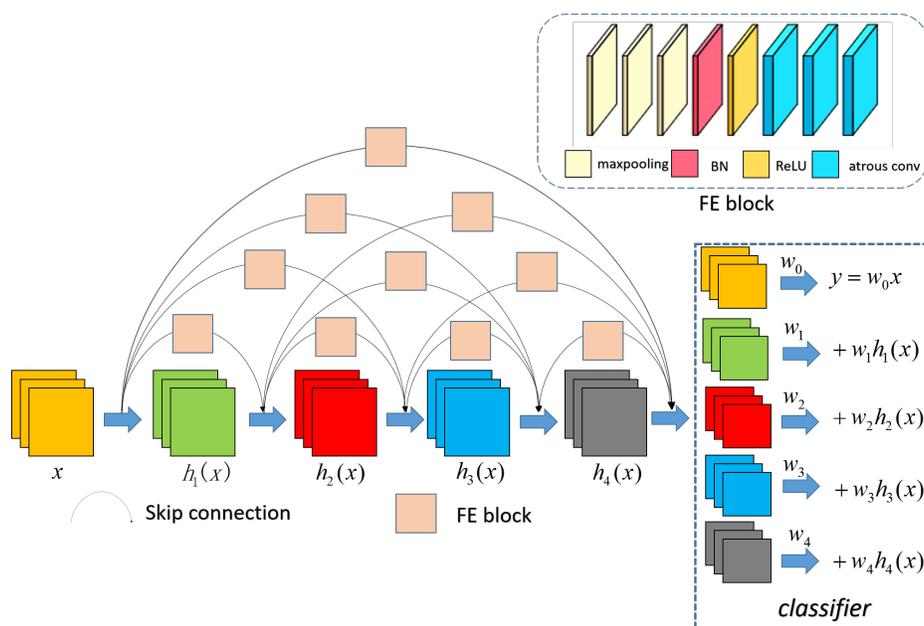


图 3.4: 改进的 DenseNet。DenseNet 可以融合高层和低层的语义特征，从而给出更平滑的决策。在上一层特征图叠加到下一层特征图之前，设计一个 FE Block，目的是保留关键纹理信息，减少其它信息对识别矩形块的干扰。

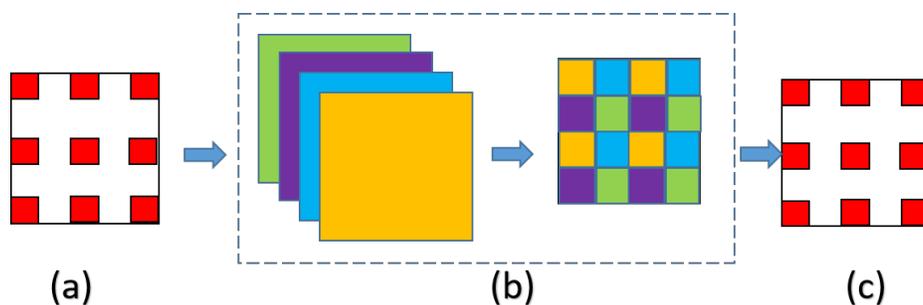


图 3.5: 过渡层的上采样。(a) 空洞卷积层；(b) sub-pixel 卷积层；(c) 空洞卷积层。

3.5 Precision Focal 损失

具有复杂纹理的材料图像数据集包含的训练样本通常较少且每相的分布往往不均匀，即训练样本数量分布不平衡。当样本分布失衡时，样本数量相对较多的类别数据集就会在损失函数中占据主导地位，训练过程中会倾向于样本多的类别，造成模型在少样本类别上的性能较差。

深度学习常用交叉熵损失 (Cross entropy loss, CE) 优化网络，交叉熵的概念源于香农定理，表示对获取信息量的期望，常用于衡量不同概率分布之间的差异 [96]，根据文献 [97]，二类别的交叉熵损失可由公式 (3.9, 3.10) 表示，其中 p 表示样本属于类别 y 的置信度。但交叉熵损失不适用处理类别被不均衡的问题，类较多的样本会主导损失函数的梯度，影响训练过程。

$$p_t = \begin{cases} p & \text{if } y = 1, \\ 1 - p & \text{otherwise,} \end{cases} \quad (3.9)$$

$$CE(p, y) = CE(p_t) = -\log(p_t) \quad (3.10)$$

平衡交叉熵损失首先添加权重因子 α_t 增大含样本数量较少的类别的权重，从而解决了类别不平衡的问题，但当负样本过多，正样本极少时，正负样本类别数量极度不平衡，权重因子失效，负样本占据损失函数的梯度，训练过程同样一般 [97]。如公式 (3.11) 所示，原 Focal 损失 (OriFL)[97] 发现即使置信度 p 很高的样本也会占据交叉熵损失梯度，实际中，置信度 p 很高的负样本通常占据总样本的主要部分，因此原 Focal 损失通过添加权重因子 $(1 - p_t)^\gamma$ 降低了置信度高样本在总损失中的比重，使得网络在学习过程中聚焦难分类样本。

$$OriFL(p_t) = -\alpha_t(1 - p_t)^\gamma \log(p_t), \quad (3.11)$$

本章从总体角度出发，对原 Focal 损失进行改进，称为 Precision Focal 损失 (PreFL)，PreFL 使用常用的机器学习分类评估标准——精度 (precision, 缩写成 pre) 替代置信度 p 。精度定义为预测正确的正例 (TP) 在所有预测正确的样本中的比值，预测正确的样本包括预测正确的正例 (TP) 和负例 (FP)。与原始的 Focal 损失相比，

精度可以更准确地反映各类别样本区分的难度，使得网络在每轮训练中都聚焦于难以区分的样本，从而加快收敛速度。

$$\begin{cases} PreFL(pre) = -\alpha_t(1 - pre)^\gamma \log(pre) \\ \text{其中, } pre = \frac{TP}{TP + FP} \end{cases} \quad (3.12)$$

3.6 缩短分割时间的方法

本章的分割算法是基于像素级分类，容易导致分割时间过长，本章提出了一种缩短分割时间的方法：在材料图像中，一个像素和它的邻域像素几乎是相同的，因此，可以设定滑动窗口的步长为 $n(n > 1)$ ，把第 i 个的矩形块表示为 b_i ，利用公式 2.4 来预测矩形块 b_i 的类别。算法 3.2 显示了我们减少分割时间的解决方案。对行和列的矩形块使用我们的解决方案可以减少需要预测的矩形块数量，从而缩短分割时间。

3.7 实验与讨论

3.7.1 实验数据集介绍

如图 3.6 所示，本文在四种数据集上评估本章方法，包括两种碳钢数据集，陶瓷 (ceramics) 数据集 (图 3.6(c)) 和木材表面 (wood) 数据集 (图 3.6(d))。其中碳钢数据集来源于公开的大型碳钢数据库 UHCSDB[98] (UltraHigh Carbon Steel Micrograph DataBase)，从数据库选择了两种碳钢数据集，分别表示为 carbon steel A(图 3.6(a)) 和 carbon steel B(图 3.6(b))。四种数据集训练集均为 2 张，carbon steel A 的测试集为 4 张，其余的数据集测试集均为 3 张。

算法 3.2 缩短分割时间的方法

```

1. Save step  $n$  and assign to variable  $k$ : set  $k = n$ ;

2. while Using the Model  $M$  to predict the categories of rectangular block  $b_i, b_{i+n}$  do
    |  $u_i = M(b_i)$ ;
    |  $u_{i+n} = M(b_{i+n})$ ;
end

3. if  $u_i == u_{i+n}$  then
    | set  $u_i, u_{i+1}, u_{i+2}, \dots, u_{i+n}$  as the same as  $u_i$ ;
    | set  $i = i + n$ ;
    | jump to step 2;
end

4. if  $u_i \neq u_{i+n}$  then
    |  $n = \lceil n/2 \rceil$ ;
    | jump to step 2;
end

5. if the categories of  $(b_i, b_{i+1}, \dots, b_{i+n})$  are all determined then
    | reset step  $n = k, i = i + n$ ;
    | jump to step 2.
end

```

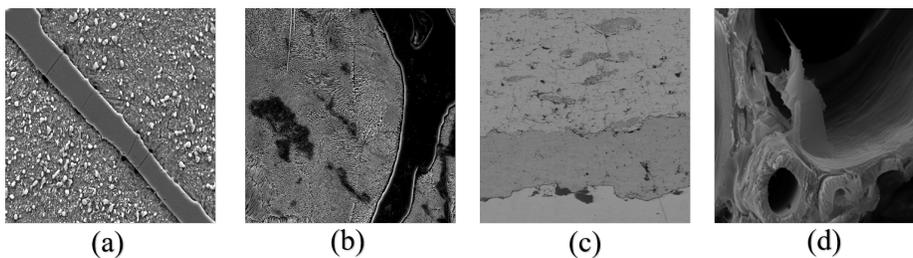


图 3.6: 数据集介绍: (a) Carbon steel A 图像: (b) Carbon steel B 图像: (c) Ceramics 图像: (d) Wood 图像。

3.7.2 参数选择

本方法基于 keras 框架实现,使用 Adam 优化器小批量的训练数据。Dense Block 的数量设置为 5, 增长率设置为 4, 使用瓶颈层, 衰减率设置为 0.01, 丢弃率 (drop out) 设置为 0.01。由于提取的矩形块大小不一, 矩形块输入到网络中训练时, 统一

大小为 32×32 像素。Precision Focal 损失中, α_t 为 0.25, γ 设置为 2。在 carbon steel A 数据集和 ceramics 数据集上, 滑动窗口大小为 11, 在 carbon steel B 数据集和 wood 数据集上的滑动窗口大小为 5。训练集有两张图像样本, 设置超像素的个数为 100, 150 和 200 并将提取的最大内接矩形块旋转 $0^\circ, 90^\circ, 180^\circ, 270^\circ$, 因此, 每张图像样本被分解为 $(100 + 150 + 200) \times 4$ 个矩形块。以 6:2:2 的比例划分训练集、验证集及测试集。

3.7.3 分割结果的对比

本节在四种数据集上评估本章的方法并和其它的方法比较, 其它方法包括 OTSU 算法、分水岭算法、Han 方法 [63]、Lai 方法 [92]、FCN、SegNet、UNet、UNet++、MSDNet[99] 和原始的 DenseNet(ODN)。如图3.7所示, 在训练集上, 首先使用 SLIC 算法 [91] 获取不同数量的超像素块, 结果显示 SLIC 输出的边界与真实的边界几乎一致, 保证提取每个矩形块中像素是同类别的, 不会对训练产生干扰。

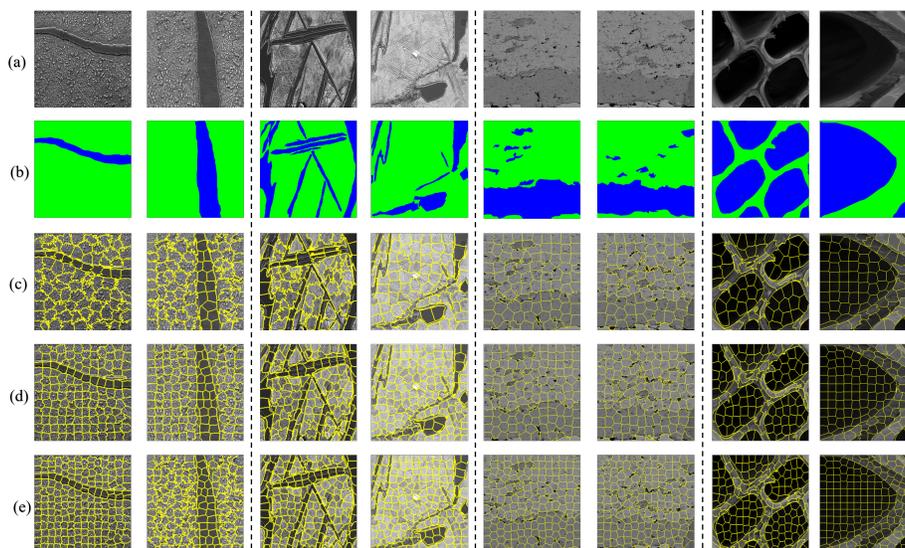


图 3.7: 在测试集上使用 SLIC 算法获取不同数量的超像素。(a) 原图; (b) 真实标注; (c) 超像素数量设置为 100 的结果; (d) 超像素数量设置为 150 的结果; (e) 超像素数量设置为 200 的结果。

图3.8展示了本章的方法和其它方法在 carbon steel A 测试集上的结果, 4 个测试样本命名为 A1、A2、A3、A4, 表3.1展示每种方法在测试集上的 mIoU 和 mDice。

图3.9展示了本章方法和其它方法在 carbon steel B 数据集上的结果，3 个测试样本命名为 B1、B2、B3，表3.2展示每种方法在测试集上的 mIoU 和 mDice。图3.10展示了本章方法和其它方法在 ceramics 测试集上的结果，3 个测试样本命名为 C1、C2、C3，表3.3展示每种方法在测试集上的 mIoU 和 mDice。图3.11展示了本章方法和其它方法在 wood 测试集上的结果，3 个测试样本命名为 D1、D2、D3，表3.4展示每种方法在测试集上的 mIoU 和 mDice。

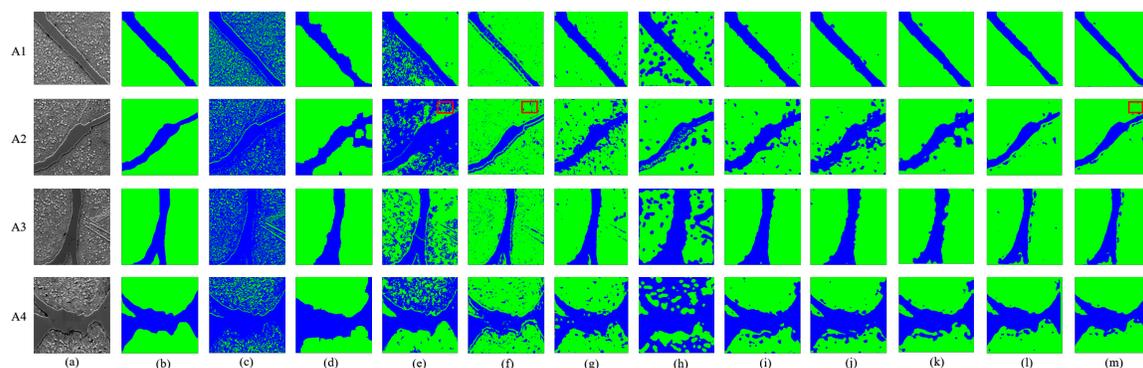


图 3.8: 本章的方法和其它方法在 carbon steel A 测试集上的结果。(a) 原图; (b) 真实标注; (c)OTSU 算法; (d) 分水岭算法; (e)Han 方法; (f)Lai 方法; (g)FCN; (h)SegNet; (i)UNet; (j)UNet++; (k)MSDNet; (l)ODN; (m) 本章的方法。

表 3.1: 本章的方法和其它方法在 carbon steel A 测试集上的 mIoU 和 mDice。

Name	A1	A2	A3	A4	Average
Method					
OTSU	0.285/0.438	0.186/0.312	0.214/0.352	0.326/0.490	0.253/0.398
Watershed	0.822/0.898	0.662/0.781	0.869/0.927	0.755/0.859	0.777/0.866
Han's method	0.525/0.673	0.209/0.344	0.572/0.714	0.550/0.709	0.464/0.610
Lai's method	0.676/0.787	0.678/0.795	0.764/0.862	0.725/0.839	0.711/0.821
FCN	0.856/0.932	0.495/0.704	0.810/0.900	0.790/0.890	0.738/0.857
SegNet	0.625/0.757	0.697/0.808	0.507/0.660	0.413/0.584	0.561/0.702
UNet	0.867/0.895	0.591/0.750	0.842/0.921	0.790/0.885	0.773/0.863
UNet++	0.852/0.884	0.478/0.657	0.810/0.903	0.748/0.862	0.722/0.827
MSDNet	0.858/0.924	0.633/0.775	0.797/0.887	0.811/0.895	0.775/0.870
ODN	0.873/0.930	0.696/0.804	0.796/0.880	0.807/0.891	0.793/0.876
本章的方法	0.849/0.903	0.724/0.828	0.819/0.891	0.845/0.906	0.809/0.882

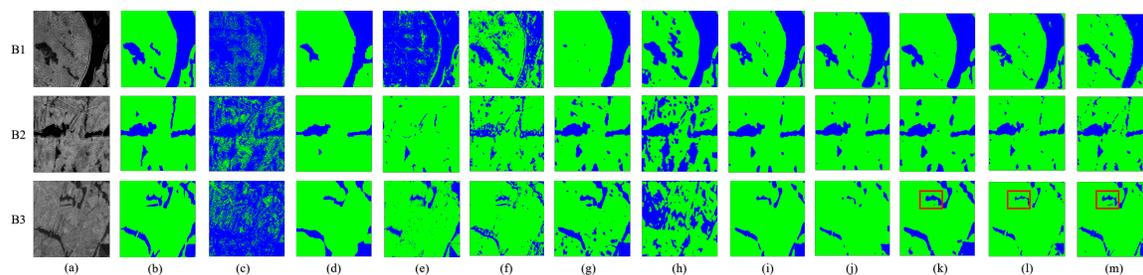


图 3.9: 本章的方法和其它方法在 carbon steel B 测试集上的结果。(a) 原图; (b) 真实标注; (c)OTSU 算法; (d) 分水岭算法; (e)Han 方法; (f)Lai 方法; (g)FCN; (h)SegNet; (i)UNet; (j)UNet++; (k)MSDNet; (l)ODN; (m) 本章的方法。

表 3.2: 本章的方法和其它方法在 carbon steel B 测试集上的 mIoU 和 mDice。

Name	B1	B2	B3	Average
Method				
OTSU	0.264/0.413	0.263/0.406	0.206/0.340	0.244/0.386
watershed	0.759/0.907	0.547/0.839	0.523/0.821	0.610/0.856
Han's method	0.216/0.349	0.505/0.579	0.653/0.764	0.458/0.564
Lai's method	0.641/0.777	0.642/0.758	0.658/0.764	0.647/0.766
FCN	0.740/0.843	0.604/0.803	0.639/0.783	0.661/0.810
SegNet	0.801/0.893	0.288/0.676	0.197/0.349	0.429/0.639
UNet	0.887/0.934	0.719/0.849	0.626/0.764	0.744/0.849
UNet++	0.869/0.924	0.750/0.855	0.356/0.516	0.658/0.765
MSDNet	0.855/0.922	0.658/0.794	0.663/0.797	0.725/0.838
ODN	0.838/0.910	0.757/0.849	0.639/0.746	0.745/0.835
本章的方法	0.854/0.920	0.739/0.836	0.715/0.817	0.769/0.858

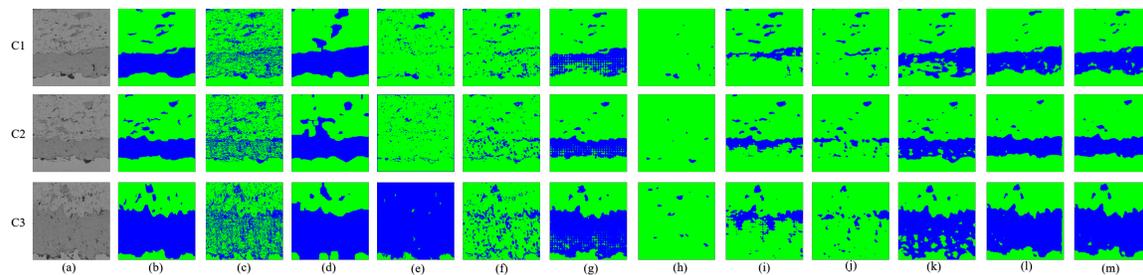


图 3.10: 本章的方法和其它方法在 ceramics 测试集上的结果。(a) 原图; (b) 真实标注; (c)OTSU 算法; (d) 分水岭算法; (e)Han 方法; (f)Lai 方法; (g)FCN; (h)SegNet; (i)UNet; (j)UNet++; (k)MSDNet; (l)ODN; (m) 本章的方法。

表 3.3: 本章的方法和其它方法在 ceramics 测试集上的 mIoU 和 mDice。

Name	C1	C2	C3	Average
Method				
OTSU	0.482/0.628	0.542/0.683	0.433/0.603	0.486/0.638
watershed	0.793/0.885	0.738/0.850	0.865/0.927	0.799/0.887
Han's method	0.351/0.445	0.358/0.438	0.341/0.454	0.350/0.446
Lai's method	0.404/0.515	0.472/0.588	0.330/0.479	0.402/0.527
FCN	0.617/0.763	0.839/0.853	0.882/0.932	0.779/0.849
SegNet	0.338/0.407	0.360/0.421	0.214/0.304	0.304/0.377
UNet	0.484/0.617	0.622/0.749	0.368/0.524	0.491/0.630
UNet++	0.399/0.502	0.480/0.596	0.270/0.393	0.383/0.497
MSDNet	0.723/0.833	0.869/0.928	0.812/0.897	0.801/0.886
ODN	0.763/0.863	0.787/0.878	0.827/0.905	0.792/0.882
本章的方法	0.783/0.876	0.805/0.890	0.858/0.923	0.815/0.896

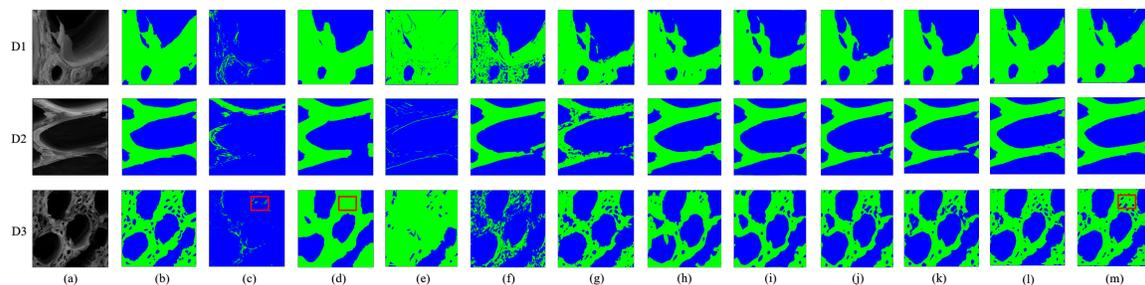


图 3.11: 本章的方法和其他方法在 wood 测试集上的结果。(a) 原图; (b) 真实标注; (c)OTSU 算法; (d) 分水岭算法; (e)Han 方法; (f)Lai 方法; (g)FCN; (h)SegNet; (i)UNet; (j)UNet++; (k)MSDNet; (l)ODN; (m) 本章的方法。

表 3.4: 本章的方法和其他方法在 wood 测试集上的 mIoU 和 mDice。

Name	D1	D2	D3	Average
Method				
OTSU	0.209/0.323	0.467/0.608	0.263/0.371	0.313/0.434
watershed	0.784/0.878	0.839/0.912	0.755/0.860	0.793/0.883
Han's method	0.388/0.500	0.334/0.436	0.362/0.514	0.361/0.483
Lai's method	0.687/0.815	0.847/0.916	0.478/0.634	0.671/0.788
FCN	0.671/0.802	0.717/0.831	0.769/0.869	0.719/0.834
SegNet	0.707/0.828	0.814/0.897	0.698/0.822	0.740/0.849
UNet	0.729/0.842	0.832/0.908	0.716/0.834	0.759/0.861
UNet++	0.712/0.831	0.845/0.915	0.754/0.859	0.770/0.868
MSDNet	0.770/0.869	0.866/0.928	0.752/0.858	0.796/0.885
ODN	0.858/0.923	0.868/0.929	0.728/0.857	0.818/0.903
本章的方法	0.911/0.953	0.875/0.924	0.768/0.869	0.851/0.915

在 4 种数据集上, 本文较为全面比较了本章提出的方法和其它方法, 包括 2 种非学习的图像分割方法 (OTSU 算法和分水岭算法), 2 种基于机器学习的图像分割方法 (Han 方法和 Lai 方法), 4 种基于深度学习的图像分割方法 (FCN、SegNet、UNet 和 UNet++), MSDNet 和原 DenseNet, 本章的方法在 4 种数据集上均取得最好的结果, 在 carbon steel A 测试集上的 mIoU、mDice 平均值为 0.809、0.882, 在 carbon steel B 测试集上的 mIoU、mDice 平均值为 0.769、0.858, ceramics 数据集

上的 mIoU、mDice 平均值为 0.815、0.896，wood 数据集的 mIoU、mDice 平均值为 0.851、0.915。与原 DenseNet 和 MSDNet 相比，本章改进的 DenseNet 结果更优，例如从 carbon steel B 数据集中的图像 B3 可以看出，如红框所示的区域，在 MSDNet 下，出现过分割现象，mIoU 为 0.663，在原 DenseNet 下，出现欠分割现象，mIoU 为 0.639，而使用本章的方法该区域的分割结果几乎与真实标注一致，mIoU 为 0.715，因为本章改进的 DenseNet 通过添加 FE block 保留关键的纹理信息、边界信息和挖掘高层语义信息同时减少其它信息对识别矩形块的干扰，Precision Focal 损失进一步提升了训练效果。与非学习的图像分割方法 (OTSU 算法和分水岭算法) 相比，本章的方法较优，例如从在 wood 数据集中的图像 D3 可以看出，在本章方法下，如红框所示的极小的蓝色区域被正确识别，mIoU 和 mDice 分别为 0.768 和 0.869，而使用 OTSU 算法和分水岭算法对于蓝色极小区域识别失败，mIoU 为 0.263 和 0.755，mDice 为 0.371 和 0.860，因为大多数非学习方法只考虑简单的像素特征，而本章的方法通过增大感受野深度挖掘和提取高层的语义特征，增强了特征表达能力。机器学习方法 (Han 方法和 Lai 方法) 通常针对特定的材料图像设计，使用范围具有一定的局限，例如从 carbon steel A 数据集中的图像 A2 可以看出，本章方法下的噪声较少 (红框所示)，mIoU 和 mDice 分别为 0.724 和 0.828，而使用 Han 方法和 Lai 方法未能消除红框所示的噪声，mIoU 为 0.209 和 0.678，mDice 为 0.344 和 0.795，因为本章的方法由于提取的像素特征信息较为丰富和全面，适用于多种数据集上。与其它经典的深度学习方法 (FCN、SegNet、UNet 和 UNet++) 相比，例如在 ceramics 数据集中的图像 C1 上，本章方法下，真实标注中大部分蓝色区域被正确识别，mIoU 和 mDice 分别为 0.783 和 0.876，比较的方法中，除了 FCN，mIoU 为 0.617，mDice 为 0.763，其余方法几乎未分割识别蓝色区域，因为本章方法紧紧抓住材料图像同相内的像素具有高度相似性的特性，使用超像素算法将整个样本分解成一定数量的超像素块，解决了材料图像中小样本的问题，此外，改进的 DenseNet 进一步融合了每个矩形块中的高层和低层语义信息，增强了特征表达能力，相比之下，经典的深度学习方法受到小样本环境的限制，训练少量样本无法充分学习和挖掘每相的纹理特征。

3.7.4 消融实验

不同损失函数下的分割结果

在材料图像中，相的分布通常是不均匀的，且有些相很难被识别，影响训练效果。Focal 损失可以很好的解决上述的两个问题。同时，本章方法改进 Focal 损失，用常见的机器学习度量——精度来代替调节因子，称之为 Precision Focal 损失，以更好的适应材料图像。图3.12显示了在原始 Focal 损失 (OriFL) 和 Precision Focal 损失 (PreFL) 下，本章提出的改进的 DenseNet 在 4 个数据集上的训练损失。与原始 Focal 损失相比，在 carbon steel A 数据集、carbon steel B 数据集和 ceramics 数据集上，经过 50 轮训练后，Precision Focal 损失值接近 0，而原始的 Focal 损失值接近 0.1-0.15，因此，在这 3 种数据集上，与原始的 Focal 损失相比，Precision Focal 损失加快了收敛速度。在 wood 数据集上，Precision Focal 损失几乎与原始的 Focal 损失同步收敛，但 Precision Focal 损失初始值较低，且整体较为平滑。图3.13显示了在原始的 Focal 损失和 Precision Focal 损失下，改进的 DenseNet 在 4 个数据集上的验证损失。结果表明，在这 4 种数据集上，最终的验证损失都接近于 0，但在 Precision Focal 损失下，初始验证损失值比原 Focal 损失值更小，验证过程更平滑。例如，在 carbon steel A 数据集和 ceramics 数据集上，验证损失值分别在 30 轮左右和前 10 轮显著波动，而在 Precision Focal 损失下，验证损失值在验证轮中保持稳定。因此，在训练过程中，与原 Focal 损失相比，Precision Focal 损失更关注 4 种数据集中难以区分的样本，从而加快了收敛速度，优化了训练效果。

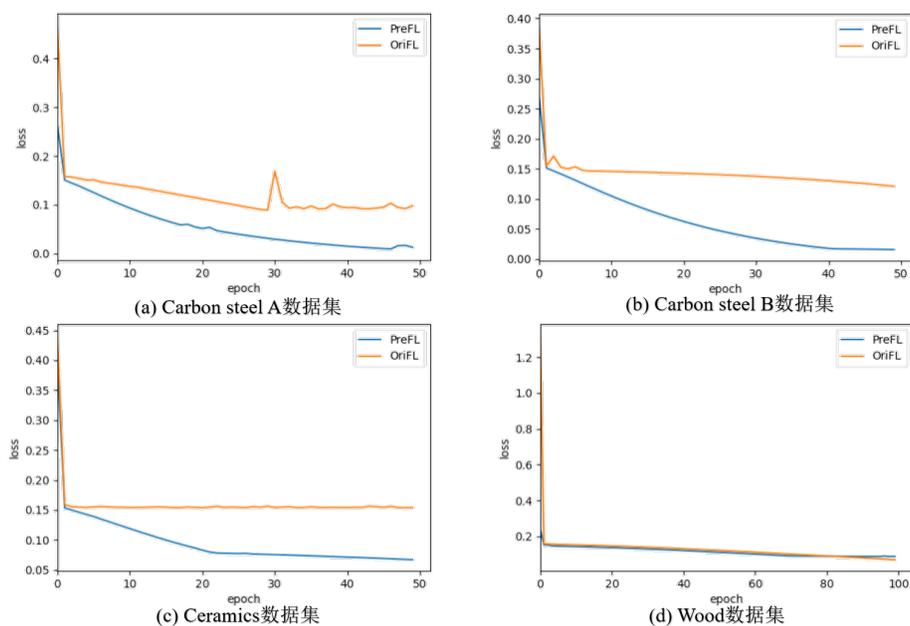


图 3.12: 在 Precision Focal 损失和原始 Focal 损失下, 改进的 DenseNet 在 4 种数据集上的训练损失。

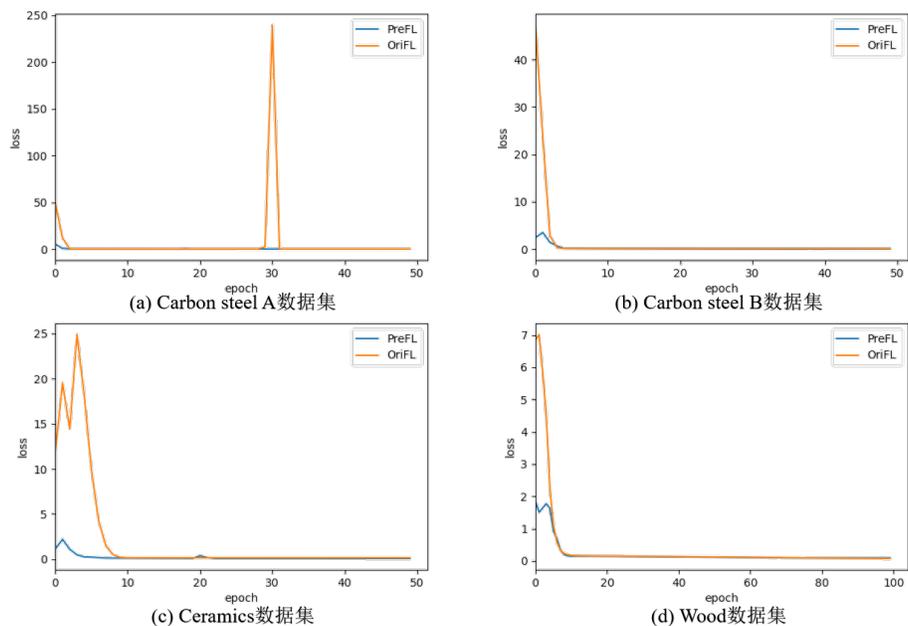


图 3.13: 在 Precision Focal 损失和原始 Focal 损失下, 改进的 DenseNet 在 4 种数据集上的验证损失。

此外, 如图3.14 (c)-(i) 和表3.5所示, 本章还评估了在不同损失函数下, 本章提出的方法在 4 种数据集上的分割结果, 损失函数包括原始的 Focal 损失 (OriFL)、交

叉熵损失、Hinge 损失、Categorical hinge 损失、Huber 损失、Cosine proximity 损失和 Precision Focal 损失 (PreFL)。结果显示，本章方法在 Precision Focal 损失下均取得最优结果。例如，尽管 carbon steel B 数据集中的相分布极不平衡，Precision Focal 损失仍然表现最优，在 Precision Focal 损失和原始的 Focal 损失下，图像 B3 的 mIoU 分别为 0.715 和 0.473，mDice 分别为 0.817 和 0.588，其它损失函数中，效果最好的是 Huber 损失，mIoU 为 0.695,mDice 为 0.799，因此，使用 *precision* 作为调节因子，可以更准确地反映每轮训练过程的情况，并反馈给网络，做出较优的决策，实现优化训练过程的效果。

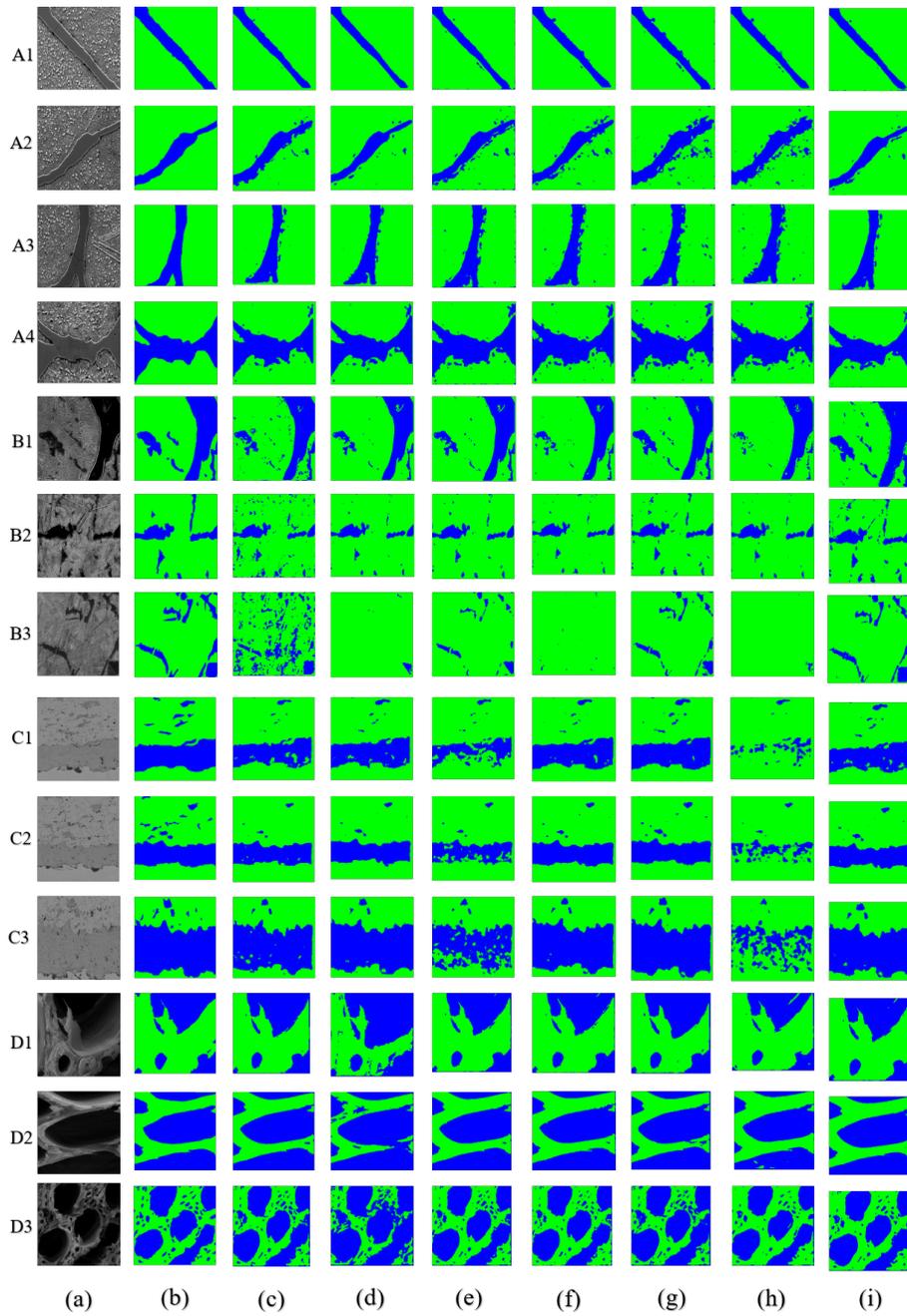


图 3.14: 改进的 DenseNet 在不同的损失函数下的结果。图像 A1-A4 属于 carbon steel A 数据集; 图像 B1-B3 属于 carbon steel B 数据集; 图像 C1-C3 属于 ceramics 数据集; 图像 D1-D3 属于 wood 数据集; (a) 原图; (b) 真实标注; (c) 在原始 Focal 损失下的结果; (d) 在交叉熵损失下的结果; (e) 在 Hinge 损失下的结果; (f) 在 Categorical hinge 损失下的结果; (g) 在 Huber 损失下的结果; (h) 在 Cosine proximity 损失下的结果; (i) 使用 Precision Focal 损失的结果。

表 3.5: 不同损失函数下, 改进的 DenseNet 在 4 种数据集上 mIoU 和 mDice 的平均值。

Loss	Data Set			
	Carbon steel A	Carbon steel B	Ceramics	Wood
原始的 Focal 损失	0.803/0.882	0.664/0.763	0.777/0.873	0.832/0.907
交叉熵损失	0.788/0.874	0.681/0.757	0.809/0.892	0.672/0.800
Hinge 损失	0.807/0.883	0.746/0.835	0.654/0.783	0.829/0.905
Categorical hinge 损失	0.807/0.887	0.658/0.730	0.802/0.888	0.837/0.910
Huber 损失	0.807/ 0.888	0.754/0.858	0.805/0.890	0.790/0.881
Cosine proximity 损失	0.792/0.878	0.655/0.733	0.475/0.613	0.826/0.904
Precision Focal 损失	0.809 /0.882	0.769 / 0.858	0.815 / 0.896	0.851 / 0.915

其它网络在 PreFL 和 OriFL 下的结果

为了进一步验证 Precision Focal 损失 (PreFL) 的性能, 本节在 4 种数据集上评估了其它深度学习网络 (Inception v3[76], MobileNet[100], ResNet, UNet, UNet++) 在原始 Focal 损失 (OriFL) 和 Precision Focal 损失的性能。图3.15和表3.6分别给出了不同网络在两种损失函数下的结果。与原始的 Focal 损失相比, 一些网络在 Precision Focal 损失下的性能显著提高。例如, 在 carbon steel A 数据集上, Inception v3 在 Precision Focal 损失下的平均 mIoU 和 mDice 分别为 0.804 和 0.886, 而原始 Focal 损失的平均 mIoU 和 mDice 分别为 0.736 和 0.858。在 carbon steel B 数据集上, ResNet 在 Precision Focal 损失的平均 mIoU 和 mDice 分别为 0.792 和 0.874, 在原始 Focal 损失下的平均 mIoU 和 mDice 分别为 0.415 和 0.569。在 ceramics 数据集上, UNet 在下 Precision Focal 损失的平均 mIoU 和 mDice 分别为 0.834 和 0.882, 在原始 Focal 损失下的平均 mIoU 和 mDice 分别为 0.500 和 0.474。在 wood 数据集上, MobileNet 在下 Precision Focal 损失的平均 mIoU 和 mDice 分别为 0.828 和 0.905, 在原始 Focal 损失下的平均 mIoU 和 mDice 分别为 0.632 和 0.823。因此, PreFL 可以有效地改善某些网络的性能。通过引入度量——精度 (precision) 来反映样本分类的难度, 使得一些网络更加关注难以分类的样本, 从而优化训练过程。

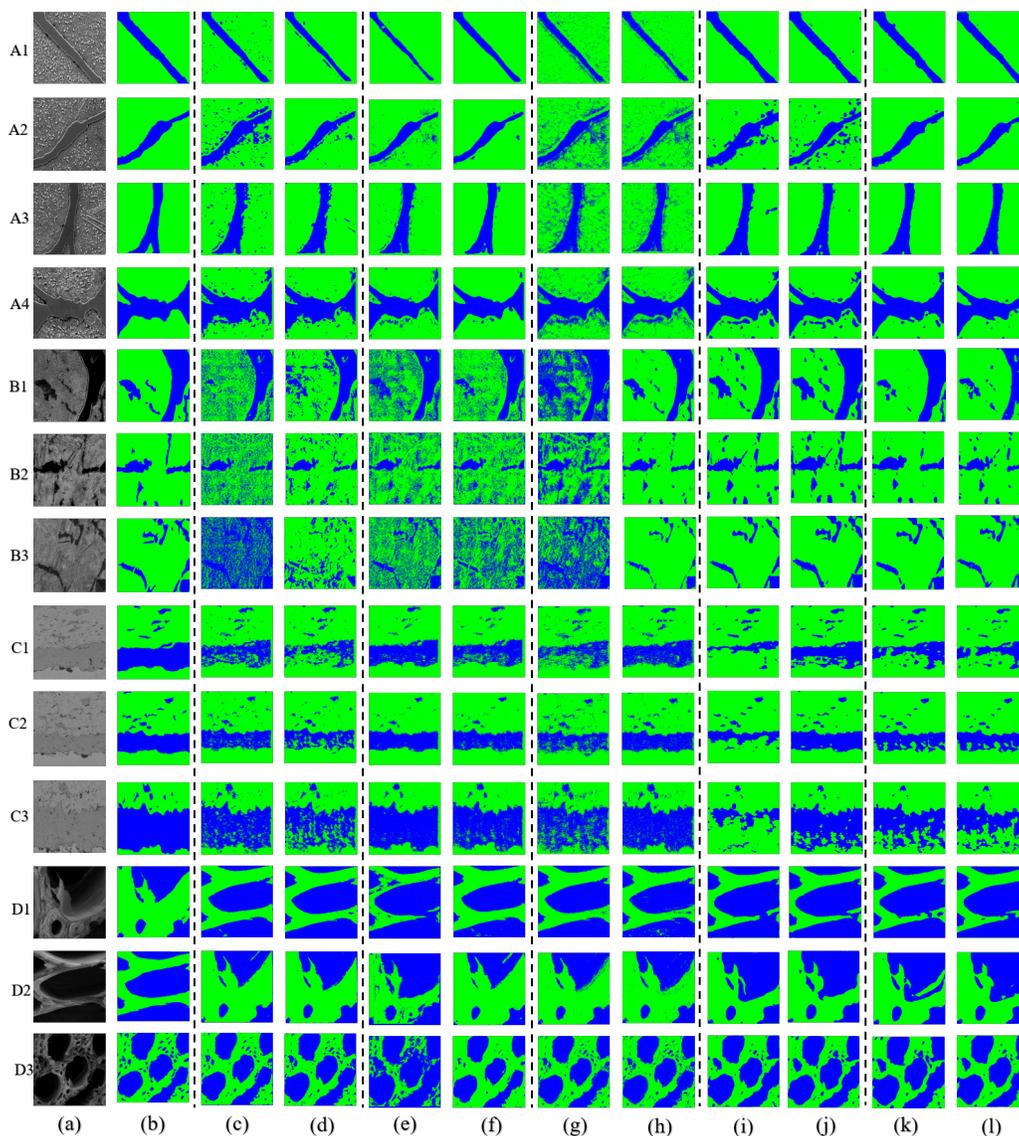


图 3.15: 其它网络使用 PreFL 和 OriFL 在 4 种数据集上的结果。图像 A1-A4 属于 carbon steel A 数据集; 图像 B1-B3 属于 carbon steel B 数据集; 图像 C1-C3 属于 ceramics 数据集; 图像 D1-D3 属于 wood 数据集;(a) 原图;(b) 真实标注;(c)Inception v3 使用原始 Focal 损失;(d)Inception v3 使用 Precision Focal 损失;(e)MobileNet 使用原始 Focal 损失;(f)MobileNet 使用 Precision Focal 损失;(g) ResNet 使用原始 Focal 损失;(h)ResNet 使用 Precision Focal 损失;(i)UNet 使用原始 Focal 损失;(j)UNet 使用 Precision Focal 损失;(k)UNet++ 使用原始 Focal 损失;(l)UNet++ 使用 Precision Focal 损失。

表 3.6: 在 OriFL 和 PreFL 下, 其它网络在 4 种数据集上 mIoU 和 mDice 的平均值

Network	Data set			
	Carbon steel A	Carbon steel B	Ceramics	Wood
Inception v3 + OriFL	0.736/0.858	0.432/0.574	0.710/0.826	0.804/0.906
Inception v3 + PreFL	0.804/0.886	0.587/0.702	0.609/0.747	0.840/0.912
MobileNet + OriFL	0.729/0.831	0.515/0.653	0.757/0.859	0.632/0.823
MobileNet + PreFL	0.795/0.878	0.505/0.632	0.703/0.818	0.828/0.905
ResNet + OriFL	0.695/0.809	0.415/0.569	0.670/0.796	0.869/0.929
ResNet + PreFL	0.739/0.841	0.792/0.874	0.791/0.881	0.848/0.917
UNet + OriFL	0.781/0.873	0.729/0.840	0.500/0.639	0.757/0.860
UNet + PreFL	0.789/0.878	0.734/0.842	0.834/0.908	0.767/0.866
UNet++ + OriFL	0.854/ 0.921	0.768/0.868	0.721/0.832	0.784/0.877
UNet++ + PreFL	0.862/0.888	0.744/0.852	0.685/0.803	0.778/0.868

有无 FE block 的结果

本章的方法中, 在上一层特征图被拼接到下一层的特征图之前, 设计了一个特征增强模块 (FE block) 进一步提取特征。图3.16 (c) 和表3.7显示了在没有 FE block 下, 改进的 DenseNet 在 4 种数据集上的分割结果。和没有有 FE block 的情况相比, carbon steel A 数据集中的图像 A2、carbon steel B 中的图像 B3 的结果更好 (在没有 FE block 的情况下, 图像 A2 和 B3 的 mIoU 和 mDice 分别为 0.661 和 0.777, 0.520 和 0.603, 而在有 FE block 的情况下分别为 0.724 和 0.828, 0.715 和 0.817)。此外, 在一个 Dense block 内没有 FE block 和有 FE block 的情况下, 图 3.17 使用类激活图 (CAM)[101] 将不同层的特征图可视化, 目的验证添加的 FE block 对网络性能的影响。展示的网络层包括 concatenate_1 layer (第 1 次拼接特征图), concatenate_15 layer (第 15 次拼接特征图), concatenate_30 layer (最后一次拼接特征图)。如图 3.16 (f) 和 (g) 所示, 除了在 ceramics 数据集, 使用 FE block 保留了关键的纹理特征, 并且在 carbon steel A 数据集中的矩形块 b_1, b_2, b_3 中, 特征关注在矩形块中心像素点附近 (图4.12(g) 所示)。表3.8显示了在没有 FE block 和有 FE

block 下, 每个矩形块属于真实类别的概率。矩形块 b_1, b_2 属于 0 类, 矩形块 b_3, b_4 属于 1 类。表 3.8 显示了有无 FE block 下, 测试集中矩形块被预测正确的概率及平均值, 结果表明使用 FE block 可以有效地提高矩形块的识别率。特别地, carbon steel A 中的矩形块 b_3, b_4 的纹理信息与 b_1, b_2 较为相似。通过使用 FE block, 所提的网络提取了更多有效的纹理特征 (如图 3.17 (f) 和 (g) 所示), 因此做决策更加全面, 提升了分类准确率。carbon steel B 和 wood 数据集中的矩形块 b_1, b_2 所含的纹理信息较少, 使用 FE block 并没有提升大幅度地提升识别率。在原 DenseNet 中, 上一层的特征图是直接拼接到下一层的特征图。在改进的 DenseNet 中, 首先对上一层的特征图施加卷积核大小为 7×7 卷积操作, 然后使用 FE block 保留关键的纹理信息, 同时减少了其它信息的干扰。FE block 中的最大池化层有如下三个优势: 首先, 它减少了特征图的维度; 其次, 它减少了需要计算的参数量; 此外, 使用最大池化层保留了关键的纹理特征, 舍弃了冗余特征。同时, FE block 中的三层空洞卷积保留了全局信息和分辨率。在 ceramics 数据集上, 并没有在很大程度上提升结果 (使用 FE block 的 mIoU 和 mDice 为 0.815 和 0.896, 不使用 FE block 的 mIoU 和 mDice 为 0.814 和 0.899), 因为 ceramics 的纹理特征相对简单, 没有必要使用 FE block。

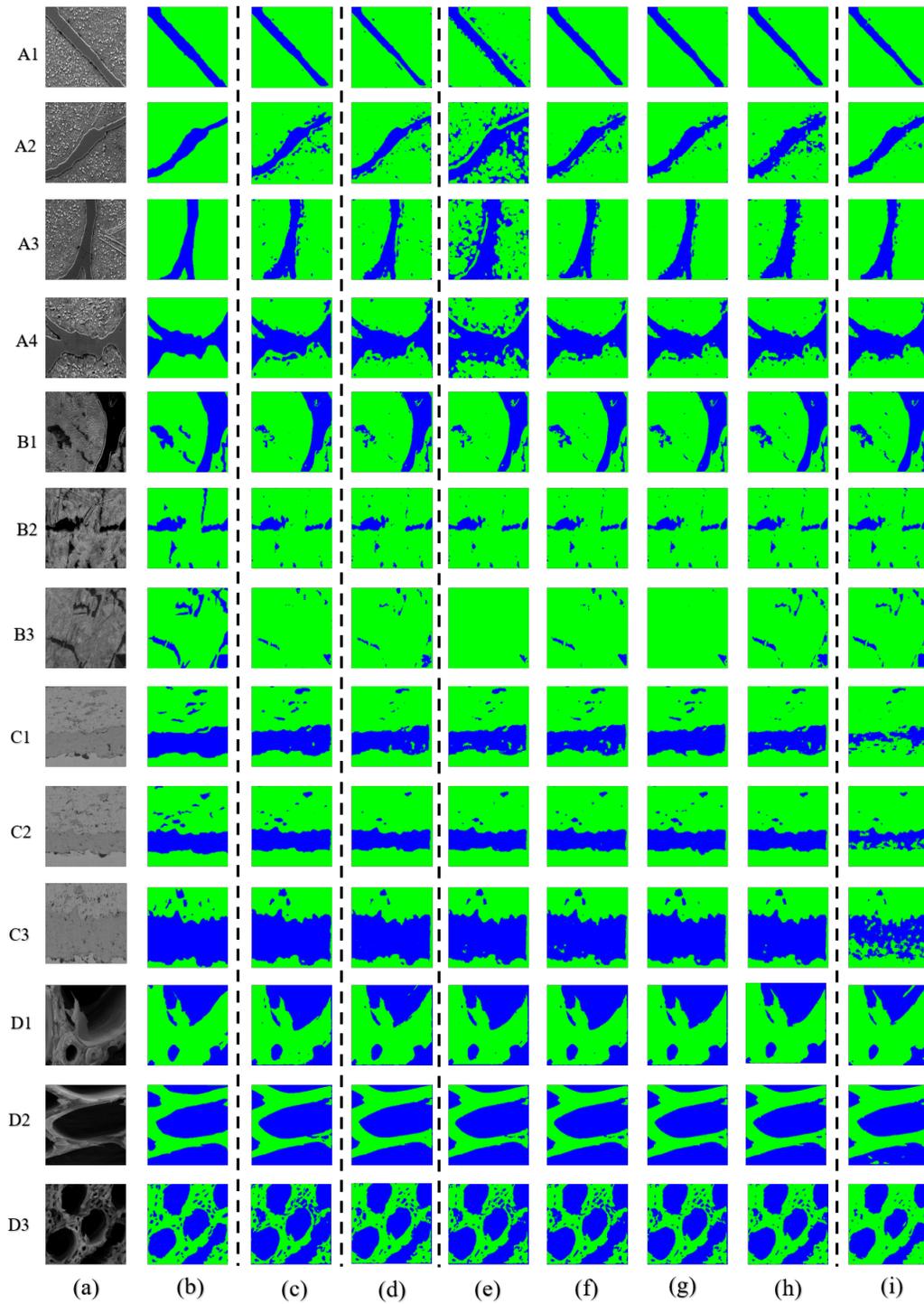


图 3.16: 验证特征增强模块和提出的过渡层上采样方法的有效性。图像 A1-A4 属于 carbon steel A 数据集; 图像 B1-B3 属于 carbon steel B 数据集; 图像 C1-C3 属于 ceramics 数据集; 图像 D1-D3 属于 wood 数据集; (a) 原图; (b) 真实标注; (c) 不使用特征增强模块的结果; (d) 使用最近邻插值的结果; (e) 使用卷积核大小为 3 的 sub-pixel 卷积的结果; (f) 使用卷积核大小为 5 的 sub-pixel 卷积的结果; (g) 使用卷积核大小为 7 的 sub-pixel 卷积的结果; (h) 使用卷积核大小为 9 的 sub-pixel 卷积的结果; (i) 使用反卷积的结果。

表 3.7: 使用 FE block 和不使用 FE block 下, 改进的 DenseNet 在 4 种数据集上的平均 mIoU/mDice。

Data set	FE block	Average
carbon steel A	without FE block	0.777/0.865
	with FE block	0.809/0.882
carbon steel B	without FE block	0.696/0.782
	with FE block	0.769/0.858
ceramics	without FE block	0.814/ 0.899
	with FE block	0.815/0.896
wood	without FE block	0.788/0.879
	with FE block	0.851/0.915

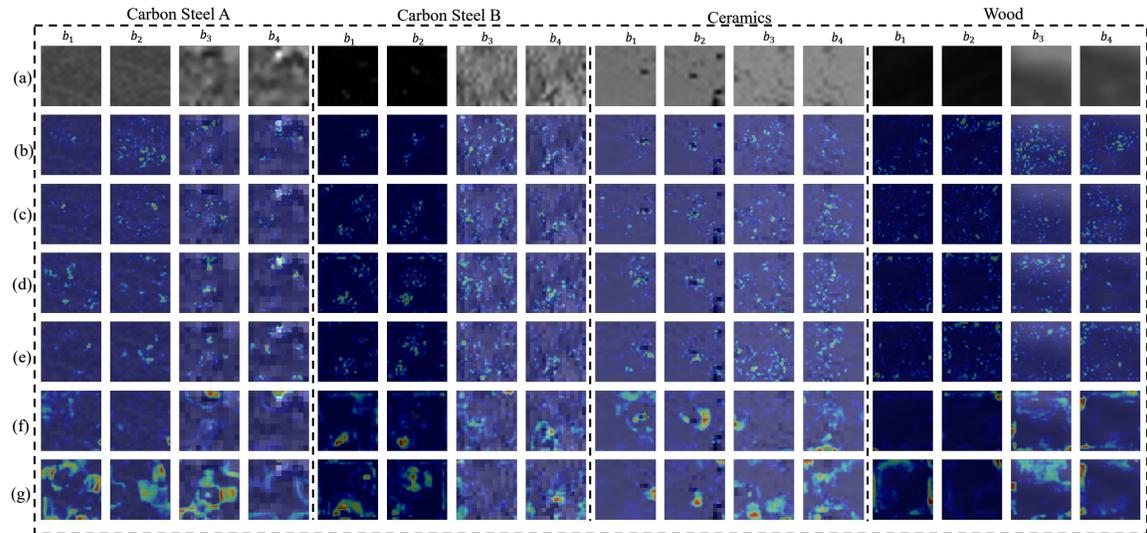


图 3.17: 有/无特征增强模块下的矩形块类激活图: (a) 测试集中的矩形块; (b) 无特征增强模块下矩形块在第一次拼接中的类激活图; (c) 有特征增强模块下矩形块在第一次拼接中的类激活图; (d) 无特征增强模块下矩形块在第 15 次拼接中的类激活图; (e) 有特征增强模块下矩形块在第 15 次拼接中的类激活图; (f) 无特征增强模块下矩形块在第 30 次拼接中的类激活图; (g) 有特征增强模块下矩形块在第 30 次拼接中的类激活图。

表 3.8: 使用 FE block 和不使用 FE block 下, 矩形块被预测正确的概率。

Data set	Block	b1	b2	b3	b4	Average
		Carbon steel A	without FE block	0.9962	0.9940	0.4658
	with FE block	0.9997	0.9997	0.9473	0.6774	0.9060
Carbon steel B	without FE block	0.9868	0.9895	0.9802	0.9769	0.9834
	with FE block	0.9976	0.9978	0.9963	0.9942	0.9965
Ceramics	without FE block	0.9994	0.9989	0.9904	0.9815	0.9926
	with FE block	0.9979	0.9976	0.9799	0.9892	0.9912
Wood	without FE block	0.9236	0.9264	0.9422	0.8402	0.9081
	with FE block	0.9393	0.9471	0.9654	0.9592	0.9528

过渡层的不同的上采样方法

改进的 DenseNet 过渡层的上采样层包括一层空洞卷积, 一层 sub-pixel 卷积 (scaling factor 为 2) 和一层空洞卷积。常用的上采样方法可分为最近邻线性插值 (Nearest-neighbor interpolation)、反卷积 (Deconvolution) 和 [70] 中提出的方法 (称为 sub-pixel conv), [70] 中的方法包括一层卷积、一层 sub-pixel 卷积 (scaling factor 为 2)、一层卷积 (设置卷积核大小分别为 3、5、7 和 9)。图3.16(d)-(i) 和表3.9中显示了最近邻插值法、sub-pixel 卷积和反卷积等 3 种上采样方法和本章的上采样方法下, 4 种数据集的分割结果及 mIoU 和 mDice, 除了反卷积在 carbon steel A 上的结果与本章的上采样方法一样好, 本章的方法在 carbon steel B 数据集、ceramics 数据集和 wood 数据集上取得了最好的结果。和 sub-pixel 卷积相比, 本章提出的上采样方法在 4 种数据集上的表现都良好。在 carbon steel A 和 ceramics 数据集上, 通过改变卷积核大小, sub-pixel 卷积和我们的上采样方法表现一样良好 (最佳的卷积大小为 7)。因此, 不需要为特定的数据集调整卷积核大小, 具有更广的应用范围。其中, carbon steel B 数据集中的图像 B3 和 wood 数据集中的图像 D1 在本章的上采样方法下表现最好 (图像 B3、D1 的 mIoU 和 mDice 分别为 0.715 和 0.817, 0.911 和 0.953), 因为本章的上采样方法可以更好地恢复特征图信息。

表 3.9: 过渡层不同的上采样方法对分割结果的影响。

Upsampling	Data set	Carbon steel A	Carbon steel B	Ceramics	Wood
	Nearest-neighbor interpolation		0.769/0.861	0.722/0.812	0.795/0.884
Sub-pixel conv(filter size=3)		0.649/0.774	0.666/0.741	0.793/0.882	0.785/0.877
Sub-pixel conv(filter size=5)		0.804/0.885	0.704/0.794	0.794/0.883	0.814/0.896
Sub-pixel conv(filter size=7)		0.806/0.886	0.680/0.755	0.813/0.896	0.823/0.901
Sub-pixel conv(filter size=9)		0.781/0.870	0.744/0.835	0.805/0.890	0.834/0.908
Deconvolution		0.808/ 0.889	0.747/0.838	0.614/0.749	0.826/0.904
本章提出的上采样方法		0.809/0.882	0.769/0.858	0.815/0.896	0.851/0.915

3.8 本章小结

在材料图像训练样本不足的前提下，深度学习的无法发挥优势，阻碍了深度学习在材料图像中的应用，且材料图像纹理复杂，非学习的图像分割方法未能进一步挖掘丰富的特征信息，分割效果较差。为此，本章抓住材料图像中同相区域内的像素特征具有高度相似性的重要特点，提出了一种基于超像素与深度学习的材料图像分割方法。方法的总体思路先使用超像素算法获取不同的超像素，提取每个超像素的最大内接矩形块并标注类别；然后在改进的 Focal 损失下，将每个最大内接矩形块放入改进的 DenseNet 网络训练，生成识别模型；最后，使用步长为 1 大小为 $l \times l$ (l 为奇数) 的滑动窗口在大小为 $n \times n$ 的像素的图像上滑动 $n \times n$ 次，获取 $n \times n$ 个矩形块，模型预测每个矩形块类别，该类别即为矩形块中间像素点类别，连接同类别的像素点，完成分割任务。通过在 4 种不同的数据集上的对比实验和消融实验验证：改进的 DenseNet 保留了关键的的纹理信息，减少其它冗余信息对识别矩形块的干扰；改进的 Focal 损失准确的关注难区分样本信息，并反馈给网络，进一步降低了损失值，提升训练效果。本方法具有一定泛化能力和通用性，可以应用于多种材料图像。和现有的分割方法相比，本方法的贡献可总结为以下几点：

- (1) 本分割方法基于材料图像的特性，解决了材料图像中存在的小样本问题；
- (2) 改进的 Focal 损失解决了材料图像中数据分布不平衡的问题；

(3) 方法具有较强的泛化能力和通用性,可应用于多种数据集,训练生成的模型可批量分割同类型数据集其它样本,节省了大量的时间和人力成本。

然而,本方法仍然需要一些不足之处,首先,本方法未能充分提取边界信息,分割结果与真实结果的边界有差异;其次,本方法是通过识别矩形块实现分割任务,而不是整张图片,感知区域较小,不含全局信息,且特征增强模块的连续三次池化操作丢失纹理信息较多,因此不适用相之间纹理相似的材料图像,针对这些问题,未来还需要进一步研究。

第四章 基于复杂纹理特征融合的材料图像分割方法

上一章所述方法的分割结果往往取决于对矩形块的识别能力，然而对于各相纹理较为相似的材料图像，上一章中的方法往往并不适用，分割效果不佳。针对各相纹理相似的材料图像，本章提出一种基于复杂纹理特征融合的材料图像分割方法，选择并改进 FCN。本章的网络包括两个阶段：编码阶段和解码阶段。在编码阶段，首先使用 VGG16 作为骨干网络；其次，将每层的特征图放入级联的特征融合模块 (CFF block)，目的是融合高低层语义信息；然后，将融合的特征图放入多尺度学习模块 (Multi-scale block) 进一步提取纹理特征。在解码阶段，对恢复每层特征图施加注意力机制 (Attention block)，减少特征图的维数，保留关键的特征图。同时，材料图像中数据分布并不均衡，本章采用并改进 Dice 损失以优化分割结果。本章研究成果已投至期刊《计算机应用研究》。

4.1 方法概述

上一章的方法使用超像素提取矩形块，然后识别矩形块获得分割结果，但在相之间纹理相似的材料图像上分割效果一般，因为超像素提取的矩形块较小 (对于输入大小为 512×512 的图片，超像素个数为 100 时，矩形块的平均大小为 5×5)，限制了感知区域范围，未能获取全局信息，且改进的 DenseNet 中的特征增强模块中的池化层会丢失信息，因此无法明显区分纹理信息，对于纹理相似的矩形块识别效果较差。针对各相纹理相似的材料图像，本章的核心思路就是使用先进的语义分割技术深度各相纹理信息，试图学习各相之间的细微差异，从而实现对各相纹理相似的材料图像的精确分割。FCN 是经典的语义分割网络，分为编码阶段和解码阶段，编码阶段借用了经典的卷积神经网络，将最后的全连接层全部替换为 1×1 卷积，生成特征图；解码阶段利用反卷积恢复特征图信息，获取分割结果。然而，原始 FCN 并未考虑全局上下文信息，解码阶段恢复的特征图较粗糙，因此无法满足对于细节信息要求较高的材料图像分割任务。

本章提出一种基于复杂纹理特征融合的材料图像分割方法，选择并改进 FCN

作为基础网络，实现了对各相纹理相似的材料图像精确分割。图4.1展示了网络框架，网络分为编码和解码阶段，使用 VGG16(VGG block) 作为骨干网络，在解码阶段，使用级联的特征融合模块融合了分别来自高层和低层的特征图的语义信息；然后将融合的特征图放入多尺度学习模块 (Multi-scale block) 进一步提取纹理信息；在解码阶段，提出一种空间与通道混合注意力机制模块 (Attention block)，对于每层恢复的特征图对特征图附加权重，通道注意力机制保留重要的特征图，空间注意力机制保留重要特征图中的关键纹理信息。同时，对于材料图像中数据分布不平衡的问题，本章选择并改进 Dice 损失，优化分割结果。

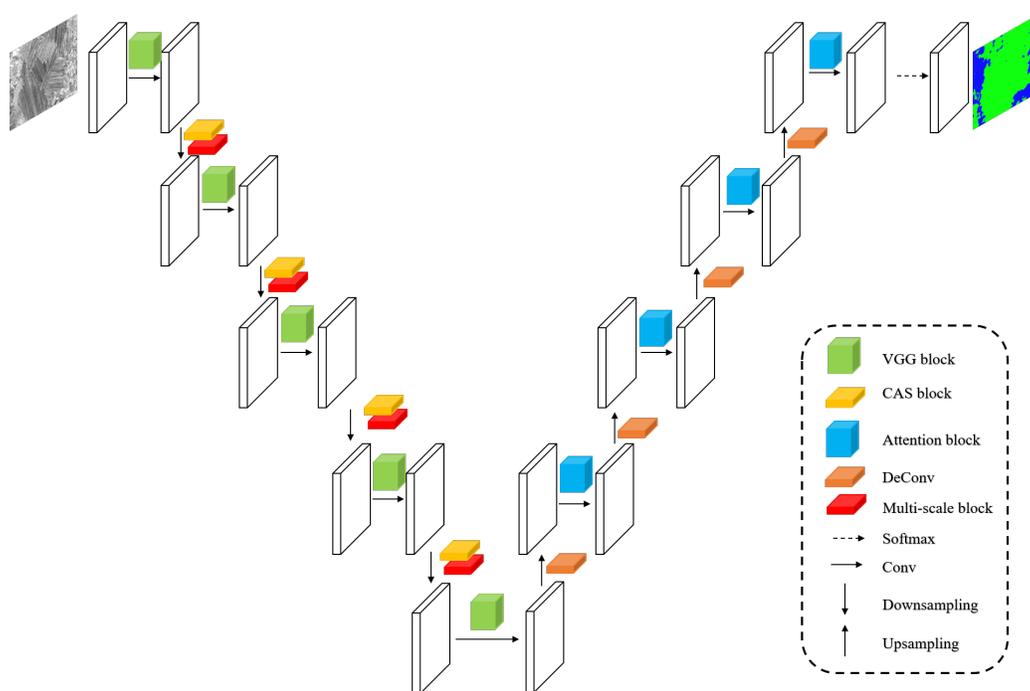


图 4.1: 改进的 FCN 的网络结构。

4.2 全卷积神经网络

如图4.2所示,FCN 是一种经典的全卷积神经网络,在编码阶段本章采用 VGG16 提取特征, VGG16 的多层结构能够自动学习不同层次的特征,浅层的卷积层感知域较小,能够提取局部信息,深层的卷积层能提取抽象的语义信息,但对输入图片大小、位置、方向等细节敏感性较低,不能识别物体具体边界、每个像素的类别,与

卷积层不同的是，全连接层的神经元只连接输入数据的局部区域且共享参数，因而从抽象特征恢复逐个像素的类别，实现像素级别的语义分割。编码阶段中经过多次卷积和池化，图像的分辨率逐步降低，最终生成一张高维的特征图；为了获取原图像尺寸的像素的稠密语义信息，解码阶段使用反卷积逐层恢复特征图信息。

由于最后输出的特征图太小，分辨率为输入图像的 $1/32$ 倍，因此上采样时丢失细节较多，为此 FCN 还引入跳过连接精细分割结果，将最后一层的特征图信息与中间层的特征图信息结合，最后一层的特征图包含较多的全局信息，中间层的特征图包含更多的局部细节信息，因此做最终决策更加全面。例如先将最后输出的特征图进行 2 倍的上采样，大小为输入图像的 $1/16$ ，然后与为输入图片大小 $1/16$ 的中间特征图叠加融合；再将融合的特征图进行 2 倍上采样，大小为输入图像的 $1/8$ ，与为输入图片大小 $1/8$ 的中间特征图叠加融合，此部分网络称为 FCN-8s。

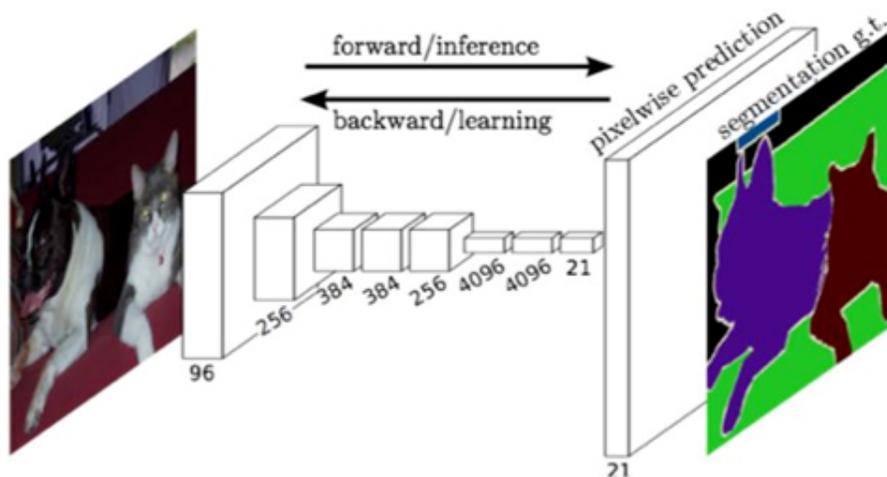


图 4.2: FCN 网络结构 [22]。

4.3 VGG 模块

VGG16 由于规律的设计、简洁可堆叠的卷积块，效果和可移植性较好，广泛使用于卷积神经网络架构 [102]，共包含 13 层卷积层和 3 层全连接层 (FC)。图展示了 VGG16 的网络框架，每个卷积层 (Conv) 结构相同，通过反复堆叠大小为 3×3 的卷积和大小为 2×2 的最大池化层，其中连续的 2 个 3×3 卷积感受野相当于 1

个 5×5 的卷积，连续的 3 个 3×3 卷积感受野相当于 1 个 7×7 的卷积，连续使用小的卷积核替代大的卷积核拥有更少的参数量和非线性变换，有助于深度提取特征。VGG16 还将输入的图片随机放大，然后将大小裁剪为 224×224 ，实现了图像增强，扩充样本数量。

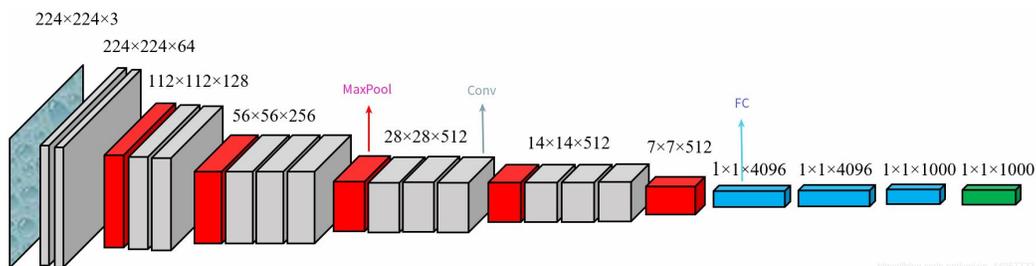


图 4.3: VGG16 的网络结构 [102]。

表 4.1: VGG16 每层的结构

层数	特征维度	过滤器大小	操作类型	个数
Layer 1	(224, 224, 3)	-		
Layer 2	(224, 224, 64)	(3, 3, 3)	卷积	64
Layer 3	(112, 112, 64)	(2, 2)	最大池化	1
Layer 4	(112, 112, 128)	(3, 3, 64)	卷积	128
Layer 5	(56, 56, 128)	(2, 2)	最大池化	1
Layer 6	(56, 56, 256)	(3, 3, 128)	卷积	256
Layer 7	(28, 28, 256)	(2, 2)	最大池化	1
Layer 8	(28, 28, 512)	(3, 3, 256)	卷积	512
Layer 9	(14, 14, 512)	(2, 2)	最大池化	1
Layer 10	(14, 14, 512)	(3, 3, 512)	卷积	512
Layer 11	(7, 7, 512)	(2, 2)	最大池化	1
Layer 12	(1, 1, 4096)	-	全连接	1
Layer 13	(1, 1, 4096)	-	全连接	1
Layer 14	(1, 1, 1000)	-	全连接	1
Layer 15	(1, 1, 1000)	-	softmax	1

4.4 级联的特征融合模块

具有复杂纹理的材料图像包含纹理信息丰富，深度挖掘纹理特征是实现高精度分割任务的关键，融合高层和低层的特征是提升分割性能的一个重要方法。低层特征具有高分辨率，主要包含位置，细节特征，但包含较少的语义信息，噪声信息较多；高层特征则相反，有较强的语义感知能力，但对细节的感知能力较差。因此，将高层特征与低层特征融合能够使得网络深度挖掘复杂的纹理特征，做决策更加全面，从而提高分割效果。

本章方法提出一种基于早融合的特征融合方法，称为级联的特征融合模块 (CFF block)。如图 4.4所示， F_1, F_2 是不同层输出的特征图，分辨率分别为 $C_1 \times H_1 \times W_1$ 和 $C_2 \times H_2 \times W_2$ ，其中 F_2 的分辨率是 F_1 的 2 倍，首先，使用 Sub-pixel 卷积 (Sub-pixel conv) 将 F_1 上采样至与 F_2 同样大小，Sub-pixel 卷积具有减少计算量和保持分辨率的优势，使得上采样时，不丢失纹理信息；其次，使用 3×3 的空洞卷积细化上采样后的特征图，空洞卷积结合了邻域像素特征信息，与常规卷积相比，只需要较小的卷积核就能实现相同的感受野 (常规卷积需要 7×7)，进一步减少了计算量。然后，使用 1×1 的卷积将 F_2 的通道数与 F_1 的通道数统一，使用两个批量归一化层处理两种特征图；最后对处理后的两种特征图叠加，使用批量归一化层处理，最终输出的融合的特征图表示为 F_3 ，大小为 $C_3 \times H_2 \times W_2$ 。通过级联的融合模块生成的特征图融合高层和低层特征图信息，且上采样保持了分辨率，不丢失纹理信息，为进一步提取纹理信息提供基础。

$$F_3 = F(F_1, F_2) \quad (4.1)$$

其中 $F(\cdot)$ 表示特征融合模块。

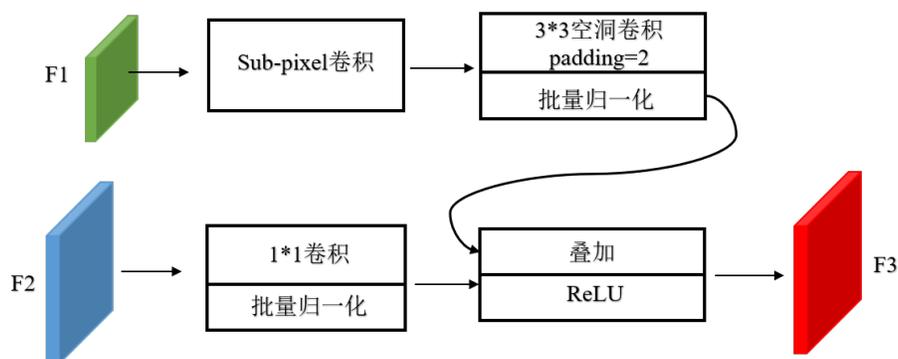


图 4.4: 级联的特征融合模块。

4.5 多尺度学习模块

经过特征融合模块输出的特征图放入两种多尺度学习模块进一步提取高低层语义信息。如图 4.5所示，多尺度学习模块包含两种的多尺度学习方式——密集的空洞卷积模块 (DAC block)[31] 及金字塔池化模块 (PSP block)[27]。如图 4.5(a) 所示，密集的空洞卷积模块基于 Inception V3 架构，包含有三种不同扩张率 (rate) 的空洞卷积，扩张率分别是 1、3 和 5，有 4 个级联的分支，每个分支的感受野分别是 3、7、9 和 19，具有大的感受野的卷积可以提取更多抽象特征，而具有小的感受野的卷积可以提取更多的细节特征，通过结合不同感受野的空洞卷积，密集的空洞卷积模块捕获了不同粒度的信息。接着，将密集的空洞卷积模块输出的大小为 $H \times W$ 的特征图放入金字塔池化模块进一步细化特征图，目的是获取不同层级的上下文语义信息。首先，使用平均池化层 (Average-pooling) 将输入的特征图下采样至原特征图的 $1/16$ ， $1/8$ ， $1/4$ ， $1/2$ ，使用大小 1×1 的卷积将特征图的通道数减少为原来的 $1/4$ ；其次，将 4 种不同的层级的特征图上采样至与输入特征图同样大小；然后，将输出的四种层级的特征图与和输入特征图做合并，即将融合得到的全局特征与原始特征图连接起来，输出特征图 F'_3 。特征合并的过程其实就是融合目标的细节特征 (浅层特征) 和全局特征 (深层特征)，该过程提取了上下文信息。

$$F'_3 = P(DA(F_3)) \quad (4.2)$$

其中 $DA(\cdot)$ 表示密集的空洞卷积模块, $P(\cdot)$ 表示金字塔池化模块。

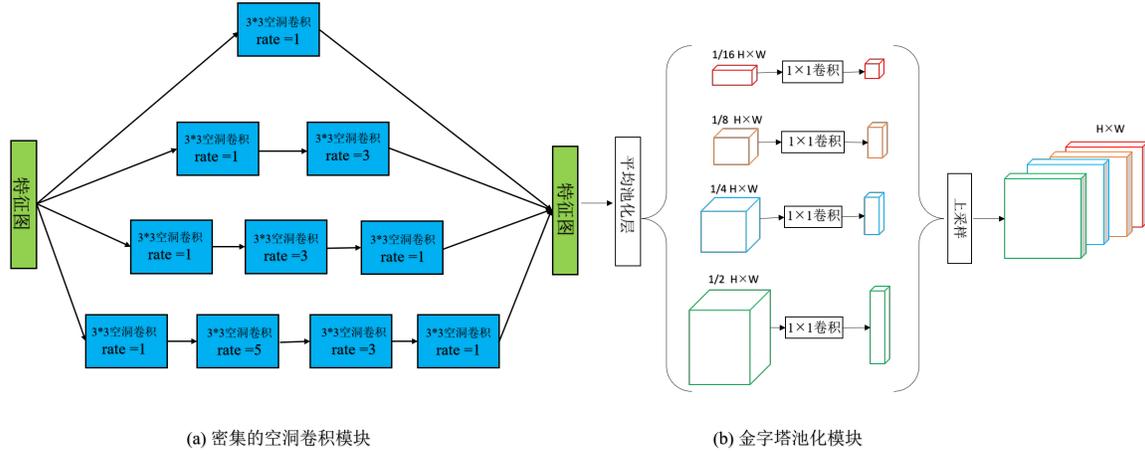


图 4.5: 多尺度学习模块。

4.6 空间与通道混合注意力机制

本章选择空间与通道混合注意力机制, 目的是对特征图附加权重, 保留关键特征图, 减少参数量。如图 4.6 (a) 所示, 首先, 对特征图施加通道注意力机制, F_l 是上一层的特征图, F_g 为下一层的特征图, 大小为 F_l 的 $1/2$, 对 F_l 进行下采样至与 F_g 同样大小; 其次, 对 F_l 和 F_g 分别使用线性变换 $W_l W_g$, 线性变换指使用 $1 \times 1 \times 1$ 卷积统一两种特征图的通道数; 然后, 然后拼接特征图 F_l 和 F_g , 经过 leaky ReLU 操作、线性变化 ψ 、Sigmoid 操作和重采样 (Resampler) 生成注意力系数 α 。最后, F'_l 表示对 F_l 施加注意力系数后的特征图,

$$F'_l = \alpha F_l \quad (4.3)$$

通过通道注意力机制, 每张特征图都获得了相应的权重, 接着使用空间注意力机制, 获取每张特征图的重要区域。如图 4.6 所示, 首先, 由通道注意力机制输出的特征图 F'_l 经过一系列的卷积操作得到特征图 \hat{F}'_l ; 然后, 由 $softmax$ 函数在空间值上生成像素级注意力图 $\mathcal{L} \in R^{H \times W}$, H 表示特征图 \hat{F}'_l 高度, W 表示特征图宽度, \hat{F}'_{l_i} 表示特征图 \hat{F}'_l 中每个像素点 i 的值:

$$\mathcal{L}_i = e^{\hat{F}'_{l_i}}, i = [1, 2, 3, \dots, H * W], \quad (4.4)$$

将注意力图 \mathcal{L} 与输入的特征图 F'_l 进行元素级乘积运算，得到最终的特征图 F''_l ：

$$F''_l = \mathcal{L} \odot F'_l \quad (4.5)$$

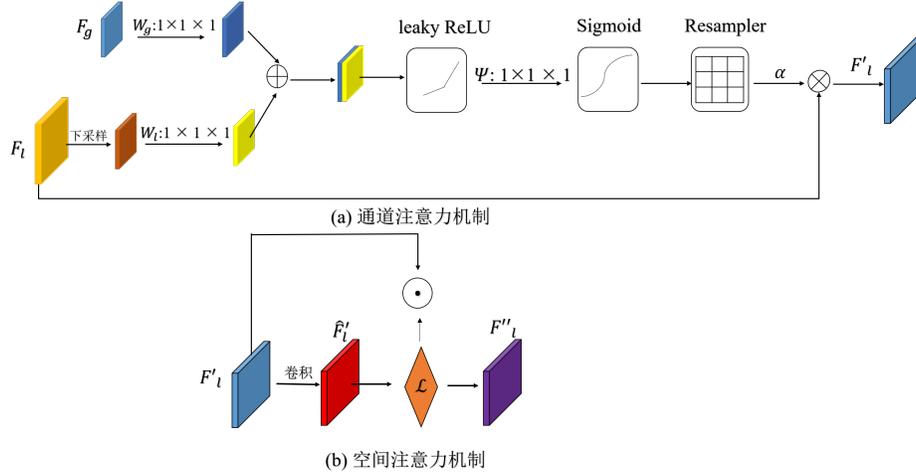


图 4.6: 注意力机制模块。

4.7 CE-Dice 损失

当网络训练结束时，softmax 函数被用来预测图像中每个像素属于各类别的概率。然而，在材料图像中，每相的分布并不均匀，关注的相只占图像非常小的部分，导致网络在学习的过程中，损失函数陷入局部最小值，只学习到分布区域较大的相的特征，影响分割结果。上一章的方法中提出的 Precision Focal 损失通过使用精度反映样本分类效果，解决了数据分布不平衡问题，然而精度常用于评估分类任务效果，本章的网络模型是用于语义分割任务，因此本章从分割任务角度解决材料图像数据不平衡的问题。Dice 系数是语义分割任务的常用评估指标，而 Dice 损失通过引入 Dice 系数，先计算真实样本和预测样本相似度，取值范围在 0 到 1 之间，取值越大表示越相似：

$$Dice = \frac{2|x \cap y|}{|x| + |y|} \quad (4.6)$$

其中 x 表示真实标注图， y 表示输出的预测结果图， $|x \cap y|$ 表示 x 和 y 之间的交集，即预测正确的像素个数， $|x|$ 和 $|y|$ 分别表示 x 和 y 的元素个数，分子乘 2 保

证了保证分母重复计算后取值范围在 $[0, 1]$ 之间，Dice 损失则可表示为：

$$L_{dice}^1 = 1 - \frac{2|x \cap y|}{|x| + |y|} \quad (4.7)$$

Dice 损失适用于正负样本严重不平衡的情况，训练过程中更关注正样本的分类性能，但训练损失容易出现震荡不稳定的现象，尤其是分割的区域较小时。假设极端情况下，只有一个类别为正样本的像素，只要该像素被分类正确，不管其它像素预测如何，损失值就接近 0，反之，损失值接近 1，导致过分割现象。相反的，对于交叉熵损失，梯度的计算仅与预测点和真实点的距离相关，预测越接近真实值，梯度越小，当网络预测接近 0 或 1 时，梯度依然保持该特性，损失值是总体求平均，更多地依赖负样本。综上所述，交叉熵损失与 Dice 损失互补，本章结合两者的优势，提出了名为 CE-Dice 损失的新的损失函数以适应材料图像：

$$CE - Dice = -(L_{dice}^1)^\alpha * \log(L_{dice}^1) - (1 - L_{dice}^1)^\alpha * \log(1 - L_{dice}^1) \quad (4.8)$$

4.8 实验与讨论

4.8.1 实验数据集介绍

如图 4.7所示，在四种数据集上评估本章方法，包括三种碳钢数据集，钛合金 (TiAl) 数据集 (图4.7(a))。其中碳钢数据集来源于公开的大型碳钢数据库 UHCSDB[98] (UltraHigh Carbon Steel Micrograph DataBase)，从数据库选择了三种碳钢数据集，分别表示为 carbon steel F(图4.7(b))、carbon steel G(图4.7(c)) 和 carbon steel H(图 4.7(d))。四种数据训练集均为 2 张，carbon steel F 的测试集为 4 张，其余的数据测试集均为 3 张。

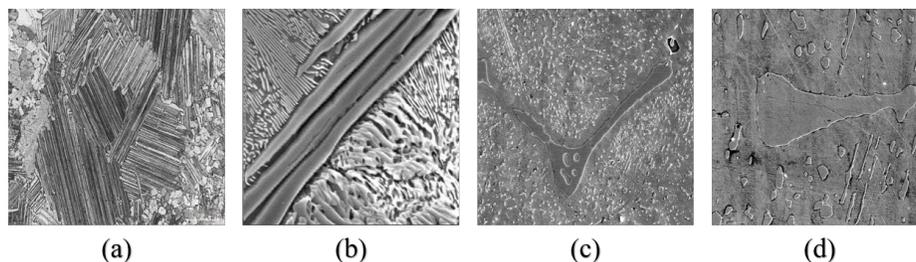


图 4.7: 实验数据集: (a)TiAl 图像; (b) carbon steel F 图像; (c) carbon steel G 图像; (d) carbon steel H 图像。

4.8.2 参数选择

本方法基于 pytorch 框架实现, 实验平台为 Ubuntu 服务器, 该服务器配有一块 Intel 至强 E5-2620 V4 2.10GHz 的 CPU 和 128G 的内存, 同时还配有 3 块 Nvidia Tesla K80 显卡, 每块显卡具有 24G 显存。为了保证提出的网络及对比网络能充分的训练, 初始学习率设置为 0.001, 使用 mIoU 监督每轮的结果。100 轮后, 在验证集上, mIoU 稳定地达到 0.95 以上, 停止训练。

4.8.3 分割结果的对比

在四种数据集上评估本章方法并和其它经典的分割方法比较, 比较的方法包括 UNet、UNet++、UNet+++、Attention-UNet、CENet、Improved DenseNet(上一章提出的方法) 和 FCN。图4.8展示了本章的方法和其它方法在 TiAl 测试集上的结果, 3 个测试样本命名为 E1、E2、E3, 表4.2展示每种方法在测试集上的 mIoU 和 mDice。图4.9展示了本章的方法和其它方法在 carbon steel F 数据集上的结果, 4 个测试样本命名为 F1、F2、F3、F4, 表4.3展示每种方法在测试集上的 mIoU 和 mDice。图4.10展示了本章的方法和其它方法在 carbon steel G 测试集上的结果, 3 个测试样本命名为 G1、G2、G3, 表4.4展示每种方法在测试集上的 mIoU 和 mDice。图4.11展示了本章的方法和其它方法在 carbon steel H 测试集上的结果, 3 个测试样本命名为 H1、H2、H3, 表4.5展示每种方法在测试集上的 mIoU 和 mDice。

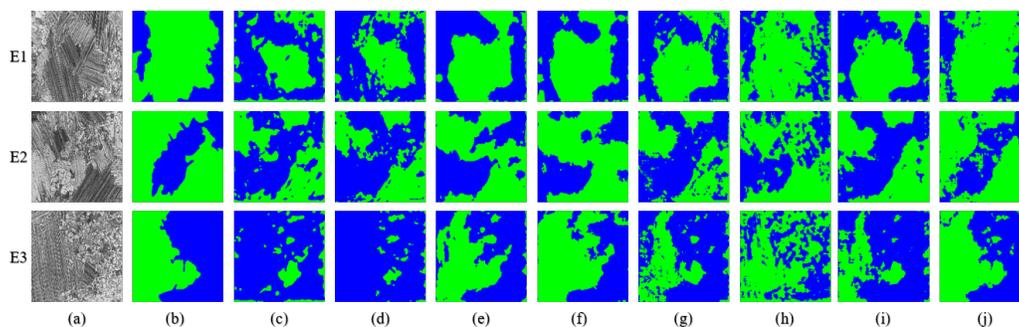


图 4.8: 本章的方法和其它方法在 TiAl 数据集上的结果: (a) 原图; (b) 真实标注图; (c) 使用 UNet 的结果; (d) 使用 UNet++ 的结果; (e) 使用 UNet+++ 的结果; (f) 使用 attention-UNet 的结果; (g) 使用 CENet 的结果; (h) 使用 improved DenseNet 的结果; (i) 使用 FCN 的结果; (j) 使用本章的方法的结果。

表 4.2: 本章的方法和其它方法在 TiAl 测试集上的 mIoU 和 mDice。

Name	E1	E2	E3	Average
Method				
UNet	0.365/0.532	0.524/0.686	0.294/0.424	0.394/0.547
UNet++	0.404/0.574	0.508/0.672	0.260/0.363	0.391/0.536
UNet+++	0.552/0.707	0.558/0.712	0.737/0.848	0.616/0.756
Attention-UNet	0.606/0.751	0.524/0.679	0.709/0.830	0.613/0.753
CENet	0.514/0.675	0.548/0.707	0.640/0.780	0.567/0.721
Improved DenseNet	0.586/0.729	0.476/0.637	0.458/0.627	0.507/0.664
FCN	0.539/0.696	0.612/0.758	0.607/0.753	0.586/0.736
本章的方法	0.694/0.814	0.721/0.835	0.834/0.910	0.750/0.853

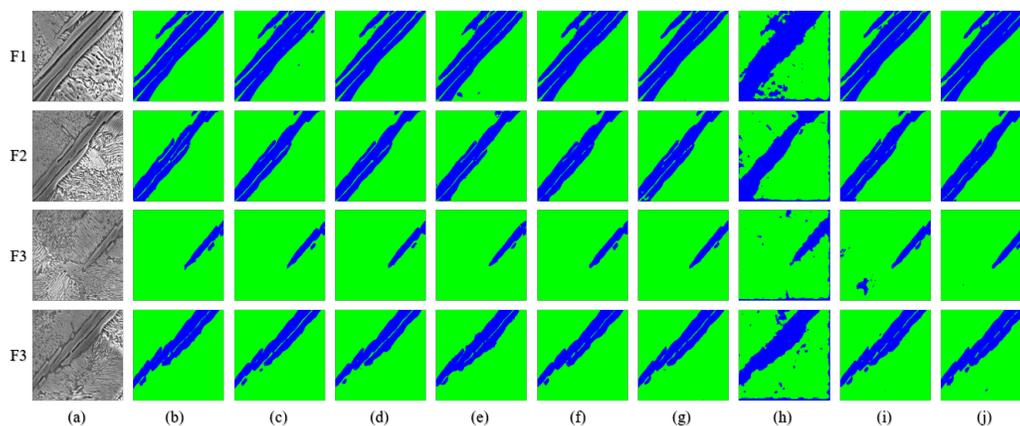


图 4.9: 本章的方法和其它方法在 carbon steel F 数据集上的结果: (a) 原图; (b) 真实标注图; (c) 使用 UNet 的结果; (d) 使用 UNet++ 的结果; (e) 使用 UNet+++ 的结果; (f) 使用 attention-UNet 的结果; (g) 使用 CENet 的结果; (h) 使用 improved DenseNet 的结果; (i) 使用 FCN 的结果; (j) 使用本章的方法的结果。

表 4.3: 本章的方法和其它方法在 carbon steel F 测试集上的 mIoU 和 mDice。

Name	F1	F2	F3	F4	Average
UNet	0.958/0.978	0.954/0.976	0.936/0.966	0.953/0.976	0.950/0.974
UNet++	0.961/0.980	0.952/0.975	0.924/0.959	0.946/0.972	0.946/0.972
UNet+++	0.937/0.967	0.919/0.957	0.919/0.956	0.918/0.957	0.923/0.959
Attention-UNet	0.961/0.980	0.945/0.971	0.939/0.968	0.940/0.969	0.946/0.972
CENet	0.958/0.978	0.945/0.972	0.936/0.966	0.961/0.980	0.950/0.974
Improved DenseNet	0.684/0.809	0.672/0.792	0.617/0.719	0.668/0.789	0.660/0.777
FCN	0.958/0.978	0.948/0.973	0.840/0.907	0.950/0.974	0.924/0.958
本章的方法	0.953/0.976	0.949/0.974	0.947/0.972	0.957/0.978	0.952/0.975

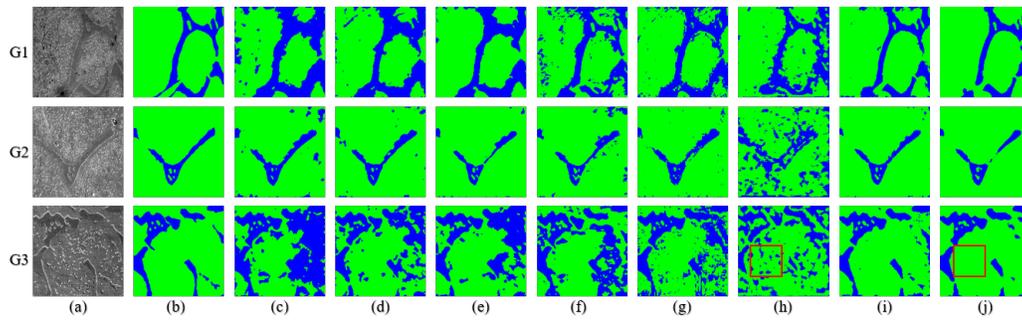


图 4.10: 本章的方法和其它方法在 carbon steel G 数据集上的结果: (a) 原图; (b) 真实标注图; (c) 使用 UNet 的结果; (d) 使用 UNet++ 的结果; (e) 使用 UNet+++ 的结果; (f) 使用 attention-UNet 的结果; (g) 使用 CENet 的结果; (h) 使用 improved DenseNet 的结果; (i) 使用 FCN 的结果; (j) 使用本章的方法的结果。

表 4.4: 本章的方法和其它方法在 carbon steel G 测试集上的 mIoU 和 mDice。

Name	G1	G2	G3	Average
UNet	0.620/0.755	0.765/0.854	0.476/0.634	0.620/0.748
UNet++	0.680/0.801	0.804/0.883	0.617/0.751	0.700/0.812
UNet+++	0.720/0.830	0.807/0.884	0.591/0.731	0.706/0.815
Attention-UNet	0.649/0.775	0.741/0.835	0.580/0.722	0.657/0.777
CENet	0.714/0.826	0.773/0.859	0.627/0.760	0.705/0.815
Improved DenseNet	0.589/0.722	0.441/0.545	0.447/0.584	0.492/0.617
FCN	0.839/0.910	0.835/0.904	0.861/0.923	0.845/0.912
本章的方法	0.874/0.931	0.843/0.909	0.866/0.926	0.861/0.922

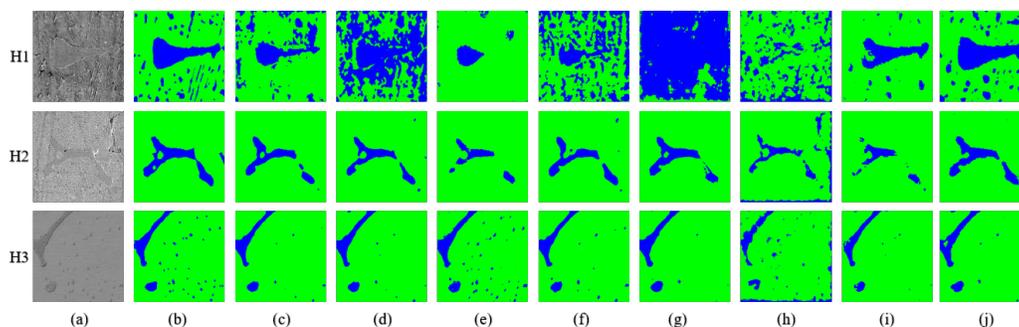


图 4.11: 本章的方法和其它方法在 carbon steel H 数据集上的结果: (a) 原图; (b) 真实标注图; (c) 使用 UNet 的结果; (d) 使用 UNet++ 的结果; (e) 使用 UNet+++ 的结果; (f) 使用 attention-UNet 的结果; (g) 使用 CENet 的结果; (h) 使用 improved DenseNet 的结果; (i) 使用 FCN 的结果; (j) 使用本章的方法的结果。

表 4.5: 本章的方法和其它方法在 carbon steel H 测试集上的 mIoU 和 mDice。

Name	H1	H2	H3	Average
UNet	0.507/0.637	0.832/0.902	0.866/0.925	0.735/0.821
UNet++	0.304/0.454	0.837/0.906	0.868/0.925	0.670/0.762
UNet+++	0.524/0.632	0.719/0.817	0.845/0.911	0.696/0.787
Attention-UNet	0.375/0.512	0.843/0.910	0.860/0.920	0.692/0.781
CENet	0.154/0.266	0.849/0.914	0.851/0.915	0.618/0.698
Improved DenseNet	0.413/0.535	0.586/0.696	0.519/0.608	0.506/0.613
FCN	0.727/0.831	0.700/0.799	0.828/0.899	0.752/0.843
本章的方法	0.744/0.846	0.865/0.924	0.832/0.902	0.814/0.891

在 4 种数据集上, 本章的方法均取得最好的结果, 在 TiAl 数据集上的 mIoU、mDice 平均值为 0.750、0.853, 在 carbon steel F 测试集上的 mIoU、mDice 平均值为 0.952、0.975, carbon steelG 数据集上的 mIoU、mDice 平均值为 0.861、0.922, carbon steel H 的 mIoU、mDice 平均值为 0.814、0.891。与经典的语义分割网络相比 (UNet, UNet++, UNet+++, FCN), 例如本章方法在 TiAl 数据集中的图像 E3 上, 真实标注图中颜色为大部分绿色的区域被正确识别, mIoU 和 mDice 分别为 0.834 和 0.910, 其余经典的语义分割网络未能正确分割绿色区域, mIoU 分别为 0.294、

0.260、0.737、0.607, mIoU 分别为 0.424、0.363、0.848、0.753, 因为图像 E3 的两相的纹理复杂, 本章方法施加特征融合模块融合高低层语义信息, 多尺度学习模块进一步挖掘纹理信息, 注意力机制保留重要纹理信息, 而经典的语义分割网络未能进一步提取纹理信息, 导致未能明显区分两相, 尤其是 UNet, UNet++ 几乎未成功的分割两相。Attention-UNet 与 CENet 都改进了 UNet 网络结构, Attention-UNet 施加了注意力机制, CENet 施加了上下文信息提取模块, 与该两种网络相比, 例如从 carbon steel H 数据集中的图像 H1 可以看出, Attention-UNet、CENet 未能成功识别与分割真实标注未蓝色的区域, mIoU 分别为 0.375、0.154, mDice 为 0.512、0.266, 在本章方法下, 蓝色区域则成功被分割, mIoU 为 0.744, mDice 为 0.846, 因为图像 H1 两种相纹理特征相似, 本章方法施加的三种模块使得决策更加全面, 而其它两种网络只单独使用注意力机制或上下文信息模块, 导致最终做决策较为片面, 分割效果不佳。与 Improved DenseNet(上一章的方法) 相比, 本章的方法大幅度提升分割效果, 例如从 carbon steel G 数据集中的图像 G3 可以看出, 红框所示的区域在 Improved DenseNet 下, 噪声较多, mIoU 为 0.447, mDice 为 0.584, 本章方法下, 较完整的识别与分割两相, mIoU 为 0.866, mDice 为 0.926, 因为首先本章的数据集纹理过于相似, 对于纹理信息的粒度大小和全局信息要求较高; 而上一章的方法输入的矩形块, 矩形块相比于整张图片, 尺寸太小, 感知区域范围非常小, 不含全局上下文信息, 且 FE block 连续使用三次池化操作, 丢失较多纹理信息, 因此不适用于各相纹理相似的材料图像; 相反, 本章的方法输入的是整张图片, 施加的三种模块提取了细粒度的纹理信息和全局上下文信息, 从而实现对纹理相似的材料数据集的精确分割。

4.8.4 消融实验

不同损失函数下的分割结果

材料图像中相的分布通常是不均衡的, 某些相分布极少, 导致训练效果一般。解决数据类别不平衡的思路主要有两种, 本章选择 Dice 损失并改进以适应材料图像数据 (名为 CE-Dice 损失)。为了验证 CE-Dice 损失对网络性能的影响, 图 4.12 展示

了本章提出的改进的 FCN 在其它损失函数, 主要是交叉熵损失、Dice 损失、Hinge 损失、Huber 损失、Lovaz-softmax 损失和 Precision Focal 损失 (上一章提出的损失函数), 其中 Dice 损失、Lovaz-softmax 损失 [103]、Precision Focal 损失是属于考虑数据分布不平衡的损失函数。表 4.6展示了改进的 FCN 使用其它损失函数在 4 种数据集上的平均 mIoU 和平均 mDice。除了在 carbon steel F 数据集上, 非考虑数据分布不平衡的损失函数 (交叉熵损失、Hing 损失、Huber 损失) 表现性能一般, 例如改进的 FCN 通过使用 CE-Dice 损失, 图像 E2、图像 G3、图像 H2 的 mIoU 分别是 0.721、0.866、0.865, mDice 分别是 0.835、0.909、0.926, 而在非考虑数据分布不平衡的损失函数下, 图像 E2 最好的 mIoU 和 mDice 分别是 0.608 和 0.754(使用 Hinge 损失), 图像 G3 最好的 mIoU 和 mDice 分别是 0.836 和 0.908(使用 Huber 损失, 图像 H2 最好的 mIoU 和 mDice 分别是 0.679 和 0.791(使用 Hinge 损失)。与考虑数据分布不平衡的损失函数相比 (Dice 损失、Lovaz-softmax 损失和 Precision Focal 损失), 图像 E2 最好的 mIoU 和 mDice 分别是 0.646 和 0.782(使用 Dice 损失), 图像 G3 最好的 mIoU 和 mDice 分别是 0.854 和 0.919(使用 Lovaz-softmax 损失), 图像 H2 最好的 mIoU 和 mDice 分别是 0.830 和 0.901(使用 Lovaz-softmax 损失)。因此, CE-Dice 损失对改进的 FCN 性能提升明显, 因为 CE-Dice 损失结合了 Dice 损失与交叉熵损失二者的优势, 解决了训练过程中梯度饱和现象, 减缓过分割现象的发生, 从而优化了分割结果。在 carbon steel F 数据集上, 使用 CE-Dice 损失对分割结果的提升不大, 因为 carbon steel F 数据集分布较为平衡, 使用非考虑数据分布不平衡的损失函数就已经取得较好的结果, 最好的平均 mIoU 和 mDic 分别为 0.952 和 0.975(使用交叉熵损失), 没有必要使用 CE-Dice 损失。

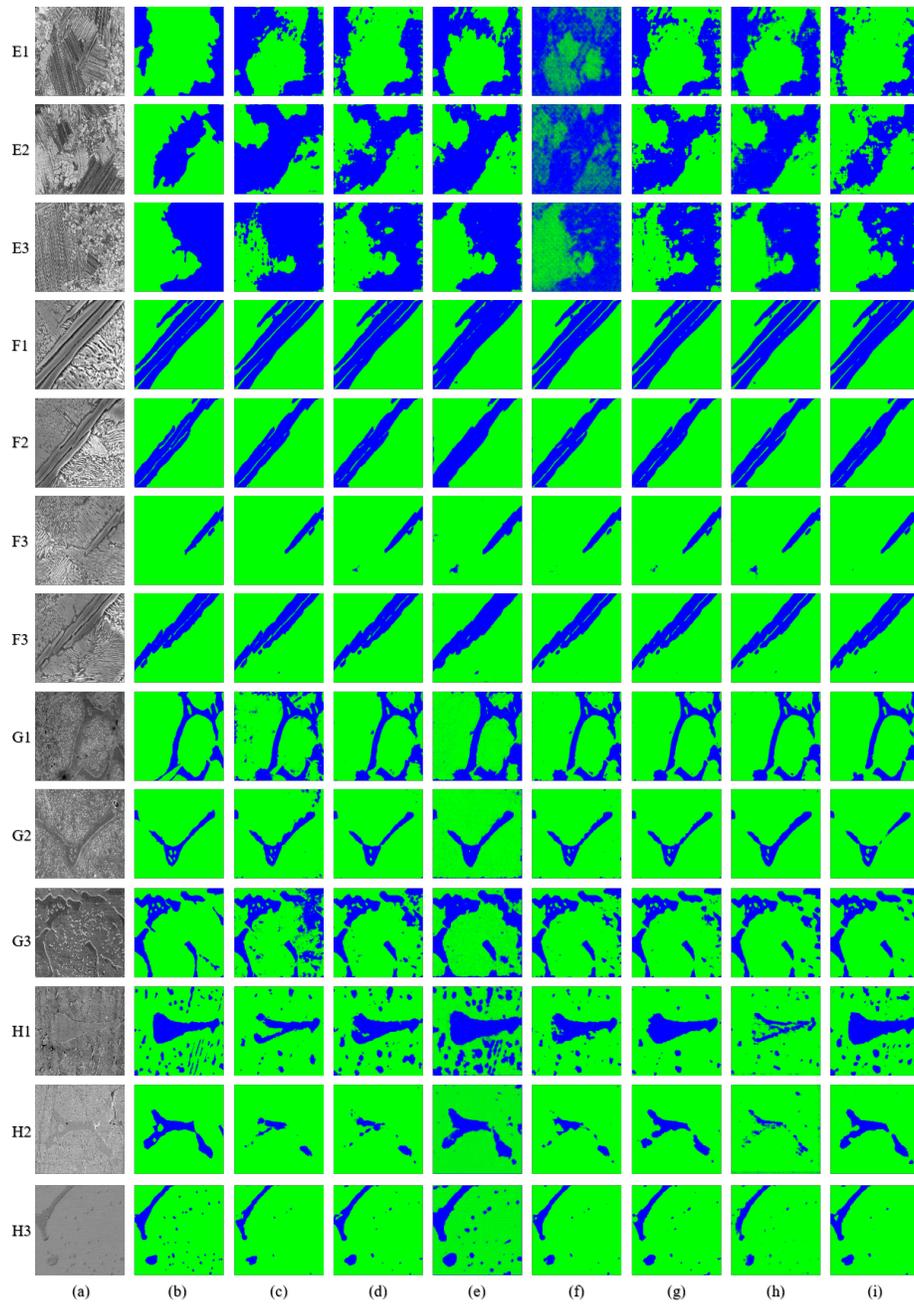


图 4.12: 改进的 FCN 使用其它损失函数和 CE-Dice 在 4 种数据集上的结果。图像 E1-E3 属于 TiAl 数据集；图像 F1-F4 属于 carbon steel F 数据集；图像 G1-G3 属于 carbon steel G 数据集；图像 H1-H3 属于 carbon steel H 数据集；(a) 原图；(b) 真实标注图；(c) 使用交叉熵损失的结果；(d) 使用 Dice 损失的结果；(e) 使用 Hinge 损失的结果；(f) 使用 Huber 损失的结果；(g) 使用 Lovasz-softmax 损失的结果；(h) 使用 Precision Focal 损失的结果；(i) 使用 CE-Dice 损失的结果。

表 4.6: 改进的 FCN 使用其它损失函数和 CE-Dice 损失在 4 种数据集上的 mIoU 和 mDice。

Loss \ Data Set	TiAl	Carbon steel F	Carbon steel G	Carbon steel H
交叉熵损失	0.567/0.721	0.952/0.975	0.746/0.846	0.703/0.800
Dice 损失	0.710/0.826	0.949/0.974	0.842/0.910	0.744/0.833
Hinge 损失	0.658/0.787	0.895/0.943	0.694/0.808	0.639/0.761
Huber 损失	0.460/0.619	0.951/0.975	0.848/0.914	0.761/0.848
Lovaz-softamx 损失	0.645/0.779	0.947/0.972	0.869/0.928	0.805/0.885
Precision Focal 损失	0.662/0.790	0.936/0.966	0.832/0.904	0.660/0.761
CE-Dice 损失	0.750/0.853	0.952/0.975	0.861/0.922	0.814/0.891

其它网络使用 Dice 损失和 CE-Dice 损失的结果

为了进一步验证 CE-Dice 损失的性能, 还评估了其它深度学习网络 (UNet、attention-UNet、CENet、FCN) 使用 Dice 损失和 CE-Dice 损失在本章 4 种数据集上的结果。图 4.13 和表 4.7 显示 CE-Dice 损失对一些网络的性能有明显的提升。例如在 TiAl 数据集上, FCN 在 CE-Dice 损失下的平均 mIoU 和 mDice 分别为 0.752 和 0.854, 而在 Dice 损失下的平均 mIoU 和 mDice 分别为 0.724 和 0.833, 特别的, 在图像 E2 上, 通过使用 CE-Dice 损失, FCN 的性能提升了 4%(CE-Dice 下的 mIoU 为 0.707, Dice 损失下的 mIoU 为 0.673), 提升效果明显。在 carbon steel F 数据集上, CENet 在 CE-Dice 损失下的平均 mIoU 和 mDice 分别为 0.954 和 0.976, 在 Dice 损失下的平均 mIoU 和 mDice 分别为 0.952 和 0.975, 提升幅度较小, 因为 carbon steel F 的相区分相对于其它数据集, 区分度明显, 因为 carbon steel F 数据集分布相对较为平衡, 不需要使用 CE-Dice 损失。在 carbon steel G 数据集上, FCN 在 CE-Dice 损失下的平均 mIoU 和 mDice 分别为 0.776 和 0.867, 在 Dice 损失下的平均 mIoU 和 mDice 分别为 0.675 和 0.793, 特别的, 在图像 G1 上, 通过使用 CE-Dice 损失, FCN 的性能提升了 20%(CE-Dice 下的 mIoU 为 0.749, Dice 损失下的 mIoU 为 0.549), 大幅度优化分割结果。在 carbon steel H 数据集上, UNet 在 CE-Dice 损失的平均 mIoU 和 mDice 分别为 0.692 和 0.780, 在 Dice 损失的平均 mIoU 和 mDice

分别为 0.596 和 0.671，特别的，在图像 H2 上，通过使用 CE-Dice 损失，UNet 的性能提升了 6%(CE-Dice 下的 mIoU 为 0.893，Dice 损失下的 mIoU 为 0.823)，提升效果明显。因此，CE-Dice 损失可以有效地改善某些网络的性能。通过结合交叉熵损失和 Dice 损失的优势，使得网络训练过程稳定，从而优化分割结果。

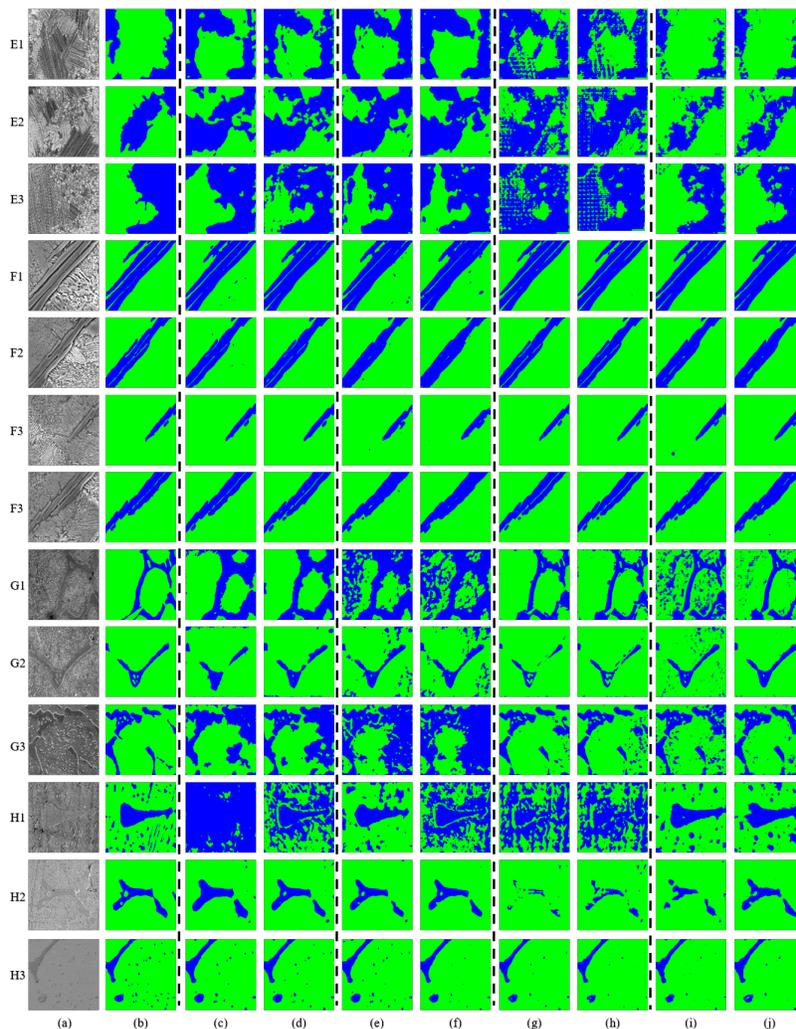


图 4.13: 其它网络使用 Dice 损失和 CE-Dice 损失的结果。图像 E1-E3 属于 TiAl 数据集；图像 F1-F4 属于 carbon steel F 数据集；图像 G1-G3 属于 carbon steel G 数据集；图像 H1-H3 属于 carbon steel H 数据集；(a) 原图；(b) 真实标注图；(c)UNet 使用 Dice 损失；(d)UNet 使用 CE-Dice 损失；(e)attention-UNet 使用 Dice 损失；(f)attention-UNet 使用 CE-Dice 损失；(g)CENet 使用 Dice 损失；(h)CENet 使用 CE-Dice 损失；(i)FCN 使用 Dice 损失；(j)FCN 使用 CE-Dice 损失。

表 4.7: 其它网络使用 Dice 损失和 CE-Dice 损失在 4 种数据集上的 mIoU 和 mDice。

Data Set	TiAl	Carbon steel F	Carbon steel G	Carbon steel H
UNet + Dice 损失	0.644/0.775	0.951/0.975	0.640/0.765	0.596/0.671
UNet + CE-Dice 损失	0.554/0.709	0.944/0.971	0.662/0.781	0.692/0.780
Attention-UNet + Dice 损失	0.579/0.729	0.912/0.952	0.504/0.647	0.784/0.865
Attention-UNet + CE-Dice 损失	0.563/0.718	0.907/0.950	0.455/0.601	0.683/0.771
CENet + Dice 损失	0.428/0.594	0.952/0.975	0.788/0.872	0.567/0.664
CENet + CE-Dice 损失	0.463/0.627	0.954/0.976	0.795/0.879	0.615/0.718
FCN + Dice 损失	0.724/0.833	0.945/0.971	0.675/0.793	0.783/0.868
FCN + CE-Dice 损失	0.752/0.854	0.943/0.971	0.776/0.867	0.805/0.885

其它模块的消融实验

本章中使用了特征融合模块 (CFF block) 融合高层与低层的语义特征, 使用多尺度学习模块 (Multi-scale block) 进一步提取细粒度信息和上下文信息, 使用注意力机制模块 (Attention block) 保留重要特征图, 优化资源配置。为了验证添加模块的重要性, 图 4.14和表 4.8展示了网络分别不使用一种模块和仅使用一种模块的结果。从总体角度上来说, 除了在 carbon steel G 数据集上, 共同使用三种模块的条件下表现较优。在 TiAl 数据集上, 使用三种模块的平均 mIoU 和 mDice 为 0.750 和 0.853, 而其它情况最优平均 mIoU 和 mDice 为 0.747 和 0.851(仅使用多尺度学习模块)。在 carbon steel F 数据集上, 使用三种模块的平均 mIoU 和 mDice 为 0.952 和 0.975, 而其它情况最优平均 mIoU 和 mDice 为 0.951 和 0.975(不使用特征融合模块)。在 carbon steel H 数据集上, 使用三种模块的平均 mIoU 和 mDice 为 0.814 和 0.891, 而其它情况最优平均 mIoU 和 mDice 为 0.763 和 0.853(不使用特征融合模块)。在 carbon steel G 数据集上, 仅使用注意力机制模块就达到最优 (平均 mIoU 和 mDice 为 0.878 和 0.932), 因为下采样阶段的模块包含卷积与池化过程, 导致丢失一些信息, 对 carbon steel G 的分割结果影响较大。

从每个模块的重要性的角度来说, 在 TiAl 数据集上, 多尺度学习模块对分割结果的影响最大 (单独使用多尺度学习模块的平均 mIoU 为 0.747 和 0.851), 这是因为 TiAl 数据集纹理信息丰富, 多尺度学习的模块能够深度挖掘纹理信息, 进而

区分各相。在 carbon steel F 数据集上, 特征融合模块对分割结果的影响最大 (单独使用特征融合模块的平均 mIoU 为 0.944 和 0.971), 因为 carbon steel F 数据集边界较为清晰, 纹理区分度明显, 特征融合保留纹理特征并融合了高层与低层语义特征。在 carbon steel G 数据集上, 注意力机制模块对分割结果的影响最大 (单独使用注意力机制模块的平均 mIoU 和 mDice 为 0.878 和 0.932), 因为 carbon steel H 数据集第二类相分布较为杂乱 (分割结果为绿色区域), 注意力机制能够保留关键纹理特征, 去除冗余特征对分割任务的干扰。在 carbon steel H 数据集上, 多尺度学习制模块对分割结果的影响最大 (单独使用多尺度学习模块的平均 mIoU 为 0.703 和 0.808), 因为 carbon steel H 数据集的两相纹理较为相似, 通过多尺度学习模块提取的每相的特征能够独特的表示各相, 进而区分各相。

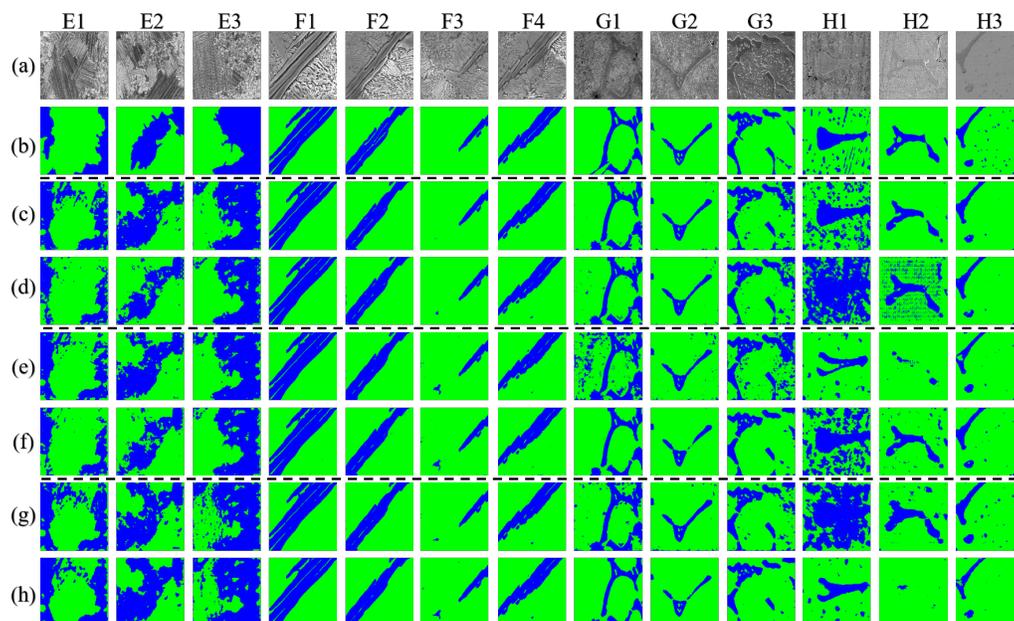


图 4.14: 其它模块的消融实验。图像 E1-E3 属于 TiAl 数据集; 图像 F1-F4 属于 carbon steel F 数据集; 图像 G1-G3 属于 carbon steel G 数据集; 图像 H1-H3 属于 carbon steel H 数据集; (a) 原图; (b) 真实标注图; (c) 不使用特征融合模块的结果; (d) 只使用特征融合模块的结果; (e) 不使用多尺度模块的结果; (f) 只使用多尺度模块的结果; (g) 不使用注意力机制模块的结果; (h) 只使用注意力机制模块的结果。

表 4.8: 消融其它模块下的平均 mIoU 和 mDice

Other ablation	Data Set	TiAl	Carbon steel F	Carbon steel G	Carbon steel H
不使用特征融合模块	0.660/0.789	0.951/0.975	0.820/0.915	0.763/0.853	
只使用特征融合模块	0.747/0.851	0.944/0.971	0.819/0.896	0.578/0.696	
不使用多尺度模块	0.680/0.804	0.932/0.963	0.710/0.819	0.678/0.774	
只使用多尺度模块	0.744/0.848	0.931/0.963	0.872/0.929	0.703/0.808	
不使用注意力机制模块	0.613/0.757	0.942/0.970	0.798/0.882	0.619/0.733	
只使用注意力机制模块	0.622/0.763	0.928/0.961	0.878/0.932	0.665/0.745	
同时使用三种模块	0.750/0.853	0.952/0.975	0.861/0.922	0.814/0.891	

4.8.5 本章方法在上一章数据集的结果

为了进一步探究本章方法和上一章方法的各自优势，图4.15和表4.9展示了本章提出的方法在上一章数据集上的结果，除了在 carbon steel A 数据集上，Improved DenseNet 表现较优，例如在 ceramics 数据集上，使用 Improved DenseNet 的 mIoU 为 0.815，mDice 为 0.896，使用本章方法的 mIoU 为 0.642，mDice 为 0.764，因为上一章的数据集尽管纹理复杂，但颜色区分明显，对全局信息不敏感，上一章的方法通过将图片分解为矩形块，扩充了训练样本数量，从而解决了材料图像小样本问题，而本章提出的方法从深度挖掘纹理信息的角度提升分割结果，但可能会受制于样本数量不足的问题未能有效训练模型。总的来说，第二章的方法解决了材料图像分割中的小样本问题和数据分布不平衡问题，但感知区域较小，不含全局信息，不适用相之间纹理相似的材料图像；本章的方法弥补上一章方法的缺陷，但可能会因样本数量不足未能有效训练模型，因此两种方法各有优劣，但优势互补，两种方法的结合全方位的实现材料图像微观结构精确分割，选择哪种方法，需要结合具体数据集使用。

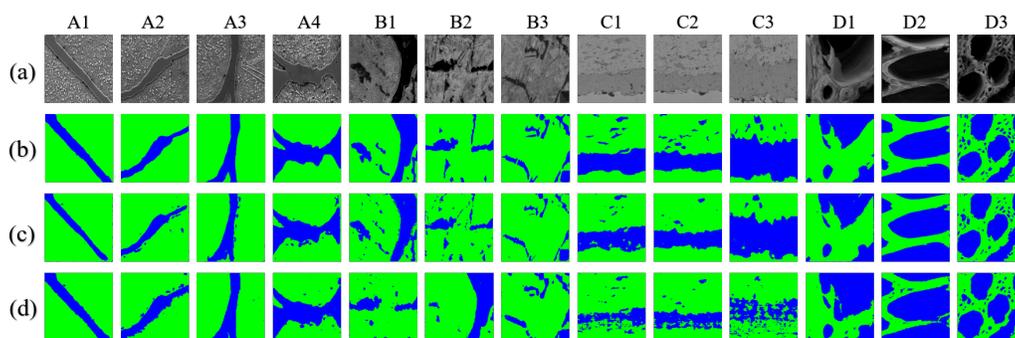


图 4.15: 本章方法在上一章数据集上的结果。图像 A1-A4 属于 carbon steel A 数据集; 图像 B1-B3 属于 carbon steel B 数据集; 图像 C1-C3 属于 ceramics 数据集; 图像 D1-D3 属于 wood 数据集; (a) 原图; (b) 真实标注图; (c) 使用 Improved DenseNet 的结果; (d) 使用本章方法的结果。

表 4.9: 本章方法在上一章 4 种数据集上的 mIoU 和 mDice。

Method/Data set	Carbon steel A	Carbon steel B	Ceramics	Wood
Improved DenseNet	0.809/0.882	0.769/0.858	0.815/0.896	0.851/0.915
本章的方法	0.871/0.931	0.732/0.835	0.642/0.764	0.761/0.862

4.9 本章小结

对于第三章提出方法的缺陷, 即在各相纹理较为相似的材料图像分割效果一般, 本章提出一种基于特征融合的材料图像分割方法。本方法基于 FCN 对材料图像进行像素级语义分割, 有效提高了材料微观结构识别的准确度。与第二章提出的方法相比, 本方法能较好的识别与分割各相纹理相似的材料图像。本方法相比现有的深度学习语义分割方法具有如下两点优势: 首先, 本方法施加的三种模块有效的优化分割结果, 提出的特征融合模块融合高层与底层语义特征, 多尺度学习模块提取细粒度信息与上下文信息, 注意力机制模块则保留关键特征图, 优化资源配置, 三种模块的结合使网络决策更加全面; 其次, 本方法针对 Dice 损失存在的梯度饱和问题和材料图像中数据量不平衡问题, 提出了一种改进的 Dice 损失, 改进的 Dice 损失结合 Dice 损失与交叉熵损失的优势, 优化训练过程, 进一步提升分割结果。通过在四种各相纹理较为相似的材料图像数据集的实验表明, 本章方法能够达到较高的分

割识别准确度。本方法的贡献点可总结为以下几点：

(1) 本分割方法通过深度挖掘复杂纹理特征，解决了第三章提出的方法感知区域较小，不含全局信息，不适用各相纹理相似的材料分割的缺陷；

(2) 改进的 Dice 损失解决了材料图像中数据分布不平衡的问题。

本章方法也存在一些不足。一方面，本方法对于极小的区域识别效果较差，因为多尺度学习模块中的空洞卷积层与池化层均会导致丢失信息，因此网络未能充分提取极小的区域的纹理特征；另一方面，本方法在边界处的分割识别上仍存在一定的误差，主要是各相之间纹理相似的数据集的边界模糊；此外，本方法不能很好的对重叠的微观结构和低分辨率的微观结构进行准确的分割和识别。上述问题值得进一步关注与分析。

第五章 总结与展望

5.1 结论

本研究结合计算机视觉技术、模式识别、机器学习和材料微观结构定量分析等领域知识，重点研究了利用计算机技术实现材料图像中微观结构自动、准确的分割识别方法。由于材料图像类型多样，纹理特征、图像分辨率等差异较大，且材料图像存在固有的分割难点，即小样本和数据分布不平衡的问题，因此难以找到一种通用的微观结构分割识别方法。本研究分析材料图像的特点，结合实际情况分别提出两种材料图像微观结构分割和识别方法。研究成果及贡献如下：

1) 针对材料图像分割中存在的小样本问题，第三章提出的方法抓住材料图像同相的像素高度相似的特性，利用超像素算法获取矩形块，解决了材料图像的小样本问题，提出的改进的 DenseNet 通过施加特征增强模块保留纹理特征，去除冗余特征的干扰，设计的过渡层上采样方法，较好的恢复特征图信息。

2) 针对材料图像中数据分布不平衡的问题，本研究使用并改进两种损失函数，第三章解决分类任务中数据分布不平衡问题，基于 Focal 损失提出 Precision Foca 损失，将置信度替换为精度，更准确的反映样本分类的难度，并反馈给网络，优化训练过程；第四章解决了分割任务中的数据分布不平衡的问题，基于 Dice 损失提出 CE-Dice 损失，CE-Dice 损失结合交叉熵损失和 Dice 损失，训练过程更平滑，优化分割结果。

3) 实现了具有复杂纹理的材料图像的精确分割，第三章提出的特征增强模块，保留关键纹理特征，去除冗余特征对矩形块识别的干扰；第四章提出改进的 FCN，弥补了第三章方法中矩形块的区域较小，不含全局信息，不适用各相纹理相似的材料图像分割任务的缺陷，级联的特征融合模块融合高层与低层语义特征，多尺度学习模块深度挖掘细粒度信息和全局上下文信息，采用注意力机制模块关注重要的特征图，优化资源配置，三种模块互为补充。

本研究对以上两种方法都进行了验证实验，实验表明本研究提出的两种方法在

对应的情况下都能取得较高的分割识别准确率，具有较强的鲁棒性与泛化性能，并且两种方法优势互补，可以全方位的应用于材料图像微观结构识别领域，为后续的材料基因工程提供数据基础。

5.2 工作展望

通过本文的研究与分析，作者认为计算机技术将为材料科学领域的发展提供重要的帮助。如今，深度学习应用广泛，因此深度学习技术将成为材料图像处理与分析的重要工具，为了更好的推动深度学习应用于材料图像领域，作者认为在以下的三个方面值得研究：

1) 其它思路解决材料图像中的小样本问题。解决材料图像中的小样本问题是实现材料图像微观结构分割与识别的关键，本研究抓住了材料图像的特性一定程度上解决了材料图像中的小样本问题，事实上，小样本问题也常见于深度学习领域，还可以将现小样本技术应用于材料图像，从而使得解决方案更全面。例如使用迁移学习，材料图像是一种复杂纹理图像，可利用在其它大型复杂纹理的图像数据集训练模型，然后迁移到材料图像中；

2) 研究材料结构与性能的内在关联。在材料科学研究中，了解材料的结构如何影响其性能有助于定量分析，但存在两个难点，一是实验产生的绝大多数数据从未被机器学习模型分析过，并且实验室数据不共享，对此可以考虑构建结构化和非结构的数据库；二是神经网络在学习如何识别结构对称性和周期性，神经网络提取的特征一般是 2D 或 3D，材料图像的纹理信息较为丰富，在常规的空间提取的特征对于挖掘纹理信息是不够的，对此，可以构建高级别的空间，用以充分挖掘特征信息，例如形状空间，为更好地理解结构-性能关系铺平道路；

3) 低分辨率、重叠与被遮挡的微观结构分割与识别。由于制备过程技术、温度等多因素的影响，导致存在低分辨率、重叠与被遮挡等低质量的材料图像。实现此类材料图像精确的分割是目前深度学习领域的难点之一。一方面，可以将专家领域知识引入深度学习领域，如交互式图像分割；另一方面，可以将基于多模态学习技术应用于材料图像领域，结合文字、图像等多种信息，重建高质量的样本。

以上问题的有效解决能够完善深度学习技术，也能推动材料图像处理领域高速发展，引领材料基因工程技术革命。

参考文献

- [1] Jain A, Ong S P, Hautier G, et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation [J]. *APL Materials*, 2013, 1 (1): 11–20.
- [2] 李冬俊, 党朋, 蔡西川, et al. 国内材料基因工程与新材料研究概况 [C]. In 2021 年中国材料大会, 2021: 381–388.
- [3] Kalidindi S R, Niezgodka S R, Salem A A. Microstructure informatics using higher-order statistics and efficient data-mining protocols [J]. *JOM*, 2011, 63 (4): 34–41.
- [4] Decost B L, Elizabeth H, A. A computer vision approach for automated analysis and classification of microstructural image data [J]. *Computational Materials Science*, 2015, 110: 126–133.
- [5] Otsu N. A Threshold Selection Method from Gray-Level Histograms [J]. *IEEE Transactions on Systems Man and Cybernetics*, 1979, 9 (1): 62–66.
- [6] Pun T. A new method for gray-level picture thresholding using the entropy of the histogram [J]. *Signal Processing*, 1985, 2 (3): 223–237.
- [7] Yen J C, Chang F J, Chang S. A new criterion for automatic multilevel thresholding [J]. *IEEE transactions on image processing*, 1995, 4 (3): 370–378.
- [8] Deng C, Ma W, Yin Y. An edge detection approach of image fusion based on improved Sobel operator [C]. In 2011 4th International Congress on Image and Signal Processing, 2011: 1189–1193.
- [9] Ulupinar F, Medioni G. Refining edges detected by a LoG operator [J]. *Computer Vision, Graphics, and Image Processing*, 1990, 51 (3): 275–298.
- [10] Er-sen L, Shu-long Z, Bao-shan Z, et al. An Adaptive Edge-Detection Method Based on the Canny Operator [C]. In 2009 International Conference on Environmental Science and Information Application Technology, 2009: 465–469.
- [11] Rosenfeld A. The Max Roberts Operator is a Hueckel-Type Edge Detector [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1981, 3 (1): 101–103.

- [12] Tremeau A, Borel N. A region growing and merging algorithm to color segmentation [J]. *Pattern Recognition*, 1997, 30 (7): 1191–1203.
- [13] Meyer B. Color image segmentation [C]. In *International Conference on Image Processing and Its Applications*, 2002: 579–587.
- [14] Cheng Y. Mean shift, mode seeking, and clustering [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995, 17 (8): 790–799.
- [15] Fukunaga K, Hostetler L. The estimation of the gradient of a density function, with applications in pattern recognition [J]. *IEEE Transactions on Information Theory*, 1975, 21 (1): 32–40.
- [16] Sheikh Y A, Khan E A, Kanade T. Mode-seeking by Medoidshifts [C]. In *2007 IEEE 11th International Conference on Computer Vision*, 2007: 1–8.
- [17] Levinshtein A, Stere A, Kutulakos K N, et al. TurboPixels: Fast Superpixels Using Geometric Flows [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31 (12): 2290–2297.
- [18] Achanta R, Shaji A, Smith K, et al. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34 (11): 2274–2282.
- [19] Boykov Y, Jolly M-P. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images [C]. In *Proceedings Eighth IEEE International Conference on Computer Vision*, 2001: 105–112.
- [20] Rother C, Kolmogorov V, Blake A. "GrabCut": Interactive Foreground Extraction Using Iterated Graph Cuts [J]. *ACM Transactions on Graphics*, 2004, 23 (3): 309–314.
- [21] Tang M, Gorelick L, Veksler O, et al. GrabCut in One Cut [C]. In *IEEE International Conference on Computer Vision*, 2013: 1769–1776.
- [22] Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 39 (4): 640–651.

- [23] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39 (12): 2481–2495.
- [24] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40 (4): 834–848.
- [25] Chen L C, Zhu Y, Papandreou G, et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation [C]. In *2018 European Conference on Computer Vision*, 2018: 833–851.
- [26] Chen L C, Papandreou G, Schroff F, et al. Rethinking Atrous Convolution for Semantic Image Segmentation [C]. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017: 743–751.
- [27] Zhao H, Shi J, Qi X, et al. Pyramid Scene Parsing Network [C]. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017: 6230–6239.
- [28] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation [C]. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015: 234–241.
- [29] Zhou Z, Siddiquee M M R, Tajbakhsh N, et al. UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation [J]. *IEEE Transactions on Medical Imaging*, 2020, 39 (6): 1856–1867.
- [30] Oktay O, Schlemper J, Folgoc L L, et al. Attention U-Net: Learning Where to Look for the Pancreas [C]. In *The International Conference on Medical Imaging with Deep Learning*, 2018: 1–10.
- [31] Gu Z, Cheng J, Fu H, et al. CE-Net: Context Encoder Network for 2D Medical Image Segmentation [J]. *IEEE Transactions on Medical Imaging*, 2019, 38 (10): 2281–2292.
- [32] Li G, Liu Z, Ling H. ICNet: Information Conversion Network for RGB-D Based Salient Object Detection [J]. *IEEE Transactions on Image Processing*, 2020, 29: 4873–4884.

- [33] Li Y, Li X, Xiao C, et al. EACNet: Enhanced Asymmetric Convolution for Real-Time Semantic Segmentation [J]. *IEEE Signal Processing Letters*, 2021, 28: 234–238.
- [34] Yang Z, Yu H, Fu Q, et al. NDNNet: Narrow While Deep Network for Real-Time Semantic Segmentation [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 22 (9): 5508–5519.
- [35] Mehta S, Rastegari M, Caspi A, et al. ESPNet: Efficient Spatial Pyramid of Dilated Convolutions for Semantic Segmentation [C]. In *European Conference on Computer Vision*, 2018: 561–580.
- [36] Chatterjee O, Das K, Dutta S, et al. Phase extraction and boundary removal in dual phase steel micrographs [C]. In *Annual IEEE India Conference (INDICON)*, 2010: 1–5.
- [37] Sun Q D, Gao S F, Huang J W, et al. Metallographical Image Segmentation and Compression [J]. *Applied Mechanics and Materials*, 2012, 152-154: 276–280.
- [38] Ananyev M, Medvedev D, Gavriljuk A, et al. Cu and Gd co-doped BaCeO₃ proton conductors: Experimental vs SEM image algorithmic-segmentation results [J]. *Electrochimica Acta*, 2014, 125: 371–379.
- [39] Lopez P, Lira J, Hein I. Discrimination of Ceramic Types Using Digital Image Processing by Means of Morphological Filters [J]. *Archaeometry*, 2014, 57 (1): 146–162.
- [40] Chen Y, Chen J. A watershed segmentation algorithm based on ridge detection and rapid region merging [C]. In *2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 2014: 420–424.
- [41] Liu J, Chen J. An Improved Iterative Watershed According to Ridge Detection for Segmentation of Metallographic Image [J]. *Metallographic Image*, 2012, 8: 1–11.
- [42] Canny J. A Computational Approach to Edge Detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, 8 (6): 679–698.
- [43] Gao W, Zhang X, Yang L, et al. An improved Sobel edge detection [C]. In *2010 3rd International Conference on Computer Science and Information Technology*, 2010: 67–71.

- [44] Rosenfeld A. The Max Roberts Operator is a Hueckel-Type Edge Detector [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1981, 3 (1): 101–103.
- [45] 廖正全, 滕奇志, 罗代升. 合金图像分割算法研究 [J]. 计算机应用, 2009, 29 (12): 3326–3328.
- [46] 蒋明星, 陈国华. 金相图像的晶界恢复与重建 [J]. 光学精密工程, 2011, 19 (10): 2541–2549.
- [47] 邓仕超, 刘铁根, 萧泽新. 应用 Canny 算法和灰度等高线的金相组织封闭边缘提取 [J]. 光学精密工程, 2010, 18 (10): 2314–2323.
- [48] Santis A D, Lacoviello F, Lacoviello D. Active Contour Segmentation for the Identification of Metallographic and Morphological Elements of Interest in ductile cast iron [J]. La Metallurgia Italiana, 2007, 99 (6): 15–24.
- [49] Han Y, Li R, Zeng Y, et al. Automatic detection of Kikuchi bands based on Radon transform and PPHT [J]. Journal of Microscopy, 2021, 285 (2): 95–111.
- [50] Zhao H, Comer M. A Hybrid Markov Random Field/Marked Point Process Model for Analysis of Materials Images [J]. IEEE Transactions on Computational Imaging, 2016, 2 (4): 395–407.
- [51] 丁贤云, 朱煜. 基于二维熵的人工鱼群材料图像分割方法 [J]. 激光与红外, 2010, 40 (2): 210–214.
- [52] 朱做涛, 穆志韬, 苏维国, et al. 基于图像处理技术的铝合金腐蚀等级评定方法 [J]. 南京航空航天大学学报, 2010, 42 (3): 383–386.
- [53] 王宝珠, 李芳影, 刘翠响, et al. 基于图像处理技术的多晶体材料金相图像分割算法研究 [J]. 铸造, 2015, 64 (11): 1078–1081.
- [54] 秦湘阁, 刘国权. 基于 Monte Carlo Potts 方法的三维大尺度晶粒组织仿真模型及定量表征 [J]. 工程科学学报, 2004, 26 (1): 49–52.
- [55] Tarabalka Y, Chanussot J, Benediktsson J A. Segmentation and classification of hyperspectral images using watershed transformation [J]. Pattern Recognition, 2010, 43 (7): 2367–2379.

- [56] Chen L, Han Y, Guan Y H, et al. Two-Dimensional Fuzzy Clustering Algorithm (2DFCM) for Metallographic Image Segmentation Based on Spatial Information [C]. In International Conference on Information Science and Control Engineering, 2015: 519–521.
- [57] Azimi S M, Britz D, Engstler M, et al. Advanced Steel Microstructural Classification by Deep Learning Methods [J]. *Scientific Reports*, 2018, 8 (1): 1–14.
- [58] Decost B L, Lei B, Francis T, et al. High throughput quantitative metallography for complex microstructures using deep learning: A case study in ultrahigh carbon steel [J]. *Microscopy and Microanalysis*, 2019, 25 (1): 21–29.
- [59] Wei L, Field K G, Dane M. Automated defect analysis in electron microscopic images [J]. *Npj Computational Materials*, 2018, 4 (1): 1–9.
- [60] Dai D, Xu T, Wei X, et al. Using machine learning and feature engineering to characterize limited material datasets of high-entropy alloys [J]. *Computational Materials Science*, 2020, 175 (1): 1–9.
- [61] Yang X, Li L, Tao Q, et al. Rapid discovery of narrow bandgap oxide double perovskites using machine learning [J]. *Computational Materials Science*, 2021, 196 (11): 1–8.
- [62] Papa J P, Nakamura R, Albuquerque V, et al. Computer techniques towards the automatic characterization of graphite particles in metallographic images of industrial material [J]. *Expert Systems with Applications*, 2013, 40 (2): 590–597.
- [63] Han Y, Lai C, Wang B, et al. Segmenting Images with Complex Textures by Using Hybrid Algorithm [J]. *Journal of Electronic Imaging*, 2019, 28 (1): 13–30.
- [64] Wang Q, Han Y, Li Q, et al. Segmenting overlapping nanoobjects in atomic force microscopy image [J]. *Journal of Nanophotonics*, 2018, 12 (1): 1–11.
- [65] 何维娜, 张丽丽. 人工神经网络在金相图像分割中的应用研究 [J]. *电子设计工程*, 2013, 21 (3): 143–147.
- [66] Sheng Y, Deng T, Qiu P, et al. Accelerating the Discovery of Cu–Sn–S Thermoelectric Compounds via High-Throughput Synthesis, Characterization, and Machine Learning-Assisted Image Analysis [J]. *Chemistry of Materials*, 2021, 33 (17): 6918–6924.

- [67] Han Y, Lai C, Wang B, et al. Segmentation and Analysis Method for Two-Phase Ceramic (HfB₂-B₄C) Based on the Detection of Virtual Boundaries [J]. *Image Analysis and Stereology*, 2019, 38 (1): 95–105.
- [68] Han Y, Song L, Wang B, et al. AtomicNet: a novel approach to identify the crystal structure of each simulated atom [J]. *Modelling and Simulation in Materials Science and Engineering*, 2020, 28 (3): 35–46.
- [69] Liu Y, Han F, Li F, et al. Inkjet-printed unclonable quantum dot fluorescent anti-counterfeiting labels with artificial intelligence authentication [J]. *Nature Communications*, 2019, 10: 1–9.
- [70] Shi W, Caballero J, Huszár F, et al. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network [C]. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016: 1874–1883.
- [71] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector [C]. In *European Conference on Computer Vision*, 2016: 21–37.
- [72] Roy A, Todorovic S. A Multi-scale CNN for Affordance Segmentation in RGB Images [C]. In *European Conference on Computer Vision*, 2016: 186–201.
- [73] Zhao Y, Han R, Rao Y. A New Feature Pyramid Network for Object Detection [C]. In *2019 International Conference on Virtual Reality and Intelligent Systems (ICVRIS)*, 2019: 428–431.
- [74] 陈科圻, 朱志亮, 邓小明, et al. 多尺度目标检测的深度学习研究综述 [J]. *软件学报*, 2021, 32 (4): 1201–1227.
- [75] Kang L, Ye P, Li Y, et al. Simultaneous estimation of image quality and distortion via multi-task convolutional neural networks [C]. In *2015 IEEE International Conference on Image Processing (ICIP)*, 2015: 2791–2795.
- [76] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the Inception Architecture for Computer Vision [C]. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016: 2818–2826.

- [77] Li Y, Chen Y, Wang N, et al. Scale-Aware Trident Networks for Object Detection [C]. In 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 6053–6062.
- [78] Liu S, Qi L, Qin H, et al. Path Aggregation Network for Instance Segmentation [C]. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 8759–8768.
- [79] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C]. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 1–9.
- [80] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift [J]. *Journal of Machine Learning*, 2015, 2 (4): 1–10.
- [81] Vaswani A, Shazeer N, Parmar N, et al. Attention is All You Need [C]. In Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 6000–6010.
- [82] 任欢, 王旭光. 注意力机制综述 [J]. *计算机应用*, 2021, 41 (S01): 1–6.
- [83] Chu X, Yang W, Ouyang W, et al. Multi-context Attention for Human Pose Estimation [C]. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 5669–5678.
- [84] Chaudhari S, Mithal V, Polatkan G, et al. An Attentive Survey of Attention Models [J], 2021, 12 (5): 1–12.
- [85] Snderby S K, Snderby C K, Maale L, et al. Recurrent Spatial Transformer Networks [J]. *Computer Science*, 2015, 3 (1): 875–878.
- [86] Hu J, Shen L, Albanie S, et al. Squeeze-and-Excitation Networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42 (8): 2011–2023.
- [87] Li X, Wang W, Hu X, et al. Selective Kernel Networks [C]. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 510–519.
- [88] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional Block Attention Module [C]. In European Conference on Computer Vision, 2018: 3–19.

- [89] Jaccard P. Etude comparative de la distribution florale dans une ´ portion des Alpes et des Jura [J]. Bulletin de la Societe Vaudoise des ´ Sciences Naturelles, 1901, 37 (1): 547–579.
- [90] Milletari F, Navab N, Ahmadi S-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation [C]. In 2016 Fourth International Conference on 3D Vision (3DV), 2016: 565–571.
- [91] Achanta R, Shaji A, Smith K, et al. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34 (11): 2274–2282.
- [92] Lai C, Song L, Han Y. Material image segmentation with the machine learning method and complex network method [J]. MRS Advances, 2018, 1 (1): 1–6.
- [93] 王楠, 李一鸣, 张瑞, et al. 基于生成对抗网络的材料组织图像增强算法 [J]. 信息技术, 2021, 45 (12): 1–7.
- [94] Huang G, Liu Z, Van Der Maaten L, et al. Densely Connected Convolutional Networks [C]. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 2261–2269.
- [95] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition [C]. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 770–778.
- [96] Li J, Liang X, Shen S, et al. Scale-Aware Fast R-CNN for Pedestrian Detection [J]. IEEE Transactions on Multimedia, 2018, 20 (4): 985–996.
- [97] Lin T-Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42 (2): 318–327.
- [98] Decost B L, Hecht M, Francis T, et al. UHCSDB: UltraHigh Carbon Steel Micrograph DataBase [J]. Integrating Materials and Manufacturing Innovation, 2017, 6 (8): 1–9.
- [99] Pelt D M, Sethian J A. A mixed-scale dense convolutional neural network for image analysis [J]. Proceedings of the National Academy of Sciences of the United States of America, 2017, 115 (2): 254–259.

- [100] Howard A, Zhu M, Chen B, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications [C]. In IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2810–2816.
- [101] Zhou B, Khosla A, Lapedriza A, et al. Learning Deep Features for Discriminative Localization [C]. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 2921–2929.
- [102] Liu S, Deng W. Very deep convolutional neural network based image classification using small training sample size [C]. In 2015 3rd IAPR Asian Conference on Pattern Recognition, 2015: 730–734.
- [103] Berman M, Triki A R, Blaschko M B. The Lovasz-Softmax Loss: A Tractable Surrogate for the Optimization of the Intersection-Over-Union Measure in Neural Networks [C]. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 4413–4421.

作者在攻读硕士学位期间公开发表的成果

期刊:

1. Recognition and segmentation of complex texture images based on superpixel algorithm and deep learning. Computational Materials Science, Vol. 209, 111398, 2022.(IF:3.300, 中科院 3 区, 导师为一作, 作者为二作)
2. 基于复杂纹理特征融合的材料图像分割. 计算机应用研究.(中文核心, 审稿中, 导师为一作, 作者为二作)

专利:

1. 发明名称: 基于机器学习的交互式图像分割方法; 申请号或专利号: 202110186749.X.(导师排名第一, 作者排名第二)

软著:

1. 软件名称: 基于机器学习的图像分割软件 V2.0; 登记号: 2020SR0686106.(导师排名第一, 作者排名第二)
2. 软件名称: 基于深度学习的材料图像分割软件 V1.0; 登记号: 2021SR1803889.(导师排名第一, 作者排名第二)

作者在攻读硕士学位期间所参与的项目

1. 国家重点研发计划：材料基因工程专用数据库平台建设与示范应用 (科技部) “陶瓷涂层材料数据自动采集、整合与应用” 课题中的子课题 “材料复杂显微结构高精度图像处理方法研究”；编号：2018YFB0704400，日期：2018.07-2022.06。
2. 上海市自然科学基金项目：小样本环境下物体自适应识别方法研究，上海市自然科学基金；编号：20ZR1419000，日期：2020.07-2023.06
3. 国家重点研发计划：材料基因工程关键技术与支撑平台 (科技部) 中 “国家新材料数据库平台建设关键技术研究” 项目中子课题：“材料大数据技术研究”；编号：2020YFB0704500，日期：2020.09.01-2022.8.31
4. 之江实验室科研攻关项目资助；编号：2021PE0AC02。

致谢

本论文的工作是在导师韩越兴副教授的严格指导与亲切关怀下完成的。韩老师治学严谨、学术态度端正、勇攀科研高峰的态度令我敬佩。在工作学习中，韩老师和蔼可亲，积极鼓励我克服学术困难、完成学业，指导我在科研的道路上大步前进，提供新颖的思路的源泉与良好的科研环境。韩老师就像一位科研的领航人，引领我顺利走完科研与学术的旅程，这会是我人生美妙的一段旅程。无论是科研态度、学术能力方面，还是人际沟通与相处方面，都注定令我受益终生，祝愿韩老师家庭幸福美满，学术成就斐然，桃李满天下！

同时，感谢研究组内的陈侨川老师的非常及时帮助，难忘在科研与课题研究中，陈老师的恰到好处的指点迷津令我豁然开朗，启发我探索与发现科研的新领地；难忘在小论文指导中，陈老师的戳到痛处的修改意见，使我的小论文投递之路较为顺利，这些都将是人生的宝贵财富。祝愿陈老师学术成绩硕果累累，在科研的道路上越走越远，越走越宽！

感谢同门张宏坤、魏惠姗、刘宇虹等同学，他们是我学习与生活中的好友，科研道路的伴侣，我们一起携手互助，在学术讨论上中开拓视野，在聚餐中畅谈人生理想，为繁忙的研究生生涯增添一抹亮丽的色彩，这是一个有温度有涵养的大家庭，也祝愿各位师弟师妹、王璐学姐、何心学姐顺利完成学业！

最后，感谢我的父母，在读研期间，他们无条件给予我经济支持与生活关怀，他们给了我一个稳定的港湾，使我全身心地投入学习中。

在此，衷心感谢所有关心与帮助我的人，祝愿你们万事胜意。