

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

基于非完全标注的复杂纹
理图像分割方法研究

作 者 李睿祺

学科专业 计算机应用技术

导 师 韩越兴

完成日期 2023.03.01

姓 名：李睿祺

学号：20721535

论文题目：基于非完全标注的复杂纹理图像分割方法研究

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。

答辩委员会签名:

主任: 李睿祺

委员: 韩进元 孙伟

导 师: 韩进元

答辩日期: 2023.04.09

姓 名：李睿祺

学号：20721535

论文题目：基于非完全标注的复杂纹理图像分割方法研究

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： 李睿祺 日期： 2023.04.09

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： 李睿祺 导师签名： 韩伟光 日期： 2023.04.09

上海大学工学硕士学位论文

基于非完全标注的复杂纹理 图像分割方法研究

姓名：李睿祺
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

2023 年 3 月

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

Research on Segmentation Method of Complex Textured Images based on Incomplete Annotations

MA Candidate: Ruiqi Li

Supervisor: Yuexing Han

Major: Computer Application Technology

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

03, 2023

摘要

材料图像的精确分割是众多材料图像数据分析工作的重要环节。在材料图像中，由于各个组织分布不均、形状各异，这导致材料图像表现出组织粘连、目标交错的情况，不利于完成材料图像的分割。随着机器学习与深度学习技术的快速发展，越来越多的机器学习与深度学习图像分割方法被引入材料图像分割任务中以取得更加精确的分割结果。然而，大部分基于学习方法的材料图像分割工作依赖于完整且精确的标注数据。考虑到材料图像内容语义信息较弱，并且图像内部结构复杂不易分辨，对于材料图像的标注工作受到领域知识的限制且代价高昂。因此，降低材料图像的标注开销并完成材料图像的精确分割是一项重要的任务。本论文针对陶瓷、碳钢、钛合金等多种材料图像的特点，结合机器学习与深度学习技术，以实时获取模型和获得精确分割结果为目标，提出了两种基于非完全标注的复杂纹理材料图像分割方法。

本论文为解决材料图像分割任务中，标注代价与模型精度二者的平衡问题，提出了基于交互式标注的机器学习材料图像分割方法。该方法以实时获取分割模型为目标，首先对图像中每个像素点提取以当前像素点为中心的 9×9 像素点矩形块作为感兴趣区域，分别使用 Gabor 滤波器、Hu 矩、方向梯度直方图、灰度共生矩阵和平均灰度值对感兴趣区域矩形块进行特征提取并用于描述当前像素点的特征；接着对图像中各物相的最显著区域进行划线，以划线标注像素点的特征构建数据集用于训练特征分类模型；最后在每张训练图像上不断迭代优化，交互式标注误分割区域并使用新的标注信息训练模型。该方法的创新贡献包括以下三方面：第一，实现了标注代价与模型精度之间的平衡，允许使用者结合实际场景进行应用；第二，实时获取能够快速迭代的分割模型；第三，设计中心点邻域特征保证分割方法的普适性。

为了进一步提升分割精度，扩展方法的普适性，实现自适应的特征提取，本论文提出了基于伪标签的深度学习材料图像分割方法。该方法提出了伪标签纹理分割网络，解决材料图像分割任务中的小样本、标注困难和算法缺乏普适

性三大问题，获取精确的分割结果。伪标签纹理分割网络的结构分为两部分：主分支和辅助分支，其中主分支用于对材料图像进行特征提取和分割预测，辅助分支用于为训练图像生成伪标签图。通过使用伪标签图对分割任务进行补充监督，实现了基于少量材料图像，训练得到具有泛化能力的图像分割模型，且模型具有普适性。在材料图像分割任务中，相比于其它现有深度学习图像分割算法，该方法实现了根据有限标注数据，得到更为精确的分割结果。

本论文基于多种材料图像进行实验，验证了两种方法在材料图像分割任务中的有效性和普适性。该研究为材料图像数据分析处理工作提供相应的理论保障与技术支持。

关键词：复杂材料图像分割，非完全标注，弱监督学习，机器学习，深度学习

ABSTRACT

Accurate segmentation of complex material images is an important step for many material image data analysis tasks. In material images, due to the uneven distribution and different shapes of individual phases, this causes the material image to exhibit multi-phase adhesion and interlaced microstructure, which hinders the segmentation of material images. With the rapid development of machine learning and deep learning techniques, more and more machine learning and deep learning image segmentation methods have been introduced into material image segmentation tasks to achieve more accurate segmentation results. However, most of the learning based material image segmentation methods rely on complete and accurate annotation data. Considering the weak semantic information and complex texture structure of material images, the annotation of material images is limited by domain knowledge and costly. Therefore, it is an important task to reduce the annotation cost of material images and to accomplish accurate segmentation results of material images. In this dissertation, two incomplete annotation based segmentation methods are proposed using machine learning and deep learning techniques for a variety of complex material images, e.g., ceramic images, carbon steel images, and titanium alloy images, with the goal of acquiring models in real time and obtaining accurate segmentation results.

In this dissertation, an interactive annotation based material image segmentation method is proposed to achieve a balance between the annotation cost and model accuracy in material image segmentation. The method aims to obtain the segmentation model instantaneously. Firstly, for each pixel, a patch with 9×9 pixels centered on the current pixel is extracted as the region of interest, and the Gabor filter, Hu moment, Histogram of Oriented Gradient, Gray Level CO-Occurrence Matrix and average gray value are used to extract features and to characterize the

current pixel. Then, the most significant regions of each phase in the image are annotated, and the features of the scribble annotated pixels are used to build a dataset for training a feature classification model. Finally, iterative optimization is performed on each training image by interactively annotating the mis-segmented regions and training the model with new annotation information. The innovative contributions of this method include the following three aspects: first, this method achieves a balance between annotation cost and model accuracy, allowing users to apply this method in conjunction with real scenes; second, this method can acquire segmentation models in real time and the segmentation models can be rapidly iterated; third, this method designs the center-environment feature to ensure the generalizability of the segmentation method.

To further improve the segmentation accuracy, extend the generalizability of the method, and achieve adaptive feature extraction, this dissertation proposes a pseudo-labeling based deep learning material image segmentation method. This method proposes a pseudo-labeling texture semantic segmentation network to solve the three major problems of material image segmentation, i.e., sample scarcity, annotation difficulty and lack of generalizability of the methods. The network structure can be divided into two parts: the master branch and the auxiliary branch, where the master branch is used for feature extraction and segmentation prediction of material images, and the auxiliary branch is used to generate pseudo labels. By supervising the material image segmentation with the help of pseudo labels, an image segmentation model with generalization capability can be obtained based on a small number of material images, and the model is generalizable. In the task of material image segmentation, this method achieves more accurate segmentation results based on limited annotation data compared to other existing deep learning image segmentation methods.

This dissertation conducts experiments on a variety of material images to verify the effectiveness and generalizability of the two proposed methods in material image

segmentation tasks. This study provides corresponding theoretical guarantees and technical support for material image data analysis and processing work.

Keywords: Complex Material Image Segmentation, Incomplete Annotation, Weakly Supervised Learning, Machine Learning, Deep 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 复杂纹理图像分割算法研究现状	4
1.4.2 基于非完全标注的图像分割算法研究现状	7
1.5 论文主要工作	9
1.6 论文组织结构	10
 第二章 相关理论与技术概述	 11
2.1 非完全标注概述	11
2.2 伪标签概述	13
2.3 弱监督语义分割概述	14
2.4 语义分割评价指标概述	16
2.5 本章小结	17
 第三章 基于交互式标注的机器学习材料图像分割方法	 19
3.1 方法概述	19
3.1.1 区域中心点邻域特征的提取	20
3.1.2 交互划线标注训练	23
3.1.3 迭代标注训练优化	24
3.2 实验分析	25
3.2.1 迭代终止轮次分析	26
3.2.2 相关方法对比	27
3.2.3 方法普适性分析	35

3.2.4 交叉验证实验	38
3.3 本章小结	40
第四章 基于伪标签的深度学习材料图像分割方法	43
4.1 方法概述	43
4.1.1 数据增强	44
4.1.2 主分支	44
4.1.3 辅助分支	46
4.1.4 双分支融和	51
4.2 实验分析	54
4.2.1 消融实验	54
4.2.2 相关方法对比	63
4.3 本章小结	68
第五章 总结与展望	70
5.1 结论	70
5.2 工作展望	71
参考文献	73
作者在攻读硕士学位期间公开发表的论文、专利与软著	85
致谢	87

第一章 绪论

1.1 课题来源

本课题得到国家重点研发计划（编号：2018YFB0704400, 2018YFB0704402, 2020YFB0704503），国家自然科学基金（面上，编号：52273228），上海市自然科学基金项目（编号：20ZR1419000），上海市“科技创新行动计划”启明星项目（扬帆专项）（编号：23YF1412900），之江实验室科研攻关项目资助（编号：2021PE0AC02）。

1.2 课题背景概述

工业领域的技术创新与新型材料的研发有着密不可分的联系。使用新型材料不仅可以解放生产力，还有助于推动人类社会的进步。为了缩短新型材料的研发周期并降低研发过程中的时间与人力成本，美国于 2011 年颁布了“材料基因计划”(Materials Genome Initiative)[1]。相应的，中国也在 2012 年启动了《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》，这标志着材料科学已经上升至我国的国家重大战略。

在研发新型材料的过程中，探索材料宏观性能、微观结构以及制备技术之间的联系是一个关键步骤 [2, 3]，其有助于优化工艺和提升材料的性能。伴随着材料研发工作的推进，相关的研究数据不断积累，这些数据中隐含着丰富的信息。为了更加高效地实现材料研究数据的处理分析工作，计算机技术被广泛应用以完成材料数据统计、材料图像数据挖掘和宏观性能分析等工作 [4, 5]。计算机技术在材料研究领域的引入，有助于提升研发速度、降低研发成本，推动材料产业的发展，加速新材料的研究与应用。

材料图像是典型的复杂纹理图像。在材料图像中，不同组织之间分布不均、形状各异，导致了材料图像表现出组织粘连、目标交错的情况，不利于对图像进行分割处理。然而，在众多材料图像数据分析工作中，对材料图像进行精确分割是一个重要环节。早期的材料图像处理主要依靠人工及传统非学习的计算机图像处理算法

完成分割，这些方法缺乏普适性且精度较差。随着机器学习与深度学习技术的快速发展，使用学习方法进行材料图像的分割可以获得更加精确的分割结果。因此越来越多的机器学习与深度学习图像分割方法被引入材料图像分割任务中。然而，大部分基于学习方法的材料图像分割工作过于依赖完整且精确标注的图像数据。考虑到材料图像内容语义信息较弱，并且图像内部结构复杂不易分辨，材料图像的标注工作受到领域知识的限制且代价高昂。因此，降低材料图像的标注开销并实现材料图像的精确分割是一项重要的任务。本研究尝试结合数字图像处理、机器学习与深度学习技术应用于材料图像微观结构的精确分割，为材料图像数据分析处理工作提供相应的理论保障与技术支持。

1.3 课题研究的目的与意义

伴随着计算机图像处理技术的发展与进步，借助计算机图像处理算法解决材料微观结构的分割问题得到了越来越多的关注。由于材料图像通常具有复杂的纹理，且同一种类的材料图像可能因为制备、拍摄等因素的影响存在一定差异，因此基于非学习方法的图像分割算法难以普遍解决材料图像的分割问题。相较于非学习方法，基于学习方法的图像分割算法具有更好的鲁棒性与普适性，因而逐渐得到了更多的应用 [6, 7, 8]。在目前的材料图像分割工作中主要具有三大难点：小样本、标注困难和算法缺乏普适性。因为上述三大材料图像分割难点的存在，材料图像的分割工作尚存在着一定的困难。

小样本是当前材料图像分割工作中的第一个难点。受限于材料的制备与拍摄条件，材料图像无法像自然场景图像一样可以被批量获取，这导致同一种类的材料图像样本数量非常有限。而大部分基于学习方法的图像分割算法都需要大规模的同一种类图像数据用于进行训练，这为材料图像的分割处理带来了一定的难度。

标注困难是当前材料图像分割工作中的第二个难点。材料图像具有复杂的纹理，其通常表现出组织粘连、目标交错的情况。因此对材料图像进行标注需要具备相关的领域知识且花费大量时间。此外，相较于自然场景图像的公共数据集，现有的材料图像数据集 [9] 往往缺乏适用于解决图像分割问题的标注数据。对于自然场景图

像，在众多经典的图像分割工作中 [10, 11, 12, 13]，机器学习与深度学习图像分割算法实现了基于完全标注的图像数据，解决公共数据集 [14, 15, 16] 的图像分割问题。而最新的研究工作 [17, 18] 已证明，在自然场景图像分割任务中，基于非完全标注进行图像分割可以在显著降低标注代价的同时，获得可观的分割精度。因此如何实现标注代价与分割精度二者之间的平衡已成为图像分割任务中值得关注的问题。在材料图像的分割任务中，由于标注代价非常高昂、标注数据不易获取，材料图像的标注代价与分割精度之间的非线性关系尤为显著。所以，如何以合理的标注代价得到精确的分割结果是材料图像分割工作的难点之一。

算法缺乏普适性是当前材料图像分割工作中的第三个难点。不同种类的材料图像由于其成分、工艺等因素的区别往往存在显著的差异。为了更精确地完成材料图像的分割工作，现今的研究通常根据特定类别材料图像的特点设计算法，这导致算法在解决其它类别材料图像的分割问题时存在失效的情况。此外，材料图像的种类非常丰富，为不同类别的材料图像单独设计算法不仅耗时费力且难以穷尽。因此，算法缺乏普适性这一问题已成为材料图像分割的最大障碍之一。

综合考虑以上因素，基于非学习方法进行材料图像分割和基于学习方法的完全标注材料图像分割都存在着诸多弊端。因此，为了克服材料图像分割工作的三大难点，本课题研究基于学习方法的非完全标注材料图像分割算法。一方面，采用非完全标注有助于解决材料图像难以标注的问题，并进一步探索材料图像分割中标注代价与分割精度之间的平衡；另一方面，采用基于学习方法的图像分割算法可以获得效果良好且更具普适性的模型。这样，本课题研究的内容可以应用于多种类别的材料图像上且降低受到标注数据的限制，在计算机图像处理领域和材料数据挖掘领域具有实际意义和应用价值。

综上，本课题基于材料图像的特点，结合数字图像处理、机器学习与深度学习技术，提出两种基于非完全标注实现材料图像分割任务的算法，为解决材料图像分割中存在的小样本、标注困难以及算法缺乏普适性的问题提供新思路，推动材料图像处理技术的发展与应用。

1.4 国内外研究现状

本论文主要采用数字图像处理、机器学习与深度学习技术，围绕基于非完全标注的复杂纹理图像分割方法进行研究。本章从复杂纹理图像分割与非完全标注图像分割两方面介绍其研究现状，这些已有的研究可以为本论文提供一定的借鉴。

1.4.1 复杂纹理图像分割算法研究现状

复杂纹理图像分割主要是指在材料图像处理、医学图像处理等领域中，对含有复杂纹理而内容语义信息较弱的图像进行分割，例如材料微观结构图像分割和医疗图像病灶分割。本文中主要研究的复杂纹理图像对象为材料微观结构图像。

材料微观结构图像分割根据方法可划分为两类，分别是基于非学习方法的材料图像分割和基于学习方法的材料图像分割。基于非学习方法的材料图像分割一般基于传统图像处理方法，而基于学习方法的材料图像分割又可分为机器学习和深度学习两大类。

基于非学习方法的材料图像分割，包括基于阈值、边缘和区域等方法。直方图双峰阈值法 [19] 和大津法 (OTSU)[20] 是两种常用的基于阈值的图像分割方法。图像的灰度直方图通常具有双峰性，即前景峰值与背景峰值。直方图双峰阈值法通过寻找灰度直方图中两个峰值之间的最小值，确定用于分割图像前景与背景的灰度阈值。大津法是一种通过计算最大类间方差得到分割灰度阈值的算法。当前景与背景之间的类间方差越大，图像中两部分的差别越显著；而当部分前景或背景在图像中被错分时，图像中两部分的类间方差就会变小。大津法被多次应用到材料图像处理中，例如 Talab 等人 [21] 使用大津法和多重滤波技术检测分割出了混凝土图像中的裂缝。基于阈值的分割方法在像素灰度值具有明显规律的情况下表现良好且算法快捷简便易于实现，适用于分割前景与背景区分度较高的图像。但是基于阈值的分割方法未充分考虑图像中像素的关联及高级语义信息，鲁棒性较差，不适用于更加复杂的纹理图像。在近期的研究中，基于阈值的图像分割方法通常被作为一种图像预处理手段进行使用。基于边缘的图像分割方法通过计算图像的一阶、二阶或高阶

导数寻找图像中变化明显的像素，即图像中的边缘，将图像中的边缘进行连接就可以确定图像各部分的轮廓从而完成图像不同区域的分割。常见的边缘检测算子包括 Sobel 算子 [22]、Canny 算子 [23]、Roberts 算子 [24]、Prewitt 算子 [25]、Laplacian 算子 [26] 和 LoG(Laplacian of Gaussian) 算子 [27] 等。在材料图像处理领域，基于边缘的图像分割方法被应用于诸多工作中，例如邓仕超等人 [28] 为了定量分析金相组织的多项组织含量，采用 Canny 算子与灰度等高线结合以产生封闭的金相组织边缘；Han 等人 [29] 为了检测并分割 EBSD 图案菊池带，采用 Radon 变换和累计概率霍夫变换对菊池带的边缘进行直线检测和线段检测，通过结合二者的检测结果完成对菊池带的分割定位。尽管基于边缘的图像分割方法充分考虑了边界信息的特征，对清晰边界的图像进行分割可以获得比较好的效果，但是难以分割边界不明显的图像。此外，基于边缘的图像分割方法可能在处理复杂图像时表现出明显的过分割问题。典型的基于区域的图像分割方法包括区域生长法 [30] 和分水岭算法 [31]。区域生长法首先选取图像中数个像素或小区域作为初始集合，随后不断向集合中添加符合生长准则的邻近像素，最终得到的每个集合包含了组成一个独立区域的全部像素。在将图像中所有的像素点都分配至所属区域后，即完成了对图像的分割。分水岭算法是根据图像中的最小灰度值设置水平面，通过不断升高水平面来提升划分阈值，使得不同区域逐渐被淹没，在每个区域被淹没之前取其分界像素即可得到该区域的分割边界。在材料图像处理领域，Chen 等人 [32] 和 Liu 等人 [33] 使用分水岭的改进算法完成了对金相图像的分割工作。基于区域的图像分割方法计算简单，但是容易受到噪声的影响易出现过分割或欠分割问题。综上所述，基于非学习方法的材料图像分割有着诸多弊端，通常仅用于对图像的预处理。

基于学习方法的材料图像分割可分为机器学习和深度学习两类。机器学习材料图像分割算法借助预先设计好的图像特征，并根据这些特征的分布建立图像处理模型，对特征数据的内在关系进行归纳与总结。常见的机器学习算法包括聚类法 [34]、支持向量机 (SVM, support vector machine) 法 [35] 等。基于像素灰度值等特征使用模糊聚类算法 (FCM, Fuzzy C-Means)[36]，将图像中的像素划分为多个类别是一种典型的机器学习聚类图像分割算法，其在迭代的过程中令划分到同一特征簇的对

象之间相似度最大，而不同特征簇的对象之间相似度最小，最终得到理想的图像分割结果。Chen 等人 [37] 在金相图像中利用空间信息使用改良的二维模糊聚类算法完成了高速且准确的球墨铸铁分割工作。支持向量机基于图像的像素特征进行划分工作，根据训练的像素特征选取不同的核函数可以构建合适的划分超平面将像素特征分为多个类别，最终得到图像的分割结果。Li 等人 [38] 基于支持向量机使用在线学习方法解决了金相图像的分割问题。此外，马尔可夫随机场 [39]、随机森林 [40]、高斯混合模型 [41] 等机器学习算法也可以用于完成图像分割工作。Lai 等人 [42] 使用四种传统图像特征描述子，利用随机森林完成了对钛合金图像的分割。Han 等人 [43] 结合多种传统图像特征描述子，使用梯度提升决策树对特征进行分类，完成多种类别材料图像的分割工作。机器学习算法相较于非学习方法，可以更充分地挖掘像素之间的关联和找到不同像素的分布，通常可以取得较为良好的分割结果。然而机器学习算法需要采用预先设计的图像特征，并非根据实际场景中待分割图像的特点进行设计，因此最终的分割效果在很大程度上依赖于所采用的图像特征是否合适。

相较于机器学习算法，深度学习算法可以根据实际场景中待分割图像的特点自适应地完成特征提取工作，无需对图像特征进行预先设计。这使得深度学习算法提取出的图像特征描述更加准确且具有针对性，从而有利于获取更加精确的分割结果。根据深度学习神经网络应用对象的不同，解决图像分割任务的深度学习神经网络可以大致分为两类，一是针对自然场景图像进行设计的神经网络，包括 FCN[10]、DeepLabv3+[44] 等；另一类是针对复杂纹理图像进行设计的，包括 Unet[45]、Unet++[46]、Unet3+[47] 等。随着深度学习分割算法的不断提出与改进，越来越多的材料图像分割工作开始尝试使用深度学习神经网络进行实现。在近期的研究中，Azimi 等人 [48] 使用 FCN 和最大投票方案对低碳钢微观结构图像进行分割，其分割精度相较于过往方法在准确率上提升了约 45%；Pan 等人 [7] 基于 DeepLabv3+ 开发了一个双注意力深度学习网络 DAN-DeepLabv3+，该网络可以在图像中对三种类型的钢铁缺陷进行分类与定位，表现出了良好的效果；Yang 等人 [8] 通过引入带有注意力门的改进 Unet++ 和用于进一步分类的残差网络 (ResNet) 作为两级结构，解决了光学显微镜下单层石墨烯图像的分割与识别问题；李原等人 [49] 就钢铁表面

缺陷图像的分割问题提出了一种改进的 Res-Unet 网络，取得了比经典 Unet 网络更好的分割效果；Han 等人 [50] 设计了基于超像素与深度学习相结合的复杂纹理图像分割算法实现了对碳钢图像、陶瓷图像和木材表面图像的分割。

由于深度学习分割算法无需人为选定特征且具有强大的分割性能，因此在近些年的研究中，深度学习分割算法逐渐在复杂纹理图像的分割问题上占据了主导地位。但是目前的大多数深度学习神经网络具有一定的局限性，例如很多网络无法良好地迁移并用于解决其它类型图像的分割，且网络的训练依赖于庞大且完全标注的数据集。因此如何更好地将深度学习分割算法应用到复杂纹理图像分割问题上仍然是一个急需解决的现实问题。

1.4.2 基于非完全标注的图像分割算法研究现状

近年来，随着深度学习技术的发展，陆续诞生了大量的深度学习神经网络，这些网络在图像分类、分割等任务中表现出了良好的效果。以图像分割任务为例，在 PASCAL VOC 数据集 [14] 上，FCN[10]、PSPNet[12]、DANet[51]、DeepLabv3+[44] 和 SANet[52] 等网络将测试集上的平均交并比 (MIoU, Mean IoU) 从 62.2% 提升至 89.0%。相较于使用完全标注的全监督图像分割任务，使用非完全标注的半监督或弱监督图像分割任务具有更深的现实意义。一方面，在现实生活中，大量的实际任务并不具备大规模完全标注的图像数据，因此较之于全监督图像分割网络，半监督或弱监督图像分割网络具有更高的实际应用价值。另一方面，标注量与分割精度之间并非线性相关，即使增加一倍的标注数据，也无法保证分割精度能够成倍提升，这使得标注工作具有明显的边际效应。近年来，越来越多的研究着眼于基于非完全标注获得高精度的图像分割结果。

在图像分割任务中，半监督学习使用大量未标注图像，同时使用标注图像进行学习。在最新的半监督学习方法中，其核心思想是基于一致性正则化原则为未标注图像构建监督。一致性正则化原则的实现形式之一是在训练过程中令模型对图像变换前后作出相同的预测结果。French 等人 [53] 对 CutMix 增强技术进行改进，提出了一种更强大的遮挡变体 CowMix，通过网络对原始图像和经过 CowMix 处理的图

像作出相同的预测，实现使用未标注的图像进行训练。Kim 等人 [54] 使用教师学生网络和 CutMix 相结合，在训练中令教师网络和学生网络对经过 CutMix 增强的训练图像作出相同的预测，从而实现以结构一致性损失对网络进行训练。Ouali 等人 [55] 令网络的解码器对编码器输出的不同扰动特征作出相同的预测，改善编码器的特征表示，通过扰动的特征嵌入利用一致性正则化原则对网络进行训练。除了基于一致性正则化原则进行半监督学习，近期的研究还提出以迭代的自训练方式实现半监督学习。Chen 等人 [56] 基于可用标注数据足以训练得到一个良好教师模型的假设，提出了 Naive-Student 模型预测未标注图像的伪标签并和标注图像一起进行迭代式的训练。

在弱监督学习中，图像数据的标注允许是不完全的，例如仅对部分像素进行标注或以较弱的语义进行标注。相较于半监督学习，弱监督学习不具备任何完全标注的训练数据，其标注量通常较之于半监督学习更少，因而更加具有挑战性。大部分近期的弱监督学习图像分割方法采用图像级标签作为监督信息，利用类激活图 (CAM, class activation map)[57] 的思想进行学习。由于类激活图仅关注并获取图像中物体最有区别的区域，因此产生了不同的方法和思路对类激活图结果进行细化和改进。Hou 等人 [58] 和 Wei 等人 [59] 采用对部分图像或特征进行对抗擦除的思想改进类激活图获得的目标区域范围，实现获得更加完整的目标区域。Oh 等人 [60]、Huang 等人 [61] 使用额外的显著性估计模型对类激活图可获取的区域范围进行改进。由于上述基于图像级标签使用类激活图思想进行弱监督学习的研究大多依赖全连接条件随机场 [62]，因此训练过程通常包含了多个阶段。除了采用图像级的标签作为监督信息，部分研究工作采用划线标注作为监督信息进行弱监督学习。相较于图像级标签，划线标注具有更为直观的空间位置和范围信息且标注代价依旧较低，因此在非完全标注形式中，划线标注通常可以获得最高的分割精度。Lin 等人 [63] 为了解决像素级标注所带来的巨大开销，提出了根据划线标注训练网络获取图像分割模型，其将划线标注的像素信息传播到未标注的像素以用于更好地学习网络参数。Ke 等人 [64] 将基于划线标注的弱监督语义分割任务表述为一个半监督的度量学习问题，通过将具有相同或不同语义的像素分别映射至相同或不同的特征，并在特征空间中根

据像素与物体的对应关系训练网络，最终完成图像的分割。此外，划线标注还常被应用于人机交互式的标注中，例如在 Graphcut[65] 和 GrabCut[66] 算法中，为了便捷地对分割结果作出反馈，二者均采用了划线标注的形式标定误分割的区域。

综合考虑各种非完全标注的标注代价和所能获得的分割精度，划线标注是一种较为均衡的非完全标注形式。划线标注在能够获得接近完全标注分割效果的同时，可以显著降低标注代价的开销。因此，如何进一步改进基于划线标注的弱监督学习图像分割并用于解决现实问题是一个值得研究的方向，这对受到标注代价限制的领域有着重要的现实意义与实用价值。

1.5 论文主要工作

为了解决复杂纹理图像中的材料图像的分割问题，本文针对材料图像分割工作的三大难点，即小样本、标注困难和算法缺乏普适性，结合机器学习与深度学习，提出基于非完全标注的分割方法。本文的主要工作和创新如下：

(1) 为了解决材料图像分割任务中标注代价与分割精度二者之间的平衡问题，实时获得分割模型，本工作设计了基于交互式划线标注与机器学习的材料图像分割算法。本方法以实时获取分割模型为首要目标，提取材料图像的中心点邻域特征并进行多轮交互式划线标注，采用增量学习方法训练得到最终的图像分割模型。

(2) 为了实现根据有限且易获取的标注数据，获得高精度的分割模型，本工作设计了基于伪标签的弱监督深度学习材料图像分割方法。本方法以基于有限标注数据获取最佳分割结果为目标，设计了新的双分支网络结构，提出新的上下文特征差异监督损失用于生成伪标签，使得模型可以仅根据划线标注的训练图像获得较高的分割准确率。本方法解决了深度学习神经网络在材料图像分割任务中面临的小样本问题、标注困难问题和测试过程中的资源浪费问题。

1.6 论文组织结构

本论文以本人攻读硕士学位期间承担的相关项目研究为基础，针对材料图像分割工作现存的三大难点，即小样本、标注困难和算法缺乏普适性，采用并发展机器学习方法和深度学习方法，与非完全标注方式相结合，研究和设计了相应的基于非完全标注的材料图像分割方法，并通过实验验证了本论文提出方法的有效性。

本论文的其它各章节内容安排如下：

第二章介绍了与本论文研究相关的理论和技术。首先，介绍了非完全标注的不同实现形式与应用领域，并分析了各个种类非完全标注的特点；然后，对伪标签的原理、应用和发展进行介绍；其次，介绍了弱监督语义分割的相关知识，对弱监督语义分割的算法思想进行了阐述；最后，对本论文中所使用的相关语义分割评价指标进行了描述。

第三章提出了一种基于交互式标注的机器学习材料图像分割方法。围绕材料图像分割任务中标注代价与分割精度之间的矛盾，以实时获取分割模型为目标，设计了基于交互式划线标注的机器学习材料图像分割方法。通过与其它方法进行对比并测试分割多种不同类别的材料图像，验证了方法的有效性和普适性。

第四章提出了一种基于伪标签的深度学习材料图像分割方法。针对第三章的方法在分割精度上存在的缺陷，以获取高精度材料图像分割模型为目标，设计了基于伪标签的深度学习材料图像分割方法。通过与其它深度学习方法在多种类别的材料图像上进行对比，并进行消融实验，验证了方法的有效性与普适性。

第五章对全文进行总结与展望。讨论并分析本论文研究成果的优势与不足，并对未来的研究方向提出展望与设想。

第二章 相关理论与技术概述

图像分割是计算机视觉领域的一个重要研究方向。伴随着机器学习与深度学习技术的成熟与进步，图像分割已被广泛应用于各个领域。材料图像分析是图像分割技术在材料学研究中的应用之一。借助于计算机图像分割技术，材料图像中的物相可以被更加精确地获取，这有助于提升材料图像分析结果的准确性。因此，图像分割技术与材料研究之间的关联正变得愈发紧密。本论文的研究目的是针对材料图像分割工作中存在的三大难点，即小样本、标注困难和算法缺乏普适性，以及机器学习和深度学习分割算法中标注代价和分割精度之间的矛盾，进一步探索适用于复杂纹理图像的非完全标注弱监督图像分割方法。因此，本章节回顾了与本论文相关的理论和技术，为后续章节的方法介绍提供铺垫。

2.1 非完全标注概述

在图像分割任务中，大部分机器学习和深度学习方法基于完全标注的数据集训练得到图像分割模型。然而，获取训练数据集的完全标注可能存在以下两方面问题。一方面，对于复杂纹理图像和大规模数据集，进行完全标注可能会受到领域知识、时间和人力等因素的限制。另一方面，标注代价与分割精度之间并非线性相关，且标注工作具有明显的边际效应，即当标注数据达到一定规模后，继续增加标注数据可能仅会带来微小的提升。为了解决上述问题，近期的研究尝试采用非完全标注数据进行训练。非完全标注可以根据形式划分为两类：第一类是对数据集中的部分训练图像进行完全标注，而不对其余训练图像进行标注；第二类是仅对训练图像中的部分像素进行标注。由于第一类非完全标注的实现与完全标注较为类似，以下主要针对第二类非完全标注的不同实现形式进行介绍。

图像类别标注通过记录图像中出现的物体或区域的类别完成标注工作。在采用图像类别对图像进行标注时，标注工作不需要关注物体或区域出现的位置。因此使用图像类别标注的训练图像仅包含了基本的类别信息，而要想获取物体或区域在图像中的位置信息则需要由模型从大量训练图像中学习得出。图像类别标注主要用于

图像分类以及图像分割等任务。图像分类任务的目标是对图像的类别进行预测，因此在图像分类任务中图像类别标注是全监督学习，经典的图像分类任务包括 ImageNet 分类 [67] 和 CIFAR 分类 [68] 等。而图像分割任务的目标是对图像中的每个像素点的类别进行预测，因此图像类别标注对于图像分割任务是弱监督学习，经典的图像分割任务包括 PASCAL VOC 图像分割 [14] 和 Cityscapes 图像分割 [15] 等。包围盒标注采用数个矩形标记图像中的物体或区域。相较于图像类别标注，包围盒标注除了包含类别信息，还对图像中的物体或区域所在的粗略位置和大致覆盖范围进行了描述。与图像类别标注完全依赖模型学习得到位置信息不同，包围盒标注由于本身涵盖了一些位置和范围信息，因此模型在训练过程中可以得到更明确的指导。在一些目标检测方法，如 YOLOv5[69] 和 Faster R-CNN[70] 等工作中，包围盒标注是最主流的标注方式。点标注通过对图像中不同物体或区域的中心/内部像素点进行标记完成标注工作。通过标注中心/内部像素点的类别，点标注为训练图像赋予了类别信息与空间位置信息 [71]。前景比例标注是一种应用较少的非完全标注形式。前景比例标注通过对图像中前景物体或区域占图像全部像素的比例进行描述完成标注工作。由于标注信息仅为前景的比例，因此该标注方式通常仅用于分割图像中某一特定类别的前景。基于前景比例标注，Zhang 等人 [72] 为了解决病理图像的组织/区域分割问题提出了 FGNet，在 H&E[73]、HER2[74] 和 KI67[75] 数据集上取得了良好的分割效果。不过前景比例标注方式仍存在着一些问题，由于该标注方式需要相对精确的前景比例，这需要将图像中对应类别的像素标注完毕后才能计算出精确数值，而这一过程仍是相对耗时的，所以在目前的研究中，基于前景比例标注设计的算法相对较少。

目前更引人注目的非完全标注方式是划线标注，相较于点标注，划线标注是对图像中的各个物体或区域采用画曲线的方式来标注像素点。由于曲线比点可以覆盖更多的像素，使用划线标注不仅可以描述训练图像中各个物体或区域的类别信息与空间位置信息，还可以大致描述物体或区域的覆盖范围。因此，划线标注的训练图像中包含了更多的有效信息，也有助于获得更加良好的分割效果。划线标注与点标注在描述方式和包含的信息上有着许多的相似之处，主要被用于解决图像分割问题。

PASCAL VOC[14] 和 DensePose[76] 等常用公共数据集均具有点标注或划线标注的数据用于研究弱监督图像分割问题。图2.1展示了在一幅图像上分别使用图像类别标注、包围盒标注、点标注和划线标注四种方式进行标注的实例。

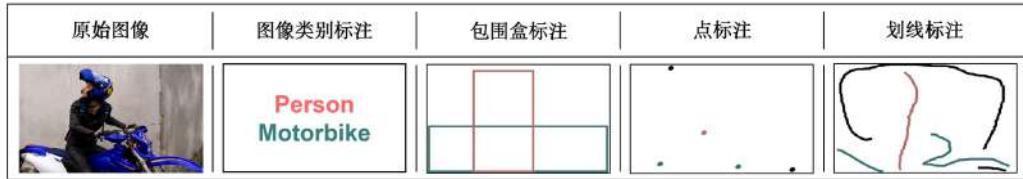


图 2.1: 图像类别标注、包围盒标注、点标注和划线标注示意图 [64]

综合考虑上述五种非完全标注方式，图像类别标注所需花费的标注代价最小，而划线标注能够取得的分割效果最佳。因此，出于最大程度节省标注代价和最大程度避免降低分割精度的考虑，采用图像类别标注或划线标注是现今研究中最主流的方案之一。

2.2 伪标签概述

伪标签概念源于半监督学习，其核心思想是借助无标签数据对基于有标签数据训练得出的模型进行性能上的改进。伪标签之所以可以令模型性能获得提升，主要是因为大部分情况下数据样本的分布存在一定的规律。在划分类别的样本空间中，不同类别样本的分类边界通常处于样本分布的低密度区域，因此在样本空间中，假设同一高密度区域中的数据样本对应于相同的标签具有合理性。这样，基于该假设规则对无标签数据赋予的伪标签是较为可靠的，而对应的图像处理模型性能也能够获得大幅提升。

不过使用伪标签监督训练模型仍可能存在一些问题。一方面，由于伪标签并不一定与真实标签的类别相符，即便样本的分布满足假设条件，仍然有可能产生类别错误的伪标签，这会导致模型在监督训练的过程中存在噪声的干扰，进而误导模型的学习。另一方面，如果为了保证伪标签的正确性，仅对少数与标注数据极为接近的无标签数据赋予伪标签，那么使用伪标签标注的数据可能无法为模型提供额外有

效的监督信息，从而导致模型仅对已知信息进行学习引发过拟合的问题。综上所述，为了实现模型性能的提升，伪标签的正确性和差异性均至关重要。

伪标签的使用通常可以描述为以下五个步骤：首先使用有标签数据对模型进行训练，其次使用训练得到的模型对无标签的数据进行预测，接着从无标签数据的预测结果中筛选出高置信度的结果作为伪标签，然后使用有标签数据和伪标签数据训练得到一个新的模型，最后不断使用新训练得到的模型重复第二到第四步，直至预测结果趋于稳定 [77]。图2.2展示了伪标签的应用流程图。

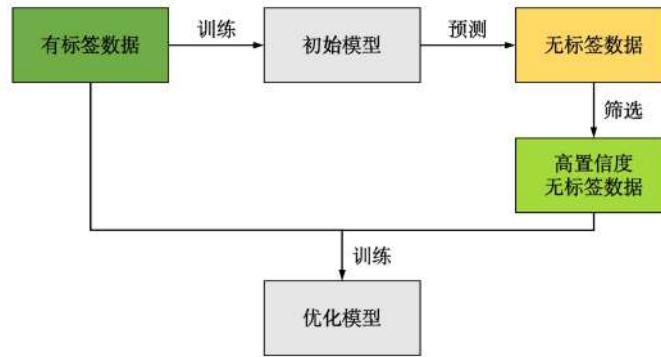


图 2.2: 伪标签应用流程示意图

伪标签的思想和技术从诞生至今已被应用于多种任务并取得了显著的效果。在早期研究中，伪标签主要用于解决图像分类工作。Lee 等人 [77] 应用伪标签，在 MNIST 手写数据集 [78] 上以极少的标注数据取得了良好的分类效果。近期研究中，MixMatch[79] 以及 FixMatch[80] 等方法进一步发展了伪标签的思想，提升了半监督图像分类问题的效果。此外，在部分图像分割的研究中 [81, 61]，伪标签思想也被引入以解决半监督或弱监督图像分割问题。迄今为止，伪标签已成为半监督学习及弱监督学习中最重要的思想之一。

2.3 弱监督语义分割概述

语义分割是指对图像中每一个像素点进行分类，确定每个点的类别，从而进行区域划分。在早期的语义分割工作中，为了获得更加精确的分割模型，通常会采用完全标注的图像数据对模型进行训练。伴随着深度学习神经网络的不断发展，在部

分公共数据集上 [14, 15]，全监督语义分割算法已经可以取得非常精细的分割结果 [44, 52]。为了更好地研究和评价新型语义分割模型的效果，最新的全监督语义分割算法需要使用更大规模、更全面的公共数据集 [82] 进行验证，这表明全监督语义分割模型的改进渐渐步入了瓶颈期。尽管在研究中全监督语义分割模型所取得的巨大进步令人倍感兴奋，但是在实际应用的过程中，全监督语义分割模型往往由于当前任务标注数据的不足，而无法取得理想的分割效果。为了克服全监督语义分割模型的缺点，令语义分割算法具有更高的可应用性，弱监督语义分割是一个值得期待的研究方向。弱监督语义分割是一种特殊的语义分割任务。相较于常规的语义分割，弱监督语义分割基于非完全标注的图像对模型进行训练，这可以在一定程度上减少标注数据的规模且不会显著降低分割效果。之所以会产生这样的现象是因为在语义分割任务中，标注工作具有明显的边际效应。最近的研究已经表明，基于少量稀疏的非完全标注，弱监督语义分割算法可以达到全监督分割结果精度的 98%[64]。

在弱监督语义分割中，根据非完全标注形式的不同，算法思路也存在着一定的差异。在基于图像类别标注进行语义分割时，通常需要依赖于类激活图 (CAM, class activation map)[57]。类激活图可以从模型中获取对应类别在图像中最有区别的区域，如图2.3中红色区域所示。基于图像类别标注，Li 等人 [83] 在 PASCAL VOC 数据集 [14] 上提出了 GAIN 网络，借助卷积神经网络的反向传播获得给定输入的注意力图，作为物体定位和语义分割的先验，实现了对自然场景图像的分割。基于包围盒标注进行语义分割主要通过循环迭代生成预测区域和训练网络学习完成。Dai 等人 [84] 基于这种思想在 PASCAL VOC 2012[14] 和 PASCAL-CONTEXT[85] 数据集上，使用包围盒标注进行监督获得了良好的分割效果。基于点标注和划线标注进行语义分割时，由于已具备精确到像素级别的标注信息，算法实现的思想较为丰富，包括基于超像素的结果进行合并 [63] 和在特征空间预测像素类别 [64] 等。不过究其根本，基于点标注和划线标注进行语义分割的过程，是在将既有的像素级标注信息向未标注的像素进行扩散，从而确定所有像素的类别最终完成分割。

对复杂纹理图像而言，大规模获取像素级标注的图像数据是难以实现的，因此在处理复杂纹理图像时，研究者们也逐渐尝试采用弱监督学习的方式解决分割问题。



图 2.3: 类激活图效果示意图 [57]

Hu 等人 [86] 在工业图像表面缺陷分割任务中,为了减轻对标注代价昂贵的像素级标签的依赖,使用大量以图像类别标注的图像数据和少量完全标注的图像数据相结合进行学习,减少了 37% 的标注代价开销。Dong 等人 [87] 为了从建筑材料图像中提取裂缝信息,提出了一种基于补丁块的弱监督语义分割网络,以图像级标注作为监督信息实现裂缝检测,将标注的工作量减少了约 80%。上述研究不仅证明了弱监督学习对复杂纹理图像分割的有效性,还展现了基于非完全标注进行弱监督图像分割时所具备的标注代价优势。

2.4 语义分割评价指标概述

为了评价二分类模型的效果,通常使用如表2.1的混淆矩阵对模型预测效果进行描述,其中 N 和 P 分别表示实际结果和预测结果,而 T 和 F 分别表示预测正确和预测错误。通过组合 N 、 P 、 T 和 F 可以得到四种情况,真正例 (TP)、真反例 (TN)、假反例 (FN) 和假正例 (FP)。

表 2.1: 二分类问题混淆矩阵

	预测为真	预测为假
标签为真	TP	FN
标签为假	FP	TN

在图像分割任务中,由于像素点的预测类别可能不止两类,因此为了评价分割效果,需要使用多分类的混淆矩阵进行描述。对于三分类问题总共包含 9 种情况,三个类别分别以 A 、 B 和 C 表示,表2.2展示了三分类问题对类别 A 的混淆矩阵。

表 2.2: 三分类问题对类别 A 的混淆矩阵

	预测为 A	预测为 B	预测为 C
标签为 A	TP_A	FN_A	FN_A
标签为 B	FP_A	TN_A	TN_A
标签为 C	FP_A	TN_A	TN_A

基于以上混淆矩阵, 常用的语义分割评价指标有准确率 (Accuracy)、平均交并比 (mIoU) 和平均 Dice 系数 (mDice), 上述三个评价指标可以反映预测分割结果与真实分割结果之间的差异性, 取值范围均为 [0, 1]。图像分割结果的准确率 (Accuracy) 由预测类别正确的像素数量和全部像素数量的比值计算得出, 其计算公式如下:

$$Accuracy = \frac{TP_A + TP_B + TP_C}{TP_A + FP_A + FN_A + TN_A} \quad (2.1)$$

图像分割结果的平均交并比 (mIoU) 由每类像素分割结果的交并比 (IoU) 取平均值得到。对于第 i 类像素的交并比 (IoU) 定义为:

$$IoU_i = \frac{TP_i}{TP_i + FP_i + FN_i} \quad (2.2)$$

平均交并比 (mIoU) 为全部 k 个类别像素交并比 (IoU) 的平均值:

$$mIoU = \frac{\sum_{i=1}^k IoU_i}{k} \quad (2.3)$$

图像分割结果的平均 Dice 系数 (mDice) 由每类像素分割结果的 Dice 系数取平均值得到。对于第 i 类像素的 Dice 系数定义为:

$$Dice_i = \frac{2TP_i}{2TP_i + FP_i + FN_i} \quad (2.4)$$

平均 Dice 系数 (mDice) 为全部 k 个类别像素 Dice 系数的平均值:

$$mDice = \frac{\sum_{i=1}^k Dice_i}{k} \quad (2.5)$$

2.5 本章小结

本章首先介绍了非完全标注的不同形式与应用, 并对各个类别的非完全标注进行了分析。接着, 对伪标签的原理、应用和发展进行了说明, 并展示了伪标签的应

用流程。然后，介绍了弱监督语义分割的相关知识，对弱监督语义分割的算法思想进行了阐述。最后，对本论文中所使用的相关语义分割评价指标进行了描述。

第三章 基于交互式标注的机器学习材料图像分割方法

材料图像的分割是进行相关统计分析工作的必要前提。在面对复杂多样的材料图像分割任务时，由使用者根据实际应用场景进行标注和训练更有利于获得理想的分割结果，然而类似的研究较为匮乏。本章针对材料图像小样本与标注困难的特点，以快速且便捷地获取任一种类的材料图像物相分割模型为目标，提出了一种基于交互式划线标注的机器学习材料图像分割方法。该方法分为三个步骤：首先提取材料图像的中心点邻域特征；然后对材料图像中的物相使用曲线进行标注并构建特征分类器学习标注信息；最后对训练图像交替进行划线标注和模型训练直至迭代终止。训练好的模型可用于对同一种类的其它材料图像进行物相分割与识别。图3.1展示了本方法的流程图。

3.1 方法概述

本方法基于交互式划线标注结合机器学习完成对材料图像中各个物相的分割，其步骤可分为区域中心点邻域特征的提取、交互划线标注训练和迭代标注训练优化三个部分。首先，对图像中每个像素点提取以当前像素点为中心的 9×9 像素点矩形块作为感兴趣区域 (ROI, region of interest)，分别使用 Gabor 滤波器 [88]、Hu 矩 [89]、方向梯度直方图 (HOG, Histogram of Oriented Gradient) [90]、灰度共生矩阵 (GLCM, Gray Level CO-Occurrence Matrix) [91] 和平均灰度值对 ROI 矩形块进行特征提取并用于描述当前像素点的特征。然后，为了降低标注代价，仅对图像中各物相的最显著区域进行划线，以划线标注像素点的特征构建数据集用于训练特征分类模型，并基于该特征分类模型获取未标注像素点的类别从而实现对当前图像的分割。最后，为了获得足够良好的图像分割模型并平衡标注代价与分割精度二者之间的关系，本方法在训练图像上不断重复标注误分割区域并使用新的标注信息训练模型，直至标注代价的增长超过模型精度的提升时终止迭代。上述过程如图3.1所示。

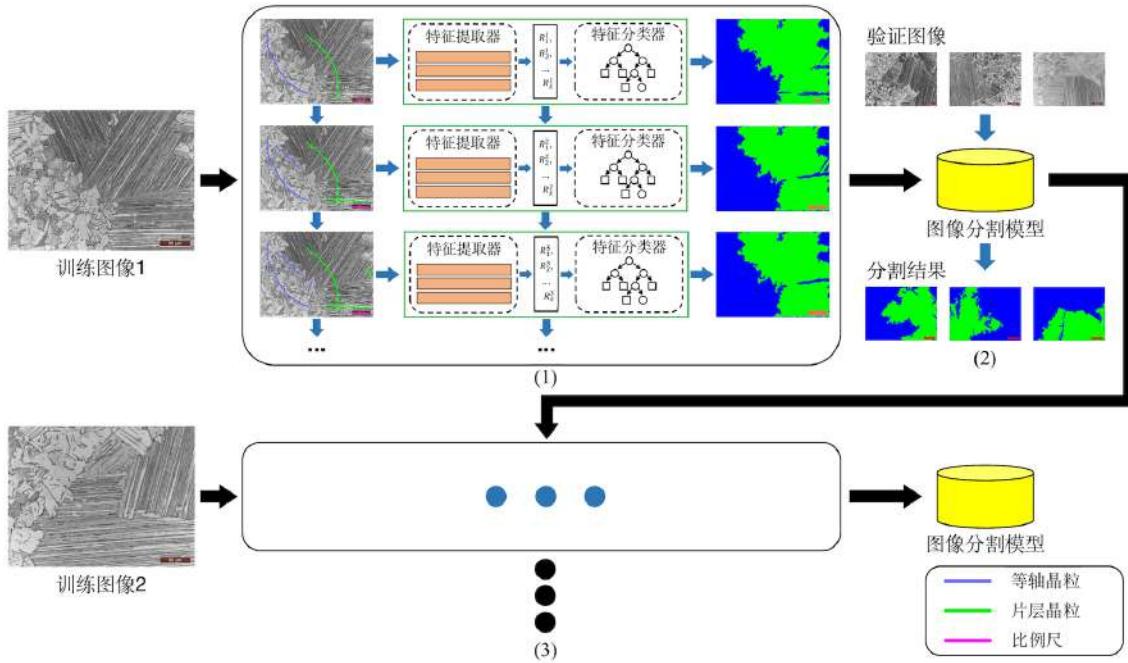


图 3.1: 方法流程图。(1) 对一张图片的训练过程, $R_1^i, R_2^i, \dots, R_k^i$ 表示第 i 轮划线标注的像素提取得到的特征向量; (2) 使用训练得到的图像分割模型对同一数据集上的其它图像进行分割, 以验证模型效果; (3) 当模型效果不佳时, 可以使用其它图像继续标注训练从而得到更好的图像分割模型。

3.1.1 区域中心点邻域特征的提取

为了精确、便捷且实时地完成对材料图像的分割, 机器学习是一种理想的实现方式。不同于深度学习, 使用机器学习进行图像分割需要预先设计好的图像特征, 并且分割结果的准确性高度依赖于选取的图像特征, 因此特征提取是机器学习图像分割过程中的重要步骤之一。由于图像分割任务的本质是对图像中的所有像素点进行分类, 因此使用机器学习需要对图像中的每个像素点提取特征。在材料图像中, 每个像素点的特征不仅取决于其自身的数值信息, 如颜色或灰度值, 还受到周围其它像素点的影响。因此即便两个像素点具有相同的数值信息, 两者的特征也可能存在差异。为了使得本步骤中所提取的特征涵盖自身像素点周围环境的信息, 选取以当前像素点为中心的 9×9 ROI 矩形用于提取当前像素点的环境特征信息, 如图3.2所示。由于在提取像素的环境特征信息时需要获取以每个像素点为中心的 9×9 ROI 矩形, 而在图像边缘的像素点无法直接从原始图像中获取其对应的 ROI 矩形。因此为了避免边

缘像素点在提取环境特征时出现问题，需要在提取特征前对原始图像进行边缘扩充。在本步骤中，原始图像的每侧边缘均被扩充 4 行或列，扩充像素点的数值以零进行填充。由于材料图像纹理复杂且缺乏统一的规律，仅采用单独某一种特征描述算子无法对像素点的特征进行良好的表示，因此本步骤提取的特征由五部分组成 [42]，分别为 Gabor 滤波器 [88]、Hu 矩 [89]、方向梯度直方图 (HOG, Histogram of Oriented Gradient)[90]、灰度共生矩阵 (GLCM, Gray Level CO-Occurrence Matrix)[91] 和平均灰度值。

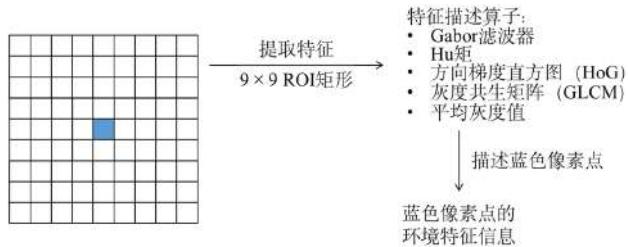


图 3.2: 像素点环境特征信息提取示意图

Gabor 滤波器是一种用于提取边缘信息的线性特征提取器，其提取的 Gabor 特征是一种可用于描述图像纹理信息的特征。Gabor 滤波器能够获得图像在不同频率下的响应，根据图像在不同频率下的响应，可以实现获取图像的特征。由于二维的 Gabor 滤波器可以很好地反映二维图像上的边界信息，因此为了得到更加详细的信息，采用六个方向 ($0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ$) 的二维 Gabor 滤波器提取原始图像不同方向上的纹理信息，从而得到每个像素点的六维特征向量。Gabor 滤波器对光照变化具有良好的适应性，它可以在一定程度上忽略图像旋转和变形所带来的影响，对姿势和光照具有鲁棒性。因此 Gabor 滤波器被广泛应用于人脸识别等工作的预处理中，被证明是一种有效的特征提取器。

Hu 矩是基于形状的特征描述算子，可以反映图像的轮廓信息。在形式上，Hu 矩是归一化矩的线性组合，具有平移不变性、旋转不变性和尺度不变性。Hu 矩中的七个不变矩 ($\Phi_1, \Phi_2, \dots, \Phi_7$) 可以利用二阶和三阶归一化中心矩计算得到，计算公式如下：

$$\Phi_1 = \eta_{20} + \eta_{02} \quad (3.1)$$

$$\Phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \quad (3.2)$$

$$\Phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \quad (3.3)$$

$$\Phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \quad (3.4)$$

$$\begin{aligned} \Phi_5 = & (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ & + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned} \quad (3.5)$$

$$\Phi_6 = (\eta_{20} - \eta_{02}) [(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (3.6)$$

$$\begin{aligned} \Phi_7 = & (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ & - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03}) [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned} \quad (3.7)$$

其中 η_{20} 、 η_{11} 和 η_{02} 表示归一化的二阶中心矩， η_{30} 、 η_{21} 、 η_{12} 、 η_{03} 表示归一化的三阶中心矩。在本步骤的特征提取过程中，七个不变矩以七维特征向量的形式被用于描述像素点的环境特征信息。

方向梯度直方图 (HOG) 最早是一种用于解决人体目标检测的图像描述子，通过对像素的梯度方向进行描述，反映了图像的局部梯度方向和梯度强度分布。方向梯度直方图的本质思想是采用边缘梯度的方向密度分布对局部目标形状进行描述。在本步骤的特征提取过程中，梯度方向被划分为九个区间，即每 20 度为一个区间。因此，方向梯度直方图对每个像素所提取的特征，记录了像素点 ROI 矩形内九个区间的梯度方向数量，该九维特征向量被用于描述像素点的环境特征信息。

灰度共生矩阵 (GLCM) 反映了共同出现像素的分布情况，它可以描述图像灰度值关于方向、间隔距离和变化幅度方面的特征。在本步骤提取像素点 ROI 矩形的 GLCM 特征向量时，首先将灰度值 (0-256) 均匀分为 8 个区间，分别以水平、垂直、斜 45 度角和斜 135 度角四个方向统计 ROI 矩形内的灰度值共生信息。由于此时得到的灰度值共生信息为四个 8×8 的矩阵（分别对应四个方向），其形式不适合直接用于特征描述。因此在特征提取过程中，基于每个矩阵进一步计算出对比度、信息熵、能量和相关度四个统计特征用于更好地描述特征。最后，计算出每个统计特征在四个方向上的方差，即可得到用于描述像素点 ROI 矩形的四维 GLCM 特征。

在对每个像素点提取其对应 ROI 矩形的特征后，图像中的每个像素点可以由六维 Gabor 特征向量、七维 Hu 矩特征向量、九维 HOG 特征向量、四维 GLCM 特

征向量和一维平均灰度值特征向量共同描述，将所提取的五种特征全部拼接组合在一起，即可得到本步骤的特征提取结果——27 维的像素点中心点邻域特征向量。

3.1.2 交互划线标注训练

在大部分机器学习与深度学习算法中，存在完全标注的图像样本是训练语义分割模型的重要前提。然而，对于材料图像的分割任务而言，获取完全标注的图像样本存在着一些问题。一方面，材料图像的语义不易被人理解，只有具备相关材料领域知识的专家才能对材料图像中的各种物相种类及边界等信息进行精确地描述，从而实现对材料图像的完全标注。另一方面，材料图像的完全标注需要花费大量的时间。相较于自然场景图像，材料图像往往表现出组织粘连和目标交错的情况，这导致了较高的标注代价。由于存在以上两方面问题，在本工作中，为了克服上述困难，本方法对材料图像使用划线标注这一非完全标注形式代替常规的完全标注。

现有的机器学习和深度学习算法在实际应用的过程中面临着标注与训练等方面的难点，这一问题在解决非计算机领域的任务时尤为突出。借助于计算机图像处理技术完成材料图像的分割，不仅依赖于设计良好的图像分割算法，还受到材料相关人员使用操作的影响。因此，选择更加简便且合理的标注方式不但有利于非计算机专业的人员进行使用，还有助于避免在标注或训练时产生偏差。基于上述因素的考虑，本方法选用了交互式划线标注的形式，由材料方面的使用人员根据实际需求和既得分割效果进行多轮的划线标注，并据此对分割模型进行训练。

本方法在对材料图像进行交互式划线标注时，图像中的每个物相均使用一条曲线进行标注。在首次标注一张材料图像时，对其中各物相最典型的区域逐个进行划线，划线标注的像素点被提取出并记录其相应的物相类别；而在后续轮次的标注中，则对当前分割结果的误分割区域进行划线，提取并记录相应像素点的物相类别。根据上一步骤中材料图像的特征提取结果，可以得到每个标注像素点所对应的中心点邻域特征向量，该部分特征即为训练模型所使用的数据集。

为了从上述构建得到的数据集中学习得到中心点邻域特征向量的分类信息，采用梯度提升决策树 (GBDT, Gradient Boosting Decision Tree)[92] 作为特征向量分

类器进行训练。GBDT 是一种基于 CART 回归树的集成学习算法，它采用 Gradient Boosting 通过梯度下降对新的学习器进行迭代，其主要流程如算法3.1所示。

算法 3.1: GBDT 算法流程

输入: 迭代轮数 M , 样本数 N , 输入数据 x_i 和输入数据的类别 y_i

$$f_0(x) = \arg \min_{\rho} \sum_{i=1}^N L(y_i, \rho)$$

for $m = 1, 2, \dots, M$ **do**

for $i = 1, 2, \dots, N$ **do**

$$r_{im} = - \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x)=f_{m-1}(x)}$$

end

$$w^* = \arg \min_w \sum_{i=1}^N (r_{im} - h_t(x_i; w))^2$$

$$\rho^* = \arg \min_{\rho} \sum_{i=1}^N L(y_i, F_{t-1}(x_i) + \rho h_t(x_i; w^*))$$

$$f_m(x) = f_{m-1}(x) + \rho^* h_t(x; w^*)$$

end

输出: 经过 M 轮迭代后获得的模型 $f_M(x)$

通过对划线标注得到的中心点邻域特征数据集使用 GBDT 进行训练，可以获得中心点邻域特征分类器模型。将该模型用于对未标注像素点的中心点邻域特征进行分类，即可得到图像中所有像素点的类别，从而完成对图像的语义分割。在使用模型对新的同一类别材料图像进行分割时，只需要先提取图像中所有像素点的中心点邻域特征，并使用训练得到的中心点邻域特征分类器模型对特征进行分类，就可以预测该图像的分割结果。

3.1.3 迭代标注训练优化

由于本方法所采用的划线标注形式在仅进行一轮标注时只能涵盖极少量的像素点，而标注信息量又与模型精度息息相关。为了实现以较低的标注代价获得较为精确的分割模型，重复进行多轮划线标注是一种合理的选择。该多轮划线标注并训练模型的过程被称为迭代优化。不同于第一轮的划线标注，在进行后续轮次的划线标注时，已经可以获得经过初步训练的分割模型。因此，在后续轮次的标注中，除了可以选取材料图像中各物相较为典型的区域进行标注，还可以参考当前模型的分割

结果对误分割区域进行标注。由于既得模型的分割精度可能受到多种因素的影响而无法达到理想效果，通过交互式标注对当前模型分割错误的区域进行反馈有助于令模型学习得到更为有效的分类规则。

在迭代优化的过程中，为了使得模型在保留已习得信息的同时，进一步学习新一轮划线标注所给予的标注信息，增量学习被用于对迭代优化过程中的模型进行训练。当不使用增量学习进行迭代优化时，历史轮次的标注数据需要始终予以保留来保证模型可以学习到先前轮次的标注信息，这导致迭代优化的轮次较多时，历史标注数据的数量将会成倍增加并大量占用存储空间。此外，将保存的大量历史标注数据全部用于训练还会显著降低后续轮次的训练时间。为了避免上述问题的发生，在本方法的迭代优化过程中，采用增量学习训练模型学习最新轮次的划线标注信息，对之前的知识进行修正和增强。

尽管迭代优化的过程可以一直持续，但是为了达到标注代价与分割精度二者之间的平衡，需要选定一个合理的终止轮次结束迭代优化。在实现过程中，划线标注像素点的数量是对于标注代价的一种直观且量化的反映，而图像的分割精度则是模型精度的直接反馈。因此，在迭代优化的过程中，当标注像素点数量的增长幅度超过图像分割精度的提升幅度时，即终止对当前模型的迭代优化过程。至此，可以得到最终的中心点邻域特征分类模型，即材料图像语义分割模型。

3.2 实验分析

为了充分展示本方法的实际分割效果，本文选择碳钢 (UHCSDB)[9]、钛合金 (*TiAl*, titanium alloy)、木材、Pt-Al 涂层横截面、WC-Co 涂层横截面和陶瓷共晶 $HfB_2 - B_4C$ 共六种不同种类的材料图像进行测试以验证所提出的方法。在六种材料图像中，我们选取来自 UHCSDB 公共数据集 [9] 的碳钢图像进行详细的方法流程展示，并与其它图像分割方法进行对比。实验中所使用的碳钢图像实例如图3.3所示。本章节采用基于交互式标注的机器学习材料图像分割方法，通过多轮交互式标注训练得到分割模型，验证提出的方法在材料图像分割问题上的性能。所有实验均在 Intel(R) Core(TM) i7-10700 CPU 2.9GHz*16 和 32G RAM 的计算机上执行。

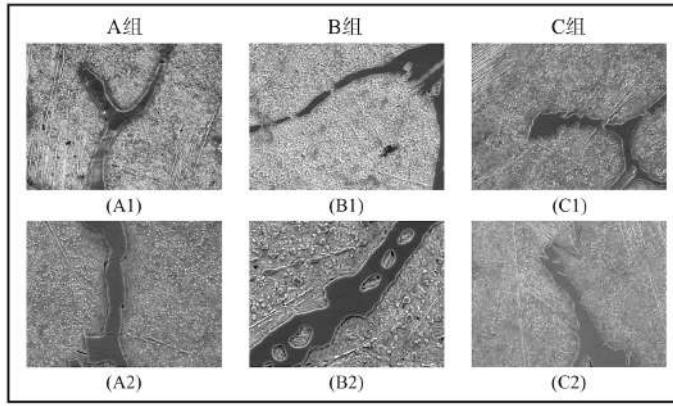


图 3.3: 碳钢图像分组示意图, 其主要成分为球状渗碳体。全部六幅碳钢图像被随机分为三组, 每组包含两张图像。

3.2.1 迭代终止轮次分析

在迭代优化的过程中, 每一轮的交互式划线标注和模型训练可以被不断重复以改善既有模型的训练效果。不过在实际应用时, 尽管使用者可以根据自身任务对模型精度和标注代价敏感程度的需要, 自适应地调整迭代优化的轮数, 但是为了尽可能地实现标注代价与分割精度二者之间的平衡, 并提供一个可参考的迭代优化轮数, 本文针对碳钢图像在迭代优化过程中标注代价与分割精度二者的变化关系进行了实验分析。

为了分析研究标注代价与分割精度之间的变化关系, 需要首先对标注代价与分割精度这两个对象进行量化地描述。对于模型精度而言, 使用常见的分割结果评价指标即可对其进行量化地描述与比较, 本文选取了前景 IoU 和全图分割结果的 mIoU 指标作为分割结果的评价指标以量化地描述模型精度。相较于模型精度, 标注代价的描述较为抽象, 标注时间、标注工具以及标注的知识门槛等因素都会对标注代价产生直接的影响。为了避免上述因素中因人为操作而导致的不确定性, 本文采用标注的像素点数量对标注代价进行描述。标注的像素点数量既可以精确地量化描述标注规模, 又可以在一定程度上反映花费在标注工作上的时间开销, 因此标注的像素点数量是标注代价的一种理想等效表示形式。图3.4展示了使用碳钢 A 组图像进行训练过程中的模型精度-标注代价变化趋势图。

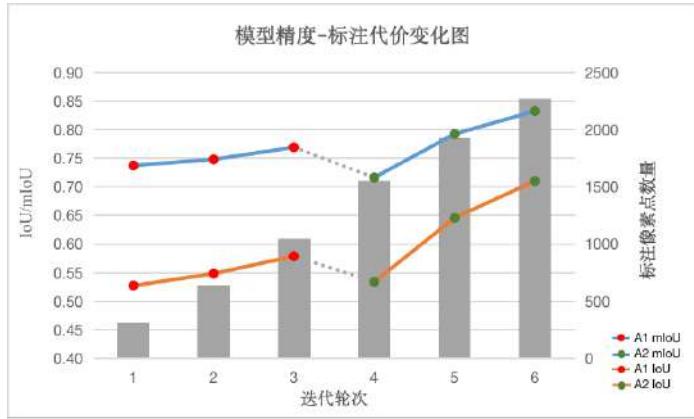


图 3.4: 碳钢 A 组图像训练过程中的模型精度-标注代价变化趋势图。蓝色折线展示了训练图像的全图分割结果随迭代轮次的 mIoU 变化情况；橙色折线展示了训练图像的前景分割结果随迭代轮次的 IoU 变化情况；灰色柱形表示累计标注像素点的数量。

在图3.4中，A 组的两幅碳钢图像分别进行了三轮迭代优化，共进行了六轮迭代优化。在每轮迭代优化完成时，累计划线标注的像素点数量以及当前模型对训练图像分割得到的前景 IoU 和全图 mIoU 三项指标被记录下来。观察图3.4的变化趋势可以发现，在进行第三轮和第六轮迭代优化时，标注像素点数量的增长幅度超过了模型分割精度的提升幅度。根据这一规律判断，在对同一幅训练图像重复交互式标注和模型训练三次后，单位标注代价下模型分割精度的提升开始放缓，标注工作的边际效应开始逐步体现。尽管继续对同一训练图像进行交互式标注和模型训练仍可能进一步改善所得模型的分割效果，但是出于平衡标注代价和模型精度二者关系的考虑，继续进行迭代优化带来的收益较低。综上所述，本方法认为对每幅训练图像进行三轮迭代优化较为合理，当具有多幅训练图像时，可以依次对每副训练图像进行三轮迭代优化以获得最终的分割模型。

3.2.2 相关方法对比

为了进一步展示本方法的分割效果，本文共选择了 18 种方法进行横向对比，包括三种基于非学习方法的图像分割算法、十一种基于机器学习的图像分割算法以及四种基于深度学习的图像分割算法。本章节通过在 A 组碳钢图像上进行迭代标注训练获得分割模型，对 B 组和 C 组碳钢图像进行测试分割，从而比较不同方法

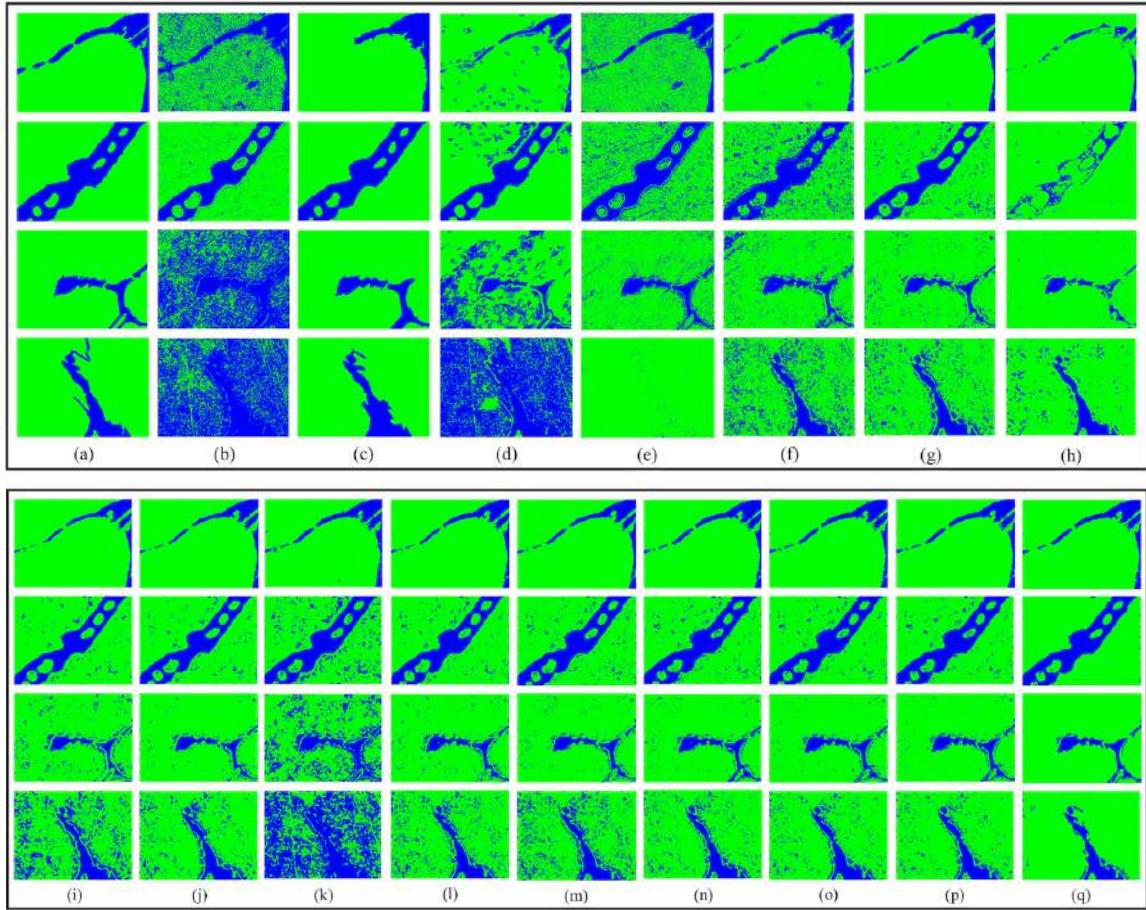


图 3.5: 本方法与非学习算法和机器学习算法的图像分割结果图。(a) 完全标注图; (b) 马尔科夫随机场; (c) 分水岭; (d) 复杂纹理混合分割算法; (e) meanshift; (f) 决策树; (g) KNN; (h) K-means; (i) 朴素贝叶斯; (j) SVM(RBF); (k) SVM(Sigmoid); (l) 随机森林; (m) AdaBoost; (n) XGBoost; (o) DART; (p) GBDT (本方法选用的分类器); (q) CES (GBDT+ 后处理)

的分割效果并进行分析。为了更好地量化各种方法的分割效果，本文采用前景物相分割结果的 IoU 和 Dice 系数、全图分割结果的 mIoU 和 mDice 系数以及准确率 (Accuracy) 共五种评价指标描述并比较不同方法的分割结果。图3.5展示了完全标注图、马尔科夫随机场 (Markov Random Field algorithm, MRF)[93] 的分割结果、分水岭算法 (watershed)[94] 的分割结果、Han 等人提出的复杂纹理混合分割算法 [95] 的分割结果、meanshift 算法 [96] 的分割结果、决策树 (decision tree, DT)[97] 的分割结果、K 最近邻 (K-Nearest Neighbor, KNN) 算法 [98] 的分割结果、K-means 算法 [99] 的分割结果、朴素贝叶斯 (Naive Bayes) 算法 [100] 的分割结果、基于

RBF 核函数的支持向量机 (Support Vector Machine, SVM)[101] 分割结果、基于 Sigmoid 核函数的支持向量机分割结果、随机森林 (random forest, RF) 算法 [40] 的分割结果、AdaBoost(Adaptive Boosting) 算法 [102] 的分割结果、XGBoost(extreme Gradient Boosting) 算法 [103] 的分割结果、DART(Dropouts meet multiple additive regression trees) 算法 [104] 的分割结果和本方法使用的 GBDT(Gradient Boosting Decision Tree) 算法 [92] 的分割效果。由于 GBDT 算法的分割结果中仍存在着部分噪声，这些噪声主要表现为离散的小块孤立区域。为了使得分割结果更加精细，本方法采用了一种简单的后处理方式改善分割结果。对于 GBDT 算法的分割结果，检测连通域面积小于设定阈值的孤立区域，并将其填充为周围像素点的类别从而减少

表 3.1: 基于碳钢 A 组图像使用不同方法训练得到的模型对碳钢 B1 图像分割结果的 IoU (蓝色前景物相)、mIoU、准确率、Dice (蓝色前景物相) 和 mDice 评价指标

算法 \ B1		IoU	mIoU	准确率	Dice	mDice
	算法					
MRF		0.315	0.502	0.727	0.479	0.647
Watershed		0.763	0.862	0.966	0.866	0.923
Han		0.529	0.721	0.921	0.692	0.823
Meanshift		0.424	0.621	0.840	0.595	0.748
DT		0.708	0.831	0.958	0.829	0.903
KNN		0.688	0.821	0.958	0.815	0.896
K-means		0.352	0.631	0.914	0.521	0.737
Naive Bayes		0.664	0.807	0.955	0.798	0.886
SVM(RBF)		0.686	0.820	0.958	0.813	0.895
SVM(Sigmoid)		0.798	0.883	0.972	0.887	0.936
RF		0.723	0.841	0.963	0.840	0.909
AdaBoost		0.733	0.847	0.964	0.846	0.913
XGBoost		0.781	0.874	0.971	0.877	0.930
DART		0.769	0.867	0.969	0.870	0.926
GBDT		0.783	0.876	0.971	0.879	0.931
CES		0.788	0.878	0.972	0.881	0.933

GBDT 算法分割结果中的噪声。经过 GBDT 算法分类和后处理可以得到图像的最终分割结果，该方法被称为中心点邻域特征分割算法 (CES)。图3.5中的 (p) 和 (q) 列展示了 GBDT 算法去噪前后的分割结果。每种算法分割结果的评价指标如表3.1、表3.2、表3.3和表3.4所示。

表 3.2: 基于碳钢 A 组图像使用不同方法训练得到的模型对碳钢 B2 图像分割结果的 IoU
(蓝色前景物相)、mIoU、准确率、Dice (蓝色前景物相) 和 mDice 评价指标

算法 \ B2	IoU	mIoU	准确率	Dice	mDice
MRF	0.674	0.781	0.908	0.806	0.873
Watershed	0.959	0.973	0.990	0.979	0.986
Han	0.677	0.777	0.903	0.808	0.871
Meanshift	0.544	0.658	0.821	0.705	0.788
DT	0.590	0.697	0.847	0.742	0.816
KNN	0.644	0.757	0.894	0.783	0.857
K-means	0.298	0.551	0.820	0.459	0.675
Naive Bayes	0.672	0.780	0.909	0.804	0.872
SVM(RBF)	0.696	0.795	0.915	0.821	0.883
SVM(Sigmoid)	0.625	0.731	0.872	0.769	0.840
RF	0.694	0.792	0.912	0.819	0.881
AdaBoost	0.723	0.811	0.920	0.839	0.893
XGBoost	0.752	0.830	0.928	0.858	0.905
DART	0.742	0.825	0.927	0.852	0.902
GBDT	0.762	0.837	0.932	0.865	0.910
CES	0.846	0.898	0.960	0.917	0.945

通过观察表3.1、表3.2、表3.3和表3.4所展示的各方法分割结果评价指标可以发现，分水岭算法和本方法的分割结果有着较好的精度，而其它算法的分割精度相对较差且并不稳定。在对四副碳钢图像进行测试分割时，分水岭算法分割结果的 IoU、mIoU、准确率、Dice 和 mDice 分别为 83.3%、90.4%、97.8%、90.7% 和 94.7%，本方法分割结果的 IoU、mIoU、准确率、Dice 和 mDice 分别为 73%、84.4%、96.2%、

表 3.3: 基于碳钢 A 组图像使用不同方法训练得到的模型对碳钢 C1 图像分割结果的 IoU
 (蓝色前景物相)、mIoU、准确率、Dice (蓝色前景物相) 和 mDice 评价指标

算法 \ C1	IoU	mIoU	准确率	Dice	mDice
算法					
MRF	0.124	0.226	0.386	0.221	0.357
Watershed	0.825	0.903	0.982	0.904	0.947
Han	0.156	0.433	0.725	0.270	0.550
Meanshift	0.407	0.646	0.894	0.578	0.759
DT	0.386	0.634	0.890	0.557	0.747
KNN	0.425	0.671	0.921	0.597	0.776
K-means	0.484	0.715	0.948	0.652	0.812
Naive Bayes	0.351	0.622	0.898	0.519	0.731
SVM(RBF)	0.486	0.711	0.939	0.654	0.810
SVM(Sigmoid)	0.245	0.496	0.765	0.394	0.624
RF	0.489	0.710	0.934	0.657	0.810
AdaBoost	0.501	0.717	0.938	0.668	0.817
XGBoost	0.550	0.745	0.943	0.710	0.839
DART	0.551	0.746	0.945	0.711	0.840
GBDT	0.569	0.756	0.948	0.725	0.848
CES	0.663	0.813	0.965	0.797	0.889

表 3.4: 基于碳钢 A 组图像使用不同方法训练得到的模型对碳钢 C2 图像分割结果的 IoU
 (蓝色前景物相)、mIoU、准确率、Dice (蓝色前景物相) 和 mDice 评价指标

算法 \ C2	IoU	mIoU	准确率	Dice	mDice
算法					
MRF	0.133	0.193	0.330	0.235	0.319
Watershed	0.786	0.878	0.973	0.880	0.933
Han	0.119	0.164	0.285	0.213	0.279
Meanshift	0.017	0.453	0.889	0.034	0.488
DT	0.298	0.542	0.803	0.459	0.669
KNN	0.360	0.603	0.859	0.530	0.723
K-means	0.476	0.693	0.917	0.645	0.799
Naive Bayes	0.251	0.500	0.768	0.401	0.629
SVM(RBF)	0.395	0.631	0.877	0.567	0.747
SVM(Sigmoid)	0.153	0.279	0.462	0.265	0.420
RF	0.362	0.604	0.859	0.532	0.724
AdaBoost	0.313	0.560	0.823	0.477	0.685
XGBoost	0.429	0.656	0.892	0.600	0.769
DART	0.450	0.671	0.901	0.620	0.782
GBDT	0.471	0.686	0.909	0.640	0.794
CES	0.623	0.785	0.951	0.767	0.870

84.1% 和 90.9%，两种方法的分割结果评价指标均显著优于其它对比方法。尽管分水岭算法的分割结果评价指标表现良好，但是在观察图3.5中 (c) 列第一行和第三行的分割结果时可以发现，前景物相的部分内容出现了大面积的缺失。为了对“使用分水岭算法获得的分割结果可能导致前景物相内容产生缺失”这一猜想作出验证，并进一步比较本方法与分水岭算法在材料图像分割任务中的差异，本文使用了碳钢图像和 WC-Co 涂层横截面图像对分水岭算法和本方法进行了进一步地比较实验，其结果如图3.6所示。

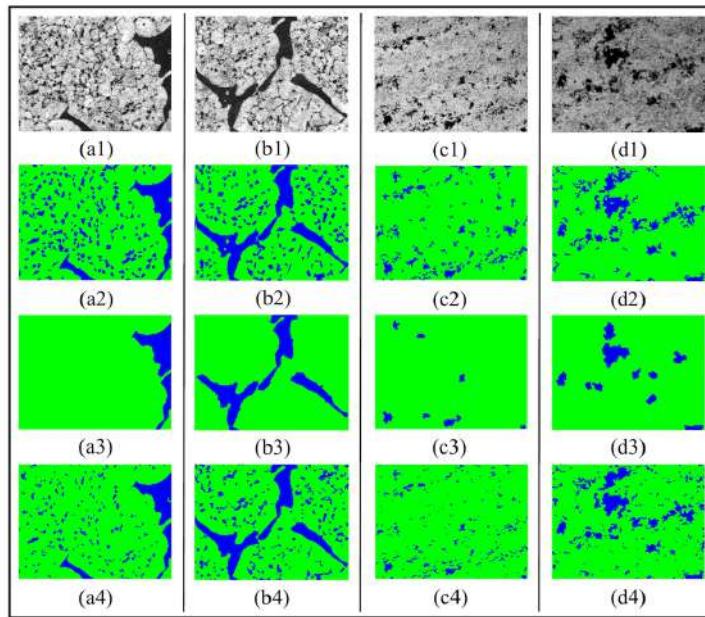


图 3.6：分水岭算法和本方法 (CES) 对碳钢图像及 WC-Co 涂层横截面图像的分割结果。
(a1)、(b1) 是主要成分为珠光体和球状渗碳体的碳钢图像，(c1) 和 (d1) 为 WC-Co 涂层横截面图像。(a2)、(b2)、(c2) 和 (d2) 为完全标注图。(a3)、(b3)、(c3) 和 (d3) 为分水岭算法的分割结果。(a4)、(b4)、(c4) 和 (d4) 为 CES 的分割结果。

根据图3.6的分割结果可以发现，分水岭算法无法很好地处理前景物相连通面积较小的情况，仅能对其中连通面积较大的区域作出分割，这对于材料图像的分割工作是无法接受的。材料图像中往往存在着大量细碎且离散的独立区域，在进行物相面积统计与定位等分析工作时，全部的独立区域都应当被检测得到。在这种情况下，使用分水岭算法获得分割结果会导致明显的偏差，因此相较于分水岭算法，本方法更适合于材料图像的分割工作。

基于非学习算法的图像分割方法之所以无法取得相对良好且稳定的分割效果，其根本原因是此类算法主要基于单个像素的数值信息和非常有限的多像素关联信息，这导致此类算法忽略了像素点周围的环境信息，无法在像素点本身数值信息相似但环境信息不同的情况下作出正确的判断。本方法由于采用了多种有效的特征描述算子，包括 Gabor 滤波器、Hu 矩、方向梯度直方图 (HOG, Histogram of Oriented Gradient)、灰度共生矩阵 (GLCM, Gray Level CO-Occurrence Matrix) 和平均灰度值，能够更加充分地对像素点及自身周围的环境信息进行描述，增强了特征表示能力，因而可以取得更好的效果。相比于非学习图像分割算法，基于机器学习的图像分割算法具备归纳与总结多个像素特征之间关联与特征分布的能力。因此，SVM 等机器学习算法在实验图像上表现出了较为良好的性能。相较于其它机器学习算法，GBDT 算法不依赖于特征值的归一化处理，并且可以获得相对更好的预测精度，这使得 GBDT 算法在对中心点邻域特征分类的过程中表现出了良好的效果。考虑到 GBDT 算法所具有的上述特性，在众多机器学习算法中，本方法选用了最适合处理中心点邻域特征的 GBDT 算法作为分类器。为了进一步改善 GBDT 算法的分割效果，本方法在以 GBDT 算法取得分割结果后，进一步对该结果进行了后处理，去除了其中面积较小的离散预测区域，这有效避免了材料图像中的噪声对于整体分割结

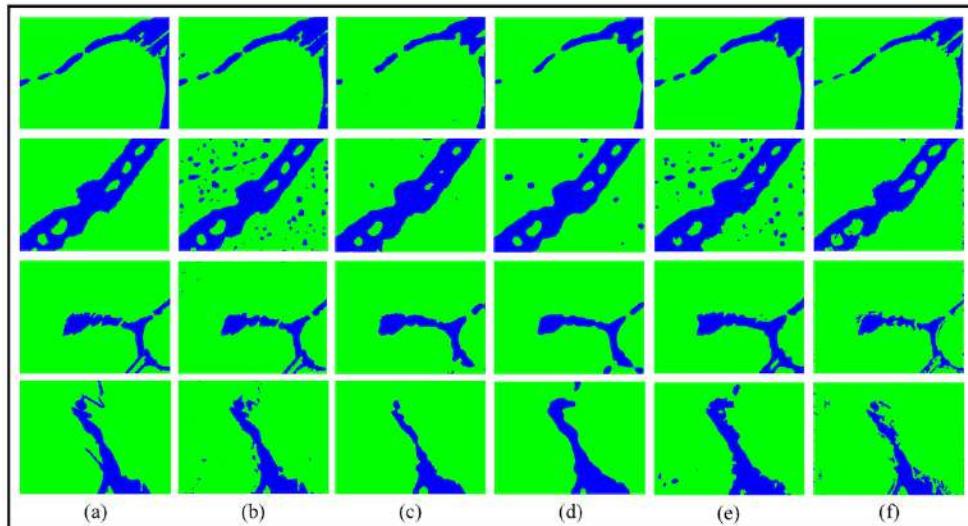


图 3.7: FCN、SegNet、PSPNet、Unet++ 和本方法 (CES) 对碳钢图像的分割结果。(a) 完全标注图; (b) FCN; (c) SegNet; (d) PSPNet; (e) Unet++; (f) CES

果产生的影响。

除了基于非学习方法的图像分割算法和机器学习图像分割算法，本文还对比了使用完全标注图进行全监督学习的深度学习图像分割算法。由于基于材料图像设计的深度学习网络数量较少，本文选取的深度学习网络主要应用于自然场景图像和医学图像的分割工作，分别为 FCN[10]、SegNet[11]、PSPNet[12] 和 Unet++[46]。图3.7展示了四种深度学习网络和本方法的分割结果对比，分割结果的评价指标如表3.5、表3.6、表3.7和表3.8所示。相较于本方法，深度学习网络采用了完全标注进行全监督训练且网络可以根据图像的特点自适应地提取特征，因此深度学习即使未经过后处理，仍然可以持平甚至超过本方法的分割效果。不过，使用深度学习网络进行分割需要更高的标注代价以及相对较长的训练时间，在标注代价高昂或难以获取完全标注的情况下，本方法仍然具有一定的可取之处。

表 3.5: 基于碳钢 A 组图像使用 FCN、SegNet、PSPNet、Unet++ 和本方法 (CES) 训练得到的模型对碳钢 B1 图像分割结果的 IoU (蓝色前景物相)、mIoU、准确率、Dice (蓝色前景物相) 和 mDice 评价指标

算法 \ B1	IoU	mIoU	准确率	Dice	mDice
FCN	0.925	0.957	0.990	0.961	0.978
SegNet	0.748	0.855	0.966	0.856	0.919
PSPNet	0.763	0.864	0.968	0.866	0.924
Unet++	0.906	0.946	0.987	0.951	0.972
CES	0.788	0.878	0.972	0.881	0.933

3.2.3 方法普适性分析

由于材料图像的种类十分丰富，且每种类别的材料图像特点各不相同，这导致一些方法仅适用于单一种类的材料图像分割工作。为了验证本方法的普适性，本文除了采用碳钢 (UHCSDB) 图像进行实验外，还选取了钛合金 ($TiAl$, titanium alloy)、木材、Pt-Al 涂层横截面、WC-Co 涂层横截面和陶瓷共晶 $HfB_2 - B_4C$ 五种不同种类的材料图像进行测试，同样对每张训练图像进行三轮迭代优化获取最终的分割模

表 3.6: 基于碳钢 A 组图像使用 FCN、SegNet、PSPNet、Unet++ 和本方法 (CES) 训练得到的模型对碳钢 B2 图像分割结果的 IoU (蓝色前景物相)、mIoU、准确率、Dice (蓝色前景物相) 和 mDice 评价指标

算法 \ B2	IoU	mIoU	准确率	Dice	mDice
算法	FCN	0.808	0.864	0.940	0.894
SegNet	0.899	0.931	0.972	0.947	0.964
PSPNet	0.907	0.937	0.975	0.951	0.967
Unet++	0.796	0.855	0.936	0.886	0.921
CES	0.846	0.898	0.960	0.917	0.945

表 3.7: 基于碳钢 A 组图像使用 FCN、SegNet、PSPNet、Unet++ 和本方法 (CES) 训练得到的模型对碳钢 C1 图像分割结果的 IoU (蓝色前景物相)、mIoU、准确率、Dice (蓝色前景物相) 和 mDice 评价指标

算法 \ C1	IoU	mIoU	准确率	Dice	mDice
算法	FCN	0.894	0.941	0.990	0.944
SegNet	0.764	0.869	0.976	0.866	0.927
PSPNet	0.772	0.874	0.977	0.872	0.929
Unet++	0.792	0.882	0.976	0.884	0.935
CES	0.663	0.813	0.965	0.797	0.889

表 3.8: 基于碳钢 A 组图像使用 FCN、SegNet、PSPNet、Unet++ 和本方法 (CES) 训练得到的模型对碳钢 C2 图像分割结果的 IoU (蓝色前景物相)、mIoU、准确率、Dice (蓝色前景物相) 和 mDice 评价指标

算法 \ C2	IoU	mIoU	准确率	Dice	mDice
算法	FCN	0.828	0.903	0.979	0.906
SegNet	0.725	0.846	0.970	0.841	0.912
PSPNet	0.737	0.849	0.964	0.848	0.914
Unet++	0.703	0.828	0.957	0.826	0.901
CES	0.623	0.785	0.951	0.767	0.870

型，其分割结果如图3.8所示。对于每种类型的图像，仍采用前景 IoU 和 Dice 系数、全图 mIoU 和 mDice 系数以及准确率五种指标量化评价分割效果，如表3.9所示。

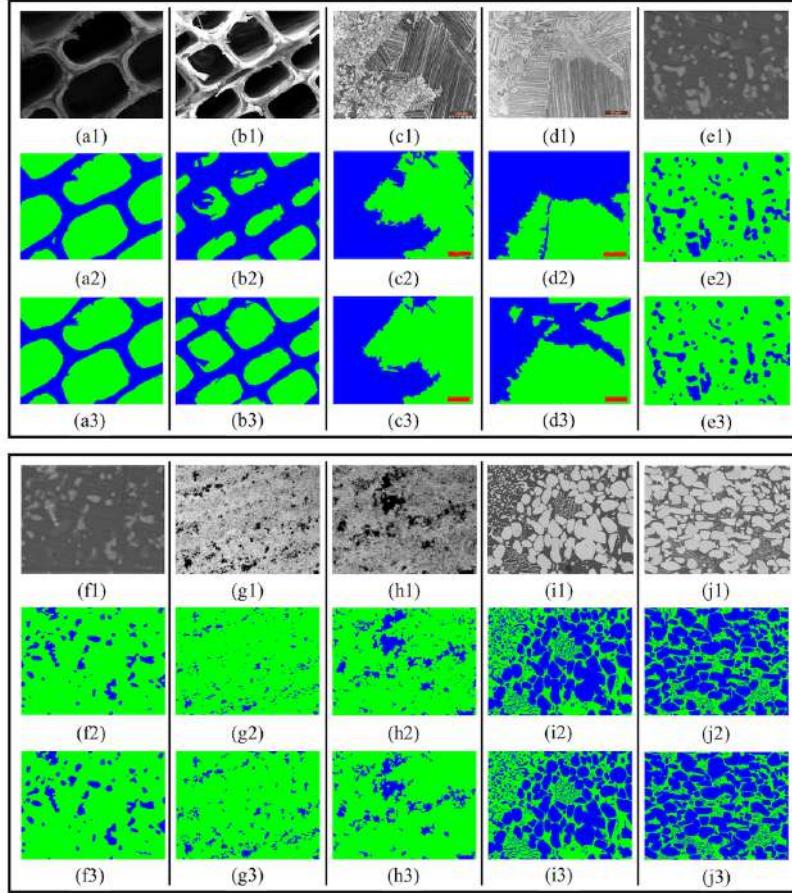


图 3.8: 本方法在不同数据集上的分割结果。(a1) 和 (b1) 为木材图像。(c1) 和 (d1) 为钛合金图像。(e1) 和 (f1) 为 Pt-Al 涂层横截面图像。(g1) 和 (h1) 为 WC-Co 涂层横截面图像。(i1) 和 (j1) 为陶瓷共晶 $HfB_2 - B_4C$ 图像。(a2)、(b2)、(c2)、(d2)、(e2)、(f2)、(g2)、(h2)、(i2) 和 (j2) 为本方法的分割结果。(a3)、(b3)、(c3)、(d3)、(e3)、(f3)、(g3)、(h3)、(i3) 和 (j3) 为完全标柱图。

通过观察不同数据集上的分割结果可以发现，本方法可以对实验中的大部分材料图像取得合理的分割效果，本方法具有普适性并可应用于多种类别的材料图像。之所以本方法能够解决不同类别的材料图像分割问题，这得益于本方法所提取的中心点邻域特征。非学习图像分割算法和使用单一特征描述子的机器学习图像分割算法由于对材料图像的特征描述能力较弱，在面对部分种类的材料图像时无法有效区分不同物相的特征，因此导致缺乏普适性。

表 3.9: 本方法对木材、钛合金、Pt-Al、WC-Co 和陶瓷共晶图像分割结果的 IoU (蓝色前景物相)、mIoU、准确率、Dice (蓝色前景物相) 和 mDice 评价指标。由于钛合金图像包含三相, 因此仅展示其 mIoU、准确率以及 mDice 指标。

数据集	IoU	mIoU	准确率	Dice	mDice
木材	0.710	0.711	0.826	0.824	0.820
钛合金	/	0.798	0.925	/	0.887
Pt-Al	0.783	0.875	0.971	0.873	0.928
WC-Co	0.542	0.743	0.947	0.694	0.832
陶瓷共晶	0.825	0.857	0.927	0.904	0.922

3.2.4 交叉验证实验

图3.3展示了六张用于交叉实验的碳钢图像。这六张图像被随机分为三组分别以 A、B、C 表示, 其中每组包含两张图像。在交叉实验的过程中共进行三组实验, 每次使用一组图像进行交互式划线标注对模型进行训练, 并对剩余两组图像进行分割测试以评价训练所得模型的效果。在每一组交叉实验中, 对于每一幅训练图像, 进行三轮划线标注。因此每一组交叉实验训练得到的模型共计进行了六轮划线标注。

图3.9展示了使用 A 组图像训练过程中的六轮划线标注信息。(a1) 和 (e1) 为原始图像, (a2) 和 (e2) 为完全标注图。(b1)、(c1)、(d1) 分别对应了 (a1) 的三轮划线标注数据; (f1)、(g1)、(h1) 分别对应了 (e1) 的三轮划线标注数据。(b2)、(c2)、(d2)、(f2)、(g2)、(h2) 分别为对应轮次根据标注信息学习得到的模型, 在当前图像上的分割结果。通过观察不同轮次的分割结果可以发现, 随着划线标注的不断进行, 模型的分割效果逐步得到改善。训练得到的模型在 B 组和 C 组图像上的测试分割结果如图3.10所示。同样的, 分别使用 B 组图像进行训练并对 A 组和 C 组图像测试分割, 使用 C 组图像进行训练并对 A 组和 B 组图像测试分割。B 组和 C 组图像的划线标注信息如图3.11和图3.13所示。图3.12和图3.14分别展示了使用 B 组和 C 组图像划线标注训练得到的模型在其余两组测试图像上的分割效果。为了量化地度量交叉实验的分割效果, 使用 IoU 和 mIoU 评价指标对前景物相和全图的分割结果进行计算, 结果如表3.10所示。

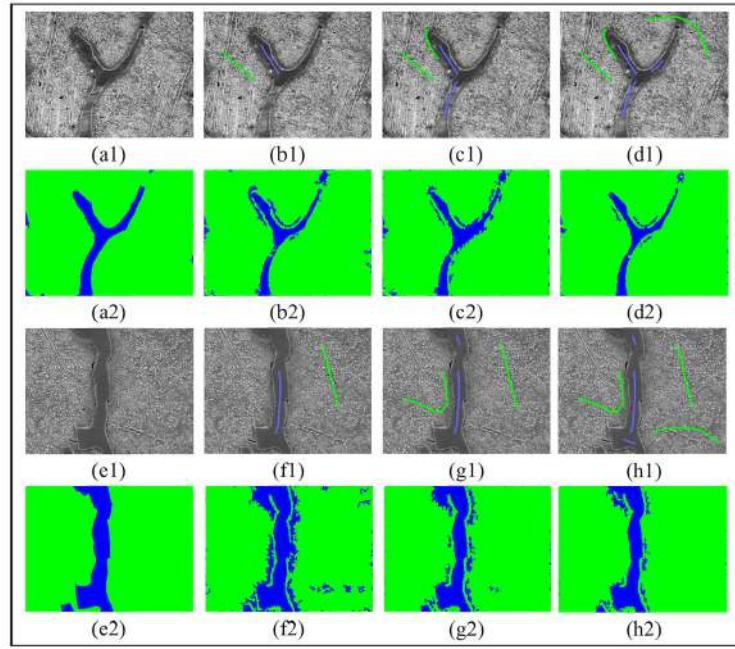


图 3.9: 碳钢图像 A 组划线标注示意图。(a1) 和 (e1) 为原始图像。(a2) 和 (e2) 为完全标注图。(b1)、(c1)、(d1) 分别对应了 (a1) 的三轮划线标注数据; (f1)、(g1)、(h1) 分别对应了 (e1) 的三轮划线标注数据。(b2)、(c2)、(d2)、(f2)、(g2)、(h2) 分别为对应轮次模型的分割结果。

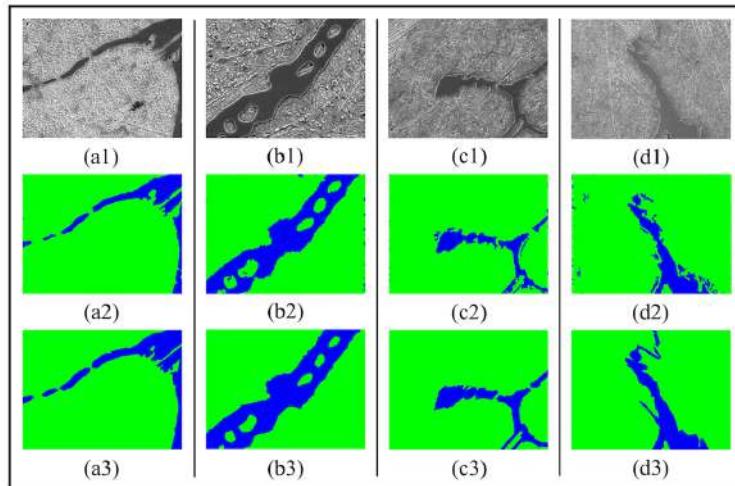


图 3.10: 使用 A 组碳钢图像训练得到的分割模型在 B 组和 C 组上的测试分割结果。(a1)、(b1)、(c1) 和 (d1) 为原始图像。(a2)、(b2)、(c2) 和 (d2) 为完全标注图。(a3)、(b3)、(c3) 和 (d3) 为模型的测试分割结果。

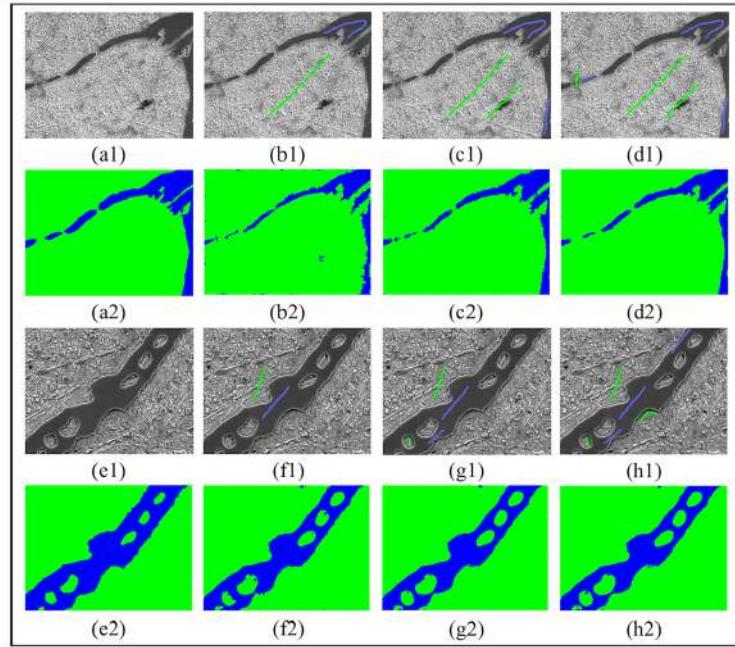


图 3.11: 碳钢图像 B 组划线标注示意图。(a1) 和 (e1) 为原始图像。(a2) 和 (e2) 为完全标注图。(b1)、(c1)、(d1) 分别对应了 (a1) 的三轮划线标注数据; (f1)、(g1)、(h1) 分别对应了 (e1) 的三轮划线标注数据。(b2)、(c2)、(d2)、(f2)、(g2)、(h2) 分别为对应轮次模型的分割结果。

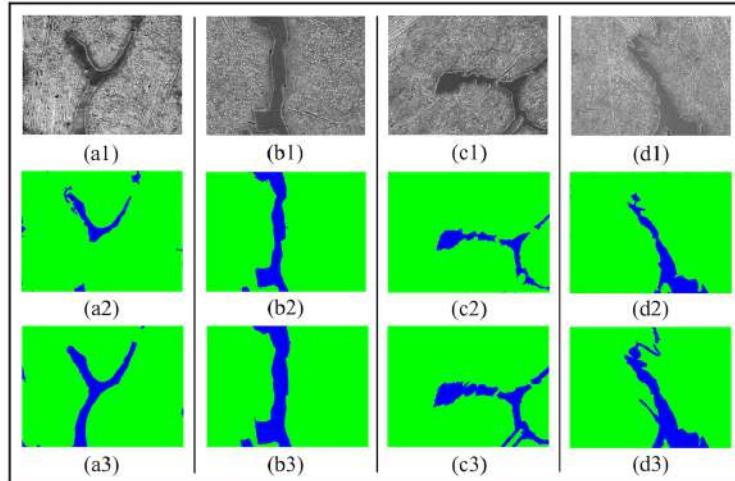


图 3.12: 使用 B 组碳钢图像训练得到的分割模型在 A 组和 C 组上的测试分割结果。(a1)、(b1)、(c1) 和 (d1) 为原始图像。(a2)、(b2)、(c2) 和 (d2) 为完全标注图。(a3)、(b3)、(c3) 和 (d3) 为模型的测试分割结果。

3.3 本章小结

本章针对具有复杂纹理的材料图像，提出了一种基于交互式标注的机器学习材料图像分割方法。首先对图像中每个像素点提取以当前像素点为中心的 9×9 像素

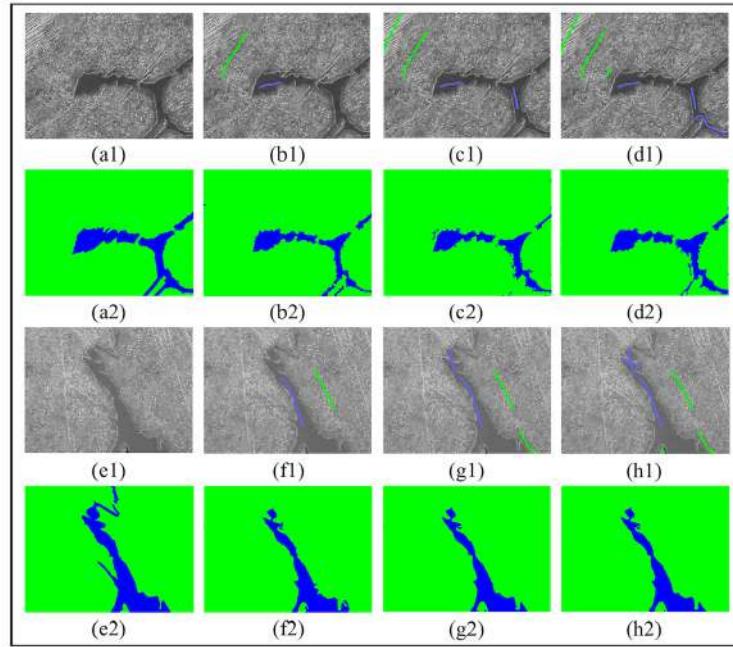


图 3.13: 碳钢图像 C 组划线标注示意图。(a1) 和 (e1) 为原始图像。(a2) 和 (e2) 为完全标注图。(b1)、(c1)、(d1) 分别对应了 (a1) 的三轮划线标注数据; (f1)、(g1)、(h1) 分别对应了 (e1) 的三轮划线标注数据。(b2)、(c2)、(d2)、(f2)、(g2)、(h2) 分别为对应轮次模型的分割结果。

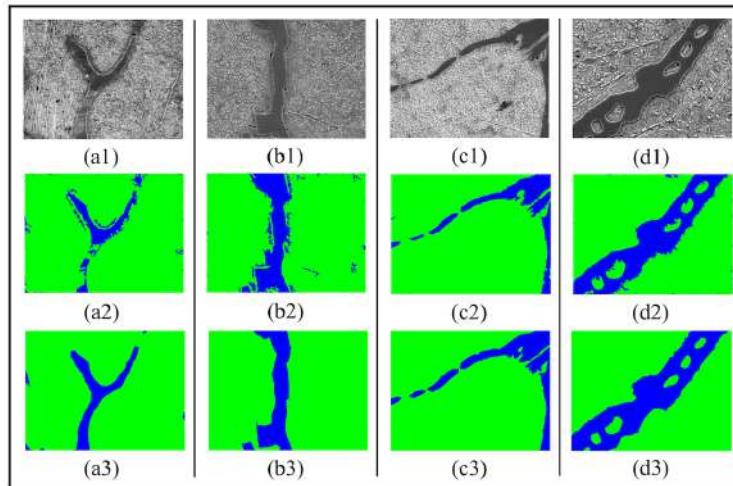


图 3.14: 使用 C 组碳钢图像训练得到的分割模型在 A 组和 B 组上的测试分割结果。(a1)、(b1)、(c1) 和 (d1) 为原始图像。(a2)、(b2)、(c2) 和 (d2) 为完全标注图。(a3)、(b3)、(c3) 和 (d3) 为模型的测试分割结果。

点矩形块作为感兴趣区域并在该区域提取多种特征，然后交互式划线标注图像中各类别物相以训练特征分类模型，最后在每张训练图像上迭代标注训练获得最终分割

表 3.10: 碳钢图像交叉实验分割结果

训练图像 IoU/mIoU	A1	A2	B1	B2	C1	C2
			0.788 / 0.878	0.846 / 0.898	0.663 / 0.813	0.623 / 0.785
A 组	/	/				
B 组	0.405 / 0.675	0.751 / 0.859	/	/	0.637 / 0.800	0.648 / 0.803
C 组	0.510 / 0.729	0.660 / 0.799	0.819 / 0.896	0.834 / 0.889	/	/

模型。本章所提出的方法具有以下三方面的贡献：首先，本方法针对材料图像标注困难这一问题，采用交互式划线标注训练的方式，既实现标注代价与分割精度二者之间的平衡，还可令使用者根据实际应用场景便捷地进行应用；其次，本方法采用 GBDT 作为分类器并使用增量学习迭代训练模型，这使得本方法分割模型的获取具有良好的实时性；最后，本方法设计了中心点邻域特征，可充分提取各种材料图像中像素点的环境信息，保证了本方法在材料图像分割任务中具有普适性。

本方法相较于非学习图像分割算法和机器学习图像分割算法具有更好的分割效果，相较于全监督深度学习算法可以显著降低标注代价，有助于解决多种材料图像的分割问题。但是在分割精度上，本方法相较于深度学习算法仍然存在一定差距，因此难以胜任对分割精度高度敏感的任务。此外，人为预先定义的图像特征描述算子始终存在一定的限制，无法保证能够良好地描述所有类别的图像，这使得不同种类材料图像的分割结果精度存在波动。针对本方法在分割精度与特征描述上的弊端，本论文将提出一种基于深度学习的方法来进一步解决材料图像的分割问题。

第四章 基于伪标签的深度学习材料图像分割方法

上一章所述方法以平衡标注代价与分割精度之间的关系为首要目标。尽管该方法可以显著降低标注代价并实现实时地获取模型，但是所使用的机器学习算法需要预先设计好的图像特征，而预先人为选定的特征存在一定的局限性，无法覆盖全部种类的材料图像特征。为了克服这一问题，并获取高精度的材料图像分割结果，本章仍然围绕着材料图像小样本、标注困难和算法缺乏普适性三大难点，提出了一种基于伪标签的深度学习材料图像分割方法。借助深度学习的自适应特征提取能力，令模型可以根据不同种类的材料图像特点提取更有效的特征表示，进而得到更为精确的图像分割结果。本方法的网络结构由两部分组成：主分支和辅助分支，其中主分支用于对材料图像进行特征提取和分割预测，辅助分支用于为训练图像生成伪标签。通过使用伪标签对分割结果进行补充监督，本方法实现了基于一到两张划线标注的复杂纹理材料图像，训练得到具有泛化能力的图像分割模型，且模型具有普适性。

4.1 方法概述

本方法基于伪标签思想，结合深度学习神经网络技术，完成对材料图像中各个物相的分割，设计的神经网络被称为伪标签纹理分割网络 (PTS, pseudo-labeling texture semantic segmentation)。标注困难是材料图像分割工作中的难点之一，PTS 网络使用划线标注代替完全标注进行监督学习以解决标注困难。在具有复杂纹理的材料图像中，不同物相之间的边界往往难以精确地划分。相较于完全标注，划线标注只需要对少数具有明确类别的像素进行标注，这显著减少了标注工作所需要的领域知识。然而，划线标注所覆盖的像素数量极少，不易对复杂的神经网络进行充分的监督学习。因此，PTS 网络通过在训练阶段生成伪标签，获取更多可用的监督信息。本章提出的 PTS 网络采用双分支结构，包含主分支和辅助分支两个部分。主分支用于提取图像特征和分割预测，辅助分支用于生成伪标签，帮助主分支进行训练，从而实现对主分支分割预测结果的双重监督（划线标注与伪标签）。在测试阶段，只有

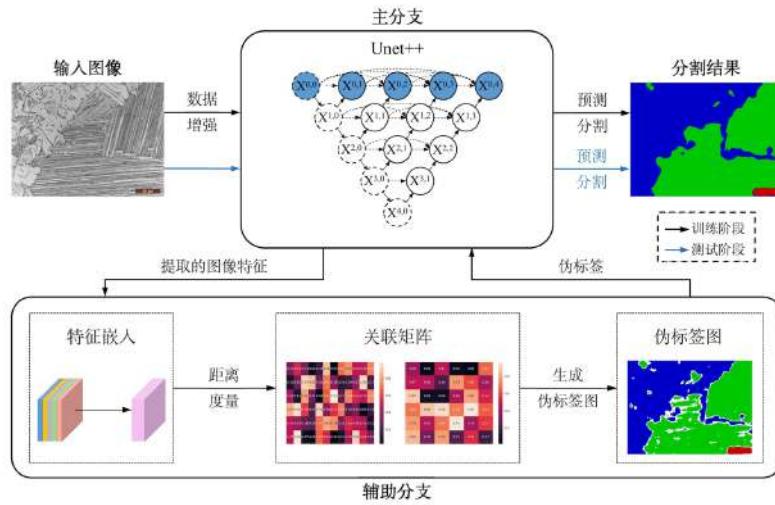


图 4.1: PTS 网络由一个主分支和一个辅助分支组成。主分支负责提取图像特征并预测分割结果; 辅助分支根据提取的图像特征生成伪标签以监督主分支的分割预测。黑色箭头描述了训练阶段的数据流, 蓝色箭头描述了测试阶段的数据流。

主分支被用于预测图像的分割结果。整个 PTS 网络的结构如图4.1所示。

4.1.1 数据增强

为了在深度学习的训练过程中, 解决材料图像样本数量不足的问题, 对训练所用的材料图像进行数据增强。在训练阶段, 每一张图像以 50% 的概率进行水平翻折, 随后再随机旋转 $n \times 10^\circ (0 \leq n < 18)$ 并输入主分支中。该数据增强操作近似等价于对原始数据集以旋转和翻折的方式进行扩充。在本方法中, 一张训练图像经过数据增强处理, 可以扩充为 36 张图像用于进行训练。

经过数据增强步骤, 不仅通过丰富训练数据的形式, 缓解了材料图像的小样本问题, 克服了深度学习训练过程中的障碍, 还可以确保模型对材料图像具备二维空间的旋转不变性, 从而进一步改善所得模型的鲁棒性。

4.1.2 主分支

在主分支中, 采用 Unet++[46] 作为基础网络对材料图像进行特征提取与分割预测, 如图4.2所示。主分支中包含三种不同类型的模块, 其中用虚线圆标注的模块

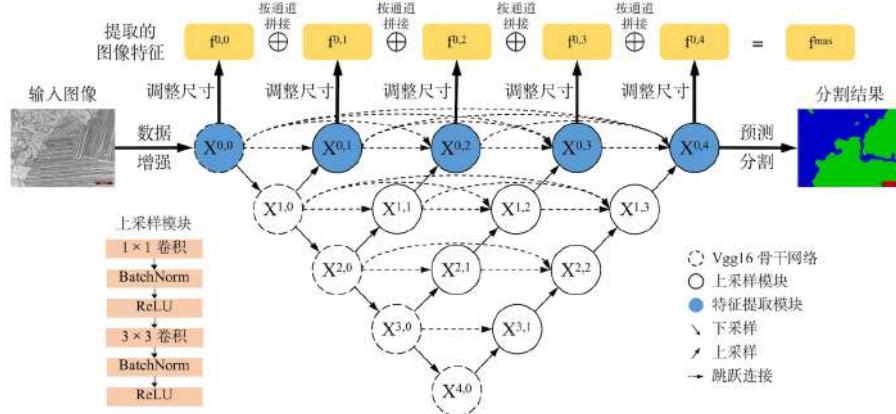


图 4.2: PTS 网络主分支结构示意图。主分支提取网络中不同层次的特征并预测输入图像的分割结果。 $f^{0,0}$ 、 $f^{0,1}$ 、 $f^{0,2}$ 、 $f^{0,3}$ 和 $f^{0,4}$ 分别表示 $X^{0,0}$ 、 $X^{0,1}$ 、 $X^{0,2}$ 、 $X^{0,3}$ 和 $X^{0,4}$ 模块的输出特征。⊕ 表示按通道拼接操作。粗实线箭头描述了主分支中的数据流动方向。

表示 Unet++ 使用的预训练骨干网络，实线圆标注的模块表示 Unet++ 的上采样模块，而使用蓝色圆圈标注的模块负责输出主分支提取的特征。从 $X^{0,0}$ 、 $X^{0,1}$ 、 $X^{0,2}$ 、 $X^{0,3}$ 和 $X^{0,4}$ 模块中提取得到的特征被输入到辅助分支中，用于生成伪标签。材料图像被输入到 Unet++ 网络中的 $X^{0,0}$ 模块，经过骨干网络和上采样模块的处理，主分支的分割预测结果可以从 $X^{0,4}$ 模块获得。

在大多数情况下，基于少量的训练样本从头开始训练，很难得到一个有效的图像分割模型。因此，主分支使用了在 ImageNet 数据集 [67] 上预训练的 Vgg16[105] 作为主分支的骨干网络。在图4.2中，主分支的骨干网络为 $X^{0,0}$ 、 $X^{1,0}$ 、 $X^{2,0}$ 、 $X^{3,0}$ 和 $X^{4,0}$ 五个使用虚线圆标注的模块。由于 ImageNet 数据集中的图像与材料图像存在明显的差异，并非全部预训练骨干网络的参数都对材料图像的分割具有积极作用。在实验中观察到，预训练的 Vgg16 骨干网络中浅层结构的参数是可迁移的，这是因为边界、纹理信息的提取在材料图像与自然场景图像中是通用的。由于材料图像中的物相类别数量较少且材料图像的数据量非常有限，本方法冻结了骨干网络的 $X^{0,0}$ 、 $X^{1,0}$ 、 $X^{2,0}$ 和 $X^{3,0}$ 四个模块，仅对骨干网络的 $X^{4,0}$ 模块基于材料图像进行微调，使 PTS 网络更适合处理材料图像的分割。

在数据增强和迁移预训练骨干网络这两种手段的帮助下，模型分割结果的稳定性与鲁棒性可以得到显著的提升。与此同时，模型因为图像样本不足而导致的过拟

合问题也借助于这两种手段得到了缓解。

从材料图像中提取用于生成伪标签的特征是主分支的另一项重要任务。在卷积神经网络中，浅层特征的特征图尺寸与原始图像大小相似，这有助于保留空间位置信息。然而，浅层特征的语义信息较弱，不利于区分不同类别。相对的，深层特征具有良好的区分度和更强的语义，可以有效地区分不同类别，但是深层特征的特征图经过了多次卷积和下采样，这使得深层特征丢失了大量的空间位置信息。为了能够最大程度地兼顾空间位置信息和特征语义信息，主分支利用 Unet++ 中的部分模块实现提取不同层次的特征作为辅助分支的输入。在图4.2中，特征提取模块 $X^{0,0}$ 、 $X^{0,1}$ 、 $X^{0,2}$ 、 $X^{0,3}$ 和 $X^{0,4}$ 以蓝色圆圈表示。提取得到的特征经过按通道拼接操作处理，组成了多维特征 f^{mas} 。由于 f^{mas} 同时涵盖了浅层特征和深层特征，因此 f^{mas} 能够充分且准确地对材料图像进行描述。 f^{mas} 可以由式 (4.1) 计算得到：

$$f^{mas} = R(f^{0,0}) \oplus R(f^{0,1}) \oplus R(f^{0,2}) \oplus R(f^{0,3}) \oplus R(f^{0,4}), \quad (4.1)$$

其中 \oplus 表示按通道拼接运算， $f^{0,0}$ 、 $f^{0,1}$ 、 $f^{0,2}$ 、 $f^{0,3}$ 、 $f^{0,4}$ 分别表示 Unet++ 中 $X^{0,0}$ 、 $X^{0,1}$ 、 $X^{0,2}$ 、 $X^{0,3}$ 、 $X^{0,4}$ 模块的输出特征，即图4.2中蓝色圆圈标注的模块， $R(\cdot)$ 表示使用双线性插值 [106] 将特征图改变到与原图相同的大小。

4.1.3 辅助分支

在辅助分支中，利用从主分支提取得到的特征可以生成伪标签。这些伪标签对主分支的分割结果进行监督，实现提升分割精度与稳定性。辅助分支的结构如图4.3所示。对于第 n 个训练批次的样本，主分支提取得到的多维特征集 f_n^{mas} 可以按照对应像素点是否存在划线标注被分为两部分，即划线标注像素对应的特征 f_n^l 与未标注像素对应的特征 f_n^u 。因此 f_n^{mas} 满足如下关系：

$$f_n^{mas} = f_n^l \cup f_n^u. \quad (4.2)$$

为了充分利用标注信息，第 $n-1$ 个训练批次样本中的划线标注像素特征 f_{n-1}^l 被额外保存，并作为第 n 个训练批次辅助分支的输入；而第 $n-1$ 个训练批次样本中的未标注像素对应的特征 f_{n-1}^u 由于并不包含标注信息，不用于第 n 个训练批次中辅

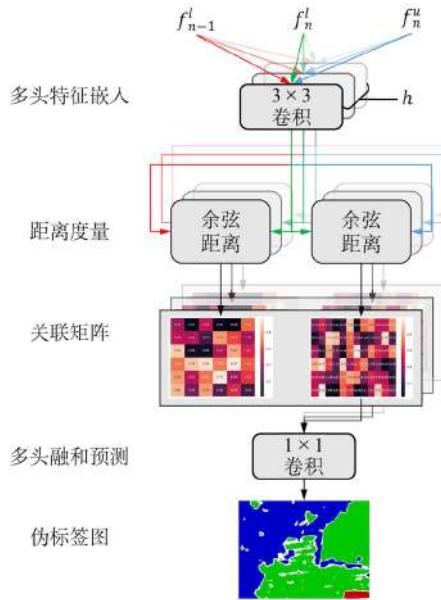


图 4.3: PTS 网络辅助分支结构示意图。辅助分支使用当前训练批次和前一训练批次样本中主分支提取的特征生成伪标签。 f_{n-1} 表示第 $n-1$ 个训练批次中主分支提取的特征集合。 f_n 表示第 n 个训练批次中主分支提取的特征集合。 f_{n-1}^l 表示第 $n-1$ 个训练批次中像素点有划线标注的特征集合。 f_n^l 表示第 n 个训练批次中像素点有划线标注的特征集合。 f_n^u 表示第 n 个训练批次中像素点无划线标注的特征集合。 h 表示辅助分支中预测头的数量。

助分支的计算。在第一个训练批次中，作为特例，令 $f_{-1}^l = f_0^l$ 。

对于主分支提取的材料图像特征， f^{mas} 由不同层次的特征信息组成。虽然该特征可用于有效地区分特征的类别，但却很难定量地度量特征之间的差异。为了使生成的伪标签更加可靠，并更准确地确定未标注像素的类别。辅助分支有必要在使用特征计算前对其进行可学习的特征嵌入。辅助分支包含 h 个预测头，每个预测头均接收 f_{n-1}^l 、 f_n^l 和 f_n^u 作为输入进行特征嵌入工作，使来自主分支不同层次的特征互相融和并处于一个适合进行度量的特征空间中。对于从主分支提取的特征 f_{n-1}^l 、 f_n^l 和 f_n^u ，辅助分支的每个预测头使用卷积核大小为 3 的卷积操作将这三种类型的特征嵌入至 64 维特征空间中。由于不同预测头的卷积相互独立，这允许每个预测头进行不同的特征嵌入，有助于更好地描述不同类别像素之间的关系。此外，每个特征嵌入头在运算中还实现了对 Unet++ 网络输出特征的融和。如图4.4所示，在主分支提取的浅层特征中，较好且稳定地保留了晶粒边界信息；而在主分支提取的深层特征中，可以较好地描述不同类别物相之间的边界信息。因此，从多层次特征中获

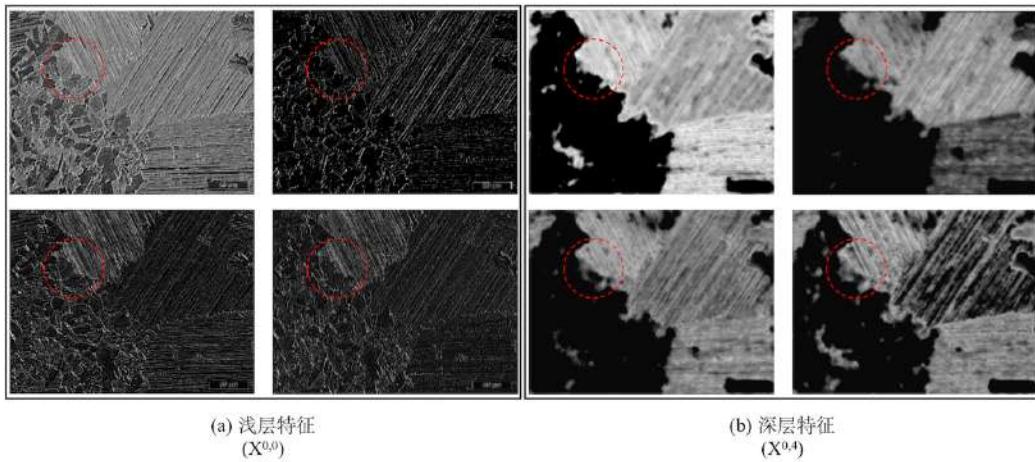


图 4.4: 主分支提取的不同层次特征的可视化实例。(a) 和 (b) 分别展示了图4.1 $X^{0,0}$ 和 $X^{0,4}$ 模块中不同通道的特征图可视化结果。红色虚线圆标注了钛合金图像中的一处不同类别物相之间的边界。特征之间的差异表现为相应像素点的灰度值差值。在浅层特征中，晶粒边界信息更为精确和稳定；而在深层特征中，不同类别的物相之间的边界更加明显。

得具有准确边界与范围的伪标签是合理的。图4.4展示了钛合金图像的主分支特征可视化实例，其中浅层特征取自 $X^{0,0}$ 模块，深层特征取自 $X^{0,4}$ 模块。

在将主分支提取的特征嵌入到合适的特征空间后，辅助分支使用余弦距离对嵌入后的 f_{n-1}^l 、 f_n^l 和 f_u 特征进行距离度量，用来确定不同像素之间的类别关系，从而进一步确定未标注像素的类别。由于两个特征向量 A 、 B 的关联性可以用向量夹角的余弦值进行描述，并且余弦距离具有一个有限的范围（在 0 到 2 之间），因此多组特征的关系方便被量化地进行比较。余弦距离的计算公式如式 (4.3) 所示：

$$d_{cos}(A, B) = 1 - s_{cos}(A, B). \quad (4.3)$$

公式 (4.3) 中的余弦相似度 $s_{cos}(A, B)$ 可由下式计算得：

$$s_{cos}(A, B) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}, \quad (4.4)$$

其中 n 表示特征向量的维度， A_i 和 B_i 分别表示特征向量 A 和特征向量 B 中的元素。由于主分支提取的特征被嵌入至 64 维特征空间中，因此每个像素所对应的嵌入后特征均由 64 个元素组成。两个嵌入后特征之间的余弦距离可以根据公式 (4.3) 和公式 (4.4) 计算得出。考虑到余弦距离的上述特性，使用余弦距离对嵌入后的特征计算关联矩阵 mat_1 和 mat_2 ，实现空间位置无关地描述不同像素之间的类别关系。

关联矩阵 mat_1 描述了前一训练批次样本的划线标注像素与当前训练批次样本的划线标注像素之间的类别关系。关联矩阵 mat_2 描述了当前训练批次样本的划线标注像素与未标注像素之间的类别关系。

为了更加充分地利用有限的划线标注信息，选择更合适的特征空间进行距离度量，本方法基于关联矩阵 mat_1 对特征嵌入进行监督，设计了一个上下文特征差异监督损失。上下文特征差异监督损失对相邻训练批次样本中划线标注像素的嵌入特征构建关联，通过优化相邻训练批次中划线标注像素经过特征嵌入后的余弦距离，调整特征嵌入的目标空间，实现对模型泛化能力的显式学习与监督，这是 PTS 网络在小样本情况下有效的一个重要原因。通过计算相邻训练批次划线标注像素嵌入特征之间的余弦距离，可以获得关联矩阵 mat_1 ，并预测成对像素点的类别是否相同。通过对比划线标注像素点的真实类别计算出上下文特征差异监督损失，并用于指导特征嵌入的学习。基于公式 (4.3)，相邻训练批次划线标注像素嵌入特征之间的余弦距离 $dist_{cos}$ 可以用如下公式表示：

$$dist_{cos}(i, j) = d_{cos}(f_{\theta 1}(x_i), f_{\theta 1}(x_j)), \quad (4.5)$$

其中 x_i 与 x_j 表示像素 i 与像素 j 由主分支提取的特征， $f_{\theta 1}(\cdot)$ 表示对特征进行特征嵌入的结果。通过比较不同像素的类别关系与余弦距离，像素类别 k 的上下文特征差异监督损失 L_k 可以用如下公式表示：

$$L_k = \frac{1}{|D_n| \times |D_{n-1}^k|} \sum_{i \in D_n} \sum_{j \in D_{n-1}^k} MSE(dist_{cos}(i, j), a_{ij}), \quad (4.6)$$

其中 a_{ij} 表示像素 i 与像素 j 的类别关系，相同类别为 0，不同类别为 2， MSE 表示均方误差函数， D_n 表示第 n 个训练批次样本的划线标注像素集合， D_{n-1}^k 表示第 $n-1$ 个训练批次样本中划线标注类别为 k 的像素集合， $|D_n|$ 表示 D_n 的元素数量， $|D_{n-1}^k|$ 表示 D_{n-1}^k 的元素数量。由于不同类别的划线标注像素点数量可能是不均衡的，为了确保划线标注像素点数量较少的类别也能得到足够的关注，完整的上下文特征差异监督损失 L_{cbl} 使用公式 (4.7) 对各个像素类别的上下文特征差异监督损失

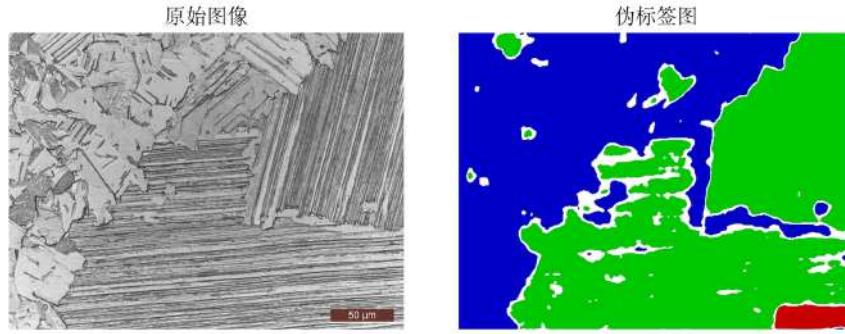


图 4.5: 钛合金图像的伪标签可视化实例。白色区域表示低置信度的像素点, 而绿色、蓝色和红色区域则对应了三种不同类别的高置信度像素点。

进行缩放后计算得出:

$$L_{cbl} = \sum_{k=0}^{N-1} \frac{|D_{n-1}|}{|D_{n-1}^k|} L_k, \quad (4.7)$$

其中 D_{n-1} 表示第 $n-1$ 个训练批次样本的划线标注像素集合, $|D_{n-1}|$ 表示 D_{n-1} 的元素数量, N 表示分割的类别数。

辅助分支中的每个预测头, 对于当前训练批次划线标注像素嵌入特征与未标注像素嵌入特征之间相关性的描述, 即关联矩阵 mat_2 , 决定了伪标签的生成结果。根据嵌入后的特征, 伪标签中的每个像素被确定为与当前训练批次划线标注像素最接近的类别。为了综合考虑不同预测头的结果, 辅助分支使用多头融和卷积合并每个预测头计算出的伪标签。忽略合并后伪标签中低置信度的像素点, 即可得到辅助分支的输出伪标签。图4.5展示了 PTS 网络对钛合金图像的伪标签预测可视化结果, 其中白色区域代表伪标签中的低置信度像素点, 而绿色、蓝色和红色区域分别对应了三种不同类别的高置信度像素点。

对于生成的伪标签, 辅助分支使用划线标注的交叉熵损失进行监督, 对特征嵌入卷积和多头融和卷积进行优化, 以确保生成伪标签的有效性。因此, 整个辅助分支的损失函数 L_{aux} 由上下文特征差异监督损失与划线标注的交叉熵损失求和得到, 如式 (4.8) 所示:

$$L_{aux} = L_{cbl} + \frac{1}{M} \sum_{i=0}^{M-1} C(y_i, f_{\theta 2}(x_i)), \quad (4.8)$$

其中 x_i 表示像素 i 的特征, y_i 表示像素 i 的真实类别, $f_{\theta 2}(\cdot)$ 表示辅助分支对像素

点类别的预测结果, $C(\cdot)$ 是交叉熵损失函数, M 表示当前训练批次中划线标注的像素点数量。

4.1.4 双分支融和

PTS 网络的主分支和辅助分支在训练中采用了分阶段冻结参数的策略进行训练, 改善分割结果并融和网络双分支。在训练材料图像分割模型的过程中, 对于一次划线标注的图像数据, PTS 网络需进行 120 个轮次的训练, 可分为三个阶段。第一个阶段是训练的前 50 个轮次。在这个阶段, 只有主分支的参数被优化更新以完成图像领域的迁移。出于训练图像数量的考虑, 本方法冻结了主分支骨干网络 Vgg16 的 $X^{0,0}$ 、 $X^{1,0}$ 、 $X^{2,0}$ 和 $X^{3,0}$ 四个模块的参数, 仅对 $X^{4,0}$ 模块的参数根据划线标注进行优化。第二个阶段, 将辅助分支加入 PTS 网络训练 50 个轮次。在这一阶段中, 主分支与辅助分支并行地进行训练。主分支使用划线标注进行监督, 主分支提取的特征被输入到辅助分支中, 而辅助分支所生成的伪标签暂不用于对主分支的参数进行监督更新。第二阶段的训练实现了特征嵌入卷积与多头融和卷积的预热, 这确保了用于监督主分支的伪标签是生成良好的。在最后一个阶段, PTS 网络被训练 20 个轮次。在这一阶段中, 进一步冻结了主分支骨干网络 Vgg16 的 $X^{4,0}$ 模块, 并使用生成的伪标签和划线标注同时监督主分支的分割结果训练 20 个轮次。第三个训练阶段使主分支获取辅助分支中的信息, 实现 PTS 网络双分支的融和, 有助于改善模型分割精度与泛化能力。

伴随着主分支参数的不断更新, 主分支提取的特征也会不断变化, 这可能导致辅助分支中学习的特征嵌入规则失效。特征嵌入规则的失效会使得生成的伪标签出现明显的错误或突变, 并导致主分支学习到错误的信息。为了避免辅助分支特征嵌入规则的失效, 在第二阶段与第三阶段的训练过程中, PTS 网络每一轮次的训练包含一次对主分支的训练迭代和十次对辅助分支的训练迭代。

伪标签的准确性和有效性是决定辅助分支能否改善主分支分割结果的关键。为了避免生成的伪标签中潜在错误预测的干扰, 在使用伪标签监督主分支分割结果的同时, 使用划线标注对主分支分割结果进行监督是必要的。为此, 本方法设计了一

一个复合监督损失函数将伪标签的信息渗透到主分支中，使得主分支的预测结果同时被划线标注像素点和伪标签两部分监督。主分支复合监督损失函数 L_{mas} 公式如下所示：

$$L_{mas} = L_{rea} + L_{pse} - \frac{L_{rea} \cdot L_{pse}}{L_{rea} + L_{pse}}, \quad (4.9)$$

其中 L_{mas} 表示主分支的复合监督损失， L_{rea} 表示划线标注损失， L_{pse} 表示伪标签损失。 L_{rea} 和 L_{pse} 分别可表示为以下形式：

$$L_{rea} = \frac{1}{M} \sum_{i=0}^{M-1} C(y_i^r, f_{\theta 3}(x_i)), \quad (4.10)$$

$$L_{pse} = \frac{\lambda}{T} \sum_{i=0}^{T-1} C(y_i^p, f_{\theta 3}(x_i)), \quad (4.11)$$

其中 x_i 表示像素 i 的特征， y_i^p 表示伪标签中像素 i 的类别， y_i^r 表示划线标注像素点 i 的类别， $f_{\theta 3}(\cdot)$ 表示主分支网络的输出预测类别， $C(\cdot)$ 是交叉熵损失函数， M 表示当前训练批次中划线标注的像素点数量， T 表示生成的伪标签中高置信度像素的数量， λ 是用于平衡 L_{rea} 和 L_{pse} 的比例因子，在实验中被设置为 3。

主分支的复合监督损失函数 L_{mas} 可以对两部分损失实现自适应地综合优化。为了说明公式 (4.9) 设计的合理性与必要性，使用公式 (4.9) 中的部分项，即公式 (4.12) 作为损失函数 L_{cmp} ，和复合监督损失函数 L_{mas} 进行比较。

$$L_{cmp} = L_{rea} + L_{pse}. \quad (4.12)$$

为了说明损失函数 L_{mas} 和损失函数 L_{cmp} 之间的差别，定义以下形式：

$$E_f = (L_{rea} - \Delta L_1) + (L_{pse} - \Delta L_2). \quad (4.13)$$

$$E_s = \frac{(L_{rea} - \Delta L_1) \cdot (L_{pse} - \Delta L_2)}{(L_{rea} - \Delta L_1) + (L_{pse} - \Delta L_2)}. \quad (4.14)$$

其中 ΔL_1 表示根据划线标注优化减少的损失值， ΔL_2 表示根据伪标签优化减少的损失值。为了便于比较 L_{mas} 和 L_{cmp} ，在此令 $\Delta L_1 + \Delta L_2 = 2\Delta L$ ，即为一定值。

尽管 ΔL_1 和 ΔL_2 不可能取值为 0，但为了突出显示 L_{mas} 和 L_{cmp} 变化的规律，对公式 (4.13) 和公式 (4.14)，令 ΔL_1 分别取值为 ΔL 、 $2\Delta L$ 和 0，相应的， ΔL_2 分

别取值为 ΔL 、0 和 $2\Delta L$ 。这样就可以得到式 (4.15-4.20):

$$E_{f1} = (L_{rea} - \Delta L) + (L_{pse} - \Delta L). \quad (4.15)$$

$$E_{f2} = (L_{rea} - 2\Delta L) + L_{pse}. \quad (4.16)$$

$$E_{f3} = L_{rea} + (L_{pse} - 2\Delta L). \quad (4.17)$$

$$E_{s1} = \frac{(L_{rea} - \Delta L) \cdot (L_{pse} - \Delta L)}{(L_{rea} - \Delta L) + (L_{pse} - \Delta L)}. \quad (4.18)$$

$$E_{s2} = \frac{(L_{rea} - 2\Delta L) \cdot L_{pse}}{(L_{rea} - 2\Delta L) + L_{pse}}. \quad (4.19)$$

$$E_{s3} = \frac{L_{rea} \cdot (L_{pse} - 2\Delta L)}{L_{rea} + (L_{pse} - 2\Delta L)}. \quad (4.20)$$

对于损失函数 L_{cmp} , 综合优化两部分损失项和单独优化某一部分损失项是等效的, 如下式所示:

$$L_{cmp} = E_{f1} = E_{f2} = E_{f3}. \quad (4.21)$$

不同于损失函数 L_{cmp} , 本方法所采用的损失函数 L_{mas} 在优化不同损失项的过程中并非等效的。当划线标注的损失与伪标签的损失二者相接近时, 即 $L_{rea} \approx L_{pse}$, 同时对划线标注与伪标签进行优化比优先优化划线标注更有可能取得较好的分割结果, 即损失函数 L_{mas} 满足如下关系:

$$L_{mas} = E_{f1} - E_{s1} < E_{f2} - E_{s2}, \quad (4.22)$$

$$L_{mas} = E_{f1} - E_{s1} < E_{f3} - E_{s3}. \quad (4.23)$$

当伪标签的损失函数较大而划线标注的损失函数较小时, 即 $L_{pse} \gg L_{rea}$, 损失函数 L_{mas} 优先根据伪标签进行优化有助于关注更多像素的特征信息:

$$L_{mas} = E_{f3} - E_{s3} < E_{f1} - E_{s1}. \quad (4.24)$$

当划线标注的损失函数较大而伪标签的损失函数较小时, 即 $L_{rea} \gg L_{pse}$, 损失函数 L_{mas} 优先根据划线标注进行优化有助于避免生成的伪标签中潜在错误预测的干扰:

$$L_{mas} = E_{f2} - E_{s2} < E_{f1} - E_{s1}. \quad (4.25)$$

由于损失函数 L_{mas} 的形式在优化过程中具备如上的大小关系，因此采用损失函数 L_{mas} 有助于达成根据两部分损失项实现自适应地综合优化，从而得到更小的损失值。

由于公式 (4.9) 中的复合监督损失函数设计，主分支将根据划线标注和伪标签自适应地进行优化，以改善分割结果。因此，主分支中的划线标注和伪标签两种监督可以得到平衡，这有助于进一步提升训练所得模型的分割性能。通过在第三阶段的训练中，对主分支使用伪标签进行监督，PTS 网络实现了隐式地将辅助分支中的信息传播到了主分支中。在测试阶段，PTS 网络仅依赖主分支进行预测，可以实现没有显卡设备的电脑上进行快速的分割。

4.2 实验分析

为了验证本方法的有效性和普适性，本章使用 PTS 网络分别对钛合金 $TiAl$ 、陶瓷共晶 $HfB_2 - B_4C$ 以及碳钢 (UHCSDB)[9] 三类材料图像进行了实验。其中钛合金图像包含等轴组织（蓝色）、片层组织（绿色）以及比例尺（红色）三相；陶瓷图像包含致密结构（绿色）与非致密结构（蓝色）两相；碳钢图像中包含珠光体（蓝色）和球状渗碳体（绿色）两相。PTS 网络使用 640×480 大小的 1 至 2 张图像进行训练，并使用 3 至 11 张图片进行测试。训练图像中的每个类别由一条领域专家绘制的曲线进行标注，曲线的绘制综合考虑了物相的典型特征和空间分布情况，每张图像中的标注像素比例不超过 0.2%。图4.6展示了钛合金图像、陶瓷图像和碳钢图像训练所使用的标注样例。

4.2.1 消融实验

为了证明 PTS 网络各部分的有效性和必要性，在实验中分别对分阶段冻结参数、数据增强、预训练骨干网络、生成伪标签和上下文特征差异监督损失进行了消融实验。此外，还将 PTS 网络分割结果与基于完全标注使用分阶段冻结参数、数据增强以及预训练骨干网络的全监督基准网络（原始 Unet++ 网络）分割结果进行

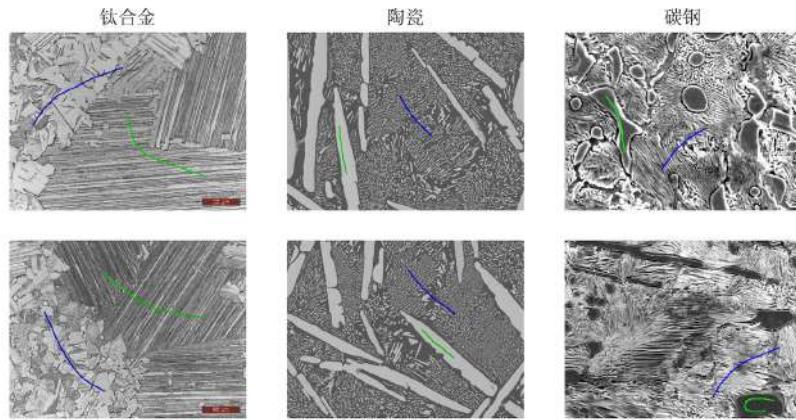


图 4.6: 钛合金图像、陶瓷图像和碳钢图像的划线标注示例图

比较。钛合金数据集的消融实验结果如表4.1和图4.7所示。通过观察实验结果可以发现，在逐步添加分阶段冻结参数、数据增强、预训练骨干网络、生成伪标签和上下文特征差异监督损失的过程中，分割结果的更多细节得到了保留，分割错误的像素不断减少。

表 4.1: PTS 网络在钛合金数据集上的消融实验结果

模型	mIoU(%)		mDice(%)	
	一张训练图像	两张训练图像	一张训练图像	两张训练图像
原始 Unet++	30.3	37.0	41.8	50.8
+ 分阶段冻结参数	31.6	55.7	42.8	69.7
+ 数据增强	58.9	63.2	71.9	76.0
+ 预训练骨干网络	70.2	73.6	79.7	83.5
+ 伪标签	73.4	75.9	83.2	85.5
+ 上下文特征差异监督损失	76.2	78.4	85.5	87.9
全监督	84.9	89.2	91.5	94.0

在钛合金数据集中，分阶段冻结参数的策略在使用一张和两张训练图片进行训练时可以带来了 1.3% 和 18.7% 的 mIoU 提升与 1.0% 和 18.9% 的 mDice 提升。分阶段冻结参数的训练策略将分割任务拆分为多个步骤，如学习特征表示和基于提取的特征进行分割预测，这有助于克服样本数量不足的问题。然而，仅依靠分阶段冻结参数策略并不能完全解决样本及标注数据不足的困扰。为此，PTS 网络进一步采

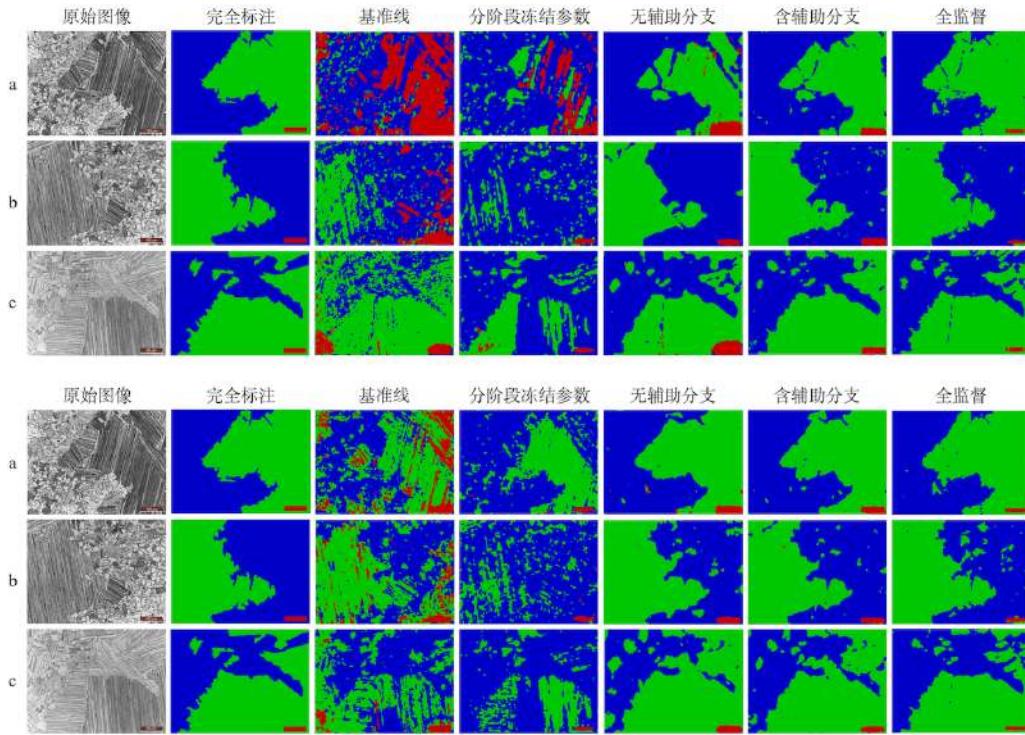


图 4.7: 钛合金数据集上的消融实验分割结果。前三行为基于一张钛合金图像进行训练得到的分割结果;后三行为基于两张钛合金图像进行训练得到的分割结果。图中的七列分别对应原始图像、完全标注、基准网络(原始 Unet++ 网络)、基准网络 + 分阶段冻结参数、不含辅助分支的 PTS 网络、含辅助分支的 PTS 网络以及全监督 Unet++ 网络的分割结果。

用了两种常用的策略,即使用数据增强和预训练骨干网络。使用数据增强与预训练骨干网络能够分别在使用一张图片进行训练和使用两张图片训练的情况下,为模型带来 38.6% 和 17.9% 的 mIoU 提升与 36.9% 和 13.8% 的 mDice 提升。这一趋势表明,通过扩充训练数据和迁移自然场景图像的通用特征表示信息,引入数据增强和预训练骨干网络有助于缓解样本和标注数据不足的问题。在对主分支应用上述训练策略后,即便训练样本与标注数据并不充足,分割模型依然能够收敛。不过,在分割结果中仍然能够观察到部分分割错误的区域,可以进一步完善。

针对生成伪标签的辅助分支消融实验,在钛合金数据集中,采用生成伪标签策略在使用一张图片进行训练和使用两张图片进行训练时,可以为最终分割结果的 mIoU 分别带来 3.2% 和 2.3% 的提升,对 mDice 带来 3.5% 和 2% 的提升。这说明,辅助分支生成的伪标签可以为主分支提供额外的监督信息。当伪标签的生成

过程仅由划线标注监督时，生成伪标签和分割预测所使用的监督信息并不存在区别。因此，仅基于划线标注生成的伪标签不足以认为分割结果带来稳定的改善。为了使生成的伪标签包含除了划线标注以外的隐式像素关联信息，辅助分支使用上下文特征差异监督损失来确定用于生成伪标签的目标特征空间。在辅助分支中添加上下文特征差异监督损失可以为使用一张图片训练和使用两张图片训练的分割结果 mIoU 带来 2.8% 和 2.5% 的提升，mDice 带来 2.3% 和 2.4% 的提升。因此，在添加了上下文特征差异监督损失后，分割结果的准确性得到了稳定的改善。这种改善得益于使用上下文特征差异监督损失监督生成的伪标签包含了隐式的像素关联信息，这些信息可能并未包含于划线标注中。

相比于使用分阶段冻结参数、数据增强和预训练骨干网络进行全监督训练的 Unet++ 网络，当使用一张图像和两张图像进行训练时，PTS 网络的分割结果在 mIoU 上分别低 8.7% 和 10.8%，在 mDice 上低 6% 和 6.1%。PTS 网络的分割精度约为全监督分割结果的 91%。考虑到实验中的划线标注在每张图像上仅包含低于 0.2% 的像素标注，该结果是可以接受的。在陶瓷数据集与碳钢数据集的实验中，同样的增益也可以被观察到，如图4.8、图4.9和表4.2、表4.3所示。这些结果充分证明了 PTS 网络中的各种结构与策略在小样本材料微观图像的分割工作中可以起到明显的改善作用，且各个结构的存在都具有必要性。

表 4.2: PTS 网络在陶瓷数据集上的消融实验结果

模型	mIoU(%)		mDice(%)	
	一张训练图像	两张训练图像	一张训练图像	两张训练图像
原始 Unet++	27.5	71.5	41.9	82.5
+ 分阶段冻结参数	49.4	71.8	65.7	82.2
+ 数据增强	61.1	80.3	74.6	88.9
+ 预训练骨干网络	80.5	83.7	88.9	91.0
+ 伪标签	80.7	82.9	89.0	90.5
+ 上下文特征差异监督损失	83.4	85.9	90.7	92.2
全监督	95.7	96.6	97.9	98.3

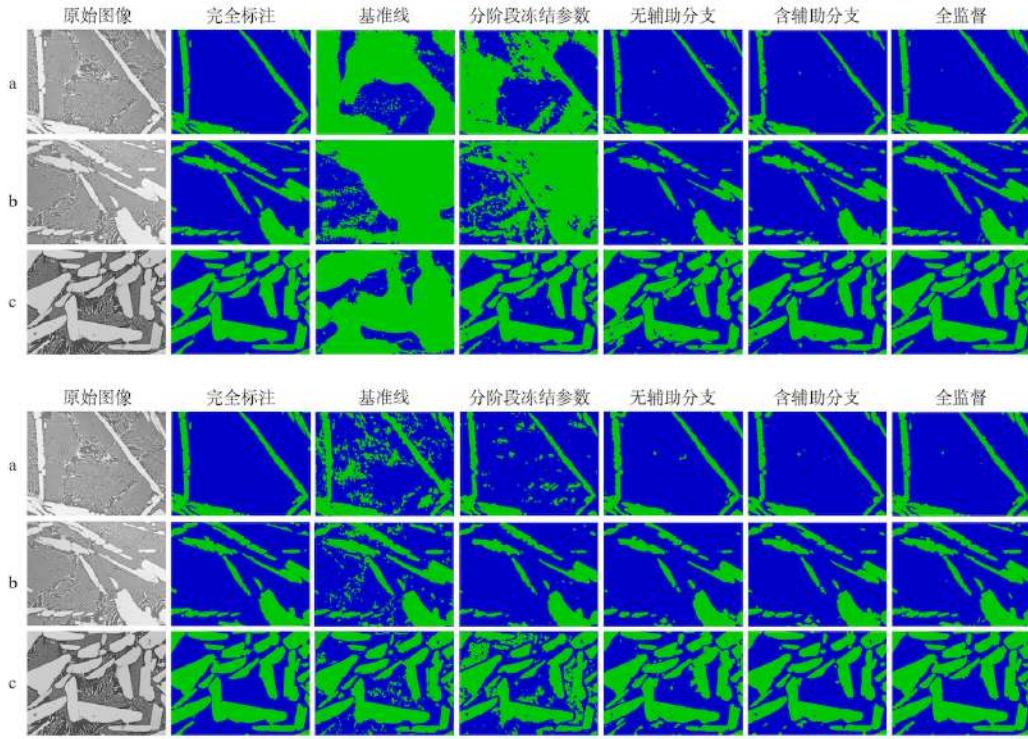


图 4.8: 陶瓷数据集上的消融实验分割结果。前三行为基于一张陶瓷图像进行训练得到的分割结果；后三行为基于两张陶瓷图像进行训练得到的分割结果。图中的七列分别对应原始图像、完全标注、基准网络（原始 Unet++ 网络）、基准网络 + 分阶段冻结参数、不含辅助分支的 PTS 网络、完整的 PTS 网络以及全监督 Unet++ 网络的分割结果。

表 4.3: PTS 网络在碳钢数据集上的消融实验结果

模型	mIoU(%)		mDice(%)	
	一张训练图像	两张训练图像	一张训练图像	两张训练图像
原始 Unet++	31.5	44.8	45.6	58.7
+ 分阶段冻结参数	42.3	46.8	56.4	59.8
+ 数据增强	59.1	61.9	71.1	70.6
+ 预训练骨干网络	68.0	68.2	79.3	79.6
+ 伪标签	70.5	72.8	81.1	82.7
+ 上下文特征差异监督损失	72.5	73.6	82.3	83.2
全监督	85.1	86.2	91.4	92.4

在陶瓷数据集使用两幅图像进行训练的实验中可以观察到，当伪标签生成过程中缺少了上下文特征差异监督损失，则不使用伪标签有可能比使用伪标签得到更好

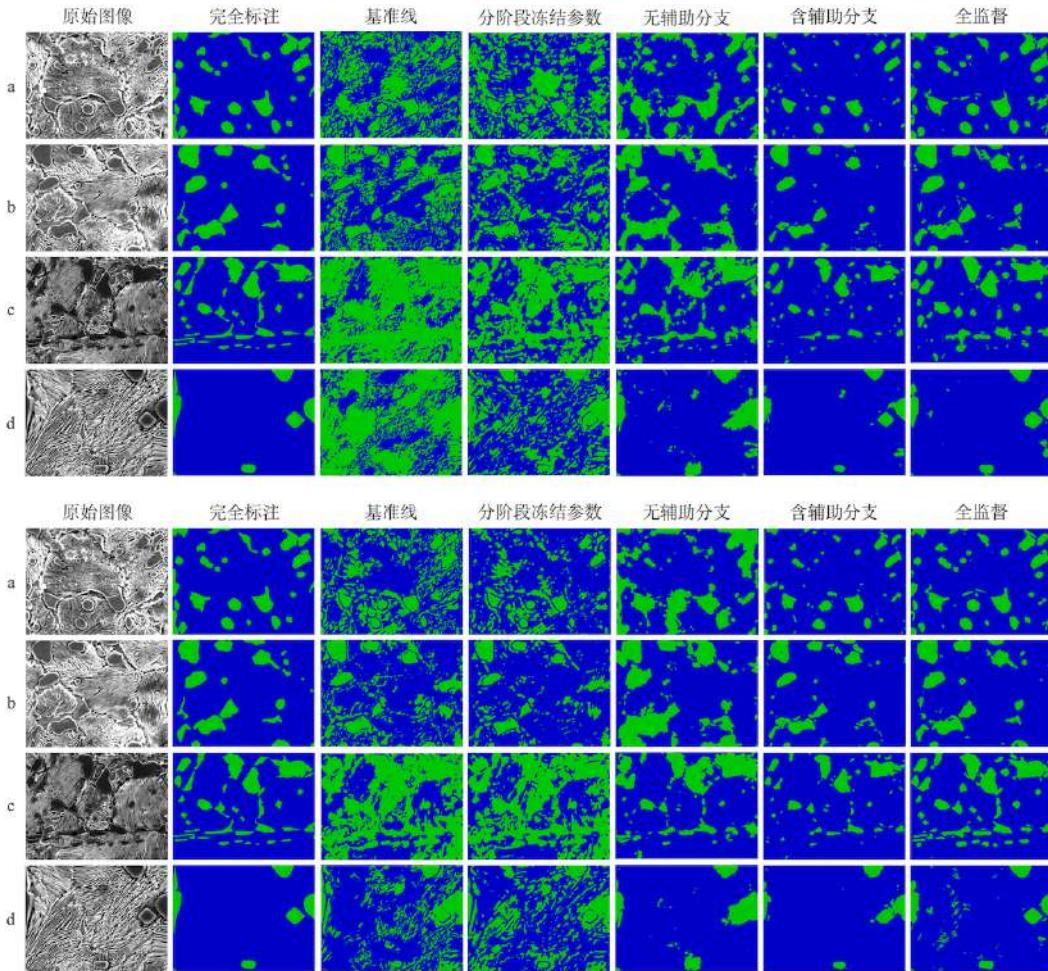


图 4.9: 碳钢数据集上的消融实验分割结果。前四行为基于一张碳钢图像进行训练得到的分割结果；后四行为基于两张碳钢图像进行训练得到的分割结果。图中的七列分别对应原始图像、完全标注、基准网络（原始 Unet++ 网络）、基准网络 + 分阶段冻结参数、不含辅助分支的 PTS 网络、完整的 PTS 网络以及全监督 Unet++ 网络的分割结果。

的分割结果。这是一个合理的现象，因为缺乏上下文特征差异监督损失的伪标签，在生成过程中没有获得额外的监督信息。因此，在对主分支的监督中，使用缺少上下文特征差异监督损失监督而生成的伪标签和使用主分支的分割结果（这代表经典的自训练策略）在获得监督信息方面没有本质的区别。实验结果见表4.4、表4.5和表4.6。

在实验中，我们采用了六种不同的经典骨干网络替换 Vgg16 进行实验，并比较使用一张钛合金图像进行训练得到的分割结果，从而进一步说明辅助分支对分割工作起到的积极作用。根据图4.10中的结果可以发现，尽管使用不同骨干网络取得的分

割结果存在明显的区别，但是每个模型在添加辅助分支后都可以取得 1.2%~6% 的 mIoU 提升，详细数据如表4.7所示。这证明 PTS 网络设计的辅助分支对于不同网络结构具有鲁棒性。由于 PTS 网络采用分阶段冻结参数策略进行训练，相对先进和新颖的骨干网络，如 ResNet、DenseNet 和 Inception，分割精度并未能够超过 Vgg 骨干网络。造成这一现象的最主要原因是跳跃连接等策略为单一阶段的网络结构所带来的增益非常有限。此外，单一阶段具有更多参数的骨干网络通常会表现得更好。由于上述原因的影响，Vgg 骨干网络在实验中的经典骨干网络中表现最好，因此被选用于 PTS 网络中。

表 4.4: PTS 网络采用伪标签策略与自训练策略在钛合金数据集上的分割结果

模型	mIoU(%)	
	一张训练图像	两张训练图像
自训练策略	72.3	75.0
伪标签策略	73.4	75.9
+ 上下文特征差异监督损失	76.2	78.4

表 4.5: PTS 网络采用伪标签策略与自训练策略在陶瓷数据集上的分割结果

模型	mIoU(%)	
	一张训练图像	两张训练图像
自训练策略	81.7	83.0
伪标签策略	80.7	82.9
+ 上下文特征差异监督损失	83.4	85.9

表 4.6: PTS 网络采用伪标签策略与自训练策略在碳钢数据集上的分割结果

模型	mIoU(%)	
	一张训练图像	两张训练图像
自训练策略	69.0	69.8
伪标签策略	70.5	72.8
+ 上下文特征差异监督损失	72.5	73.6

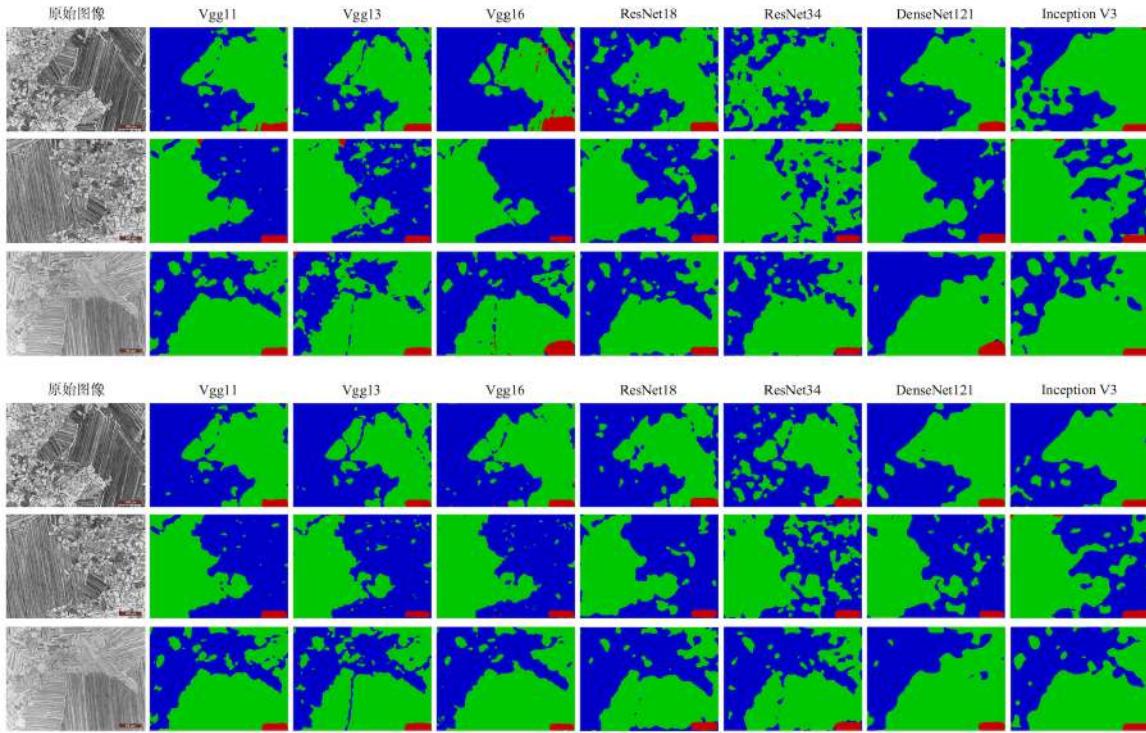


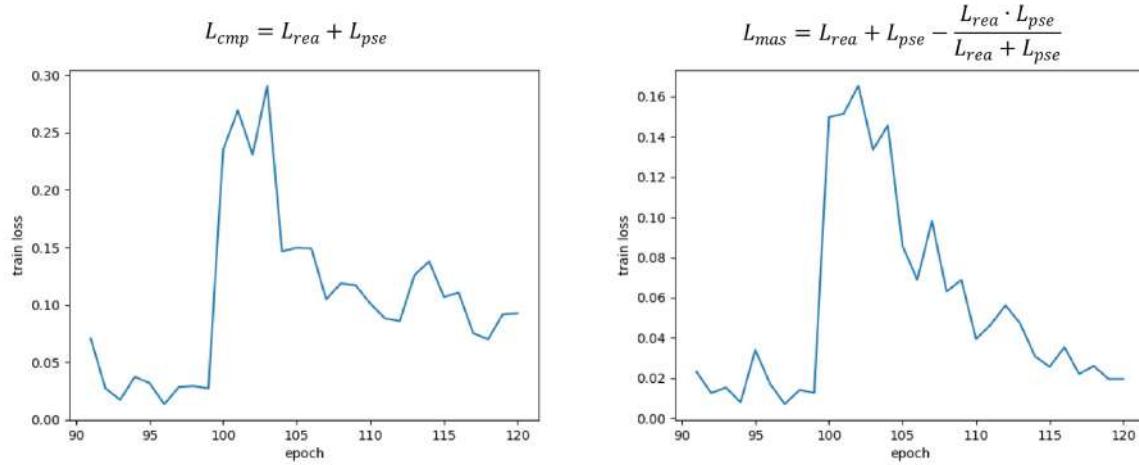
图 4.10: 基于不同骨干网络的 PTS 网络分割结果。前三行为不使用辅助分支时的分割结果；后三行为应用辅助分支后的分割结果。

表 4.7: 基于不同骨干网络的 PTS 网络在钛合金数据集上的分割结果

骨干网络	无辅助分支 (mIoU%/mDice%)	含辅助分支 (mIoU%/mDice%)
ResNet18	62.6/75.2	68.2/80.2
ResNet34	61.8/75.6	66.8/79.1
DenseNet121	67.8/79.6	69.0/81.1
Inception V3	62.0/75.7	65.7/77.9
Vgg11	70.6/81.1	73.7/83.8
Vgg13	71.0/81.8	76.0/85.4
Vgg16	70.2/79.7	76.2/85.5

表 4.8: PTS 网络在钛合金图像上对融和损失函数的消融实验结果

损失函数	mIoU(%)
$L_{cmp} = L_{rea} + L_{pse}$	75.3
$L_{mas} = L_{rea} + L_{pse} - \frac{L_{rea} \cdot L_{pse}}{L_{rea} + L_{pse}}$	76.2

图 4.11: 第 90 轮次至第 120 轮次之间 L_{cmp} 和 L_{mas} 的训练损失曲线图。 L_{cmp} 和 L_{mas} 在第 100 轮次至第 120 轮次之间使用。

在主分支中，本方法使用复合监督损失函数 L_{mas} 对主分支进行监督，并根据划线标注和伪标签共同优化分割结果。 L_{mas} 避免了模型仅基于划线标注或伪标签进行有限的优化。因此， L_{mas} 有助于不断优化主分支中的参数以获得更理想的分割结果。为了证明使用该损失函数的有效性，在实验中对比了使用 L_{cmp} (公式 (4.12)) 和 L_{mas} (公式 (4.9)) 作为主分支损失函数的分割精度。实验结果如表4.8和图4.11所示。

此外，为了说明 PTS 网络在小样本训练集上获得的模型具有泛化能力，在实验中使用两幅钛合金图像进行训练，将 PTS 网络的分割结果与使用分阶段冻结参数、数据增强以及预训练骨干网络的全监督 Unet++ 网络分割结果进行比较，实验结果如图4.12所示。尽管 PTS 网络在每张图像中仅含 0.2% 的标注像素进行训练，但其分割效果与全监督模型的分割效果并没有显著的差异，并且基于这两种方法获得的分割边界基本一致。这证明了 PTS 网络能够基于非常有限的训练样本和稀疏的划线标注，获得一个具有泛化能力的分割模型。

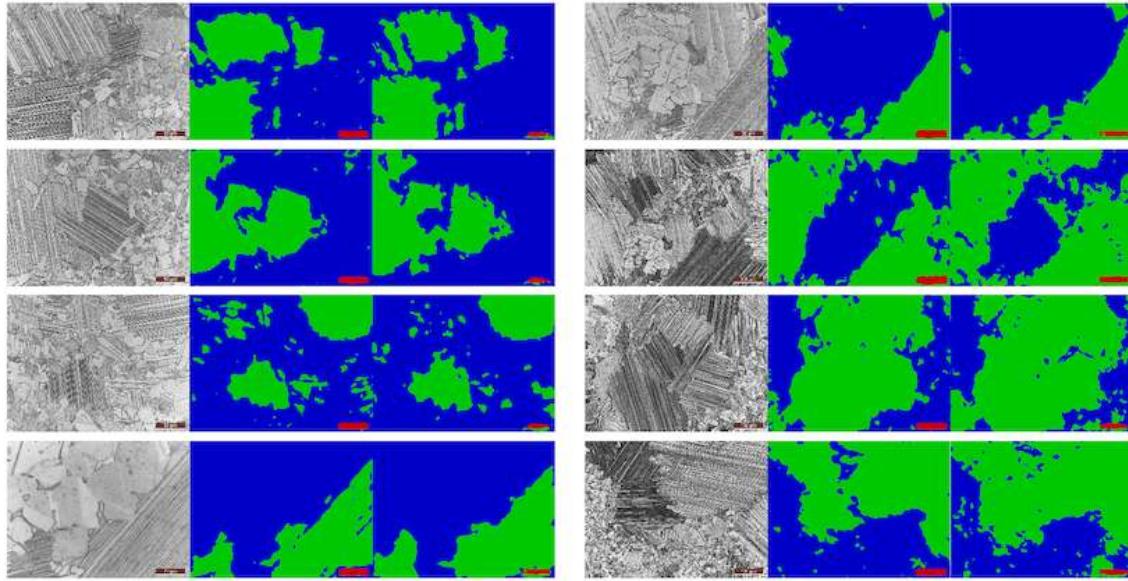


图 4.12: PTS 网络和全监督 Unet++ 网络在更多钛合金图像上的分割结果对比。每组三张图像中, 左侧为原始钛合金图像, 中间为 PTS 网络的分割结果, 右侧为全监督 Unet++ 网络的分割结果。

4.2.2 相关方法对比

为了横向比较不同方法对材料图像的分割效果与可用性, 我们将本方法与基于划线标注训练的 Unet++[46]、Unet 3+[47]、Ce-net[107]、SPML[64]、TEL[108] 和上一章节所提出的方法 (CES)[43] 进行比较。分割结果如图4.13、图4.15和图4.16所示。表4.9、表4.10和表4.11展示了分割结果的 mIoU 和 mDice 评价指标。在陶瓷数据集和碳钢数据集中, PTS 网络的分割结果优于其它的方法, 取得了最好的分割结果。在钛合金数据集中, TEL 网络的分割精度评价指标比 PTS 网络更优。但是, PTS 网络对等轴组织 (蓝色) 和片层组织 (绿色) 的分割结果比 TEL 网络更好。

观察图4.13、图4.15和图4.16中的分割结果可以发现, Unet++、Unet 3+ 和 Ce-net 的分割结果中均有较大面积的预测错误, 这证明了常见的全监督图像分割网络无法直接基于划线标注取得有效的分割效果。SPML 的结果丢失了过多的边界细节, 仅保留了极少的空间范围和大致空间位置, 这说明了处理自然场景图像的弱监督分割网络在不经过细致修改的情况下可能难以直接应用于材料图像的分割任务中。与 SPML 相比, TEL 在钛合金图像和陶瓷图像上取得了良好的分割结果。然而, TEL

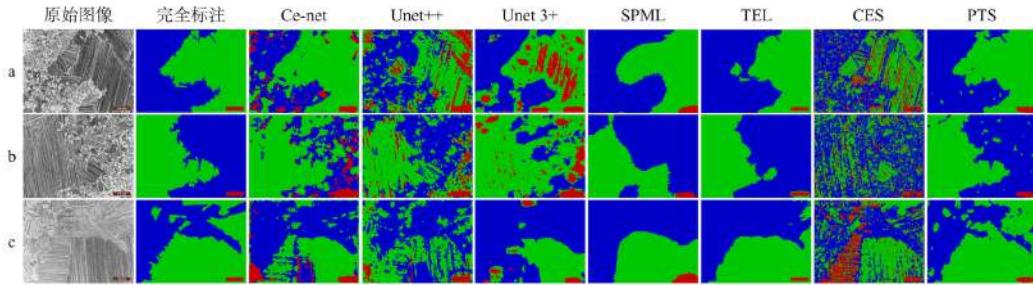


图 4.13: Ce-net、Unet++、Unet 3+、SPML、TEL、CES 和 PTS 在钛合金数据集上基于划线标注的分割结果。为了定量地比较 CES 与其它方法的分割结果, CES 未进行多轮交互式划线标注。钛合金图像中的三种物相分别以绿色、蓝色和红色表示。

并未能对碳钢图像实现有效的分割, 这表明基于自然场景图像设计的弱监督分割网络在分割不同种类的材料图像时可能存在问题, 对材料图像分割任务缺乏普适性。相较于 PTS 和其它五种深度学习神经网络图像分割方法, 上一章节提出的 CES 由于采用了机器学习方法, 因此在训练和分割的用时上有着显著的优势。为了定量地比较 CES 与其它方法的分割结果, 在本实验中, CES 并未进行多轮交互式划线标注。在对陶瓷和碳钢图像进行分割时, CES 在未经交互式划线标注的情况下仍然表现出了较为良好的分割效果。不过, 在对钛合金图像进行分割时, CES 由于缺少了交互式划线标注进行迭代训练的过程, 其分割结果表现较差。与上述方法相比, 本章提出的 PTS 网络的分割结果具有较大优势, 它的分割结果噪音较少、边界比较准确, 并且能够在不同种类的材料图像上获得有效的分割结果。

此外, 在钛合金数据集上, 我们对比了基于划线标注的 PTS 网络和基于完全标注训练的 FCN[10]、Unet++、Unet 3+、Ce-net、DeepLab V3+[44] 网络, 如图4.14和表4.9所示。在每张训练图仅包含 0.2% 像素标注的情况下, PTS 网络的分割精度仅比使用全监督训练表现最好的 Ce-net 网络低 2% 的 mIoU 和 1% 的 mDice。观察图4.14中的分割结果, Unet++、Unet 3+、Ce-net、DeepLab V3+ 的大部分全监督分割结果中可以观察到各物相的范围与轮廓, 而得益于伪标签的引入, PTS 网络尽管只使用划线标注进行训练, 但是也取得了相近的效果。

为了进一步说明 PTS 网络相比于现有深度学习方法在处理材料图像分割任务上的优势, 我们分别对 PTS 网络与 Unet++ 网络在最终预测层之前的特征图 (即

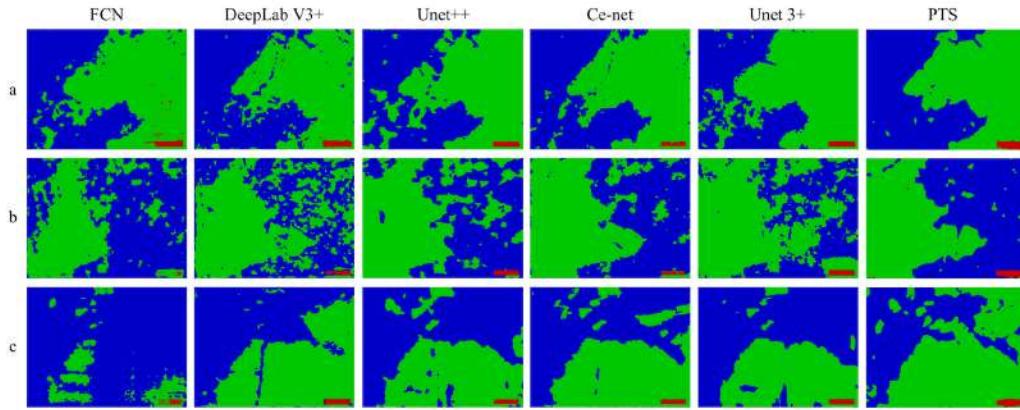


图 4.14: 全监督的 FCN、DeepLab V3+、Unet++、Ce-net 和 Unet 3+ 与基于划线标注的 PTS 在钛合金数据集上的分割结果

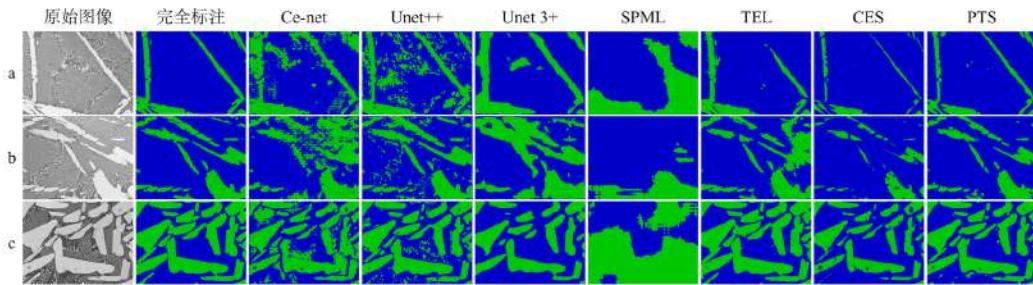


图 4.15: Ce-net、Unet++、Unet 3+、SPML、TEL、CES 和 PTS 在陶瓷数据集上基于划线标注的分割结果。为了定量地比较 CES 与其它方法的分割结果, CES 未进行多轮交互式划线标注。

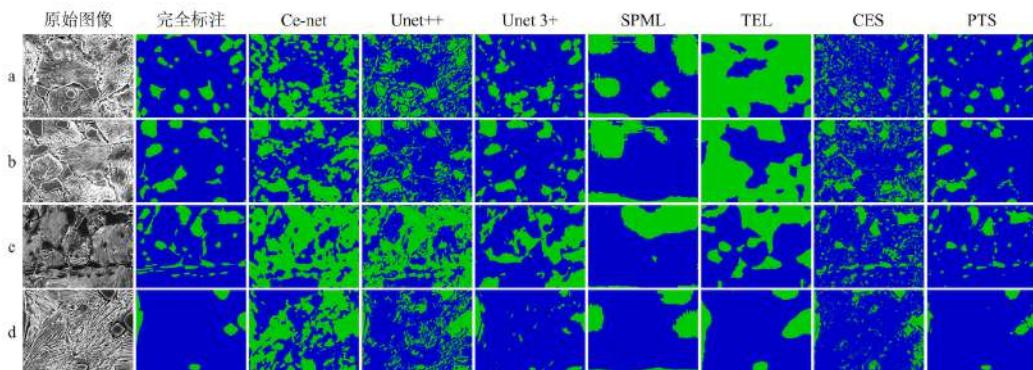


图 4.16: Ce-net、Unet++、Unet 3+、SPML、TEL、CES 和 PTS 在碳钢数据集上基于划线标注的分割结果。为了定量地比较 CES 与其它方法的分割结果, CES 未进行多轮交互式划线标注。

表 4.9: Ce-net、Unet++、Unet3+、SPML、TEL、CES、FCN、DeepLab V3+ 和 PTS 在钛合金数据集上的分割结果。为了定量地比较 CES 与其它方法的分割结果, CES 未进行多轮交互式划线标注。

模型	标注方式	mIoU(%)	mDice(%)
Ce-net	划线标注	46.1	58.2
Unet++	划线标注	37.0	50.8
Unet3+	划线标注	41.6	53.9
SPML	划线标注	62.4	75.4
TEL	划线标注	84.5	91.5
CES	划线标注	32.9	45.3
FCN	完全标注	55.8	68.6
DeepLab V3+	完全标注	75.6	85.6
Unet++	完全标注	78.8	87.8
Ce-net	完全标注	80.4	88.9
Unet 3+	完全标注	75.8	85.6
PTS	划线标注	78.4	87.9

表 4.10: Ce-net、Unet++、Unet3+、SPML、TEL、CES 和 PTS 网络在陶瓷数据集上的分割结果。为了定量地比较 CES 与其它方法的分割结果, CES 未进行多轮交互式划线标注。

模型	标注方式	mIoU(%)	mDice(%)
Ce-net	划线标注	62.5	76.2
Unet++	划线标注	71.5	82.5
Unet3+	划线标注	71.3	82.7
SPML	划线标注	48.1	63.0
TEL	划线标注	85.3	91.7
CES	划线标注	69.5	80.8
PTS	划线标注	85.9	92.2

表 4.11: Ce-net、Unet++、Unet3+、SPML、TEL、CES 和 PTS 网络在碳钢数据集上的分割结果。为了定量地比较 CES 与其它方法的分割结果, CES 未进行多轮交互式划线标注。

模型	标注方式	mIoU(%)	mDice(%)
Ce-net	划线标注	40.7	55.3
Unet++	划线标注	44.8	58.7
Unet3+	划线标注	53.8	65.2
SPML	划线标注	50.6	63.4
TEL	划线标注	44.4	57.9
CES	划线标注	54.3	65.5
PTS	划线标注	73.6	83.2

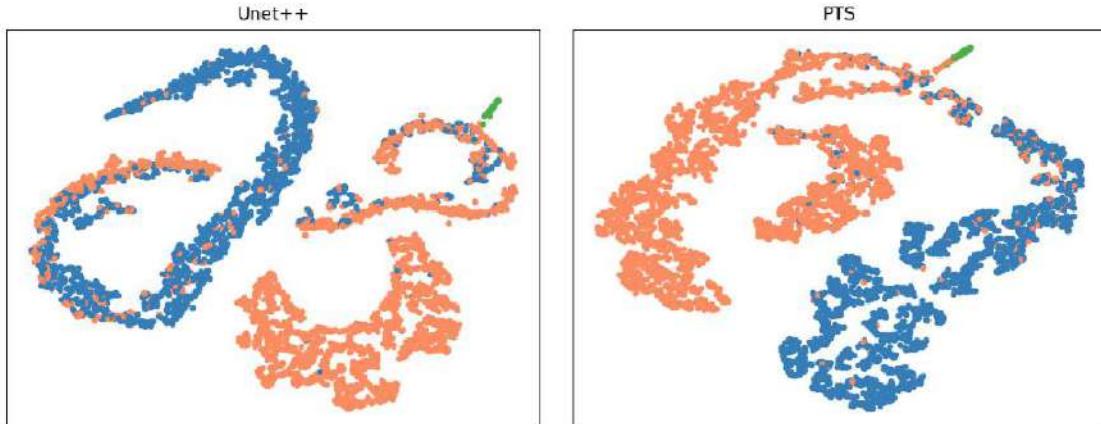


图 4.17: PTS 网络和 Unet++ 网络的特征分布可视化结果

图4.1中的 $X^{0,4}$ 模块的特征图)进行降维并可视化。在对 PTS 网络的特征和 Unet++ 网络的特征分别使用 t-SNE 方法 [109] 进行降维后, 可以得到如图4.17所示的特征分布。图4.17中的蓝色、橙色和绿色分别对应了钛合金微观图像中的等轴组织、片层组织与比例尺三相。通过比较 PTS 网络与 Unet++ 网络的特征分布可以发现, Unet++ 网络的不同类别特征相比于 PTS 网络存在互相混淆的情况。图4.17中 Unet++ 网络的蓝色和橙色特征点的分布相互重叠, 可能限制了 Unet++ 网络最终的分割结果。相比之下, PTS 网络的特征分布不同类别之间的界限更加明显, 有助于更好地进行预测分割。

综上所述，PTS 的分割结果优于其它使用划线标注训练的网络，具有 14.4%~19.8% 的 mIoU 提升和 9.5%~18% 的 mDice 提升。相较于全监督网络，本方法仅比分割结果最好的全监督网络（Ce-net）在 mIoU 和 mDice 两项评价指标上低 2% 和 1%，但在标注的便捷性上，PTS 网络有着显著的优势。本方法之所以在材料图像分割任务中表现较为突出，主要得益于 PTS 中生成的伪标签和设计的双分支网络结构。这两个策略不但关注了基于划线标注进行弱监督图像分割的任务性质，还关注了材料图像这一目标对象的特点，通过选用更为合适且有效技术手段，本方法实现了根据有限标注数据，得到更为精确的分割结果。

4.3 本章小结

本章针对材料图像分割任务中现存的小样本、标注困难以及算法缺乏普适性三大问题，提出了一种基于伪标签的深度学习材料图像分割方法，称为伪标签纹理分割网络（PTS）。该网络结构可分为两部分：主分支和辅助分支，其中主分支用于对材料图像进行特征提取和分割预测，辅助分支用于为训练图像生成伪标签。通过使用伪标签对分割任务进行补充监督，实现了基于少量材料图像，训练得到具有泛化能力的图像分割模型，且模型具有普适性。本章所提出的方法具有以下三方面贡献：首先，本方法设计了上下文特征差异监督损失来显式监督网络的泛化能力，这使得 PTS 网络可以依靠极少的样本进行训练；其次，PTS 网络基于划线标注进行训练，显著降低了训练网络所需的标注代价，缓解了材料图像分割中面临的标注困难问题，同时借助于生成的伪标签取得了良好的分割结果；最后，PTS 网络的模型结构不依赖于特定种类的材料图像特点，在不修改网络结构的情况下，PTS 网络可以应用于不同种类的材料图像分割任务。

通过实验验证发现，本方法相较于现有深度学习图像分割方法，在材料图像分割任务中具有更好的分割效果。此外，相较于解决自然场景图像的弱监督语义分割网络，本方法的网络设计充分考虑了材料图像与自然场景图像二者之间的差异与特点，因此本方法相较于现有最好的弱监督语义分割网络更适合解决材料图像分割任务。尽管本方法在实验中表现出了良好的效果，但在实验过程中仍反映了本方法的

一些问题。在训练阶段，PTS 网络的辅助分支占用了大量的存储空间，这使得模型在训练阶段对设备有着一定的要求。此外，PTS 网络的部分分割结果中还表现出了过分割问题，这也值得在后续的研究中进一步完善改进。

第五章 总结与展望

5.1 结论

材料图像通常包含复杂的纹理，其本身的特点导致材料图像的分割任务存在着诸多问题。这些问题可以被概括为三大难点，分别是小样本、标注困难和算法缺乏普适性。材料的制备与拍摄条件决定了材料图像难以被批量获取，这导致同一种类的材料图像样本数量非常有限；材料图像中组织粘连、目标交错，这决定了其标注工作有着高昂的代价；现今的研究为了更精确地完成材料图像的分割，通常根据特定类别材料图像的特点设计算法，这导致了算法缺乏普适性。为了获得更好的分割结果，最新的图像分割算法主要采用机器学习与深度学习分辨材料图像中复杂的微观结构，因此标注代价与分割精度二者之间的矛盾已成为材料图像分割任务中无法忽视的问题之一。本论文针对多种具有复杂纹理的材料图像，提出不同的非完全标注方法，对其进行处理，即：

1. 基于交互式标注的机器学习材料图像分割方法。首先，对图像中每个像素点提取以当前像素点为中心的 9×9 像素点矩形块作为感兴趣区域，分别使用 Gabor 滤波器、Hu 矩、方向梯度直方图、灰度共生矩阵和平均灰度值对 ROI 矩形块进行特征提取并用于描述当前像素点的特征；接着，对图像中各类别物相的最显著区域进行划线，以划线标注像素点的特征构建数据集用于训练特征分类模型；最后，在每张训练图像上不断迭代优化，交互式标注误分割区域并使用新的标注信息训练模型。该方法实现了标注代价与分割精度二者之间的平衡，借助交互式标注的形式，实时地获取分割模型，能够解决多种不同类别的材料图像分割问题。
2. 基于伪标签的深度学习材料图像分割方法。该方法提出了伪标签纹理分割网络(PTS)以克服材料图像分割任务中的小样本、标注困难和算法缺乏普适性三大问题。其网络结构可分为两部分：主分支和辅助分支，其中主分支用于对材料图像进行特征提取和分割预测，辅助分支用于为训练图像生成伪标签。通过使

用伪标签对分割任务进行补充监督，实现了基于少量材料图像，训练得到具有泛化能力的图像分割模型，且模型具有普适性。该方法通过引入伪标签思想，结合弱监督深度学习等技术，改善了有限标注下材料图像分割的准确性，可获得较精确的材料图像分割结果。

本论文对以上两个方法进行了实验验证，实验表明上述方法在基于非完全标注解决具有复杂纹理的材料图像分割问题时，有着良好的准确性和普适性，具有较高的实用价值。这为获取并统计材料图像中各种类别物相的数据信息提供了技术支持，也对新材料的研发分析和探究材料图像与宏观性能之间的关联起到了一定的促进作用。

5.2 工作展望

尽管本论文提出的方法能够基于非完全标注解决具有复杂纹理的材料图像分割问题，并从获取模型的实时性和分割精度两个角度提出了相应方法，但是目前方法仍然存在着一定的局限性。为了进一步改善本论文方法的有效性，本论文可以在如下几个方面进一步深入研究：

1. 基于自适应提取的图像特征实时地获取分割模型。在本论文的机器学习方法中，模型采用人为预先选定的特征用于描述材料图像。但是人为预先定义的特征描述算子的特征表示能力无法覆盖所有类别的材料图像，且分割结果的准确性很大程度上取决于提取的图像特征。因此构建自适应特征描述算子，结合机器学习方法实时地获取分割模型，不仅有助于进一步改善分割模型的精度，还可进一步提升分割模型的普适性。
2. 进一步改善基于非完全标注图像分割结果的精度。在本论文的深度学习方法中，模型在对材料图像的分割中取得了良好的结果。但是，通过对比现有公共数据集上基于划线标注的弱监督图像分割方法可以发现，在大规模的自然场景图像数据集中，弱监督网络可以达到全监督网络分割结果的约 98%。相比之下，本论文所提出的深度学习方法尚未达到这一效果，其最主要原因是数据集规模的

限制。因此如何在小样本场景下，进一步改善基于非完全标注图像分割结果的精度具有着较高的研究价值与实用意义。

上述研究方向仍然需要大量的实验和探索才能取得相应的成果，未来任重而道远。

参考文献

- [1] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, “Commentary: The materials project: A materials genome approach to accelerating materials innovation,” *APL Materials*, vol. 1, no. 1, p. 011002, 2013.
- [2] S. V. Kalinin, B. G. Sumpter, and R. K. Archibald, “Big–deep–smart data in imaging for guiding materials design,” *Nature Materials*, vol. 14, no. 10, pp. 973–980, 2015.
- [3] Y. Xu, Y. Zhang, M. Zhang, M. Wang, W. Xu, C. Wang, Y. Sun, and P. Wei, “Quantitative analysis of metallographic image using attention-aware deep neural networks,” *Sensors*, vol. 21, no. 1, 2021.
- [4] V. Korniy, B. Rusyn, Y. Lysak, and O. Lutsyk, “Quantitative analysis of metallographic images,” in *2006 International Conference - Modern Problems of Radio Engineering, Telecommunications, and Computer Science*, pp. 251–252, 2006.
- [5] S. Broderick, C. Suh, J. Nowers, B. Vogel, S. Mallapragada, B. Narasimhan, and K. Rajan, “Informatics for combinatorial materials science,” *JOM*, vol. 60, no. 3, pp. 56–59, 2008.
- [6] V. H. C. de Albuquerque, P. C. Cortez, A. R. de Alexandria, and J. M. R. Tavares, “A new solution for automatic microstructures analysis from images based on a backpropagation artificial neural network,” *Nondestructive Testing and Evaluation*, vol. 23, no. 4, pp. 273–283, 2008.
- [7] Y. Pan and L. Zhang, “Dual attention deep learning network for automatic steel surface defect segmentation,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 37, no. 11, pp. 1468–1487, 2021.
- [8] B. Yang, M. Wu, and W. Teizer, “Modified unet++ with attention gate for graphene identification by optical microscopy,” *Carbon*, vol. 195, pp. 246–252, 2022.
- [9] B. L. DeCost, M. D. Hecht, T. Francis, B. A. Webler, Y. N. Picard, and E. A. Holm, “Uhcsdb: Ultrahigh carbon steel micrograph database,” *Integrating Materials and Manufacturing Innovation*, vol. 6, no. 2, pp. 197–205, 2017.

- [10] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, 2015.
- [11] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [12] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2881–2890, 2017.
- [13] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.
- [14] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The pascal visual object classes (voc) challenge,” *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.
- [15] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The cityscapes dataset for semantic urban scene understanding,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3213–3223, 2016.
- [16] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft coco: Common objects in context,” in *Computer Vision – ECCV 2014*, (Cham), pp. 740–755, Springer International Publishing, 2014.
- [17] S. Jo and I.-J. Yu, “Puzzle-cam: Improved localization via matching partial and full features,” in *2021 IEEE International Conference on Image Processing (ICIP)*, pp. 639–643, 2021.
- [18] S. Lee, M. Lee, J. Lee, and H. Shim, “Railroad is not a train: Saliency as pseudo-pixel supervision for weakly supervised semantic segmentation,” in *Proceedings of the*

- IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5495–5505, 2021.
- [19] K. R. Castleman, *Digital Image Processing*. USA: Prentice Hall Press, 1996.
 - [20] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.
 - [21] A. M. A. Talab, Z. Huang, F. Xi, and L. HaiMing, “Detection crack in image using otsu method and multiple filtering in image processing techniques,” *Optik*, vol. 127, no. 3, pp. 1030–1033, 2016.
 - [22] I. Sobel, G. Feldman, *et al.*, “A 3x3 isotropic gradient operator for image processing,” *a talk at the Stanford Artificial Project in*, pp. 271–272, 1968.
 - [23] J. Canny, “A computational approach to edge detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, 1986.
 - [24] L. G. Roberts, *Machine perception of three-dimensional solids*. PhD thesis, Massachusetts Institute of Technology, 1963.
 - [25] J. M. Prewitt *et al.*, “Object enhancement and extraction,” *Picture processing and Psychopictorics*, vol. 10, no. 1, pp. 15–19, 1970.
 - [26] R. Jain, R. Kasturi, B. G. Schunck, *et al.*, *Machine vision*, vol. 5. McGraw-hill New York, 1995.
 - [27] D. Marr, E. Hildreth, and S. Brenner, “Theory of edge detection,” *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 207, no. 1167, pp. 187–217, 1980.
 - [28] 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,” *光学精密工程*, vol. 18, no. 10, pp. 2314–2323, 2010.
 - [29] Y. Han, R. Li, Y. Zeng, and M. Liu, “Automatic detection of kikuchi bands based on radon transform and ppht,” *Journal of Microscopy*, vol. 285, no. 2, pp. 95–111, 2022.
 - [30] R. Adams and L. Bischof, “Seeded region growing,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 641–647, 1994.

- [31] J. B. T. M. Roerdink and A. Meijster, “The watershed transform: Definitions, algorithms and parallelization strategies,” *Fundamenta Informaticae*, vol. 41, no. 1-2, pp. 187–228, 2000.
- [32] Y. Chen and J. Chen, “A watershed segmentation algorithm based on ridge detection and rapid region merging,” in *2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, pp. 420–424, 2014.
- [33] J. Liu and J. Chen, “An improved iterative watershed according to ridge detection for segmentation of metallographic image,” *Metallogr. Image*, vol. 8, pp. 8–13, 2012.
- [34] T. S. Madhulatha, “An overview on clustering methods,” 2012.
- [35] M. Hearst, S. Dumais, E. Osuna, J. Platt, and B. Scholkopf, “Support vector machines,” *IEEE Intelligent Systems and their Applications*, vol. 13, no. 4, pp. 18–28, 1998.
- [36] J. C. Bezdek, R. Ehrlich, and W. Full, “Fcm: The fuzzy c-means clustering algorithm,” *Computers & Geosciences*, vol. 10, no. 2, pp. 191–203, 1984.
- [37] L. Chen, Y. Han, B. Cui, Y. Guan, and Y. Luo, “Two-dimensional fuzzy clustering algorithm (2dfcm) for metallographic image segmentation based on spatial information,” in *2015 2nd International Conference on Information Science and Control Engineering*, pp. 519–521, 2015.
- [38] M. Li, D. Chen, S. Liu, and D. Guo, “Online learning method based on support vector machine for metallographic image segmentation,” *Signal, Image and Video Processing*, vol. 15, no. 3, pp. 571–578, 2021.
- [39] H. Zhao and M. Comer, “A hybrid markov random field/marked point process model for analysis of materials images,” *IEEE Transactions on Computational Imaging*, vol. 2, no. 4, pp. 395–407, 2016.
- [40] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [41] D. A. Reynolds, “Gaussian mixture models.,” *Encyclopedia of biometrics*, vol. 741, no. 659-663, 2009.

- [42] C. Lai, L. Song, Y. Han, Q. Li, H. Gu, B. Wang, Q. Qian, and W. Chen, “Material image segmentation with the machine learning method and complex network method,” *MRS Advances*, vol. 4, no. 19, p. 1119–1124, 2019.
- [43] Y. Han, R. Li, S. Yang, Q. Chen, B. Wang, and Y. Liu, “Center-environment feature models for materials image segmentation based on machine learning,” *Scientific Reports*, vol. 12, no. 1, p. 12960, 2022.
- [44] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 801–818, 2018.
- [45] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, (Cham), pp. 234–241, Springer International Publishing, 2015.
- [46] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: A nested u-net architecture for medical image segmentation,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, (Cham), pp. 3–11, Springer International Publishing, 2018.
- [47] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, “Unet 3+: A full-scale connected unet for medical image segmentation,” in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1055–1059, 2020.
- [48] S. M. Azimi, D. Britz, M. Engstler, M. Fritz, and F. Mücklich, “Advanced steel microstructural classification by deep learning methods,” *Scientific Reports*, vol. 8, no. 1, p. 2128, 2018.
- [49] 李原, 李燕君, 刘进超, 范衡, 王庆林, “基于改进 res-unet 网络的钢铁表面缺陷图像分割研究,” *电子与信息学报*, vol. 44, no. 05, pp. 1513–1520, 2022.
- [50] Y. Han, S. Yang, and Q. Chen, “Recognition and segmentation of complex texture images based on superpixel algorithm and deep learning,” *Computational Materials Science*, vol. 209, p. 111398, 2022.

- [51] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, “Dual attention network for scene segmentation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3146–3154, 2019.
- [52] Z. Zhong, Z. Q. Lin, R. Bidart, X. Hu, I. B. Daya, Z. Li, W.-S. Zheng, J. Li, and A. Wong, “Squeeze-and-attention networks for semantic segmentation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13065–13074, 2020.
- [53] G. French, S. Laine, T. Aila, M. Mackiewicz, and G. Finlayson, “Semi-supervised semantic segmentation needs strong, varied perturbations,” 2020.
- [54] J. Kim, J. Jang, H. Park, and S. Jeong, “Structured consistency loss for semi-supervised semantic segmentation,” 2021.
- [55] Y. Ouali, C. Hudelot, and M. Tami, “Semi-supervised semantic segmentation with cross-consistency training,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12674–12684, 2020.
- [56] L.-C. Chen, R. G. Lopes, B. Cheng, M. D. Collins, E. D. Cubuk, B. Zoph, H. Adam, and J. Shlens, “Naive-student: Leveraging semi-supervised learning in video sequences for urban scene segmentation,” in *Computer Vision – ECCV 2020*, (Cham), pp. 695–714, Springer International Publishing, 2020.
- [57] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, “Learning deep features for discriminative localization,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2921–2929, 2016.
- [58] Q. Hou, P. Jiang, Y. Wei, and M.-M. Cheng, “Self-erasing network for integral object attention,” in *Advances in Neural Information Processing Systems*, vol. 31, Curran Associates, Inc., 2018.
- [59] Y. Wei, J. Feng, X. Liang, M.-M. Cheng, Y. Zhao, and S. Yan, “Object region mining with adversarial erasing: A simple classification to semantic segmentation approach,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1568–1576, 2017.

- [60] S. J. Oh, R. Benenson, A. Khoreva, Z. Akata, M. Fritz, and B. Schiele, “Exploiting saliency for object segmentation from image level labels,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5038–5047, 2017.
- [61] Z. Huang, X. Wang, J. Wang, W. Liu, and J. Wang, “Weakly-supervised semantic segmentation network with deep seeded region growing,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7014–7023, 2018.
- [62] P. Krähenbühl and V. Koltun, “Efficient inference in fully connected crfs with gaussian edge potentials,” in *Advances in Neural Information Processing Systems*, vol. 24, Curran Associates, Inc., 2011.
- [63] D. Lin, J. Dai, J. Jia, K. He, and J. Sun, “Scribblesup: Scribble-supervised convolutional networks for semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3159–3167, 2016.
- [64] T.-W. Ke, J.-J. Hwang, and S. X. Yu, “Universal weakly supervised segmentation by pixel-to-segment contrastive learning,” 2021.
- [65] Y. Boykov and M.-P. Jolly, “Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images,” in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 1, pp. 105–112 vol.1, 2001.
- [66] C. Rother, V. Kolmogorov, and A. Blake, “Grabcut: Interactive foreground extraction using iterated graph cuts,” *ACM Trans. Graph.*, vol. 23, no. 3, p. 309–314, 2004.
- [67] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.
- [68] A. Krizhevsky, G. Hinton, *et al.*, “Learning multiple layers of features from tiny images,” 2009.
- [69] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, “Tph-yolov5: Improved yolov5 based on transformer prediction head for object detection on drone-captured scenarios,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pp. 2778–2788, 2021.

- [70] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” in *Advances in Neural Information Processing Systems*, vol. 28, Curran Associates, Inc., 2015.
- [71] H. Qu, P. Wu, Q. Huang, J. Yi, G. M. Riedlinger, S. De, and D. N. Metaxas, “Weakly supervised deep nuclei segmentation using points annotation in histopathology images,” in *Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning*, vol. 102 of *Proceedings of Machine Learning Research*, pp. 390–400, PMLR, 2019.
- [72] J. Zhang, Z. Hua, K. Yan, K. Tian, J. Yao, E. Liu, M. Liu, and X. Han, “Joint fully convolutional and graph convolutional networks for weakly-supervised segmentation of pathology images,” *Medical Image Analysis*, vol. 73, p. 102183, 2021.
- [73] K. Sirinukunwattana, J. P. Pluim, H. Chen, X. Qi, P.-A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez, A. Böhm, O. Ronneberger, B. B. Cheikh, D. Racoceanu, P. Kainz, M. Pfeiffer, M. Urschler, D. R. Snead, and N. M. Rajpoot, “Gland segmentation in colon histology images: The glas challenge contest,” *Medical Image Analysis*, vol. 35, pp. 489–502, 2017.
- [74] Y. Xu, Z. Jia, L.-B. Wang, Y. Ai, F. Zhang, M. Lai, and E. I.-C. Chang, “Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features,” *BMC Bioinformatics*, vol. 18, no. 1, p. 281, 2017.
- [75] K. Bera, K. A. Schalper, D. L. Rimm, V. Velcheti, and A. Madabhushi, “Artificial intelligence in digital pathology —new tools for diagnosis and precision oncology,” *Nature Reviews Clinical Oncology*, vol. 16, no. 11, pp. 703–715, 2019.
- [76] R. A. Güler, N. Neverova, and I. Kokkinos, “Densepose: Dense human pose estimation in the wild,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7297–7306, 2018.
- [77] D.-H. Lee *et al.*, “Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks,” in *Workshop on challenges in representation learning, ICML*, vol. 3, p. 896, 2013.

- [78] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [79] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. A. Raffel, “Mixmatch: A holistic approach to semi-supervised learning,” in *Advances in Neural Information Processing Systems*, vol. 32, Curran Associates, Inc., 2019.
- [80] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C.-L. Li, “Fixmatch: Simplifying semi-supervised learning with consistency and confidence,” in *Advances in Neural Information Processing Systems*, vol. 33, pp. 596–608, Curran Associates, Inc., 2020.
- [81] Y. Zou, Z. Zhang, H. Zhang, C.-L. Li, X. Bian, J.-B. Huang, and T. Pfister, “Pseudoseg: Designing pseudo labels for semantic segmentation,” in *International Conference on Learning Representations*, 2021.
- [82] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, “Scene parsing through ade20k dataset,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 633–641, 2017.
- [83] K. Li, Z. Wu, K.-C. Peng, J. Ernst, and Y. Fu, “Tell me where to look: Guided attention inference network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [84] J. Dai, K. He, and J. Sun, “Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1635–1643, 2015.
- [85] R. Mottaghi, X. Chen, X. Liu, N.-G. Cho, S.-W. Lee, S. Fidler, R. Urtasun, and A. Yuille, “The role of context for object detection and semantic segmentation in the wild,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 891–898, 2014.
- [86] B. Hu, X. Wang, and W. Yu, “Joint weakly and fully supervised learning for surface defect segmentation from images,” *Signal Processing: Image Communication*, vol. 107, p. 116807, 2022.

- [87] Z. Dong, J. Wang, B. Cui, D. Wang, and X. Wang, “Patch-based weakly supervised semantic segmentation network for crack detection,” *Construction and Building Materials*, vol. 258, p. 120291, 2020.
- [88] D. Gabor, “Theory of communication. part 1: The analysis of information,” *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering*, vol. 93, pp. 429–441(12), 1946.
- [89] M.-K. Hu, “Visual pattern recognition by moment invariants,” *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.
- [90] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886–893 vol. 1, 2005.
- [91] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973.
- [92] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [93] J. L. Chen, S. R. Gunn, M. S. Nixon, and R. N. Gunn, “Markov random field models for segmentation of pet images,” in *Information Processing in Medical Imaging*, (Berlin, Heidelberg), pp. 468–474, Springer Berlin Heidelberg, 2001.
- [94] F. Meyer, “Color image segmentation,” in *1992 International Conference on Image Processing and its Applications*, pp. 303–306, 1992.
- [95] Y. Han, C. Lai, B. Wang, and H. Gu, “Segmenting images with complex textures by using hybrid algorithm,” *Journal of Electronic Imaging*, vol. 28, no. 1, p. 013030, 2019.
- [96] Y. Cheng, “Mean shift, mode seeking, and clustering,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790–799, 1995.
- [97] W.-Y. Loh, “Classification and regression trees,” *WIREs Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 14–23, 2011.

- [98] G. Baudat and F. Anouar, “Generalized Discriminant Analysis Using a Kernel Approach,” *Neural Computation*, vol. 12, no. 10, pp. 2385–2404, 2000.
- [99] I. S. Dhillon, Y. Guan, and B. Kulis, “Kernel k-means: Spectral clustering and normalized cuts,” in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’04, (New York, NY, USA), p. 551–556, Association for Computing Machinery, 2004.
- [100] N. Friedman, D. Geiger, and M. Goldszmidt, “Bayesian network classifiers,” *Machine Learning*, vol. 29, no. 2, pp. 131–163, 1997.
- [101] C. J. C. Burges, “A tutorial on support vector machines for pattern recognition,” *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
- [102] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.
- [103] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’16, (New York, NY, USA), p. 785–794, Association for Computing Machinery, 2016.
- [104] R. Korlakai Vinayak and R. Gilad-Bachrach, “DART: Dropouts meet Multiple Additive Regression Trees,” in *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, vol. 38 of *Proceedings of Machine Learning Research*, (San Diego, California, USA), pp. 489–497, PMLR, 2015.
- [105] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2015.
- [106] E. J. Kirkland, *Bilinear Interpolation*, pp. 261–263. Boston, MA: Springer US, 2010.
- [107] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, “Cen-net: Context encoder network for 2d medical image segmentation,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2281–2292, 2019.

- [108] Z. Liang, T. Wang, X. Zhang, J. Sun, and J. Shen, “Tree energy loss: Towards sparsely annotated semantic segmentation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16907–16916, 2022.
- [109] L. Van der Maaten and G. Hinton, “Visualizing data using t-sne.,” *Journal of machine learning research*, vol. 9, no. 11, 2008.

作者在攻读硕士学位期间公开发表的论文、专利与软著

攻读硕士期间论文发表情况：

1. Han, Y., Li, R., Zeng, Y., & Liu, M. (2022). Automatic detection of Kikuchi bands based on Radon transform and PPHT. *Journal of Microscopy*, 285: 95–111. <https://doi.org/10.1111/jmi.13079> (导师一作, 本人二作, 已发表, SCI 三区)
2. Han, Y., Li, R., Yang, S. et al. Center-environment feature models for materials image segmentation based on machine learning. *Sci Rep* 12, 12960 (2022). <https://doi.org/10.1038/s41598-022-16824-w> (导师一作, 本人二作, 已发表, SCI 三区)
3. Han, Y., Li, R., Wang, B., Chen, Q. A Pseudo-labeling Based Weakly Supervised Segmentation Method For Few-Shot Texture Images. (导师一作, 本人二作, submitted: *Expert Systems With Applications*)
4. Han, Y., Li, R., Fu, J., Wang, B., Wang, H., Xu, X. Statistics and Analysis of Lath Martensite Transformation based on in-situ observation and video processing. (导师一作, 本人二作, submitted: *Journal of Materials Research and Technology*)

攻读硕士期间专利获取情况：

1. 专利名称：基于检测直线和线段的菊池花样几何信息提取方法，发明人：韩越兴、李睿祺、王冰。专利号：ZL 2020 1 0297634.3，授权日：2023 年 2 月 10 日，授权公告号：CN 111612738 B。(导师一作, 本人二作)
2. 专利名称：基于少量样本划线标注的弱监督神经网络的图像分割方法，发明人：韩越兴、李睿祺、钱权、王冰。申请号或专利号：2023101577354，申请日：2023 年 2 月 23 日，公开日：2023 年 2 月 23 日。(导师一作, 本人二作)

攻读硕士期间软著获取情况：

1. 软件名称：上海大学菊池花样自动识别检测软件 V1.0，开发人：韩越兴、李睿祺、王冰。登记号：2020SR0441105，申请人：上海大学，开发完成日期：2020 年 3 月 12 日，登记日期：2020 年 5 月 12 日。

致谢

光阴荏苒，日月如梭。转眼间已度过研究生的三年宝贵时光，人生的求学之路也即将告一段落。回想在校学习研究的时光，离不开身边帮助过我的老师与同学，在此特别对诸位报以最诚挚的感谢。

首先，我要感谢我的导师韩越兴老师。本论文和研究生期间的相关研究学习工作都是在导师韩越兴老师的悉心指导下完成的。可以说，没有韩老师的认真指导和教诲就没有我今天的收获。韩老师是一位认真负责的导师，他对于学术认真严谨，始终在学术讨论和日常交谈中为我们发掘新思路、新构想。此外，韩老师工作兢兢业业、一丝不苟，为我们树立了良好的榜样起到了带头作用，为我们的研究的学习提供了良好的氛围与环境。在学习工作之余，韩老师还总是关心我们的日常生活，正值疫情三年，韩老师的关心是稳定的在校生活最大的保障。在此衷心祝愿韩老师身体健康、前程万里、桃李满天下。

其次，我要感谢我的父母，是他们给了我机会让我能够在校求学研究，也感谢他们对我二十多年的养育之恩。此处寥寥几笔难表其中深情，祝你们福寿双全、年年有余。

接着，还要感谢陈侨川老师、王冰老师等各位合作老师对我的指导与支持，没有你们的帮助我无法完成研究生期间的研究工作。在我的研究生三年期间与你们相逢是我最大的幸运，在此祝各位老师幸福安康、心想事成。

然后，我要感谢我的同门，博士学姐、学长以及师兄师弟们，感谢我的舍友和其它所有帮助过我的同学，是你们与我风雨同舟，给了我不断坚持学习研究的动力，成就了现在的自己。

最后，向百忙中抽出时间的盲审专家和诸位答辩老师之以最诚挚的感谢，谢谢你们。