

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	基于图卷积神经网络的材料图像分割研究
--------	--------------------

作 者 魏惠姍

学科专业 计算机应用技术

导 师 韩越兴

完成日期 2022.04.01

姓名：魏惠姗

学号：19721562

论文题目：基于图卷积神经网络的材料图像分割研究

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：

委员：

导 师：

答辩日期：

姓名：魏惠姗

学号：19721562

论文题目：基于图卷积神经网络的材料图像分割研究

原创性声明

本人声明：所提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：_____日 期：_____

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：_____导师签名：_____日期：_____

上海大学学术硕士学位论文

基于图卷积神经网络的材料图像
分割研究

姓 名：魏惠姗

导 师：韩越兴

学科专业：计算机应用技术

上海大学计算机工程与科学学院

2022 年 5 月

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

**Research on Material Image
Segmentation Based on Graph
Convolutional Neural Network**

MA Candidate: Huishan Wei

Supervisor: Yuexing Han

Major: Computer Applications Technology

School of Computer Engineering and Science

Shanghai University

May, 2022

摘 要

随着多学科间的交叉和融合，科学工作者开始尝试将计算机视觉技术运用到材料科学领域，使传统的材料研究手段得以巨大的改观。在材料科学研究中，材料图像分割是材料图像分析的重要步骤之一。材料图像处理方法主要分为基于传统的图像处理方法、基于机器学习的图像处理方法和基于深度学习的图像处理方法。基于传统的图像处理方法和基于机器学习的图像处理方法存在适用范围小、依赖人工参与和分割不准确等缺点。目前的大多数深度学习方法可以对特定场景的图像进行有效分割，但是材料数据样本过少、材料图像纹理复杂和材料结构形状各异等问题限制了深度学习的广泛应用，缺乏一个通用的方法对不同结构的材料图像进行准确的分割。为了提高对材料图像分割任务的处理能力，本文针对材料图像的特点，结合多种深度学习技术，提出了两种材料图像的分割方法。

首先，针对材料图像中存在的数据集规模小、纹理复杂和结构尺寸差异大等问题，本文提出了一种基于图卷积结合多维特征融合的材料图像分割方法。该方法利用材料图像语义信息丰富的特点，先用编码器提取高级特征图；然后使用双重注意力机制学习通道维度和空间维度的注意力为特征图分配权重，将图卷积结合空间注意力的权重矩阵再次为各像素分配注意力权重；紧接着，使用不同感受野范围的空洞卷积和不同尺寸维度的池化操作进行特征图的多尺度融合；最后，使用解码器输出精确的分割结果。该方法将注意力机制和图卷积相结合增加对重要像素的关注，通过残差连接和多尺度融合提升对图像细节分割的能力。

接着，针对上述方法堆叠过多卷积层导致特征信息损失的问题，本文提出了基于跳跃连接的 Graph-UNet。该方法以 UNet 为骨干网络，在下采样阶段获取高层语义特征，使用图编码器将特征图转化为图结构；然后，使用图注意力模块学习图结构中节点的注意力权重，加入超参数避免节点特征趋于全局化，依靠跳跃连接融合不同维度的节点特征；紧接着，使用图解码器将节点特征更

新后的图结构转化为特征图；最后，在上采样阶段输出精确的分割结果。该方法利用多层图卷积和图注意力层有效组合代替卷积层，提升了网络对材料图像的分割精度，并提出一套通用的图编码器和图解码器，实现图像和图结构无障碍的相互转化，推动了图卷积神经网络在语义分割领域的应用和发展。

本文提出的两种深度学习方法可以在材料图像中实现端到端的训练，并从不同角度在多种材料的电子显微镜图像数据集上对提出的方法进行了详细的对比实验和消融实验。利用这两种方法，本文可以准确识别材料图像中的多种结构，为小样本的材料图像分割方法的发展提供了新的思路，为材料科学领域结构分析的新机制提供了重要的见解。

关键词：材料图像，深度学习，注意力机制，图卷积神经网络，小样本

ABSTRACT

With the cross and fusion of multiple disciplines, scientists began to try to apply computer vision technology to the field of materials science, which greatly changed the traditional materials research methods. In material science research, material image segmentation is one of the important steps in material image analysis. Material image processing is mainly divided into traditional image processing methods, image processing methods based on machine learning and image processing methods based on deep learning. Traditional image processing methods and machine learn-based image processing methods have some disadvantages such as small scope of application, dependence on manual participation and inaccurate segmentation. Most of the current deep learning methods can effectively segment the images of specific scenes, but the wide application of deep learning is limited by problems such as too few material data samples, complex texture of material images and different structure shapes of materials, and there is a lack of a general method for accurate segmentation and analysis of material images with different structures. In order to improve the processing ability of material image segmentation task, this paper proposes two material image segmentation methods according to the characteristics of material images and combined with a variety of technologies based on deep learning.

Firstly, aiming at the problems of small size of data set, complex texture and large size difference of structure in material image, a material image segmentation method based on graph convolution and multi-dimensional feature fusion is proposed in this paper. Based on the rich semantic information of material images, the advanced feature maps are extracted by encoder. Then, the dual attention mechanism is used to learn the attention of channel dimension and spatial dimension to assign weight to feature maps, and the graph convolution is

combined with the weight matrix of spatial attention to assign attention weight to each pixel again. Next, the multi-scale fusion of feature images is performed by using the void convolution of different receptive field ranges and the pooling operation of different dimensions. Finally, the decoder outputs the exact segmentation results. This method combines the attention mechanism with graph convolution to increase the attention of important pixels. Multi-dimensional feature images were obtained by residual connection and multi-scale fusion to improve the segmentation ability of image details.

Then, aiming at the problem of feature information loss caused by stacking too many convolutional layers in the above method, Graph-UNet based on skip connections is proposed in this paper. In this method, UNet is used as the backbone network to obtain high-level semantic features in the down-sampling stage, and graph encoder is used to transform feature map into graph structure. Then, the graph attention module is used to learn the attention weight of nodes in the graph structure, adding hyperparameters to avoid the node features tend to be global, and integrating node features of different dimensions by skipping connections. Then, graph decoder is used to transform the updated graph structure into feature map. Finally, the accurate segmentation results are output in the up-sampling stage. The method based on the multilayer diagram layer effective combination instead of convolution convolution and attention, improve the network on the material image segmentation accuracy, and puts forward a set of general image encoder and decoder, realize the mutual transformation between image and figure structure accessible, promoted the figure convolution neural network application and development in the field of semantic segmentation.

The two deep learning methods proposed in this paper can achieve end-to-end training in material images, and detailed comparison experiments and ablation experiments of the proposed methods are carried out on electron

microscope image datasets of various materials from different angles. Using these two methods, multiple structures in material images can be accurately identified in this paper, which provides a new idea for the development of material image segmentation methods for small samples, and an important insight for the new mechanism of structure analysis in the field of materials science.

Keywords: material image, deep learning, attention mechanism, graph convolutional neural network, small sample

目录

第一章 绪论	1
1.1 课题来源	1
1.2 课题研究背景与意义	1
1.3 国内外研究概况	3
1.3.1 基于传统图像处理技术的分割方法	3
1.3.2 基于机器学习技术的分割方法	4
1.3.3 基于深度学习技术的分割方法	5
1.3.4 基于图卷积神经网络的分割方法	8
1.4 论文主要工作	9
第二章 相关理论和技术概述	11
2.1 全卷积神经网络的基本概念	11
2.2 本文相关的卷积神经网络框架	14
2.3 经典的基于语义分割的深度学习方法	16
2.4 图卷积	20
2.4.1 应用于图结构的图卷积	20
2.4.2 图卷积和卷积的区别	23
2.4.3 应用于语义分割的图卷积	25
2.5 常见的图像分割评价指标	27
2.6 本章小结	28
第三章 基于图卷积与深度学习的编解码网络	29
3.1 方法概述	29
3.2 基于图卷积结合多维特征融合的网络模型	30
3.2.1 编码器和解码器	30
3.2.2 图注意力	32

3.2.3	多尺度特征融合模块	37
3.3	实验和讨论	39
3.3.1	实验环境与评价标准	39
3.3.2	现有方法和 FMGNet 的对比实验	40
3.3.3	不同骨干网络的 FCN 和 FMGNet 的对比实验	44
3.3.4	消融实验	46
3.4	本章小结	50
第四章	基于跳跃连接的 Graph-UNet	52
4.1	方法概述	52
4.2	语义分割中的图模型	53
4.2.1	图编码器	54
4.2.2	基于跳跃连接的图注意力模块	54
4.2.3	图解码器	57
4.3	实验和讨论	58
4.3.1	实验环境与评价标准	58
4.3.2	现有方法和 Graph-UNet 的对比实验	59
4.3.3	验证图注意力模块的有效性	63
4.3.4	消融实验	64
4.3.5	Graph-UNet 和 FMGNet 的对比实验	67
4.4	本章小结	69
第五章	结论和展望	71
5.1	结论	71
5.2	展望	72
	参考文献	73
	作者在攻读硕士学位期间公开发表的论文	85

第一章 绪论

1.1 课题来源

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

1.2 课题研究背景与意义

随着计算机视觉技术的蓬勃发展，以及多学科间的深入交叉和融合，科学工作者开始尝试将图像分割为代表的计算机图像处理技术运用到材料科学领域 [1]。计算机视觉与材料科学的融合使传统的材料研究方法在理论和实验两方面取得了很大的改进，不仅使理论研究从分析推导的束缚中解放出来，而且使实验研究方法得到根本性的变革 [2,3]。两个领域间的有效融合有利于后续的分析、处理建立在更加严谨的基础上，有利于从可视化现象中揭示客观规律和从少量的样本数据中探索丰富的潜在信息。材料图像分割是统计材料微观组织参数进行后续处理的基础，分割后的处理任务都取决于图像分割的质量，而计算机图像处理技术的引入，大大提高了晶粒尺寸测量、夹杂物分级、相分析、显微硬度、孔隙率、球化率和涂层厚度等定量分析的质量。因此，材料图像中目标区域的准确分割对于结构分析、材料性能研究等都具有重要意义。

不同于自然场景图像分割任务，材料图像处理具有以下特点：1) 小样本特性：制作实验易受环境、设备和操作方式的影响造成材料制备过程困难，图像标注需要

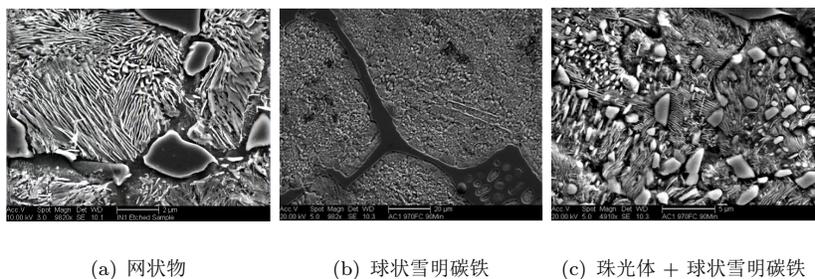


图 1.1: 不同组织结构的材料图像 [4]。(a) 的主要显微成分是网状物, (b) 的主要显微成分是球状雪明碳铁, (c) 的主要显微成分是球状雪明碳铁和珠光体。

专业领域的专家为像素赋予标签导致标签数据的稀少, 设备昂贵和跨领域人才匮乏导致投入科研成本巨大。2) 色彩单调特性: 材料的微观结构一般是由电子显微镜拍摄, 均是单色素图像, 例如灰度图像。根据材料结构呈现的明暗, 灰度图用黑白灰的形式显示。3) 纹理复杂特性: 材料结构本身有形状结构各异、纹理表征差别不明显和边界不连续甚至模糊等特性。这三个特点的存在, 导致研究者们需要在样本数据少量的前提条件下实现对多种材料图像准确分割, 增加了材料图像分割算法的设计难度, 提升了对网络模型泛化性能的需求。

上述三个材料图像处理的特点导致基于深度学习的方法在分割精度上难以提升。如图1.1, 第一张图片的主要显微成分是网状物 (Network), 它的结构错综复杂, 并且包含大量的噪声。第二张图的主要显微成分是球状雪明碳铁 (Spheroidite), 相的形状较为明显, 但是对比度低, 相的边界部分模糊导致难以清楚区分目标相与背景的边界。第三张图的主要显微成分是球状雪明碳铁 (Spheroidite) 和珠光体 (Pearlite), 可以清楚看到相同物相的大小差别很大, 这些特点都大大提高了材料图像分割的难度。

自动化分割小样本材料图像具有巨大的难度, 而在研究中材料结构和性能的分析又迫切需要高精度的自动化分割方法帮助研究者为后续的工作奠定基础。深度学习技术的发展为迎接这一挑战提供了契机。不同于图像分割中的传统的全卷积神经网络, 解决不规则数据分类或分割任务的图卷积神经网络凭借灵活的节点连接和特有的图卷积方式也被研究者开始应用于语义分割任务。因此, 本文尝试结合多种基于卷积神经网络的图像分割技术和图卷积神经网络的方法去探索更适用的材料图像

分割方法，为计算机视觉的研究和小样本材料图像分割的发展献出一份力量。

1.3 国内外研究概况

图像分割是图像分析的重要步骤之一，是计算机视觉的基础之一，同时也是图像处理的难点之一。如今计算机视觉技术已应用于各个领域，如医学图像分割 [5]、材料图像分割 [6] 等。本节将国内外的材料图像分割方法分为三类：基于传统图像处理、机器学习和深度学习技术的分割方法。

1.3.1 基于传统图像处理技术的分割方法

传统的图像分割方法主要分为基于阈值方法 [7]、区域方法 [8]、边缘检测方法 [9–11]、图论方法 [12]、能量泛函方法 [13] 和数学形态学方法 [14] 等。阈值分割法 [7] 的基本思想是通过设置不同的特征阈值，对图像中的所有像素进行分类，其优点是实现简单、计算量小、性能稳定，适用于目标和背景灰度值差距较大的图像。区域分割法 [8] 的基本思想是用划分区域的方式集合相似像素，对于连通均匀的目标具有计算简单、分割效果好的优点，但是需要人为选取种子，对噪声比较敏感。边缘检测法的基本思想是连接图像中灰度值变化梯度较大的点，构成线条，如 Roberts [9]、Sobel [10] 和 Canny 算子 [11] 都是效果非常好的边缘检测算子，但是该方法只能产生边缘点，无法完整的实现图像分割。图论法的基本思想是通过去除特定的边的方式，将图划分为若干子图进行分割。例如 GrabCut [12] 方法是按颜色分布和边缘对比度来分割图像的，只需要简单的交互就可以得到不错的分割图，但是计算量大且分割速度慢。能量泛函法的基本思想是将分割过程转变为求解能量泛函的最小值的过程。例如 Snake [13] 是一种基于参数曲线和相应能量函数的分割方法，在边缘检测、角点识别、动态跟踪和立体匹配等方面都取得了很好的效果，但是其经常陷入局部最小状态，容易忽略微小特征，且计算时间长。数学形态学的基本思想是对复杂形状进行分解，从无用信息中提取有意义的形状分量，适用于噪声抑制、特征提取、边缘检测、图像分割等问题。该方法的基本形态操作是腐蚀与膨胀，腐蚀具有

收缩目标的能力，增加目标的内孔和消除外部隔离噪声的效果，而膨胀是增加目标和减小孔径，可以在目标中添加空间，使其形成一个连通域。但是，该方法经过图像处理，容易存在大量与目标不一致的短线和孤立点，导致运算速度显著下降。

目前已经有多种改进的传统图像处理技术应用于材料图像分割。Chen 等人 [15] 和 Liu 等人 [16] 改进了传统的分水岭方法，针对金相图像分割任务解决过分割问题，这两种方法均以传统的图像处理技术为基础，但更适用于边界清晰的图像分割任务。Han 等人 [17] 提出了一种混合算法，将高斯滤波、均值漂移法、泛洪填充法、改进的基于流量的高斯差分法和聚类算法相结合来解决金相组织图像分割问题，然而这些方法用于解决特定的材料结构取得了不错的分割结果，却很难适用于大多数复杂材料图像。Shao 等人 [18] 提出了一种改进的自适应加权均值滤波方法，先对图像中的脉冲噪声点进行检测，再用改进均值法对检测到的噪声点进行滤除，实现了金相图像中的噪点分割，然而该方法仅适用于需要噪点检测的图像。传统方法通常需要人工参与并且有较多的方法局限性，导致基于传统图像处理技术的分割方法难以适用于材料图像的自动化分割任务。

1.3.2 基于机器学习技术的分割方法

随着机器学习技术的发展和广泛应用，常用于解决图像分割任务的方法主要有 K 近邻 [19]、KMeans [20]、支持向量机 [21]、随机森林 [22] 和人工神经网络 [23] 等算法。 K 近邻算法 [19] 的基本思想是归并目标样本的 K 个最近样本的大多数，适用于划分类域内重叠较多的样本集，但因为每个待分类样本与所有已知样本之间的距离必须计算，才能得到其 K 个最近邻点，所以计算量大。KMeans 算法 [20] 的基本思想是将没有标注的多个样本聚集成指定个数的簇，适用于处理大数据集，可以达到不错的聚类效果，但是需要用户有一定的先验知识自定义参数，并且算法依赖于初始的聚类中心。支持向量机算法 [21] 的基本思想是求出一个最优的线性分类超平面，使支持向量点到最优分类超平面距离趋于最大，适用于样本数据线性可分且噪声较小的场景，但需要人工参与核函数的选择和参数调整。随机森林算法 [22] 的基本思想是通过多个不相同的决策树对随机选取的特征进行训练，该算法有着准

确率高、抗噪能力强、不易发生过拟合和适合处理高维数据的优点，但输出的结果是无法连续的，并且对于小数据或低维数据的处理结果不够稳定。神经网络算法 [23] 的基本思想是通过众多的神经元连接去处理和学习数据，该算法有着大规模并行处理、分布式信息存储和良好的自学习能力的优点，但是神经网络的“黑盒子”性质也被认为是其主要缺点之一。

目前已经有多种改进的机器学习技术应用于材料图像分割。Li 等人 [24] 利用在线支持向量机对复杂多变条件下的模型参数进行实时更新实现对金相图像的分割，该方法能够有效地提取局部和抽象特征，并且基于动态样本的实时更新模型具有最先进的分割性能，但是由于实时性和样本数量的条件约束，该方法不适用于样本多样的场景。Lai 等人 [25] 分别利用随机森林和复杂网络分割钛合金和陶瓷图像，但是复杂网络需要在大型图结构上运算，具有计算时间长的缺点。Zhu 等人 [26] 提出了一种基于粒子群优化的 K 均值聚类算法，根据粒子群适应度方差函数，选择结合粒子群优化和 KMeans 算法计算的插入点，实现对印章材料图像的分割和检测印章表面的缺陷，然而该方法适用场景单一且需要一定的专业知识。这些方法用于解决简单的材料结构取得了不错的分割结果，但对于大多复杂纹理的材料图像适用性不够广泛，导致基于机器学习技术的分割方法难以适用于材料图像的自动化分割任务。

1.3.3 基于深度学习技术的分割方法

深度学习是人工智能研究的基础技术之一，为计算机视觉带来了令人瞩目的改变。由于深度学习的快速发展，许多研究领域都大量地使用了深度学习技术，例如自动驾驶 [27]，人机交互 [28]，材料图像分割 [29] 等。

全卷积网络（Fully Convolutional Networks, FCN）[30] 的提出使得深度学习技术在语义分割任务上带来了巨大的成就。该方法阐释了网络模型从图像分类向语义分割上过渡的具体过程，利用经典的分类模型 [31, 32] 转化为全卷积网络，从而实现在语义分割任务上端对端的自动化训练。相比于传统图像分割方法和机器学习方法，FCN 在标准数据集如 PASCAL VOC [33] 等上的分割准确率取得了极大的进

步。由于上述及更多显著的贡献，FCN 成为了众多研究者参考和借鉴的网络原型。

除了 FCN 之外，还有 UNet [34]、SegNet [35]、Deeplab 系列 [36–38] 和 OCRNet [39] 等都是经典的应用于语义分割任务的网络模型。Ronneberger 等人 [34] 提出了一种应用于医学图像的网络模型 UNet，由下采样捕捉上下文信息去收缩路径，由上采样还原原始图像的细节和分辨率去扩张路径。SegNet [35] 通过一系列的上采样及卷积层组成，最终接上一个 Softmax 分类器来预测像素级别的标签，输出一张与输入图像大小相同的特征图。相比于 FCN，SegNet 引入了更多的连接，通过索引的方式提高存储效率。Deeplab v1 [36] 结合深度卷积神经网络 (DCNNs) 与概率图模型 (DenseCRFs) [40, 41] 提出了两点重大改进，第一，把控制感受野大小的方法化成空洞卷积 (Atrous Convolution)，第二，加入条件随机场 (CRF)，利用像素之间的关联信息。随后，Deeplab v2 [37] 在 v1 的基础上进一步做出了改进，引入了 ASPP (Atrous Spatial Pyramid Pooling) 的结构，ASPP 层就是通过选择不同扩张率空洞卷积去处理特征图，融合不同级别的语义信息。Deeplab v3 [38] 在 v2 基础上改进了 ASPP 模块，引入残差块，丢弃条件随机场，加入了批归一化和全局平均池化。不同于 ASPP 的融合方式，OCRNet [39] 提出了另一种获取上下文信息的方式，利用目标区域表示来增强像素的表示，根据粗糙的语义分割结果和网络深层输出的像素特征得到物体的上下文特征表示。这些研究都大力促进了语义分割方法的发展和进步。

小样本图像分割方法包括基于数据增强 [42]、模型微调 [43] 和迁移学习 [44] 等分割方法。其中数据增强通过增加样本多样性，微调模型通过对大数据集的预训练模型再对小样本数据集的参数进行微调，迁移学习通过旧知识学习新知识提升分割性能。Chen 等人 [45] 根据图像中各类别语义特征丰富的特点，利用标签语义空间和图像特征空间的相互映射对图像的特征进行增强。Nakamura 等人 [46] 提出一种微调模型，将较低的学习率用在小样本类别上再训练的过程中，将自适应梯度优化器用于微调阶段，当原数据集和目标数据集差异较大时，实现对整个网络的调整。Shaban 等人 [47] 通过侧分支的方式，以标注图像作为输入，辅助调整主分支的网络参数，对查询图像进行前景预测。Wang 等人 [48] 提出 PANet 从标签图像学习特

定类的原型表示，然后将每个像素与学习到的原型进行匹配，并对查询图像实施分割。Liu 等人 [49] 提出交叉参考机制，找到两张图片中共同出现的对象，由支持图像和查询图像互相引导完成分割。小样本图像分割方法对于小样本材料图像分割有着意义深远的启发和指导意义。

如今，深度学习的方法逐渐应用在材料图像分割的领域中，并取得了一定的进展。Shi 等人 [50] 提出了一种基于多注意力的编解码网络，利用钢筋材料固有的特点，在 U 型网络的解码路径中加入行均值注意模块，实现对钢筋材料图像的准确分割，然而该方法仅适用于钢筋及类似材料的图像。Huang 等人 [51] 针对热轧带钢材料的表面图像提出了一种深度可分的 U 形网络，采用深度可分离卷积代替传统的卷积层降低计算复杂度，使用多尺度模块提取不同尺度的上下文，提高了该种材料图像的分割性能，然而该方法增加卷积操作也会伴随着额外的特征信息损失。Li 等人 [52] 提出了一种基于超像素算法结合迁移学习的模型，利用在生物细胞数据上预训练的 UNet 实现金相图像中合金的精确分割，但是该算法不适用于边界和区域特征不明显的图像。Bertoldo 等人 [53] 提出了一种模块化 UNet (Modular UNet) 的解释方法，对三相玻璃纤维的三维断层微观成像图像进行分割。然而，他们主要是针对特定的小众应用场景设计网络模型。

基于深度学习技术的语义分割方法，在许多可视化任务中取得了良好的效果，但这些方法的缺点也很明显。第一，针对常见的卷积操作，每个像素点都需要取一个包含邻域的方块作为感受野，这就导致非常多的信息冗余。第二，分类准确率和定位精度不可兼得，当感受野过大时，卷积层的步长就需要增大，定位精度就会降低，感受野过小时，卷积层的步长就需要减小，分类精度会降低。第三，在样本数据少的前提下，网络无法学习到有用的特征信息，过多的卷积操作还会损失大量的像素和空间信息，分割精度难以提升。解决这些难题都能够有效提高深度学习方法在图像分割任务中的效率。

1.3.4 基于图卷积神经网络的分割方法

随着图卷积神经网络的快速发展，图卷积神经网络的方法也被逐渐改进以应用在图像分割任务 [54, 55]。

在语义分割任务中，由于在卷积过程中，特征图需要卷积核计算局部区域的特征信息，导致损失了很多局部特征的像素级信息和空间信息，从而使分割精度下降。为了使特征保留局部位置信息，研究者引入了图卷积神经网络使节点在逐层传播的过程中不会消失，缓解了局部特征的像素级特征和空间位置信息的损失。Lu 等人 [54] 提出 Graph-FCN 将图卷积应用到语义分割任务中，将语义分割任务转化为节点分类任务，利用卷积神经网络产生的特征图构建成图结构输入到图卷积神经网络。传统的语义分割方法是把图片当成一个图网络，然后切割图网络来实现语义分割。Graph-FCN 有更加灵活的跳跃连接，不像卷积网络中上一层卷积层的一个像素只连接下一层卷积层的像素，可以探索到图网络中不同节点间的关系。Graph-FCN 的消息传播机制在节点间交换信息，每个节点的感受野变得更大，并且所有节点都不会消失，保证局部位置信息不会丢失。该方法在自然场景的数据集上实现了 1.5 个百分点的提升。

在语义分割任务中，利用远程上下文信息也是语义分割等像素级预测任务的关键。与以往使用多尺度特征融合或扩展卷积的工作相比，Zhang 等人 [56] 提出的对偶图卷积网络 (DGCNet) 就是将处理不规则数据的图卷积融合到双重注意力模型中解决语义分割任务。对偶图卷积网络通过在一个单一的框架中建模两个正交图来建模输入特征的全局上下文。网络包含两个主要的组件，第一个组件是坐标空间的图卷积神经网络建模像素的空间关系，使得网络能够产生连续的预测并考虑图中的所有对象。第二个组件是特征空间用图卷积神经网络捕获更抽象的特征间的相关关系，例如物体的局部部分。推理后两个部分的特征映射到原始坐标空间并加到原始特征上，这是通过将特征投影到一个新的低维空间来有效地实现的，在这个空间中所有成对交互都可以建模，然后重新投影到原始空间。与强大的基线相比，该方法提供了实质性的好处，并且在城市景观等数据集上都取得了最先进的结果。

近年来,国内外许多研究机构和科研人员不断探索利用图卷积神经网络解决图像分割等多种计算机视觉任务,并取得很多优秀成果。Aizawa 等人 [57] 利用图卷积神经网络的思想设计了新的层次、上下文和多尺度的金字塔表示来捕获从输入图像的属性,其关键思想是基于预定义的区域数量和这些区域的递归分割和这些区域的上下文聚合,最后通过从递归聚合的上下文中构造金字塔表示,获得了多尺度和层次化的属性。Huang 等人 [58] 提出了一种新的边界感知半监督医学图像分割网络 Graph-BAS3Net,该网络结合了边界信息,学习了图域中语义和几何之间的对偶约束。具体而言,该方法由多任务学习框架 BAS3Net 和基于图的跨任务模块组成。BAS3Net 通过添加一个边界检测任务,编码了更丰富的对象形状和表面特征。基于图的跨任务模块进一步探索了语义分割和边界检测任务之间的共存关系,使网络从标记和未标记数据中学习更强的语义和几何对应。

图卷积神经网络在图像分割任务中的提出和拓展为广大研究学者开拓了探索的眼界,也为后续的多领域技术交叉和融合的研究方法提供了宝贵的思路。

为了解决深度学习技术中出现的普遍问题,本文将多种深度学习技术结合图卷积神经网络,提出了两种应用于材料图像分割的方法。在第一个方法中,本文提出了基于图卷积的多维特征融合的网络模型,以特征图为主、图为辅展开一系列特征学习和提取,结合图卷积和以卷积操作为主的双重注意力机制增加对重要特征的学习,利用残差连接、多尺度融合和加深网络深度等方法实现多层次、多维度的特征信息的丰富。在第二个方法中,本文提出了基于跳跃连接的 Graph-UNet,在没有引入任何卷积操作的前提下针对图展开一系列特征学习和提取,有效结合了多层图卷积和图注意力,在增加节点注意力的同时大大缓解过多卷积操作带来的空间信息、像素信息的损失,利用跳跃连接融合节点的多维特征,增强对特征的学习能力。

1.4 论文主要工作

本论文是以作者攻读硕士学位期间承担的课题工作为基础,针对材料图像的特点,研究多种深度学习技术相结合的小样本材料图像自动化分割方法。本文的组织结构如下:

第二章介绍了本研究相关的理论和技术概况。首先描述了全卷积神经网络以及相关组成结构；接着介绍与本文相关的两种经典卷积网络结构，即 FCN 和 UNet；然后说明了与本文工作相关的技术方法，着重对图卷积做了详细的描述；最后罗列了常用的评估度量方法。

第三章针对材料图像的小样本、纹理复杂和色彩单调等特性，提出了基于图卷积与深度学习的编解码网络模型。首先描述了模型的框架和介绍了各个模块；然后说明了实验环境和材料数据集；最后进行了对比实验和消融实验的验证，同时对实验结果进行了充分的说明和分析。

第四章针对第三章方法存在的缺陷，提出了基于跳跃连接的 Graph-UNet。首先阐述了模型框架的设计原理；然后详细介绍了图编码、图注意力模块和图解码器；最后进行了对比实验和消融实验的说明，对实验结果进行了说明和分析。

第五章总结和展望。本章对本文所提出的两种方法做出了总结，并展望了未来的研究内容。

第二章 相关理论和技术概述

随着深度学习的迅速发展，越来越多的网络模型在计算机视觉等一系列领域都取得了突破性的成就。本文的研究目标是结合现有技术思想和优点，设计和探索更有效的深度学习方法实现小样本材料图像的自动化分割。因此本章回顾了与本工作相关的深度学习技术，为后续章节的方法和实验提供理论和技术基础。

2.1 全卷积神经网络的基本概念

语义分割可以理解为像素级分类任务，将不同类别的标签分配给图像中的所有像素，并区分不同标签的区域。全卷积神经网络首次将分类网络转化成用于分割任务的网络结构应用在语义分割任务中，输入是正常的图片，经过一系列的卷积、池化和反卷积等操作，对每一个像素进行类别预测，如图2.1所示。网络通常包含卷积层、池化层、归一化层、激活层和反卷积层等。

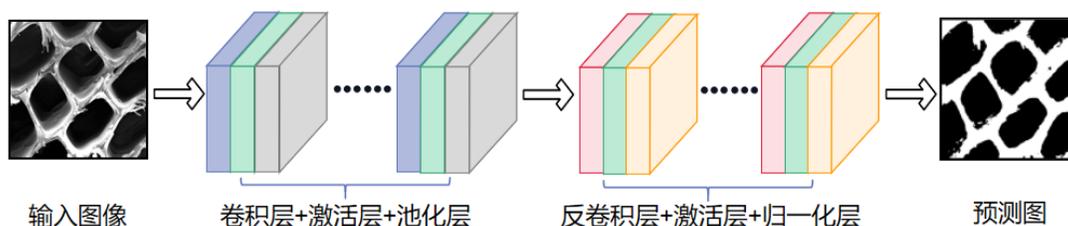


图 2.1: 全卷积神经网络的结构。输入图像经过多个卷积层、激活层和池化层提取到高级特征图，再经过反卷积层、激活层和归一化层得到与输入图像相同大小的预测图。

卷积层的运算是两个矩阵进行计算的过程。两个矩阵分别是图像的特征图所生成的矩阵和用于移动计算的卷积核矩阵。如图2.2所示，卷积操作是全卷积神经网络中卷积层最基础的运算，左边是输入图像生成的矩阵，中间是卷积核，右边是输出矩阵。卷积核在输入矩阵上以一定的步幅移动，在卷积核的大小范围内进行点积运算，得到最终的输出矩阵。卷积操作通过一个范围的局部特征进行卷积整合，可以学习到更高层和更抽象的语义特征，不仅减少了网络的运行时间成本和空间存储成本，同时也保留了大部分的空间信息和特征信息。

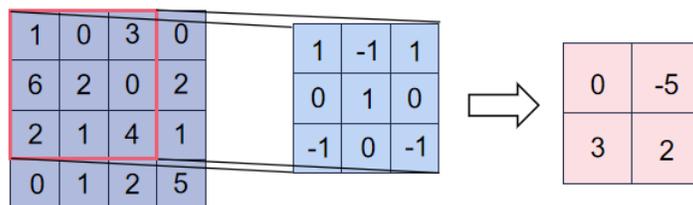


图 2.2: 卷积操作示意图。左边是输入图像生成的矩阵, 中间是卷积核, 右边是输出矩阵。卷积核在输入矩阵上以一定的步幅移动, 在卷积核的大小范围内进行点积运算, 得到输出矩阵。

池化层通常放置在连续的卷积层之间, 以压缩矩阵减少参数量, 降低过拟合。如果输入的数据是图像, 那么池化层最主要的作用就是缩小图像尺寸, 其具体操作与卷积层的操作方式基本相同, 池化层的卷积核是只取对应位置的最大值或平均值等(最大池化和平均池化), 不同的池化方法对矩阵的运算规律不一样。最大池化和平均池化的实现原理如图2.3所示。

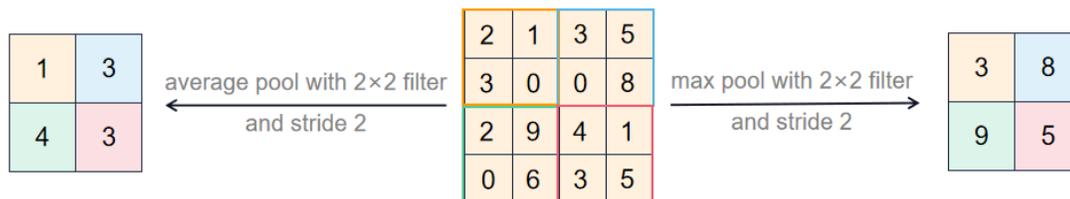


图 2.3: 最大池化和平均池化的实现原理示意图。池化操作通过取对应位置的最大值或平均值等(最大池化和平均池化)方式缩小图像尺寸。

归一层是将输入的各个维度归一化为标准正态分布, 通常放在卷积层之后。深层神经网络中, 中间某一层的输入是前一层的神经层的输出, 因此, 其之前的神经层的参数变化会导致当前层输入的分布有很大的差异。归一层的加入使得每一个神经层的输入的分布在训练过程保持一致, 很好的避免了梯度弥散和爆炸等问题, 在遇到 Sigmoid 或者 Tanh 等激活函数时, 不会使数据落在饱和区域导致梯度弥散, 并且归一化降低了权重初始化的困难。

激活层是对输入数据进行函数变换。在 AlexNet [59] 之前, 神经网络往往采用 Sigmoid 作为激活函数, 然而饱和激活函数存在梯度消失的缺陷, 严重限制了神经网络的深度。非饱和激活函数 ReLU 在一定程度上缓解了梯度消失问题, 使得神经

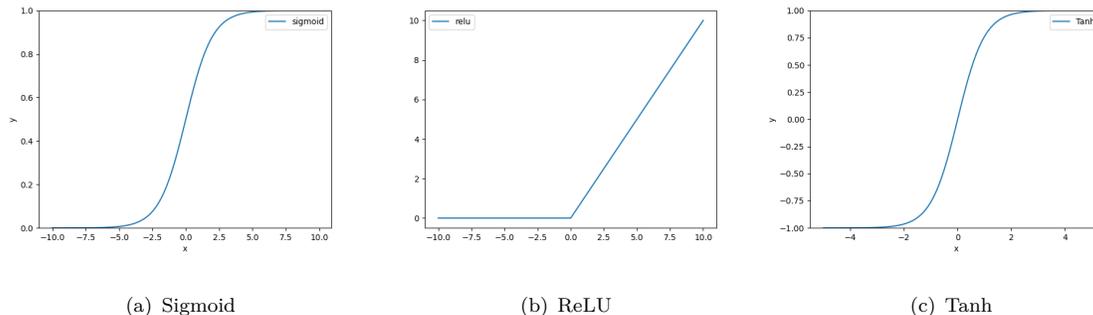


图 2.4: 常见的激活函数可视化。(a) 是 Sigmoid 函数, (b) 是 ReLU 函数, (c) 是 Tanh 函数。

网络可以变得更深。常见的激活函数有 Sigmoid、ReLU 和 Tanh 等, 具体的函数曲线如图 2.4 所示, 一般 Sigmoid 更多用于二分类预测, Softmax 更多用于多分类预测, ReLU 和 Tanh 则是更多的紧跟着卷积层配合使用, 用于增强网络的非线性表达 [60]。

反卷积层是卷积的逆过程。反卷积也是一种需要学习权重的卷积, 但是缺点就是需要在行列进行补零, 这使得该操作不是一个有效的计算接口。反卷积层的作用是增加输入的维数, 即输出维数大于等于输入维数, 反卷积的具体过程如图 2.5 所示。虽然反卷积的本意是还原输入, 但实际上是无法完全还原输入的特征。

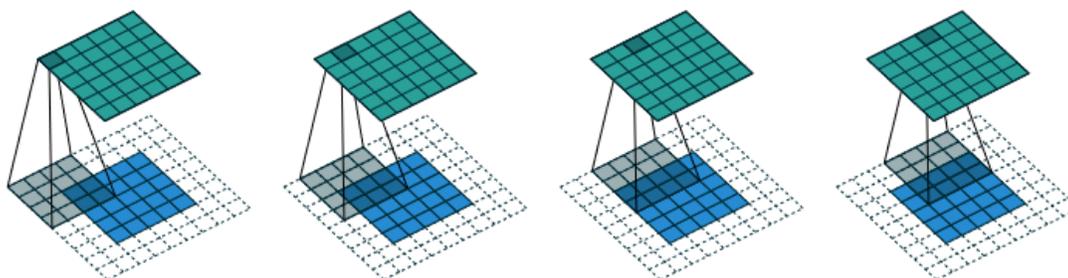


图 2.5: 反卷积过程示意图。反卷积通过在行列补零扩大特征图尺寸 [61]。

传统的基于卷积神经网络的分割方法通常对像素进行分类, 并使用像素周围的局部图像块作为网络的输入进行训练和预测。这种方法存在一些弊端, 第一, 存储开销很大, 第二, 计算效率低下, 冗余信息过多, 第三, 像素块的大小限制了感受

野的大小。

针对这三个问题，研究学者设计了用于图像分割的全卷积神经网络，试图从抽象特征中恢复每个像素的类别。随着图像分类进一步扩展到像素分类，全卷积神经网络对语义分割任务的后续发展做出了重要贡献。

2.2 本文相关的卷积神经网络框架

随着语义分割任务中深度学习技术的兴起，计算机领域出现了很多优秀的网络结构。FCN [30]、UNet [34]、SegNet [35] 和 PSPNet [62] 等都是经典的网络框架，拥有优越的网络性能，大多研究者都受到这些网络的启发构建自己的分割框架。本章详细介绍了与本文研究相关的两种卷积网络结构，即 FCN 和 UNet。

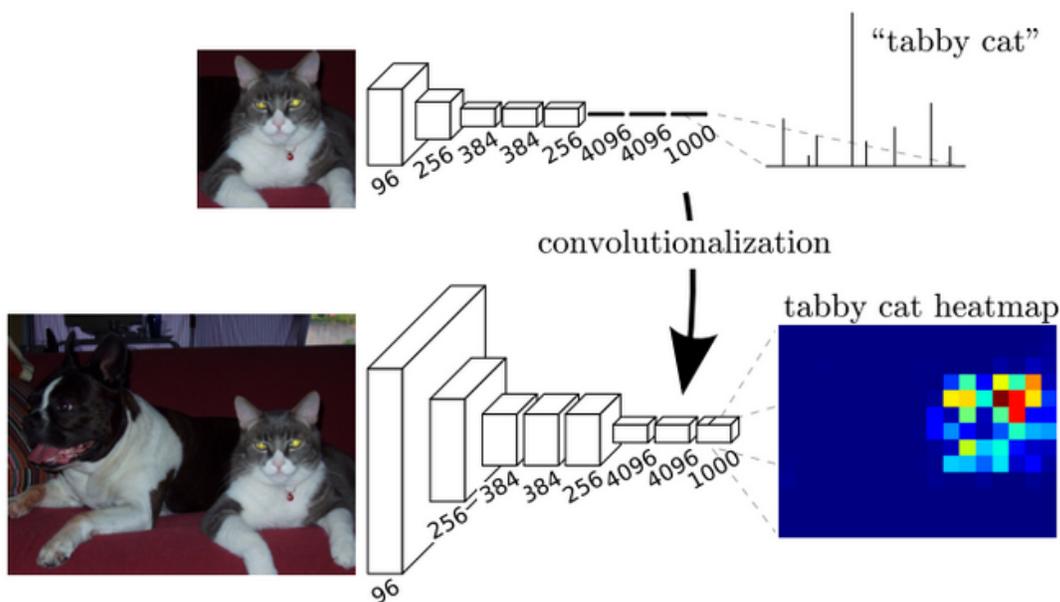


图 2.6: FCN 网络结构与传统卷积神经网络的区别。图中上方网络模型为传统卷积神经网络，包含了全连接层；图中下方为全卷积神经网络，不包含全连接层 [30]。

FCN [30] 用卷积层代替全连接层，恢复抽象特征图中每个像素的类别，如图2.6所示。在传统的卷积神经网络结构中，第 1 至 5 层为卷积层，第 6 层和第 7 层分别是特定长度的一维向量，第 8 层是以分类类别为长度的一维向量，对应各个类别的概率。FCN 将后 3 层全连接层替代为卷积层，卷积核的大小分别为 $(L, 1, 1)$ 、

$(L, 1, 1)$ 和 $(N, 1, 1)$ ，其中卷积核的第一维 L 和 N 表示通道数，第二维数值表示卷积核的宽，最后一维数值表示卷积核的高。

由于 FCN [30] 经过 5 次卷积和池化以后，图像的分辨率依次降低到原图像的 2、4、8、16 和 32 倍，分辨率出现越来越低的现象。为了从低分辨率的图像恢复原始图像的分辨率，FCN 使用上采样操作，对卷积的特征图进行 32 倍的上采样，最终得到与输入图像相同的尺寸。

相比于传统的卷积神经网络进行图像分割的方法，FCN [30] 具有两个明显的优势：一是可以输入任意大小的图像，并且训练图像不需要与测试图像有相同的大小；二是更加高效，因为避免了使用像素块带来的卷积重复存储和计算的问题。同时 FCN 也有明显的缺点。第一，网络最终输出的图像不够精细，虽然倍数少的网络上采样效果要比倍数多的网络好得多，但上采样结果仍然丢失了大量的特征，导致图像模糊、平滑；第二，网络忽略了上下文关系，全卷积神经网络对每个像素进行分类，只关注局部特征，而忽略全局背景下像素之间的关系，缺乏空间一致性。

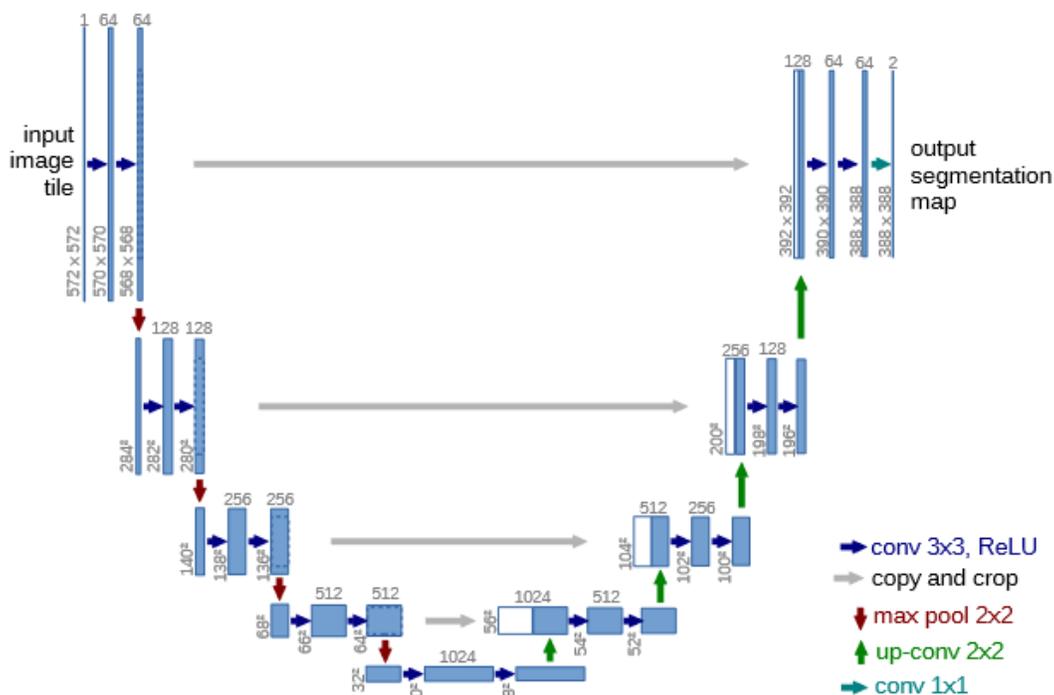


图 2.7: UNet 网络结构示意图。UNet 是 U 型结构，左边部分是编码器，右边部分是解码器 [34]。

UNet [34] 紧接着 FCN 之后被提出,为解决语义分割任务提供新的思路。UNet 刚出现时应用在生物医学图像上,主要处理样本规模偏小的灰度图像。UNet 是 U 型结构,该结构即使使用很少的训练图片,分割的准确度也保持很高的水平,UNet 的网络结构如图 2.7 所示。UNet 的编码器部分共需要下采样 4 次,输入图像的大小会减少到原来的 16 倍,相应的,解码器部分也需要进行 4 次上采样,将编码器中得到的高级语义特征图恢复到原始图像的分辨率。

相比于 FCN, UNet [34] 共进行了 4 次上采样,并在同一个分辨率的特征图中使用了跳跃连接,而不是直接监督高级语义特征和反向传播损失,这确保了最终恢复的特征包含了更多的低级特征,也使得不同尺度的特征得到了融合,从而可以进行多尺度预测和深层监督。4 次上采样和跳跃连接使分割图的边缘信息等的恢复更加细化。

UNet [34] 已经被广泛地应用到医学图像分割上,然而医学图像语义较为简单,结构较为固定,存在明显的缺点。第一,由于每个像素点都需要取一个包含邻域的方块,这会导致大量冗余;第二,分类准确率和定位精度不可兼得,当感受野过大时,池化层的降维倍数就会增大,定位精度就会降低,感受野过小时,分类精度会降低;第三,网络结构较为简单,限制了网络的非线性表达能力,不适用于大规模和类别过多的样本数据。

2.3 经典的基于语义分割的深度学习方法

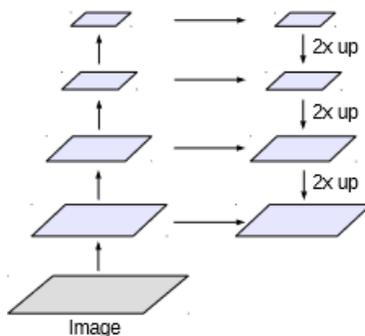


图 2.8: 编码器和解码器的结构示意图。图中左边是编码器,将图像转化为高维语义特征图。右边是解码器,将高维语义特征图转化为像素级别分类的预测图 [38]。

在语义分割领域中，通常会存在很多经典的深度学习方法去实现对图像的精准分割，例如编码器和解码器 [63]、多尺度融合 [64] 和注意力机制 [65] 等。因为卷积层的堆叠存在大量卷积运算，图像中的信息不可避免的出现粗糙化导致很多细节信息的损失。因此，为了提升网络模型的分割性能，许多科学研究者提出了各式各样的方法和思路去降低由卷积而导致的特征损失 [66–68]。

首先，在网络中使用编码器和解码器是一个主流的方法 [69–72]。编码器和解码器的结构如图2.8所示。编码器-解码器一开始不是应用在语义分割中，自动编码器用于特征提取和数据降维。现在针对密集预测，编码器用于将图像转化为高维语义特征图，解码器用于将高维语义特征图转化为像素级别分类的预测图。经典的 UNet 网络就是通过 U 型结构实现编码器和解码器的操作。紧接着，UNet++ [69]、UNet3+ [70]、EfficientFCN [71] 和 Attention-UNet [72] 等网络模型被提出，都是先编码（下采样），再解码（上采样），将特征图还原为与原始图像相同的大小，再进行像素分类。DEDNet（Double Encoder-Decoder Networks）[73] 同样是编码器和解码器的思想，该框架按顺序堆叠两个编码器-解码器网络，其中第一个网络进行初始的预测，第二个网络作为一种注意力机制，专注于图像中重要的区域，从而提高分割性能。

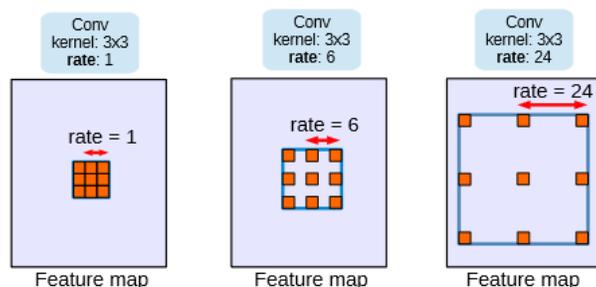


图 2.9: 核大小为 3×3 ，不同空洞率值的空洞卷积感受野范围。采用不同的空洞率值（分别为 1、6 和 24）改变了卷积核的感受野大小 [38]。

空洞卷积是一种用于密集预测的卷积方法 [74–76]。空洞卷积的出现是因为在图像分割中需要保证特征图尺寸不变和不引入额外参数的情况下增加感受野范围，具体实现方式如图2.9所示。CENet [77] 加入了空洞卷积，得到了在不损失特征图分辨率的情况下成倍地扩展感受野的好处。Zhao 等人 [78] 提出 D2A-UNet，在模型解码器中引入了混合扩张卷积，并得到了卷积输出包含更大范围信息和优化解码过程

的好处。空洞卷积很好的避免了 FCN 的感受野局部受限的缺点，在解决图像需要全局信息和语音文本对信息依赖性较长的问题时，空洞卷积是被优先考虑的一种方法 [79,80]。但是空洞卷积存在一定的缺陷，在空洞率值越来越大的同时，卷积核中的不连续会造成卷积核中的参数不能全部用来计算，降低特征提取的效果。

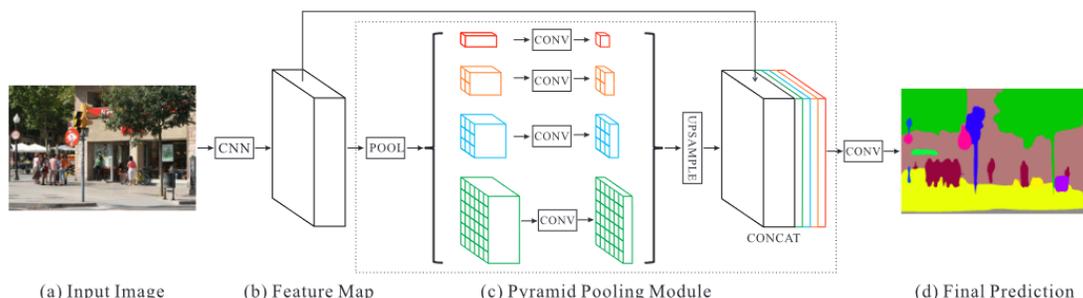


图 2.10: 经典的多尺度网络之一: PSPNet 的网络结构。图中 (a) 是输入的图片; 图中 (b) 是由卷积操作生成的特征图; 图中 (c) 是金字塔解析模块获取不同的子区域表示, 然后上采样和串联层形成最终的特征表示, (c) 中同时携带局部和全局上下文信息; 图中 (d) 是最终的预测结果图 [62]。

多尺度特征的融合是增强特征学习和提高分割性能的一个重要手段 [81]。低层特征具有分辨率高和特征信息丰富的特点, 但是参与卷积运算少、语义低和噪声多容易影响最终的分割结果; 高层特征具有分辨率低和特征信息抽象的特点, 但是参与卷积运算多、语义性高和噪声少容易导致分割结果粗糙。将低级特征和高级特征有效融合是改善分割模型和提高网络性能的关键。PSPNet [62] 是一个经典的多尺度融合网络, 将高分辨率的特征图与低分辨率的特征图相结合, 局部和全局的特征共同作用以使最终的预测更加准确, 具体网络结构如图2.10所示。Zhu 等人 [82] 利用金字塔纹理特征提取模块, 得到了一组具有不同感受野大小的特征图, 也可有效地从多个尺度上提取和统计纹理特征再去融合输出结果。Tao 等人 [83] 提出了一种基于多尺度关注特征图的语义分割网络, 利用平均池化或最大池化来得到多种尺度图像的方式提高网络性能, 并在 Camvid 数据集 [84] 上得到了验证。Mohammed 等人 [85] 提出了一种上下文多尺度多层次网络 (CMM-Net), 通过在解码器部分的每个收缩卷积网络层融合多个空间尺度的全局上下文特征, 从而获得更丰富的特征信息以提高网络的分割性能。

残差连接作为一种经典方法，可以通过加深网络深度提高网络的非线性表达能力。在 VGG [31] 中，卷积网络深度达到了 19 层，在 GoogLeNet [86] 中，网络达到了 22 层。网络层数的增加通常伴随着几个问题，一是计算机资源的消耗；二是容易出现过拟合，这是网络非线性表达能力提升所带来的弊端；三是会产生梯度消失和梯度爆炸的问题。为了缓解反向传播过程中的梯度消散问题，使深度网络易于训练，He 等人 [32] 提出了残差网络，详细的残差块结构如图 2.11 所示。Gudhe 等人 [87] 提出了一种新的多级扩展残差神经网络，在网络的跳跃连接中加入一个非线性的多层次残差块，以减少语义差距，并恢复在编码器到解码器单元连接特征时丢失的信息。

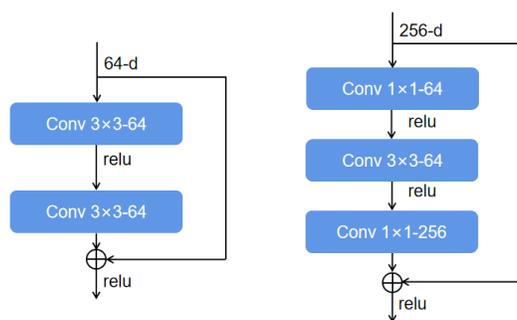


图 2.11: 残差网络两种构建块结构的示意图。左边的残差结构是输入特征图经过两层卷积核为 3×3 、通道数为 64 的卷积层后的输出加和原始输入特征图；右边的残差结构是输入特征图经过两层卷积核分别为 1×1 和 3×3 、通道数都为 64 的卷积层，再经过一层卷积核为 1×1 、通道数为 256 的卷积层的输出加和原始输入特征图。

注意力机制是一种常见的提升网络性能的方法，可以选择性地关注特征图的一部分。注意力模型最初被用于机器翻译，现在已成为神经网络领域的一个重要概念，该方法的基本思想是使系统忽略不相关的信息，关注重要的信息，主要分为通道注意力和空间注意力。SENet [88] 通过对通道域上的特征处理，用 Sigmoid 函数在通道维度上生成 0-1 之间的注意力权重增强对重要特征的关注度。Non-local Neural Networks [89] 通过计算任意两个位置之间的交互，从全局的角度考虑，捕捉远距离特征间的关联，而不是只考虑相近点。CBAM (Convolutional Block Attention Module) [90] 进一步改进了 SENet，通过通道注意力模块和空间注意力模块提高两个维度的关注度。DANet [91] 融合了 CBAM 和 Non-local 两种方法并做了改进，将空间维度和

通道维度的自我注意应用于高层特征图，最后将两个结果进行融合。相比于 DANet 中注意力图计算所有像素之间的相似性，CCNet [92] 巧妙的减少了参数量，只计算每个像素与其所在像素在交叉方向上的相似度，通过循环间接计算到每个像素与其他像素间的相似性，将空间复杂度由 $(H \times W) \times (H \times W)$ 降为 $(H \times W) \times (H + W - 1)$ 。极化自注意力机制 (Polarized Self-Attention, PSA) [93] 用于解决像素级的回归任务，在设计上主要有两个改进点：1) 极化滤波：在通道和空间维度保持比较高的分辨率，缓解降维造成的信息损失问题；2) 非线性增强：采用细粒度回归输出分布的 Sigmoid 非线性函数，PSA 的结构如图2.12所示。

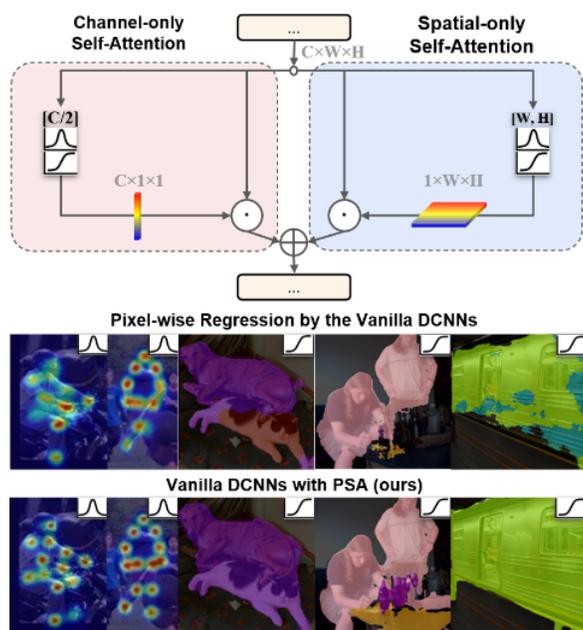


图 2.12: 经典的注意力模块之一：极化注意力的结构、模块的人关键点热图和语义分割标注图。极化注意力的结构包括通道自注意力和空间自注意力 [93]。

2.4 图卷积

2.4.1 应用于图结构的图卷积

实际应用场景中，除了解决规则化数据的方法，还存在许多应用于不规则化数据的方法。常见的不规则数据的应用场景有网络分析 [94]、推荐系统 [95]、生物化

学 [96] 和交通预测 [97] 等。解决这一类任务的方法和图嵌入密切相关，属于表示学习的范畴。

图嵌入的方式主要有三种，即矩阵分解 [98]、随机游走 (DeepWalk) [99] 和图神经网络 (Graph Neural Network, GNN) [100]。基于矩阵分解的方法是将节点之间的关系用矩阵去描述，然后分解矩阵得到嵌入向量，如邻接矩阵、拉普拉斯矩阵等。随机游走是基于 Word2vec [101] 词嵌入提出来的，以节点为单词，以随机游走得到的节点序列作为句子，然后输入至 Word2vec，同时利用节点的嵌入表示作为下游任务的初始化参数可以优化下游任务的效果。基于图神经网络的方法则是结合图数据的特点，采用深度学习的方法构建网络，处理与图结构相关的任务。

其中，图神经网络可以细分为图卷积网络 (Graph Convolution Networks, GCN) [102]、图注意力网络 (Graph Attention Networks, GAT) [103]、图自编码器 (Graph Auto-Encoders) [104]、图生成网络 (Graph Generative Networks) [105] 和图时空网络 (Graph Spatial-Temporal Networks) [106] 等。图卷积网络是将卷积运算从规则化数据推广到不规则的图数据，其核心思想是通过函数映射学习节点特征。图注意力网络是基于空间的图卷积网络，利用注意机制聚集特征信息，同时确定节点之间的权值。图自动编码器是利用神经网络结构将图的节点表示为低维向量。图生成网络是在给定一组已知图数据生成新的图。图时空网络是计算时空图的时空相关性，适合解决交通网络或预测未来的节点标签等场景任务。

图卷积 [102] 是众多复杂图神经网络模型的基础，可以分为基于谱 [107] 和基于空间 [108] 两大类方法。基于谱的方法从图信号处理的角度定义图卷积，其中图卷积是通过引入滤波器，从图信号中摒弃噪声的干扰。基于空间的方法是将图卷积定义为来自邻居的特征信息的聚合，当在节点层面执行图卷积网络算法时，图池化模块可以与图卷积层交叠，将图转换为粗略的高层表示，适合于抽取各级图表示和对图进行分类等任务。

图卷积处理的数据是图。对于图 $G = (V, E)$ ， V 是节点的集合， E 是边的集合，对于每个节点 i ，均有其特征 x_i ，也可以用矩阵 $X_{N \times M}$ 表示，其中 N 表示节点数量， M 表示每个节点的特征数，即特征向量的维度。

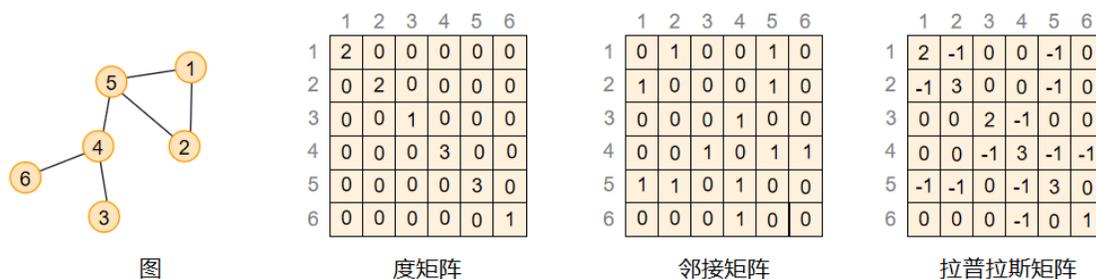


图 2.13: 一个简单图的度矩阵、邻接矩阵和拉普拉斯矩阵。

图的描述方式一般有度矩阵、邻接矩阵和拉普拉斯矩阵等，如图2.13所示。度矩阵是描述每个节点与其他节点之间的连接密度，邻接矩阵和拉普拉斯矩阵用于度量节点和邻居节点之间的关系，其中邻接矩阵用于表示节点之间的连接关系。拉普拉斯矩阵有不同的表达方式，常用的一种运算方式是将度矩阵减去邻接矩阵，如图2.13所示，拉普拉斯矩阵在对角线上数值表示了每个节点与其他节点连接的疏密程度，数值越大说明越多的节点需要与中心节点交换特征信息，在非对角线上数值表示了节点间是否存在消息传播。

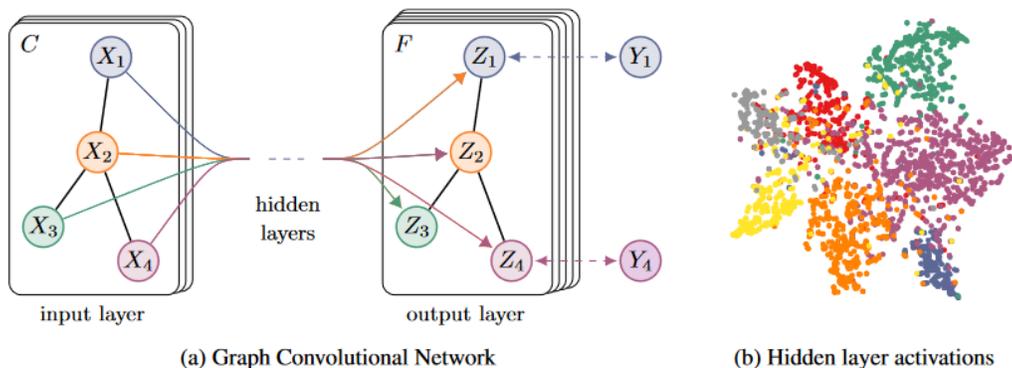


图 2.14: 基于维度为 C 的输入通道和维度为 F 的特征映射的多层图卷积网络 (GCN) 的半监督学习原理描述图。图的结构 (用黑线表示的边) 在图层上共享，标签用 Y_i 表示。右边是使用部分的标签，在 Cora 数据集上训练两层 GCN 的隐藏层激活的可视化，颜色表示文档类 [102]。

图卷积 [102] 依靠拉普拉斯矩阵的实现公式如下：

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}) \tag{2.1}$$

其中 $H^{(l)}$ 是第 $l+1$ 层的输入, D 和 A 分别是图的度矩阵和邻接矩阵。 \tilde{D} 和 \tilde{A} 是指图的每个节点添加了自连接, 并构建了相应的度矩阵和邻接矩阵。 $W^{(l)}$ 是第 l 层图卷积的权值矩阵。 σ 是激活函数。 $H^{(l+1)}$ 是第 $l+1$ 层图卷积操作的输出, 图卷积神经网络的示意图如图2.14所示。

图卷积在解决不规则化数据的相关任务上有着巨大的优势。Yuan 等人 [109] 研究基于细粒度交通数据的深度学习框架, 根据道路现状预测各路段近期的洪水状态。该研究设计并实现了 3 个时空图卷积网络 (STGCN) 模型, 以预测洪水事件期间美国德克萨斯州哈里斯县哈维飓风背景下路段级的路网状态。Zeng 等人 [110] 研究了基于图的方法来提取句子之间的关系, 通过迭代地结合来自邻居节点的信息, 对存在于不同句子中的实体提及的交互进行建模。该方法从非结构化文本中自动提取化学诱导疾病关系对疾病治疗和药物开发具有重要意义。Fu 等人 [111] 提出了一种新的多视图图卷积网络 (MVGCN), 采用自监督学习策略获取节点属性作为初始嵌入, 通过在每个视图中聚合来自域间和域内邻域的信息来迭代更新节点的嵌入, 最后集成多个视图以获得最终的节点嵌入, 该模型对于预测药物-疾病相关性、药物-靶标相互作用等领域做出了积极的推动。

2.4.2 图卷积和卷积的区别

图卷积在图卷积神经网络中的意义相当于卷积对于卷积神经网络。对于数字图像来说, 图像是一个二维数据, 图像的卷积运算实际上是将卷积核在表示图像的矩阵上滑动, 计算卷积核中间像素对应的像素值, 最后滑动整个图像的过程。图像的每个像素都有一个固定的规则邻接结构, 卷积核也是一个共享矩阵用来处理整个图像。初始随机的公共卷积核用来计算像素的加权和, 从而提取压缩特征, 然后通过反向传播学习卷积核的参数来获得最优参数去提取特征, 这是卷积神经网络的一个基础操作。

对于不规则结构的数据, 例如社交网络信息、知识图谱、蛋白质网络、万维网等, 与图像领域的卷积相似, 图卷积需要一个通用的范式来提取图的特征, 这就是图卷积在图卷积神经网络中的意义。图卷积的方式是根据图的特性来定义生成的, 图有

两个基本的特性。第一，每个节点都有自己的特征信息；第二，图的每个节点还具有不同的结构信息。图中的每个节点会一直被相连的邻居和其他有关联的点影响而改变自己的状态持续到全局平衡，越近的邻居影响越大，越远的节点影响越小。

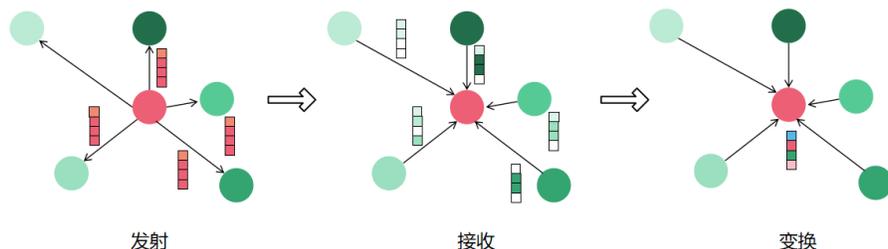


图 2.15: 图卷积学习局部特征的过程示意图。

根据图的特性，图卷积的设计是通过三个步骤去学习局部的特征信息，流程示意图如图2.15所示。第一步，每个节点变换自己的特征信息并将其发送给相邻节点；第二步，每个节点通过边的信息流收集其邻居的特征，这一步是融合节点的局部结构信息；第三步，收集信息后进行非线性变换，以提高模型的表达能力。

在解决图结构数据的任务中，图卷积神经网络以端到端的方式同时学习节点特征信息和结构信息，是目前对图数据学习任务的最佳选择。卷积神经网络无法解决节点分类与边预测等任务，然而图卷积神经网络可以轻松解决这类问题，在公开数据集上效果要远远优于卷积神经网络。

随着模型的融汇互通，各个特定领域的模型也被逐渐泛化，图卷积已经趋向建立通用的模型应用到各个领域。Ma 等人 [112] 提出了一种新的注意图卷积网络来对遥感图像数据进行超像素分割，模型将超像素块构建成图结构，引导图卷积层将注意力集中在最相关的节点上，从而做出决策，如图2.16所示。Yang 等人 [113] 提出了一种用于手绘草图语义分割和标注的图卷积神经网络 SketchGCN，将输入草图视为二维点集，并将笔划结构信息编码为图形节点/边缘表示，利用图卷积和全局分支网络结构提取笔画内部和笔画之间的特征。Shi 等人 [114] 将身体部位特征编码到一个基于人的时空图中，并使用一个轻量级的图卷积模块来显式建模身体部位之间的依赖关系，以识别动态的人体行为。

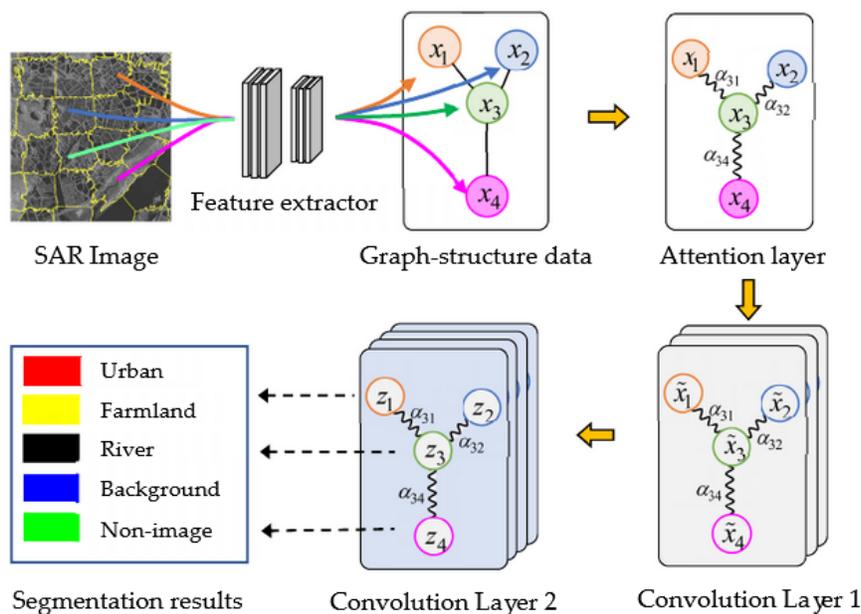


图 2.16: 基于超像素的图注意力网络应用于遥感图像。本模型使用超像素分割方法处理输入图像, 通过特征提取器构建图结构的数据, 将数据放入图注意力层和图卷积层从而做出决策 [112]。

2.4.3 应用于语义分割的图卷积

近年来, 图卷积网络不仅在解决图数据任务上取得巨大的成就, 还在语义分割任务中逐渐占据重要的地位。Chen 等人 [115] 提出了基于图的全局推理网络, 该方法将一组特征在坐标空间中全局聚集, 然后映射到相互作用空间中的节点上, 形成一个全连通图, 对图进行图卷积实现关系推理后, 再将节点特征投影回坐标空间完成下游任务, 网络结构如图2.17所示。Wu 等人 [116] 提出图交互单元 (Graph Interaction Unit) 和语义上下文损失 (Semantic Context Loss) 去探索如何结合语言知识来促进图像区域的上下文推理, 该方法首先将基于数据集的语言知识引入到视觉图单元中, 以促进视觉图上的上下文推理, 然后将视觉图的进化表示映射到每个局部表示, 以增强场景解析的区分能力。He 等人 [117] 提出一种边界感知图卷积模块在目标物体内部传播特征, 以粗分割特征图作为节点特征和边界预测来指导图的构建, 经过图卷积后, 将推导出的特征与输入特征进行融合, 得到更精细的特征, 输出更准确的分割结果。Hu 等人 [118] 提出了一种基于类的动态图卷积模块实现信息的自适应传输,

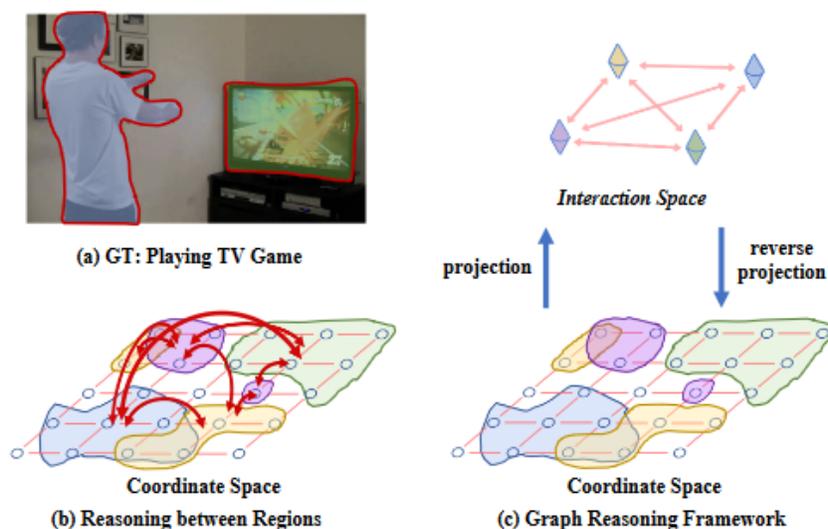


图 2.17: 基于图的全局推理网络的结构示意图。(a) 是“玩电视游戏”的标注图, (b) 是不同区域间的推理过程, (c) 是图推理框架。本方法在获取整个输入空间上任意区域之间的关系 (用不同的颜色表示), 提出了一种新的全局推理方法 (如图 (c) 所示), 将坐标空间中的有色区域的特征投影到交互空间的节点中, 形成一个全连通图。对图进行推理后, 将节点特征投影回坐标空间 [115]。

设计了同类像素之间的图推理, 实现语义分割从粗糙到精细的过程。Pan 等人 [119] 解决了基于图像级类标签的弱监督图像语义分割, 克服了随机游走的前馈特性不能正则化所生成的完全伪标签的质量的问题, 将完全伪标签的生成看作是一个半监督学习任务, 通过反向传播拉普拉斯算子和熵正则化损失分别对每幅训练图像学习两层图卷积。

在使用图卷积方法解决图像处理任务中, 最关键的一点是如何将图像或者特征图转化为图卷积网络可以处理和操作的图数据。在手绘草图预测任务中, 研究者会将草图视为二维点集, 使用节点和边来描绘点集的结构从而预测草图的标签; 在行为识别任务中, 研究者会将人体的骨骼转化为节点去分析由节点构成的图如何变化从而判别行为的标签; 在语义分割任务中, 研究者将图像以像素为节点或者以区域为节点等思路用自定义的连接方式构建图从而实现分割的目的。如何将任务中的样本数据转化为图结构是图卷积应用于各类场景的重点和难点。总之, 将图卷积应用于语义分割任务为研究者们打开了跨领域交叉和融合的新思路, 为迎接其他的分割任务挑战贡献了举足轻重的力量。

2.5 常见的图像分割评价指标

深度学习在图像语义分割上已经取得了不凡的进步，产生了许多应用于语义分割的模型与基准数据集，这些基准数据集提供了一套标准的评价指标。在大多数情况下，研究者从算法精度等方面来评价模型的性能。

精度是评价图像分割网络最主要也是最流行的技术指标，这些精度估算方法各种不同，但是主要可以分为两类，一类是基于像素精度，另外一类是基于 IoU。

评价指标中 True Positives (TP)、True Negatives (TN)、False Negatives (FN) 和 False Positive (FP) 都是必备的条件。TP 是被模型预测为正的正样本，TN 是被模型预测为负的负样本，FP 是被模型预测为正的负样本，FN 是被模型预测为负的正样本。

假定一定有 $k + 1$ 类（包括 k 个目标类和 1 个背景类）， p_{ij} 表示原属于 i 类却预测为 j 类的像素点总数，具体地， p_{ii} 表示 TP， p_{ij} 表示 FP， p_{ji} 表示 FN。

像素准确率 (Pixel Accuracy, PA) 是指正确分类的像素点数和所有的像素点数的比例。具体计算公式如下：

$$PA = \frac{\sum_{i=0}^k p_{ii}}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}} \quad (2.2)$$

平均像素准确率 (Mean Pixel Accuracy, MPA) 是指计算每一类正确分类的像素点数和该类的所有像素点数的比例求平均。具体计算公式如下：

$$MPA = \frac{1}{k + 1} \frac{\sum_{i=0}^k p_{ii}}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}} \quad (2.3)$$

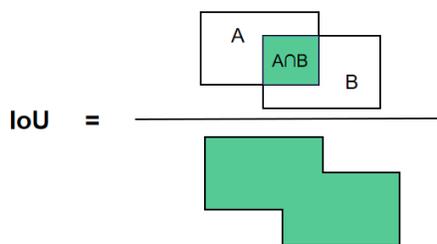


图 2.18: IoU 计算的示意图。

平均交并比 (Mean Intersection over Union, MIoU) 是计算每一类的 IoU 然后求平均。一类的 IoU 计算方式如下, p_{ii} 表示 TP, 即原属于 i 类且预测也为 i 类, p_{ij} 表示原属于 i 类却预测为其他类的像素点数, p_{ji} 表示原属于其他类却预测为 i 类的像素点数, 在分母处 p_{ii} 计算了两次所以要减去一个 p_{ii} , IoU 计算示意图如图2.18所示。MIoU 具体计算公式如下:

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}} \quad (2.4)$$

2.6 本章小结

本章首先介绍了语义分割中全卷积神经网络中的基本概念和组成结构。接着, 详细回顾了与本文工作相关的两种经典网络结构, 即 FCN 和 U-Net, 描述了从分类网络到分割网络的变化和创新。本章还介绍了很多经典的深度学习模块或者方法在语义分割任务的应用, 例如编码器和解码器、多尺度的特征融合、注意力机制和图卷积等。然后回顾了图卷积的实现方式和相关理论知识, 并讨论了图卷积在各个领域的应用, 着重介绍了在语义分割领域的应用。最后, 介绍了语义分割领域中常见的评价指标。

第三章 基于图卷积与深度学习的编解码网络

针对材料图像数据集规模小、纹理复杂和结构尺寸大小不一等问题，本文提出了一种基于图卷积结合多维特征融合的材料图像分割方法。具体而言，针对材料图像的小样本特性，该方法利用残差连接缓解特征的损失；针对材料图像的纹理复杂和结构尺寸大小不一，该方法利用多尺度特征融合模块通过不同感受范围的感受野和不同尺度的特征图融合学习更丰富准确的材料结构特征，并使用注意力机制结合图卷积提高重要特征的注意力。该方法分为四个步骤：编码器提取图像的高级特征图；图注意力模块计算注意力权重实现局部聚合；多尺度融合模块融合不同尺度特征图；解码器输出图像的像素级预测结果，最终实现多种材料图像的分割。

3.1 方法概述

材料图像分割主要面临三个难点。第一，材料图像通常使用显微镜拍摄，呈现单调的灰度图，各区域具有同相结构差异大、纹理表征差别不明显且边界不连续甚至模糊等特点；第二，深度学习方法在解决这类任务时大都对各个像素进行分类，没有考虑到像素之间的关系，忽略了上下文信息，不能捕捉图像中像素之间的更长距离关系；第三，材料图像数据的获得往往需要投入大量的时间成本、人力成本和设备成本，导致数据规模较小。

为了克服以上难点，本文提出了基于图卷积结合多维特征融合的网络（Multi-dimensional Feature Fusion Network based on Graph Convolution, FMGNet），为解决材料图像中的小样本分割问题提供了新的思路。图3.1展示了该网络的结构，输入图像先经过包含残差连接的编码器获得高阶语义信息；再经过一个图注意力模块，将特征图基于注意力进行局部聚合；再经过一个多尺度特征融合模块，融合不同尺度的特征图；随后特征图进入解码器，对缩小后的特征图像进行上采样；最后输出分割结果图。FMGNet 的主要思想是融合多层次、多维度的特征信息，关注局部特征和全局特征更好的融合，通过反复补充和丰富不同分辨率的图像特征，在少量样本的条件下得到更优质的分割结果。下面对方法的各个部分一一描述。

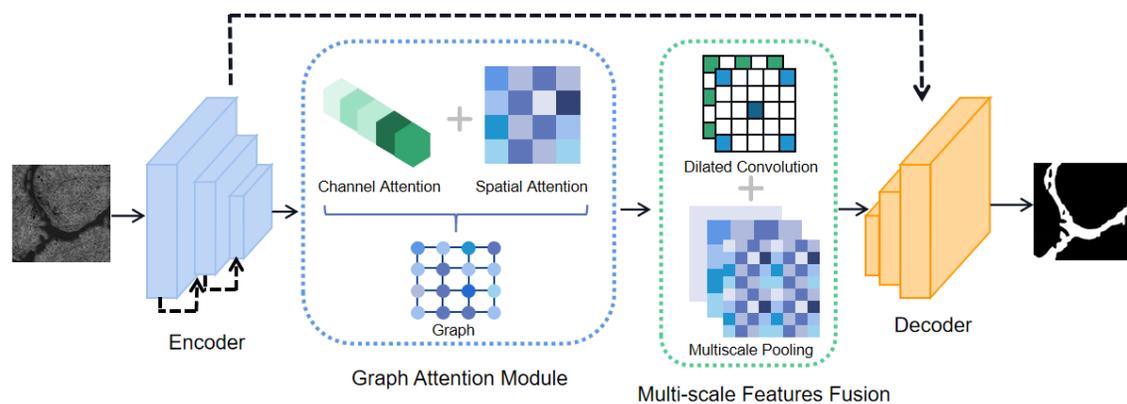


图 3.1: FMGNet 网络结构。输入图像先经过包含残差连接的编码器获得高阶语义信息；再经过一个图注意力模块，将特征图基于注意力进行局部聚合；再经过一个多尺度特征融合模块，融合不同尺度的特征图；随后特征图进入解码器，对缩小后的特征图像进行上采样；最后输出分割结果图。

3.2 基于图卷积结合多维特征融合的网络模型

网络的详细结构见图3.1。编码器可以提取和分析图像的低层特征像素值，获得高阶语义信息；图注意力模块用来增强特征图的通道和空间双重注意力并根据注意力权重对局部特征进行关注；多尺度特征融合模块用来整合从编码器的特征图不同感受野和不同尺度的信息；解码器收集语义信息，并将相同类别结构的标签映射到相应的像素。

3.2.1 编码器和解码器

在深度学习骨干网络中，网络浅层生成的高分辨率低级特征图对细粒度结构的编码效果更好，而网络深层生成的低分辨率高级特征对像素级分类预测的影响更大。两者是实现语义分割的重要且互补的方法。为了结合这两种特征的优点，本章提出了一种改进的编码器-解码器结构 [30]，从不同分辨率的特征图中集成和重构具有丰富语义信息的特征。

网络中的编码器-解码器的详细结构见图3.2。编码器部分是基于 VGG16 [31] 前 13 层进行改进。VGG16 中卷积层包含五个部分，每个卷积层卷积核大小为 3×3 、步长为 2。前两个部分中，每个部分各有两个卷积层；后面的三个部分，每个部分

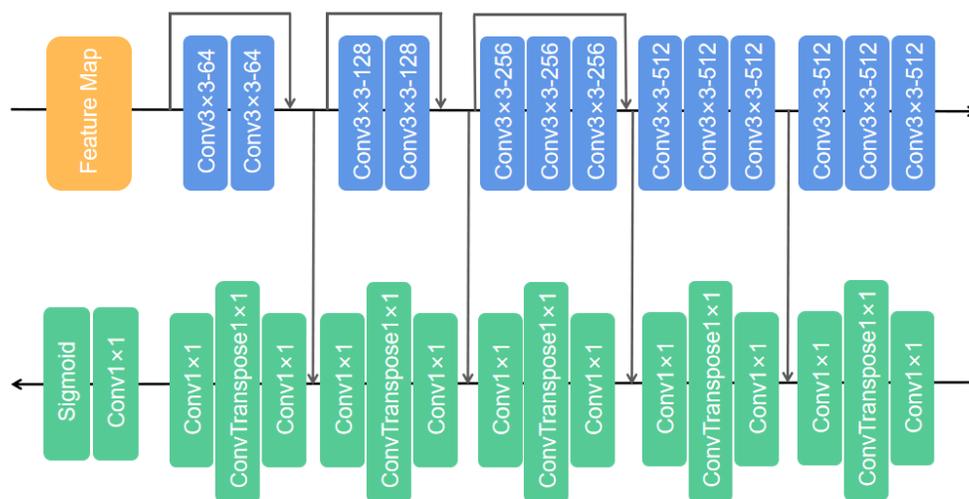


图 3.2: 编码器和解码器结构。编码器-解码器属于 FMGNet 网络结构中的 Encoder 和 Decoder 部分。

各有三个卷积层。为了丰富语义信息，本文在每个部分添加一层残差连接操作，用来融合不同卷积层提取的特征图，具体计算见公式3.1。每一层得到的特征图与下一层特征图进行融合，实现不同分辨率和不同感受野的信息充分补充。最后，网络通过编码器会得到五个维度的特征图。

$$y = F(x) + x \quad (3.1)$$

其中， x 是编码器部分中每一层的输入特征图， $F(\cdot)$ 是每一层的具体卷积等操作， y 是每一层的输出特征图。残差连接在实践中有大的优势，有改善梯度消失的作用，需要注意的是方程中 x 和 F 的维度必须是相等的。

解码器分为五个部分，每个部分包含一个反卷积层和两个卷积层。反卷积层的作用是放大特征图的尺寸，卷积层的作用是提高非线性函数的表达能力。在反卷积过程中，其中四个反卷积层依次持续放大特征图，每次放大后都与编码器相应维度的特征图融合，以补充特征图中的精细特征，缓解特征图中像素信息和空间信息的损失。解码器输出的特征图再通过一个卷积层和 Sigmoid 激活函数得到最终的像素级分类结果。

3.2.2 图注意力

图注意力模块（Graph Attention Module, GAM）是网络的重要组成部分，主要起到了提高对图像细节关注的作用，对图像中的一些不明显的噪点或者边界更加灵敏、准确。模块结构如图3.3所示。

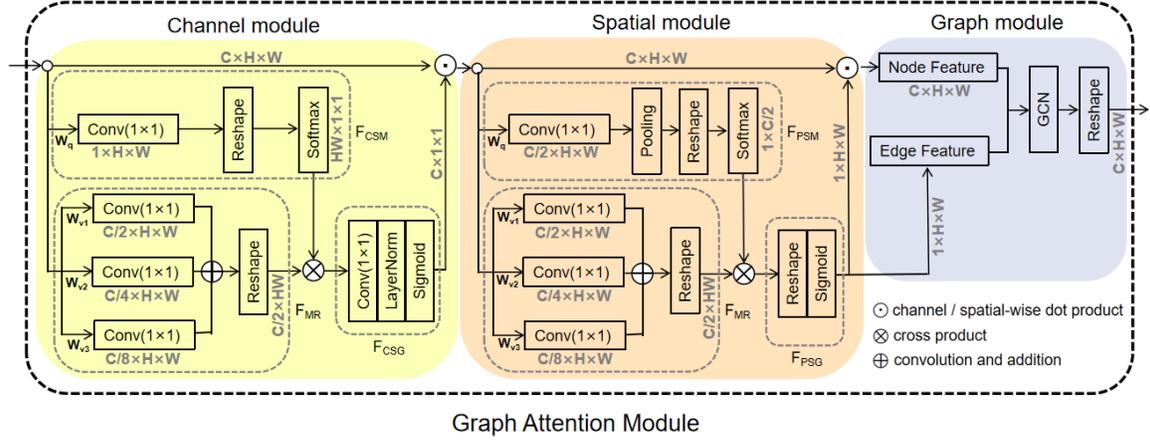


图 3.3: 图注意力模块的结构。

双重注意力分为两个部分，通道模块和空间模块，通道模块的权重 $M^{ch}(X)$ 计算公式如下：

$$M^{ch}(X) = F_{CSG}(F_{CSM}(X) \otimes F_{MR}(X)) \quad (3.2)$$

X 是编码器部分输出的特征图，维度为 $C \times W \times H$ 。 \otimes 是矩阵之间的叉乘， $F_{CSM}(X)$ 是降维函数，将输入特征图的维度由 $C \times H \times W$ 降到 $HW \times 1 \times 1$ ，具体公式如下：

$$F_{CSM}(X) = F_{SM}(R_1(W_q(X))) \quad (3.3)$$

W_q 是核为 1×1 的卷积层，将通道维度由 C 降为 1。 R_1 是 Reshape 操作，将特征图的维度由 $1 \times H \times W$ 变为 $1 \times HW$ 。 $F_{SM}(\cdot)$ 是 Softmax 激活函数。因为 W_q 的通道维度被压缩，如 PSA [93] 所说的那样，就需要通过高动态范围（HDR）进行特征信息的增强，因此用 Softmax 对 W_q 的特征以分配权重值的方式进行了强化。

$F_{MR}(X)$ 是融合函数，用于融合不同通道维度的特征图，具体公式如下：

$$F_{MR}(X) = R_2(F_M(W_{v1}(X), W_{v2}(X), W_{v3}(X))) \quad (3.4)$$

W_{v1} 、 W_{v2} 和 W_{v3} 都是核为 1×1 的卷积层，分别将特征图的通道维度由 C 降为 $C/2$ 、 $C/4$ 和 $C/8$ 。 $F_M(\cdot)$ 是融合函数，用于融合 W_{v1} 、 W_{v2} 和 W_{v3} 卷积后的三个特征图，即将不同通道维度 ($C/2$ 、 $C/4$ 和 $C/8$) 的特征图先通过核为 1×1 的卷积操作统一维度为 $C/2 \times H \times W$ 的特征图再加和。 R_2 是 Reshape 操作，将特征图维度由 $C/2 \times H \times W$ 变为 $C/2 \times HW$ 。 W_{v1} 、 W_{v2} 和 W_{v3} 卷积后分别代表高分辨率、中分辨率和低分辨率的特征图。将 W_{v1} 、 W_{v2} 和 W_{v3} 进行高低分辨率的特征融合，增强特征的代表能力。

$F_{CSG}(X_{input})$ 是权重生成函数，将输入特征图 X_{input} 的维度由 $C/2 \times 1 \times 1$ 转化为 $C \times 1 \times 1$ 再生成通道维度的权重向量，具体公式如下：

$$F_{CSG}(X_{input}) = F_{SG}(W_z(X_{input})) \quad (3.5)$$

X_{input} 是 $F_{CSM}(X)$ 和 $F_{MR}(X)$ 输出特征图叉乘的结果，维度为 $C/2 \times 1 \times 1$ 。 W_z 是核为 1×1 的卷积层，将特征的通道维度由 $C/2$ 升为 C 。 $F_{SG}(\cdot)$ 是 Sigmoid 激活函数，使得所有的参数光滑的映射到 0-1 之间。通过这种方式，模块得到通道维度的权重向量，并将通道维度的权重向量与输入的特征点乘，得到包含通道注意力的特征图。

本文的通道模块将输入特征压缩到通道维度大小，通过 Sigmoid 激活函数给每个通道的特征中添加一个表示通道与关键信息相关度的权重。权重越大，特征之间的关联度越高。然后，通道模块得到的通道注意力向量作用于语义信息路径对应的特征图上。

由通道注意力模块输出的特征图将会进入空间注意力模块。空间注意力模块计算空间维度权重 $M^{sp}(X_c)$ 的公式如下：

$$M^{sp}(X_c) = F_{PSG}(F_{PSM}(X_c) \otimes F_{MR}(X_c)) \quad (3.6)$$

X_c 是通道注意力模块输出的特征图。 F_{MR} 和公式3.4相似， W_{v1} 、 W_{v2} 和 W_{v3} 特征的空间维度保持在一个比较大的水平 ($H \times W$)。然后将 W_{v1} 、 W_{v2} 和 W_{v3} 卷积后的特征进行高、低分辨率的特征融合。 $F_{PSM}(X_c)$ 是降维函数，将 X_c 的维度由

$C \times H \times W$ 降到 $1 \times C/2$ ，具体公式如下：

$$F_{PSM}(X_c) = F_{SM}(R_3(F_{GP}(W_q(X_c)))) \quad (3.7)$$

W_q 是核为 1×1 的卷积层，将通道维度由 C 降为 $C/2$ 。 F_{GP} 是全局池化函数，将维度由 $C/2 \times H \times W$ 降到 $C/2 \times 1 \times 1$ 。 R_3 是 Reshape 操作，将维度由 $C/2 \times 1 \times 1$ 变为 $1 \times C/2$ 。 $F_{SM}(\cdot)$ 是 Softmax 激活函数，因为 W_q 的空间维度被严重压缩，所以就用了 Softmax 进行信息增强。

$F_{PSG}(X_{input1})$ 生成了维度为 $1 \times H \times W$ 的权重向量，具体公式如下：

$$F_{PSG}(X_{input1}) = F_{SG}(R_4(X_{input1})) \quad (3.8)$$

X_{input1} 是 $F_{PSM}(X_c)$ 和 $F_{MR}(X_c)$ 叉乘的结果，维度为 $1 \times HW$ 。 R_4 是 Reshape 操作，维度由 $1 \times HW$ 变为 $1 \times H \times W$ 。 $F_{SG}(\cdot)$ 是 Sigmoid 激活函数，由此该模块得到了空间维度的权重向量。将空间维度的权重向量和输入特征点乘便得到了包含空间注意力的特征图。

考虑到图像中不同空间位置信息的重要性不同，空间注意力从空间信息路径的特征图中提取空间注意力矩阵，然后将提取的空间注意力矩阵应用于语义信息路径对应的特征图。通道注意力和空间注意力的融合是为了同时激发通道和空间两个维度上的重要信息，对目标特征的通道、语义信息进行侧重分配，增强高层特征图语义信息的表达能力。融合后的特征图将对后续的分类和预测产生重要作用。相比于极化注意力 (PSA) [93]，本文注意力模块主要的亮点：在通道和空间维度保持更丰富的分辨率的特征融合，即在通道和空间上分别使用通道数为 $C/2$ 、 $C/4$ 和 $C/8$ 的特征融合。这一步不仅能够很大程度上减少降维造成的信息损失，还通过不同分辨率的特征图丰富特征的代表能力。

本文提出了一种基于注意力的图卷积网络。图卷积网络 (Graph Convolutional Network, GCN) 是一种可以直接对图结构数据进行操作的特征提取器，通过图卷积运算，可以充分利用节点之间的依赖关系和每个邻居节点的特征信息来分析图结构 [102]。因此，与卷积神经网络不同，GCN 可以直接应用于不规则数据。

图结构的构建需要每个节点的节点特征和所有节点之间的连接关系。图结构构

建后可以用于图结构的分类和逐点分割，通过节点聚合的特征预测节点的标签的方法。由于图卷积网络只能直接作用于图结构数据，本文将图结构中的节点类比到图像的像素点，图结构的邻接矩阵表示图像像素点之间的重要联系程度。本文将注意力机制和图卷积神经网络结合为一体，探究图像像素与像素之间的联系，发现局部特征之间的关联。这样，基于 GCN，本文将欧式空间的数据结构应用到非欧式空间的结构，用一种新的思维方式解决语义分割的挑战。

本文设计图结构主要有两个特点。第一，图结构有更加灵活的跳跃连接，与卷积网络以一个像素为中心，建立一个 $n \times n$ 的局部区域进行卷积不同，GCN 可以探索到图网络中不同节点间的关系。第二，图卷积利用消息传播机制在节点间实现信息交换，节点的感受野增大，实现更大范围的特征得以充分捕捉。

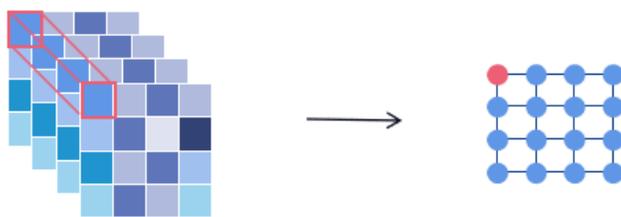


图 3.4: 特征图以四邻域的连接方式转化成图的示意图。

图结构类型的数据可以用 $G(N, E, U)$ 表示， $N \in R^K$ 表示图结构的节点集合，其中 K 表示节点个数。 $E \in R^M$ 表示图结构的边集合，其中 M 表示边的数量。 $U \in R^{K \times C}$ 表示图结构的特征集合，其中 C 为特征向量的维数。显然，图的定义和结构在特征传播中起着关键作用。本文提出的框架的动机是通过不同的图投影策略构建特征图，以学习更好的特定任务表示。在图结构中，本文根据像素点的空间位置关系让每个节点按照四邻域的方式建立边的连接，连接方式示意图如图3.4所示。每个节点都有属于自己的特征，由此得到了节点的特征矩阵。然后，就可以用 GCN 对这些节点进行聚合。

图结构的建立依赖于特征图的维度。特征图经过通道注意力模块和空间注意力模块后维度为 $C \times W \times H$ ， C 是通道数， W 和 H 是特征图的长和宽。本文将特征图划分为 $W \times H$ 个节点。每个节点的节点特征维度是 $1 \times C$ ，邻接矩阵的维度

是 $(W \times H) \times (W \times H)$ 。本文提出的模块通过不同的图构建策略应用在语义分割中，一般来说采用 $R(\cdot)$ ，即 Reshape 操作，将特征图转换为新的维度结构得到 X_r 。

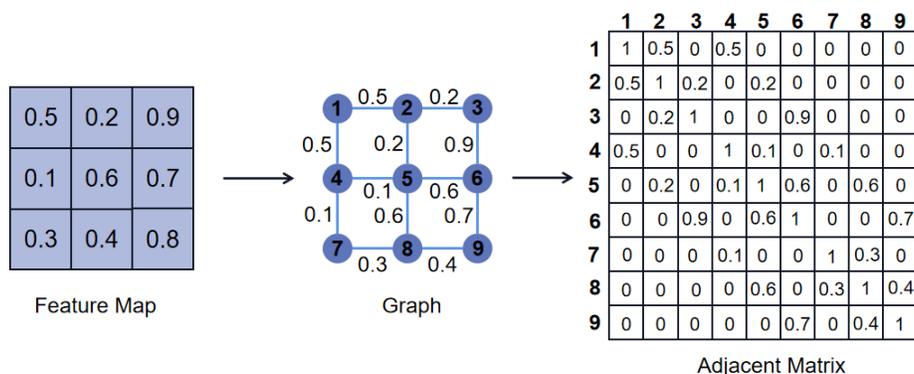


图 3.5: 一个邻接矩阵的构建例子。带有空间维度权重的特征图，以特征块为节点，权重值为边值构建成图结构，每个节点再添加了自连接，形成邻接矩阵 \tilde{A}_s 。

图注意力模块将空间注意力中获取的权重作为邻接矩阵的值，即特征图的每个像素点的权重值作为该像素点的右方和下方邻近像素点的边权重。一个权重值设定的例子如图3.5所示，邻接矩阵 A_s 再通过规范化操作添加自连接，因此得到了一个新的邻接矩阵 \tilde{A}_s 。每个像素点的局部区域都会有侧重的进行特征聚合，通过邻接矩阵，对于节点间不同的连接边值，来划分对节点的关注程度。中心节点与其他节点的边权重越大，表示节点间的信息流动越流畅，相连的邻近节点的特征越紧密的聚合于中心节点；中心节点与其他节点的边权重越小，表示节点间的信息流动越堵塞，中心节点的特征会以自身节点特征为主。因此，分配边权重的本质还是分配邻近节点特征的权重。图结构数据中，由于不相似节点可能直接相连，也就是连通节点之间的相似度可能低于未连通节点。对于这种情况，给邻接矩阵分配可学习的权重，通过网络不断地学习和训练，优化权重，即使连通节点之间的相似度低，也可以通过分配小的权重来抑制节点特征的聚合。模块用邻接矩阵 A_s 构建一个轻量级的全连通图，用于在节点间传播信息，图的节点聚合来自像素“簇”的信息，而边度量这些簇的重要程度。基于注意力的图卷积层后，局部特征聚合更新完成，提高了局部信息的丰富和补充。基于 [102] 图卷积公式，本文的图卷积操作得到新的公式，公式如下：

$$Z^{l+1} = \sigma(\tilde{D}_s^{-\frac{1}{2}} \tilde{A}_s \tilde{D}_s^{-\frac{1}{2}} X_r W^l) \quad (3.9)$$

X_r 是由空间注意力模块输出转换的图结构。 D_s 和 A_s 分别是 X_r 的度矩阵和邻接矩阵。 \sim 是指图的每个节点添加了自连接, 并构建了相应的度矩阵 \tilde{D}_s 和邻接矩阵 \tilde{A}_s 。 W^l 是第 l 层图卷积的权值矩阵。 σ 是 ReLU 激活函数。 Z^{l+1} 是第 $l+1$ 层图卷积操作的输出。

本文将提出的图注意力模块合并到编码器-解码器主干架构中, 在编码器和多尺度特征融合模块之间插入图注意力模块。由通道注意力和空间注意力融合输出的新特征再经过图卷积模块, 不改变它原本的维度, 也不改变它的空间位置关系。因此, GCN 明确地模拟了图像中像素之间的空间关系, 使网络能够产生连贯的预测, 考虑图像中的所有对象。图卷积层实现节点特征的更新后, 会再封装回特征图, 进行后续的学习和训练过程。

3.2.3 多尺度特征融合模块

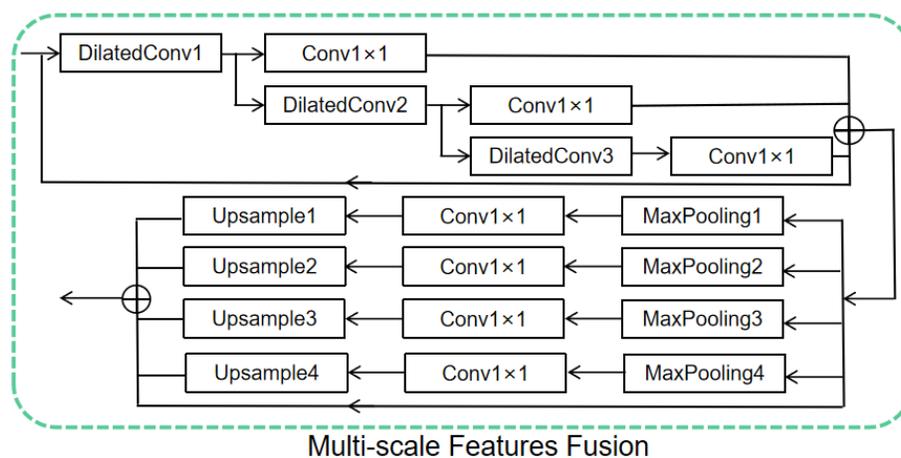


图 3.6: 多尺度特征融合模块。

多尺度特征融合模块 (Multi-scale Feature Fusion Module, MFFM) 主要由空洞卷积层和池化层构成, 具体结构如图3.6所示。当空洞卷积在高层特征图使用的空洞指数过大时, 或许只对一些大面积对象有良好的分割效果, 而对小面积对象来说

可能有弊无利。由于空洞指数大，对输入信息的采样会变得非常稀疏，不利于学习，导致一些局部信息的丢失，一些远距离信息可能不产生有效的帮助。当空洞指数过小时，采样的感受野小，容易忽略了全局的信息。

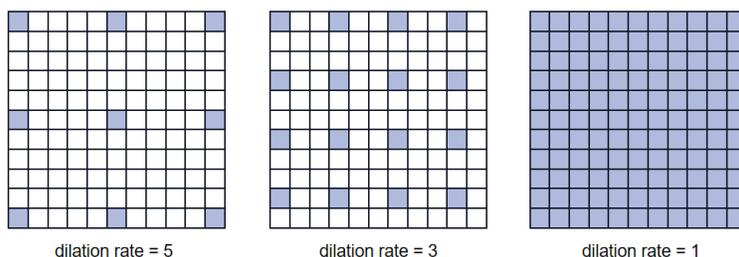


图 3.7: 不同空洞指数的卷积核。

为了弥补感受野过小或过大的弊端，本文使用混合空洞卷积，将不同空洞指数采样的特征信息完全融合。如图3.7，该模块使用空洞指数分别为 1、3 和 5 的空洞卷积层，分别获取局部和全局信息。首先该模块将输入的特征图通过一层空洞指数为 1 的空洞卷积层（DilatedConv1）输出 Z_1 。然后该模块将 Z_1 放入空洞指数为 3 的空洞卷积层（DilatedConv2）输出 Z_2 。紧接着，该模块的 Z_2 放入空洞指数为 5 的空洞卷积层（DilatedConv3）输出 Z_3 。 Z_1 、 Z_2 和 Z_3 再分别进行一次卷积核为 1×1 的卷积操作，最后将三个输出和输入的特征图加和得到混合空洞卷积操作的结果。模块同时满足小物体和大物体的分割要求，近距离信息由小的空洞指数的空洞卷积来收集，远距离信息由大的空洞指数的空洞卷积来收集。

融合不同感受野的特征图后，模块进入不同分辨率的特征图融合阶段。当池化层的下采样倍数较小时，特征感受野小，适合处理小目标，大尺度的特征图可以获得更多关于小目标的有用信息。当池化层的下采样倍数较大时，特征感受野大，适合处理大目标，小尺度的特征图获得更多大目标的有用信息。为了结合大尺度和小尺度的优点，该模块使用多尺度池化，将不同尺度的特征图结合，如图3.6所示。模块使用 MaxPooling1、MaxPooling2、MaxPooling3 和 MaxPooling4 池化层处理混合空洞卷积的输出特征图，核分别为 2×2 、 3×3 、 5×5 和 6×6 ，步长分别为 2、3、5 和 6。输入特征图的维度由 $C \times W \times H$ 分别降维为 $C \times W/2 \times H/2$ 、 $C \times W/3 \times H/3$ 、 $C \times W/5 \times H/5$ 和 $C \times W/6 \times H/6$ 。四个池化层的输出分别

使用卷积核为 1×1 的卷积将通道压缩为 1。然后四个特征图通过上采样分别恢复到原来水平的大小 ($1 \times W \times H$)。最后，四个上采样层输出的特征图和混合空洞卷积输出的特征图进行拼接得到了多尺度特征融合模块的输出，输出特征图的维度是 $(C + 4) \times W \times H$ 。特征图的通道数增加了 4，正是因为拼接了四个不同尺度的特征图。

3.3 实验和讨论

3.3.1 实验环境与评价标准

在本章的实验中，本文使用的计算机配置为 Intel(R) Core(TM) i7-7700K CPU@4.20Hz，NVIDIA GeForce GTX 1060 3GB，以及 Python3.8 的开发工具。本文的网络是通过 Pytorch 实现的，Pytorch 是 Torch 的 Python 版本，由 Facebook 开源的一个神经网络框架。与 Tensorflow 的静态计算图不同，Pytorch 的计算图是动态的，可以根据需要实时更改计算图 [120]。

本文将图片大小统一调整为 512×512 作为网络输入，将二元交叉熵损失函数作为网络的损失函数。本文使用带动量的梯度下降优化算法 Adam 训练网络模型，迭代次数为 100，动量和衰减系数为 0.9 和 0.0005，批处理大小为 1，初始学习率设为 0.001。

表 3.1: 数据集分布情况。

class	Spheroidite	Wood	Pearlite1	Pearlite2	Ceram
train set	4	3	4	3	3
test set	2	1	2	1	2
data set	6	4	6	4	5

实验的数据集一共五类材料结构，分别是 Spheroidite, Wood, Pearlite1, Pearlite2 和 Ceram。其中，Spheroidite, Pearlite1 和 Pearlite2 来源于 [4]，是超高碳钢 (UHCS) 的微观结构数据集。碳钢铸造时强度低，脆性大，经热处理和热处理后，可获得均匀分布有细小渗碳体颗粒的超细铁素体，钢的组织在不同的加工和处理条件下会发

生很大的变化。木材微观结构是木材的超微组织结构，其特征是排列规则，主要由轴向薄壁组织和树脂通道组成。陶瓷微结构是可以显微镜观察到的陶瓷体的内部结构。

数据集由于制备复杂和标注成本高昂，每类有 4 到 6 张图片，数据集分布见表3.1。本文会将每类的图片根据数据集的大小选择训练集的规模用于训练，一般训练集和测试集的比值是 3: 1 或者 2: 1。本文主要用 MIoU 作为语义分割网络的评价指标。

3.3.2 现有方法和 FMGNet 的对比实验

为了进一步证明提出的模型有效的应用在材料图像分割任务中，本文比较了在五个数据集上当前先进方法的分割性能，其中传统方法包括 KMeans [20] 和 Region-Grow [121]，以及深度学习方法，包括 FCN [30]、R2-UNet [122]、Deeplab3+ [38]、CENet [77]、DFANet [123]、DGCNNet [56]、UNet3+ [70]、BiSeNet [124] 和 DANet [91]。

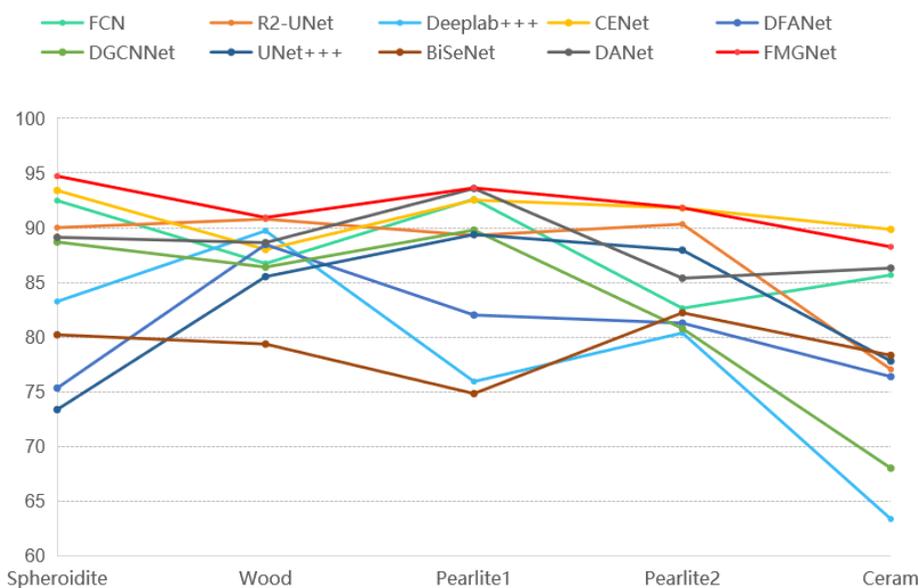


图 3.8: 深度学习方法的对比实验折线图。折线图的曲线表示 10 种深度学习方法，包括 FCN [30]、R2-UNet [122]、Deeplab3+ [38]、CENet [77]、DFANet [123]、DGCNNet [56]、UNet3+ [70]、BiSeNet [124]、DANet [91] 和本章方法。纵轴表示 MIoU (%)，横轴表示五个材料图像的测试集。

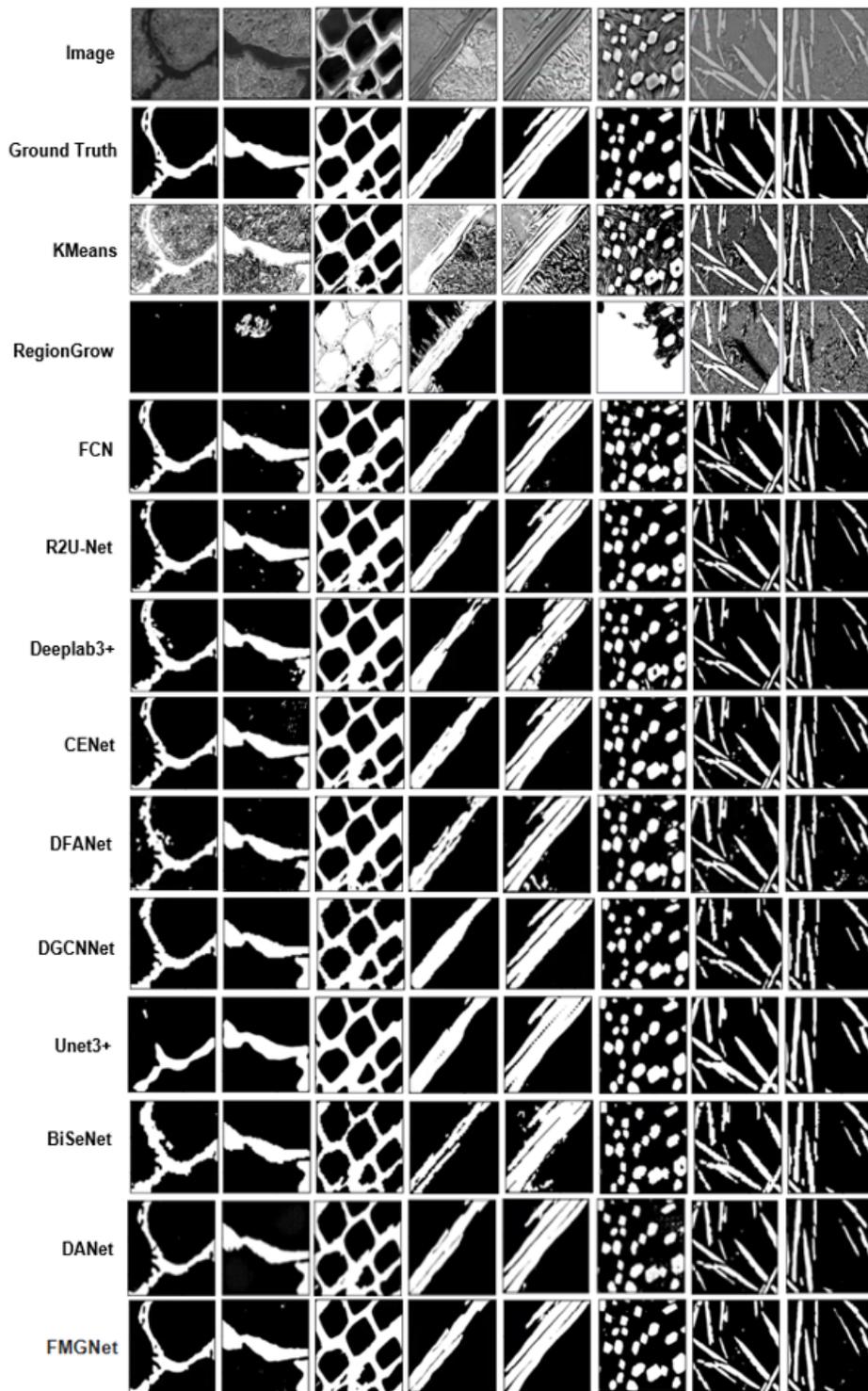


图 3.9: 现有方法对比可视化结果。图中分别是原始输入图像、输入图像的标注图, 传统方法中 KMeans [20] 和 RegionGrow [121] 的分割结果图, 还有深度学习方法 FCN [30]、R2-UNet [122]、Deeplab3+ [38]、CENet [77]、DFANet [123]、DGCNNet [56]、UNet3+ [70]、BiSeNet [124] 和 DANet [91] 的分割结果图。

表 3.2: 与现有方法的对比实验。本章的方法与先进的图像分割方法进行对比实验, 主要对比传统方法和深度学习方法, 评价指标为 MIoU (%)。

Methods	Spheroidite	Wood	Pearlite1	Pearlite2	Ceram
KMeans [20]	38.0144	88.4259	35.5587	70.0082	62.8760
RegionGrow [121]	39.7494	29.4655	53.5102	25.9609	47.8345
FCN [30]	92.4857	86.7264	92.6187	82.6316	85.6721
R2-UNet [122]	90.0087	90.7879	89.2969	90.3309	77.0384
Deeplab+++ [38]	83.2498	89.7209	75.9259	80.3745	63.3483
CENet [77]	93.4119	88.0033	92.5342	91.8024	89.8484
DFANet [123]	75.3247	88.4672	82.0128	81.2751	76.3730
DGCNNNet [56]	88.6951	86.3972	89.7884	80.7797	68.0016
UNet+++ [70]	73.3576	85.5251	89.3767	87.9503	77.7987
BiSeNet [124]	80.20605	79.3578	74.8235	82.22057	78.3223
DANet [91]	89.1257	88.6333	93.5909	85.3818	86.3052
FMGNet	94.7262	90.9345	93.6414	91.8192	88.2500

图3.8以折线图的形式展示了深度学习方法在测试集上的 MIoU 对比结果, 图3.9展示了本章方法和其他方法在测试集上的可视化结果, 表3.2展示本章方法和其他方法在测试集的 MIoU。如图3.9所示, KMeans 的分割结果不准确。首先对噪声过于敏感, 导致绝大多数的噪声都被分割出来; 其次, KMeans 不能得到较好的区域结构, 造成了区域的过度分割, 甚至将材料结构的纹理细节全部勾勒出来。Region-Grow 在材料图像分割中性能不稳定, 对不均匀的连通目标没有很好的分割效果, 如图3.9中 RegionGrow 第 1、2、5 和 6 图中, 基本没有分割出目标结构甚至出现了分割图大面积白色或全黑的错误结果。各个网络模型在材料结构数据集上表现因为样本过少, 无法得到充分数据集训练, 导致对材料图像的特征学习十分匮乏。FCN 的实验结果出现不同程度的噪声和过度分割, 如图3.9所示 FCN 的第 3 张图和第 6 张图, 在一个相的完整区域内出现不同相。可视化结果说明 FCN 的感受域偏小, 没有很好的结合上下文信息, 导致没有学习到全局的特征信息。R2U-Net 的实验结果出现了局部分割错误, 如图3.9中 R2U-Net 第 1 张图和第 6 张图, 错误的分割了

Spheroidite 和 Pearlite 的形状, 可视化结果说明 R2U-Net 对于特征的学习和训练与真实标签偏差较大, 没有在少量样本的前提下学习到正确的特征表示。Deeplab3+、DFANet 和 BiSeNet 分割结果与 R2U-Net 雷同甚至更差, 在颜色相近和纹理相似的情况下, 很难智能的分割出不同的材料结构。如 Deeplab3+ 的第 1、5 张图, 由于两相之间对比度较低, 边界难以区分, 导致边界处出现大量分割错误。如 DFANet 的第 4 张分割结果图, 相与相之间的边界很好的分割, 却偏向过度分割, 将一些整体结构也孤立成两个结构, 还将部分噪声也识别成错误的类别。BiSeNet 的分割错误更加明显, 并且对于边界的划分十分不平滑。CENet 的实验结果很好, 但是细节处理不够精细, 如图3.9中的 CENet, 在部分图像上表现优异, 然而不能稳定的保证每一类材料图像的分割结果准确, 对相与相之间的边界划分不太敏感。可视化结果说明 CENet 中高分辨率的特征图经过多次卷积或其他操作, 损失了重要的边界细节, 低分辨率的高级特征图占据了主导地位。例如, 在表3.2的 Pearlite1 数据集中, CENet 的指标比 FMGNet 只低了 1.1 百分点。但是在图3.9中 CENet 的第 4 张图, 相的边界完全没有体现出来, 基本上划分为了一类。DGCNNNet 和 UNet3+ 的实验结果出现大范围的边界分割错误, 表3.2中 DGCNNNet、UNet3+ 和本章方法的 MIoU 相对比, 本章方法的性能大幅度超过他们。如图3.9中 DGCNNNet 虽然很好的找到了每相的区域, 去掉了几乎所有的噪声, 但是它对整体相的边界分割十分迟钝, 类似锯齿状, 表现得不够平滑, 没有很好的平衡对噪声和区域分割。UNet3+ 虽然边界分割的十分平滑, 但是目标的空间距离十分靠近的情况下, 很难准确识别到目标的边界。可视化结果说明 DGCNNNet 和 UNet3+ 中过度的丰富特征或者过多的跳跃连接反而适得其反, 容易导致信息冗余。DANet 在一些数据集上有很好的分割效果, 但是在一些数据集上出现了很多琐碎的噪声, 如图3.9中 DANet 的第 1 张图, 分割出现了明显的边界错误。更详细的统计可以看表3.2, FMGNet 在 Spheroidite, Wood, Pearlite1, Pearlite2 和 Ceram 的测试集上 MIoU 分别是 94.7262%、90.9345%、93.6414%、91.8192% 和 88.2500%, 大多指标超过现有方法 2-10 个百分点。CENet 的 MIoU 在 Ceram 数据集表现突出, 但是其他数据集上 FMGNet 的指标表现更优异。在折线图3.8中可以看出, 本章方法所表示的曲线在前四个测试集上稳定的处于

所有曲线上方，在 Ceram 数据集上十分接近最高点 CENet 的 MIoU。先进的方法优点和缺点并存，大都不能做到优点兼并，即既可以对边界完好分割又可以把每一类完整呈现。

我们提出的 FMGNet 网络模型对多类数据集的分割效果呈现稳定的性能。首先，材料结构的大部分噪点都被很好的过滤，对大面积的连通区域分割准确；其次，材料结构之间的狭窄边界分割优于绝大多数方法；最后，在颜色对比度低的情况下，网络依然展现强大的学习能力。本章方法较好的权衡了图像中大物体和小物体的关注度。

3.3.3 不同骨干网络的 FCN 和 FMGNet 的对比实验

本文探究了基于不同骨干网络的 FCN 和 FMGNet 的性能，对比了骨干网络为 VGG16、ResNet34、ResNet50 和 UNet 的网络模型。图3.10展示了 VGG16、ResNet34 和 ResNet50 的网络结构（ResNet 的网络结构没有显示残差连接）。

本文选择 VGG16、ResNet34、ResNet50 和 UNet 作为主要比较的网络模型是因为 FMGNet 的编码器部分是基于 VGG16 和残差结构改进所得。ResNet34 和 ResNet50 都拥有残差连接并且网络层数和 VGG16 相差不多，这样对比可以在缩小编码器部分的差异的情况下，探究编码器部分中层数和残差连接对于分割材料图像结果的影响。本文选择对比 UNet 是因为 UNet 的提出最先应用于医学图像的，而医学图像与材料图像非常相似。为了探索所提模块在改变骨干网络后是否也能提高分割性能，本文选择了四个有代表性的网络结构。

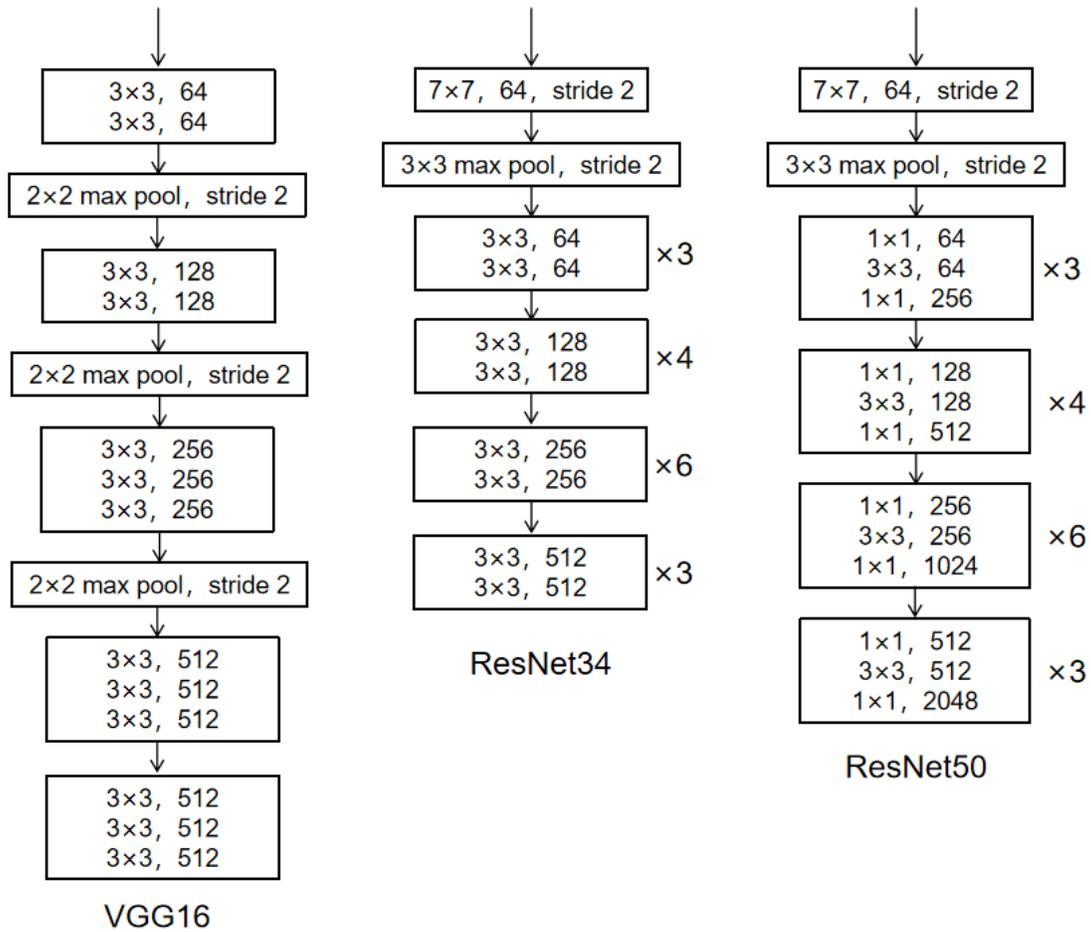


图 3.10: VGG16、ResNet34 和 ResNet50 的网络结构示意图 (ResNet 的网络结构没有显示残差连接)。

表 3.3: 不同骨干网络的 FCN 和 FMGNet 的对比实验。本文对比了骨干网络为 VGG16、ResNet34、ResNet50 的 FCN 和 UNet 与 FMGNet 在五个数据集上的 MIoU (%), 并且计算了每个网络在所有数据集的 MIoU 平均值。

Methods	Spheroidite	Wood	Pearlite1	Pearlite2	Ceram	Average
FCN(ResNet34)	93.5685	87.0280	93.0097	91.4508	88.7177	90.7549
FMGNet(ResNet34)	94.2877	88.4998	93.8427	91.4191	88.7084	91.3515
FCN(ResNet50)	91.5149	88.8333	93.5988	92.0617	87.9175	90.7852
FMGNet(ResNet50)	93.4291	89.0097	93.4171	91.5057	88.1058	91.0935
UNet	90.9038	87.1089	93.5368	92.7355	89.4665	90.7503
FMGNet(UNet)	93.7649	86.4137	93.7351	92.8730	90.9544	91.5482
FCN(VGG16)	92.4857	86.7264	92.6187	82.6316	85.6721	88.0269
FMGNet(VGG16)	94.7262	90.9345	93.6414	91.8192	88.2500	91.8743

表3.3展示了不同骨干网络和 FMGNet 对比的 MIoU 结果。实验发现基于 ResNet 的 FCN 由于编码器部分层数更多, 网络结构更加复杂, 能实现很好的分割效果; 基于 ResNet 的 FCN 加上图注意力、多尺度特征融合模块和增加反卷积的非线性表达后提升的性能幅度不是很显著。在 FMGNet(ResNet50) 中, 在 Spheroidite 数据集相比 FCN(ResNet50) 提高了 1 个百分点左右, 整体网络性能和 FCN(ResNet50) 几乎持平, 提升了很微弱的效果。基于 VGG16 的 FCN 的编码器部分层数较少, 网络结构较简单, 相比与基于 ResNet 的 FCN 的网络性能较差; 然而, 基于 VGG16 的 FCN 加上跳跃连接、图注意力、多尺度特征融合模块和增加反卷积的非线性表达后, 整体网络性能提升了将近 4 个百分点。基于 UNet 的 FMGNet 在部分数据集上具有优秀的性能, 整体上具有良好的分割性能。总体而言, 在基于不同的编码器结构的 FMGNet 中, FMGNet(VGG16) 的网络性能最优。

3.3.4 消融实验

本文的工作在 VGG 和 FCN 的框架结构中改进了很多地方。为了了解每个改进点对网络模型的影响力, 本文做了消融实验, 第 1 组是 FMGNet 去除解码器中每个 Decoder 层的卷积操作, 第 2 组是 FMGNet 去除编码器中的残差连接, 第 3 组

是 FMGNet 去除多尺度特征融合模块，第 4 组是 FMGNet 去除图注意力模块，第 5 组是 FMGNet 在材料数据集上的分割结果。

表 3.4: 消融实验。探究残差连接、多尺度特征融合模块、图注意力模块和解码器中卷积层对于网络性能的影响。编码器是基于 VGG16，评价指标为 MIoU (%)。

Methods	Spheroidite	Wood	Pearlite1	Pearlite2	Ceram
FMGNet-Decoder	90.6920	89.0181	91.8056	90.8772	86.8469
FMGNet-Residual	92.7541	90.1095	92.4580	90.0861	85.7249
FMGNet-GAM	93.2660	90.8698	92.9921	90.9822	86.3872
FMGNet-MFF	92.9712	90.8214	93.0137	89.0103	84.6033
FMGNet	94.7262	90.9345	93.6414	91.8192	88.2500

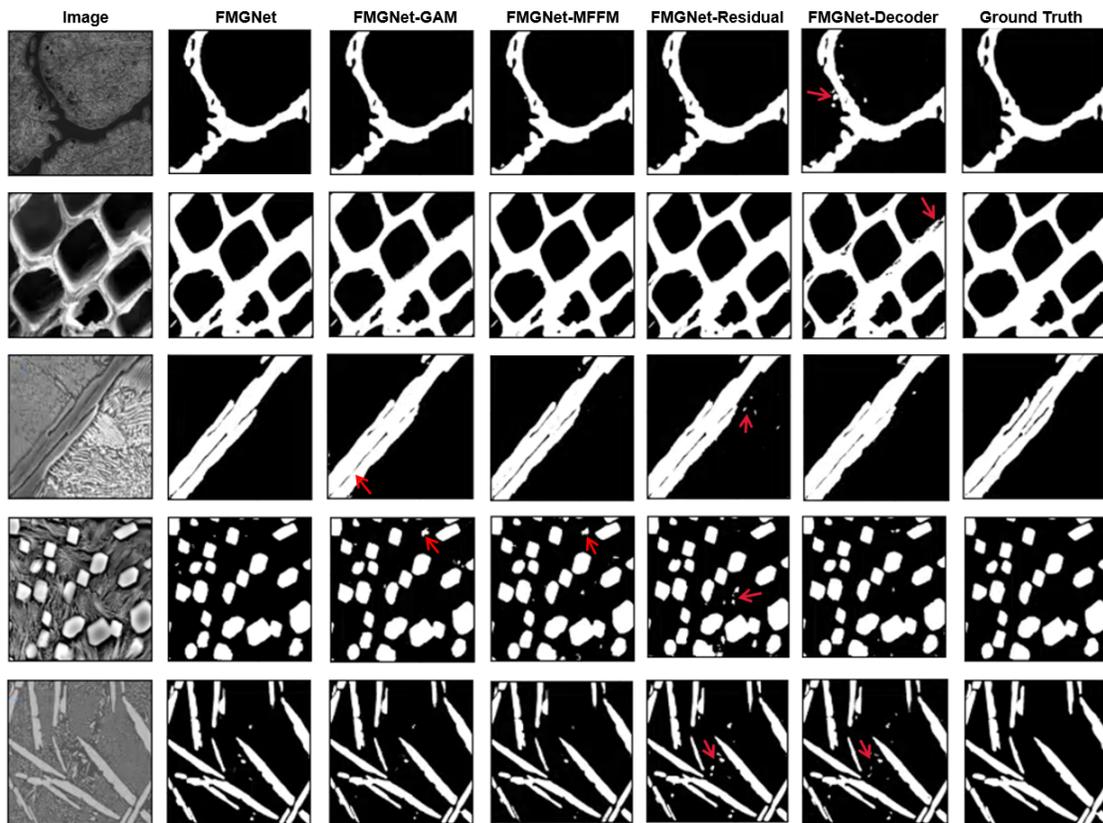


图 3.11: 消融实验。探究残差连接、多尺度特征融合模块、图注意力模块和解码器中卷积层对于网络的分割结果的影响。图中箭头是指分割结果包含大量噪声的具体位置。

表3.4展示了五组消融实验的 MIoU。从表格中可得，每个改进部分都对模型性能有着不同的积极影响。例如在 Spheroidite 数据集中，FMGNet 去除 Decoder 中的卷积层，性能下降了 4 个百分点。在 Ceram 数据集中，FMGNet 去除残差连接性能下降了 2.5 个百分点，FMGNet 去除图注意力模块下降了将近 2 个百分点。在 Pearlite2 数据集中，FMGNet 去除多尺度特征融合模块，性能下降了将近 2 个百分点。

图3.11展示了消融实验的可视化结果，可以发现多种消融实验的分割结果都包含了大量的噪音，已用箭头指示出来。实验发现 FMGNet 中残差连接和 Decoder 操作可以去除明显的噪点和瑕疵，使得材料结构中某一相完整且平滑的分割出来。FMGNet 与去除残差连接和 Decoder 卷积层的实验结果对比表明，经典的编码器-解码器性能通常低于 FMGNet，关键原因是他们的上采样特征图更关注局部特征图的信息提取。相比于传统的 VGG 作为编码器，残差连接使信息可以在高层和低层之间顺利传导，提高了网络的泛化能力。在解码器中，FMGNet 有效地融合编码器中高分辨率语义特征。

图注意力是对细节的把控更加精致。见图3.11第一行的第 3 张图和第三行的第 3 张图，可以看出一些孤立的或者毛刺的边界得到了更好的雕琢，一些较难分割的边界都得到了分离。这说明，基于空间信息的权重矩阵加强了相似度高的节点之间的聚合，抑制了相似度低的节点之间的信息传递。FMGNet 明显对材料结构的细节补充十分有效，一些结构瑕疵和粗糙的边界问题都得到了很好的修补。

多尺度特征融合起到了去除微小噪点的修饰作用，特别是边界附近的小噪点，见图3.11第一行的第 4 张图。多尺度特征融合模块分别经过不同的融合路径集成了不同层、不同空洞指数的语义特征，再经过四个池化层。该模块利用混合空洞卷积和不同分辨率的特征图很好的处理小物体和大物体之间的关联。

表 3.5: 消融实验。在图注意模块中, 研究了空间注意力、通道注意力和图卷积对网络性能的影响, 评价指标为 MIoU (%)。

Methods	Spheroidite	Wood	Pearlite1	Pearlite2	Ceram
FMGNet-GCN	92.9075	90.5226	93.3624	91.2454	86.0154
FMGNet-DA	94.0909	90.3254	93.4447	91.2344	85.3344
FMGNet-PA	93.7860	90.4208	92.8115	89.4884	87.6865
FMGNet-CA	92.9402	89.3765	93.1495	92.2886	85.0421
FMGNet	94.7262	90.9345	93.6414	91.8192	88.2500

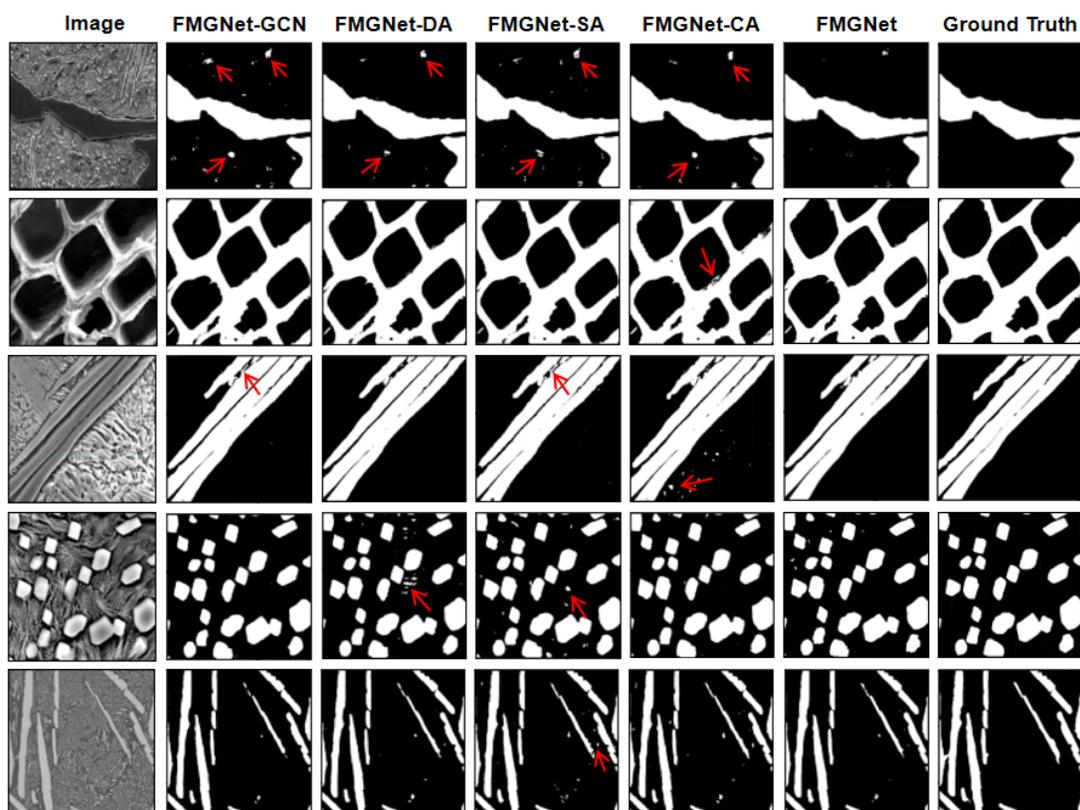


图 3.12: 消融实验。在图注意模块中, 研究了空间注意力、通道注意力和图卷积对网络性能的影响。图中箭头是指分割结果包含大量噪声的具体位置。

为了验证所提出的基于双重注意力的图卷积模块的有效性，本文在五个材料数据集上进行了消融实验。图注意模块分为三个部分：空间注意力（Spatial Attention, SA）、通道注意力（Channel Attention, CA）和图卷积（GCN）。本文做了五个实验。第一组的网络模型是 FMGNet 去除图卷积，第二组是 FMGNet 去除空间注意力和通道注意力两个部分（Dual Attention, DA），第三组是 FMGNet 去除空间注意力，第四组是 FMGNet 去除通道注意力，第五组是 FMGNet。

表3.5展示了图注意力模块的消融实验的 MIoU。图3.12展示了有关图注意力模块消融实验的可视化结果。图注意力模块在测试集上对整个网络的 MIoU 和分割效果都有积极的影响，使网络在五种材料图像上大多提升了 1-3 个百分点。可视化结果可以看出，图注意力模块的加入减少了大量的噪声，使材料结构更好的学习。在图3.12的第一行中，可以明显的看到，不管去掉图注意力模块的任何一部分，图像都呈现出大量的噪声点。然而，当图注意力模块完全加入到网络中时，噪声点明显消失了很多。

3.4 本章小结

在小样本材料图像的前提条件下，材料图像纹理复杂，缺乏成熟的深度学习技术解决材料图像分割问题。为此，本章利用图卷积灵活的连接方式和节点不消失的优势，结合双重注意力提出了基于图卷积结合多维特征融合的网络框架。该网络通过编码器对图像的低层特征像素值进行提取与分析，从而获得高阶语义信息；通过图注意力模块增强特征图的通道和空间双重注意力并且根据注意力权重重点聚合局部的特征；通过多尺度特征融合模块将编码器特征图中不同感受野和不同尺度的信息进行融合；通过解码器收集语义信息，并将同一类别结构的标签对应到相应的像素点上。

本方法利用图像训练得到十分准确的分割结果，主要有两个优点：

(1) 提出了一种新型的图注意力模块，在通道和空间维度保持更丰富的分辨率的特征融合的同时，将图卷积结合双重注意力机制加强对重要特征的关注。

(2) 提出了一个通用的应用于小样本材料图像的网络模型，解决一般材料图像

分割困难的问题。

但是，本章提出的方法也存在不足。第一，由于图卷积本身不适宜层数过多的特点，图卷积层数的加深会导致网络性能的降低。第二，加入多个以卷积操作为基础的模块，无法避免像素信息和空间信息的损失。本研究后续会为了解决以上问题而进一步探索。

第四章 基于跳跃连接的 Graph-UNet

在上一章，本文基于全卷积神经网络提出了一种新的基于图卷积结合双重注意力并融合了残差连接和多尺度融合技术的网络模型，可以有效的探索材料图像的上下文特征信息和图片中的细节信息。然而，过多的卷积操作不可避免伴随着材料图像中材料结构间的空间信息和像素信息的丢失。对于小样本的材料数据集而言，网络结构不应该存在过多的卷积操作。为了不增加网络的卷积操作的同时提升网络的表达能力，在不改变特征图尺寸的同时增加图注意力，本文在第四章提出了基于跳跃连接的图注意力网络，摒弃双重注意力机制，减少卷积导致的空间和像素信息损失，以多维特征融合和跳跃连接的思想将多层图卷积和图注意力有利的组合，解决材料图像分割问题。针对图像与图结构的数据形式的不一致问题，本文还提出了一种新颖的图编解码框架，实现特征图与图结构的互相转换。

4.1 方法概述

为了不增加网络的卷积操作的同时提升网络的表达能力，不改变特征图尺寸和空间位置信息的同时增加图注意力，本方法结合 UNet 和图卷积神经网络提出了一个通用的网络模型，Graph-UNet。相比上一章的网络模型，本章方法摒弃了以卷积层为主的双重注意力机制，利用图注意力层提升对重要节点的注意力，将卷积神经网络中常用的跳跃连接和多维特征融合的思想迁移到图卷积神经网络中来解决材料图像的语义分割任务。该模型将卷积神经网络中处理的特征图构建为一个图结构，使用跳跃连接融合多维特征的图注意力模型（Graph Attention Model，GAM）在邻居节点之间传播特征，再将信息更新后的图结构封装为特征图进行后续的卷积层操作。

本章的网络包括五个部分：编码器、图编码器、图注意力模块、图解码器和解码器。首先，输入图像通过编码器获得高级特征图；其次，图编码器将特征图转化为图结构；然后，图注意力模块融合不同维度的节点特征；随后，图解码器将图结构转化为特征图；最后，解码器输出预测图。网络结构见图4.1。

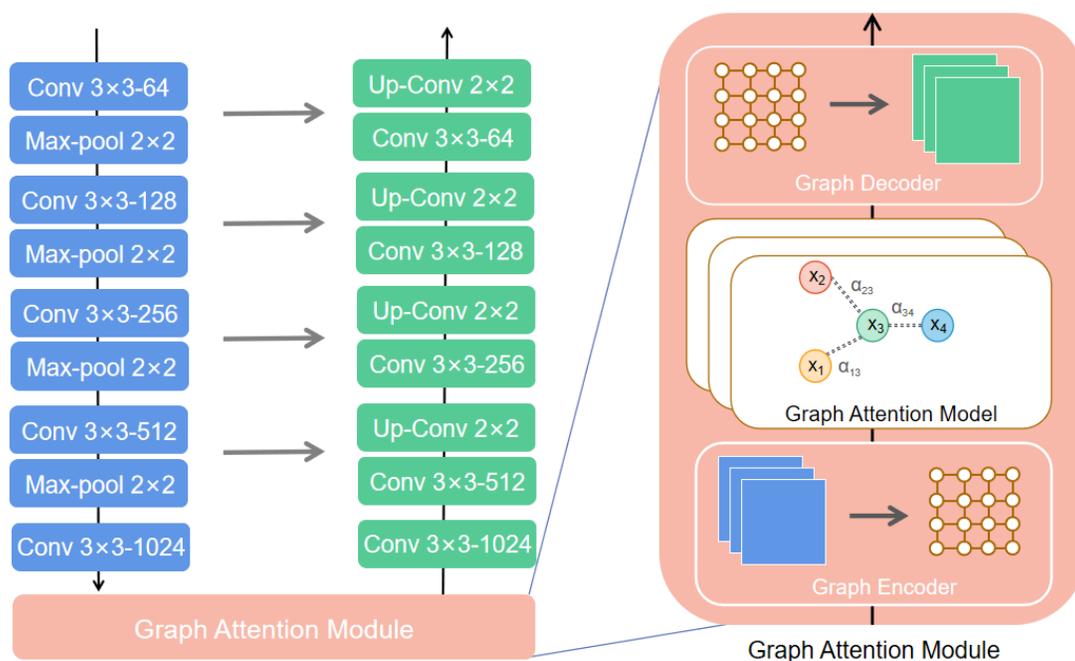


图 4.1: 网络结构。

4.2 语义分割中的图模型

本章方法将基于跳跃连接和多维特征融合的图注意力模块加入 UNet [34] 中解决材料图像的语义分割问题。本方法选择 UNet 作为骨干网络的初衷是因为 UNet 的提出最开始应用于医学图像，而材料图像与医学图像有很大的相似之处，例如纹理复杂。现在，对于 UNet 经典的改进方法有 UNet++ [69] 和 Attention-UNet [72] 等。UNet++ 主要通过增加跳跃连接增强对图像特征的补充，而 Attention-UNet 通过在解码器部分增加注意力机制提高网络性能。本方法受到图卷积和注意力机制在图像分割任务中应用的启发，提出了图注意力模块。相比于上一章的方法，本章采用不会因为图卷积操作而改变图结构大小和位置关系的图注意力。

网络结构如图4.1所示，输入图像首先经过五层卷积层，前四层随后各加一层池化层。在第五层卷积层之后，特征图进入基于跳跃连接和多维特征融合的图注意力模块，图编码器将提取的特征图转化为图结构，图注意力模块进行节点特征的传播和学习，图解码器将节点特征更新后的图结构再转化为特征图。随后，网络将特征图放

入后面的四层上采样层，不断地将特征图的尺寸放大。每一次上采样操作后都会将当前的特征图和编码器中相同尺寸大小的特征图进行拼接。最后，网络通过 Sigmoid 函数输出预测图，得到分割结果。

4.2.1 图编码器

图注意力模块的输入必须是图结构。为了将特征图输入到图注意力模块，本文创建了图编码器把特征图转化为能够执行图卷积和图注意力层操作的图结构。

在模型中，节点特征由 UNet 的编码器部分初始化。UNet 编码器部分得到步幅为 16 的特征图，尺寸为 $H \times W \times C$ ， H 和 W 是特征图的高和宽， C 是通道维度。本文把特征图依序划分为有 $H \times W$ 个节点的图结构，每个节点的特征维度是 $1 \times C$ ，节点特征包含了在局部接受域上提取的特征。

在图模型中，用邻接矩阵表示节点间边的连接，每个节点都连接到它最近的节点。本文根据像素点的空间位置关系让每个节点按照四邻域的方式建立边的连接。节点间的连接表示特征可以通过图注意力模块中的边进行传播。

4.2.2 基于跳跃连接的图注意力模块

本文提出了一种基于跳跃连接的图注意力模块，该模块起到了关注重要节点的作用，增强重要节点的特征流动，抑制无关节点的特征流动。模块结构如图4.2所示。

GCN [102] 是处理图结构的深度学习方法之一，它的具体实现公式如下：

$$Z^{l+1} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} X W^l) \quad (4.1)$$

X 是由 UNet 编码器的输出转换的图结构。 D 和 A 分别是 X 的度矩阵和邻接矩阵。 \sim 是指图的每个节点添加了自连接，并构建了相应的度矩阵 \tilde{D} 和邻接矩阵 \tilde{A} 。 W^l 是第 l 层图卷积的权值矩阵。 σ 是 ReLU 激活函数。 Z^{l+1} 是第 $l+1$ 层图卷积操作的输出。

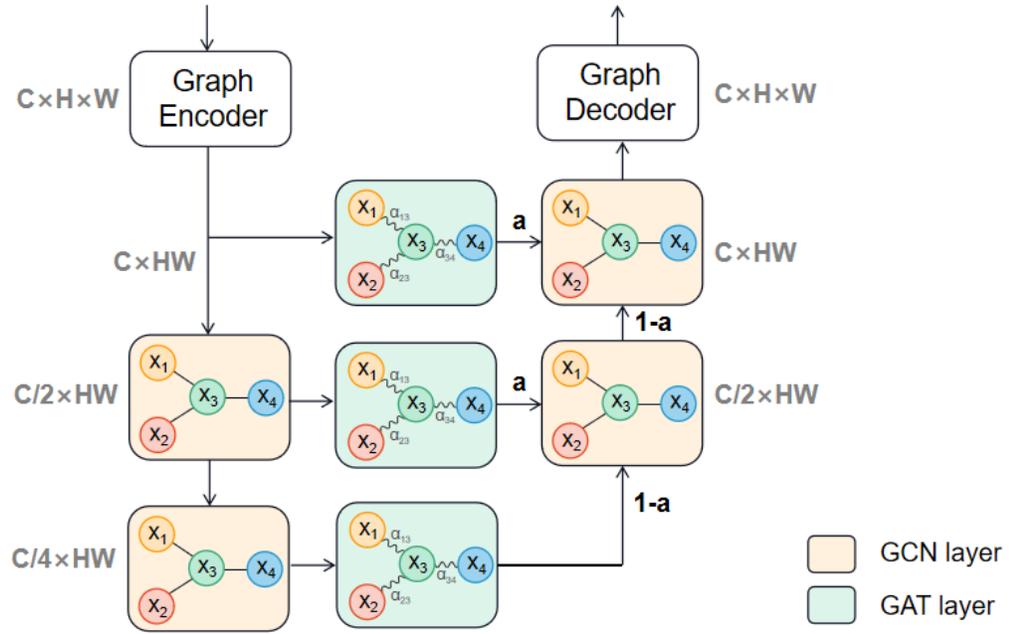


图 4.2: 图注意力结构。

GAT [103] 是应用于图结构的注意力机制之一，它的具体实现公式如下：

$$\vec{h}_i(K) = \sigma\left(\frac{1}{K} \sum_{k=1}^K \sum_{j \in N_i} \alpha_{ij}^k W^k \vec{h}_j\right) \quad (4.2)$$

K 表示注意力头的个数， N_i 是节点 i 的邻居节点集合， α_{ij}^k 是第 k 个注意力头中节点 i 和它的第 j 个邻居节点之间的注意力系数， W^k 第 k 个注意力头的权重矩阵， \vec{h}_j 是第 j 个节点的特征向量， σ 是激活函数， \vec{h}_i 是节点 i 经过多头图注意层的聚合后的输出特征。其中 α_{ij} 的具体公式如下：

$$\alpha_{ij} = \frac{\exp(LR(\vec{a}^T [Wh_i || Wh_j]))}{\sum_{r \in N_i} \exp(LR(\vec{a}^T [Wh_i || Wh_r]))} \quad (4.3)$$

LR 是 *LeakyReLU* 激活函数， \vec{a} 和 W 都是权重矩阵， h_i 、 h_j 和 h_r 分别是节点 i 、节点 j 和节点 i 的第 r 个邻居节点。

GCN 和 GAT 分别利用拉普拉斯矩阵和注意力系数，聚合邻居节点的特征到中心节点，实现节点特征的更新。GCN 在卷积时对所有邻居节点赋予相同的权重，不能根据重要程度分配不同的权重。然而，GAT 是将节点特征之间的相关性很好的融

合到了模型，根据分配不同的注意力值来进行更新节点的表示，是局部网络的运算。为了避免单次注意力的结果不准确，GAT 利用多头注意力以稳定学习过程，多头注意力的思想是利用 K 个独立的关注机制来计算隐藏状态，然后以拼接或计算平均值的方式结合每个关注机制的特征，实现操作如图4.3所示。

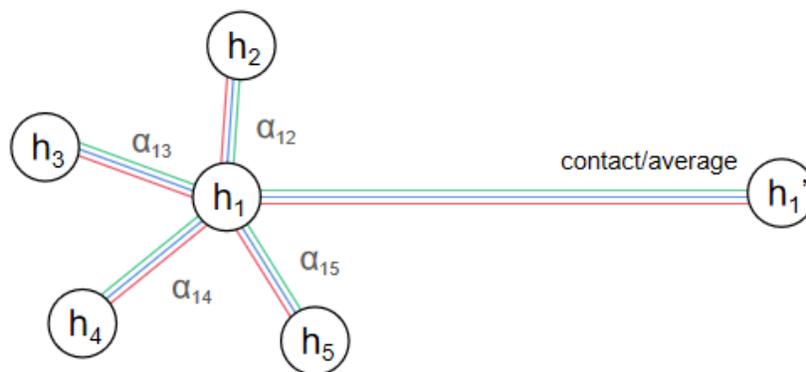


图 4.3: 图注意力的多头注意力机制。图中 h_1 、 h_2 、 h_3 、 h_4 和 h_5 分别是节点特征，经过不同的关注机制计算注意力系数 α_{ij} ，然后将节点特征连接起来计算更新后的节点特征 h_1' 。GAT 利用多头注意力以稳定学习过程，它应用 K 个独立的关注机制来计算隐藏状态，然后将其特征连接起来或计算平均值。

一般来说，图卷积神经网络解决节点分类或分割问题采用一到两层，过多层的图卷积层堆叠会导致过平滑的问题。一层图卷积相当于对一阶邻居节点信息进行聚合，而多层堆叠就相当于不断的增大聚合的半径，类比 CNN 相当于感受野不断的增大。当图卷积层数达到一定深度时，图中的一个节点就几乎包含全图的信息了，这种全局的信息相对有利于图层面的任务，例如图分类，但现实还有很多任务仅需要探索图的局部关系并不需要很深层。

然而，卷积神经网络模型中，一般深层的卷积层堆叠会展现更好的性能和效果。更深的模型，意味着更好的非线性表达能力，有潜力学习更复杂的转换，从而适合更复杂的特征输入，例如 ResNet [32] 和 DenseNet [125] 等。为了借鉴 CNN 的思想，本方法通过增加超参数 a 控制每层节点特征输入下一层图卷积层的比例，加深了图卷积模型的深度。这种思想借鉴了 ResNet 的残差连接，但这种连接仅部分缓解了过度平滑的问题。当网络堆叠更多层时，模型的性能仍然会下降。

基于跳跃连接的图注意力模块具体结构见图4.2。该模块使用两层 GCN 一边对

节点特征进行消息传播，一边对节点特征降维操作，再通过两层 GCN 对节点特征升维操作且加和降维过程中相应维度的节点特征。首先，输入图结构有 $H \times W$ 个节点，每个节点特征维度为 C 。该模块经过一层图卷积层，将节点特征的维度降到 $C/2$ 。然后，再将节点特征更新过一次的图输入第二层图卷积层，节点特征维度降为 $C/4$ 。紧接着，对图进行升维操作，分别再进入两层图卷积层。降维和升维后的特征一一对应，每次降维后的节点特征连接一层图注意力 GAT 再加和到升维的图卷积层。节点维度为 C 、 $C/2$ 和 $C/4$ 的图结构分别对应图像领域中的高分辨率、中分辨率和低分辨率的特征图。通过代表低分辨率特征图的图结构学习到粗糙的特征，代表高分辨率特征图的图结构学习到精细的细节特征，将不同分辨率的图结构融合达到对图像的准确学习和表示。在图经过图注意力层后本模块通过设置一个超参数 a 控制节点特征流动到下一层图卷积层的比例。设置超参数 a 的目的是为了避免过多的图卷积层导致节点特征最终都几乎倾向等于全局特征。特征升维部分的具体融合实现方式见公式4.4。

$$H_{add}^{l+1} = a \cdot \vec{h}^l + (1 - a) \cdot (Z^{l+1}) \quad (4.4)$$

H_{add}^{l+1} 是图注意力层输出和图卷积层升维操作的输出的加和结果， Z^{l+1} 是第 $l+1$ 层图卷积的输出特征， a 是超参数， \vec{h}^l 是图注意力层的输出。

4.2.3 图解码器

图注意力模块的输出依然是图结构。为了将图注意力模块的输出可以继续放入卷积神经网络中继续学习和训练，本文创建了图解码器将图结构转化为特征图。

图解码器将图注意力模块输出的图结构维度规模由 $C \times HW$ 转化为 $C \times H \times W$ 。抛弃图编码器构建的邻接矩阵，只保留节点特征，将节点特征按照原先的排序还原成特征图。还原后的特征图以备接下来的卷积神经网络使用。

基于跳跃连接的图注意力模块是网络模型的重要组成模块，是图卷积和图注意力组合的结果。它主要有三个优点：1) 通过跳跃连接和多维度特征融合的思想加强网络对特征信息的学习能力；2) 图注意力的使用，增加了对重要特征的关注；3) 图

编解码器可以实现特征图和图结构的相互转换，以便图卷积和图注意力层可以无障碍的应用于语义分割任务。

4.3 实验和讨论

4.3.1 实验环境与评价标准

在实验中,本文使用的计算机配置为 Intel(R) Core(TM) i7-7700K CPU@4.20Hz, NVIDIA GeForce GTX 1060 3GB, 以及 Python3.8 的开发工具。本文的网络是通过 Pytorch 实现的。

本文将图片大小统一为 512×512 的大小作为网络输入,将二元交叉熵损失函数作为网络的损失函数,函数计算公式如 (4.5)。本文使用带动量的梯度下降优化算法 Adam 训练网络模型,动量和衰减系数为 0.9 和 0.0005,迭代次数为 100,批处理大小为 1,初始学习率设为 0.001。

$$L(y|q) = -y\log(q) - (1 - y)\log(1 - q) \quad (4.5)$$

y 是真实样本的标签, q 是预测样本标签, \log 是指对数, $L(y|q)$ 是二元交叉熵损失函数的输出值。当真实样本标签和预测样本标签数值相等时,函数输出为 0;当真实样本标签和预测样本标签数值不相等时,函数输出不为 0。网络训练的目标是将损失函数的输出值无限趋向 0。

表 4.1: 数据集分布情况。

Class	Spheroidite	Wood	Pearlite1	Pearlite2
train set	4	4	4	3
test set	2	2	2	1
data set	6	6	6	4

实验的数据集一共四类材料结构,分别是 Spheroidite, Wood, Pearlite1 和 Pearlite2。其中, Spheroidite, Pearlite1 和 Pearlite2 来源于 [4]。材料数据集规模较

小，数据集分布见表4.1。本文主要用 MIoU 作为语义分割网络的评价指标。

4.3.2 现有方法和 Graph-UNet 的对比实验

为了进一步证明提出的模型有效的应用在材料图像分割任务中，本文比较了当前先进的方法，其中传统的图像分割方法，包括 KMeans [20] 和 Watershed [126] 的分割结果图，还有深度学习方法，包括 FCN [30]、UNet [34]、UNet++ [69]、CENet [77]、R2U-Net [122]、Attention-UNet [72] 和 UNet3+ [70]。

表 4.2: 与现有方法的对比实验。Graph-UNet 与主流或者先进的图像分割方法进行对比实验，主要对比传统方法和深度学习方法，评价指标为 MIoU (%)。

Methods	Spheroidite	Wood	Pearlite1	Pearlite2	Average
KMeans [20]	41.5487	88.4293	32.3180	58.7838	55.2700
Watershed [126]	91.8501	63.6468	58.7721	35.0380	62.3268
FCN [30]	94.3108	88.3342	92.4786	88.1496	90.8183
UNet [34]	95.6969	89.3567	93.7345	91.6389	92.6068
UNet++ [69]	96.1455	89.9543	94.1003	91.4282	92.9071
CENet [77]	94.0800	89.0585	93.3075	91.3135	91.9399
R2U-Net [122]	83.2993	88.2415	85.8098	87.0114	86.0905
Attention-UNet [72]	91.9393	91.2300	94.8990	89.8639	91.9831
UNet+++ [70]	92.8708	89.2140	93.6977	87.5113	90.8235
Graph-UNet	96.9200	90.3073	94.6179	91.5946	93.3600

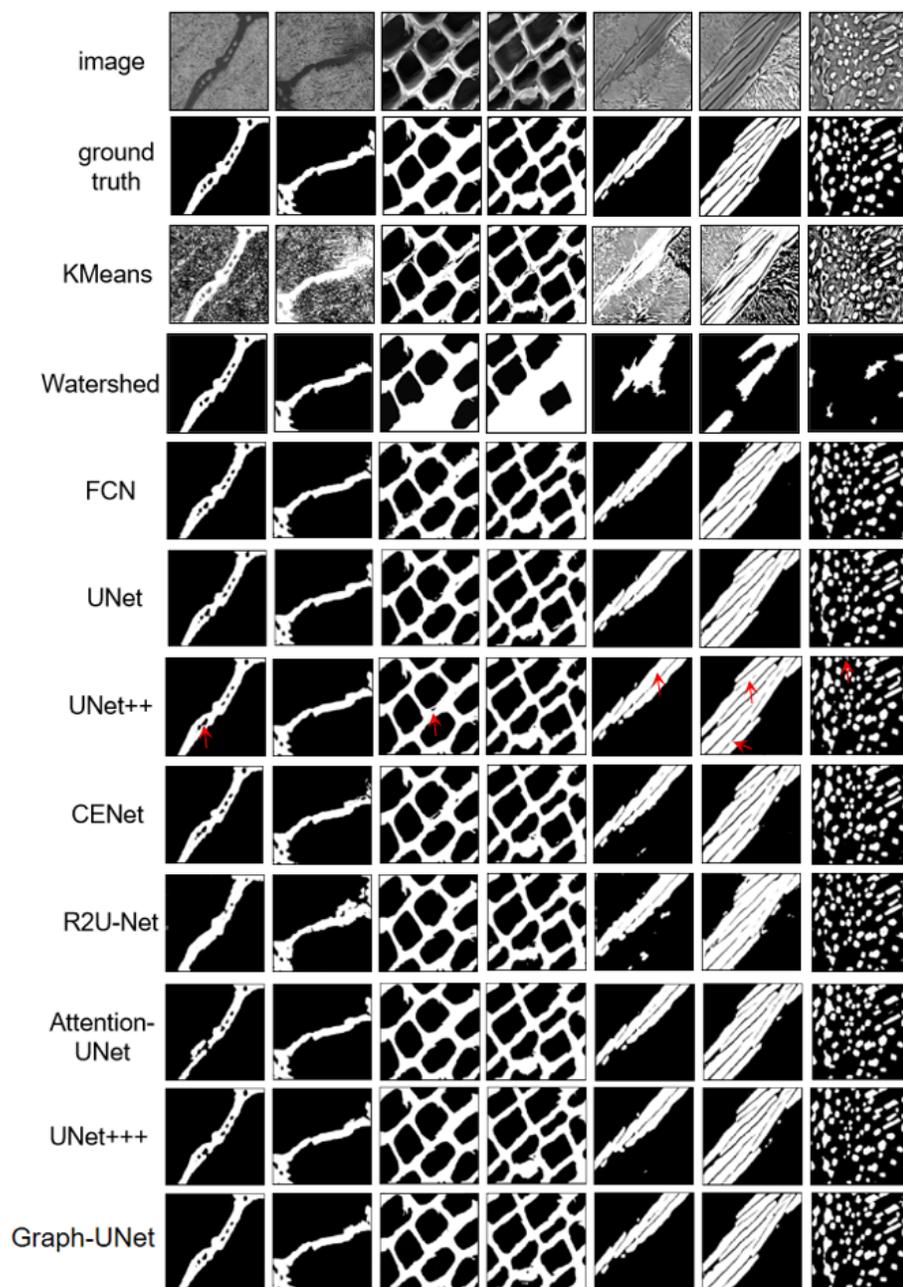


图 4.4: 现有方法对比可视化结果。图中分别是原始输入图像、输入图像的标注图, 传统方法中 KMeans [20] 和 Watershed [126] 的分割结果图, 还有深度学习方法 FCN [30]、UNet [34]、UNet++ [69]、CENet [77]、R2U-Net [122]、Attention-UNet [72] 和 UNet3+ [70] 的分割结果图 (图中箭头指出 UNet++ 出现明显分割错误的位置)。

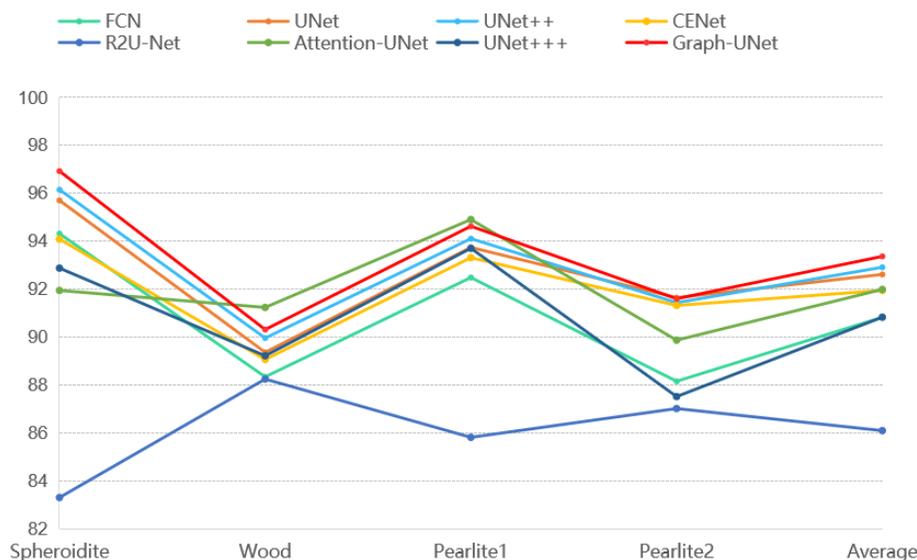


图 4.5: 深度学习方法的对比实验折线图。折线图的曲线表示 8 种深度学习方法, 包括 FCN [30]、UNet [34]、UNet++ [69]、CENet [77]、R2U-Net [122]、Attention-UNet [72]、UNet3+ [70] 和本章方法。纵轴表示 MIoU (%), 横轴表示四个材料图像的测试集和平均 MIoU。

表4.2展示了本章方法与其他方法在测试集上的 MIoU。图4.4展示了本章方法与其他方法在测试集上的可视化结果（图中箭头指出 UNet 明显分割错误的位置）。图4.5以折线图的方式展示了深度学习方法在测试集上的 MIoU 和平均 MIoU。KMeans 对于噪声比较敏感, 容易出现过分割, 如图4.4所示, KMeans 分割结果中绝大多数纹理的细节特征都被分割出来, 不能得到完整的材料结构, 各相被过度分割。Watershed 的分割结果悬殊较大, 如图4.4中 Watershed 第 1 和 2 图, 很好的分割出两相, 边界部分也分割的十分清晰, 但是如图4.4中 Watershed 第 5、6 和 7 图, 基本没有分割出目标结构甚至出现了分割图大面积呈现黑色的错误结果。由可视化结果可得, Watershed 对于对比度不强烈的灰度图, 难以实现较好的分割效果。KMeans 和 Watershed 在材料数据集评价指标的数值上都较低, 适用范围较窄, 在图像分割任务上缺乏通用性。基于深度学习的网络模型在小样本材料图像分割任务上的表现呈现出多元化。FCN 的实验结果出现了不同程度的噪声, 如图4.4中 FCN 的第 3 张图, 边界附近出现了许多微小但明显的噪声, 如图4.4中 FCN 的第 7 张图, 在块状的材料结构存在明显的颜色差异, 即中心趋向黑色和边缘呈现白色, FCN 没有很好的将

属于同一块的结构分割为一块，分割出很多不存在的边界。由可视化结果可得，FCN 没有很好的结合上下文信息，分割结果较为粗糙，对于边界附近的特征学习还有所欠缺。UNet、UNet++ 和 UNet+++ 的分割结果都不错，但是三个网络相比之下，UNet+++ 的分割可视化效果略欠一筹，出现了一些噪声。可视化结果说明，对于材料数据集来说，过多的跳跃连接不一定是百利而无一害。虽然增加跳跃连接会加强不同层的特征相互改善和弥补损失的图像信息，但是过于追求增加跳跃连接会导致一些细节信息的误判。CENet 的评价指标和分割图都趋于稳定，但依然存在小范围的分割错误，如图4.4中 CENet 的第 2 张和第 7 张图。R2U-Net 的实验结果出现了局部分割错误，如图4.4中 R2U-Net 第 1、2、5 和 6 张图，R2U-Net 对于特征的学习和训练与真实标签偏差较大，错误的分割了材料结构的形状。Attention-UNet 的分割结果也出现了层次不齐，在评价指标 MIoU 上，明显第 1 个和第 4 个数据集的 MIoU 偏低，另外两个数据集偏高。Attention-UNet 在不同数据集上展现不太稳定的分割效果，如图4.4中 Attention-UNet 的第 1 张、第 2 张和第 7 张图，Attention-UNet 没有正确的分割出结构的轮廓，主要呈现出将白色相预测为黑色相，然而其他几张图则分割的非常优秀。更详细的统计可以看表4.2，Graph-UNet 在 Spheroidite, Wood, Pearlite1 和 Pearlite2 测试集上的 MIoU 分别为 96.9200%、90.3073%、94.6179% 和 91.5946%，大多超过现有方法 2-5 个百分点左右。Attention-UNet 的 MIoU 在 Wood 和 Pearlite1 数据集表现突出，但是其他数据集和平均 MIoU 上 Graph-UNet 的表现更优异。UNet++ 的平均 MIoU 最接近 Graph-UNet，但是可视化结果已用箭头指出 UNet++ 明显分割错误的地方，在边界分割和细节处理的部分，Graph-UNet 的效果更优。从折线图4.5中可以看出，Spheroidite 数据集上 Graph-UNet 的 MIoU 是最高点，在 Pearlite2 上 Graph-UNet 和 UNet++ 等折点十分接近都处于最高位置的微小范围内，在 Wood 和 Pearlite1 数据集 Attention-UNet 是最高点，Graph-UNet 处于稍下方，但是在平均 MIoU 的一系列中，本章方法处于最高点，性能最优。大多深度学习方法在小样本材料图像上都展现出了不同的缺点，能满足对边界完好分割又可以把每一类完整呈现的技术研究者们需要进一步的探索和研究。

Graph-UNet 对多类数据集的分割效果呈现稳定的性能。材料结构的大部分噪

点都被很好的过滤，对大面积的连通区域分割准确，材料结构之间的狭窄边界分割优于绝大多数方法。在小样本数据集和图像颜色对比度低的情况下，依然展现强大的学习能力和特征提取能力。本文的方法较好的权衡了分类准确率和定位精度。

4.3.3 验证图注意力模块的有效性

本文探究了基于不同骨干网络的模型性能，对比各个网络添加图注意力模块和不添加图注意力模块的网络在小样本材料图像上的分割结果，验证图注意力模块的有效性。本文对比了骨干网络为 ResNet34-UNet、ResNet34-FCN 和 UNet++ 的网络模型的性能。

表 4.3: 不同骨干网络和添加图注意力模块 (MGAM) 的对比实验。本文对比了骨干网络为 ResNet34-UNet、ResNet34-FCN 和 UNet++ 的网络在四个数据集上的 MIoU (%), 并且计算了每个网络在所有数据集的 MIoU 平均值。

Methods	Spheroidite	Wood	Pearlite1	Pearlite2	Average
ResNet34-UNet [34]	95.0883	88.0642	92.6186	90.5320	91.5758
ResNet34-UNet-GAM	95.4993	88.5567	93.6210	91.0782	92.1888
ResNet34-FCN [30]	94.1122	88.8613	93.4565	90.9899	91.8550
ResNet34-FCN-GAM	95.5784	89.5392	94.0044	91.1701	92.5730
UNet++ [69]	96.1455	89.9543	94.1003	91.4282	92.9071
UNet++-GAM	96.5721	90.0362	94.1026	91.4533	93.0411
UNet [34]	95.6969	89.3567	93.7345	91.6389	92.6068
Graph-UNet	96.9200	90.3073	94.6179	91.5946	93.3600

表4.3展示了不同骨干网络中是否添加图注意力模块 (GAM) 的对比实验的 MIoU 结果。图注意力模块对于提高网络性能有着积极的作用。在 FCN 的基础上加上图注意力模块，四个数据集的分割指标都得到了改进。在 UNet 网络模型中，基础的网络加上图注意力模块提升的 MIoU 指标幅度最大，提升了将近一个百分点。UNet 的编码器换成 ResNet34 后图注意力的提升效果略有削弱。在 UNet 基础上改进的 UNet++ 由于本身网络就很优秀，添加图注意力模块后材料图像分割也有了小幅度提升。基于不同的网络中，UNet 添加图注意力模块的网络性能最优。

为了证明提出的图注意力模块对于材料图像分割的积极作用，本文对比了以往经典的注意力方法，包括 SENet [88]、CBMA [90]、DANet [91]、CCNet [92] 和 PSANet [93]。

表 4.4: 不同注意力方法的对比实验。本文对比了 SENet [88]、CBMA [90]、DANet [91]、CCNet [92] 和 PSANet [93] 在四个数据集上的 MIoU (%)，并且计算了每个网络在所有数据集的 MIoU 平均值。

Methods	Spheroidite	Wood	Pearlite1	Pearlite2	Average
SENet [88]	93.4806	88.3570	93.1956	89.8086	91.2105
CBAM [90]	96.3472	88.2062	93.3316	90.3746	92.0649
DANet [91]	96.3901	88.3589	94.5993	90.9791	92.5819
CCNet [92]	95.8607	88.4541	92.0037	88.0546	91.0933
PSANet [93]	94.6511	89.4653	93.9191	90.2729	92.0771
Graph-UNet	96.9200	90.3073	94.6179	91.5946	93.3600

表4.4展示了五种注意力方法作为五个不同的模块和本章提出的图注意力模块分别添加到 UNet 中在测试集上的 MIoU 结果。实验结果表明单纯的通道注意力 SENet 并不能有效的关注到主要的特征，各个数据集上的 MIoU 都较低。空间注意力的辅助会让网络的性能得到提升，例如 DANet 在四个数据集上都表现了优秀的分割结果。本文提出的模块没有改变特征图的空间信息，并对特征图的每个像素进行关注，在所有的数据集都得到了很好的分割结果，平均 MIoU 也是最高的。实验证明本文提出的模块对于材料图像的分割有着积极影响。

4.3.4 消融实验

本文提出了一个新颖的基于跳跃连接的图注意力模块。为了了解图卷积、图注意力和超参数 a 对网络模型的影响，本文做了详细的消融实验，第一组的网络模型是 UNet 加一层图卷积层，第二组是 UNet 加一层图注意力层，第三组是 UNet 加上基于跳跃连接的图卷积模块 (Graph Convolution Module, GCM)，第四组是 UNet 加上没有超参数 a 的基于跳跃连接的图注意力模块，第五组是 Graph-UNet。五组实验中的图卷积模块结构如图4.6所示。

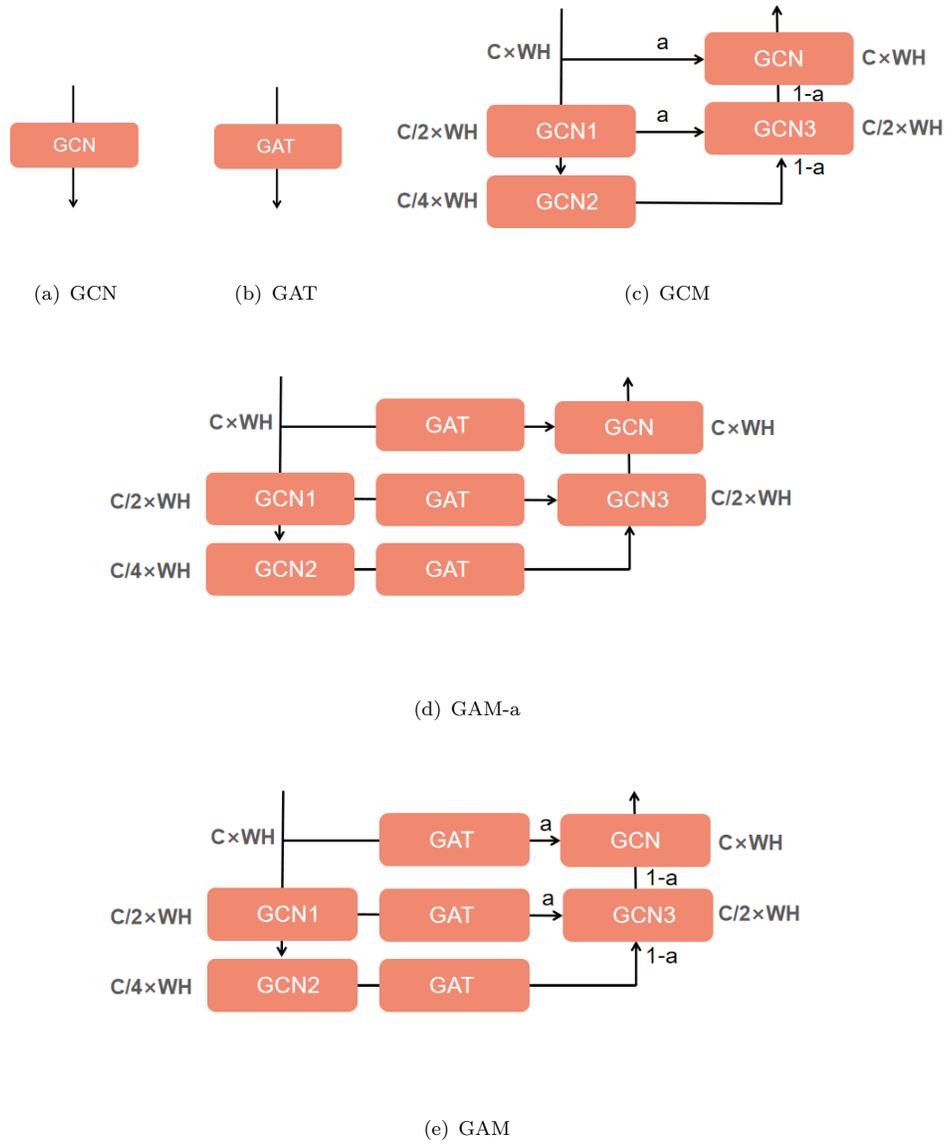


图 4.6: 消融实验中五组图卷积模块的具体结构。

表 4.5: 消融实验。探究单层图卷积、单层图注意力、跳跃连接的图注意力模块去除图注意力层和跳跃连接的图注意力模块去除超参数 a 对于网络的分割结果的影响, 评价指标为 MIoU (%)。

Methods	Spheroidite	Wood	Pearlite1	Pearlite2
UNet+GCN	96.2792	89.1921	94.2530	90.0736
UNet+GAT	92.6897	88.3697	93.9177	92.3606
UNet+GCM	95.8542	90.1897	94.5596	90.2766
UNet+GAM-a	96.4528	89.8272	93.8442	89.9251
Graph-UNet	96.9200	90.3073	94.6179	91.5946

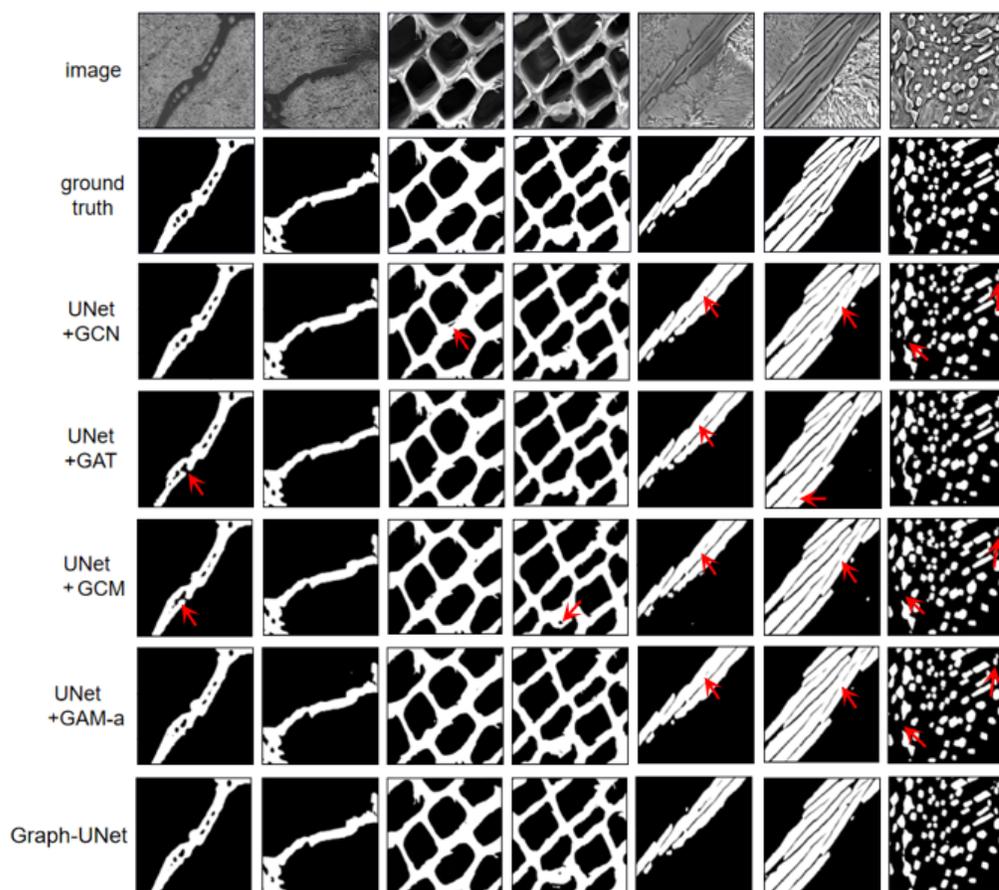


图 4.7: 消融实验。探究单层图卷积、单层图注意力、跳跃连接的图注意力模块去除图注意力层和跳跃连接的图注意力模块去除超参数 a 对于网络的分割结果的影响。图中箭头是指出现明显分割错误的位置。

表4.5展示了消融实验的 MIoU。从表格中，本文提出的基于跳跃连接的图注意力模块对于模型性能有着明显积极的影响。实验表明在 UNet 的基础上单纯加一层图注意力层，网络性能有明显的下降，但是在 Pearlite2 数据集中，对于材料结构呈现明显近似圆的块状时，分割指标比 Graph-UNet 高出将近 1 个百分点，相比其他组网络在 Pearlite2 数据集上的评价指标，高出甚至 1-3 个百分点。这说明图注意力层在一定范围内对于明显聚堆的目标起到了积极的关注作用，中心节点和邻近节点之间的注意力系数决定节点之间特征流动。图注意力层促进了明显属于一类标签的像素点之间的特征传播。相比于图注意力层，图卷积层可以更稳定的学习节点之间的特征，适用于大多数不同种类的材料结构。相比于第二组实验不包含图卷积层，其他组包含图卷积层的模型在各个数据集上比第二组模型的 MIoU 大多高出 1-2 个百分点。

图4.7展示了消融实验的可视化结果，图中箭头是指出现明显分割错误的位置。实验发现在 UNet 的基础上仅加一层 GAT 对材料结构的分割有所欠缺，如图4.7第四行的第 1 张图中，白色相出现了明显的缺失，并且对于靠近边界部分的特征学习也出现较大的错误，第四行的第 3 张图中，出现了肉眼可见的黑色小噪点。每一组的网络模型对于材料结构的分割结果都是不错的，人眼较难去发现不同组之间的区别。然而，在边界的分割中，结果发现在每一组实验的第 5 张和第 6 张图，本文提出的模块对于识别边界有很好的积极影响，第五组实验分割出的边界相比于其他组更加连续和明显。在每一组实验的第 7 图右上角小而成堆的相中，第五组的分割结果最靠近标注图。由此可得，基于跳跃连接的图注意力模块提升了对于微小细节的分割效果。

4.3.5 Graph-UNet 和 FMGNet 的对比实验

为了进一步探究本章方法 Graph-UNet 和上一章方法 FMGNet 的各自优缺点，本文在本章测试集上对比了两种方法的分割结果。

表 4.6: FMGNet 和 Graph-UNet 的对比实验, 评价指标为 MIoU (%)。

Methods	Spheroidite	Wood	Pearlite1	Pearlite2	Average
FMGNet	95.4744	90.3561	93.9524	91.0510	92.7085
Graph-UNet	96.9200	90.3073	94.6179	91.5946	93.3600

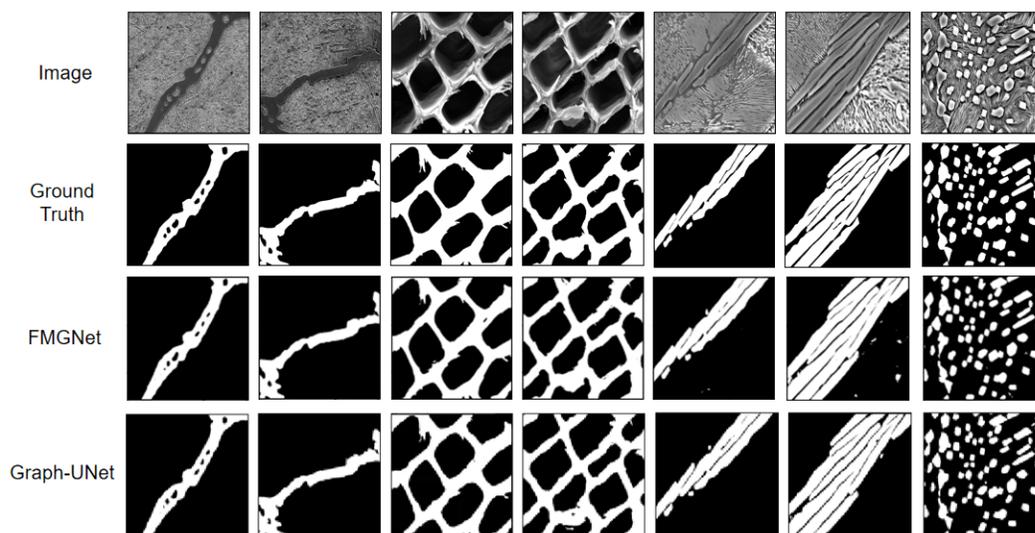


图 4.8: FMGNet 和 Graph-UNet 的对比实验。

表4.6展示了 FMGNet 和 Graph-UNet 在测试集上的 MIoU。图4.8展示了 FMGNet 和 Graph-UNet 在测试集上的可视化结果。在 Wood 数据集上, FMGNet 在 MIoU 的指标上以微弱的优势高于 Graph-UNet, 但是在 Spheroidite, Pearlite1 和 Pearlite2 这三个数据集上, Graph-UNet 都高于 FMGNet。FMGNet 的平均 MIoU 为 92.7085%, Graph-UNet 的平均 MIoU 为 93.3600%。因为材料数据集的样本数量过少, FMGNet 的网络结构虽然复杂, 但是在小样本数据上无法突显网络模型的优势, 过多的卷积层容易丢失了图像中的像素特征和空间特征, 如图4.8第 6 列图中 FMGNet 的分割结果有出现明显分割错误。Graph-UNet 在 UNet 上仅添加了一个图注意力模块, 没有增加卷积层, 避免了像素特征和空间特征的过多损失。总而言之, FMGNet 提升了网络在小样本材料图像分割任务上的性能和稳定, 但是过多的模块和卷积层使网络在小样本数据上容易丢失信息, Graph-UNet 避免添加过多的卷积操作, 不改变原始 UNet 的特征图像素信息和空间信息, 以图注意力的方式关

注重要特征提升了分割精度。在解决小样本材料图像分割问题上，Graph-UNet 的优势更加明显。

4.4 本章小结

在这项工作中，本文提出了一种针对材料图像的语义分割方法。为了加深网络模型并且不损失像素点之间的位置关系，本文将图卷积加入 UNet 中；为了学习不同分辨率的特征图，本文将 C 、 $C/2$ 和 $C/4$ 的图结构分别对应图像领域中的高分辨率、中分辨率和低分辨率的特征图进行特征融合，利用不同分辨率的特征图很好的处理小物体和大物体之间的关联；为了提升模型对重要特征的关注度，本文在不同分辨率的特征图中都加入了图注意力层加强对重要节点的学习；为了防止多层图卷积层堆叠导致节点特征出现过平滑，本文加入超参数 α ，抑制多层图卷积导致感受野过大，从而每个节点最终特征都过于趋向全局特征的融合。该方法在加深网络模型的同时，不进一步损失空间位置上的信息，利用图像训练得到十分准确的分割结果。该方法主要有两个优点：1) Graph-UNet 在不改变特征图空间位置的同时扩大感受野，结合图注意力加强对重要特征的关注。2) Graph-UNet 使用图卷积的方式加深了网络深度，提高了网络的拟合能力。

针对第三章提出方法的不足，即图卷积层的堆叠会使节点特征趋向全局化导致网络性能的降低和过多的卷积操作增加了像素信息和空间信息的损失，本文提出了基于跳跃连接的 Graph-UNet。本文基于 UNet 对小样本材料图像实现语义分割，提高了对材料各相区域的分割精确度。与第三章提出的方法相比，本方法在保持 UNet 良好分割性能的基础上，加深网络深度，不改变特征图的空间信息，提升多种材料图像的识别和分割结果。本章方法主要有两点贡献：

(1) 提出了一种新型的图注意力模块，利用跳跃连接和多维特征融合的思想将图卷积和图注意力进行堆叠，提升网络模型学习特征信息的能力。

(2) 提出了一套完整详细的应用于语义分割的图编解码模型，实现图结构和特征图的相互转换。

本章提出的方法同样也存在不足。一方面，图卷积的加入需要构建图结构额外

耗费了更多的时间，且图卷积操作必然伴随着计算量的增加；另一方面，材料结构不明显的边界细节还是不能很好的分割出来。在未来的工作中，我们会加深对于边界细节的探究。

第五章 结论和展望

5.1 结论

本文结合多种深度学习技术重点研究了利用卷积神经网络和图卷积神经网络实现小样本材料图像端到端的自动化分割方法。由于材料图像样本规模较小、结构各异、纹理复杂和色彩单调等特点，缺乏通用的分割方法解决这类问题。本文针对材料图像的特点，结合现有的理论知识分别提出两种材料图像的分割和识别方法。本文的研究成果如下：

(1) 针对材料图像的小样本和纹理复杂问题，第三章提出了基于图卷积与深度学习的材料图像分割方法。该方法利用残差连接和多尺度融合模块丰富特征图的信息，利用基于图卷积的双重注意力机制增加关键特征的关注，添加反卷积部分的卷积层提高网络的非线性表达能力。

(2) 针对小样本数据集上卷积层过多容易丢失特征的问题，第四章以 UNet 为骨干网络设计了一种基于跳跃连接的图注意力模块。该方法结合了卷积神经网络的思想将图卷积和图注意力层连接起来，致力从图结构的角度融合多种维度的节点特征，在加深加深网络深度的同时，减少像素级信息和空间信息的损失，以达到提高网络分割性能的目的。

(3) 实现了图卷积技术跨领域应用于语义分割任务。第三章提出的基于图卷积的双重注意力机制，将空间注意力的权重矩阵转化为图结构的边权重，利用灵活的节点连接方式打破了传统卷积核的局部卷积范围。第四章提出了将特征图转换为图结构的图编码器和图解码器，使卷积过程中的特征图可以结合自身维度转变为相应节点数的图结构，有利于促进图卷积神经网络应用于语义分割任务中。

本文对两种方法都进行了对比和消融实验，实验表明两种方法在小样本材料图像分割任务上都能取得较高的分割结果，具有较好的泛化性和通用性，为未来的材料图像分割和材料科学发展提供一定的贡献和帮助。

5.2 展望

本文主要结合了应用于不同数据形式的深度学习方法提出了两种网络模型解决材料图像分割问题。实验充分证明了基于深度学习技术的图像分割方法是实现自动化分割材料图像的优秀工具。为了更好的促进深度学习技术应用于材料图像任务，作者认为在以下两个方面值得深入探究：

(1) 其他思路利用图卷积解决图像分割问题。图卷积神经网络的卷积范围取决于图结构本身的连接方式，这添加了卷积核尺寸和范围的多样性，并且图卷积过程中不会改变图结构的连接方式和节点个数，避免了空间信息的损失。在语义分割任务中，图卷积的有效加持为网络模型性能提升提供了另一条巧妙的思路。

(2) 其他思路解决语义分割中的小样本问题。不论是材料图像的小样本问题，还是其他应用场景样本规模较小的情况，小样本问题一直是计算机视觉领域上的难点之一。本文结合多尺度融合和注意力机制等方法缓解了样本数据少导致的分割精度低的问题，在其他深度学习方法中，可利用元学习将问题转化为学习如何学习材料图像特征的问题实现对小样本图像的分割。

以上两个方面的深入研究能够进一步完善深度学习技术，更能推动材料图像处理在材料科学领域的飞速发展，为材料工程技术革命添砖添瓦。

参考文献

- [1] LI S, LIU Y, CHEN D, et al. Encoding the atomic structure for machine learning in materials science[J]. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 12(1):1558–1577, 2022.
- [2] XU J, LI Y, MA K, et al. Real-time observation of heterogeneous nucleation and grain growth of hypoeutectic al-si alloy inoculated by Al-Ti-Nb-B master alloy[J]. *Scripta Materialia*, 2020.
- [3] RODRIGUES J F, FLOREA L, DEOLIVEIRA M, et al. Big data and machine learning for materials science[J]. *Discover Materials*, 1(1):1–27, 2021.
- [4] DECOST B, HECHT M, FRANCIS T, et al. UHCSDB: Ultrahigh carbon steel micrograph database[J]. *Integrating Materials and Manufacturing Innovation*, 6(8):1–9, 2017.
- [5] HATAMIZADEH A, TANG Y, NATH V, et al. Unetr: Transformers for 3d medical image segmentation[C]. *In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, PP:574–584, 2022.
- [6] FU K, SHI W, KE J, et al. Image restoration and quantitative metallographic tissue based on machine vision[J]. *Journal of Electronic Imaging*, 31(5):051412, 2022.
- [7] OTSU N. A threshold selection method from gray-level histograms[J]. *IEEE Transactions on Systems Man and Cybernetics*, 9(1):62–66, 1979.
- [8] ROLF A, LENNE B. Seeded region growing[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6):641–647, 1994.
- [9] ROSENFELD A. The max roberts operator is a hueckel-type edge detector[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3(1):101–103, 1981.
- [10] KANOPOULOS N, VASANTHAVADA, BAKER R L. Design of an image edge detection filter using the sobel operator[J]. *IEEE Journal of Solid-State Circuits*, 23(2):358–367, 1988.
- [11] JOHN C, MEMBER. A computational approach to edge detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1987.

- [12] CARSTEN R, VLADIMIR K, ANDREW B. GrabCut : Interactive foreground extraction using iterated graph cut[J]. *ACM Transactions on Graphics*, 23:309–315, 2004.
- [13] MICHAEL K, ANDREW W, DEMETRI T. Snake active contour models[J]. *International Journal of Computer Vision*, 4:321–331, 1988.
- [14] ALBUQUERQUE V, TAVARES J, CORTEZ P. Quantification of the microstructures of hypoeutectic white cast iron using mathematical morphology and an artificial neural network[J]. *International Journal of Microstructure and Materials Properties*, 5(1):52–64, 2012.
- [15] YAN C, CHEN J. A watershed segmentation algorithm based on ridge detection and rapid region merging[C]. *2014 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 2014.
- [16] LIU J, CHEN J. An improved iterative watershed according to ridge detection for segmentation of metallographic image[J]. 1(1):8–13, 2012.
- [17] HAN Y, LAI C, WANG B, et al. Segmenting images with complex textures by using hybrid algorithm. *IEEE Transactions on Image Processing*[J], 28(1):13–30, 2019.
- [18] SHAO C, KAUR P, KUMAR P. An improved adaptive weighted mean filtering approach for metallographic image processing[J]. *Journal of Intelligent Systems*, 30(1):470–478, 2021.
- [19] PETERSON L. K-nearest neighbor[J]. *Scholarpedia*, 4(2):1883, 2009.
- [20] MACQUEEN J. Some methods for classification and analysis of multivariate observations[C]. *In Proc of Berkeley Symposium on Mathematical Statistics and Probability*, 1965.
- [21] SAUNDERS C, STITSON M, WESTON J, et al. Support vector machine[J]. *Computer Science*, 1(4):1–28, 2002.
- [22] BREIMAN L. Random forest[J]. *Machine Learning*, 45:5–32, 2001.
- [23] GLASS J, REDDICK W, GOLOUBEVA O, et al. Hybrid artificial neural network segmentation of precise and accurate inversion recovery (PAIR) images from normal human brain[J]. *Magnetic Resonance Imaging*, 18(10):1245–1253, 2000.

- [24] LI M, CHEN D, LIU S, et al. Online learning method based on support vector machine for metallographic image segmentation[J]. *Signal Image and Video Processing*, 28(12):1807–1817, 2020.
- [25] LAI C, SONG L, HAN Y, et al. Material image segmentation with the machine learning method and complex network method[J]. *MRS Advances*, 4:1119–1124, 2019.
- [26] ZHU J, LI X, SHI H, et al. Surface defect detection of seals based on k-means clustering algorithm and particle swarm optimization[J]. *Scientific Programming*, 2021.
- [27] SISTU G, LEANG I, YOGAMANI S. Real-time joint object detection and semantic segmentation network for automated driving[J]. *ArXiv:1901.03912*, 2019.
- [28] XU L, HOU J, GAO J. A novel smart depression recognition method using human-computer interaction system[J]. *Wireless Communications and Mobile Computing*, 2021(20):1–8, 2021.
- [29] DECOST B, LEI B, FRANCIS T, et al. High throughput quantitative metallography for complex microstructures using deep learning: A case study in ultrahigh carbon steel[J]. *Microscopy and Microanalysis*, 25(1):21–29, 2019.
- [30] LONG J, SHELFHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):640–651, 2015.
- [31] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]. In *2015 3rd IAPR Asian Conference on Pattern Recognition*, pp:730–734, 2015.
- [32] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[J]. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [33] EVERINGHAM M, ESLAMI S, GOOL L, et al. The pascal visual object classes challenge: A retrospective[J]. *International Journal of Computer Vision*, 111(1):98–136, 2015.
- [34] RONNEBERGER O, FISHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation[J]. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp:234–241, 2015.

- [35] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39:2481–2495, 2017.
- [36] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs[J]. *ArXiv:1412.7062*, 2014.
- [37] CHEN L C, PAPANDREOU G, MEMBER S, 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*, 40(4):834–848, 2018.
- [38] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[C]. *In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp:743–751, 2017.
- [39] YUAN Y, CHEN X, WANG J. Object-contextual representations for semantic segmentation[C]. *In European Conference on Computer Vision*, PP:173–190, 2020.
- [40] KRHENBVHL P, KOLTUN V. Efficient inference in fully connected CRFs with gaussian edge potentials[J]. *ArXiv:1210.5644*, 2012.
- [41] TEICHMANN M, CIPOLLA R. Convolutional CRFs for semantic segmentation[J]. *ArXiv:1805.04777*, 2018.
- [42] SANTORO A, BARTUNOV S, BOTVINICK M, et al. One-shot learning with memory-augmented neural networks[J]. *ArXiv:1605.06065*, 2016.
- [43] KOCH G, RICHARD Z, RUSLAN S. Siamese neural networks for one-shot image recognition[J]. *ICML Deep Learning Workshop*, 2015.
- [44] RAVI S, HUGO L. Optimization as a model for few-shot learning[J]. *ICLR*, 2017.
- [45] CHEN Z, FU Y, ZHANG Y, et al. Semantic feature augmentation in few-shot learning[J]. *ArXiv:1804.05298*, 2018.
- [46] NAKAMURA A, HARADA T. Revisiting fine-tuning for few-shot learning[J]. *ArXiv:1910.00216*, 2019.

- [47] SHABAN A, BANSAL S, LIU Z, et al. One-shot learning for semantic segmentation[J]. *ArXiv:1709.03410*, 2017.
- [48] WANG K, LIEW J, ZOU Y, et al. PANet: Few-shot image semantic segmentation with prototype alignment[C]. *In Proceedings of the IEEE/CVF International Conference on Computer Vision*, PP:9197–9206, 2019.
- [49] LIU W, ZHANG C, LIN G, et al. CRNet: Cross-reference networks for few-shot segmentation[C]. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp:4165–4173, 2020.
- [50] SHI J, WU K, YANG C, et al. A method of steel bar image segmentation based on multi-attention u-net[J]. *IEEE Access*, 9:13304–13313, 2021.
- [51] HUANG Z, WU J, XIE F. Automatic surface defect segmentation for hot-rolled steel strip using depth-wise separable u-shape network[J]. *Materials Letters*, 301:130271, 2021.
- [52] LI M, CHEN D, FU S, et al. Metallographic image segmentation method based on superpixels algorithm and transfer learning[C]. *In 32nd China Control and Decision Conference*, 2020.
- [53] BERTOLDO J, DECENCIRE E, RYCKELYNCK D, et al. A modular U-Net for automated segmentation of X-ray tomography images in composite materials[J]. *ArXiv:2107.07468*, 2021.
- [54] LU Y, CHEN Y, ZHAO D, et al. Graph-FCN for image semantic segmentation[J]. *International Symposium on Neural Networks*, 11554:97–105, 2019.
- [55] KIM S J, KIM S Y, LEE J, et al. A low-power graph convolutional network processor with sparse grouping for 3d point cloud semantic segmentation in mobile devices[J]. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2022.
- [56] ZHANG L, LI X, ARNAB A, et al. Dual graph convolutional network for semantic segmentation[J]. *ArXiv:1909.06121*, 2019.
- [57] Aizawa H, Domae Y, Kato K. Hierarchical pyramid representations for semantic segmentation. *ArXiv:2104.01792*, 2021.

- [58] Huang H, Lin L, Zhang Y, et al. Graph-BAS3Net: Boundary-aware semi-supervised segmentation network with bilateral graph convolution. *In Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp:7386–7395, 2021.
- [59] KRIZHEVSKY A, SUTSKEVER I, HINTON G. Imagenet classification with deep convolutional neural networks [J]. *NIPS*, 2012.
- [60] HAYOU S, DOUCET A, ROUSSEAU J. On the impact of the activation function on deep neural networks training[C]. *In International Conference on Machine Learning*, pp:2672–2680, 2019.
- [61] DUMOULIN V, VISIN F. A guide to convolution arithmetic for deep learning[J]. *ArXiv:1603.07285*, 2016.
- [62] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]. *In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, PP:6230–6239, 2017.
- [63] MUBASHAR M, ALI H, GRONLUND C, et al. R2U++: a multiscale recurrent residual u-net with dense skip connections for medical image segmentation[J]. *Neural Computing and Applications*, pp:1–17, 2022.
- [64] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[J]. *ArXiv:1612.03144*, 2017.
- [65] ZHANG X, DU B, WU Z, et al. LAANet: lightweight attention-guided asymmetric network for real-time semantic segmentation[J]. *Neural Computing and Applications*, pp:1–15, 2022.
- [66] HU J, CHEN Z, YANG M, et al. A multiscale fusion convolutional neural network for plant leaf recognition[J]. *IEEE Signal Processing Letters*, 25(6):853–857, 2018.
- [67] CHEN L, YANG Y, WANG J, et al. Attention to scale: Scale-aware semantic image segmentation[C]. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp:3640–3649, 2016.
- [68] POHLEN T, HERMANS A, MATHIAS M, et al. Full-resolution residual networks for semantic segmentation in street scenes[C]. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, PP:4151–4160, 2017.

- [69] ZHOU Z, SIDDIQUEE M, TAJBAKHSI N, et al. UNet++: A nested U-Net architecture for medical image segmentation[J]. *IEEE Transactions on Medical Imaging*, 39(6):1856–1867, 2020.
- [70] HUANG H, LIN L, TONG R, et al. UNet 3+: A full-scale connected unet for medical image segmentation[J]. *ArXiv:2004.08790*, 2020.
- [71] LIU J, HE J, ZHANG J, et al. EfficientFCN: Holistically-guided decoding for semantic segmentation[J]. *ArXiv:2008.10487*, 2020.
- [72] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention U-Net: Learning where to look for the pancreas[C]. In *The International Conference on Medical Imaging with Deep Learning*, pp:1–10, 2018.
- [73] GALDRAN A, CARNEIRO G, BALLESTER M. Double encoder-decoder networks for gastrointestinal polyp segmentation[J]. *International Conference on Pattern Recognition*, PP:293–307, 2021.
- [74] QAMAR S, AHMAD P, SHEN L. Dense encoder-decoder-based architecture for skin lesion segmentation[J]. *Cognitive Computation*, 13(2):583–594, 2021.
- [75] MENG T, GHIASI G, MAHJORIAN R, et al. Revisiting multi-scale feature fusion for semantic segmentation[J]. *ArXiv:2203.12683*, 2022.
- [76] HE S, LU X, GU J, et al. RSI-Net: Two-stream deep neural network for remote sensing images-based semantic segmentation[J]. *IEEE Access*, 10:34858–34871, 2022.
- [77] GU Z, CHENG J, FU H, et al. CE-Net: Context encoder network for 2d medical image segmentation[J]. *IEEE Transactions on Medical Imaging*, 38:2281–2292, 2019.
- [78] ZHAO X, ZHANG P, SONG F, et al. D2A U-Net: Automatic segmentation of covid-19 ct slices based on dual attention and hybrid dilated convolution[J]. *Computers in biology and medicine*, 135:104526, 2021.
- [79] ZHANG J, LU C, WANG J, et al. Concrete cracks detection based on fcn with dilated convolution[J]. *Applied Sciences*, 9(13):2686–2698, 2019.

- [80] RAVIKIRAM P, CHAKKARAVARTHY M. Improved efficiency of semantic segmentation using pyramid scene parsing deep learning network method[J]. In *Intelligent Systems and Sustainable Computing*, volume PP, pages 175–181. Springer, 2022.
- [81] ZHANG L, NAN Q, BIAN S, et al. Real-time segmentation method of billet infrared image based on multi-scale feature fusion[J]. *Scientific Reports*, 12(1):1–7, 2022.
- [82] ZHU L, JI D, ZHU S, et al. Learning statistical texture for semantic segmentation[J]. *ArXiv:2103.04133*, 2021.
- [83] TAO A, SAPRA K, CATANZARO B. Hierarchical multi-scale attention for semantic segmentation[J]. *ArXiv:2005.10821*, 2020.
- [84] BROSTOW G. Computer vision segmentation and recognition using structure from motion point clouds[J]. pp:44–57, 2008.
- [85] MOHAMMED A, DONGHYUN K. CMM-Net: contextual multi-scale multi-level network for efficient biomedical image segmentation[J]. *Scientific reports*, 11(10191):1–18, 2021.
- [86] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[J]. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [87] GUDHE A, BEHRAVAN H, SUDAH M, et al. Multi-level dilated residual network for biomedical image segmentation[J]. *Scientific Reports*, 11(14101):1–18, 2021.
- [88] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(8):2011–2023, 2020.
- [89] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp:7794–7803, 2018.
- [90] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. In *European Conference on Computer Vision*, PP:3–19, 2018.
- [91] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation[C]. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

- [92] HUANG Z, WANG X, WEI Y, et al. CCNet: Criss-cross attention for semantic segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp(99):603–612, 2020.
- [93] LIU H, LIU F, HUANG D. Polarized self-attention: Towards high-quality pixel-wise regression[J]. *ArXiv:2107.00782*, 2021.
- [94] WEN G, CAO P, BAO H, et al. Mvs-gcn: A prior brain structure learning-guided multi-view graph convolution network for autism spectrum disorder diagnosis[J]. *Computers in Biology and Medicine*, PP:105239, 2022.
- [95] HELL F, TAHA Y, HINZ G, et al. Graph convolutional neural network for a pharmacy cross-selling recommender system[J]. *Information (Switzerland)*, 11(11):525, 2020.
- [96] SONG X, ZHOU F, FRANGI A, et al. Graph convolution network with similarity awareness and adaptive calibration for disease-induced deterioration prediction[J]. *Medical Image Analysis*, 69(3):101947, 2021.
- [97] YANG H, ZHANG Z, LI Z, et al. Region-level traffic prediction based on temporal multi-spatial dependence graph convolutional network from gps data[J]. *Remote Sensing*, 14, 2022.
- [98] MAHONEY M, DRINEAS P. CUR matrix decompositions for improved data analysis[J]. *Proceedings of the National Academy of Sciences*, 106(3):697–702, 2009.
- [99] PEROZZI B, ALRFOU R, SKIENA S. DeepWalk: Online learning of social representations[J]. *SIGKDD explorations*, CD/ROM:697–706, 2014.
- [100] SCARSELLI F, GORI M, TSOI A C, et al. The graph neural network model[J]. *IEEE transactions on neural networks*, 20(1):61–80, 2008.
- [101] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. *ArXiv:1301.3781*, 2013.
- [102] KIPF T, WELLING M. Semi-supervised classification with graph convolutional networks[J]. *ArXiv:1609.02907*, 2016.

- [103] VELIKOVI P, CUCURULL G, CASANOVA A, et al. Graph attention networks[J]. *ArXiv:1710.10903*, 2017.
- [104] KIPF T, WELING M. Variational graph auto-encoders[J]. *ArXiv:1611.07308*, 2016.
- [105] WANG H, WANG J, WANG J, et al. Graph representation learning with generative adversarial nets[J]. *AAAI, Graphgan*, 2018.
- [106] YAN S, XIONG Y, LIN D. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]. In *Thirty-second AAAI conference on artificial intelligence*, 2018.
- [107] WANG C, SAMARI B, SIDDIQI K. Local spectral graph convolution for point set feature learning[c]. In *Proceedings of the European Conference on Computer Vision (ECCV)*, PP:52–66, 2018.
- [108] FEINBERG E, SUR D, HUSIC B, et al. Spatial graph convolutions for drug discovery[J]. *ArXiv:1803.04465*, 2018.
- [109] YUAN F, XU Y, LI Q, et al. Spatio-temporal graph convolutional networks for road network inundation status prediction during urban flooding[J]. *ArXiv:2104.02276*, 2021.
- [110] ZENG D, ZHAO C, QUAN Z. CID-GCN: An effective graph convolutional networks for chemical-induced disease relation extraction[J]. *Frontiers in Genetics*, 12:624307–624315, 2021.
- [111] FU H, HUANG F, LIU X, et al. MVGCN: data integration through multi-view graph convolutional network for predicting links in biomedical bipartite networks[J]. *Bioinformatics*, 38(2):426–434, 2022.
- [112] MA F, GAO F, SUN J, et al. Attention graph convolution network for image segmentation in big sar imagery data[J]. *Remote Sensing*, 11(21):2586, 2019.
- [113] YANG L, ZHUANG J, FU H, et al. SketchGCN: Semantic sketch segmentation with graph convolutional networks[J]. *ArXiv:2003.00678*, 2020.
- [114] SHI L, ZHANG Y, CHENG J, et al. Action recognition via pose-based graph convolutional networks with intermediate dense supervision[J]. *Pattern Recognition*, 121:108170, 2022.

- [115] CHEN Y, ROHRBACH M, YAN Z, et al. Graph-based global reasoning networks[C]. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, PP:433–442, 2019.
- [116] WU T, LU Y, ZHU Y, et al. GINet: Graph interaction network for scene parsing[C]. In *European Conference on Computer Vision*, pp:34–51, 2020.
- [117] HE H, CUI J, ZHA H. Boundary-aware graph convolution for semantic segmentation[C]. *2020 25th International Conference on Pattern Recognition (ICPR)*, 2020.
- [118] HU H, JI D, GAN W, et al. Class-wise dynamic graph convolution for semantic segmentation[J]. *Springer*, pp:1–17, 2020.
- [119] PAN S, LU C, LEE S, et al. Weakly-supervised image semantic segmentation using graph convolutional networks[C]. In *2021 IEEE International Conference on Multimedia and Expo (ICME)*, PP:1–6, 2021.
- [120] KETKAR N. *Introduction to PyTorch[M]*. Deep Learning with Python, 2017.
- [121] PRASAD S, PEDDOJU S, GHOSH D. Mask region grow segmentation algorithm for low-computing devices[C]. In *2016 Twenty Second National Conference on Communication (NCC)*, 2016.
- [122] CHRIS Y, MAHMUDUL H, MD Z, et al. Recurrent residual convolutional neural network based on u-net (R2U-Net) for medical image segmentation[J]. *ArXiv:1802.06955*, 2018.
- [123] LI H, XIONG P, FAN H, et al. DFANet: Deep feature aggregation for real-time semantic segmentation[C]. 2020.
- [124] YU C, WANG J, PENG C, et al. BiSeNet: Bilateral segmentation network for real-time semantic segmentation[C]. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp:325–341, 2018.
- [125] HUANG G, LIU Z, LAURENS V, et al. Densely connected convolutional networks[C]. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp:4700–4708, 2017.

- [126] BIENIEK A, MOGA A. An efficient watershed algorithm based on connected components[J]. *Pattern Recognition*, 33(6):907–916, 2000.

作者在攻读硕士学位期间取得的成果

1. 论文：应用于材料图像分割的 Graph-UNet. 计算机应用研究. (第一作者, 中文核心)
2. 论文：“Material Structure Segmentation Method Based on Graph Attention”. Neural Computing & Applications. (SCI, 导师为一作, 本人为二作, 审稿中)
3. 专利：基于多维特征融合的图注意力的材料图像分割方法, 专利号：202210318948.6. (导师排名第一, 本人排名第二)
4. 软著：复杂纹理的材料结构图像分割软件 V1.0, 登记号：2022SR0859786. (导师排名第一, 本人排名第二)

作者在攻读硕士学位期间所参与的项目

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

致谢

在这篇论文完成的同时，首先我要衷心的感谢导师韩越兴老师。在韩老师的悉心指导下，我的研究才能如火如荼的继续下去，我的论文才能如此顺利的得到完成。在研究生三年的学习中，韩老师对我提出了很多建设性的指导意见，耐心地和我探讨研究过程中出现的难题和挑战，给予专业上的指导和信心上的鼓励。我能顺利完成这篇论文和老师的认真指导与辛苦批改密不可分，在此我由衷的向韩老师表示感谢。研究生期间，韩老师渊博的专业知识，严谨的科研态度，精益求精的工作作风和严以律己的崇高风范对我的影响十分深远。“授人以鱼不如授人以渔”，韩老师不只是传授给我们知识，更是传授给我们学习知识的方法。三年中，每周一次的组会让我们对自己的学习任务更加有规划，组会上的细心指导和对科研的探讨让我们打开眼界，平时的交流和督促让我们不荒废大好的时光。我相信在未来的工作和生活中，我依然会向韩老师学习，把优秀的品质和认真的态度贯彻和延续下去。

同时我也要感谢学校给我们创造良好的学习环境，让我们能够在良好的氛围中完成自己的学业。我还要感谢我的同学们和其他老师们，在学习和生活中对我的热情帮助让我的研究生时光更加丰富多彩。我还要感谢我的家人们，大力支持我的学业让我安心的完成学业。