

中图分类号：TP391

单位代号：10280

密 级：公开

学 号：21721559

上海大学



硕士学位论文

SHANGHAI UNIVERSITY
MASTER'S DISSERTATION

题 目	小样本图像分割方法研究及其 在材料性能挖掘的应用
--------	-----------------------------

作 者 池淑婷

学科专业 计算机应用技术

导 师 韩越兴

完成日期 二〇二四年四月

姓 名：池洳婷

学号：21721559

论文题目：小样本图像分割方法研究及其在材料性能挖掘的应用

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 席：张博峰

委 员：王水 邵国栋

导 师：韩越兴

答辩日期：2024年6月2日

姓名：池洳婷

学号：21721559

论文题目：小样本图像分割方法研究及其在材料性能挖掘的应用

上海大学学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师指导下，独立进行研究工作所取得的成果。除了文中特别加以标注和致谢的内容外，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研究者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。

学位论文作者签名：池洳婷

日期：2024 年 6 月 2 日

上海大学学位论文使用授权说明

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)

学位论文作者签名：池洳婷

导师签名：韩兆兴

日期：2024 年 6 月 2 日

日期：2024 年 6 月 2 日

上海大学工学硕士学位论文

小样本图像分割方法研究及其在材料 性能挖掘的应用

作者: 池涵婷
导师: 韩越兴
学科专业: 计算机应用技术

计算机工程与科学学院

上海大学

2024 年 4 月

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

**Research on Small Sample Image
Segmentation Method and Its
Application in Material Property
Mining**

Candidate: Ruting Chi

Supervisor: Yuexing Han

Major: Computer Application Technology

**School of Computer Engineering and Science
Shanghai University
April, 2024**

摘 要

近年来，深度学习技术的快速发展带动了人工智能在各个领域的广泛应用。在材料科学领域，图像分割技术可以用来挖掘材料图像的信息，快速并准确地获得材料微观结构与性能的内禀关系。深度学习分割材料图像的主要难点是小样本问题及预测结果在材料科学中解释的问题。以解决材料图像分割的小样本问题和挖掘材料微观结构与性能的内禀关系为目标，本论文提出多种深度学习技术，实现材料图像处理 and 材料性能挖掘。具体工作如下：

（1）针对材料图像语义分割中存在的小样本和微观结构特征复杂的问题，提出基于特征金字塔和十字交叉注意力的双分支语义分割网络。该网络分为主分支和辅助分支。主分支使用特征金字塔模型聚合多层级的图像特征来增强细节信息；辅助分支使用骨干网络的低层特征分割图像，辅助网络学习纹理和边界信息。在多任务监督和多尺度特征的协作下，相较于对比模型，该方法在多个小样本材料图像数据集上取得最佳性能。

（2）针对材料图像实例分割中显著的小样本问题，从提高现有数据的利用率出发，提出基于多模态融合和伪标签技术的实例分割方法。该方法通过融合图像和文本的多模态数据，提高网络对物体分类、定位和分割的准确度。本论文将训练分为两阶段：全监督训练和半监督训练。在半监督训练阶段，采用伪标签技术使未标注数据参与监督模型寻优。在多模态信息和两阶段训练的协作下，相较于对比模型，该方法在多种小样本实例分割场景中取得优越的性能。

（3）针对2205双相不锈钢显微图像，设计统计分析方法来完成材料图像分割结果和材料性能的内禀关系挖掘。本论文分别基于语义分割结果以及实例分割结果，使用材料经验公式，高效且精确地获得和人工分析结果相近的微观结构的生长速率和材料性能变化情况。

关键词：语义分割；实例分割；小样本；材料性能挖掘；图文融合

ABSTRACT

In recent years, the rapid development of deep learning technology has driven the widespread application of artificial intelligence in various fields. In the field of material science, image segmentation techniques can be used to mine information from material images and obtain the accurate intrinsic relationship between material microstructure and property quickly. The main difficulty of deep learning in segmenting material images is the problem of small sample size and the interpretation of prediction results in material science. With the goal of solving the problem of small sample size in material images and mining the intrinsic relationship between microstructure and property, the thesis proposes various deep learning techniques to achieve material image processing and material property mining. The specific tasks are as follows:

(1) In response to the problems of small sample size and complex features of microstructure in semantic segmentation of material images, dual branch semantic segmentation network based on feature pyramid network and criss-cross attention are proposed. The network is divided into the main branch and the auxiliary branch. The main branch uses a feature pyramid network to aggregate multi-level image features to enhance the detailed information of the image. The auxiliary branch directly uses the low-level features of the backbone to segment images, and assists the network in learning image texture and boundary information. Compared to the comparative models, the method achieves the best performance on multiple small sample material image datasets under the collaboration of multi-task supervision and multi-scale features.

(2) In response to the more significant small sample problem in material image instance segmentation, instance segmentation method based on multimodal fusion and pseudo labeling technology is proposed. This method improves the accuracy of object classification, localization, and segmentation in the network by fusing multimodal data from images and text. This thesis divides training into two stages: fully-supervised training and semi-

supervised training. In the semi-supervised training stage, pseudo labeling techniques are used to enable unlabeled data to supervise model optimization. Under the collaboration of multimodal information and two-stage training, the method achieves superior performance in various small sample instance segmentation scenarios compared to the comparative models.

(3) Statistical analysis methods for the intrinsic relationship between material image segmentation results and material properties in the microscopic images of 2205 duplex stainless steel are designed. Based on the semantic segmentation results and instance segmentation results respectively, using material empirical formulas, the thesis efficiently obtain the growth rate and material performance changes of microstructures that are similar to the results of manual analysis.

Keywords: Semantic Segmentation; Instance Segmentation; Small Sample; Material Property Mining; Image-text Fusion

目 录

摘 要	I
ABSTRACT	II
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 小样本学习研究现状	4
1.4.2 图像分割研究现状	6
1.4.3 图像文本融合研究现状	9
1.4.4 材料图像和性能的内禀关系挖掘研究现状	9
1.5 论文主要工作	10
1.6 论文组织架构	10
第二章 相关理论与技术概述	12
2.1 残差网络	12
2.2 图像分割	13
2.2.1 语义分割	14
2.2.2 实例分割	16
2.3 注意力机制	18
2.4 伪标签	20
2.5 评价指标	21
2.6 本章小结	23
第三章 基于双分支语义分割网络的图像分割和材料性能挖掘	24
3.1 方法概述	24
3.1.1 骨干网络	24

3.1.2	主分支	26
3.1.3	辅助分支	27
3.1.4	损失函数	29
3.2	实验分析	30
3.2.1	2205双相不锈钢图像数据集介绍	30
3.2.2	实验设置	31
3.2.3	消融实验	32
3.2.4	对比实验	36
3.2.5	普适性分析	37
3.2.6	基于图像语义分割的材料性能挖掘	40
3.3	本章小结	45
第四章	基于信息融合的图像实例分割和材料性能挖掘	46
4.1	方法概述	46
4.1.1	基模型	47
4.1.2	语义融合模块	48
4.1.3	融合伪标签的两阶段训练	49
4.1.4	损失函数	51
4.2	实验分析	53
4.2.1	实验设置	53
4.2.2	数据集介绍	53
4.2.3	消融实验	55
4.2.4	对比实验	61
4.2.5	基于图像实例分割的材料性能挖掘	65
4.3	本章小结	69
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		72

攻读硕士学位期间取得的科研成果	84
攻读硕士学位期间参与的科研项目	85
致 谢	86

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），之江实验室科研攻关项目（编号：2021PE0AC02），云南省科技重点项目（202302AB080022）的资助。

1.2 课题背景概述

得益于计算机算力的提升以及大数据时代的到来，数据驱动型的人工智能算法在各种任务中展示出强大的能力。计算智能、感知智能和认知智能是人工智能的三大核心能力^[1]。Alpha Go的成功标志着计算机处理海量数据的能力已经具备，即计算智能已经实现。认知智能要求机器可以自主学习并作出决策，是人工智能未来发展的趋势之一。当前的人工智能研究主要处在实现感知智能的阶段，即使得机器能够模拟人类的视觉、听觉等感知能力。人类视觉系统瞬时处理海量信息的能力吸引了许多研究聚焦于以图像作为数据来源的计算机视觉技术，由此发展出了多种视觉任务：识别、分割、跟踪、场景重建等，相应的研究成果也在现实场景中得到广泛的应用：人脸识别、无人驾驶、医疗诊断等。

随着多学科交叉研究的深入，计算机视觉技术逐渐成为材料科学领域的重要工具，应用在微观结构的分割、缺陷检测等任务中。传统的材料科学研究需要通过大量的实验，来探究材料性能变化的各种影响因素，存在实验周期长、实验条件严苛等问题。人工智能技术可以帮助科研人员从耗时耗力的工作中解脱出来从而转入专业性更强的研究中，还可以克服资源限制带来的困难。图1.1统计了2000年至2020年期间关于机器学习以及机器学习结合材料科学的已发表作品的数量，可以看出越来越多的材料科学研究借助机器学习模型来完成^[2-6]。近年来，深度学习算法在许多领域上取得了令人瞩目的成就，虽然大部分深度学习方法是在自然图像上验证的，但是由于显微图像与自然图像的相似性——它们都有边缘、纹理等特征，所以将这

些方法迁移到材料显微图像中是可能的^[7]。

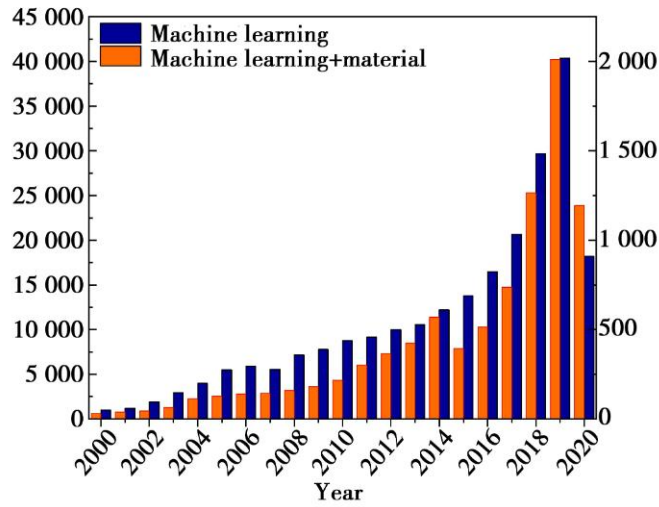


图 1.1 2000-2020年间“机器学习”和“机器学习”+“材料”的已发表作品数量^[6]

但是，当前的材料科学和计算视觉的交叉研究还存在一些共性问题。首先，小样本问题突出。在材料研究中，数据的获取依托对相关物质进行的制备实验，材料制备周期长的特点使得材料数据相较于自然场景数据更为匮乏。同时，对于有监督学习，数据的标注是指导算法学习的重要依据，而材料数据的标注工作需要专业人员参与，当标注粒度细化时（如从图像级至像素级），投入的时间和精力就急剧增加，导致带标签的材料图像数据稀缺。其次，材料图像和自然图像的差异带来算法迁移的障碍。材料图像更加关注纹理、边界等底层特征，深度神经网络提取的高层特征丢失了这些信息，因此需要针对性地设计相关算法。最后，关于计算机视觉的处理结果和材料性能相联系的研究还较为稀少。材料性能是材料科学研究的重要动力之一，而计算机研究人员通常将材料图像视为一般图像，使用图像分割、分类等算法处理数据，对于计算机得到的结果缺少和材料性能的相关性挖掘。

1.3 课题研究的目的与意义

机器学习的图像处理任务包括分类、分割、跟踪等，图像分割是计算机视觉能够理解图像细节信息的关键，也是材料性能挖掘中至关重要的环节。在计算机视觉发展的早期，利用像素点的灰度值和梯度设计了一系列特征描述符和匹配算法，例如定向梯度直方图（Histogram of Oriented Gradients, HOG）^[8]。由于这类方法需要

预先设计特征且关注的信息有限，无法满足精确分割材料图像的要求。机器学习的主要分支——深度学习可以自适应地获取丰富的特征，成为材料图像处理的重要技术。许多深度学习模型在处理自然图像上具有高性能，但是在材料图像上的适用性并未得到充分发挥，特别是在材料图像分割上仍存在许多挑战：

- 材料图像的小样本问题和大规模数据驱动的深度学习算法之间存在矛盾；
- 材料图像特征和自然图像特征差异大，需要设计合适的特征学习策略；
- 基于图像分割结果的材料性能挖掘尚缺乏广泛的研究。

为了克服上述挑战，本论文分别解决图像分割的两大任务——语义分割和实例分割在材料图像应用上的困难。一方面，设计特殊的网络和训练方式缓解数据匮乏和特征差异带来的负担；另一方面，设计统计方法挖掘分割结果和材料性能的联系。

材料科学的研究门类细致且复杂，不同材料之间的研究大不相同。同时，带有语义分割和实例分割标注以及性能信息的材料图像数据集十分罕见，因此本论文针对2205双相不锈钢（Duplex Stainless Steel, DSS）开展相关研究。双相不锈钢具有良好的机械性能和耐腐蚀性，已广泛应用于海洋建筑、石化工业和电厂^[9-13]。众所周知，这种良好的特性归因于在双相不锈钢中形成的铁氧体（ α 相）和奥氏体（ γ 相）。然而，制备双相不锈钢的过程中容易产生不需要的二次相，如 χ 相、 σ 相、 γ_2 相等，降低了双相不锈钢的性能。对双相不锈钢的微观结构演化已有广泛的研究^[10-11,14-24]，可以总结为铁氧体 α 相的分解： $\alpha \longrightarrow \text{M}_{23}\text{C}_6(\text{Cr}_2\text{N}) + \gamma_2$ 和 $\alpha \longrightarrow \sigma(\chi) + \gamma_2$ ^[10,14,16,18-24]。二次相的尺寸、形态和分布对钢的力学性能有很大的影响，根据奥斯特瓦尔德熟化机制^[25]，可以通过计算二次相与铁氧体或奥氏体之间的平均界面能^[26]预测钢的性能变化。然而，粗化过程中测量的困难导致了只有少量关于2205双相不锈钢中二次相大小和分布的定量研究^[27-28]。由于双相不锈钢的良好特性和广泛应用，材料研究人员期望可以掌握其性能的变化规律，更好地控制材料的制备。通过反复的控制变量实验来完成这一工作显然是耗时、耗力、耗材的，因此，将高效的计算机技术用来指导生产至关重要。

综上，本论文以2205双相不锈钢作为示范材料，借助深度学习算法提出了两种小样本材料图像微观结构分割的方法，并设计统计分析方法完成分割结果和材料性能的联系，推动深度学习技术与材料科学的融合和发展。

1.4 国内外研究现状

本论文结合信息融合的思想，实现小样本图像分割，挖掘材料微观结构和性能的内禀关系。本节介绍小样本学习、图像分割、图像和文本的多模态融合以及材料图像和性能内禀关系挖掘的研究现状。

1.4.1 小样本学习研究现状

在深度学习中，数据的规模和质量是决定模型表现能力的关键因素。但是，在许多领域很难提供大量的训练样本，如材料、医学等，使得数据驱动型的深度学习模型无法发挥作用。由此衍生出小样本学习任务，专注于让深度学习模型可以更好地适应数据匮乏的场景。现有的小样本学习方法主要分为数据扩充和模型寻优。

数据扩充包含线性变换和基于生成式模型的方法。在深度学习模型的训练中，数据加载阶段会涉及是否对数据进行旋转、平移、裁剪等简单的线性变换。由于下游模型学习的特征倾向于具备尺度、旋转、平移不变性，基于线性变换的数据扩充方法取得的效果并不显著，激发了更多的研究。Mixup^[29]是一种发源于计算机视觉的数据增强技术，它的思想非常简单：随机选择两个训练样本以及对应的标签，使用线性插值的方法生成一个新的向量和对应的标签作为增强的数据，由此提高模型的泛化能力。基于Mixup发展出了多种数据增强方法，例如基于显著性的最优化Mixup^[30]、风格和内容分离的Mixup^[31]、最近随着Transformer兴起的TokenMixup^[32]，这些改进带来了更为显著的效果提升。但基于Mixup的方法假设特征线性插值对应标签线性插值，所以这类数据增强方式不能很好地描述非线性关系的真实分布。生成式模型是一类以学习数据分布为目标的模型，可以大规模生成数据，例如生成对抗网络（Generative Adversarial Networks, GAN）^[33]和扩散模型^[34]。在单样本学习中，多尺度图像生成框架SinGAN^[35]可以基于单张图像训练，生成与训练图像有相似的几何结构与纹理信息的新图像。GeoAug^[36]是一种用于神经辐射场（Neural Radiance Fields, NeRF）的数据扩充方法，丰富了基于多视图几何约束的训练数据。医学图像处理是一个典型的缺少数据并且标注成本较高的领域，因此生成式模型在其中有较多的应用。例如通过基本的数据增强方法和生成对抗网络框架联合训练，用于肝损伤医学图像生成，然后将生成的图像用于训练分类任

务^[37]。以及,通过输入图像和分割掩码监督生成视网膜光学相干断层扫描图像,然后用于分割任务^[38]。但是生成式模型的工作原理与小样本数据的特点存在矛盾,即生成式模型学习的是数据分布,而少量的样本数据呈现的分布和真实的数据分布有很大偏差,理论上很难从这个带偏差的分布中学习到正确的数据分布。所以,即便生成对抗网络表现强势,但其需要大规模训练数据,训练不稳定、易产生梯度消失和模式崩溃以及可解释性差的问题使得基于模型的数据增强方法不能很好地解决小样本问题。

模型寻优相关的研究聚焦于如何更好地缩小解空间,寻找最优解。例如,基于度量学习的方法将模型当作一个特征生成器,使用神经网络学习特征,如原型网络^[39]、匹配网络^[40]等。这类方法通常适用于分类任务,不直接利用网络进行分类,而是通过度量样本和原型的相似度进行预测,需要大量人工设计和实验来确定合适的度量方法。迁移学习是目前热门的研究方向,这类方法先基于额外的训练数据训练网络,再在目标小样本域上微调模型参数,源域和目标域的距离会影响迁移的有效性。其中最具代表性的是使用元学习预训练网络。元学习的预训练阶段使用和目标小样本数据集不重合的数据进行相同的训练任务,该阶段使用的数据称为 D_{Base} 或支撑集 (Support Set),包含的类别称为基类 (Base Class),相应地,正式训练阶段的数据称为 D_{Novel} 或查询集 (Query Set),包含的类别称为新类 (Novel Class)。元学习方法提供了特殊的训练技巧,需要精心设计以最大化支撑集的利用率,而且其中任务强相关的预训练阶段需要耗费大量的训练成本和额外的数据。在一般的计算机视觉任务中,有很多可以共享的先验知识,这也是迁移学习可以有效工作的主要原因。但是在特定的任务上,例如材料、医学、工业场景,很难找到与之有大量共享先验知识的数据,所以这些方法很难用到这些场景上。模型寻优不仅可以通过充分使用有标注数据,还可以加入无标签数据的信息。半监督学习是基于有标签数据虽然很少,但无标签数据容易获取的事实上进行的研究。因此,半监督学习的首要目标是如何利用无标签数据的信息。现有的半监督学习方法主要有一致性正则化和伪标签方法。正则化方法基于数据增强的思想,通过最小化模型对加噪前后的数据预测结果的差异来避免模型陷入过拟合,从而提高模型的泛化能力。例如 Π -model^[41]通过对同一数据点进行随机变换(例如随机最大池化)来最小化两

次预测的差异。II-model要求网络对每个样本运行两次，因此其计算成本加倍。为了降低计算成本，时间集成（Temporal Ensembling）^[42]将保持模型预测的指数变化平均值（Exponential Moving Average, EMA）作为学习目标，对每个训练样本仅在每一轮中评估和更新一次，由于标签预测只会在每个轮次发生变化，当训练数据集很大时会使这种方法变得笨拙。因此，平均教师（Mean Teacher）^[43]提出通过跟踪模型权重的变化平均值而不是模型输出来克服目标更新缓慢的问题。受对抗训练（Adversarial Training）^[33]在输入中加入噪声使得模型对攻击具有鲁棒性的启发，虚拟对抗训练（Virtual Adversarial Training, VAT）^[44]将对抗训练策略扩展到半监督学习中。除了加噪的策略，基于插值的正则化方法也有广泛应用。插值一致性训练（Interpolation Consistency Training, ICT）^[45]遵循MixUp的思想通过添加更多数据点插值来增强数据集，期望模型在混合样本上产生一个预测标签，以匹配对应的预测插值。一致性正则化方法通过约束模型对扰动前后的无标签数据输出一致的预测结果，使决策边界位于低密度区域，从而缓解过拟合现象。伪标签方法使用教师模型生成无标签数据的伪标签，将已有数据和带有伪标签的无标签数据混合在一起训练学生模型。Noisy Boundary^[46]使用在已有数据上训练好的模型一次性生成伪标签，这种离线生成方式不能更新伪标签，带来了局限性。元伪标签（Meta Pseudo Labels）^[47]根据学生模型在标记数据集上的表现反馈不断调整教师模型。教师和学生模型并行训练，教师模型可以学习生成更好的伪标签，学生可以从伪标签中学习。伪标签通过最小化无标注数据的类概率条件熵，以支持类之间的低密度分离。众多方法中，结合一致性正则化和伪标签的方法表现出了良好的性能，例如MixMatch^[48]和FixMatch^[49]。

1.4.2 图像分割研究现状

图像分割包含语义分割、实例分割和全景分割。语义分割和全景分割都为图像的每个像素点预测一个类别，但是后者还需要区分可数类别（Thing）中的各个对象；实例分割区分每个可数物体并预测类别和掩码，而不考虑无定形区域（Stuff）。本小节介绍语义分割和实例分割的发展现状。

材料图像的语义分割算法可以分为三大类：基于领域知识和规则的方法、经典机器学习方法以及基于深度神经网络的方法^[50]。基于领域知识和规则的方法融合

传统的图像处理方法和专家领域知识来提取和分析材料图像中的关键信息，依赖于每个图像的固有属性。这类方法需要人工定义多种参数，不仅费时费力，而且也很难确定准确的参数值。例如阈值法^[51]根据灰度值的双峰分布进行图像分割，当灰度值不满足双峰分布时就失去了效果。随着机器学习的发展，许多工作利用机器学习方法对材料图像进行分割。Han等人提出了一种结合领域知识和梯度提升决策树（Gradient Boosting Decision Tree, GBDT）模型对多种材料图像进行分割的方法^[50]，这种通过邻域特征识别中心像素的方法大大提升了传统机器学习方法在图像分割上的性能。这类方法用传统的分类模型进行图像分割，一般是两阶段式的，需要确定邻域大小以及需要手工设计特征，不能充分利用整张图片的信息，有很大的局限性。近年来，神经网络凭借强大的学习能力在材料图像分割任务上超越了传统方法。语义分割模型SegNet^[52]及其变体DefectSegNet^[53]在树突生长图像和高质量的钢缺陷图像上取得了优越的分割性能。利用超像素技术挖掘图像特征并结合DenseNet，Han等人实现了小样本高分辨率复杂纹理材料图像的分割^[54]。此外，也有一些学者对神经网络这种黑盒模型进行解释。如Maksov等人利用高斯混合模型挖掘卷积神经网络检测到的晶体缺陷结构与特定的物理结构的联系^[55]，实现了对深度学习模型解释性的探索，推动深度学习在材料科学上的应用。神经网络方法凭借优越的性能已经成为现下最热门的材料图像处理方法，但是因为材料图像的种类繁多，许多方法只适用于单一类别的材料图像，缺少普适性。

实例分割不仅需要学习类间差异，还要学习类内差异，相较于语义分割，是一个更复杂的任务。目前，实例分割有两种解决策略：自底向上和自顶向下。自底向上的方式先得到语义分割的结果，通过聚类的方法来确定属于各个物体的像素点。相似群建议网络（Similarity Group Proposal Network, SPGN）^[56]是自底向上的一个典型工作，它通过PointNet++^[57]学习像素点的特征，接着使用三个分支处理这些特征，其中一个根据特征进行语义分割，还有一个用来生成相似度矩阵，用来聚类确定各个实例的像素点，最后一个分支是生成置信度，用来过滤不属于实例的点。自底向上的方法需要使用聚类后处理，其中计算相似度矩阵的复杂度为 $O(n^2)$ ，不适合大规模的推广，因此现有的研究大都遵循自顶向下的策略。自顶向下的方式需要先检测出各个物体，再对物体进行分类和分割，现有方法可以分为三大类。第

一类是基于锚（Anchor）的模型，如Mask-RCNN^[58]。这类方法是两阶段方法，最突出的特点是需要区域候选网络（Region Proposal Network, RPN）根据事先设定的锚得到候选框并完成定位，后续基于这些候选框的特征进行筛选、分类和生成掩码。这类方法简单有效，但锚的丰富性以及候选框的质量会影响后续任务，有一定的局限性。第二类是基于网格（Grid）的模型，如SOLO^[59]。这类方法受目标检测模型YOLO^[60]的启发，在YOLO的基础上增加掩码分支以完成实例分割任务。基于网格的模型不需要使用锚对图像进行滑动窗口来得到候选框，有更快的推理速度。由于每个网格负责预测一个实例，当有多个目标落在同一个网格中时，容易发生漏检的情况。此外，实例分割需要得到掩码，基于掩码的非极大抑制（Non-maximum Suppression, NMS）急剧增加了后处理的时间。第三类方法是基于查询（Query）的方法，如Mask2Former^[61]。这类方法受目标检测模型DETR^[62]的启发，得益于视觉Transformer强大的解码能力，将特定数量的可学习查询和编码器得到的图像特征一起解码得到实例分割的结果。这类方法不再受限于锚和网格，也不需要后处理，是单阶段实例分割方法，具有很强的学习能力，虽然目前的研究相较于前两者还不够成熟，但是可以预见其强大的潜能。

实例分割方法在材料图像上的研究还较为稀缺，但是小样本实例分割已经有许多研究成果，可以为小样本材料图像实例分割提供指导。完全指导网络（Fully Guided Network, FGN）^[63]是一种遵循元学习策略的小样本实例分割方法，它在Mask-RCNN中引入了三种不同的引导机制，分别为注意力引导区域候选网络（Attention-guided RPN），关联引导检测器（Relation-guided Detector）以及注意力引导全卷积网络（Attention-guided FCN），通过这些机制可以更充分地利用支撑集的信息。RefT^[64]设计了一个基于掩码的动态加权模块来增强元学习支撑集的特征，并通过交叉注意力链接目标查询以更好地校准特征，在RefT中，查询特征能从特征层面和实例层面被增强两次。FAPIS^[65]提出了不需要锚的小样本实例分割框架，对训练对象类之间共享的潜在对象部分显式建模，旨在回归前景边界框的位置并得到对应的分数，同时估计每个框内潜在部分的相对重要性。可以看出，许多小样本实例分割的研究和元学习息息相关。

1.4.3 图像文本融合研究现状

随着多模态学习的兴起，图像和文本作为最重要的两种信息常常被用来联合训练。这些方法需要两个编码器，一个用来编码文本信息，另一个用来编码图像信息。例如对比语言-图像预训练^[66]（Contrastive Language-Image Pre-Training, CLIP）模型通过增加图像的文本描述来帮助提升图像识别任务的性能。在图像分割领域，GroupViT^[67]使用图像-文本对来监督模型训练。在数据增强领域，Chen等人利用编码TriNet将特征空间映射到语义空间中，并在此进行数据增强，实现了小样本数据扩充^[68]。在许多零样本视觉学习任务中，语义信息被编码后输入网络以提高图像特征的质量。现有的许多方法通过构建知识图谱来辅助完成任务，ADS^[69]和SSR-FSD^[70]是小样本目标检测中为数不多使用语义嵌入的模型，但前者的类别信息是依次嵌入的，没有考虑相关性；后者是通过语义解释模块进行语义空间增强，带来了繁重的参数量。OneFormer^[71]在MaskFormer的基础上增加了一个文本编码器，为模型增加了一个基于文本条件（实例、语义或全景）的输入。在小样本实例分割中，由于需要对实例分割出细致的掩码，需要学习更为精确的特征。跨模态训练虽然带来更多的信息，但是也使训练难度上升，如何合理地融合不同模态的数据仍然是研究的难点之一。

1.4.4 材料图像和性能的内禀关系挖掘研究现状

材料性能是材料科学研究的关注点之一，而图像作为具有丰富信息的数据来源，帮助了材料性能的挖掘工作。图像分类和回归技术在材料性能挖掘上已经有广泛的研究和应用。这类方法将图像或图像特征作为输入，然后使用机器学习方法构建图像与性能之间的关系。例如，Song等人使用卷积神经网络提取材料的微观结构特征，使用多层感知机提取成分特征，并使用神经网络对自适应融合的高级特征完成复合材料的力学性能预测^[72]；根据6061铝合金激光振荡焊接实验的金相显微图，Ai等人采用图像识别技术计算焊缝横截面积，利用线性回归、多项式回归等机器学习算法，建立了不同工艺条件下焊缝面积的预测模型^[73]；Balaji等人利用卷积神经网络模型和处理后图像的像素强度直方图，根据材料去除率对扫描电子显微图像进行分类^[74]。为了丰富性能预测可利用的信息，Ren等人构建了一个深度学习框架耦合成

分的数值信息和多源显微组织的图像信息，用于预测双相钢的拉伸性能^[75]。基于图像分类和回归的材料性能挖掘方法适用于不需要挖掘图像细节信息的性能分析工作。相较于图像分类和回归，图像分割可以获得细节信息，适用于需要对物相统计分析的性能挖掘工作。虽然已有许多分割方法应用在材料图像上，但是由于材料的多样以及相关数据集的匮乏，构建材料分割结果和性能之间关系的工作还很稀缺。

1.5 论文主要工作

为了在小样本环境下从材料图像挖掘材料性能，基于计算机视觉技术，本论文提出两种解决方法，如图1.2所示。主要工作和创新如下：

（1）本论文提出基于特征金字塔和十字交叉注意力的双分支语义分割网络，该网络包含主分支和辅助分支。针对深度学习处理材料图像时遇到的小样本和底层特征丢失的问题，本论文在主分支使用特征金字塔结构融合多层次特征；在辅助分支以较少的计算代价设计了带有十字交叉注意力机制的结构，该分支直接使用骨干网络得到的特征，帮助进一步监督细节信息的提取。通过实验，该方法在多种小样本材料图像分割任务上的有效性和普适性得到了验证。

（2）本论文提出基于图文多模态融合和伪标签技术的实例分割方法。该方法使用两阶段的训练方式：全监督训练和半监督训练，在半监督训练中采用伪标签技术提高对未标注数据的利用率。本论文设计的语义模块融合文本的语义特征和图像的视觉特征，提高实例分割的性能。该方法不仅可以应用在2205双相不锈钢数据集上，还可以在多种公共数据集的小样本场景中发挥作用。

（3）本论文提出一种统计分析方法来挖掘双相不锈钢的微观结构和性能的关系，基于语义分割和实例分割模型的输出，实现了图像分割结果和材料性能的关系挖掘。

1.6 论文组织架构

针对现有的小样本材料图像和性能内禀关系挖掘的问题，本论文充分发挥深度学习的优势，提出相应的小样本材料图像分割方法和材料性能的统计分析方法，通过实验验证了提出的方法的有效性，达到了解决小样本图像分割问题以及挖掘材料

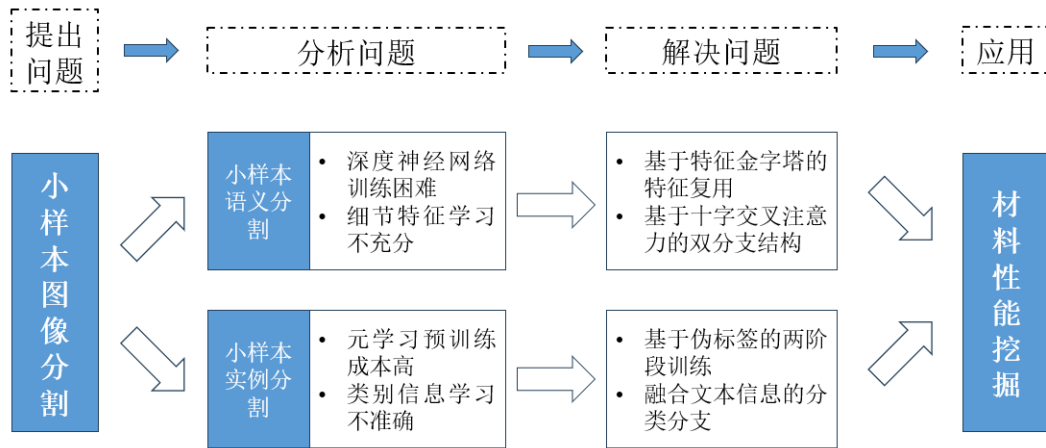


图 1.2 本论文的技术路线图

图像和性能内禀关系的目的，减少了人力投入，提高材料研究效率。

本论文其他章节的内容安排如下：

第二章介绍了有关的理论和技术。主要包括以下内容：语义分割和实例分割的相关理论、注意力机制、伪标签的相关知识和图像分割的评价指标。

在第三章中，提出基于特征金字塔和交叉交叉注意力的双分支语义分割网络，设计统计分析方法完成针对2205双相不锈钢的性能挖掘。首先是对于方法的概述，包括本方法使用的骨干网络、主分支和辅助分支的工作方式以及设计的多任务损失函数；其次，展示本方法在2205双相不锈钢数据集上进行的消融实验，验证各个模块的有效性和部分超参数对语义分割性能的影响；接着，在多个小样本材料图像数据集上分析本方法的普适性；最后，使用分割结果完成材料性能挖掘。

在第四章中，提出基于图文多模态融合和伪标签技术的实例分割方法，完成小样本图像实例分割及预测结果和材料性能内禀关系的挖掘。本章首先介绍多模态信息融合的方式，以及伪标签的工作原理；然后，在多个小样本数据集上进行消融实验和对比实验，验证方法的有效性；最后，使用实例分割结果完成材料性能挖掘。

第五章是对全文的总结与展望：分析本论文方法的优势与不足，并且对未来的研究提出展望。

第二章 相关理论与技术概述

深度学习在图像上的研究历史非常悠久，图像处理任务丰富多样，从图像分类到图像分割，处理的颗粒度更加细化，获得的信息更加丰富。随着深度学习的愈发成熟，图像分割技术已经得到了广泛的应用，例如在自动驾驶、医学辅助诊断等领域。材料科学是重要的支撑学科，材料图像也是当前深度学习研究的对象之一。通过利用计算机技术，材料图像的处理不再纯粹依赖人工，可以更便捷、高效地得到更精确的结果。针对深度学习处理材料图像面临的问题：小样本、特征表示不准确、图像分割结果缺少和性能的联系，本论文探索小样本图像分割和材料性能挖掘的方法。本章回顾了本论文使用的相关理论和技术，为后续章节提供理论基础。

2.1 残差网络

对于深度神经网络，输入的数据首先需要经过骨干网络的处理，得到初步的特征，骨干网络的特征质量奠定了后续学习到的特征的质量。残差网络^[76]（Residual Network, ResNet）已经成为当前深度学习模型发展的重要基础。在一般的认知下，当深度神经网络的层数增加时，网络结构的复杂度也随之增加，其拟合复杂表达式的能力也得到提升，因此对于数据有更强的表征能力，预测性能也相应地提升。但是，实验发现随着深度神经网络层数的增加，在层数达到一个临界值以后，它的训练误差反而有上升的趋势，即新增加的网络使原网络的性能恶化，产生了“负优化”的现象。这主要是因为网络层数的增加使得在反向传播中产生了更多的信息损失。由于反向传播需要计算模型输出结果和真实标签的误差，因此计算得到的梯度信息是从网络的最顶层传向最底层。当梯度的绝对值介于0和1之间，根据梯度的链式法则和小数乘积的性质，网络底层得到的梯度信息会逐渐减小，甚至消失，导致从该层到最底层的权重不能得到更新，出现梯度消失的现象。相应地，当梯度绝对值大于1时，会出现梯度爆炸的问题。

最初，残差网络的提出是为了解决“负优化”问题，但同时它也巧妙地解决了梯度消失的问题，残差模块是构成残差网络的基石，图2.1是残差模块的示意图。一

般的深度神经网络学习的是输入到输出的映射：

$$f : X \rightarrow Y \quad (2.1)$$

在残差模块中，输入不仅通过权重层往前传播，还通过跳跃连接作为恒等部分加入到输出中，因此，残差模块得到的输出表示为：

$$Y = \sigma(\mathcal{F}(X) + X) \quad (2.2)$$

其中， σ 是激活函数， $\mathcal{F}$ 是残差模块学习的映射。因此，可以看出该模块学习的是输入和输出之间的残差 $\mathcal{F}(X)$ ，这也是其名字的由来。通过跳跃连接，网络层和顶层之间有了更短的传播路径，从而可以接收到更多来自顶层的梯度信息，梯度消失的问题有效地得到了解决。基于残差模块，可以灵活地设计其中的网络层以及叠加的残差模块数量。目前的残差网络中残差模块的数量多为4，而模块中的网络层不固定。根据网络层的不同，存在以下几种主流残差网络：ResNet18、ResNet34、ResNet50、ResNet101和ResNet152，数字代表网络的卷积层数。

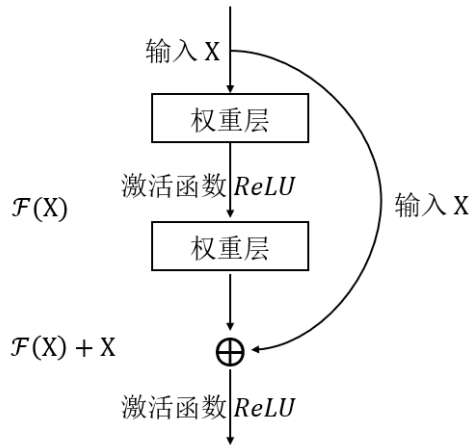


图 2.1 残差模块

2.2 图像分割

图像分割包含三个不同的任务：语义分割（Semantic Segmentation）、实例分割（Instance Segmentation）和全景分割（Panoramic Segmentation）。语义分割任务是对图像中的每一个像素点分配一个类别，这个类别属于一个有限、事先约定好的集合，

因此可以将语义分割看作是一个像素级的分类问题。实例分割任务需要标出实例的类别标签和实例标签，还要得到实例在图像中的像素区域，即掩码。全景分割任务是语义分割和实例分割的结合，需要对物体进行实例分割，并且利用语义分割得出非物体的区域。由于本论文只涉及到语义分割和实例分割，本小节着重介绍语义分割和实例分割的基本理论和技术。

2.2.1 语义分割

语义分割和图像分类息息相关，不同点在于语义分割的标签是一个和图像等大的矩阵，而图像分类的标签是一个标量。因此，可以通过修改用于图像分类的神经网络的输出结构，将图像分类的相关研究迁移到图像分割上。

全卷积神经网络^[77]（Fully Convolution Networks, FCN）是利用深度学习解决语义分割任务的开山之作，也是当前众多研究的基础。如图2.2，全卷积神经网络在分类网络的基础上去掉全连接层，使用上采样还原高分辨率的图像以得到分割的结果。下采样的部分称为编码器，上采样的部分为解码器。可以看出，语义分割是在

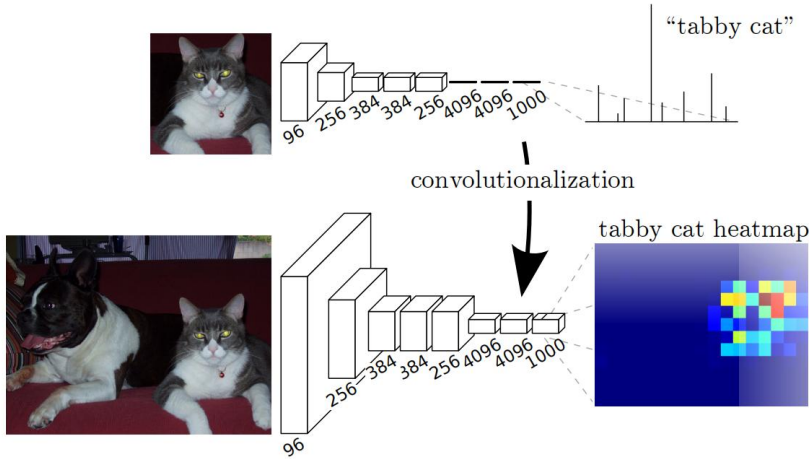


图 2.2 全卷积神经网络将全连接层转换为卷积层^[77]

学习一个映射：

$$f: X^{C \times H \times W} \rightarrow Y^{K \times H \times W} \quad (2.3)$$

其中 C 表示输入的通道数， K 表示数据集的类别数（包含背景）， H 和 W 是输入的高度和宽度。在语义分割中，输入和输出的通道数可能不同，但是大小是一致的，这

是其相较于图像分类最大的不同。在全卷积神经网络中，为了更好地还原输出的尺寸以及得到准确的结果，使用了跳跃连接（Skip）和插值上采样的方式实现。跳跃连接可以同时利用浅层和深层的特征，有效地结合局部信息和全局信息。根据上采样过程的不同，全卷积神经网络可以细分为FCN-8s、FCN-16s和FCN-32s。全卷积神经网络包含五次下采样过程，每个过程由若干个卷积层、激活层和池化层组成。其中，卷积层的作用在于利用不同的卷积核由低级特征合成更为复杂的特征，激活层可以为网络带来非线性因素，使得神经网络可以拟合更复杂的任务，池化层的作用在于增大感受野。每次下采样会将输入图像缩小，将第 x 次下采样的输出称为 $pool_x$ 。下采样可以扩大卷积的感受野，使模型学习到更丰富的全局信息。但是，模型的输出是一个和输入图像等高等宽、维度为类别数的张量，因此需要对下采样得到的高维特征图上采样。图2.3展示了全卷积神经网络三种不同的上采样方式，表2.1描述了对应的公式，其中， UP_n 表示 n 倍上采样， Out 为模型输出，每个像素由一个 K 维向量表示，代表像素点属于各个类别的概率。

表 2.1 全卷积神经网络的三种上采样方式

模型名称	上采样公式描述
FCN-8s	$Out = UP_8(UP_2(UP_2(pool_5) + pool_4) + pool_3)$
FCN-16s	$Out = UP_{16}(UP_2(pool_5) + pool_4)$
FCN-32s	$Out = UP_{32}(pool_5)$

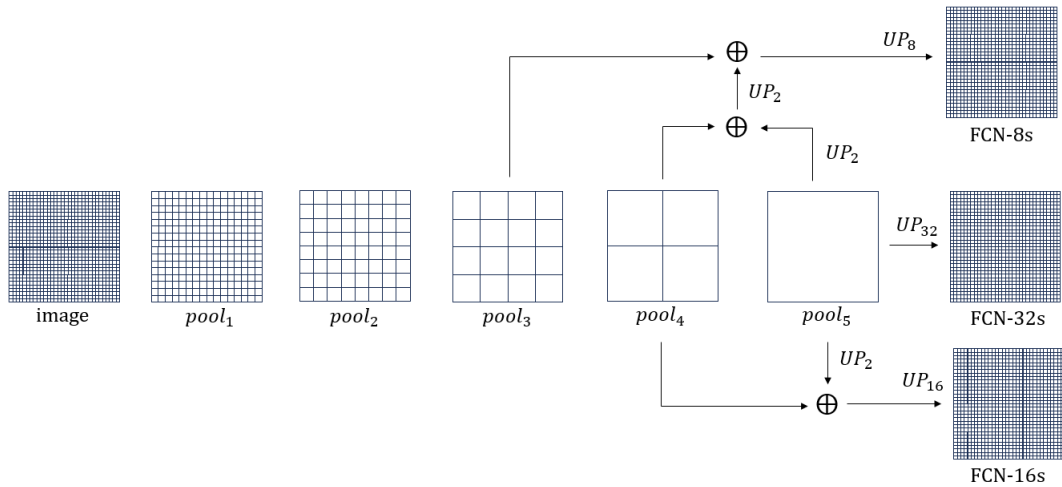


图 2.3 全卷积神经网络的三种上采样方式

同时期还有另一个非常具有影响力的语义分割模型——UNet^[78]。它的名字来源于它的U型结构，它在上采样过程也使用了跳跃的连接方式，不过不同于全卷积神经网络使用加和的方式融合浅层和深层的特征，UNet使用的是拼接（Concatenate）的方式，增加了特征的维度。U形网络架构能够更充分地融合浅层特征和深层特征，这也是UNet在医学影像分割上的性能优于FCN的主要原因。语义分割中如何在增大感受野的同时保留细节信息一直是研究的重点，例如DeepLab^[79]系列模型提出采用空洞卷积代替传统的卷积加池化的操作来扩展感受野，与此同时不会降低特征图的分辨率。DeepLab系列设计了空洞金字塔池化（Atrous Spatial Pyramid Pooling, ASPP）模块，组合不同空洞率的空间卷积产生的特征图，获取更加丰富的上下文信息。后续的许多语义分割模型依旧是遵循编码器-解码器的结构，提出一系列特征提取和融合的方式。

2.2.2 实例分割

现有的实例分割方法大都是遵循自顶向下的策略，需要先定位出实例，再预测实例的类别及掩码。实例分割方法可以分成两阶段和单阶段的。两阶段的方法先使用检测模型框出实例的位置，再对候选框中的像素点赋予类别信息和掩码信息，如Mask R-CNN^[58]（简称MRCNN）。单阶段的方法不再依赖检测的结果，掩码和类别的处理是并行的，如SOLO^[59]和Mask DINO^[80]。

如图2.4, MRCNN是基于Faster R-CNN^[81]的实例分割模型。Faster R-CNN进行目标检测任务并输出检测框（Box）和类别（Class），MRCNN在此基础上增加掩码（Mask）分支，通过感兴趣区域对齐（ROI Align）得到候选框的特征，由掩码分支输出分割结果。这类方法需要基于预设的锚（Anchor）和缩放比例，使用区域建议网络（Region Proposal Network, RPN）扫描图像，得出可能存在实例的区域，使得后续的分类、分割都十分依赖于检测的精度。同时，由于自然界的物体大小范围是无法限定的，预先设定好的锚和缩放比例无法满足所有需求，而为了不遗漏任意大小的物体去无限增加锚的数量，带来的计算量激增是难以忍受的。此外，由于检测框的数量往往是大于真实包含的实例数量，因此需要通过非极大值抑制（Non-Maximum Suppression, NMS）的后处理操作进行检测框的过滤。

受目标检测模型YOLO^[60]的启发，SOLO^[59]应运而生，SOLO的框架如图2.5。

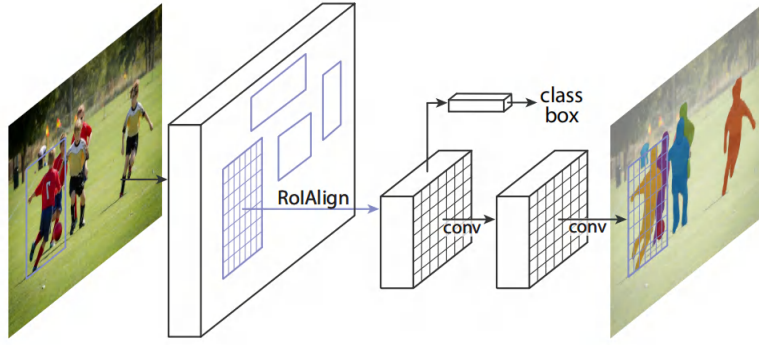


图 2.4 MRCNN框架 [58]

和YOLO一致，它将图像分成 $S \times S$ 的网格（Grid），每个网格负责预测一个对象，如果一个目标的中心落在某个网格内，该网格就负责预测该目标的类别和掩码。不同之处在于，YOLO输出检测框和类别，而SOLO输出掩码和类别。区分图像中相同类别的实例很大程度上取决于实例的中心点坐标和形状，为了处理同类别的实例重叠或粘连的问题，SOLO将不同的实例分配到不同的输出通道上。通过这种方式，和RCNN系列相比，计算速度提高了许多。但是SOLO也继承了YOLO的缺点：当一个格子内有多个物体时，存在漏检的问题；由于去掉了检测框的步骤，对结果的非极大值抑制的后处理需要基于掩码进行，相较于基于检测框的非极大抑制增加了非常多的计算量。

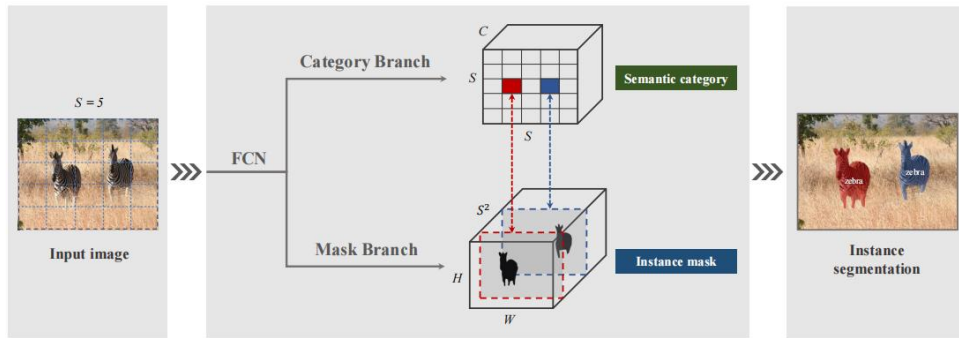


图 2.5 SOLO框架 [59]

得益于Transformer^[82]强大的编码能力和解码能力，DETR^[62]模型首次将Transformer引入目标检测任务，提供了一个目标检测新范式。Transformer编码器的输入是图像经过卷积神经网络后得到的特征与位置编码的加和，通过Transformer的注意力机制，可以获取全局上下文信息。Transformer的解码器接收编码器的输出和目标查询（Object Queries），每个查询对应图像中的一个物

体。最后，前馈网络（Feed-forward Network, FFN）根据各个查询的特征得到目标的类别和位置。后续涌现了许多基于DETR改进的模型，例如DINO^[83]。Mask DINO^[80]则是在DINO的基础上加上掩码分支得到的实例分割模型。可以看出，基于查询的模型框架更为简单，不再依赖于锚和网格，扩展性很强，具有很大的潜力，因而成为现在研究的热点。

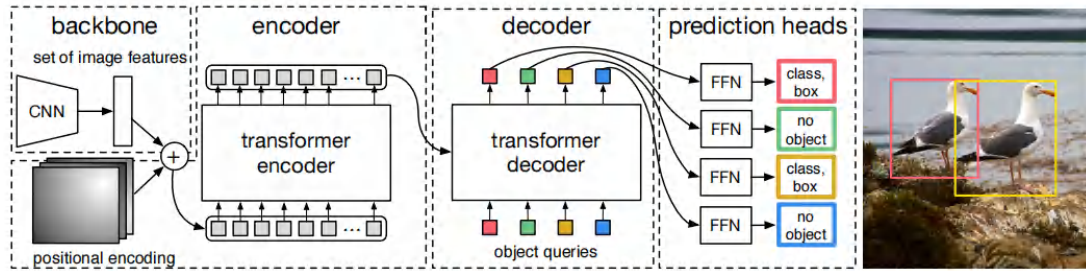


图 2.6 DETR框架^[62]

2.3 注意力机制

感知智能的终极目标是可以像人类一样处理接收的信息，人类的视觉系统可以瞬间接收海量的数据并在霎那间筛选出感兴趣区域，这启发了注意力机制的产生。注意力机制在深度神经网络中已经得到了广泛的应用，通过注意力机制，神经网络可以关注到更多重要的信息，提高模型的性能。根据关注度的来源，可以将经典的注意力机制分为三种：空间注意力、通道注意力和混合注意力。

由于文本信息天然的序列性和前后相关性，注意力机制起初在自然语言处理领域取得了很大的成功。空间注意力受自然语言处理的自注意力机制启发，由自然语言迁移到图像上。元素经过神经网络得到三个新的向量，分别称为查询（Query）、键（Key）和值（Value）。查询特征用来“主动地”和其他向量匹配，键特征则“被动地”接收其他向量的匹配，值代表当前向量的重要特征。自注意力机制是每一个元素使用自己的查询特征与其他所有元素的键计算相似度，根据各个匹配过程的相似度来加权匹配的值从而更新自己的特征的过程。多头注意力机制是自注意力机制的变体，通过使用多个独立的注意力头，每个元素可以有有组查询、键和值，分别计算注意力权重，并将它们的结果融合，从而获得更丰富的表示，增强模型的表达能力和泛化能力。非局部注意力^[84]（Non-local Attention）是将自注意力机制引入

图像的重要工作，它的非局部主要是为了和卷积神经网络中卷积核局部的感受野进行比较，通过将图像中的像素当作独立的元素，使用自注意力机制可以获取其他所有像素点对自身特征更新的贡献。它的计算方式如图2.7，其中 θ 、 ϕ 和 g 用来获得查询、键和值，后续操作和自注意力机制如出一辙。

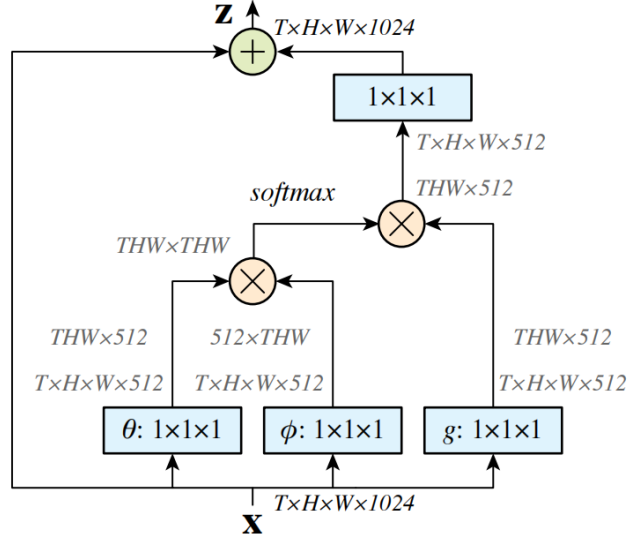
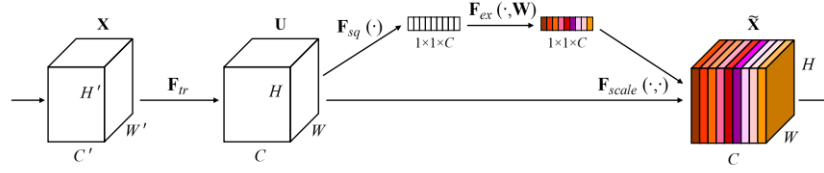


图 2.7 非局部注意力^[84]

对于图像，经过网络处理后的特征各个维度往往包含了不同的信息，例如在UNet的跳跃连接中，不同通道的信息来自不同层级的特征。因此，通道注意力机制将各个通道视为独立的元素，计算各个通道的重要性。例如，图2.8的SENet^[85] (Squeeze-and-Excitation Networks)，其工作可以分为三步。首先，SENet将得到的特征跨越空间维度进行压缩产生通道描述符，压缩后保留了通道数，但各个通道上只有一个元素。该通道描述符嵌入了通道特征响应的全局分布，使来自网络全局感受野的信息能够被其较低层利用。之后，通过神经网络的映射为每个通道学习特定采样的激活，控制每个通道的激励值。最后特征映射被激励值重新加权以生成通道注意力的输出。通常，SENet会与其他注意力机制共同使用，来增强表达能力。

由于空间信息和通道信息都十分重要，因此诞生了混合注意力机制，不仅能学习不同区域的重要性，还能学习不同通道特征的重要性，例如CBAM^[86]和DANet^[87]。

图 2.8 SENet的压缩和激活模块^[85]

2.4 伪标签

伪标签是半监督学习中提出的一个概念，指的是模型给无标签数据赋予的用于监督训练的标签，和已有数据的真实标签进行区分。在小样本学习的任务中，强调了有标签数据的稀缺。在一些情形下，原始数据是丰富的，但是为这些数据作标记的成本是很昂贵的，这种情况十分常见。例如，对于医学图像和材料图像，数据标注需要专业的背景知识。在这种情况下，半监督学习应运而生。半监督学习介于有监督学习和无监督学习之间，前者使用的数据都是带有标签的数据，而后者则相反。半监督学习使用有标签数据预先对模型进行训练，由此得到的模型就有了一定的预测能力，可以给相同领域的无标签数据进行较可靠的预测。对于这些预测，选择置信度高的预测及其原始数据对模型进行监督训练。使用伪标签技术的关键点在于设置合适的阈值筛选伪标签以及设计模型的训练更新方式。

伪标签的工作流程如图2.9，可以分为以下几个步骤：

- 使用带标签数据训练目标模型H；
- 使用模型H对无标签数据预测伪标签；
- 使用带标签数据和无标签数据一起有监督地训练模型H'；
- 根据模型H'更新模型H直至稳定。

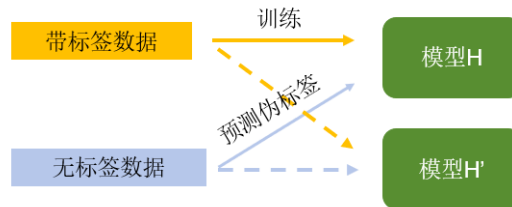


图 2.9 伪标签工作流程

目前，伪标签技术已经得到一定程度的发展，例如Match系列：MixMatch^[48]、ReMixMatch^[88]和FixMatch^[49]。

2.5 评价指标

平均交并比（Mean Intersection over Union, mIoU）、平均准确率（Mean Accuracy, mAcc）、平均Dice系数（mDice）和平均精度（Average Precision, AP）是图像语义分割和实例分割常用的评价指标。不论是语义分割还是实例分割，都涉及分类的过程，因此可以在分类的评价指标上拓展延伸得到相应的分割评价指标。上述评价指标基于混淆矩阵计算得到，混淆矩阵如表2.2。由混淆矩阵可以直接得到两个

表 2.2 混淆矩阵

	预测为正例	预测为负例
实际为正例	真正例 (TP)	假反例 (FN)
实际为负例	假正例 (FP)	真反例 (TN)

评价指标：精度（Precision, p ）和召回率（Recall, r ）。精度表示真正例占有所有预测正例的比例：

$$p = \frac{TP}{TP + FP} \quad (2.4)$$

召回率表示真正例中被预测为正例的比例：

$$r = \frac{TP}{TP + FN} \quad (2.5)$$

交并比计算的是预测和真实标注之间重合的比例，平均交并比是对数据集中所有类别（包含背景）的交并比计算平均值后的结果。mIoU的计算可以表示为：

$$mIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{FN + FP + TP} \quad (2.6)$$

其中， $k+1$ 表示数据集中类别数。

平均准确率是指所有类别的像素分类正确率的平均值：

$$mAcc = \frac{1}{k+1} \sum_{i=0}^k \frac{TP + TN}{FN + FP + TP + TN} \quad (2.7)$$

Dice系数可以用来衡量两个集合的相似度，它的形式主要有两种：

$$Dice = \frac{2|X \cap Y|}{|X| + |Y|} \quad (2.8)$$

和

$$Dice = \frac{2 \sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \quad (2.9)$$

在公式(2.8)中, $|X \cap Y|$ 表示集合 X 和集合 Y 的交集包含的元素个数, $|X|$ 和 $|Y|$ 分别表示集合 X 和集合 Y 的元素个数。在公式(2.9)中, N 表示元素个数, p_i 表示预测值, g_i 表示真实标签, g_i 是二进制元素。和混淆矩阵比较, 结合图2.10可以发现,

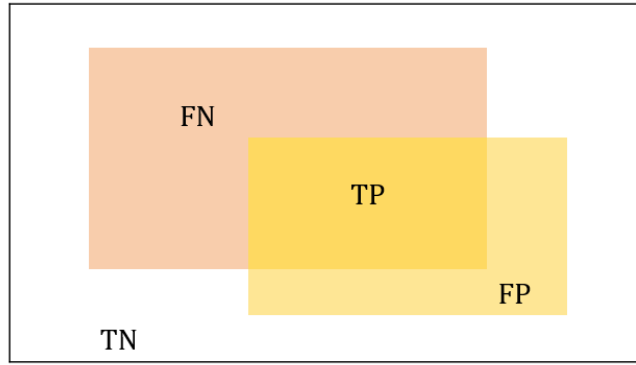


图 2.10 混淆矩阵可视化

$|X \cap Y|$ 等价于真正例 TP , 而 $|X|$ 和 $|Y|$ 等价于 $2TP + FP + FN$ 。所以, 基于混淆矩阵的Dice系数表示为:

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (2.10)$$

可以看出, 本质上, 在衡量两个集合的相似度时, Dice系数结合了精度和召回率, 是精度和召回率的调和平均数, 等价于 F_1 分数:

$$\begin{aligned} F_1 &= 2 \frac{pr}{p+r} \\ &= 2 \frac{\frac{TP}{TP+FP} \frac{TP}{TP+FN}}{\frac{TP}{TP+FP} + \frac{TP}{TP+FN}} \\ &= \frac{2TP}{2TP + FP + FN} \end{aligned} \quad (2.11)$$

其中, p 和 r 分别表示精度和召回率, 它们的计算如公式(2.4)和公式(2.5)。因为分母中存在重复计算, 为了保证Dice系数在 $[0, 1]$ 之间, 将分子的系数设为2。与mIoU的计算类似, mDice是多个类别Dice系数的平均值。

实例分割最常见的评价指标是平均精度，最早来源于Pascal VOC Challenge^[89]，平均精度是精度-召回曲线（ $p-r$ 曲线）下的面积，是一种综合评价指标。通过改变决策阈值，即预测为正类的最小分数，可以得到一组精度和召回率，以精度为纵轴，召回率为横轴，在二维坐标系中标出每组精度和召回率的坐标点，按顺序连接这些点，即可得到 $p-r$ 曲线。曲线与横轴的面积采用化整为零的思想计算，也就是拆解成多个以相邻召回率为底、精度为高的矩形的面积和。计算公式如下：

$$AP = \sum_n (r_n - r_{n-1}) p_n \quad (2.12)$$

对所有类别的AP值求平均可以得到mAP。平均精度的计算是通过划分置信度计算得来的，由平均精度AP演化出的AP₅₀和AP₇₅不仅考虑置信度还考虑了交并比IoU。AP₅₀和AP₇₅先计算预测结果和真实结果的交并比IoU，高于阈值（50或75）的预测结果被筛选进入置信度划分的环节。为了衡量实例分割模型对不同尺寸物体的分割性能，由AP还衍生出了AP_S、AP_M和AP_L，分别对应小目标、中等目标和大目标的精度。

2.6 本章小结

本章阐述了本论文使用到的关键技术和基础理论。首先介绍广泛使用的骨干网络——残差网络的相关理论；其次，介绍最基础的语义分割和实例分割模型；接着，是对几种注意力机制的工作原理的介绍；然后，对小样本学习中的伪标签技术进行介绍；最后，为了客观地评价本论文的方法，本章介绍了使用的评价指标。

第三章 基于双分支语义分割网络的图像分割和材料性能挖掘

深度学习中关于图像的任务有分类、检测、分割等，图像分类的任务粒度在图像级，图像检测的任务粒度在物体级，而图像分割的任务粒度达到了像素级。因此，图像分割对于图像中信息的挖掘是最深入和细致的。为了得到材料图像和材料性能的关系，本论文基于图像分割方法开展了一系列研究。针对本论文的示范材料——2205双相不锈钢的微观组织形貌图像，本章提出基于十字交叉注意力的双分支语义分割网络，可以在小样本环境下准确地提取2205双相不锈钢中二次相的信息，还原出材料性能的变化。

3.1 方法概述

本章基于材料图像的材料性能挖掘方法可以分成三步：图像语义分割、二次相尺寸信息提取以及材料性能挖掘，图3.1展示了本章方法的流程。首先，本章提出了一个适用于小样本2205双相不锈钢图像数据集的双分支语义分割网络用来得到二次相在图像中的分布。接着，针对语义分割结果的特点，即只区分类别、不区分个体，边界跟踪算法^[90]被用来区分各个二次相个体，在边界跟踪的过程中，各个二次相的尺寸信息也得到采集。最后，基于这些尺寸信息借助奥斯特瓦尔德熟化机制^[25]分析双相不锈钢的性能变化。

图3.2展示了本章提出的双分支语义分割网络，上半部分是主分支，使用了特征金字塔和金字塔池化结构；下半部分是辅助分支，使用了十字交叉注意力机制。通过特征金字塔丰富的层级语义信息以及轻量级注意力机制的辅助分割，本章提出的方法以最小的参数量代价提高语义分割的性能。本小节将对使用的骨干网络、双分支的实现细节以及使用的损失函数进行介绍。

3.1.1 骨干网络

本章的语义分割网络使用空洞残差网络^[91]（Dilated Residual Network, DRN）作为骨干网络，如图3.2。残差网络最初面向的是分类任务，通过卷积和池化操作不

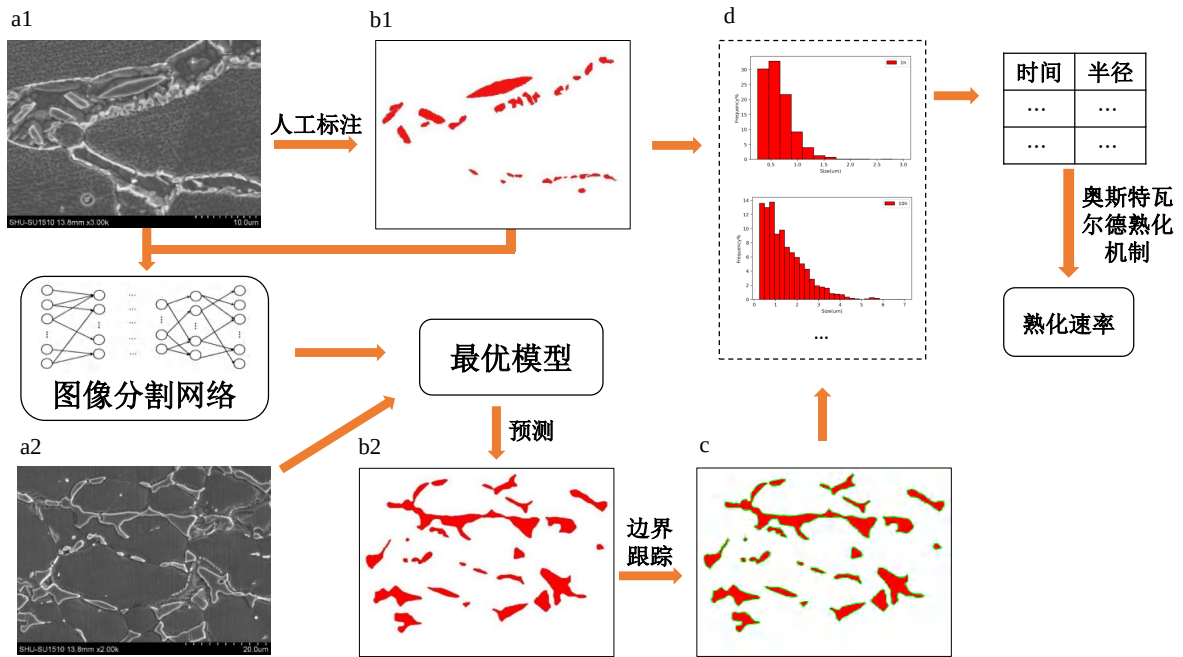


图 3.1 基于图像分割的材料性能挖掘方法流程图：a1：带有人工标注的图像样本；b1：a1对应的标注图；a2：无标注样本；b2：由最优模型预测的二次相的分布图；c：通过边界跟踪算法得到的二次相的轮廓；d：各个时刻的半径直方图。标注图像中的红色部分代表二次相，也就是前景。

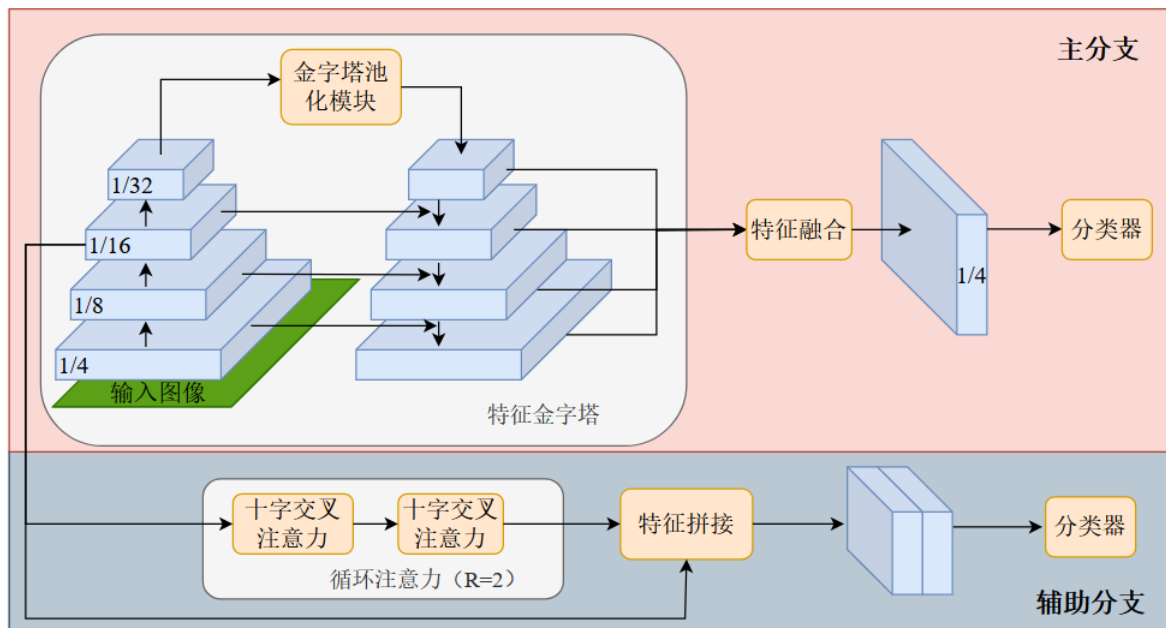


图 3.2 双分支语义分割网络

断缩小图像，直到其变为一个尺寸微小的特征图，最后通过全连接操作预测类别概率。在下采样的过程中，各层网络得到的特征图中的每个元素蕴含的全局信息不断增加，网络的感受野不断变大。然而，获得全局信息就意味着不可避免地需要舍弃细节信息。为了保留材料图像的底层纹理和边界信息，需要额外的技术弥补残差网络中细节丢失的缺陷。相较于普通的卷积，空洞卷积既可以保证获得全局感受野，又不损失图像空间的分辨率^[92]，如图3.3是不同空洞率下的空洞卷积示意图，绿色区域代表感受野，橙色点标出的元素代表参与卷积运算的元素。可以看出，当空洞率（也称作膨胀率）为1时，空洞卷积和一般的卷积操作相同，所以，可以将一般的卷积看作是空洞卷积的空洞率为1时的特殊情况。当空洞率为2时，感受野由原来的 3×3 变为 7×7 ，但是参与卷积运算的元素个数不变。借助空洞卷积，可以避免池化操作带来的信息损失，还可以在不增加参数量的情况下扩大感受野，增强网络的表征能力，保留图像中的细节信息。空洞残差网络在残差网络的基础上，将最后两个残差模块中的普通卷积替换成空洞卷积^[91]，空洞率分别为2和4。

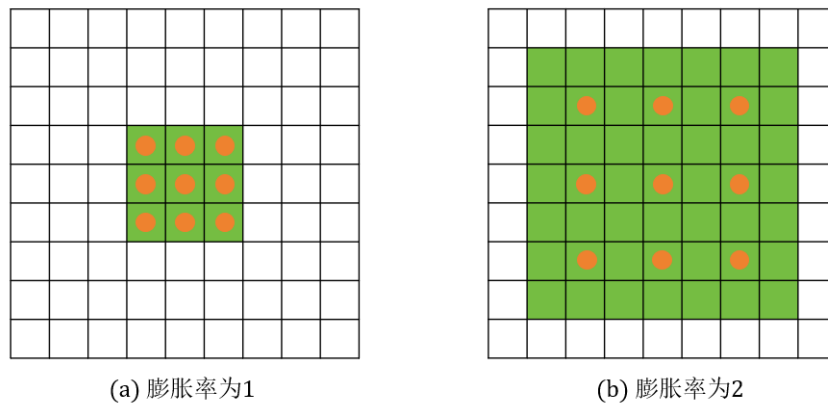


图 3.3 不同空洞率下的空洞卷积

3.1.2 主分支

在主分支中设计了特征金字塔架构，对特征金字塔的各层输出进行聚合，得到分割前的特征图，如图4.1。特征金字塔是一种有效的特征提取器，在一个固有的金字塔式的层次结构中充分利用多层特征。它使用带有横向连接的自顶向下体系结构，将高级语义信息融合到中低级别实现特征解码。基于特征金字塔，网络可以保留高层特征的语义信息和低层特征的细节信息，为语义分割提供更全面的视觉信息，提

高分割的精度。Zhou等人发现，虽然深度卷积神经网络的理论感受野足够大，但实际的感受野其实达不到理论值^[93]。因此，参考统一感知解析网络（Unified Perceptual Parsing Network, UperNet）的设计思想^[94]，本章应用金字塔池化模块（Pyramid Pooling Module, PPM）^[95]处理骨干网络得到的最高层特征，之后将得到的结果送入金字塔自上而下的结构中，金字塔池化模块的实现细节如图3.4，特征图经过自适应池化后的大小分别为 1×1 、 2×2 、 3×3 和 6×6 ，卷积用来改变维度。

为描述方便，通常将残差网络中每个残差块得到的特征图表示为 C_2 、 C_3 、 C_4 和 C_5 ，将特征金字塔得到的特征图表示为 P_2 、 P_3 、 P_4 、 P_5 ，其中，降采样率分别为4、8、16、32。 P_5 的下采样率比较大，使得全局平均池化后的特征更加关注高级语义，因此，融合特征金字塔的所有特征图比只使用特征金字塔分辨率最高的特征图（ P_2 ）更好。

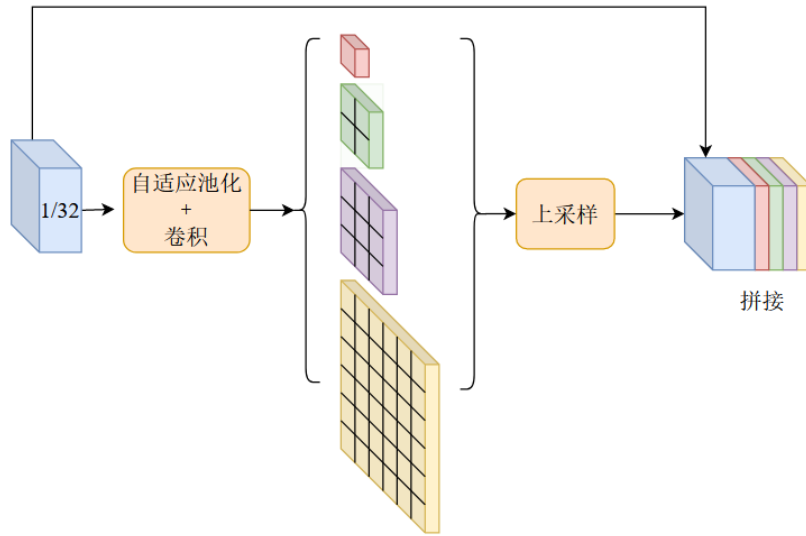


图 3.4 金字塔池化模块

3.1.3 辅助分支

辅助分支是进行跳跃连接的分支，直接对骨干网络的第三个残差块输出的特征 C_4 使用计算代价小的循环十字交叉注意力机制提取图像特征，并和输入的原始特征拼接，最后完成分割。辅助分支基于低层特征实现语义分割，使得低层网络的梯度更新路径更短，加强对边界、纹理等细节信息的学习。材料图像和自然图像关注的特征存在一些差异，具体表现为，自然图像的分割更多关注物体的形状信息，而

对于材料图像分割，相的纹理信息和边界信息至关重要。因此，辅助分支可以进一步提升材料图像语义分割的准确性。

受自然语言处理的启发，注意力机制获得长程信息的能力已得到证实。当前，在计算机视觉任务中，注意力机制已经广泛地应用在编码器中，用来引导神经网络关注特定区域的信息。因此，为了增加特征包含的全局信息，辅助分支使用空间注意力机制进一步处理骨干网络的低层特征 C_4 。2.3节介绍的非局部空间注意力机制中，某个位置的响应是其他所有元素特征的加权和，通过这种方式，每个元素会获得其他所有元素的信息，从而获得全局上下文信息。其中，加权的权重值是匹配对的键和值之间的相似度，一般以标准化内积相似度进行衡量。对一张特征图 $X \in R^{C \times H \times W}$ ，以非局部注意力机制的方式获得其中一个位置的响应，需要计算 HW 个对应的权重，而完成整张特征图的注意力计算，则需要计算 $(HW)^2$ 个对应的权重。显然，这样的时间消耗是很难容忍的。因此，在本工作中，使用的是交叉交叉注意力机制（Criss-cross Attention, CCA）^[96]，可以有效减少计算量，但不衰减性能。

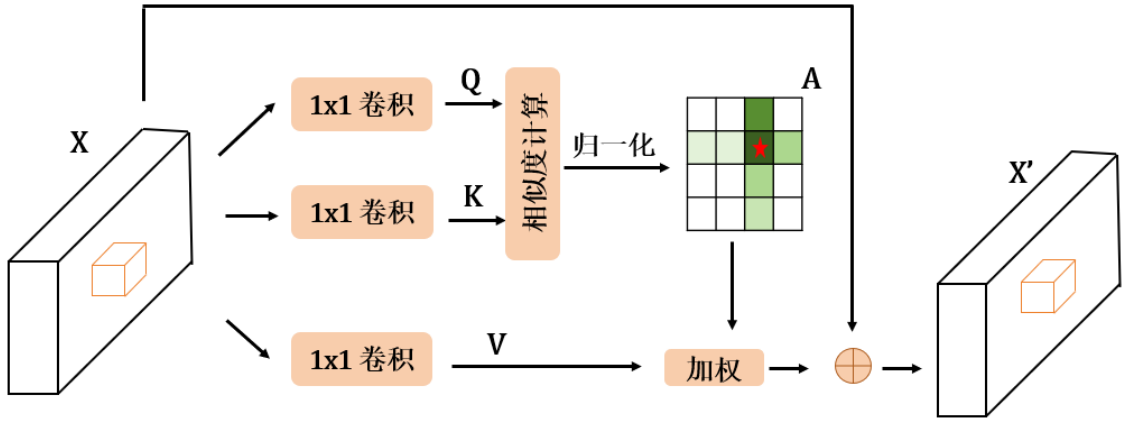


图 3.5 十字交叉注意力机制

图3.5展示了十字交叉注意力机制的计算流程。对于输入 X ，十字交叉注意力机制首先遵循自注意力机制的原则，通过卷积层得到相应的矩阵，它们分别是查询矩阵 $Q \in R^{C \times H \times W}$ 、键矩阵 $K \in R^{C \times H \times W}$ 和值矩阵 $V \in R^{C \times H \times W}$ ，权重的计算和元素的响应机制也与非局部注意力机制相同，即通过前两个矩阵计算相似度得到注意力图（Attention Map），然后通过值矩阵上加权求和来更新元素。不同之处在于，十

字交叉注意力机制并不计算其余所有元素的权重，只计算和其空间上位于同行或同列的元素的权重。在图3.5中，用★表示一个待计算响应的位置 $u = (x, y)$ ，绿色区域则是得到该响应需要计算权重的位置，用 Ω_u 表示该区域。对于 $p \in \Omega_u$ ，其权重值的计算为：

$$A_{u,p} = Q_u K_p^T \quad (3.1)$$

而其注意力图的空白区域权重可以视作0。位置 $u = (x, y)$ 的响应按如下方式计算：

$$X'_u = \sum_{p \in \Omega_u} A_{u,p} V_p + X_u \quad (3.2)$$

通过这样的定义，每个位置的响应只需要计算 $(H + W) - 1$ 个对应的权重，大大地减少了计算量。但是，这样的注意力机制无法获得全局的上下文信息。当通过十字交叉注意力机制后，对于特征中的每个元素，其同行、同列的元素也得到了更新，获得了它们对应的同行、同列的元素信息，从而得到新的特征图 X' 。具体地说，和其处于同行的元素，在更新为新的响应的过程中，会分别获得图像每一列的信息，而和其处于同列的元素，在更新为新的响应的过程中，会分别获得图像每一行的信息。根据传递性原理，当以新的特征图 X' 再次进行十字交叉注意力时，每个位置都间接地得到了全局的上下文信息。因此，在得到空洞残差网络输出的特征图以后，以尽可能少地增加计算量且得到全局上下文信息为目标，辅助分支重复两次十字交叉注意力机制以得到最终的特征图，如图3.2。

3.1.4 损失函数

本章提出的双分支语义分割网络以多任务的方式监督网络的训练，其中，辅助分支仅在训练阶段增加一个辅助损失^[95]，对预测结果不产生影响。在总的损失函数中，使用权重来平衡辅助损失：

$$L = L_{main} + \alpha L_{aux} \quad (3.3)$$

其中， L_{main} 是主分支的损失， L_{aux} 是辅助分支的损失， α 是辅助分支损失的权重。

在香农信息论中^[97]，从物理中粒子的混乱程度吸收了关于熵的理论：微观粒子活动越混乱，系统的熵越高。衡量机器学习分类算法的预测和真实标注之间的差

异，等价于衡量对于输入变量 X 预测的标签分布 $q(x)$ 和真实标签分布 $p(x)$ 之间的距离。在信息论中，两个分布的距离可以使用KL散度衡量：

$$\begin{aligned} D_{KL}(p||q) &= \sum_i p(x_i) \log \left(\frac{p(x_i)}{q(x_i)} \right) \\ &= \sum_i p(x_i) \log p(x_i) - \sum_i p(x_i) \log q(x_i) \end{aligned} \quad (3.4)$$

KL散度也称作相对熵。因此，模型的任务可以等价为最小化KL散度。对于确定的需要学习的数据集，真实标签分布已经固定，所以 $p(x)$ 是确定的常量。因此最小化KL散度等价于最小化交叉熵（Cross Entropy，CE）^[98]：

$$H(p, q) = - \sum_i p(x_i) \log q(x_i) \quad (3.5)$$

交叉熵损失函数有良好的数学特性，如可导性、凸性等，非常便于使用梯度下降法寻找最优解。语义分割是像素级的分类任务，借鉴分类任务的经验，本章的主分支和辅助分支都使用交叉熵衡量预测结果与真实标签的差异，对于二分类任务，其表示为：

$$L_{ce} = - \frac{1}{N} \sum_{i=1}^N y_i \log q_i + (1 - y_i) \log (1 - q_i) \quad (3.6)$$

其中， N 代表训练样本的数量， y_i 是第 i 个样本的真实标签（负类0或正类1）， q_i 是模型预测第 i 个样本属于正类的概率。公式(3.6)与公式(3.5)是等效的。

3.2 实验分析

3.2.1 2205双相不锈钢图像数据集介绍

本论文的材料图像来自15mm热轧态的2205双相不锈钢，所有样品均在1100°C下进行固相溶液处理30分钟，然后立即在水中淬灭，在850°C时效不同时间（10min、20min、30min、1h、2h、4h、10h、20h、50h、100h、200h），利用扫描电子显微镜/能量色散光谱仪（Scanning Electron Microscope/Energy Dispersive Spectrometer，SEM/EDS）和透射电镜（Transmission Electron Microscope，TEM）采集其显微组织形貌图像。该数据集一共包含352张图像，均为1280 × 960的灰度图，其中211张图像手工标注了目标相的分布，其余141张图像未标注。在观测前

期, χ 相占主导地位, 随着时间的推移, σ 相的增长速度快于 χ 相, 导致 χ 相会转换成 σ 相。所以, 在这些采样时刻中, 30min、1h和2h标注的是二次相 χ 相, 而其余时刻标注的是 σ 相, 为了得到更准确的结果, 本论文以此为依据将2205双相不锈钢数据集分为两部分, 分别训练语义分割网络并进行预测。各个时刻的有标签数据和无标签数据分布如表3.1。

表 3.1 2205双相不锈钢图像数据分布

	30min	1h	2h	4h	10h	20h	50h	100h	150h	200h	总计	
											χ 相	σ 相
有标签	12	18	12	18	22	25	31	25	8	40	42	169
无标签	22	17	6	21	16	11	1	27	9	11	45	96

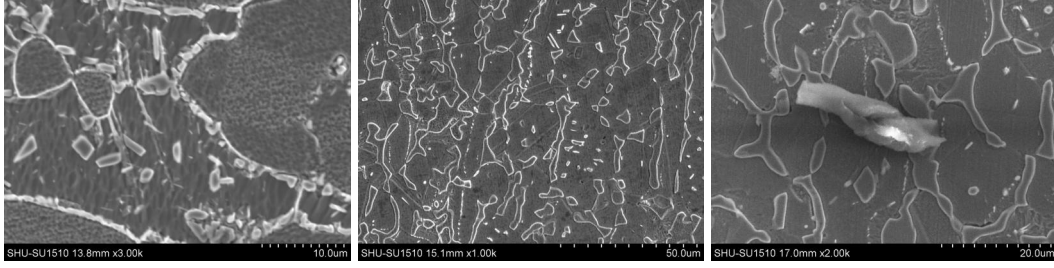
图3.6展示了一组图像及其手工标注的目标相分布图, 可以看出, 图像中一共有两个类别, 红色标注的目标相所在区域为前景, 白色部分为背景, 原图中最下方的黑底白字部分不做删除。在手工计算方法中, 材料领域的研究人员直接通过这211张图像统计目标相的尺寸分布。在本工作中, 通过分层(根据不同时刻分层)抽样的方式将其中 10%作为验证集, 剩余的作为训练集, 按照图3.1所示流程完成实验, 训练得到的模型在剩余的141张未标注图像上预测目标相的分布, 最后使用所有图像的二次相分布图统计相的尺寸分布。

3.2.2 实验设置

本实验在RTX 3090完成, 显存24GB, 使用的操作系统是Ubuntu20.04。在训练过程中, 由于显存不足, 训练图像统一裁剪为 512×512 后再送入模型计算, 每个批次有4张图像一起输入, 即batch size为4。实验采用了线性下降(PolyLR)的学习率下降策略, 初始学习率 LR_{Base} 为0.01。第 i 次迭代的学习率为:

$$LR_i = (LR_{Base} - LR_{min}) \times (1 - i/Iter)^{power} + LR_{min} \quad (3.7)$$

其中, LR_{min} 是最小学习率, 为0.0001, $Iter$ 是迭代的次数, 在 χ 相的训练中, $Iter = 2000$, 在 σ 相的训练中, $Iter = 8000$, $power$ 控制学习率下降的速度, 设为0.9。训练过程使用的优化器是随机梯度下降(Stochastic Gradient Descent, SGD), 动量为0.9。



(a) 图像示例



(b) 对应的标签可视化

图 3.6 三组2205双相不锈钢图像和标签示例

本章使用的评价指标是交并比IoU、平均准确率mAcc和Dice系数。

3.2.3 消融实验

为了探究本章提出的各个模块的有效性以及模块中超参数的影响，本章设计了一系列消融实验，具体包括各个模块的有效性、辅助分支的损失权重和十字交叉注意力的循环次数。

为了探究加入的各个模块的有效性，本章对空洞残差网络、双分支结构、特征金字塔以及十字交叉注意力机制几个部分进行消融实验。 χ 相的消融实验结果如表3.2所示， σ 相的消融实验结果如表3.3所示，括号中的数字代表相较于上一行模型的指标变化。首先使用一个基模型进行分割实验，该基模型以残差网络作为骨干网络，使用全卷积神经网络的分割头进行语义分割。在此基础上，依次进行以下操作改变网络结构：将骨干网络改为空洞残差网络、加入额外的全卷积神经网络分割头构造双分支结构、将主分支的全卷积分割头替换为由特征金字塔和金字塔池化结构组成的分割头以及将辅助分支的全卷积分割头改为基于循环十字交叉注意力的分割头。在模块的消融实验中，辅助分支的损失权重设为0.4，十字交叉注意力机制的循环次数设为2。可以看出每一次的操作使得网络在两个子数据集的五个定量衡量指标上都有提升，验证了本章做出的改进的有效性。通过每一次的性能增加还可以发

现，不论是在 χ 相还是在 σ 相的分割实验中，每次增加的模块对IoU和Dice系数的提升是最大的。IoU和Dice系数衡量了模型对前景类的分割性能，它们的提升意味着在基模型中，对前景的分割是欠缺的，而增加的模块非常准确地关注到了对前景相的学习还不足的问题，取得了期望的结果，即模型可以更加关注对性能分析有用的前景相的分割而不是背景。

表 3.2 不同模块在 χ 相的消融实验结果

模块	IoU	mIoU	mACC	Dice	mDice
基模型	49.02	72.89	80.04	65.79	82.07
+空洞残差网络	50.17(+1.15)	73.54(+0.65)	80.04(+0)	66.82(+1.03)	82.62(+0.55)
+双分支	51.01(+0.84)	73.98(+0.44)	80.53(+0.49)	67.56(+0.74)	83.01(+0.39)
+特征金字塔	53.23(+2.22)	75.21(+1.23)	80.96(+0.43)	69.48(+1.92)	84.03(+1.02)
+十字交叉注意力	56.23(+3)	76.77(+1.56)	83.19(+2.23)	71.98(+2.5)	85.31(+1.28)

表 3.3 不同模块在 σ 相的消融实验结果

模块	IoU	mIoU	mACC	Dice	mDice
基模型	63.19	78.26	87.2	77.45	87
+空洞残差网络	64.4(+1.21)	78.9(+0.64)	88.44(+1.24)	78.34(+0.89)	87.47(+0.47)
+双分支	65.88(+1.48)	79.78(+0.88)	89.37(+0.93)	79.43(+1.09)	88.08(+0.61)
+特征金字塔	68.87(+2.99)	81.65(+1.87)	90(+0.63)	81.56(+2.13)	89.35(+1.27)
+十字交叉注意力	71.65(+2.78)	83.28(+1.63)	91.6(+1.6)	83.48(+1.92)	90.44(+1.09)

在图3.7中展示了基模型和本章提出的模型得到的特征可视化图，由于特征是高维数据，本章使用主成分分析技术（Principal Component Analysis, PCA）和t分布式随机邻居嵌入（t-distributed Stochastic Neighbor Embedding, t-SNE）^[99]对特征进行降维和可视化。第一行是 χ 相及其背景的特征分布，第二行是 σ 相及其背景的特征分布，用绿色代表前景，橙色代表背景。可以直观地看出本章提出的模型准确地将两个类别分开，而基模型学习的特征分布十分混乱，不利于构建决策超平面。

对模块的消融实验的可视化结果如图3.8所示，该图展示了上述消融实验在两张图像上的预测结果，使用蓝色标记的是包含 χ 相的图，使用绿色标记的是包含 σ 相的图。在真实标注中，用红色方框标出了最值得关注的区域。在包含 χ 相的图中可以看出，随着每一步操作，未被预测出的前景相逐渐被挖掘并得到更精确的掩码。在

包含 σ 相的图中，每一步操作都使红框区域的分割结果逐步接近真实标注，残缺的部分逐渐得到补足。

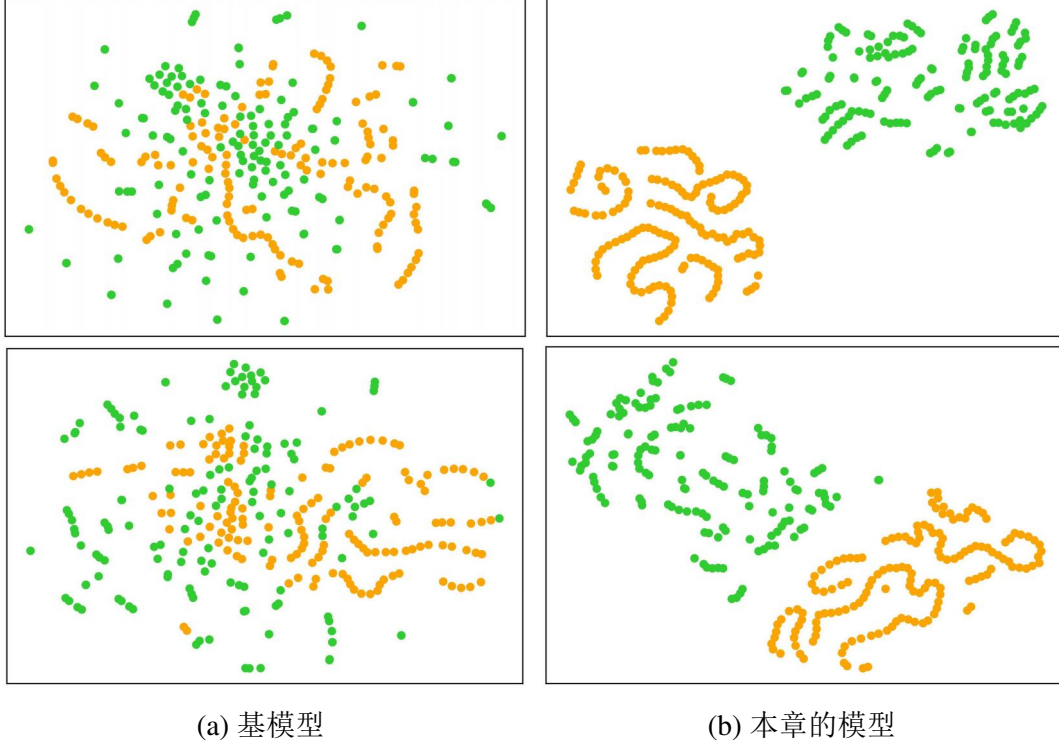


图 3.7 基模型和本章的模型在 χ 相和 σ 相的特征分布可视化

辅助分支的加入带来了新的损失，公式(3.3)使用 α 调节辅助损失的权重。本章通过实验探讨了 χ 相和 σ 相上辅助分支损失值权重对分割性能的影响，图3.9展示的是在0到1之间以0.1的步长遍历权重取值后得到的结果。从图中并未发现明显的规律，但是可以找到权重的最优值，分别是0.2和0.5。同时，从图中可以发现，对于 χ 相的分割，当辅助分支的权重不合适时，会带来负面的影响，但对于 σ 相的分割没有发现这种情况。

以一张 512×512 的灰度图像作为输入，对十字交叉注意力机制的循环次数进行消融实验，实验结果如表3.4。其中，训练 χ 相时辅助分支的损失权重设为0.2，训练 σ 相时辅助分支的损失权重设为0.5。由于循环十字交叉注意力机制中是重复使用一个十字交叉注意力模块，因此空间占用上并不会有很大增长。在推理阶段，只需使用主分支，所以无需考虑推理速度。从表中可以看出，当循环次数到达2时，继续增加至3已经无法带来很高的增益，因此在本章中将循环次数设置为2。

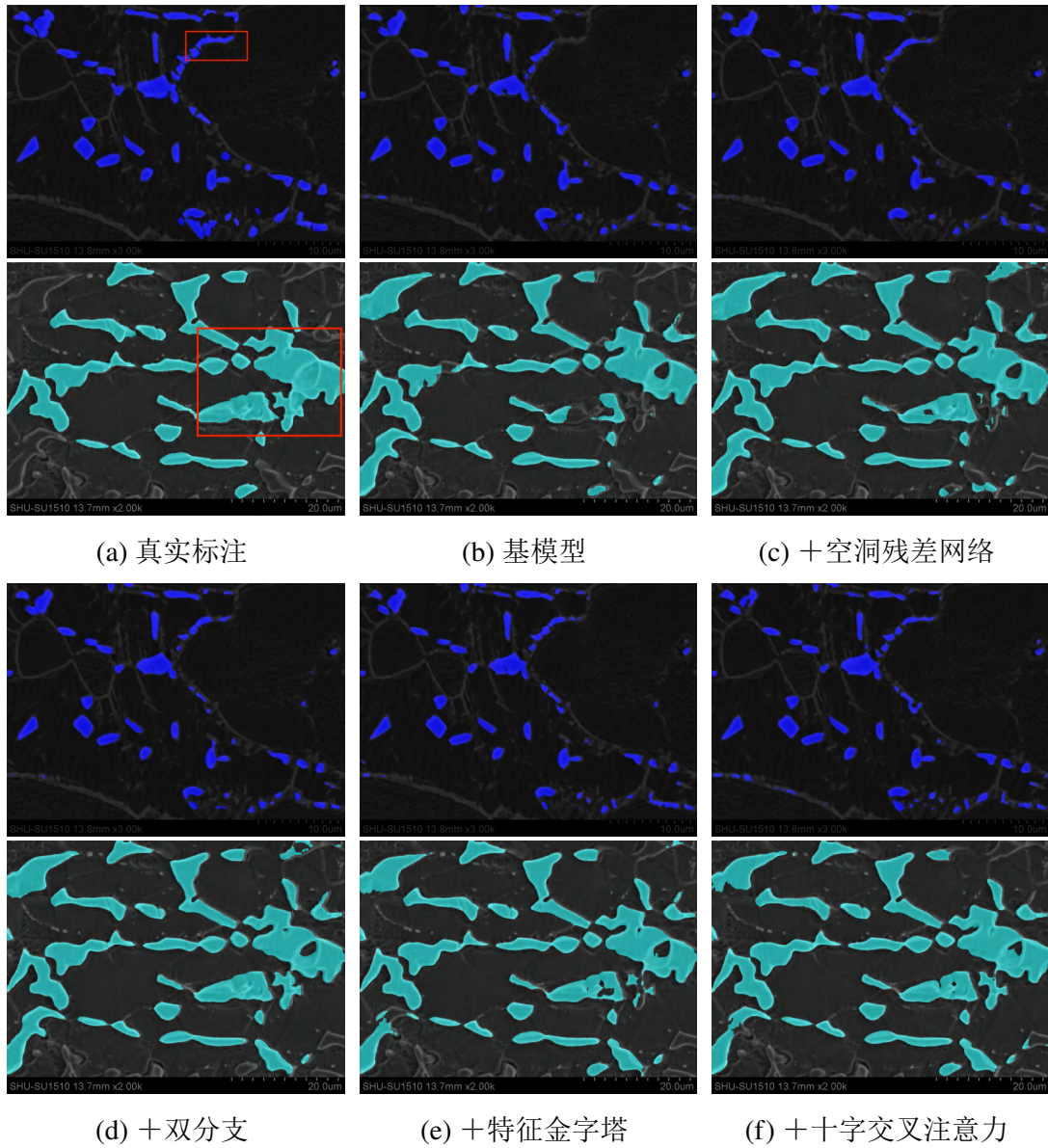


图 3.8 不同模块组合的分割结果展示

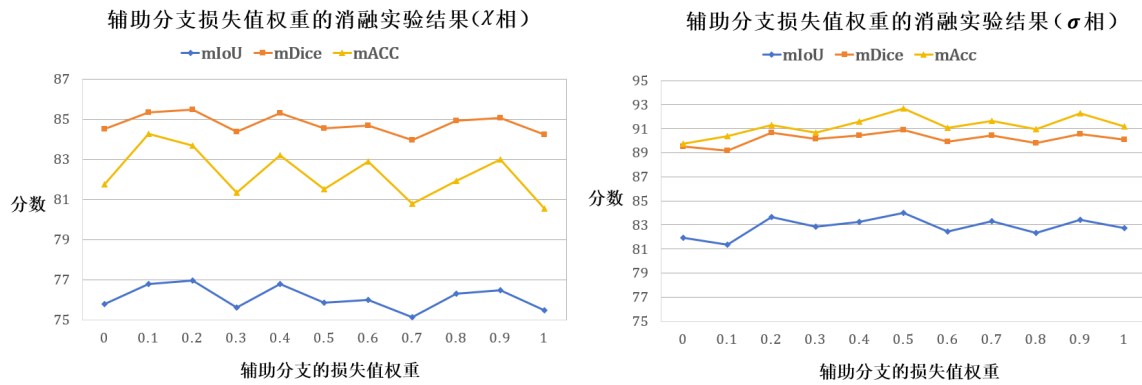


图 3.9 辅助分支损失函数值权重的消融实验结果

表 3.4 十字交叉注意力机制循环次数的消融实验结果

循环次数	占用空间 (kB)	mIoU		mAcc		mDice	
		χ 相	σ 相	χ 相	σ 相	χ 相	σ 相
0	7123	75.63	83.15	83.96	91.57	85.24	90.35
1	7144	75.76	83.64	81.81	91.97	84.48	90.55
2	7146	76.96	83.98	83.68	92.71	85.46	90.89
3	7147	76.56	82.1	83.11	90.04	85.14	89.65

3.2.4 对比实验

为了体现本章提出的语义分割模型在性能上的优越性，本章选取一些新颖且高效的语义分割模型进行对比实验。表3.5和表3.6展示了DeepLabV3+^[100]、NonLocalNet^[84]、UperNet^[94]、CCNet^[96]、PointRend^[101]、KNet^[102]及Segformer^[103]和本章的方法在 χ 相和 σ 相上的分割性能。可以看出，在5个不同的评价指标上，本章的方法都优于这些对比模型。其中，CCNet使用了十字交叉注意力机制，可以看出它的性能和NonLocalNet相当，但是其计算量相较于后者更小。值得注意的是，其中的KNet和Segformer是基于Transformer的语义分割模型。尽管这两个模型搭载强大的骨干网络和预训练权重，依旧没能在本章的任务上取得很好的性能，体现了材料图像分割的特殊性。

表 3.5 不同模型在 χ 相的分割评价

网络结构	IoU	mIoU	mACC	Dice	mDice
DeepLabV3+ ^[100]	54.25	75.71	82.44	70.34	84.45
NonLocalNet ^[84]	49.46	73.14	79.91	66.18	82.29
UperNet ^[94]	54.47	75.85	82.03	70.53	84.56
CCNet ^[96]	51.07	74.01	80.8	67.61	83.03
PointRend ^[101]	53.7	75.46	81.16	69.88	84.24
KNet ^[102]	51.79	74.28	83.16	68.24	83.3
SegFormer ^[103]	54.02	75.63	81.49	70.15	84.37
Ours	56.23	76.77	83.19	71.98	85.31

在图3.10中展示了不同模型在一张包含 χ 相和一张包含 σ 相的图像上的分割结果，蓝色代表 χ 相，绿色代表 σ 相，使用红色框线标出了最值得关注的区域。可以看出在的第一张图中，其余模型的过分割问题严重，误将背景视作前景，而本章的方

表 3.6 不同模型在 σ 相的分割评价

网络结构	IoU	mIoU	mACC	Dice	mDice
DeepLabV3+ ^[100]	69.55	82.04	90.51	82.04	89.61
NonLocalNet ^[84]	65.86	79.83	88.75	79.42	88.75
UperNet ^[94]	69.54	82.04	90.48	82.04	89.61
CCNet ^[96]	67.2	80.61	89.57	80.38	88.65
PointRend ^[101]	69	81.7	90.33	81.65	89.39
KNet ^[102]	64.83	79.32	87.34	78.66	87.73
SegFormer ^[103]	67.06	80.63	88.59	80.29	88.65
Ours	71.65	83.28	91.6	83.48	90.44

法最接近真实标注。在第二张图中，大部分模型都没能还原出红色框线内所有的前景像素点，而本章的方法的预测结果和真实标注非常相似。

结合5个定量评价指标的结果和可视化结果，可以看出本章提出的方法超越了当前广泛使用的语义分割模型，在2205双相不锈钢微观组织形貌图像分割上取得了精确的结果。通过本章提出的语义分割模型，可以为后续的性能挖掘奠定良好的基础。

3.2.5 普适性分析

材料图像丰富多样，本章提出的语义分割方法在2205双相不锈钢图像上的有效性已经得到检验。为了验证本章提出的语义分割模型的普适性，额外选取了钛合金 $TiAl$ 和陶瓷共晶 $HfB_2 - B_4C$ 两个材料数据集进行实验。钛合金和陶瓷共晶数据集分别包含3张图像，对这两个数据集，本小节选取其中2张作为训练集，剩下的1张为验证集，经过三折交叉验证得到实验结果。对比模型是上一小节中的7个语义分割模型以及本章的方法，分割结果如图3.11。在普适性分析中，本章依旧采用交并比、Dice系数等评价指标进行定量分析，结果如表3.7和表3.8。钛合金数据集中的两个相均是前景相，因此在表3.7中只列出了平均指标：mIoU、mAcc和mDice。

通过观察不同数据集上的分割结果，可以发现本章提出的方法可以在其他材料图像上取得优越的效果，证实了本章方法的有效性和普适性。

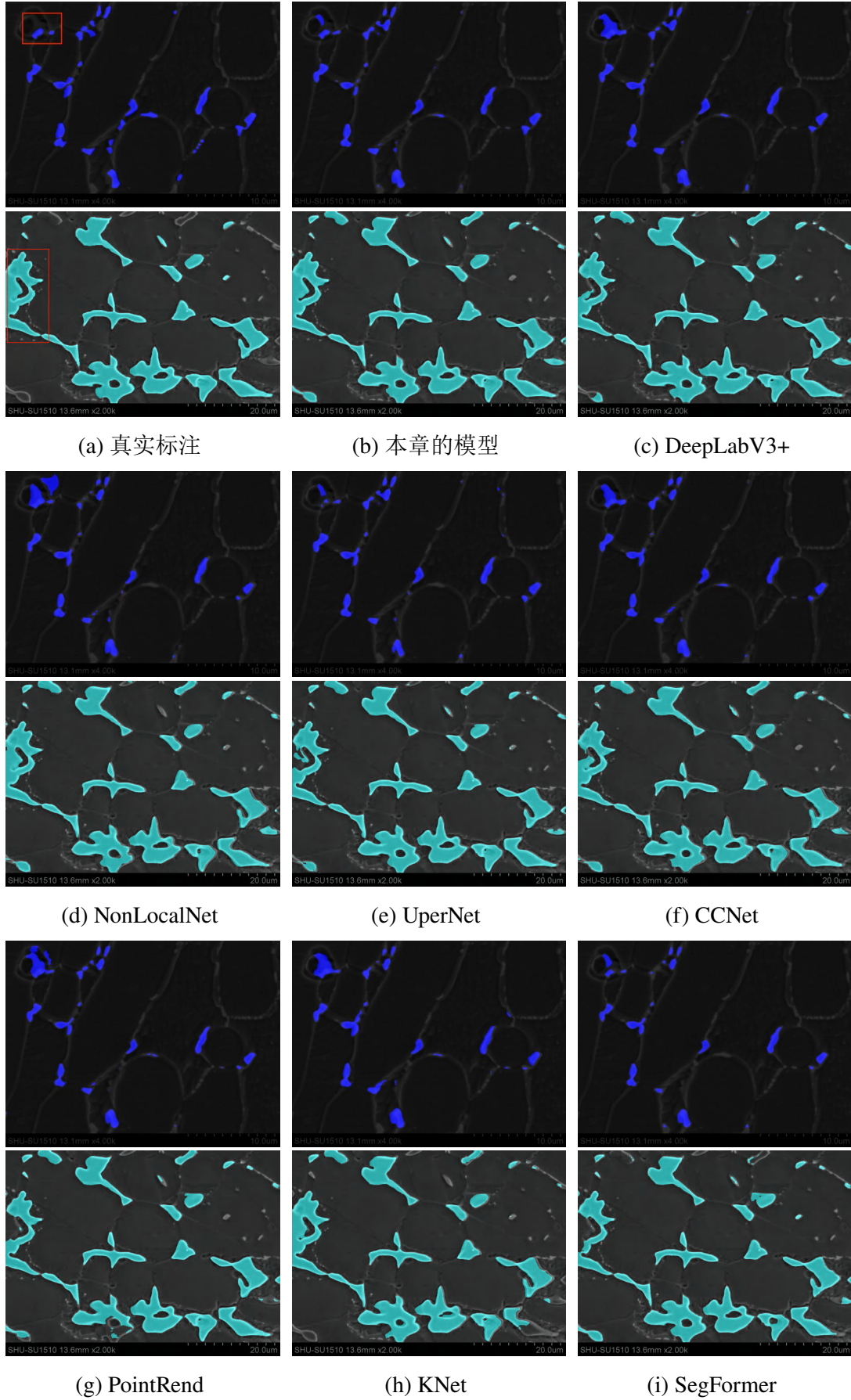


图 3.10 不同模型在2205双相不锈钢数据集上得到的分割结果展示

表 3.7 不同模型在钛合金的分割评价

网络结构	mIoU	mACC	mDice
DeepLabV3+ ^[100]	87.61	93.47	93.29
NonLocalNet ^[84]	88.43	93.78	93.81
UperNet ^[94]	88.4	93.85	93.81
CCNet ^[96]	88.28	93.66	93.68
PointRend ^[101]	89.82	94.58	94.6
KNet ^[102]	88.94	94.13	94.1
SegFormer ^[103]	85.15	91.93	91.93
Ours	90.07	94.8	94.73

表 3.8 不同模型在陶瓷共晶的分割评价

网络结构	IoU	mIoU	mACC	Dice	mDice
DeepLabV3+ ^[100]	82.92	87.93	92.61	90.52	93.43
NonLocalNet ^[84]	78.65	84.4	90.87	88	91.37
UperNet ^[94]	86.73	90.56	94.58	92.86	91.64
CCNet ^[96]	78.22	83.95	90.35	87.77	91.1
PointRend ^[101]	88.1	91.53	94.64	93.66	95.53
KNet ^[102]	83.5	88.20	92.98	90.97	93.63
SegFormer ^[103]	87.75	91.17	95.05	93.46	95.13
Ours	88.18	91.56	95.6	93.69	95.54

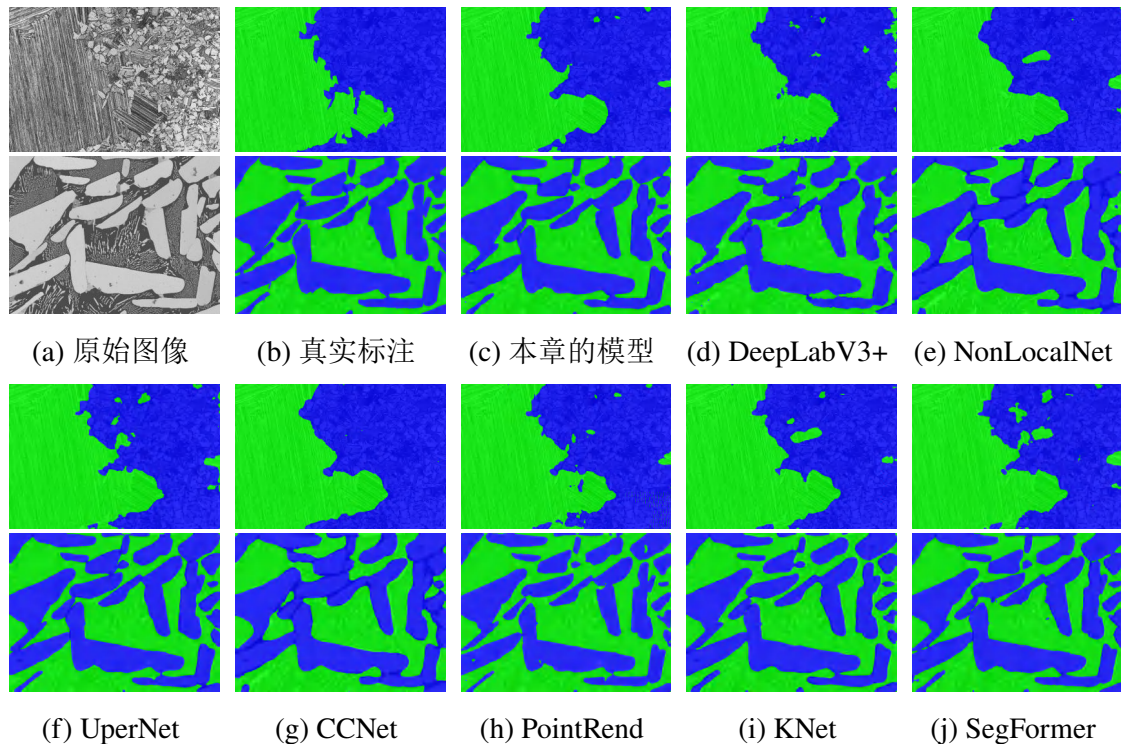


图 3.11 不同模型在钛合金以及陶瓷图像数据集上得到的分割结果展示

3.2.6 基于图像语义分割的材料性能挖掘

奥斯特瓦尔德熟化^[25]（Ostwald Ripening）机制从热力学角度描述了非均匀结构随时间流逝所发生的变化，颗粒长大受微合金元素的扩散控制，可以用来分析双相不锈钢中二次相的生长行为。奥斯特瓦尔德熟化机制公式具体为：

$$\bar{r}_t \approx \left(\frac{8\sigma V_p^2 D c_o}{9V_B c_p R T} \right)^{1/3} \cdot t^{1/3} \quad (3.8)$$

也可以写成：

$$k = \frac{\bar{r}_t^3}{t} \approx \frac{8\sigma V_p^3 D x_o}{9V_B V_M x_p R T} \quad (3.9)$$

其中 $\bar{r}_t$ 是二次相的平均半径， σ 是二次相与基质之间的相界面能， V_p 是二次相的摩尔体积， D 是基质中扩散控制元素的扩散系数， c_o 是基质中扩散控制元素的平衡摩尔浓度， V_B 是扩散控制元素的摩尔体积， c_p 是扩散控制元素在二次相中的平衡摩尔浓度， R 是气体常数， T 是华氏温度， x_o 是扩散控制元素在基质中的平衡摩尔分数， x_p 是扩散控制元素在二次相中的平衡摩尔分数， V_M 是基质的摩尔体积， $k^{1/3}$ 表示二次相的熟化速率。奥斯特瓦尔德熟化机制描述的是一般性现象，并不能完全契合双相不锈钢的熟化行为，需要通过实验进行对比分析。

对于2205双相不锈钢中二次相对性能影响的探究，应该测量足够数量的二次相进行统计分析，根据前文的介绍，本论文在30min、1h和2h的时效时刻统计 χ 相的尺寸分布，而4h、10h、20h、50h、100h、150h和200h的时效时刻统计 σ 相的尺寸分布。二次相尺寸信息主要是其占有的面积，基于前文介绍的图像分割算法可以区分出材料图像中二次相占据的像素区域。由于语义分割方法得到的前景没有区分不同的个体，无法直接得到各个二次相的信息。所以，需要借助数字图像处理技术——边界跟踪算法^[90]区分各个二次相。边界跟踪算法基于二值图像的拓扑结构，得到各个连通域的信息。在计算机中图像以矩阵表示，像素间的连通主要分为4连通和8连通。假设有两个取值相同的像素 p 和 q ，如果 q 在 p 的4邻域（距离为1的4个像素）中，则称这两个像素是4连通的，如果 q 在 p 的8邻域（包围 p 的8个像素，距离为1或 $\sqrt{2}$ ）中，则称这两个像素是8连通的。在4（8）连通场景中，如果一个取值为1的像素在它的8（4）连通域中有取值为0的像素存在，那么这个1像素就是一个边界点。在查找轮廓

时，通过光栅扫描（从左至右，从上至下）访问每一个像素，并且根据该像素的邻域判断它是否为边界点，同时根据边界跟踪过程中维护的编号（当前跟踪的边界的编号和最近遇到的边界的编号）来给这个像素点赋值，这个值包含了它是否为边界点、是否为右边界点、属于第几个轮廓等信息。如此往复，直至扫描到右下角的像素点。通过边界跟踪算法，可以得到二次相的面积等信息。

从公式(3.9)可知，需要的尺寸信息是 $\bar{r}_t$ ，边界跟踪算法得到的二次相面积不能直接使用。根据Schwind等人的研究^[104]，可以将二次相等效为球体，相应地在二维图像上则是圆形。所以，使用等效圆半径的理论可以得到二次相的等效圆半径：

$$r = \sqrt{\frac{A}{\pi}} \quad (3.10)$$

其中，A表示二次相面积， π 表示圆周率。 $\bar{r}_t$ 描述的是 t 时刻下所有二次相的平均半径，为了使结果更具统计特性，本论文使用直方图进行统计代替直接加和的方式求取平均值：

$$h_t^i = \frac{|\{r \mid b_{i-1} \leq r < b_i\}|}{|\{r \mid s_t \leq r < e_t\}|} \quad (3.11)$$

式中， t 表示对应的时刻， i 是 $[1, N_t]$ 上的组数， N_t 是直方图划分的组数， h_t^i 是当前第 i 组的比例， $||$ 表示集合中的元素数。 $b_i = \frac{i(e_t - s_t)}{N_t} + s_t$ ，其中 s_t 和 e_t 分别表示直方图统计的最小值和最大值。直方图的横轴表示预定义半径统计范围，纵轴表示二次相尺寸属于每个区间内的比例。 χ 相的半径统计直方图如图3.12， σ 相的半径统计直方图如图3.13。

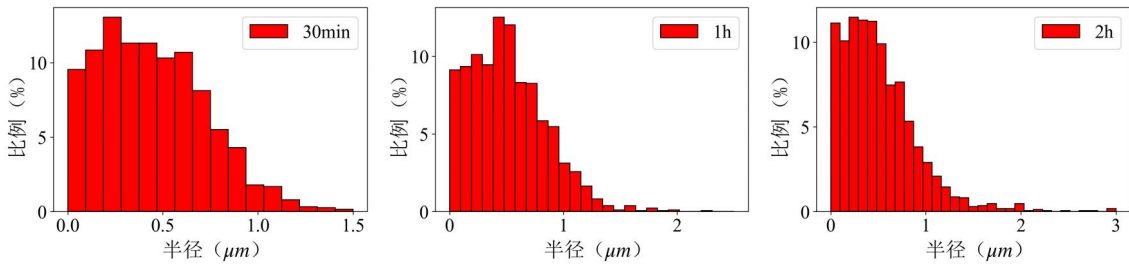


图 3.12 不同时刻的 χ 相尺寸分布

基于直方图得到二次相在 t 时刻的平均半径：

$$\bar{r}_t = \sum_{i=1}^{N_t} h_t^i \times (i - 1/2) \times ((e_t - s_t) / N_t + s_t) \quad (3.12)$$

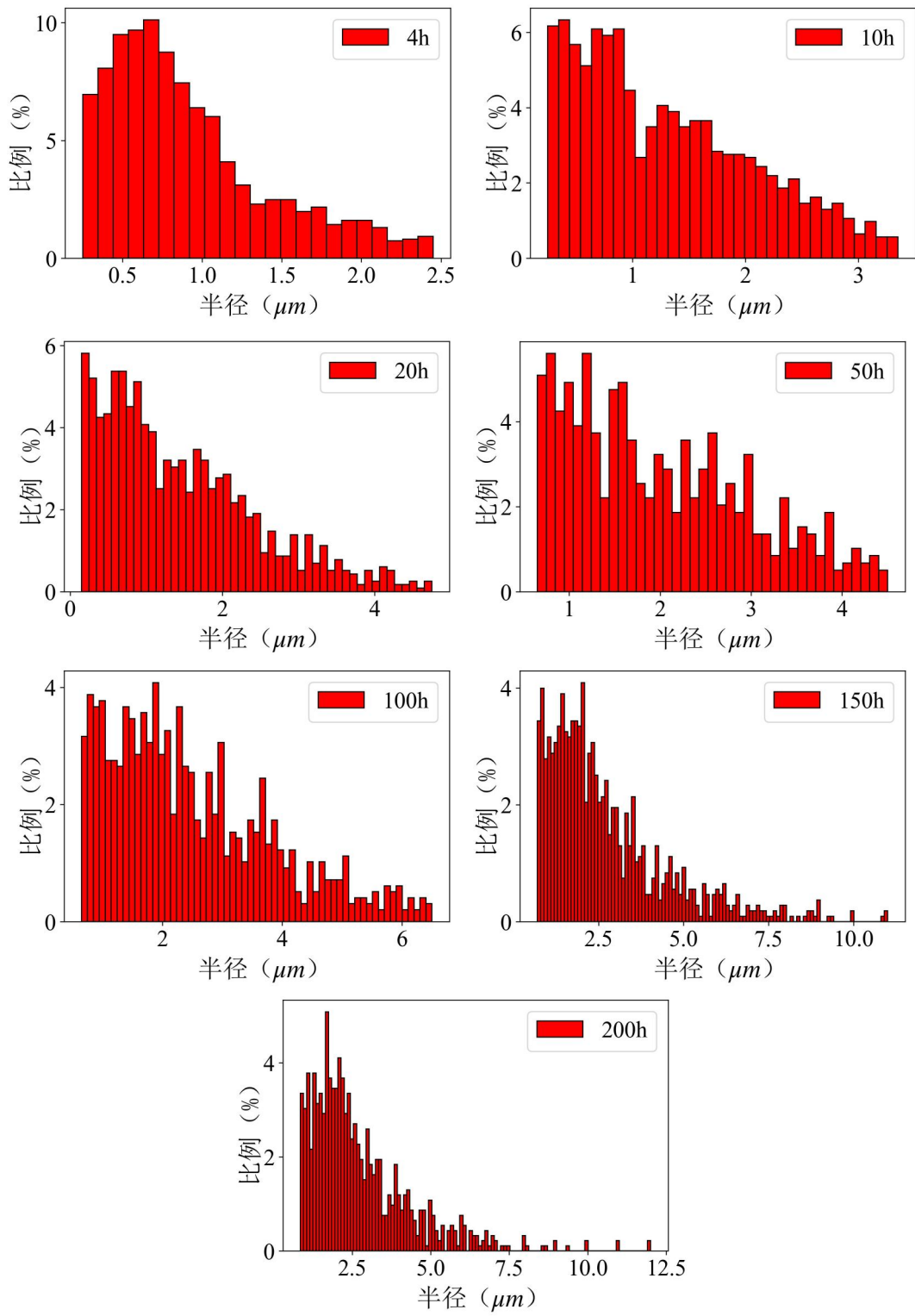


图 3.13 不同时刻的 σ 相尺寸分布

如图3.14, 本论文使用不同的标注来源来计算平均尺寸, 来源主要由两个变量控制, 一个是参与统计的图像来源, 分别是带标注图像和无标注图像, 另一个变量是图像标注的获取方式, 分别为人工标注和模型预测。通过组合不同的图像来源和图像标注的获取方式, 可以得到5条不同的尺寸变化曲线。在图3.14(a)中, 首先对人工标注的42张 χ 相图和169张 σ 相图的标注图使用轮廓查找和直方图统计, 得到该图的曲线, 用此作为本章实验结果的比较依据。在图3.14(b)中, 使用训练得到的模型预测上述人工标注图像的二次相分布, 并且进行相应的统计得到该图。通过比较图3.14(a)和图3.14(b), 可以发现它们的曲线走势有很高的相似度, 这意味着本章的模型准确地学习了标注数据的信息。此外, 为了增加参与统计的图像数量, 使用得到的模型对未标注图像进行预测, 使得这部分图像可以被利用。在图3.14(c)、图3.14(d)和图3.14(e)中, 分别是仅使用模型对无标注图像的预测、使用人工标注和模型对无标注图像的预测以及使用模型对所有图像的预测得到的尺寸变化曲线。可以发现, 新数据的加入使得曲线发生了微小的变化, 但是整体的趋势依然保持不变。由图3.14(a)-图3.14(e)可以看出在4h到10h之间发生了跳跃, 表明在20h后材料进一步熟化。而150h和200h后的 σ 相大小基本保持不变, 说明200h的观察周期内包含了整个熟化过程。为了更加直观地看出各个时刻不同标注来源得到的平均尺寸变化曲线的差距, 本章绘制了图3.14(f), 可以由该图看出曲线间的差距是较小的, 这意味着本章的方法是有效的。这些结果在二次相尺寸的变化趋势有高度的相似性: 二次相的平均直径在前几个时效时间迅速上升, 即二次相迅速膨胀, 然后逐渐趋于平稳, 其结果与预期的结果相一致。

得到各个时刻的二次相平均半径后, 可以分析二次相的熟化速率, 熟化速率反映了双相不锈钢性能的变化。由公式(3.9), 可以推出熟化速率的计算方式:

$$S_c = k^{1/3} = \frac{\overline{r_t}}{t^{(1/3)}} \quad (3.13)$$

在本章使用的2205双相不锈钢200h时效时间内, σ 相占主导地位。根据公式(3.13), 对 σ 相尺寸变化曲线经过坐标变换, 得到850°C下二次相的平均半径 $\overline{r_t}$ 和变换后的时效时间 $t^{(1/3)}$ 之间的关系图。如图3.15所示, 本章基于人工标注图像和模型对所有图像的预测的变化曲线分别进行了上述坐标变换。对于 $\overline{r_t}$ 和 $t^{(1/3)}$, 本章使用最小二乘法进行直线拟合, 得到基于人工标注图像的二次相熟化速率为 $29.366154nm/s^{(1/3)}$,

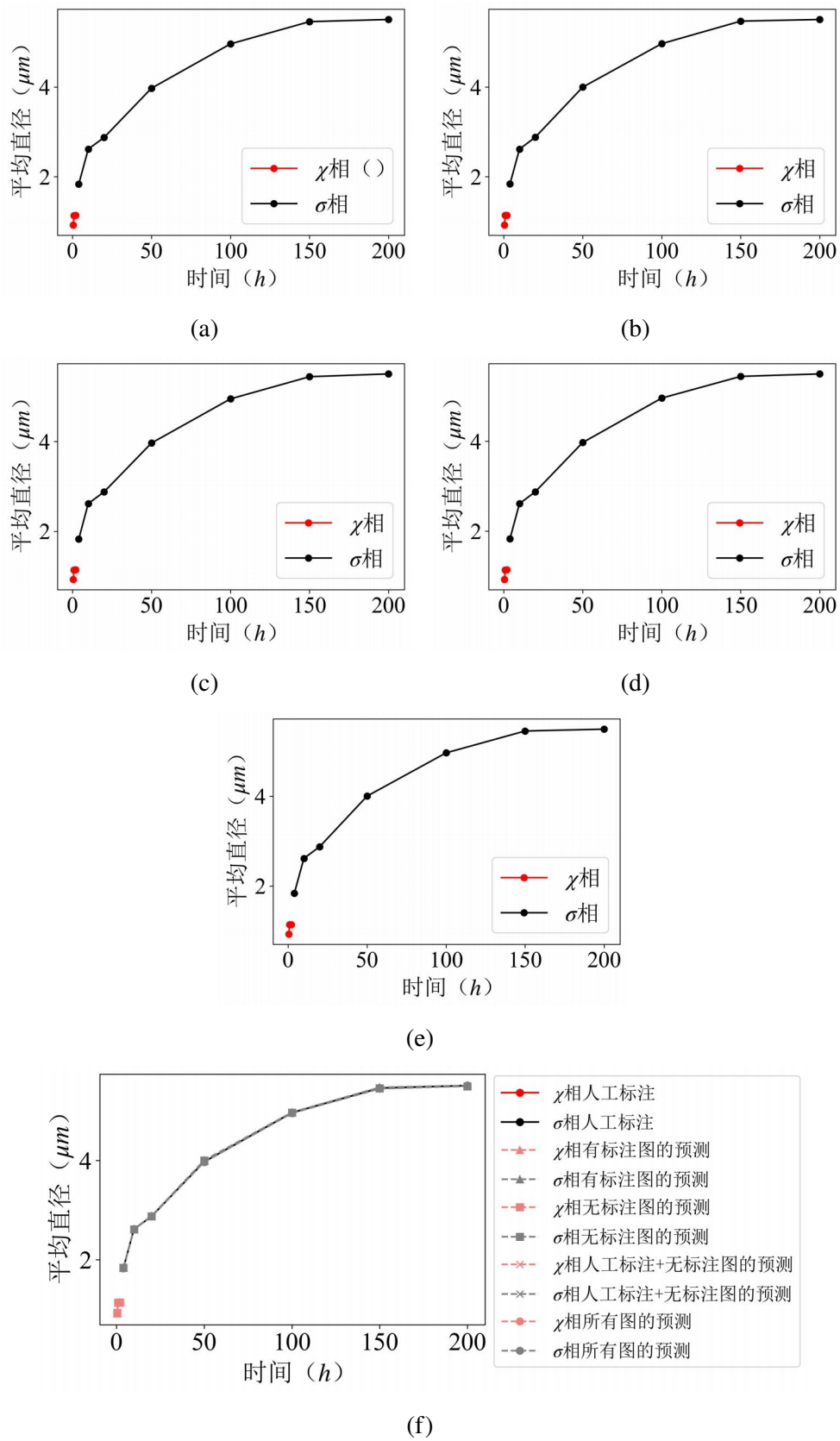


图 3.14 使用不同标注来源得到的二次相平均直径变化图：(a)仅使用人工标注；(b)仅使用模型对有标注图像的预测；(c)仅使用模型对无标注图像的预测；(d)使用人工标注和模型对无标注图像的预测；(e)使用模型对所有图像的预测；(f)整合所有直径变化曲线。

基于模型预测的所有图像得到的二次相熟化速率为 $29.266475\text{nm}/\text{s}^{(1/3)}$ 。可以看出，两种方法得到的结果非常相近，这意味着本章的方法可以有效地分割本论文面向的2205双相不锈钢微观结构形貌图像，可以帮助进行准确的性能挖掘。由熟化速率以及公式(3.9)中熟化速率和相界面能 σ 之间的正比例关系，可以推断相界面能相对较高^[26]。本论文使用奥斯特瓦尔德熟化机制估算相界面能，该公式是经验公式，将二次相处理为球形来近似尺寸，同时假设二次相的大小遵循 $1/3$ 次幂的规则。除了统计误差外，上述因素可能对我们的实验结果产生影响。由于相界面能难以测量，因此无法直接用来评估所获结果的准确性。本章的方法得到的结果和人工方法得到的结果具有高度相似性，这种人工方法与自动化方法相结合的方案不仅提高了研究的效率，而且使研究结果相互得到印证。

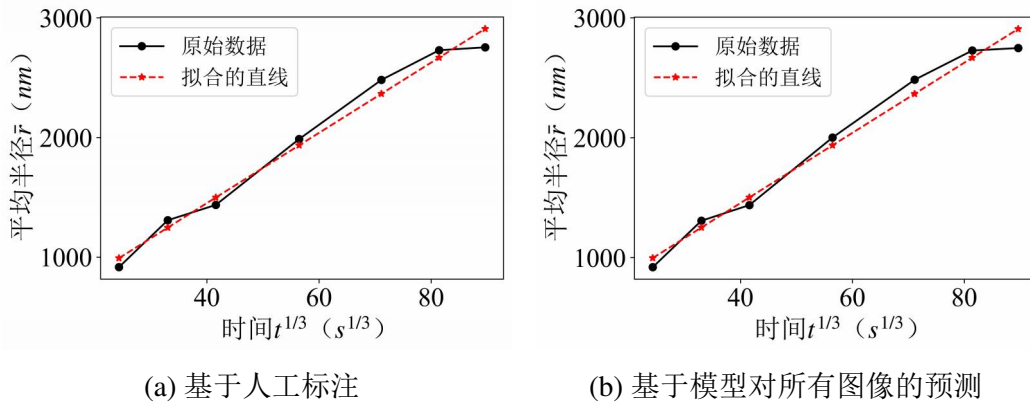


图 3.15 σ 相熟化速率拟合

3.3 本章小结

针对材料图像存在的特点：小样本、相较于自然图像更关注纹理信息，本章设计了基于特征金字塔的双分支语义分割网络来提取图像中的微观结构。为了使模型更加关注纹理和边界信息，本章进一步采用十字交叉注意力机制引导模型学习细节信息。本章提出的方法在2205双相不锈钢上取得了良好的分割效果，其普适性在多种小样本材料图像数据集上得到了验证。鉴于语义分割的输出不区分单个实例，对于模型预测的掩码信息，本章采用数字图像处理技术和统计方法进一步获得二次相的尺寸信息。最后，借助材料经验公式——奥斯特瓦尔德熟化机制分析材料性能的变化，完成分割结果在材料性能上的解释，研究的完整性和实用性得到了提升。

第四章 基于信息融合的图像实例分割和材料性能挖掘

在第三章中，介绍了一种基于语义分割模型的材料性能挖掘方法，针对材料图像样本匮乏、更加关注纹理信息的特点设计了语义分割网络。基于语义分割结果，本论文得到了和人工标注非常相近的二次相尺寸变化曲线，拟合了二次相的熟化速率，由此分析材料性能的变化。但是，语义分割网络输出的掩码包括了同类别的所有个体，无法直接得到各个二次相的信息，因此，在第三章中借助了边界跟踪算法。这样的材料性能挖掘流程是较为复杂的，多个步骤的叠加会带来的误差累加。本章提出一种基于多模态融合和伪标签技术的材料图像实例分割方法，通过语义模块和两阶段训练来克服小样本问题。基于实例分割的结果，可以直接得到各个二次相的尺寸信息，完成材料性能挖掘。

4.1 方法概述

相较于语义分割，实例分割任务更加复杂，需要学习更精细的信息，样本匮乏与高性能的大模型之间的矛盾更为突出。如图4.1，本章在基模型上增加语义模块和融合伪标签的两阶段训练，通过类别的语义信息编码进一步提高实例分割中分类的准确性，并且使用伪标签对无标注数据加以利用。本节介绍了本章使用的实例分割基模型、本章提出的语义模块和融合伪标签的两阶段训练以及使用的损失函数。

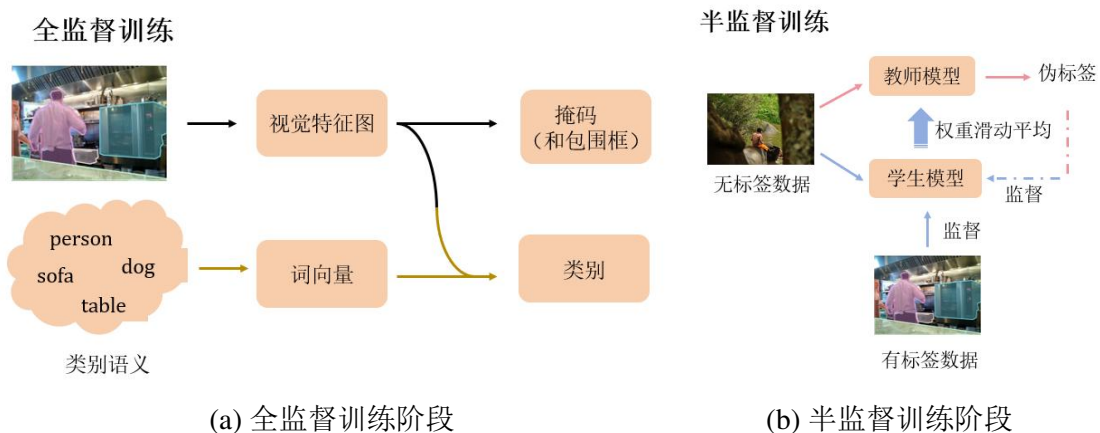


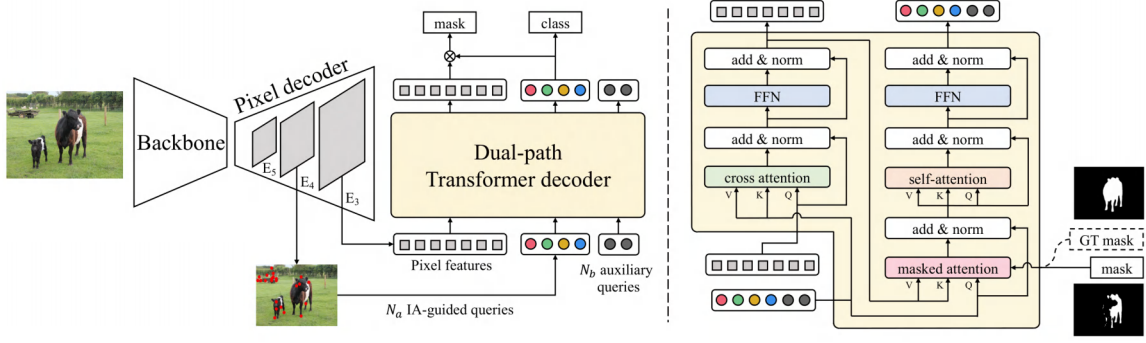
图 4.1 带有语义模块的实例分割模型和融合伪标签的两阶段训练

4.1.1 基模型

目前，实例分割模型根据是否基于候选框可以分为两类。本章提出的语义模块和融合伪标签的两阶段训练不涉及候选框的提取，适用于任何一种实例分割模型，因此，本章选取基于候选框（Box-dependent）的实例分割网络MRCNN^[58]和不基于候选框（Box-free）的实例分割网络FastInst^[105]分别作为基础模型，验证方法的普适性和有效性。

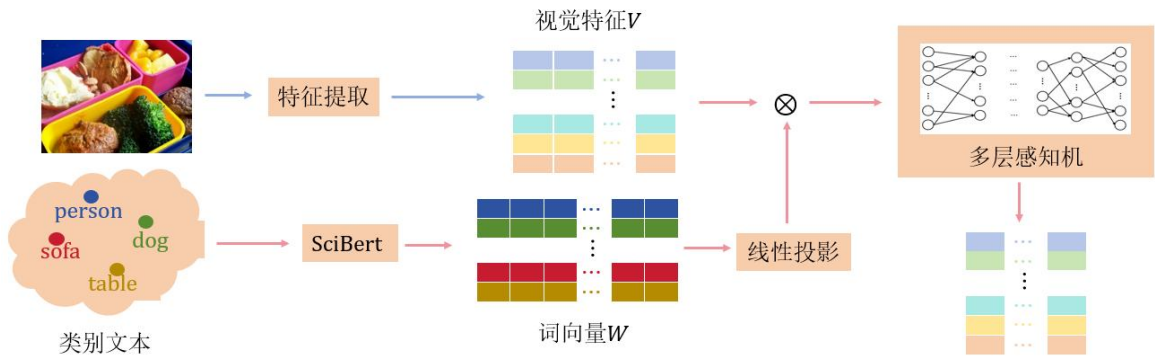
MRCNN的模型框架如图2.4，可以看出在MRCNN中，图像首先经过区域候选网络（Region Proposal Network, RPN），区域候选网络使用预定义的锚（Anchor）通过滑动窗口的方式在图像上生成一系列的候选区域（即Region Proposal），然后利用这些候选区域作为感兴趣区域（Region of Interest, ROI）送入分类器或检测器中进行后续的工作。为了避免卷积和池化操作对图像带来的偏移，MRCNN使用感兴趣区域对齐（ROI Align）的策略对图像的候选框进行调整，以获得对应区域的特征图，然后使用特征图进行分类和掩码生成。

FastInst^[105]是一个可以快速实现实例分割的模型，基于Mask2Former^[61]设计了一个简洁高效的实例分割框架，同时保持了相当的性能。这类实例分割模型遵循DETR^[62]的思想，利用功能强大的Transformer，结合一组可学习的查询，来预测实例的类别和分割掩码。对比图2.6中展示的DETR模型，FastInst的查询（Queries）由两部分组成，一部分是从具有高前景概率的特征图 E_4 中选择 N_a 个像素嵌入作为实例激活引导查询（Instance Activation Guided Queries），另一部分是辅助查询（Auxiliary Queries），二者连接作为总查询，如图4.2所示。此前，基于查询的方法使用的查询与图像无关，因此需要许多Transformer解码层来细化，实例激活引导查询在初始化时保存了关于潜在对象的丰富信息，提高了Transformer解码器中查询迭代的效率。FastInst还设计了双路径Transformer解码结构提高运算效率，如图4.2右半部分。与传统的单路径更新策略相比，像素特征和查询可以同时得到更新，从而减少对繁重像素解码器的依赖，获得更细粒度的特征。

图 4.2 不基于候选框的实例分割模型FastInst^[105]

4.1.2 语义融合模块

基于候选框的实例分割模型，例如MRCNN，需要先得到实例的候选框，基于候选框区域完成实例的分类和实例掩码的分割。不基于候选框的实例分割模型，例如SOLO，仅包含实例分类和掩码分割。可以看出，这两类方法都需要对实例进行分类。本节提出的语义模块主要是通过改变实例分割模型的分类分支，从而提高分类的精度。在对实例进行分类时，传统的单模态实例分割网络只利用了编码后的图像特征，本章加入类别的语义信息，并将语义信息与图像特征进行多模态融合。从海量文本数据训练得到的语义信息，潜在地包含了不同类别之间的关系以及各个类别的特征，通过融入这些信息，有助于提高模型的分类精度。语义信息模块和原始模型的融合如图4.3所示。

图 4.3 语义模块与图像特征融合的框架示意。蓝线表示图像视觉特征 V 的获取流，红线表示语义信息参与的数据流。

类别的语义信息通常用自然语言表示，不能直接参与模型的计算。本章使用在大规模文本上预训练后的SciBert^[106]来提取类别的嵌入表示，用 $W \in \mathbb{R}^{N \times d_w}$ 表示类别嵌入组成的语义空间。其中， N 是数据集中包含背景类别数； d_w 是由SciBert生

成的每个单词嵌入的维度。由于无法保证图像的视觉特征和单词嵌入具有相同的维数，本章使用一个投影层来改变 W 的维数。通过该操作，训练模型将视觉信息投影到语义空间上，由此得到新的嵌入 $W_{se-vi} \in \mathbb{R}^{N \times d}$ 。维度 d 与图像特征 $V \in \mathbb{R}^{q \times d}$ 的维度一致。在单阶段方法中， q 表示查询的数量，在两阶段方法中 q 表示最终预测框的数量。

接着，本章融合从模型中得到的图像特征 V 和新的词嵌入 W_{se-vi} 。最后，经过多层感知机后得到分类结果。这些操作使图像特征投影在语义空间上。上述过程可以表示为：

$$\begin{aligned} W_{se-vi} &= WP \\ Pred_{cat} &= Softmax(MLP(V \times W_{se-vi}^T)) \end{aligned} \quad (4.1)$$

式中， $Pred_{cat} \in \mathbb{R}^{q \times N}$ 为预测的各个类别概率， $P \in \mathbb{R}^{d_w \times d}$ 为投影层的权重。利用额外的类别文本信息，模型不仅可以学习图像信息，还可以从稀缺的数据中学习语义信息。与SRR-FSD^[70]相比，本章的语义模块从SciBert而不是ImageNet API^①中获得嵌入，并且只需要简单的卷积层而不是计算复杂的知识图就可以对嵌入进行投影。相较于CLIP等图文融合方法，本章的语义模块不需要额外的图片描述提示符，仅在已有的文本信息上进行多模态融合，更加简便和高效。

4.1.3 融合伪标签的两阶段训练

由于在训练开始时获得的预测结果不准确，本章将整个训练分为两个阶段，设计了全监督和半监督相结合的两阶段训练方式，提高对未标注数据的利用率。在全监督训练阶段仅使用标注数据训练图4.1(a)的模型。在半监督训练阶段同时采用两种技术来充分利用未标注数据的信息，提高模型性能。首先，基于一致性正则化的思想对输入图像加噪，期望模型保持输出的稳定性。其次，利用伪标签技术实现未标注数据对模型的监督学习。半监督训练的详细描述如算法1。

如图4.1(b)所示，本章的半监督训练阶段包含两个子模型，分别为教师模型和学生模型，它们具有与图4.1(a)相同的结构，但负责不同的任务。教师模型负责为未标注数据生成伪标签，而学生模型使用有标注数据和带有伪标签的无标注数据进行有

① <http://image-net.org/download-API>

算法 1: 一轮半监督训练流程

输入: 带标签数据 D_l , 无标签数据 D_{un} , 阈值 t , 比例 β , 权重 α

输出: 经过一轮迭代训练后获得的教师模型 M_{tea} 和学生模型 M_{stu}

// 数据增强

- 1 $D_{l_w}, D_{un_w} \leftarrow WeakAug(D_l, D_{un})$
- 2 $D_{l_s}, D_{un_s} \leftarrow StrongAug(D_l, D_{un})$
- // 为无标签数据生成伪标签
- 3 $pred_{D_{un}} \leftarrow M_{tea}(D_{un_w})$
- 4 $pseudo_label \leftarrow filter(pred_{D_{un}}, t)$
- 5 $all_D_{un} \leftarrow add_label(D_{l_w}, pseudo_label) + add_label(D_{l_s}, pseudo_label)$
- // 使用有标签数据和带有伪标签的无标签数据监督训练学生模型
- 6 $all_D_l \leftarrow D_{l_w} + D_{l_s}$
- 7 $loss = M_{stu}(all_D_l) + \alpha M_{stu}(all_D_{un})$
- 8 $loss.backward()$
- // 使用权重滑动平均的方式更新教师模型
- 9 $\theta_{tea} \leftarrow \beta \theta_{tea} + (1 - \beta) \theta_{stu}$

监督学习，它们的初始权重都是全监督训练阶段得到的最终权重的拷贝。在半监督训练阶段，首先对输入图像进行弱增强和强增强。弱增强只会对图像进行随机翻转和尺寸变化，而强增强则会改变图像色彩、增加噪声和遮挡部分区域。弱增强的未标注图像输入到教师模型中，生成伪标签；经过强、弱增强后的有标注数据和伪标签标注的四种数据输入到学生模型中。通常情况下，即便是教师模型也不能确保所有的输出都是可靠的。因此，本章设置了一个阈值来过滤教师模型的输出，只有具有高置信度的输出会作为监督学生模型训练的未标注数据的伪标签。相较于学生模型，教师模型接收的输入图像更简单，可以输出更可靠的预测结果，从而指导学生模型训练。

学生模型总的损失来源于标注数据和未标注数据，使用梯度下降法对学生模型的参数 θ_{stu} 进行更新。教师网络 θ_{tea} 的参数使用权重滑动平均（Exponential Moving

Average, EMA) 更新:

$$\theta_{tea} = \theta_{tea} * \beta + \theta_{stu} * (1 - \beta) \quad (4.2)$$

其中, β 为保留的旧参数的比例。该更新方法不仅一定程度上保留了第一阶段的训练参数, 还利用了未标注图像的信息, 可以更加稳定地更新教师模型。因此, 教师模型可以对未标注图像持续提供更加可靠的伪标签, 用于学生模型的进一步训练。周而复始, 两个模型逐渐获得更准确的分割结果。

Noisy Boundary^[46] 是半监督学习的一个知名模型, 它的伪标签是一次性离线生成并保存的。在 Noisy Boundary 中, 首先使用有标签数据训练模型, 然后使用该模型对未标注数据进行预测。接着, 模型视这些预测为无标签数据的“真实标签”, 同有标签数据一起从头进行有监督训练。与 Noisy Boundary 不同, 本章在半监督训练的过程中同时更新教师网络和学生模型, 实时生成伪标签。

4.1.4 损失函数

本章在基于候选框的实例分割网络 MRCNN 和不基于候选框的实例分割网络 FastInst 分别进行了新模型的搭建和实验验证, 两种基础模型的输出不同, 因此在训练阶段使用的损失函数也不同。

基于候选框的实例分割方法 MRCNN 输出预测的包围框、掩码以及类别, 它的损失函数由三部分组成:

$$L_{box-dependent} = L_{box} + L_{mask} + L_{cls} \quad (4.3)$$

其中, 掩码损失 L_{mask} 和类别损失 L_{cls} 都是分类损失, 使用交叉熵函数计算, 预测框的损失 L_{box} 是回归损失, 使用平滑 L_1 损失函数 (Smooth L_1 Loss) 来计算:

$$Smooth_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases} \quad (4.4)$$

在回归问题中通常选择均方误差即 L_2 范数作为损失函数, 当误差很大时, 容易产生梯度爆炸。相较于 L_2 范数, L_1 范数和误差呈线性关系, 具有更平缓的梯度。但是 L_1 范数不是处处可导的, 可能会影响模型的收敛。平滑 L_1 损失函数结合 L_1 范数

和 L_2 范数的优点，在0点附近使用平方损失函数，缓解了 L_1 范数不够平滑的问题，得到更平缓的梯度。

不基于候选框的实例分割模型FastInst的输出分为两部分：类别和掩码，它的损失函数包含了对这两部分输出的监督：

$$L_{box-free} = L_{cls} + L_{mask} \quad (4.5)$$

其中，使用交叉熵损失函数衡量预测类别的损失，使用交叉熵和Dice Loss协同衡量预测掩码的误差。基于候选框的实例分割方法使用候选框对应的感兴趣区域预测实例的掩码；不基于候选框的实例分割方法使用整个图像分别预测各个实例的掩码，单个实例占有的像素一般远少于背景占有的像素，即在分割单个实例掩码时存在类别不平衡问题。Dice Loss是一种常见的分割损失计算方式，由VNet^[107]首次提出，旨在应对语义分割中正负样本极其不平衡的场景，被广泛应用在医学图像分割上。Dice Loss是基于Dice系数的损失函数：

$$L_{dice} = 1 - Dice \quad (4.6)$$

通过分析Dice系数的两种形式，即公式(2.8)和公式(2.9)，可以发现Dice系数是基于区域计算的，它只关注正样本的区域，因此当正负样本不平衡时，可以减小背景的梯度。但这种基于区域的损失也带来新的问题——当正样本为小目标时会使损失值大幅度震荡，导致训练不稳定。因此，Dice Loss一般和别的损失函数一起使用，本章结合交叉熵和Dice Loss一起衡量预测掩码的误差。

在全监督训练阶段，根据基础模型的不同，使用 $L_{box-dependent}$ 或 $L_{box-free}$ 作为损失函数。在半监督训练阶段，有标签数据和伪标签标注的数据一起监督模型训练，损失函数为：

$$L_{semi} = L_{label} + \alpha L_{unlabel} \quad (4.7)$$

L_{label} 和 $L_{unlabel}$ 的函数形式保持一致， α 是无标注数据的损失权重。

4.2 实验分析

本节首先介绍实验设置和数据集，接着，展示在多种数据集上的实验结果，最后，基于实例分割结果，完成2205双相不锈钢的性能挖掘。

4.2.1 实验设置

本章分别基于MRCNN和FastInst实现语义模块和两阶段训练，骨干网络是ResNet50，初始权重是ImageNet上预训练的结果。模型使用AdamW优化器训练，基于MRCNN的模型初始学习率是0.01，基于FastInst的模型初始学习率为0.0001，批大小为4。全监督学习的迭代轮次是2000轮，在半监督训练阶段，学生模型和教师模型一起训练了100个轮次，教师模型参数更新使用的比例 β 为0.999。

4.2.2 数据集介绍

由于带有实例分割标注的材料数据集十分匮乏，为了验证方法的普适性，本章增加了在两个公共数据集上的实验：COCO2017^[108]和TrashCan^[109]。COCO2017包括从自然场景中收集的80类图像，这些类别用英文来表示，分别是：

- person, bike, car, motorcycle, airplane, bus, train, truck, boat, traffic light, fire hydrant, stop sign, parking meter, bench, bird, cat, dog, horse, sheep, cow, elephant, bear, zebra, giraffe, backpack, umbrella, handbag, tie, suitcase, frisbee, skis, snowboard, sports ball, kite, baseball bat, baseball glove, skateboard, surfboard, tennis racket, bottle, wine glass, cup, fork, knife, spoon, bowl, banana, apple, sandwich, orange, broccoli, carrot, hot dog, pizza, donut, cake, chair, couch, potted plant, bed, dinning table, toilet, tv, laptop, mouse, remote, keyboard, cell phone, microwave, oven, toaster, sink, refrigerator, book, clock, vase, scissors, teddy bear, hair drier, toothbrush

训练集包含118278张图像，验证集包含5000张图像，测试集包含40640张图像。该测试集的标注是不公开的，因此本章使用COCO2017的验证集衡量模型的性能。数据集TrashCan包含22个类别，这些类别用英文来表示，分别是：

- ro, plant, animal fish, animal starfish, animal shells, animal crab, animal eel, animal etc, trash clothing, trash pipe, trash bottle, trash bag, trash snack wrapper, trash can, trash cup, trash container, trash unknown instance, trash branch, trash wreckage, trash trap, trash rope, trash net

训练集中有6065张图像, 验证集中有1147张图像。本章使用科学文本的预训练语言模型SciBert获取这些类别文本的词嵌入。在自然语言处理中, 通常使用余弦相似度来衡量两个词嵌入之间的相似度, 图4.4是COCO2017数据集和TrashCan数据集的类别嵌入的余弦相似度的可视化, 从图中可以看出, 有些类别在语义上与其他类别有很大的差异, 即呈现整行或整列几乎都是蓝色的现象。对照图4.4, 可以发现COCO2017数据集中的类别“sheep”和“banana”与其他类别的语义相似度较低, 由于在COCO2017数据集中存在和这两个类别视觉上相近的类别, 它们词嵌入的低相似性意味着语义模块可以带来新的信息。

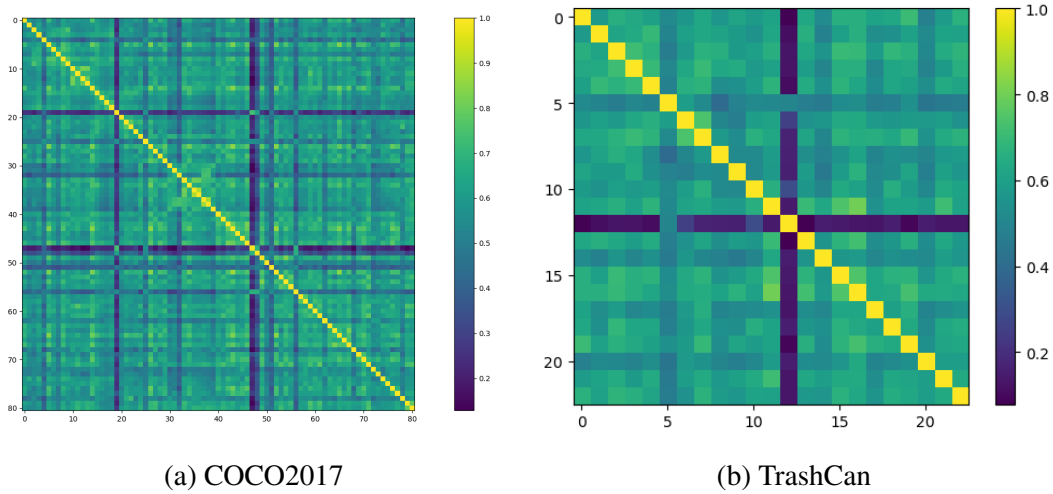


图 4.4 COCO2017数据集和TrashCan数据集类别嵌入的相似度矩阵

在实验中, 为了模拟小样本场景, 从COCO2017的训练集中随机抽取0.3%、0.5%和1%的数据, 从TrashCan的训练集中随机抽取1%、2%和3%的数据, 分别作为本章模型的六个训练集。各个训练集中没有被抽取的部分作为未标注的数据。此外, 2205双相不锈钢数据集也用来验证本章的方法, 其设置和第三章一致。本章通过比例随机抽取数据, 相较于为每个类别抽取K个样本的设置, 即K-shot, 一方面可以有很好的随机性, 另一方面可以更好地保持原始数据的分布。在图4.5中分别展示

了COCO2017数据集和TrashCan数据集的3张带标注图像，可以看出，这两个数据集和2205双相不锈钢数据集差异非常大：COCO2017包含的是自然光照下的图像，通过人眼就可以进行辨认；TrashCan是海底垃圾数据集，拍摄的光照条件很差，采集的图像较难辨认，相较于COCO2017图像识别难度更大；2205双相不锈钢数据集由显微图像组成，其中微观组织的辨认需要专业的知识，异类组织的视觉相似性要求深度学习模型敏锐地学习到类别差异。本章在这三种风格迥异的数据集上进行消融实验和对比试验，验证语义模块和两阶段训练的有效性。



(a) COCO2017数据集示例



(b) TrashCan数据集示例

图 4.5 本章额外使用的数据集示例

4.2.3 消融实验

本章基于两种不同的基础架构，即MRCNN和FastInst，搭建了实例分割的框架，分别对语义模块和两阶段训练、半监督训练中伪标签的过滤阈值以及每个批次中有标签数据和无标签数据的比例进行消融实验。

表4.1、表4.2以及表4.3展示了在TrashCan、COCO2017以及2205双相不锈钢数据集上对语义模块和两阶段训练的消融实验的定量结果。可以看出，添加的两个方法在三个评价指标上均有助于提高模型的性能。语义模块更倾向于提高 AP_{75} ，而两阶段训练对 AP_{50} 的提升更有帮助。这意味着语义模块更准确地提高了具有更高IoU的预测性能，而两阶段训练带来更均匀的提升。在两个分支的合作下，评价指标AP、

AP_{50} 和 AP_{75} 与基础模型相比有不同程度的提升。

表 4.1 不同模块组合在TrashCan的定量结果

语义模块	两阶段训练	1%			2%			3%		
		AP	AP ₅₀	AP ₇₅	AP	AP ₅₀	AP ₇₅	AP	AP ₅₀	AP ₇₅
基于FastInst										
✓		1.51	3.63	1.24	4.15	11.21	2.61	6.01	14.87	3.94
		1.83	3.73	1.75	4.40	10.95	2.82	6.29	14.59	4.56
	✓	1.81	4.42	1.26	4.32	11.29	2.80	6.03	14.82	4.13
✓	✓	1.87	3.75	1.94	4.48	11.15	2.87	6.44	14.73	4.82
基于MRCNN										
✓		1.13	2.98	0.85	4.01	10.17	2.16	6.28	13.70	4.63
		1.71	3.68	1.45	4.32	10.08	3.45	6.51	14.46	4.53
	✓	1.21	3.16	0.92	4.12	10.36	2.71	6.34	13.97	4.73
✓	✓	1.77	3.80	1.53	4.36	10.31	3.39	6.54	14.46	4.49

表 4.2 不同模块组合在COCO2017的定量结果

语义模块	两阶段训练	0.3%			0.5%			1%		
		AP	AP ₅₀	AP ₇₅	AP	AP ₅₀	AP ₇₅	AP	AP ₅₀	AP ₇₅
基于FastInst										
✓		4.41	8.96	3.96	6.48	12.65	6.00	9.25	17.85	8.58
		4.57	8.98	4.21	6.72	12.79	6.33	9.29	17.57	8.77
	✓	4.51	9.09	4.06	6.53	12.91	6.00	9.31	18.10	8.64
✓	✓	4.67	9.13	4.32	6.76	12.93	6.37	9.49	18.17	8.79
基于MRCNN										
✓		4.21	10.11	3.03	6.71	14.79	5.23	9.53	19.98	8.17
		4.75	11.02	3.45	6.99	14.96	5.82	10.14	20.49	8.91
	✓	4.51	10.93	2.99	6.86	15.1	5.45	9.58	20.24	8.03
✓	✓	5.03	11.45	3.75	7.15	15.68	5.75	10.26	20.85	8.86

图4.6和图4.7是在不同的模块组合下的实例分割结果，分别选取三个数据集中的一个例子展示，为了便于描述，将它们称为第一个、第二个和第三个例子。图4.6展示了基于FastInst的模块消融实验结果。在第一个例子红框标记的区域中，加入的语义模块将基础模型FastInst中错误的分类“trash snack wrapper”转变为“trash

表 4.3 不同模块组合在2205双相不锈钢的定量结果

语义模块	两阶段训练	χ 相			σ 相		
		AP	AP ₅₀	AP ₇₅	AP	AP ₅₀	AP ₇₅
基于FastInst							
✓		0.99	4.59	0.1	13.69	34.38	7.99
		1.64	6.38	0.44	16.14	37.95	11.34
	✓	1.13	5.4	0.1	13.92	34.98	7.71
✓	✓	1.7	6.8	0.4	16.16	38.29	11.44
基于MRCNN							
✓		36.2	68.79	33.39	59.13	82.9	68.22
		39.84	69.34	40.91	65.7	84.72	76.44
	✓	37.1	70	35.31	63.61	86.6	73.48
✓	✓	40.53	70.21	42.78	69.7	89.2	79.53

unknown instance”；在第二个例子中，语义模块初步挖掘出了基础模型没有检测出的实例“suitcase”，而在语义模块的基础上加入两阶段训练可以提供更精细的掩码，如图4.6(e)所示。图4.7中展示了基于MRCNN的模块消融实验结果。在第一个例子中，语义模块帮助发现了基础模型中未识别出的实例“animal fish”，而两阶段训练帮助得到更精细的实例“rov”的分割掩码。在第二个例子中，语义模块过滤掉了错误的预测，通过两阶段训练挖掘出了基础模型未检测到的“baseball bat”。从实验结果可以总结出二者的作用：语义模块可以提高正确分类的得分，减少错误分类得分，通过这种调整，它可以帮助过滤不正确的分割，缓解基模型FastInst中过度分割的问题；两阶段训练使分割掩码更精确，并揭示在基础模型中没有被检测到的实例。通过二者的结合，可以生成高置信度的精细掩码。

在半监督训练阶段，伪标签需要通过阈值筛选后才能对学生模型进行监督。本章在COCO2017抽取0.3%的训练集上完成过滤阈值的消融实验，结果如表4.4。可以看出，当阈值过低时，即为0.3时，不正确的伪标签会带来错误的指导，模型的性能有所降低。而当阈值过高时，即为0.9时，只有少量伪标签可以通过筛选从而参与到训练中，导致两阶段训练的效果不显著。从定量结果看出，对于基于FastInst的方法，伪标签阈值设置为0.7有最好的效果，而对于基于MRCNN的方法，阈值设置为0.5有最好的效果。



(a) 真实标注



(b) 基础模型FastInst



(c) 基础模型+语义模块



(d) 基础模型+两阶段训练



(e) 基础模型+语义模块+两阶段训练

图 4.6 基于FastInst的不同模块组合得到的分割结果展示



(a) 真实标注



(b) 基础模型MRCNN



(c) 基础模型+语义模块



(d) 基础模型+两阶段训练



(e) 基础模型+语义模块+两阶段训练

图 4.7 基于MRCNN的不同模块组合得到的分割结果展示

表 4.4 不同过滤阈值下的定量结果

阈值	AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L
基于FastInst						
0.3	4.56	9	4.2	1.26	3.59	7.07
0.5	4.62	9.08	4.17	1.33	3.63	7.20
0.7	4.67	9.13	4.32	1.36	3.67	7.06
0.9	4.54	8.98	4.19	0.94	3.72	6.99
基于MRCNN						
0.3	5.06	11.49	3.78	1.63	5.16	7.77
0.5	5.07	11.49	3.81	1.62	5.12	7.80
0.7	5.03	11.45	3.75	1.62	5.15	7.72
0.9	5.05	11.49	3.80	1.61	5.14	7.47

在半监督训练阶段，有标签数据和无标签数据的比例也会给模型的性能带来影响。本章依旧在COCO2017抽取的0.3%的训练集上进行消融实验，结果如表4.5所示。可以看出，不论是否基于候选框，每个训练批次中有标签数据和无标签数据的数量相同的设置下可以取得更好的性能。

表 4.5 不同有标签数据和无标签数据比例下的定量结果

比例	AP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L
基于FastInst						
1:1	4.67	9.13	4.32	1.36	3.67	7.06
1:2	4.57	9	4.2	0.93	3.69	7.13
1:3	4.56	9.02	4.11	0.95	3.62	7.08
1:4	4.62	9.08	4.17	1.33	3.63	7.20
基于MRCNN						
1:1	5.07	11.49	3.81	1.62	5.12	7.80
1:2	5.06	11.48	3.76	1.63	5.15	7.76
1:3	4.81	11.26	3.39	1.61	5.05	7.28
1:4	4.81	11.25	3.4	1.62	5.05	7.30

4.2.4 对比实验

本章提供了与其他方法的对比实验结果，包括MRCNN、FastInst、iFS-RCNN^[110]、Noisy Boundary^[46]和PAIS^[111]。这些模型可以分为三类：MRCNN和FastInst代表一般的实例分割模型，iFS-RCNN是小样本实例分割模型，Noisy Boundary和PAIS是半监督模型。iFS-RCNN在提出时依照元学习的训练方式，在两个训练阶段中使用两个不重叠的数据集，即 D_{Base} 和 D_{Novel} ，因为本章没有将数据集分成 D_{Base} 和 D_{Novel} ，所以在对比实验中对iFS-RCNN仅进行一阶段的全类别训练。

表4.6和图4.8是在TrashCan数据集上的实验结果，加粗表示的是最高值，下划线表示次高值。从表4.6可以看出，本章的模型与其他模型相比，达到了最好的或第二好的性能。本章的模型可以达到AP的最佳值，但经过高IOU阈值过滤后，其竞争力减弱。对于 AP_{50} ，基于大规模数据提出的通用模型FastInst^[105]的性能最好。对比的两个半监督模型——Noisy Boundary^[46]和PAIS^[111]，保持相对稳定的分割性能。由于没有两阶段的训练，iFS-RCNN^[110]失去了在TrashCan数据集上的竞争力。综合看来，本章的模型可以基于不同比例的训练集在TrashCan上获得相对稳定的高性能。基于TrashCan的3%的训练集，图4.8展示了对验证集实例分割的可视化结果，可以看出本章的模型可以有准确的类别预测和精细的掩码。

表 4.6 不同模型在TrashCan的定量结果

方法	1%			2%			3%		
	AP	AP_{50}	AP_{75}	AP	AP_{50}	AP_{75}	AP	AP_{50}	AP_{75}
MRCNN ^[58]	1.13	2.98	0.85	4.01	10.17	2.16	6.28	13.70	4.63
FastInst ^[105]	1.51	3.63	1.24	4.15	11.21	2.61	6.01	14.87	3.94
iFS-RCNN ^[110]	1.08	2.62	0.76	3.52	8.04	2.69	4.59	10.18	3.31
Noisy Boundary ^[46]	1.20	2.90	0.80	<u>4.46</u>	9.92	3.88	6.28	12.53	5.61
PAIS ^[111]	1.74	4.54	1.26	4.09	9.45	<u>3.45</u>	6.21	<u>14.79</u>	3.99
Ours（基于FastInst）	1.87	<u>3.75</u>	1.94	4.48	<u>11.15</u>	2.87	<u>6.44</u>	14.73	<u>4.82</u>
Ours（基于MRCNN）	<u>1.77</u>	3.80	<u>1.53</u>	4.36	10.31	3.39	6.54	14.46	4.49

本章展示了前面提到的所有模型在COCO2017数据集的80个类别上的实例分割性能，如表4.7和图4.9。与TrashCan相比，COCO2017的难度在于数据集包含

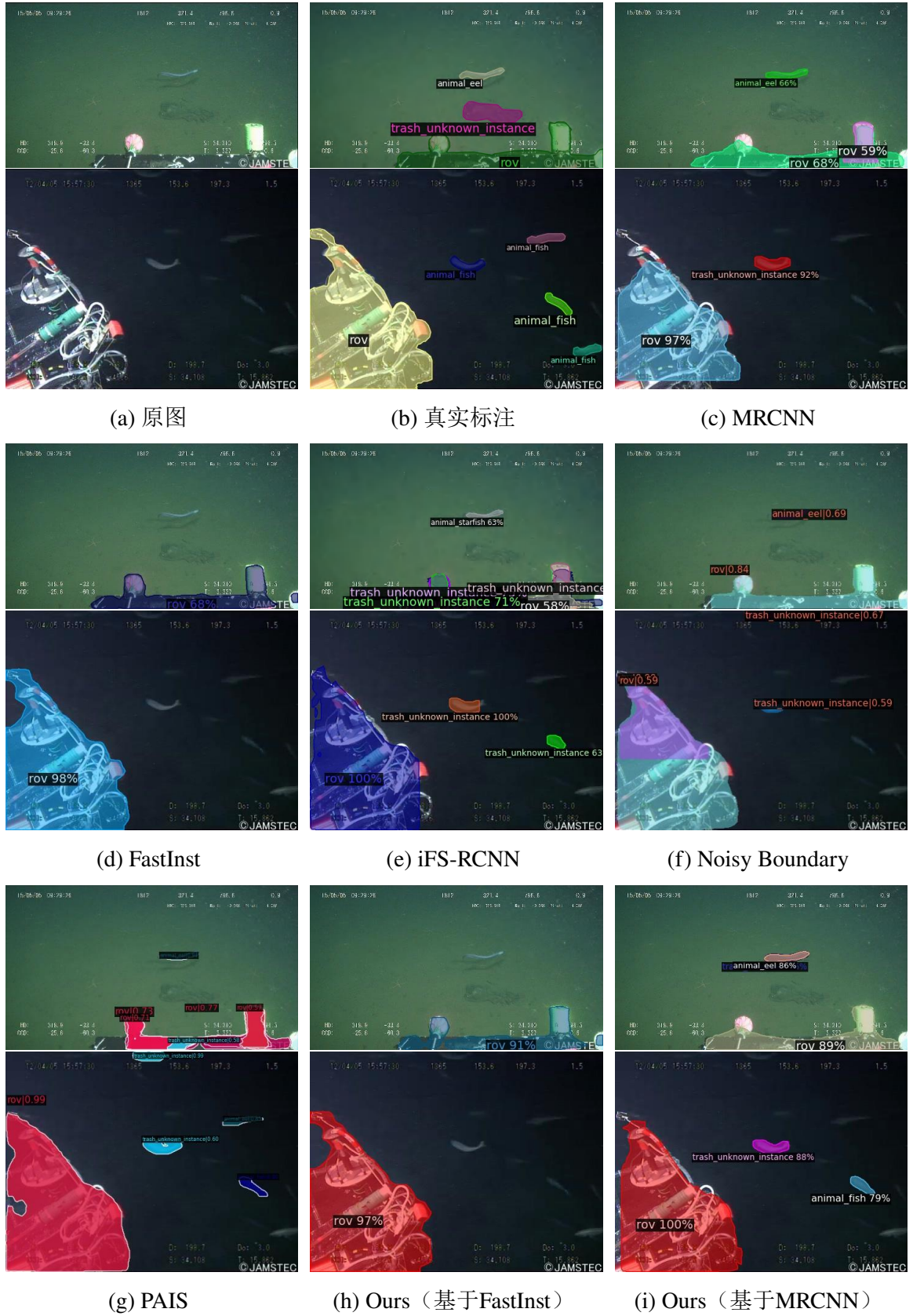


图 4.8 不同模型在TrashCan上得到的分割结果展示

的类别数量多。可以发现，iFS-RCNN^[110]作为一个专门设计用于小样本实例分割的模型不能在COCO2017数据集上实行有效的预测。因为相较于TrashCan以及从COCO2017分离的 D_{Novel} ——iFS-RCNN的工作对象，完整的COCO2017数据集有更复杂的类别，所以iFS-RCNN的失效是可预见的。在没有仔细的元学习训练设计的情况下，iFS-RCNN^[110]在面对更简单的训练策略和更复杂的数据集时，性能降低了很多。从表中还可以发现，尽管Noisy Boundary和PAIS都是基于半监督学习的模型，但PAIS在COCO2017数据集上的性能要比Noisy Boundary差得多。总体来说，本章的模型在COCO2017数据集上取得了最好的性能。图4.9展示了在COCO2017上进行实例分割的可视化结果。从可视化结果中可以得出结论，本章的模型仍然优于其他模型，其中，只有图4.9(h)的方法较为准确地获得了第二个例子中的“suitcase”，而在第一个例子中只有图4.9(f)和图4.9(i)的方法检测出了人群中的“bottle”，并且图4.9(i)中的人群掩码更加精细。

表 4.7 不同模型在COCO2017的定量结果

方法	0.3%			0.5%			1%		
	AP	AP ₅₀	AP ₇₅	AP	AP ₅₀	AP ₇₅	AP	AP ₅₀	AP ₇₅
MRCNN ^[58]	4.21	<u>10.11</u>	3.03	6.71	<u>14.79</u>	5.23	<u>9.53</u>	<u>19.98</u>	8.17
FastInst ^[105]	4.41	8.96	<u>3.96</u>	6.48	12.65	<u>6</u>	9.25	17.85	8.58
iFS-RCNN ^[110]	0.20	0.51	0.16	0.31	0.81	0.22	0.95	2.23	0.64
Noisy Boundary ^[46]	4.43	9.51	3.48	6.68	13.71	5.83	9.31	18.29	8.77
PAIS ^[111]	2.78	5.94	2.32	4.33	9.22	3.53	6.11	11.83	5.56
Ours（基于FastInst）	<u>4.67</u>	9.13	4.32	<u>6.76</u>	12.93	6.37	9.49	18.17	<u>8.79</u>
Ours（基于MRCNN）	5.03	11.45	3.75	7.15	15.68	5.75	10.26	20.85	8.86

表4.8展示了在2205双相不锈钢数据集的 χ 相和 σ 相上不同模型的分割性能。在图4.10中展示了2205双相不锈钢数据集的两个相上不同模型的分割可视化结果图。可以发现在第一张图中，小目标物体占主导地位，基于MRCNN的方法有更强的优势，而在第二张图中，大目标物体占主导地位，基于FastInst的性能有所提升，但是总的来说，还是基于MRCNN的方法更有优势。2205双相不锈钢数据集使得基于FastInst的模型失效，和第三章中基于Transformer的语义分割网络失去性能优势有相同之处，大规模的深度神经网络在材料图像的处理上的乏力进一步表明了图像分

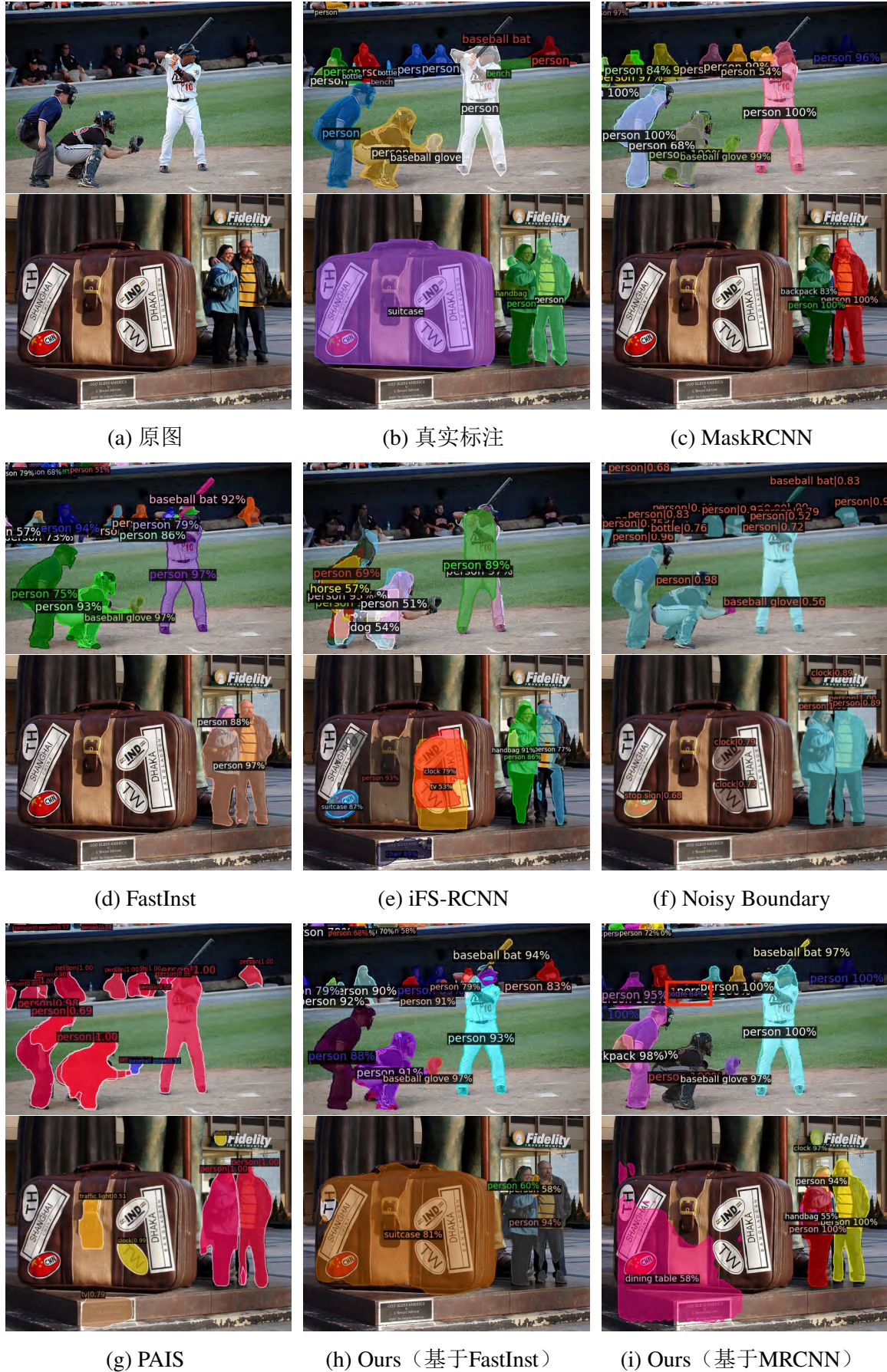


图 4.9 不同模型在COCO2017上得到的分割结果展示

割算法应用于材料图像上存在的迁移困难问题。

表 4.8 不同模型在2205双相不锈钢的定量结果

方法	χ 相			σ 相		
	AP	AP ₅₀	AP ₇₅	AP	AP ₅₀	AP ₇₅
MRCNN ^[58]	36.2	68.79	33.39	59.13	82.89	68.21
FastInst ^[105]	0.99	4.59	0.1	13.69	34.38	7.99
iFS-RCNN ^[110]	20.96	40.20	20.91	53.47	80.6	60.28
Noisy Boundary ^[46]	39.96	<u>69.61</u>	39.03	62.74	84.65	74
PAIS ^[111]	<u>40.23</u>	69.33	46.28	<u>65.79</u>	<u>88.86</u>	<u>78.58</u>
Ours（基于FastInst）	1.7	6.8	0.4	16.16	38.29	11.44
Ours（基于MRCNN）	40.53	70.21	<u>42.78</u>	68.7	89.2	79.53

4.2.5 基于图像实例分割的材料性能挖掘

基于在MRCNN上实现的本章的实例分割模型，可以得到各个二次相的掩码，通过计算各个掩码中非0元素的个数得到二次相所占的像素点个数，非常简便地计算出二次相的尺寸信息。接着，使用直方图统计对实例尺寸的分布进行模拟。 χ 相的半径统计直方图如图4.11， σ 相的半径统计直方图如图4.12。可以看出部分图像中出现了突出的柱，这主要是因为MRCNN是基于候选框的模型，预定义的候选框会导致实例的大小集中在某些数值附近。

基于直方图得到各个时刻的平均尺寸，绘制尺寸变化图，如图4.13所示，和第三章保持一致，本章也使用了不同的数据来源绘制尺寸变化图。可以直观地看出，不同的数据来源得到的变化趋势是非常相近的，这说明本章提出的基于实例分割的方法学习到了人工标注的信息，可以可靠地进行后续的性能分析工作。

同样地，本章对基于模型预测的所有图像得到的图4.13(e)进行尺寸变化，并使用最小二乘法拟合出熟化速率，得到的二次相熟化速率为 $29.366155nm/s^{(1/3)}$ ，和基于人工标注图像的二次相熟化速率 $29.366154nm/s^{(1/3)}$ 仅有微小的差距。和第三章的基于语义分割的方法相比，本章的方法不仅减小了误差还简化了得到分割结果后的材料性能挖掘流程。

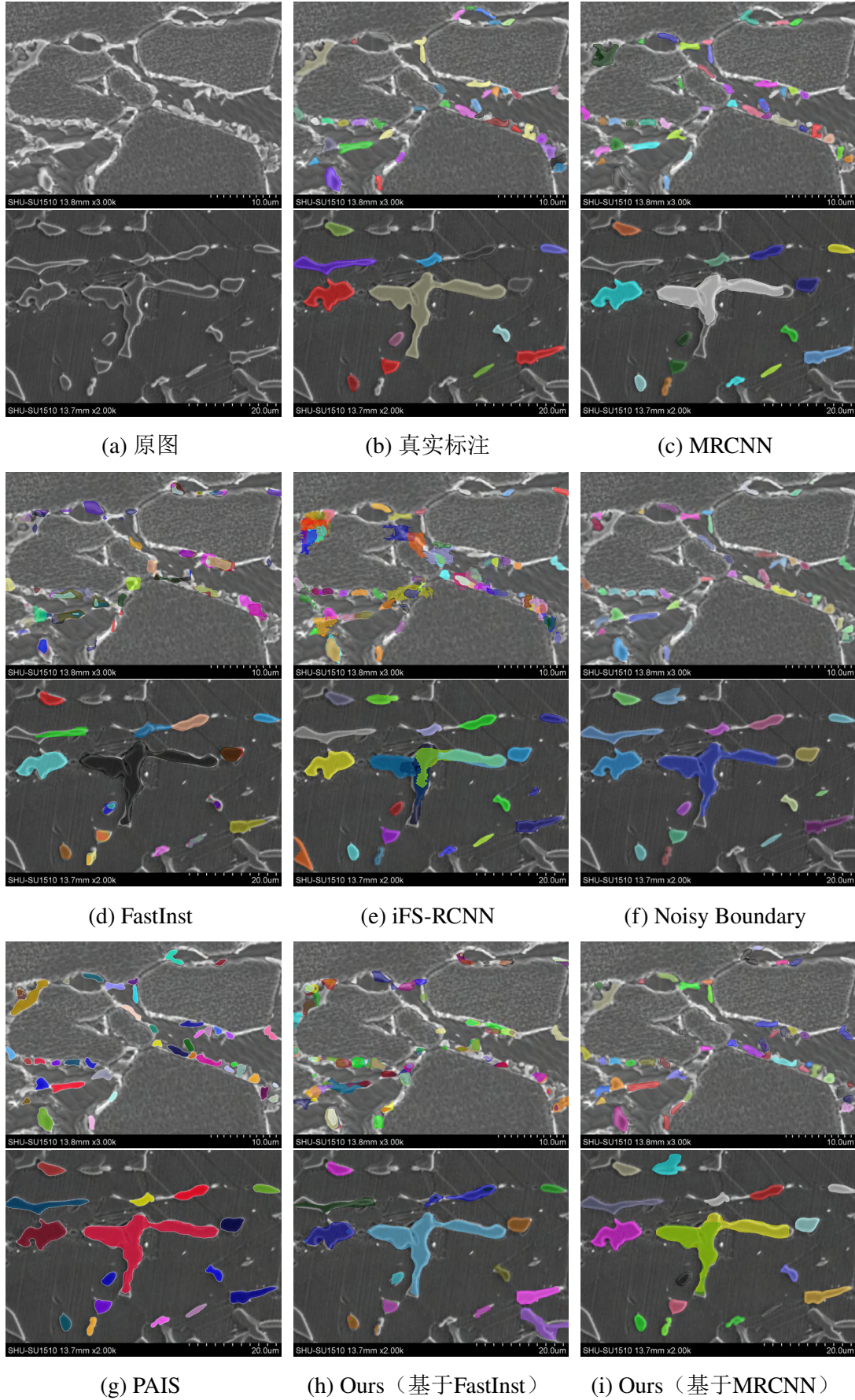


图 4.10 不同模型在2205双相不锈钢上得到的分割结果展示

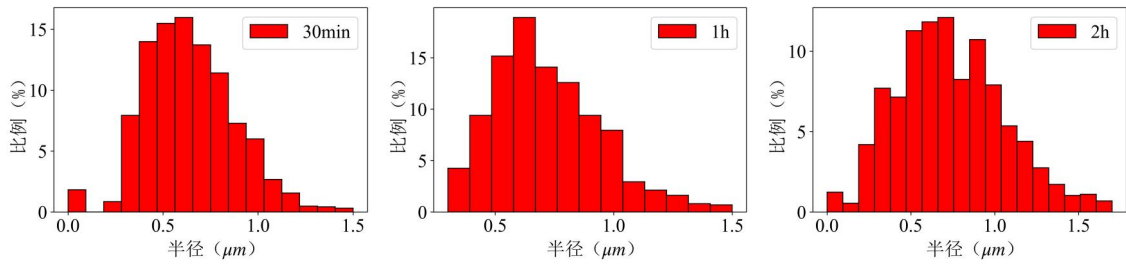


图 4.11 不同时刻的 χ 相尺寸分布

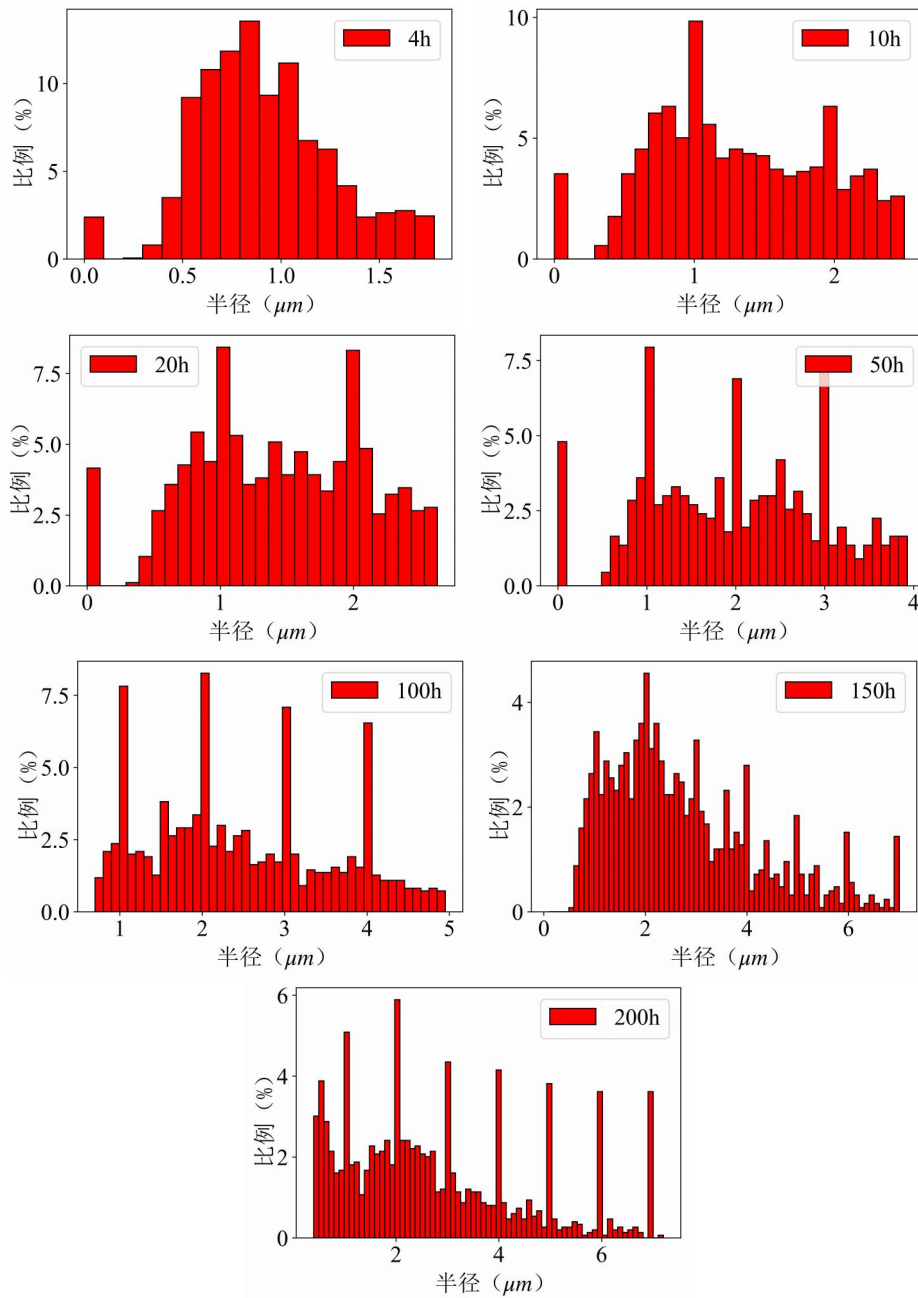


图 4.12 不同时刻的 σ 相尺寸分布

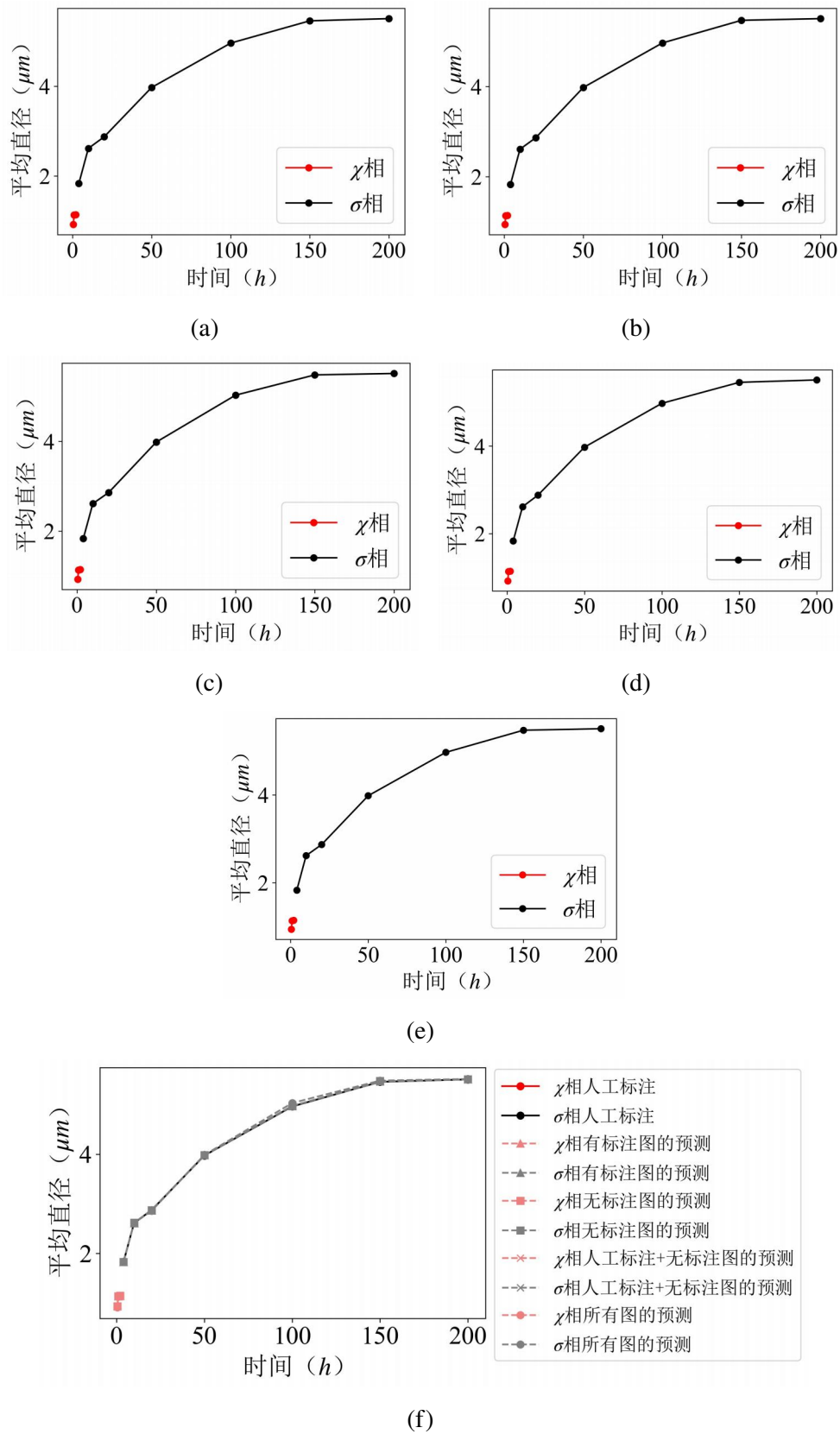


图 4.13 使用不同标注来源得到的二次相平均直径变化图：(a)仅使用人工标注；(b)仅使用模型对有标注图像的预测；(c)仅使用模型对无标注图像的预测；(d)使用人工标注和模型对无标注图像的预测；(e)使用模型对所有图像的预测；(f)整合所有直径变化曲线。

4.3 本章小结

本章主要关注的是矛盾突出的小样本问题，提出了两个解决方案。一个是加入类别的语义信息，经过多模态数据的投影，可以使得对类别的预测更为准确，有效地提高正确预测的置信度、抑制错误预测的置信度。另一个是设计两阶段的训练方式，通过在半监督训练阶段加入动态伪标签，使得未标注样本可以得到利用，增加了模型可用的数据来源，从实验结果看出，伪标签的加入可以帮助精细化分割的掩码、挖掘潜在的目标。本章在三个截然不同的数据集上进行了消融实验和对比实验，证实了本章提出的解决模块的有效性、稳定性和普适性。最后，依照前文提出的材料图像性能挖掘流程，不需要借助额外的数字图像处理技术，本章直接使用实例分割结果得到各个二次相的尺寸信息，并对尺寸进行统计，完成性能分析。

相较于上一章的方法，本章的基于图像实例分割的材料性能挖掘方法虽然在训练阶段较为繁琐，但性能挖掘流程更为简洁，精确性也得到了验证。通过实验可以发现，对于小目标的物体，基于候选框的方法有更好的表现，因为在这类方法中，通过候选框得到的感兴趣区域分割实例掩码，可以缓解因小目标物体和背景因尺寸悬殊带来的类别不平衡问题。而在不基于候选框的方法中，实例分割参照的是整个图像，导致小物体容易被忽略。

第五章 总结与展望

5.1 结论

受益于深度学习的发展，人工智能在许多领域取得了突破性进展，吸引各个领域开展和人工智能的交叉研究。材料科学和深度学习的交叉研究呈现出逐年递增的趋势，但是目前这种交叉研究存在以下问题：材料数据中突出的小样本问题和数据驱动型深度学习算法之间的矛盾；两个学科天然的知识屏障使得以计算机为主导的交叉研究过于侧重算法的提升而忽略了对下游任务的解释。图像分割是计算机获取材料图像微观结构的重要途径之一，本论文分别在小样本图像语义分割和实例分割上提出解决方案，并设计统计分析方法挖掘预测结果和材料性能的内禀关系，即：

(1) 基于特征金字塔和十字交叉注意力的双分支语义分割网络。为了避免高层特征丢失过多的细节信息导致边界分割的不准确，该网络使用空洞卷积、特征金字塔实现网络层的跳跃连接以及双分支结构，使得低层细节可以在更短的路径内得到更新。为了进一步引导网络提升特征的质量，该网络在辅助分支上以较低的计算代价加入循环十字交叉注意力机制，有效地提高了分割的精度。通过大量的实验，证明了本论文提出的语义分割网络在多种小样本材料图像上都可以取得很好的分割效果，超过了当前广泛使用的新模型。

(2) 基于图文多模态融合和伪标签技术的实例分割方法。在材料图像的实例分割任务中，小样本问题更为严峻，本方法聚焦小样本问题的处理，提出两个增强模块：语义模块和基于伪标签的两阶段训练。语义模块通过自然语言模型将类别信息进行编码，通过投影层，实现语义信息和图像信息的融合，引导模型学习更准确的类别信息。在两阶段训练的半监督训练阶段，首先使用数据增强技术对图像进行增强，然后通过筛选高置信度的预测得到伪标签，使未标注数据可以参与模型的训练，增加可学习的信息。由于带有实例分割标签的材料图像数据集十分匮乏，本论文额外增加两个不同场景的公共数据集来验证方法的有效性。在消融实验中，发现了语义模块对分类置信度的影响以及伪标签在探索潜在目标以及精细化掩膜上的能力，对比实验证实了本方法的有效性和普适性。

(3) 根据材料经验公式, 本论文设计了统计分析方法来挖掘图像分割结果和材料性能的内禀关系。根据材料经验公式——奥斯特瓦尔德熟化机制, 性能分析需要对单独的物体进行尺寸统计。由于语义分割任务不区分个体, 本论文使用边界跟踪算法完成对语义分割结果的个体区分, 而实例分割结果不需要额外的处理。基于图像分割结果, 本论文设计使用直方图统计和最小二乘法分析二次相的尺寸变化趋势以及生长速率, 完成材料图像微观结构和材料性能内禀关系的挖掘。通过比较基于深度学习方法和人工方法得到的分析结果, 本论文提出的两种图像分割技术在材料性能挖掘上的可靠性得到验证。

5.2 工作展望

尽管本论文从两个角度提出的两种材料图像性能挖掘方案都可以很好地还原材料中目标物尺寸的变化情况, 为计算机技术在材料领域的深入应用奠定了基础、提供了参考方案, 但是在目前两种方案都存在着局限性。为了进一步提高解决本论文问题的能力, 可以在以下方面继续开展相关研究:

(1) 设计更合适的损失函数监督语义分割网络的训练。材料性能的挖掘需要落到每一个实例上, 由于语义分割将同类别的所有像素点一起进行计算, 不能很好地关注每一个实例的分割效果。具体来说, 对于一个完整的实例, 在语义分割中如果出现断连, 那么只有导致断开的部分会得到惩罚, 但是在后续的统计中, 原本属于一个“大物体”的目标被统计为属于若干个“小物体”的目标, 会导致尺寸统计得到的分布发生偏差, 但这种错误不能很好地在当前的损失函数中反映出来。

(2) 设计更为有效的语义信息和伪标签使用方案。本论文提出的语义模块较为严格, 会删去低置信度的实例, 而伪标签挖掘出的潜在样本多是难区分的低置信度样本, 因此二者有时会出现矛盾, 导致许多实例没有被挖掘出来, 如何使得二者更为协调地合作是可以继续深挖的工作。

(3) 虽然基于实例分割的材料性能挖掘方案使得整个流程更为简洁, 但是后续的分析是基于材料经验公式进行的, 该公式并不能完全准确地描述2205双相不锈钢微观组织的熟化过程。如果可以获得足够的样本, 可以基于深度学习技术探索更准确的公式。

参考文献

- [1] 马俊. 人工智能中美PK: 海量用户是中国独特优势[EB/OL]. 2017[2017-11-22]. https://news.china.com/internationalgd/10000166/20171122/31699726_1.html.
- [2] SUN W, ZHENG Y, YANG K, et al. Machine learning-assisted molecular design and efficiency prediction for high-performance organic photovoltaic materials[J]. Science Advances, 2019, 5(11): eaay4275.
- [3] JIE J, HU Z, QIAN G, et al. Discovering unusual structures from exception using big data and machine learning techniques[J]. Science Bulletin, 2019, 64(9): 612-616.
- [4] KE G, MENG Q, FINLEY T, et al. Lightgbm: A highly efficient gradient boosting decision tree[C]//Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS). 2017: 3149-3157.
- [5] DI S, FENG Y, QU T, et al. Data-driven stress-strain modeling for granular materials through deep reinforcement learning[J]. Chinese Journal of Theoretical and Applied Mechanics, 2021, 53(10): 2712-2723.
- [6] 侯腾跃, 孙炎辉, 孙舒鹏, 等. 机器学习在材料结构与性能预测中的应用综述[J]. 材料导报, 2022, 36(6): 5-12.
- [7] HOLME A, COHN R, GAO N, et al. Overview: Computer vision and machine learning for microstructural characterization and analysis[J]. Metallurgical and Materials Transactions A, 2020, 51: 5985-5999.
- [8] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2005: 886-893.
- [9] LLORCA-ISERN N, LÓPEZ-LUQUE H, LÓPEZ-JIMÉNEZ I, et al. Identification of sigma and chi phases in duplex stainless steels[J]. Materials Characterization, 2016, 112: 20-29.

- [10] BADJI R, BOUABDALLAH M, BACROIX B, et al. Effect of solution treatment temperature on the precipitation kinetic of σ -phase in 2205 duplex stainless steel welds[J]. Materials Science and Engineering: A, 2008, 496: 447-454.
- [11] PAULSEN C O, BROKS R L, KARLSEN M, et al. Microstructure evolution in super duplex stainless steels containing σ -phase investigated at low-temperature using in situ sem/ebsd tensile testing[J]. Metals, 2018, 8(7): 478.
- [12] GAO L, ZHANG X, ZHA X, et al. Effect of mechanical shock treatment on microstructure and corrosion properties of manual argon arc welding joints of 2205 duplex stainless steel[J]. Materials, 2022, 15(9): 036106.
- [13] MIAO L, ZHANG Y, WANG G, et al. Qualitative and quantitative analysis of precipitate phases for 750°C aged 2205 duplex stainless steel[J]. Metallurgical Analysis, 2010, 30: 6-13.
- [14] SIEURIN H, SANDSTRÖM R. Sigma phase precipitation in duplex stainless steel 2205[J]. Materials Science and Engineering: A, 2007, 444: 271-276.
- [15] CHEN T, WENG K, YANG J. The effect of high-temperature exposure on the microstructural stability and toughness property in a 2205 duplex stainless steel[J]. Materials Science and Engineering: A, 2002, 338: 259-270.
- [16] MÉSZÁROS I, SZABO P. Complex magnetic and microstructural investigation of duplex stainless steel[J]. NDT & E International, 2005, 38: 517-521.
- [17] TROCOLI-MONTESINO P, RODRÍGUEZ-PRATO E, ROSALES-MENDOZA A, et al. Ultrasonic evaluation of the formed phases in a duplex stainless steel 2205 heat treated at 750°C and corrosion behavior[J]. Revista De Metalurgia, 2019, 55(1): e135.
- [18] POHL M, STORZ O, GLOGOWSKI T. Effect of intermetallic precipitations on the properties of duplex stainless steel[J]. Materials Characterization, 2007, 58(1): 65-71.
- [19] HASKEL H L, SANCHES L S, PONTE H D A. A new methodology of nondestructive testing for quantitative evaluation of sigma phase in duplex stainless steels[J]. Materials Research, 2019, 22(3): e20180682.

- [20] DJAMA M, SAIDI D, KADRI A, et al. Correlation between the pitting potential evolution and σ phase precipitation kinetics in the 2205 duplex stainless steel[J]. Journal of Materials Engineering and Performance, 2018, 27: 3911-3919.
- [21] YAMASHITA S, IKE K, YAMASAKI K, et al. Relationship between ferrite austenite phase transformation and precipitation behavior of sigma phase in super duplex stainless steel weldment[J]. Welding in the World, 2022, 66: 351-362.
- [22] NESLUŠAN M, TROJAN K, HAUŠILD P, et al. Monitoring of components made of duplex steel after turning as a function of flank wear by the use of Barkhausen noise emission[J]. Materials Characterization, 2020, 169: 110587.
- [23] BIEZMA M, MARTIN U, LINHARDT P, et al. Non-destructive techniques for the detection of sigma phase in duplex stainless steel: A comprehensive review[J]. Engineering Failure Analysis, 2021, 122: 105227.
- [24] JACKSON E, VISSER P, CORNISH L. Distinguishing between chi and sigma phases in duplex stainless steels using potentiostatic etching[J]. Materials Characterization, 1993, 31: 185-190.
- [25] VOORHEES P W. The theory of Ostwald ripening[J]. Journal of Statistical Physics, 1985, 38: 231-252.
- [26] 雍岐龙. 钢铁材料中的第二相[M]. 北京: 冶金工业出版社, 2006: 418-494.
- [27] LU L, ZHU N Q, HE Y L, et al. Study on coarsening behavior of carbide in Fe-Cr-C alloy during high temperature aging[J]. Transactions of Materials and Heat Treatment, 2011, 32: 53-56.
- [28] WANG M, WANG J, HE Y, et al. Experimental study and calculation of the precipitation behavior of carbides in Fe-5.78Cr-0.35C alloy during high-temperature aging [J]. Advanced Materials Research, 2014, 936: 1184-1188.
- [29] ZHANG H, CISSE M, DAUPHIN Y N, et al. mixup: Beyond empirical risk minimization[C]//International Conference on Learning Representations (ICLR). 2018.
- [30] DABOUEI A, SOLEYMANI S, TAHERKHANI F, et al. Supermix: Supervising the mixing data augmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021: 13794-13803.

- [31] HONG M, CHOI J, KIM G. Stylemix: Separating content and style for enhanced data augmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021: 14862-14870.
- [32] CHOI H K, CHOI J, KIM H J. Tokenmixup: Efficient attention-guided token-level data augmentation for transformers[C]//Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS). 2022: 14224-14235.
- [33] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[C]//Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS). 2014: 2672-2680.
- [34] DHARIWAL P, NICHOL A. Diffusion models beat gans on image synthesis[C]//Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS). 2021: 8780-8794.
- [35] SHAHAM T R, DEKEL T, MICHAELI T. Singan: Learning a generative model from a single natural image[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 4570-4580.
- [36] CHEN D, LIU Y, HUANG L, et al. Geoaug: Data augmentation for few-shot nerf with geometry constraints[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2022: 322-337.
- [37] FRID-ADAR M, DIAMANT I, KLANG E, et al. Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification[J]. Neurocomputing, 2018, 321: 321-331.
- [38] MAHAPATRA D, BOZORGTABAR B, SHAO L. Pathological retinal region segmentation from oct images using geometric relation based augmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 9611-9620.
- [39] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning [C]//Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS). 2017: 4080-4090.

- [40] VINYALS O, BLUNDELL C, LILLICRAP T, et al. Matching networks for one shot learning[C]//Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS). 2016: 3637-3645.
- [41] SAJJADI M, JAVANMARDI M, TASDIZEN T. Regularization with stochastic transformations and perturbations for deep semi-supervised learning[C]//Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS). 2016: 1171-1179.
- [42] LAINE S, AILA T. Temporal ensembling for semi-supervised learning[C]//International Conference on Learning Representations (ICLR). 2016.
- [43] TARVAINEN A, VALPOLA H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[C]//Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS). 2017: 1195-1204.
- [44] MIYATO T, MAEDA S I, KOYAMA M, et al. Virtual adversarial training: a regularization method for supervised and semi-supervised learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 41(8): 1979-1993.
- [45] VERMA V, KAWAGUCHI K, LAMB A, et al. Interpolation consistency training for semi-supervised learning[J]. Neural Networks, 2022, 145: 90-106.
- [46] XIE Q, LUONG M T, HOVY E, et al. Self-training with noisy student improves imagenet classification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 10687-10698.
- [47] PHAM H, DAI Z, XIE Q, et al. Meta pseudo labels[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021: 11557-11568.
- [48] BERTHELOT D, CARLINI N, GOODFELLOW I, et al. Mixmatch: A holistic approach to semi-supervised learning[C]//Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS). 2019: 5049-5059.

- [49] SOHN K, BERTHELOT D, CARLINI N, et al. Fixmatch: Simplifying semi-supervised learning with consistency and confidence[C]//Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS). 2020: 596-608.
- [50] HAN Y, LI R, YANG S, et al. Center-environment feature models for materials image segmentation based on machine learning[J]. Scientific Reports, 2022, 12: 12960.
- [51] SEZGIN M, SANKUR B. Survey over image thresholding techniques and quantitative performance evaluation[J]. Journal of Electronic Imaging, 2004, 13(1): 146-168.
- [52] STAN T, THOMPSON Z T, VOORHEES P W. Optimizing convolutional neural networks to perform semantic segmentation on large materials imaging datasets: X-ray tomography and serial sectioning[J]. Materials Characterization, 2020, 160: 110119.
- [53] ROBERTS G, HAILE S Y, SAINJU R, et al. Deep learning for semantic segmentation of defects in advanced stem images of steels[J]. Scientific Reports, 2019, 9: 12744.
- [54] HAN Y, YANG S, CHEN Q. Recognition and segmentation of complex texture images based on superpixel algorithm and deep learning[J]. Computational Materials Science, 2022, 209: 111398.
- [55] MAKSOV A, DYCK O, WANG K, et al. Deep learning analysis of defect and phase evolution during electron beam-induced transformations in WS_2 [J]. npj Computational Materials, 2019, 5: 12.
- [56] WANG W, YU R, HUANG Q, et al. Sgpn: Similarity group proposal network for 3d point cloud instance segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 2569-2578.
- [57] QI C R, YI L, SU H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[M]//Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS). 2017: 5105-5114.
- [58] HE K, GKIOXARI G, DOLLAR P, et al. Mask r-cnn[C]//Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017: 2961-2969.

- [59] WANG X, KONG T, SHEN C, et al. Solo: Segmenting objects by locations[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2020: 649-665.
- [60] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 779-788.
- [61] CHENG B, MISRA I, SCHWING A G, et al. Masked-attention mask transformer for universal image segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022: 1290-1299.
- [62] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2020: 213-229.
- [63] FAN Z, YU J G, LIANG Z, et al. Fgn: Fully guided network for few-shot instance segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 9172-9181.
- [64] HAN Y, ZHANG J, XUE Z, et al. Reference twice: A simple and unified baseline for few-shot instance segmentation[EB/OL]. 2023. <https://arxiv.org/pdf/2301.01156>.
- [65] NGUYEN K, TODOROVIC S. Fapis: A few-shot anchor-free part-based instance segmenter[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021: 11099-11108.
- [66] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C]//Proceedings of the International Conference on Machine Learning (ICML). 2021: 8748-8763.
- [67] XU J, DE MELLO S, LIU S, et al. Groupvit: Semantic segmentation emerges from text supervision[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022: 18134-18144.
- [68] CHEN Z, FU Y, ZHANG Y, et al. Multi-level semantic feature augmentation for one-shot learning[J]. IEEE Transactions on Image Processing, 2019, 28(9): 4594-4605.

- [69] RAHMAN S, KHAN S, BARNES N, et al. Any-shot object detection[C]//Proceedings of the Asian Conference on Computer Vision (ACCV). 2020: 89-106.
- [70] ZHU C, CHEN F, AHMED U, et al. Semantic relation reasoning for shot-stable few-shot object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021: 8782-8791.
- [71] JAIN J, LI J, CHIU M T, et al. Oneformer: One transformer to rule universal image segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2023: 2989-2998.
- [72] SONG L, WANG D, LIU X, et al. Prediction of mechanical properties of composite materials using multimodal fusion learning[J]. Sensors and Actuators A: Physical, 2023, 358: 114433.
- [73] AI Y, LEI C, CHENG J, et al. Prediction of weld area based on image recognition and machine learning in laser oscillation welding of aluminum alloy[J]. Optics and Lasers in Engineering, 2023, 160: 107258.
- [74] BALAJI V N S. Optimization of wire-edm process parameters for Ni–Ti–Hf shape memory alloy through particle swarm optimization and cnn-based sem-image classification[J]. Results in Engineering, 2023, 18: 101141.
- [75] REN D, WANG C, WEI X, et al. Building a quantitative composition-microstructure-property relationship of dual-phase steels via multimodal data mining[J]. Acta Materialia, 2023, 252: 118954.
- [76] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 770-778.
- [77] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2015: 3431-3440.
- [78] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2015: 234-241.

- [79] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40 (4): 834-848.
- [80] LI F, ZHANG H, XU H, et al. Mask dino: Towards a unified transformer-based framework for object detection and segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2023: 3041-3050.
- [81] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: towards real-time object detection with region proposal networks[C]//Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS). 2015: 91-99.
- [82] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS). 2017: 5998-6008.
- [83] ZHANG H, LI F, LIU S, et al. Dino: Detr with improved denoising anchor boxes for end-to-end object detection[C]//International Conference on Learning Representations (ICLR). 2022.
- [84] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 7794-7803.
- [85] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 7132-7141.
- [86] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 3-19.
- [87] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 3146-3154.

- [88] BERTHELOT D, CARLINI N, CUBUK E D, et al. Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring[C]//International Conference on Learning Representations (ICLR). 2019.
- [89] EVERINGHAM M, VAN GOOL L, WILLIAMS C, et al. The pascal visual object classes (voc) challenge[J]. International Journal of Computer Vision, 2010(2): 303-338.
- [90] SUZUKI S, et al. Topological structural analysis of digitized binary images by border following[J]. Computer Vision, Graphics, and Image Processing, 1985, 30(1): 32-46.
- [91] YU F, KOLTUN V, FUNKHOUSER T. Dilated residual networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 472-480.
- [92] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[C]//International Conference on Learning Representations (ICLR). 2016.
- [93] ZHOU B, KHOSLA A, LAPEDRIZA A, et al. Learning deep features for discriminative localization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2016: 2921-2929.
- [94] XIAO T, LIU Y, ZHOU B, et al. Unified perceptual parsing for scene understanding [C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 418-434.
- [95] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 2881-2890.
- [96] HUANG Z, WANG X, HUANG L, et al. Ccnet: Criss-cross attention for semantic segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 603-612.
- [97] SHANNON C E. A mathematical theory of communication[J]. The Bell System Technical Journal, 1948, 27(3): 379-423.

- [98] RUBINSTEIN R. The cross-entropy method for combinatorial and continuous optimization[J]. *Methodology and Computing in Applied Probability*, 1999, 1: 127-190.
- [99] VAN DER MAATEN L, HINTON G. Visualizing data using t-sne[J]. *Journal of Machine Learning Research*, 2008, 9(11): 2579-2605.
- [100] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//*Proceedings of the European Conference on Computer Vision (ECCV)*. 2018: 801-818.
- [101] KIRILLOV A, WU Y, HE K, et al. Pointrend: Image segmentation as rendering[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020: 9799-9808.
- [102] ZHANG W, PANG J, CHEN K, et al. K-net: Towards unified image segmentation [C]//*Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*. 2021: 10326-10338.
- [103] XIE E, WANG W, YU Z, et al. Segformer: Simple and efficient design for semantic segmentation with transformers[C]//*Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*. 2021: 12077-12090.
- [104] SCHWIND M, KÄLLQVIST J, NILSSON J O, et al. σ -phase precipitation in stabilized austenitic stainless steels[J]. *Acta Materialia*, 2000, 48(10): 2473-2481.
- [105] HE J, LI P, GENG Y, et al. Fastinst: A simple query-based model for real-time instance segmentation[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2023: 23663-23672.
- [106] BELTAGY I, LO K, COHAN A. Scibert: A pretrained language model for scientific text[C]//*Proceedings of the Conference on Empirical Methods in Natural Language Processing and the International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 2019: 3615-3620.
- [107] MILLETARI F, NAVAB N, AHMADI S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]//*Proceedings of the International Conference on 3D Vision (3DV)*. 2016: 565-571.

- [108] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2014: 740-755.
- [109] HONG J, FULTON M, SATTAR J. Trashcan: A semantically-segmented dataset towards visual detection of marine debris[EB/OL]. 2020. <https://arxiv.org/pdf/2007.08097>.
- [110] NGUYEN K, TODOROVIC S. ifs-rcnn: An incremental few-shot instance segmenter [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022: 7010-7019.
- [111] HU J, CHEN C, CAO L, et al. Pseudo-label alignment for semi-supervised instance segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2023: 16337-16347.

攻读硕士学位期间取得的研究成果

一、论文

1. Han Y, Chi R, Chen Q, et al. Microstructural evolution and coarsening behavior of the precipitates in 2205 duplex stainless steel aged at 850°C[J]. Journal of Materials Research and Technology, 2023, 26: 2560-2574. (已发表, 导师一作, 本人二作, 中科院一区, JCR一区)
2. Han Y, Chi R, Chen Q. Integrated Image-Text Based on Semi-supervised Learning for Few-shot Instance Segmentation. (导师一作, 本人二作, submitted: IEEE Transactions on Pattern Analysis and Machine Intelligence)

二、软著

1. 软件名称: 相间能量计算平台V1.0, 开发人: 韩越兴、池洳婷、王冰。登记号: 2022SR1402923, 登记日期: 2022.10.14, 申请人: 上海大学。(导师一作, 本人二作)
2. 软件名称: 双相不锈钢析出物种类和尺寸统计平台V1.0, 开发人: 韩越兴、池洳婷、王冰。登记号: 2023SR1539665, 登记日期: 2023.11.30, 申请人: 上海大学。(导师一作, 本人二作)

攻读硕士学位期间参与的科研项目

1. 项目来源：国家自然科学基金面上项目

项目名称：基于小样本机器学习的Mg-Gd系合金图像-组织-性能的内禀关系研究

项目编号：52273228

执行期限：2023.01-2026.12

2. 项目来源：科学技术部

项目名称：知识与数据双驱动的树脂基复合材料智能设计共性关键技术研发与空间应用

项目编号：2022YFB3707800

执行期限：2022.11-2025.10

3. 项目来源：科技部国家重点研发计划

项目名称：材料大数据技术研究

项目编号：2020YFB0704500

执行期限：2020.09-2022.08

4. 项目来源：上海市自然科学基金

项目名称：小样本环境下物体自适应识别方法研究

项目编号：20ZR1419000

执行期限：2020.07-2023.06

5. 项目来源：科技部国家重点研发计划

项目名称：材料复杂显微结构高精度图像处理方法研究

项目编号：2018YFB0704400 2018YFB0704402

执行期限：2018.07-2022.06

致 谢

短暂的研究生学习时光悄然而逝，相识于落叶纷飞的秋季，而在生机盎然的初夏挥手告别。在此，我要向给予我帮助和支持的人们致以最真挚的感谢。

首先我要感谢我的导师韩越兴老师，从论文选题到实验设计再到终稿的完成，都离不开韩老师的悉心指导和帮助。科研上，韩老师秉持着一丝不苟的治学风气，感谢韩老师不厌其烦地为我的研究和论文撰写提出宝贵的意见，他对学术的一腔热血极大地鼓舞了迷茫的我，敦促我严谨治学，使我坚韧不屈。生活中，韩老师为人幽默诙谐、平易近人，创造了轻松愉悦的科研氛围。同时，我还要感谢陈侨川老师精准的指点和创新的思路，以及刘微老师提供的材料数据以及文献资料。在此衷心地祝愿各位老师身体健康、学术长青。

其次，我要感谢我的家人们。感恩我的父母教会我不轻言放弃的态度以及给我探索无限可能的动力、永远真诚和坦荡的底气；感恩他们让我站在他们的肩膀上看到了更为广阔的世界。感谢我的弟弟，总是用最实际的行动和最贴心的话语分担我的压力。感谢在异乡求学的年岁里始终伴我左右的另一半，感谢他在我无数次的失落中给予的肯定和支持，让我始终相信美好，充满对未来的期待和向往。感谢我的所有亲人们，唯愿你们长安久乐、身体康健，余生平安顺遂。

学贵得师，亦贵得友。感谢师兄师姐们在过去时光里给予的包容以及提供的宝贵经验，感谢同窗的互帮互助，还要感谢师弟师妹们带来的欢声笑语。从他们的身上，我学到了勤奋、细致，学到了善良、热情，学到了在科研的同时要享受生活的态度，感谢你们让我度过了舒心悦的三年时光。

再回首，轻舟已过万重山。再次感谢所有伸出援手的人，愿天高海阔，万事胜意。愿自己，身怀赤诚，于万物众生中磊落做人。