

中图分类号：TP391

单位代号：10280

密 级：公开

学 号：20721516

上海大学



硕士学位论文

SHANGHAI UNIVERSITY

MASTER'S DISSERTATION

题
目

面向小样本材料图像的增
强方法研究

作 者 刘宇虹

学科专业 计算机应用技术

导 师 韩越兴

完成日期 2023.05

姓名：刘宇虹

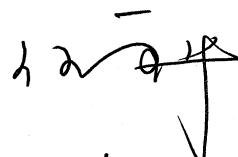
学号：20721516

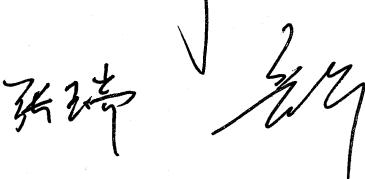
论文题目：面向小样本材料图像的增强方法研究

上海大学

本论文经答辩委员会全体委员审查，确认符合上海大学硕士学位论文质量要求。

答辩委员会签名：

主任： 

委员： 

导师： 

答辩日期： 2023.6.7

姓 名：刘宇虹

学号：20721516

论文题目：面向小样本材料图像的增强方法研究

原创性声明

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

签 名：刘宇虹 日期：2023.6.7

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：刘宇虹 导师签名：韩以光 日期：2023.6.7

上海大学工学硕士学位论文

面向小样本材料图像的增强方法 研究

姓 名：刘宇虹

导 师：韩越兴

学科专业：计算机应用技术

上海大学计算机工程与科学学院

2023 年 5 月

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

Research on Augmentation Methods for Small Sample Material Images

Candidate: Yuhong Liu

Supervisor: Yuexing Han

Major: Computer Application Technology

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

May, 2023

摘要

人工智能技术的发展，加速了机器学习、深度学习以及计算机视觉方法在各个领域的应用，推动了这些领域的进步。例如使用人工智能技术对材料数据进行处理和分析，有效促进了材料的智能化研发。然而，人工智能中深度学习方法的研究极度依赖大规模标注数据和高性能计算设备，在数据量匮乏的领域进行图像识别、图像分类等任务时面临着很大挑战。由于材料图像的获取受到采集成本、技术困难、隐私保护等多方面的限制，导致材料图像数据稀少，这也造就了材料图像的小样本特性。此外，材料图像种类繁多、微观纹理复杂、显微结构差异性大，难以使用统一模式对其进行处理。因此，如何在小样本材料图像数据集的场景下，顺利完成对复杂材料图像的处理至关重要。

为了提高小样本材料图像的处理效率和性能，本论文通过对不同小样本材料图像进行分析，融合图像增强、图像生成、深度学习等方法，分别提出了针对热障涂层形貌特征识别任务的图像增强方法和针对钢铁显微图像分类任务的数据增强方法，有效满足了各自的任务需求，具体工作如下：

(1) 对于样本数量极少的热障涂层图像，提出了基于图像增强的形貌特征识别方法。利用色彩空间转换、中值滤波、Otsu 阈值处理和形态学运算等多种操作，去除图像噪声、增大图像特征间的差异、加强图像的纹理细节和轮廓边缘，从而提升了图像质量、实现了初步的图像增强。然后借助图像作差去除热障涂层中的细微孔洞，使图像中只包含裂纹这一种形貌特征。接着使用基于闭运算的断线修复增强裂纹，最后结合骨架提取算法和遍历搜索算法完成对热障涂层中裂纹的自动识别与长度计算。该方法在数量仅有几张的热障涂层图像上进行了大量增强，以降低裂纹识别的难度，并使用传统的闭运算方法连接像素点缺失的裂纹断线，增强了裂纹特征。相比于人工识别裂纹，该方法更加便捷高效，实验也验证了方法的有效性。针对此任务，还设计开发了可供材料研究者直接使用的软件，提高了热障涂层微结构分析的研究效率。

(2) 传统的图像增强方法效果有限，对深度学习所需的大数据集来说杯水车

薪。在利用卷积神经网络对小样本材料图像分类的任务中，提出了基于单样本生成模型的材料图像数据增强方法。使用融合注意力机制的 HP-VAE-GAN (Hierarchical Patch VAE-GAN) 生成超高碳钢显微组织图像，实现对训练数据集的增广。改进的 HP-VAE-GAN 在 PatchVAE 中引入 CBAM (Convolutional Block Attention Module)，CBAM 融合了通道注意力机制和空间注意力机制，使网络的特征提取和表示能力都强于原始的 HP-VAE-GAN。实验表明，改进后的 HP-VAE-GAN 其性能优于原始模型。针对具体的超高碳钢显微组织图像分类任务，使用结合 CBAM 的 HP-VAE-GAN 生成图像，并对训练集扩增后，测试集上的分类准确率得到了显著提升，有效缓解了小样本场景下分类模型的过拟合问题。

在材料图像数据稀缺的情况下，本文提出的两种增强方法都有效辅助了目标任务的完成，为小样本材料图像的研究奠定了基础，加速了材料图像的研究进程。此外，本研究对其他学科领域内的小样本问题同样具有借鉴意义，为小样本图像的研究提供了新思路。

关键词： 材料图像处理；图像增强；数据增强；图像生成；小样本

ABSTRACT

The development of artificial intelligence technology accelerates the application of machine learning, deep learning and computer vision methods in various fields, promoting the progress of these fields. For example, using artificial intelligence technology to process and analyze materials data effectively promotes the intelligent research and development of materials. However, research on deep learning methods in artificial intelligence heavily relies on large-scale labeled data and high-performance computing devices. There are great challenges in performing visual tasks such as image recognition and image classification when data is scarce. Due to the limitations of the acquisition cost, technical difficulties, privacy protection and other aspects, material image data is scarce, which also results in the small sample characteristics of the material image. In addition, because of the variety of material images, the complexity of micro-texture, and the big differences in microstructure, it is difficult to process them in a uniform mode. Therefore, it is critical to complete the processing of complex material images on small sample material image data sets.

To improve the processing efficiency of small sample material images, this paper proposes image enhancement methods for thermal barrier coating morphology feature recognition and data augmentation methods for steel micro image classification by analyzing different small sample material images, integrating image enhancement, image generation, deep learning, and other methods. The specific work is as follows:

(1) A morphology feature recognition method based on traditional image enhancement is proposed for thermal barrier coating images with very few samples. By utilizing various operations such as color space conversion, median filtering, Otsu threshold processing and morphological operations, the image noise is removed, the difference between image features is increased, the texture details and contour edges of the image are strengthened. Thus, the image quality is improved and preliminary image enhancement is achieved. Next, with the help of image subtraction, the fine holes in the

thermal barrier coating are removed, so that the image only contains the cracks. Then, a closed operation based broken line repair is used to enhance the crack feature. Finally, the automatic identification and length calculation of cracks in thermal barrier coatings are completed by combining skeleton extraction algorithm and traversal search algorithm. This method uses extensive enhancement on only a few thermal barrier coating images to reduce the difficulty of crack detection, and uses traditional closed operation methods to connect missing pixel crack lines, enhancing crack features. Compared to manual detection, this method is more convenient and efficient, and experiments have also verified the effectiveness of the method. For this task, the software is developed for materials researchers to improve the research efficiency of microstructure analysis of thermal barrier coatings.

(2) Traditional image enhancement methods have limited effectiveness and are inadequate for the large datasets required for deep learning. In the task of using convolutional neural network to classify small sample material images, a data augmentation method based on single sample generative model is proposed. HP-VAE-GAN (Hierarchical Patch VAE-GAN) combined with attention mechanism is used to generate ultrahigh carbon steel microstructure images, expanding the training dataset. The improved HP-VAE-GAN adds CBAM(Convolutional Block Attention Module) to PatchVAE, which integrates channel attention mechanism and spatial attention mechanism. The feature extraction and representation capabilities of the network are stronger than that of the original HP-VAE-GAN. Experiments show that the performance of the improved HP-VAE-GAN is better than the original model. For the specific classification task of ultrahigh carbon steel microstructure image, the HP-VAE-GAN combined with CBAM is used to generate the image, and after the training set is expanded, the classification accuracy on the test set is significantly improved, effectively alleviating the overfitting problem of the classification model on small sample data sets.

In the case of scarce material image data, the two enhancement methods proposed

in this paper effectively assist in completing the target task, laying the foundation for the research of small sample material images and accelerating the research process of material images. Besides, this work also has reference significance for small sample problems in other interdisciplinary fields, providing new ideas for the study of small sample images.

Keywords: Material Image Processing; Image Enhancement; Data Augmentation;
Image Generation; Small Sample

目 录

| | |
|------------------------------------|-----------|
| 第一章 绪论 | 1 |
| 1.1 课题来源..... | 1 |
| 1.2 研究背景及意义..... | 1 |
| 1.3 国内外研究现状..... | 3 |
| 1.3.1 传统图像增强..... | 4 |
| 1.3.2 图像生成模型 | 9 |
| 1.3.3 材料图像增强..... | 13 |
| 1.4 论文研究内容..... | 14 |
| 1.5 论文组织结构..... | 15 |
| 第二章 相关理论与技术概述 | 16 |
| 2.1 本文相关的传统图像增强方法..... | 16 |
| 2.1.1 图像滤波 | 16 |
| 2.1.2 数学形态学 | 16 |
| 2.2 无监督图像生成模型基础..... | 17 |
| 2.2.1 生成对抗网络 | 17 |
| 2.2.2 变分自编码器 | 20 |
| 2.3 计算机视觉中的注意力机制..... | 25 |
| 2.4 本章小结..... | 29 |
| 第三章 基于图像增强的热障涂层形貌特征识别 | 30 |
| 3.1 热障涂层形貌特征识别概述..... | 30 |
| 3.2 基于图像增强的热障涂层裂纹识别..... | 32 |
| 3.2.1 孔隙轮廓增强与图像去噪 | 32 |
| 3.2.2 孔隙去除与裂纹修复 | 34 |
| 3.2.3 裂纹识别与长度计算 | 35 |
| 3.3 实验验证..... | 37 |
| 3.4 热障涂层裂纹识别软件的研发..... | 42 |

| | |
|-------------------------------------|-----------|
| 3.4.1 图像增强与裂纹细化软件的设计与实现..... | 42 |
| 3.4.2 裂纹识别与长度计算软件的设计与实现..... | 45 |
| 3.5 本章小结..... | 49 |
| 第四章 基于单样本生成模型的材料图像数据增强 | 50 |
| 4.1 材料图像数据增强概述..... | 50 |
| 4.2 融合 CBAM 的 HP-VAE-GAN..... | 51 |
| 4.2.1 模型整体架构 | 51 |
| 4.2.2 编码器和解码器 | 54 |
| 4.2.3 损失函数..... | 55 |
| 4.3 小样本材料图像数据增强实验..... | 56 |
| 4.3.1 数据集和实验设置..... | 56 |
| 4.3.2 实验分析与讨论 | 57 |
| 定性分析..... | 57 |
| 定量分析..... | 64 |
| 4.3.3 消融实验 | 66 |
| 4.3.4 CBAM 与其他注意力机制的对比..... | 66 |
| 4.4 本章小结..... | 67 |
| 第五章 总结与展望 | 68 |
| 5.1 工作总结..... | 68 |
| 5.2 未来展望..... | 69 |
| 参考文献 | 70 |
| 作者在攻读硕士学位期间取得的成果 | 81 |
| 作者在攻读硕士学位期间参与的项目 | 82 |
| 致 谢 | 83 |

第一章 绪论

1.1 课题来源

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

1.2 研究背景及意义

随着计算能力和数据收集能力的日益增强,现代科学已经进入密集数据融合人工智能的时代。图像理解、语音理解和文字理解,这三大部分基本构成了人工智能,而在这些领域中,图像理解又是核心。研究如何组织输入到计算机中的图像信息,以及识别图像中的物体和场景,并对图像内容给予解释,正是计算机视觉的核心问题。计算机视觉借助计算机及相关设备模拟生物视觉,对目标进行分割、分类、识别和跟踪,对图像进行处理、分析和理解。随着学科间的交叉融合,作为人工智能领域的重要组成部分,计算机视觉技术也逐渐被运用到材料领域的钢铁金相分析、表面缺陷检测等任务中。

目前材料科学相关的研究大多仍以实验和模拟方法为主,因此工作量巨大、耗时耗力,材料信息学的发展为以上问题提供了解决方案。作为材料基因工程的核心和材料研发的新范式[1],材料信息学将材料科学与工程、计算机视觉、机器学习、图像处理、人工智能相结合,推动了材料科学的研究进程和材料科学的工程化应用。

针对将人工智能技术应用到材料领域这一课题,国内外都积极谋篇布局。2011年,美国首次提出了“材料基因组计划(Materials Genome Initiative, MGI)”;同年,中国科学院和中国工程院也提出要开展中国版的“MGI”,欧盟也启动了“加速冶金学(Accelerated Metallurgy, ACCMET)计划”;2012年,欧洲科学基

金会推出了“欧洲冶金复兴计划”[2]; 2016 年, 中国科技部开启“材料基因工程关键技术与支撑平台”重点专项; 2020 年, 欧洲又发布了“欧洲电池 2030+计划”; 2021 年底, 中国工信部发布了《“十四五”原材料工业发展规划》,《规划》也提出要实施材料基因工程计划, 探索材料研发的新模式, 并把握好新材料技术与信息技术的融合发展趋势。无一例外, 这些项目和规划都是借助人工智能技术加速材料科学研究, 通过将人工智能与材料领域专业知识融合, 缩短材料的研发周期并降低研究成本。

众所周知, 人工智能技术对数据量有着较高的要求, 一个优质的机器学习模型通常需要在大量样本上训练才能达到较高的精度, 在深度学习领域更是如此。然而材料数据的一个突出特点就是小样本, 即材料数据样本量少[3], 尽管世界各地研究机构收集的材料数据数以亿计, 但在材料问题的繁复性和多样性面前仍是九牛一毛。材料数据通常从材料微观组织图像中获得, 但由于设备成本、实验环境等因素的限制, 某些材料微观图像的样本数量甚至只有几张。又因为大量的材料数据分散于不同组织机构或个人手中, 受技术机密和经济利益的约束, 这些数据往往难以被共享, 这也造成了特定研究中材料数据不足的问题。

作为驱动材料科研的动力, 对材料微观图像的分析至关重要。人工观察材料微观图像以及人工统计分析材料图像参数的方法速度缓慢, 且容易出现误差。数字图像处理技术和计算机视觉方法为分析材料微观结构提供了强有力的处理工具, 但在小样本材料图像数据的场景下, 这些方法难以达到预期的效果。虽然在某些任务中材料图像的样本量很少, 但是如果能对已有图像进行增强或扩充大量标注数据, 就可以利用这些增强后的图像或扩充后的数据辅助目标任务的进行, 而图像数据增强[4]正是这样一种行之有效的方法。因此, 为了在材料图像样本稀缺的情况下, 也能顺利完成相关视觉任务, 本文在分析了不同小样本材料图像的基础上, 结合图像处理、图像增强、图像生成等技术, 分别实现了小样本热障涂层图像的特征增强以及钢铁微观图像的数据增强, 有效满足了各自任务中的需求, 为小样本材料图像的研究提供了借鉴意义。

1.3 国内外研究现状

数据增强又称数据增广或数据扩增，是一种常用的扩大数据规模的方法，能够在非实质性增加数据的情况下，让有限的数据产生等价于更多数据的价值。而广义的数据增强包含两方面的内容，具体指的是以图像的先验知识为基础，通过对原始图像适当变换来增强图像特征或创建合成新数据来扩充样本量，从而简化图像的处理流程或得到泛化能力更强的网络。因此在计算机视觉任务中，数据增强方法经常被用于解决图像处理中的小样本问题。

常用的图像数据增强方法从不同的角度来看也有不同的分类方式，比较有代表性的分类方式有：有监督数据增强和无监督数据增强；单样本数据增强和多样本数据增强；基于图像空间的增强和基于特征空间的增强。有监督的数据增强通常使用一些基础的图像处理手段对图像进行变换，如几何变换、颜色变换以及图像混叠；而无监督的数据增强多使用基于深度学习的图像生成模型来生成部分样本[5][6]，如基于生成对抗网络（Generative adversarial network, GAN）的图像生成、基于变分自编码器（Variational Autoencoders, VAE）的图像生成、基于流模型的图像生成、基于扩散模型的图像生成和基于 SinGAN 及其变种的单样本图像生成。这些不同的数据增强方式按照操作样本的数量又可以分为单样本数据增强和多样本数据增强。具体分类如图 1.1 所示。

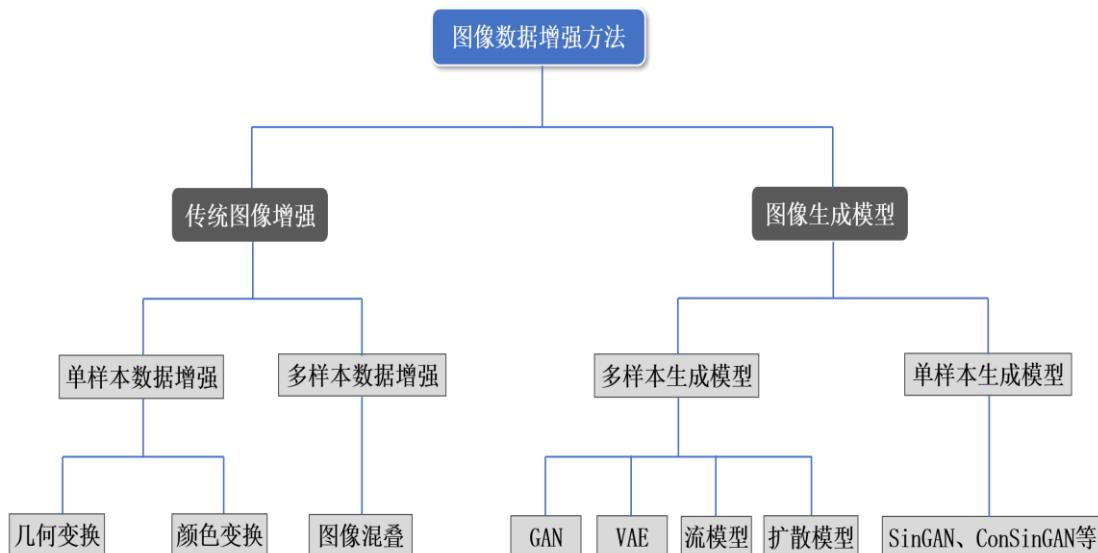


图 1.1 图像数据增强方法分类

除此之外，数据增强方法还可以分为基于图像空间的增强和基于特征空间的增强。基于图像空间的增强是直接在图像空间进行变换，而基于特征空间的增强则是先对图像进行特征提取，然后在图像的特征空间内使用一些通用的数据扩充技术进行变换。目前流行的大多数图像增强方法，多基于图像空间，但也有一些基于特征空间的算法能够实现较好的数据增强效果，如 SMOTE（Synthetic Minority Over-sampling Technique）[7]算法。

综上，本节从传统图像增强（有监督数据增强）和图像生成模型（无监督数据增强）两个角度，对图像数据增强的研究现状进行综述，同时分析和阐述材料图像研究中使用的数据增强方法。

1.3.1 传统图像增强

传统的图像增强方法多基于数字图像处理技术，基础的图像操作包括几何变换、颜色变换以及图像混叠。通过对输入的图像进行各种操作的组合变换，产生众多新图像，从而实现数据扩充。此外，这些方法还能通过增强图像特征、增强图像对比度等方式实现图像增强和图像质量的提升。

几何变换类的图像增强方法比较基础也相对容易实现，该类方法包括对图像翻转、旋转、平移、裁剪以及缩放等操作，主要从图像形态出发进行增强。图像翻转包括对图像水平翻转、垂直翻转和对角线翻转。在 ImageNet[8]、Omniglot[9]、VOC[10]以及 STL-10[11]等数据集上，该操作经常被用于扩充图像数据集。图像旋转通常是图像绕自身中心点旋转一定角度得到一张新图像，图像旋转和图像翻转类似，对于有些对方向不敏感的计算机视觉任务，如图像分类，这都是常用的增强操作。如 Han 等人[12]在对纳米物体识别时，对预训练的图像数据按照设定的角度进行旋转，在消除角度影响的同时增强了图像特征，从而提高了目标物体的识别精度。图像平移是将图像进行水平或垂直方向上的移动，移动产生的空白区域可以使用指定的像素值填充。平移和旋转都属于对图像的仿射变换，仿射变换能够保持图像的平行性和平直性。图像裁剪是按照随机定义的兴趣区域对图像裁剪，该操作会改变图像的大小，如 $(32, 32) \rightarrow (25, 25)$ ，这一点和平移操作略有相似，但平移操作对图像的空间尺寸没有影响。随机裁剪也是深度学习模

型训练时经常使用的增强方法，在一些经典的深度学习网络如 LeNet[13]、AlexNet[14]以及 VGG16[15]中均被使用。图像缩放和上述几种操作最大的不同之处在于缩放会导致图像失真，因为该操作是按照设定的比例缩小或放大图像。通常情况下，神经网络要求训练的图像数据尺寸大小固定，针对大小不一的图像，就可以对图像裁剪或缩放来保证尺寸相同。但图像缩放导致的图像失真也会使模型效果相对较差，而全卷积神经网络 FCN (Fully Convolutional Networks) [16]对于输入的图像尺寸没有严格要求。

以上五种几何变换类的图像增强方法虽然简单易用，但仍存在一些缺点，如使用此类方法容易导致标签数据的改变。在手写数字分类任务中，若对 MNIST[17]数据集使用翻转或旋转，有些图像的标签可能不会被保留。在目标检测任务中，若使用翻转或平移操作，则要对应调整 GT (Ground Truth) 框，即改变待检测物体的标签。因此使用几何变换需要事先确保图像标签是否会被改变。另一个常见的问题是几何变换类方法扩充的图像样本过于单一，产生新图像样本的同时并没有考虑到这些样本是否具有实际价值。因此几何变换类方法的适用场景和操作范围也相对有限。除平移外，上述几何变换类操作并未更改图像本身的内容，只是对局部图像或图像的像素进行重分布，故增强后的图像多样性不高。而改变图像内容的操作，大多可归入颜色变换类的数据增强方法中。

颜色变换类方法的本质是像素变换，比较常用的像素变换方法包括：添加噪声、色彩空间转换、随机擦除与像素填充、图像平滑、直方图均衡化、阈值处理和形态学操作等。添加噪声的增强方法主要是在原始图像上随机叠加部分噪声，如在图像中添加高斯噪声、泊松噪声或椒盐噪声。除此之外，还有其他较复杂的基于噪声的数据增强方法，如 SimplexNoiseAlpha、FrequencyNoiseAlpha 和 Coarse Dropout[18]。SimplexNoiseAlpha 通过将连续单一噪声的掩模与原图像混合来产生新图像；FrequencyNoiseAlpha 首先在频域中用随机指数对噪声映射进行加权，然后转换到空间域，以此产生新图像；Coarse Dropout 通过将图像中位置随机但面积大小可设定的矩形区域的信息丢失来实现图像变换，若矩形区域中丢失的是所有通道的信息，那么该矩形区域会成为黑色矩形块，若丢失的是部分通道的信息，则会产生彩色噪声。Moreno-Barea 等人[19]对 UCI 数据库[20]中的图像数据

注入噪声，结果发现该操作能使卷积神经网络（Convolutional Neural Networks, CNN）学习到更鲁棒的特征。

色彩空间转换即转换图像的色彩空间类型，除了常见的 RGB 色彩空间类型，还有 GRAY 色彩空间、XYZ 色彩空间、YCrCb 色彩空间、HSV 色彩空间、HLS 色彩空间等。在计算机视觉任务中，经常要转换图像的色彩空间，如提取图像特征时，通常需要将图像从 RGB 图转换为灰度图。在其他应用中，也会把图像在不同的色彩空间之间进行转换。如邢静等人[21]研究了基于 HSV 色彩空间的改进红外与可见光图像融合算法，首先将图像的色彩空间从 RGB 转为 HSV，并分离 H、S、V 三通道，然后对 H、S、V 通道建立数学模型相互映射融合，再将新的 HSV 图像转换到 RGB 图像，结果表明，该方法产生的新图像拥有更加明亮的目标信息和环境信息。而另一种色彩变换是通过调整图像的亮度、对比度、饱和度、锐度及色调等基本属性实现图像增强。Mikołajczyk 等人[22]在皮肤黑色素瘤诊断任务中使用了提高图像对比度和锐度等不同的色彩变换方法增强图像，有效辅助了皮肤黑色素瘤分类任务的完成。除此之外，RGB 颜色扰动也常被用于图像增强，通过在某个颜色空间中增加或减少颜色分量，或更改颜色通道的顺序完成对图像的颜色变换。王舒雅等人[23]提出了一种基于颜色模型扰动的语义对抗样本生成方法，该方法生成的样本对于图像背景信息的修改相对较多，但是图像中物体的形状特性保持的相对完整，有效保留了图片的语义信息。从前人的实验结果来看，对图像进行色彩空间转换能起到一定的增强效果，但也增加了一定的计算成本。

随机擦除由 Zhong 等人[24]提出，该方法常结合像素填充一起使用。具体做法是随机选取图像的某块区域，用设定的像素值或随机像素值来覆盖此区域，以此擦除选定区域的图像信息。Devries 等人[25]提出的 cutout 方法就是随机将样本中的部分区域删除，并用 0 像素值填充。由于选定区域的信息被删除，卷积神经网络在学习时就会被迫关注其它特征，这在一定程度上防止了模型对某个特征过拟合，cutout 方法的使用多见于图像识别和图像检测。He 等人[26]提出的 MAE (Masked Autoencoders) 也使用了随机擦除的思想，模型训练时先对输入图像的大比例随机区域予以遮蔽，再重建丢失的像素，结果表明即使随机遮蔽了 75% 的

图像区域，模型依旧能很好地重建图像。此外，随机擦除在图像分类、目标检测、行人重识别等任务中都取得了较好的增强效果，但该类方法也容易造成图像标签的改变，因此，在具体任务中可能需要对数据集进行手工干预。

图像平滑是在尽量保留图像原有信息的情况下，过滤掉图像的噪声。其基本原理是减少各像素点值的差异实现像素的平滑，这同时也导致了图像模糊，因此图像平滑又称图像模糊。图像平滑的方式很多，不同的核滤波器对应不同的平滑方式，常用的有高斯滤波、中值滤波、均值滤波、方框滤波、以及双边滤波等，因此图像平滑也可称图像滤波，多数滤波方式的卷积核大小和数值都可以灵活设定。为了提高低照度图像去噪处理的整体效果，罗丹[27]提出了基于模糊均差的低照度图像平滑去噪方法，该方法不仅去噪效果好，且图像信息的损失较低。而 He[28]等人提出的引导滤波在去噪、去雾[29]、抠图以及细节增强等领域也被频繁使用，由于其优异的性能和较低的时间复杂度，引导滤波也成为三大保边平滑算法之一。而 Simard 等人[30]提出的弹性变换也能根据扭曲场的平滑度与强度逐一地移动局部像素点来实现图像模糊。在医疗影像分割任务中，由于医疗图像样本较少，对 U-Net[31]的训练也使用了弹性变换方法增强图像并取得了不错的效果，且弹性变换的数据增强方法亦有助于解决线的分割问题。

直方图均衡化能够增加图像整体的对比度，对那些对比度较低的灰度图像来说，该操作能够使图像的阴影部分也产生细微的差异。直方图均衡化常用于 CT 图像分割、去雾去雨、指纹识别、手指静脉识别等任务中，均衡化后的图像其视觉效果得到改善，使用价值也有所提高。如李少荣[32]提出了一种改进直方图均衡化的红外图像增强技术，提高了图像的清晰度和整体对比度，降低了图像的细节损失。

阈值处理和形态学运算在各类视觉任务中也比较常见。阈值处理在图像分割中的使用尤为广泛，通过剔除图像内像素值高于或低于阈值的像素点，便可得到一张增强后的图像。二值化阈值处理、截断阈值处理、自适应阈值处理以及 Otsu[33]算法等都可以实现图像增强，从而推动后续任务的进行。而形态学运算的基础是腐蚀操作和膨胀操作，将腐蚀和膨胀组合，就可以实现开运算、闭运算、礼帽运算、黑帽运算等多种不同形式的运算。形态学操作在文字识别、车牌识别、

材料图像处理、红外图像处理等领域都有非常重要的应用。如唐守锋等人[34]为提升红外图像边缘检测的性能，提出一种改进 Canny 算子与多尺度形态学融合的红外图像边缘检测方法，并使用 Otsu 算法对阈值进行选取，以提高算法的自适应性，此方法提取的边缘轮廓更加清晰平滑，细节边缘突出且层次分明，有效抑制了红外图像噪声，提高了边缘检测效果。阈值处理和形态学操作都改变了图像的像素值，因此两者都属于颜色变换类的数据增强。颜色变换类的图像增强方法大多都易于实现，也能够产生大量样本，具备一定的增强效果。但对于颜色特征非常敏感的任务来说，此类方法便不再适用，因为上述方法很可能导致颜色信息的丢失。

几何变换类的增强方法和颜色变换类的增强方法都属于单样本数据增强，因为其操作对象都是图像本身，而多样本数据增强使用多个图像样本来产生新图像，其中较为常见的就是对图像的混合或叠加，代表性的方法有 SamplePairing[35]、mixup[36]、Cutmix[37]和 SMOTE。Ionue 等人提出的 SamplePairing 方法用于图像分类任务中的数据增强，实际操作时，先将从训练集中随机抽取的两张图像随机裁剪到设定的尺寸，再对图像随机水平翻转，最后平均每个 RGB 通道的像素值来生成新图像。该方法思路简单，增强效果较好，在医学图像处理中比较有效。但由于随机选取的样本标签有可能不同，从而导致模型训练容易出现误差，并且 SamplePairing 方法缺乏可解释性，没有足够的理论支持。作为对 SamplePairing 的拓展延伸，mixup 将随机的两张图像按比例插值生成新样本，实验结果表明使用此方法进行数据增强能够降低深度学习模型在某些图像数据集、语音数据集以及表格数据集上的泛化误差。与 mixup 不同的是，Cutmix 需要将图像的部分区域裁剪掉，然后随机填充训练集中其他图像区域的像素值，Cutmix 与前文提到的 cutout 的区别在于两者填充的像素值不同。需要注意的是，mixup 和 Cutmix 仅适用于图像分类任务中的数据增强。还有一种基于插值思想的 SMOTE 算法能为数据集中的小样本类合成新样本，以此缓解数据集样本不平衡的问题，与上述众多方法最大的不同之处在于，SMOTE 是基于特征空间的数据增强方法。由于 SMOTE 算法能够很好地降低样本不平衡的比例，因此后续又涌现出众多对该算法的改进，如 DBSMOTE[38]、Borderline-SMOTE[39]和 Safe-level-smote[40]等。

但以上方法合成的新样本在特征空间中仍位于已知样本的附近，如果能在已知样本附近以外的区域插值，有可能更好地提升数据增强效果。综上，图像混叠类的增强方法能在特定任务中发挥作用，但也存在一个显著问题，即此类方法带来的性能提升缺乏可解释性。

1.3.2 图像生成模型

传统图像增强方法大多是有监督的，且操作简单、易于实现。由于这类方法增强效果有限，因此对大多数任务而言，仅用它们进行增强是远远不够的。为了获得更优质的模型，研究者们把目光投向无监督的数据增强，而无监督数据增强方法中最重要的便是使用图像生成模型扩充训练样本。图像生成模型的基本原理是：向网络中输入一组图像，模型从这组图像中学习图像分布的信息，然后使用训练好的模型生成一组与输入数据类似的图像，因此拟合出图像分布是图像生成模型的关键。目前仍被频繁使用的两种图像生成模型分别是 GAN[41]和 VAE[42]，而使用流模型生成图像的应用相对较少[43]。近两年，扩散模型[44]在图像生成任务上的表现已经超越 GAN，似乎该模型已经能够代替 GAN 主导图像生成领域，但其优异性能背后的代价是极大的算力和超长的训练时间，这使得多数研究者在实际工作中仍无法使用扩散模型。因此本节将详细介绍 GAN 和 VAE 及其变体的研究现状，对流模型和扩散模型不在详细叙述。从训练样本的角度来看，基于 GAN 和 VAE 的模型大多属于多样本图像生成模型，与其对应的单样本图像生成模型的相关研究本节也会做出介绍。

GAN 自 2014 年提出后，便一直被改进并应用在各领域，迄今为止 GAN 相关的衍生模型至少有上百种。理想情况下，如果能用 GAN 生成大量样本作为对训练集的补充，那么对于小样本数据集下的模型训练来说就能获得良好的数据增强效果。但原始的 GAN 并不容易使用，因为 GAN 的训练很不稳定，容易产生梯度消失和模式崩塌等问题。改进后的 GAN 如 CGAN[45]、DCGAN[46]、RoCGAN[47]、BiGAN[48]、EBGAN[49]、WGAN-GP[50]、SAGAN[51]以及 DivGAN[52]等诸多模型要么是在一定程度上解决了 GAN 训练不稳定的问题，要么是进一步提高了生成图像的质量。其中 CGAN 的提出使模型的生成结果不再

只是逼近真实图像，而是让生成图像与真实图像相似的同时也符合给定条件的描述。Triple-GAN[53]和 StackGAN[54]两者都优化了 CGAN 的模型结构，使生成图像的质量更好。原始 GAN 模型使用的是全连接神经网络，而图像处理中应用最广泛的是卷积神经网络，DCGAN 将 GAN 和 CNN 结合，把生成器和判别器的网络都换成卷积神经网络，并对结构做出一些改变，从而提高了生成样本的质量和模型的收敛速度。RoCGAN 通过添加隐式约束使模型对噪声更具鲁棒性，并且该模型在自然场景和人脸图像等多个领域的表现都优于之前的 CGAN 架构。WGAN 提出用 Wasserstein 距离取代 JS 距离，以此来更好地衡量真实图像分布和生成图像分布之间的差异，WGAN-GP 是对 WGAN 的加强。SAGAN 在模型中引入了自注意力机制，DivGAN 则致力于缓解原始 GAN 模式崩塌的问题。

由于 GAN 的快速发展和优异表现，GAN 在医学图像、人脸图像[55][56]、行人检测[57]和异常检测[58]等各场景中都得到了广泛应用。Frid-Adar 等人[59]使用 DCGAN 生成肝脏病变图像来增广数据，最终将模型的分类准确率提升了 7%。Liu 等人[60]使用 Wasserstein GAN 生成癌症数据，以扩充训练集的样本，从而提高了样本识别的准确率。对于人脸图像生成任务，ProGAN[61]基于渐进式增长的思想生成高分辨率的图像，效果较好。Choi 等人[62]提出的 StarGAN 能使生成的人脸图像在多种风格之间转换，并且模型小巧高效。Sun 等人[63]在 StarGAN 的基础上，修改了模型的目标函数，改进后的模型生成的人脸图像更自然，并且使用生成样本进行数据扩增后训练的模型分类准确率更高，这表明改进后的 StarGAN 在视觉感受和量化指标上都略胜一筹。Zhong 等人[64]提出一种基于 GAN 的数据增强策略，缓解了行人数据集的小样本问题。孔祥斌等人[65]也使用 GAN 生成铁路周界行人图像，生成图像的视觉效果较好且能兼顾周边环境，并且使用生成图像扩充样本后的检测准确率得到提升。

GAN 的另一个重要应用是图像到图像的风格迁移，Leon Gatys 等人[66]在 2015 年就提出使用神经网络迁移图像画风。随着 GAN 的不断发展，基于 GAN 的风格迁移算法也逐渐涌入大众视野，如 CycleGAN[67]、InstaGAN[68]、IcGAN[69]、Modular GAN[70]以及前文提到的 StarGAN。图像的风格迁移技术也能用于数据增强，马征等人[71]使用 CycleGAN 将游戏图像转换为遥感仿真图像，

扩充了标签数据，缓解了遥感图像数据不足导致模型训练效果差的问题。朱仲贤等人[72]提出了一种基于 CycleGAN 的联合对比学习方法生成光伏巡检图像，该方法提升了生成图像的质量，生成图像的主体内容不会产生形变和缺失，并且能够有效辅助真实样本稀少情况下的光伏巡检任务。虽然 GAN 被广泛应用于各领域，但仍然不能忽视它存在的一些本质问题，如训练不稳定、易产生梯度消失、易发生模式崩溃以及可解释性差等。

VAE 和 GAN 目标一致，都是希望构建一个能从隐变量生成目标数据的模型。尽管 VAE 与自编码器（AutoEncoder，AE）的名称十分类似，并且都由编码器和解码器组成，但两者的原理和应用大不相同。AE 中编码器的作用是将输入数据压缩成潜在空间表征，而解码器的作用是重构来自潜在空间表征的输入数据，因此自编码器具有表征学习算法的功能，常用于降维和去噪。而 VAE 中的编码器是用神经网络拟合出输入数据分布的均值和方差，然后从分布中采样，再经过解码器产生新数据。因此 VAE 常用于图像生成，但 VAE 生成的图像比较模糊。

和 GAN 一样，VAE 的衍生模型也数量众多，这些模型在原始 VAE 的基础上引入各种技巧，弥补 VAE 生成图像模糊的缺陷。对 VAE 改进的经典做法是将 VAE 与 GAN 结合，如 CVAE-GAN[73]和 IntroVAE[74]等，其中 CVAE-GAN 能够生成具有细粒度类别标签的真实且多样的样本，IntroVAE 则实现了高清图像等高维数据的无条件生成。这类工作虽然借助 GAN 提升了性能，但也容易引入 GAN 训练不稳定等缺点。另一种改进方法是将 VAE 与流模型结合，如 IAF-VAE[75]和 f-VAEs[76]。NVIDIA 的 Arash 和 Jan[77]把当前计算机视觉领域的众多优秀成果，如多尺度架构、可分离卷积、swish 激活函数及流模型等融合到 VAE 中，提出了 NVAE 这一模型。众多技巧的叠加组合使 NVAE 的生成效果达到新高度，但 NVAE 的训练成本很大，对个人来说，复现或使用 NVAE 并不容易。

除了生成图像模糊这一缺点，VAE 还存在后验坍塌问题，这通常是由 KL 散度消失导致的。为了防止后验坍塌，很多方法从 KL 损失和重构损失这两个方向出发改进 VAE。如 δ -VAE[78]和 BN-VAE[79]等模型都是从改进 KL 散度损失项入手阻止其消失。而从重构损失项出发缓解后验坍塌问题的核心思想是弱化解码器，这个方向的改进工作多见于自然语言处理领域[80]，在图像生成领域较少。

虽然 VAE 不如 GAN 的应用广泛，但在图像生成领域也有一席之地。Zhang 等人[81]提出一种基于 SENet (Squeeze-and-Excitation Network) 和 FBN (Fixed Batch Normalization) 的 VAE 模型，即 SE-FBN-VAE。融合 SENet 后，模型对通道特征的敏感性提高，网络的特征提取能力也被加强，而 FBN 的使用则减少了网络的收敛时间，后验坍塌问题也被有效缓解，结合上述技巧后的 SE-FBN-VAE 能有效完成 3D 多孔介质的重建。和前一个工作不同，Zhang 等人[82]提出一种将 VAE 和 GAN 结合的生成模型来重构 3D 多孔介质，模型接收 2D 图像切片作为输入，输出结果是生成的 3D 多孔介质。在该模型中，VAE 的编码器对输入图像的统计信息和形态信息进行表征，然后将解码出的低维特征向量送入生成器，得益于该模型是 VAE 和 GAN 的混合，故模型训练时更加稳定，生成能力也得以增强。杜娟等人[83]设计了基于 VAE 的风格迁移算法，并将其应用在现代服饰的中国风格迁移中，使用该算法合成的图像不仅服饰更加精细，并且能适应不同的中国传统风格，满足了服饰风格迁移场景的应用要求。Huang 等人[84]把 VAE 嵌入到 WGAN-GP 中，设计了 VAE-WGANGP 网络，以生成合成孔径雷达（Synthetic Aperture Radar, SAR）图像，在生成 SAR 图像方面，VAE-WGANGP 网络能同时兼顾生成图像的质量和生成图像的多样性。由既有相关文献可知，基于 VAE 模型的图像生成在交叉学科中的应用不十分常见，并且 VAE 模型在做生成任务时常与 GAN 或注意力机制等其他技术一起使用，这也是对 VAE 生成图像模糊这一缺点的改进。

前文所述的基于 GAN 或 VAE 的生成模型大都属于多样本生成模型，即训练集中包含多张或多组图像。训练集中只包含一张图像的则属于单样本生成模型，如 SinGAN[85]、ConSinGAN[86]以及 HP-VAE-GAN[87]等，这些模型多用于图像的超分辨率重建。程国建等人[88]使用 SinGAN 对岩石薄片图像进行超分辨率重建，无需输入大量数据即可重构出高质量图像。虽然单样本生成模型在图像增广方面的应用较少，但对小样本数据集来说确实能够起到数据增强的效果。彭晏飞等人[89]将 ConvNeXt[90]与 SinGAN 结合提出了一种基于改进 SinGAN 的遥感图像数据增强方法，使用该方法生成的图像在三种不同的评价指标上均有所提升，并且生成的遥感图像还可用于图像分割任务中。在工业缺陷检测领域，胡文杰等

人[91]使用注意力机制和结构相似度损失函数改进了 ConSinGAN，改进后的模型生成效果较好且具有实用价值，同时生成的众多样本也有效地扩充了训练集。

使用单样本生成模型增强图像的最大优点是训练网络无需大量数据。对诸如医学图像、材料图像及遥感图像等特定领域而言，由于受到隐私保护和收集成本的限制，这些领域中往往存在着由数据量极度缺乏导致的小样本问题，而使用GAN 或 VAE 生成图像都需要有足够的数据来训练网络使其收敛。又因为 GAN 和 VAE 都是基于真实图像的数据分布实现的，在真实图像很少的情况下，图像的数据分布大概率也较差，基于差的真实数据分布生成图像，其生成结果也并不理想。单样本生成模型则不受上述问题的困扰，并且生成样本的数量可以自行设定，因此对小样本数据集来说，使用单样本生成模型进行数据增强具有现实意义。

1.3.3 材料图像增强

材料微观组织图像属于纹理图像，由于材料在热加工处理过程中受温度和其他条件的影响，因此材料微观图像中的晶体数量和形状存在很大的多样性，且图像中的纹理纷繁复杂。材料信息学背景下的材料图像研究往往需要使用图像处理技术和计算机视觉方法，对大部分材料微观图像来说，把原始图像直接送入计算机处理往往行不通，通常需要借助预处理手段对图像增强，比如转换色彩空间、去除图像噪声等。又由于材料图像数据集的小样本特性，使用深度学习方法常会使模型过拟合或难收敛，所以对图像数据集进行增广十分必要。因此对材料图像的增强可分为两类，一类是使用图像处理方法增强材料图像的颜色特征、纹理特征、形状特征、空间关系特征等属性，另一类是扩增材料图像的样本数量。近年来，材料研究领域涌现出众多将图像增强方法应用于材料图像的识别、分割和性能预测等任务中的工作，并取得了优秀的成果。

Wang 等人[92]在研究喷涂功率对钇稳定氧化锆（Yttrium-Stabilized Zirconia, YSZ）飞溅形貌和热障涂层微观结构的影响这一问题时，提出了一种利用机器视觉自动化识别热障涂层中的孔隙并计算孔隙实度的方法。该方法使用滤波算法、图像二值化以及形态学操作对图像进行增强，增强后图像中的孔隙轮廓更容易被提取，有效提升了孔隙识别的效率。Han 等人[93]在分割纹理复杂的金相图像时，

先使用高斯滤波平滑图像，增强了图像的纹理特征，使材料图像中相的分割更加准确。Wang 和 Han 在各自的工作中都使用了传统图像增强方法加强图像的某些特征，改善了图像的质量，属于广义上图像增强的范畴，还有一些工作通过生成新样本来增广图像数据。

Ye 等人[94]在原始高熵合金图像数据中加入高斯噪声，以生成更多的“伪样本”，在高熵合金的相分类任务和硬度预测任务中，该方法显著改善了机器学习模型的相分类结果和硬度预测结果。Nawa 等人[95]使用高斯噪声注入的方式对磁性多层膜 CoFe 数据集进行增强，扩充后的训练集提升了神经网络模型的泛化性能，并提高了对自旋磁矩和形成能这两个目标量的预测精度。前面两项工作都用到了基于高斯噪声的传统数据增强方法，和它们不同的是 Yang 等人[96]基于生成对抗网络对高熵合金数据进行双步骤增广，即先使用 GAN 生成特征数据，再用 GAN 生成标签数据，将生成数据添加到原始训练集后，训练后的模型表现出较之前更好的预测性能。而 Ma 等人[97]提出了一种基于风格迁移的数据增广策略，通过融合模拟图像和真实图像来创建合成图像，解决了纯铁多晶粒组织图像分割任务中的训练数据不足的问题。

上述工作中的增强方法都对应解决了各自领域中的任务，但小样本材料图像的特异性和稀缺性导致了对不同材料图像的处理没有统一模式，这也限制了某些计算机视觉算法在材料图像上的应用。因此，如何在小样本场景中使用具有高度鲁棒性和泛化性的算法处理材料微观图像是急需解决的问题，而图像增强和数据增强正是解决此问题的关键。

1.4 论文研究内容

为了解决小样本场景下材料图像难处理的问题，本论文针对不同任务中材料图像的特点，提出了两种不同的增强策略，分别完成了对材料微观图像中目标对象的识别和多类别材料微观图像的分类。

在第一个任务中，提出了基于图像增强的热障涂层形貌特征识别方法。利用色彩空间转换、图像平滑、阈值处理和形态学运算等多种操作对热障涂层进行图像增强后，大大降低了裂纹识别的难度。进而使用基于闭运算的断线修复增强裂

纹特征，并结合骨架提取算法和遍历搜索算法完成对热障涂层中裂纹的自动识别与长度计算。相比于人工识别，该方法对裂纹的自动识别更加便捷高效，极大节约了时间成本和人力成本。针对此任务，本文还设计开发了可供材料科研人员直接使用的软件，提高了热障涂层微结构分析的研究效率。

在第二个任务中，使用融合注意力机制的单样本生成模型 HP-VAE-GAN 生成材料图像，实现对小样本材料图像数据集的增广。改进的 HP-VAE-GAN 在 PatchVAE 中引入 CBAM，在通道注意力和空间注意力的双重加持下，模型的特征提取和特征表示能力都强于原始的 HP-VAE-GAN。针对具体的超高碳钢显微组织图像分类任务，使用生成图像对训练集进行数据增强后，测试集上的分类准确率得到了显著提升，这也缓解了小样本场景下分类模型训练时的过拟合问题。

1.5 论文组织结构

本文一共分为五个章节，组织结构如下：

第一章：绪论。介绍了课题来源以及课题的研究背景和意义，分析梳理了传统图像增强方法、图像生成模型以及材料图像增强领域的研究现状。

第二章：相关理论与技术概述。本章主要对论文中涉及的相关理论与技术进行介绍，包括传统图像增强方法中的图像滤波技术、形态学理论以及 GAN 和 VAE 的工作机制，并阐述了计算机视觉领域中常用的几种注意力机制。

第三章：基于图像增强的热障涂层形貌特征识别。针对样本极少的热障涂层微观图像，提出了基于传统图像增强的形貌特征识别方法。首先介绍了方法的整体流程，进而详细说明了增强方法的使用目的，并开展实验以验证方法的有效性。最后，针对本章中的任务，还设计了热障涂层裂纹识别软件，并做了详细说明。

第四章：基于单样本生成模型的材料图像数据增强。首先概述了整体方法，其次介绍了使用融合 CBAM 的 HP-VAE-GAN 的模型结构，接着使用改进后的 HP-VAE-GAN 生成材料微观图像，再用生成图像对训练集进行扩充，并对材料图像进行分类实验，最后验证了所提数据增强方法的有效性。

第五章：总结与展望。本章主要是总结全文，回顾并分析存在的问题及不足，并对后续研究和未来工作做出展望。

第二章 相关理论与技术概述

2.1 本文相关的传统图像增强方法

图像增强是指通过增强图像中的有效信息满足图像分析的需要，突出图像中的感兴趣区域和特征。使用传统图像增强方法处理图像，能在一定程度上提高后续图像分析的效率和准确度。本节概述了论文第一个工作中使用到的传统图像增强方法，包括图像滤波和数学形态学。

2.1.1 图像滤波

图像滤波即图像平滑，对图像滤波能够过滤掉图像中的噪声和伪影，并最大化保留图像的原有信息。因此图像滤波方法常用于图像去噪，以实现对图像的增强。常见的滤波算法包括方框滤波、均值滤波、高斯滤波，中值滤波和双边滤波。

中值滤波对抑制脉冲噪声非常有效，对图像边界信息的保留也较好，主要原因在于中值滤波没有求均值这一操作，所以不会出现均值滤波等滤波方式引起的细节模糊问题。相比于其他几种常见的滤波算法，中值滤波能够在减弱或消除图像的高频分量上实现较好的效果，因此中值滤波在各场景中的使用比较频繁。但中值滤波存在运算量较大这一缺点，因为操作时需要对像素值排序。

2.1.2 数学形态学

数学形态学[98]简称形态学，该操作能从图像中提取对描绘图像形状十分有用的分量信息，有助于图像理解。形态学中的基本运算是腐蚀和膨胀，腐蚀运算能使图像边缘向内收缩，膨胀能使图像边界向外扩张。将腐蚀和膨胀组合能实现开运算、闭运算、黑帽运算及礼帽运算等。开运算是对图像先腐蚀后膨胀，闭运算是先膨胀后腐蚀；黑帽运算是用闭运算图像减去原始图像，礼帽运算是用原始图像减去开运算图像。形态学运算能有效分析图像的形状和结构，因此在字符识别、文字识别、材料显微图像分析以及指纹检测等方面都取得了良好的应用[99]。形态学在各领域的成功使用与其具备的图像增强作用密不可分，如开运算能用于

去除图像噪声，闭运算能关闭图像前景物体内部的小孔，也能连接不同的前景图像。在给定场景中，使用形态学运算能显著增强图像中的有效信息，突出图像的整体或局部特征，从而改善图像的视觉效果，方便后续任务中的图像处理或分析。

本文在第一个工作中除了用到膨胀和腐蚀操作，还使用了闭运算增强材料微观图像中目标对象的整体特征，应用闭运算能够填补断裂的轮廓线这一特性提高了材料微观图像的质量。

2.2 无监督图像生成模型基础

2.2.1 生成对抗网络

生成对抗网络采用对抗学习的思想，参与对抗的双方是生成器（Generator）和判别器（Discriminator），GAN 的主要框架如图 2.1 所示。隐变量 $z \sim p_z(z)$ （服从高斯分布的随机噪声）作为输入被送入生成器 G 中，生成样本记为 $G(z)$ ，生成样本的数据分布记为 p_g ；相应地，真实样本 x 的数据分布记为 p_d 。当真实样本 x 或生成样本 $G(z)$ 被送入判别器后，判别器输出概率值 $D(x)$ 或 $D(G(z))$ 来判断输入图像的真伪，概率值为 1 代表是真实图像，概率值为 0 代表是生成器生成的图像。

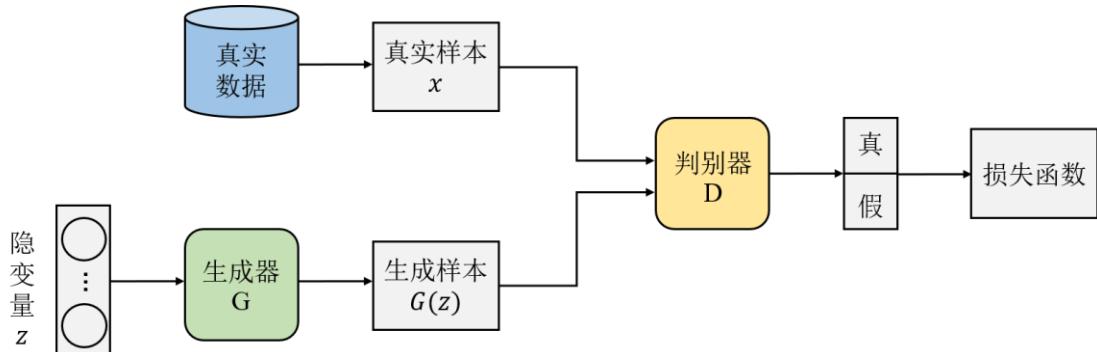


图 2.1 GAN 模型框架

简言之，生成器负责生成逼近真实样本的图像，并且生成的图像要使判别器难以辨别，而判别器需要尽量准确判断出输入图像的真假。当生成器生成的图像使判别器难以辨明真伪时，GAN 就达到了最优状态，即纳什均衡。网络训练时， G 和 D 交替优化：首先初始化 G 和 D ，在每次训练迭代中，先固定生成器 G ，更新判别器 D ；再固定判别器 D ，更新生成器 G 。具体的训练过程如下：

(1)训练判别器 D 时,先固定生成器 G ,然后从随机噪声 z 中采样生成 $G(z)$,将生成样本 $G(z)$ 和真实样本 x 送入判别器 D ,根据判别器输出的概率值 $D(x)$ 或 $D(G(z))$ 及样本标签来计算误差,最后通过反向传播更新判别器 D 的参数 θ_d ;

(2)训练生成器 G 时,先固定判别器 D ,生成器 G 随机生成样本 $G(z)$ 送入判别器 D 中,然后根据判别器输出的概率值 $D(G(z))$ 和样本标签来计算误差,最后通过反向传播更新生成器 G 的参数 θ_g 。

根据训练过程,可以推导出 GAN 的目标函数。对判别器而言,如果输入图像 x 来自于真实的数据分布 p_d ,就希望输出的概率值 $D(x)$ 越大越好,即 $\log D(x)$ 越大越好;相反,如果输入图像来自于生成数据的分布 p_g ,也即输入图像为 $G(z)$,则希望输出的概率值 $D(G(z))$ 越小越好,等同于 $\log(1 - D(G(z)))$ 越大越好。经分析得出判别器需要优化的目标函数如公式(2.1)所示:

$$\max_D V(G, D) = E_{x \sim p_d}[\log D(x)] + E_{z \sim p_z}[\log(1 - D(G(z)))] \quad (2.1)$$

对生成器而言,生成的图像越真实越好,也即若输入图像来自 p_g ,则希望输出的概率值 $D(G(z))$ 越大越好,等同于 $\log(1 - D(G(z)))$ 越小越好,不难得出生成器待优化的目标函数如公式(2.2)所示:

$$\min_G V(G, D) = E_{z \sim p_z}[\log(1 - D(G(z)))] \quad (2.2)$$

由公式(2.1)和公式(2.2)可以得出 GAN 训练时需要优化的总目标函数如下:

$$\min_G \max_D V(G, D) = E_{x \sim p_d}[\log D(x)] + E_{z \sim p_z}[\log(1 - D(G(z)))] \quad (2.3)$$

从上述推导可知 GAN 类似于双人 MinMax 游戏,理想状态是游戏达到纳什均衡,此时的最佳生成器和最佳判别器分别记为 G^* 和 D^* 。对 G^* 、 D^* 以及值函数的求解可视为求极大极小值问题。由于 GAN 训练时首先会训练判别器使其达到最优,其次才是训练生成器,故求最优生成器是在得出最优判别器的基础上进行的。因此,在未达到纳什均衡时,固定生成器 G ,求解此时最优的判别器 D_G^* 是首要任务。令 $G(z) = x$,再将公式(2.1)中的数学期望写成等价的积分形式,经推导可得:

$$\begin{aligned} \max_D V(G, D) &= E_{x \sim p_d}[\log D(x)] + E_{z \sim p_z}[\log(1 - D(G(z)))] \\ &= E_{x \sim p_d}[\log D(x)] + E_{x \sim p_g}[\log(1 - D(x))] \\ &= \int [p_d(x) \log D(x) + p_g(x) \log(1 - D(x))] dx \end{aligned} \quad (2.4)$$

为了使 $V(G, D)$ 最大，等价于 $f = p_d(x) \log D(x) + p_g(x) \log(1 - D(x))$ 最大，此时 G 被固定，故 $p_d(x)$ 和 $p_g(x)$ 均为常数。 f 对 D 求导，令导数为零，再进行适当变换，即可求出固定生成器 G 时最优的判别器 D_G^* 。具体求解过程如下所示：

$$\frac{df(x)}{dD(x)} = p_d(x) \frac{1}{D(x)} + p_g(x) \frac{(-1)}{1 - D(x)} = 0 \quad (2.5)$$

$$D_G^*(x) = D(x) = \frac{p_d(x)}{p_d(x) + p_g(x)} \quad (2.6)$$

求解出 D_G^* 后，将其带入公式(2.3)中，经等价变换可得公式(2.7)所示的最终结果：

$$\begin{aligned} \min_G \max_D V(G, D) &= \min_G \max_D V(G, D_G^*) \\ &= E_{x \sim p_d} [\log \frac{p_d(x)}{p_d(x) + p_g(x)}] + E_{x \sim p_g} [\log \frac{p_g(x)}{p_d(x) + p_g(x)}] \\ &= \int [p_d(x) \log \frac{\frac{p_d(x)}{2}}{(p_d(x) + p_g(x))/2} + p_g(x) \log \frac{\frac{p_g(x)}{2}}{(p_d(x) + p_g(x))/2}] dx \\ &= \int [p_d(x) \log \frac{p_d(x)}{(p_d(x) + p_g(x))/2} + p_g(x) \log \frac{p_g(x)}{(p_d(x) + p_g(x))/2}] dx \\ &\quad - \log 2 \int [p_d(x) + p_g(x)] dx \end{aligned} \quad (2.7)$$

根据 KL 散度和 JS 散度的定义，可将上式转换成 JS 散度的形式，即：

$$\begin{aligned} \min_G \max_D V(G, D) &= \min_G \max_D V(G, D_G^*) \\ &= KL(p_d(x) || \frac{p_d(x) + p_g(x)}{2}) \\ &\quad + KL(p_g(x) || \frac{p_d(x) + p_g(x)}{2}) - \log 4 \\ &= 2 \min_G JS(p_d(x) || p_g(x)) - \log 4 \end{aligned} \quad (2.8)$$

又因为 KL 散度的值恒大于等于 0，因此上式的值大于等于 $-\log 4$ ，当 $KL = 0$ ，即 $p_d(x) = \frac{p_d(x) + p_g(x)}{2} = p_g(x)$ 时，公式(2.8)可以取得最小值 $-\log 4$ 。即对最优的生成器 G^* 而言， $p_g(x) = p_d(x)$ ，再将其代入公式(2.6)中可得，最优判别器 $D_G^* = \frac{1}{2}$ 。由此可知，GAN 本质上是在最小化生成样本分布与真实数据分布间的 JS 距离。综上，GAN 训练过程的伪代码如算法 2.1 所示，其中梯度的更新可以使用任意基于梯度的学习策略，GAN 的训练过程中使用了基于动量的梯度更新算法。

算法 2.1: GAN 的训练过程

输入: 随机噪声 z , 真实数据 x , 对抗训练迭代次数 T ,

判别器的训练迭代次数 K

- 1 随机初始化判别器的参数 θ_d 和生成器的参数 θ_g ;
 - 2 **for** $t \leftarrow 1$ **to** T **do**
 - 3 **for** $k \leftarrow 1$ **to** K **do**
 - 4 从噪声先验分布 $p_g(z)$ 中采样 m 个小批量噪声样本 $\{z^{(1)}, \dots, z^{(m)}\}$;
 - 5 从真实数据分布 $p_d(x)$ 中采样 m 个小批量真实样本 $\{x^{(1)}, \dots, x^{(m)}\}$;
 - 6 通过随机梯度上升更新判别器的参数 θ_d :
 - 7
$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\log D(x^{(i)}) + \log (1 - D(G(z^{(i)})))]$$
 - 8 **end for**
 - 9 从噪声先验分布 $p_g(z)$ 中采样 m 个小批量噪声样本 $\{z^{(1)}, \dots, z^{(m)}\}$;
 - 9 通过随机梯度下降更新生成器的参数 θ_g :
 - 10
$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \log (1 - D(G(z^{(i)})))$$
 - 10 **end for**
- 输出:** 生成器 $G(z; \theta_g)$
-

2.2.2 变分自编码器

在介绍变分自编码器 (VAE) 的原理前先对自编码器 (AE) 稍作说明, 自编码器是一种利用无标签数据来学习低维特征的无监督学习, 并不生成数据。自编码器包括一个编码器 (Encoder) 和一个解码器 (Decoder), 图像输入到编码器后会被映射到低维空间, 得到一个低维向量, 解码器使用编码出的低维向量重构原始图像, 重构图像与输入图像间的重构误差越小就说明重构效果越好, 图 2.2 展示了这一过程。在自编码器中, 无需先验约束, 通过编码器网络就可以自动学习

到低维向量，实现了无监督表示学习。但有些情况下，希望学习出的低维向量符合某项分布，即实现有约束的无监督表示学习，VAE 即采用了这一思想。VAE 在重构损失的基础上添加了一个正则项，用以约束潜在变量满足高斯分布，其算法实现比较简单，但数学推导十分复杂。

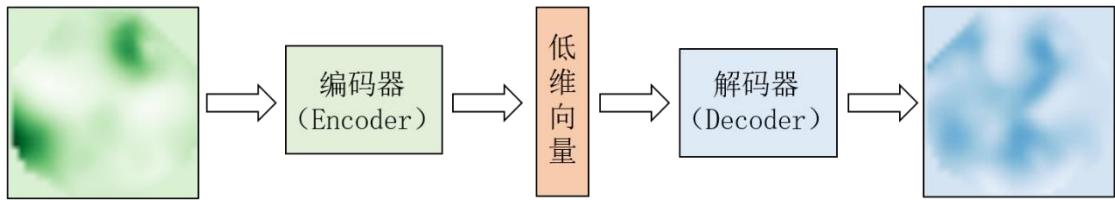


图 2.2 自编码器框架图

和 GAN 一样，VAE 也属于生成模型，生成模型的基本流程是：（1）接收从原始数据分布中抽取的样本作为训练集；（2）模型学习到一个能够模拟训练数据分布的概率分布；（3）从学习到的概率分布中采样生成新样本。由此可知，生成模型的关键问题是估计样本的概率分布，即模型想知道样本服从的数据分布是什么。对样本的数据分布进行估计又称概率密度估计，简称密度估计。概率密度是观测样本与其概率之间的关系，概率密度的总体形状称为概率分布，常见的概率分布有均匀分布、正态分布、指数分布等。对观测样本的概率计算是通过概率密度函数来完成的，但通常情况下很难获取观测样本的概率密度函数，因此需要不断逼近这个概率密度函数，这个过程就是概率密度估计。图 2.3 展示了使用高斯密度函数拟合一维样本点的概率分布的过程，左图中是采样的数据，右图是根据采样数据拟合出的高斯分布的概率密度函数。



图 2.3 概率分布的估计[100]

概率密度估计又分为显式密度估计和隐式密度估计：显式密度估计能够显式地构建出观测样本的概率密度函数，并使用极大似然估计（Maximum Likelihood Estimate, MLE）求解出参数；而隐式密度估计无需估计出概率密度函数，只需拟合出一个模型，使其能生成符合观测样本数据分布的样本即可。2.2.1 节中所述

的 GAN 通过对抗训练直接生成服从真实数据分布的样本，而不关注对真实数据分布的估计，故 GAN 属于隐式密度模型。VAE 则通过显示地定义和求解真实数据的分布，并从所求分布中采样得到新样本来实现生成功能，因此 VAE 属于显式密度模型。

作为生成模型，VAE 的需求是根据一批数据样本 $\{x_1, \dots, x_i\}$ 得到 x 的数据分布 $P(x)$ ，再从 $P(x)$ 中采样得到新样本，因此 VAE 的目标就是估计数据的概率分布。但数据分布的估计并不容易，于是 VAE 借助一个隐向量 z ，将 z 通过一个复杂映射得到 x ，其中 $z \sim N(0, I)$ 。公式(2.9)展示了数据分布 $P(x)$ 在 z 作用下的变换：

$$P(x) = \sum_z P(x|z)P(z) = \int_z P(x|z)P(z)dz \quad (2.9)$$

其中 $P(x|z)$ 描述了一个由 z 生成 x 的模型，且 $x|z \sim N(\mu(z), \sigma(z))$ 。

因此只需学习到 z 服从的高斯分布的参数（均值 μ 和标准差 σ ）和从 z 到 x 的映射函数，即可得到原始数据的分布。但隐变量 z 无法直接获得，因此还需从已有样本 x 中得到对应的 z ，即求解后验分布 $P(z|x)$ ，又因为后验概率也无法求出，所以不能完成基于后验概率的推理，而 VAE 使用分布函数 $Q(z|x)$ 近似 $P(z|x)$ 巧妙地解决了这一难题，其中 $Q(z|x)$ 可以是任意分布。在此基础上，就可以继续求解公式(2.9)。由于 $P(x)$ 通常十分复杂，导致 μ 和 σ 难以计算，因此引入神经网络来辅助求解，一个神经网络是编码器，对应 $Q(z|x)$ ，初始化网络参数为 \emptyset ；另一个神经网络是解码器，对应 $P(x|z)$ ，初始化网络参数为 θ 。

由于显式生成模型的目的是学习一个模型来最大化似然观测样本，所以希望 $P(x)$ 越大越好，等价于 $\sum_x \log P(x)$ 越大越好， $\log P(x)$ 又可写成公式(2.10)的形式。

$$\log P(x) = \log \frac{P(x, z)}{P(z|x)} \quad (2.10)$$

对公式(2.10)积分并代入 $Q(z|x)$ ，因为 $\int_z Q(z|x) dz = 1$ ，所以变换后结果不变。继续变换求解，由于 KL 散度的值恒大于等于 0，因此最终求得 $\log P(x)$ 的变分下界为 $L_b = \int_z Q(z|x) \log \left(\frac{P(x|z)P(z)}{Q(z|x)} \right) dz$ ，公式(2.11)展示了详细的求解过程，具体步骤如下所示：

$$\begin{aligned}
\log P(x) &= \int_z \log P(x) \cdot Q(z|x) dz \\
&= \int_z Q(z|x) \log \left(\frac{P(z,x)}{P(z|x)} \right) dz \\
&= \int_z Q(z|x) \log \left(\frac{P(z,x) Q(z|x)}{Q(z|x) P(z|x)} \right) dz \\
&= \int_z Q(z|x) \log \left(\frac{P(z,x)}{Q(z|x)} \right) dz + \int_z Q(z|x) \log \left(\frac{Q(z|x)}{P(z|x)} \right) dz \\
&= \int_z Q(z|x) \log \left(\frac{P(z,x)}{Q(z|x)} \right) dz + KL(Q(z|x)||P(z|x)) \\
&\geq \int_z Q(z|x) \log \left(\frac{P(x|z)P(z)}{Q(z|x)} \right) dz = L_b
\end{aligned} \tag{2.11}$$

最终的 $\log P(x)$ 就可以写成公式(2.12)的形式，最大化 $\log P(x)$ 就等价于最大化 L_b ，因此，问题转变成求最大的 L_b 。

$$\log P(x) = L_b + KL(Q(z|x)||P(z|x)) \tag{2.12}$$

对于求解 L_b 的最大值，首先要对 L_b 求导，才能使用梯度下降或梯度上升更新网络参数。那么下一步要做的就是推导出 L_b 的导数，求导前需要先对 L_b 变换，变换后的 L_b 如公式(2.13)所示，这也正是 VAE 的目标函数。

$$\begin{aligned}
L_b &= \int_z Q(z|x) \log \left(\frac{P(x|z)P(z)}{Q(z|x)} \right) dz \\
&= \int_z Q(z|x) \log \left(\frac{P(z)}{Q(z|x)} \right) dz + \int_z Q(z|x) \log P(x|z) dz \\
&= -KL(Q(z|x)||P(z)) + E_{Q(z|x)}[\log P(x|z)]
\end{aligned} \tag{2.13}$$

于是，最大化 L_b 就等价于最小化公式(2.13)中的 $KL(Q(z|x)||P(z))$ 并最大化 $E_{Q(z|x)}[\log P(x|z)]$ 。对这两项分开计算，KL 散度项实际就是约束编码器，使 x 编码出的隐向量 z 尽可能服从标准正态分布 $P(z)$ ，也即 $Q(z|x)$ 向 $N(0, I)$ 逼近。由于已经假设先验分布 $P(z)$ 是正态分布，即 $P(z) = N(0, I)$ ，因此可以直接将 KL 散度项代入正态分布的 KL 散度计算公式，从而得到公式(2.14)中的结果：

$$-KL(Q(z|x)||P(z)) = \frac{1}{2} \sum_{j=1}^d (1 + \log \sigma_{(j)}^2 - \mu_{(j)}^2 - \sigma_{(j)}^2) \quad (2.14)$$

其中 d 是隐变量 z 的维度, $\mu_{(j)}$ 和 $\sigma_{(j)}^2$ 分别代表正态分布的均值向量和方差向量的第 j 个分量。接下来根据求导法则对 $-KL(Q(z|x)||P(z))$ 求导, 这一步实际上就是对编码器的参数 θ 求导, 然后用梯度下降法进行反向传播更新参数即可。

公式(2.13)中的第二项是一个对数似然期望, 机器学习中常使用蒙特卡洛 (Monte Carlo, MC) 采样求解复杂积分期望问题, 但 MC 采样过程不可微, 因此也无法使用随机梯度下降来更新参数, 导致后续难以进行反向传播。故 VAE 中使用重参数技巧使该过程变得可微, 从而保证网络能够反向传播。重参数技巧其实就是参数变换, 即先从标准正态分布中采样一个高斯噪声 ε , 然后乘上编码器输出的标准差 σ 再加上均值 μ , 即 $z = \varepsilon * \sigma + \mu$ 。这样一来, 就可以对 $E_{Q(z|x)}[\log P(x|z)]$ 求导, 也就是对解码器中的参数 θ 求导, 然后进行梯度更新, 从而完成对整个模型的训练。由于模型是从编码器输出的 μ 和 σ 中采样 z , 再将 z 通过解码器解码出来, 理论上解码器生成数据的期望是真实数据, 因此可以用 $L2$ 损失来估计 $E_{Q(z|x)}[\log P(x|z)]$, 这也类似于自编码器中的重构误差, 如公式(2.15)所示:

$$E_{Q(z|x)}[\log P(x|z)] = \|\hat{x} - x\|^2 \quad (2.15)$$

其中 \hat{x} 是解码器生成的数据, x 是真实数据。

至此, 就完成了对公式(2.13)所示的 VAE 目标函数的变换推导, 在实际编程中使用的也是变换后的目标函数, 即:

$$L_b = \frac{1}{2} \sum_{j=1}^d (1 + \log \sigma_{(j)}^2 - \mu_{(j)}^2 - \sigma_{(j)}^2) + \|\hat{x} - x\|^2 \quad (2.16)$$

图 2.4 展示了 VAE 的整体架构。经上述分析可知, VAE 的本质就是在常规自编码器的基础上, 对编码器输出的向量加入高斯噪声, 从而使解码器对噪声具有鲁棒性, 额外添加的 KL 损失项则希望编码出的向量逼近标准正态分布。自编码器中编码出的向量, 在 VAE 中对应了神经网络拟合出的均值 μ , 而 VAE 中神经网络拟合出的标准差 σ , 则用来动态调节噪声的强度。当解码器还没有训练好时, 就适当降低噪声强度, 使解码器网络拟合时容易一些; 反之, 如果解码器训练得比较好, 就适当增加噪声强度, 使解码器网络拟合时困难一些, 这时解码器

就会尽力提高自身的生成能力。简言之，当图 2.4 中的高斯噪声 ε 为 0 时，VAE 会退化为常规的自编码器，目标函数中的 KL 散度项也将变得无意义，因此在实际使用 VAE 时要防止该问题的出现。

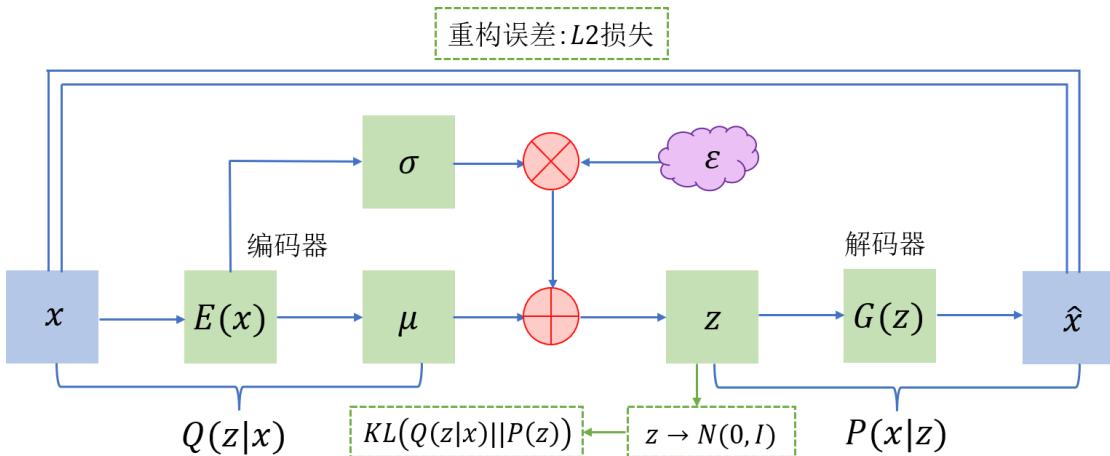


图 2.4 VAE 的整体架构

综上，VAE 训练过程的伪代码如算法 2.2 所示。

算法 2.2：VAE 的训练过程

输入： m 个小批量真实样本集 $X = \{x^{(1)}, \dots, x^{(m)}\}$

- 1 随机初始化编码器网络参数 ϕ 和解码器网络参数 θ ；
- 2 **repeat**
- 3 编码器对 X 编码得到均值 μ 和标准差 σ
- 4 使用重参数技巧 $z = \varepsilon * \sigma + \mu, \varepsilon \sim N(0, 1)$ 得到隐向量 z
- 5 解码器对 z 解码，得到生成数据集 $Y = \{y^{(1)}, \dots, y^{(m)}\}$
- 6 最小化目标函数 L_b ，反向传播更新参数 ϕ 和 θ ；
- 7 **until** 达到最大迭代次数

输出： 训练后的编码器网络参数 ϕ^* 和解码器网络参数 θ^*

2.3 计算机视觉中的注意力机制

人类通过眼睛观察万物，感知世界上的大量信息，由于人具备选择能力，因此可以免受海量信息的干扰，这也是因为人眼处理信息时通常会选择重要信息，

而忽略不重要信息。人类在观看图像时，并非能够对图像的所有信息给予相同的关注度，而是将注意力放在感兴趣区域，这种处理信息的机制称为注意力机制。注意力机制的核心思想就是让网络模拟人类视觉系统可以迅速高效地关注到重点区域的特性，其本质就是对网络的输入进行加权再输出[101]。

在计算机视觉领域，按照注意力关注域的不同，可以将其分为空间注意力机制、通道注意力机制、时间注意力机制、混合注意力机制和自注意力机制等。对视觉任务而言，最常使用的还是通道注意力、空间注意力以及二者混合的注意力。通道注意力关注图像的众多特征里哪些是有意义的，特征图中的每个通道都被视作是一种特征检测器，这些特征包括但不限于图像中的点、线、面、明暗度等。而空间注意力关注的是图像中哪些区域的特征是有意义的，图像中的不同区域对任务的贡献并非同等重要，只有任务相关的区域才需要格外注意，即空间注意力关注的是特征的位置信息。混合注意力则是通道注意力和空间注意力的结合。

通道注意力最具代表性的模型是 Hu 等人[102]在 2017 年提出的 SENet (Squeeze-and-Excitation Networks)，该模型使用了 SE 模块，主要包括三个部分：squeeze(压缩)、excitation(激励)和 scale(缩放)。对于输入的大小为 $H' \times W' \times C'$ 的特征图 X ，首先使用卷积操作 F_{tr} 将其变换为大小为 $H \times W \times C$ 的特征图 U ；然后对特征图 U 做压缩操作，即 $F_{sq}(\cdot)$ ，将特征图变成 $1 \times 1 \times C$ 大小，并得到通道级的全局特征，其中压缩操作由全局平均池化实现；接下来是对全局特征进行激励操作，即 $F_{ex}(\cdot, W)$ ，该操作用以捕捉通道间的关系，通过将 $1 \times 1 \times C$ 大小的特征图送入神经网络和 sigmoid 函数来生成每个通道的权重；最后是缩放操作 $F_{scale}(\cdot, \cdot)$ ，将 $1 \times 1 \times C$ 大小的特征权重图与特征图 U 在通道上相乘，从而完成在通道维度上对原始特征的重新标定，图 2.5 展示了 SENet 的模型结构。SENet 提出后，后续很多工作又基于它做了改进，GSoP-Net[103]和 FcaNet[104]都改进了 SENet 中的 squeeze 部分；ECA-Net[105]对 SENet 中的 excitation 部分做了改进；SRM[106]和 GCT[107]对 SENet 中的 scale 部分进行改进，这些工作都实现了对 SENet 性能的提升。

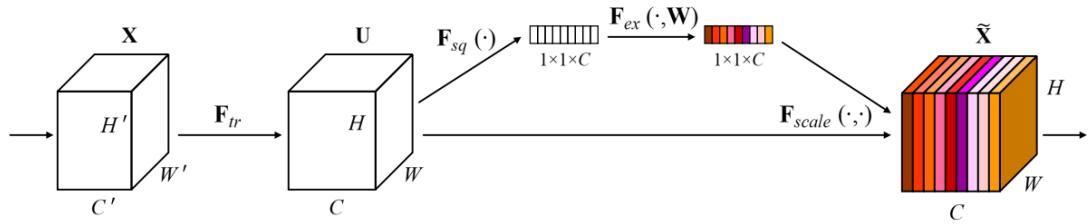


图 2.5 SENet 的模型结构[102]

空间注意力机制最具代表性的模型是 STN (Spatial Transformer Networks, STN) [108]，STN 引入了一个可学习的空间转换模块 ST (Spatial Transformer)，该模块能够对各种形变数据在空间中进行转换，并且能够自动捕获重要区域的特征，使模型具有更鲁棒的空间不变性。ST 模块由 Localisation net、Grid generator 和 Sampler 三个子模块构成， U 和 V 分别是输入图和输出图，如图 2.6 所示。Localisation net 通过卷积神经网络提取图像的特征来预测变换矩阵 θ ；Grid generator 利用 Localisation net 模块预测出的 θ 对图片中的位置进行变换，需要说明的是，图片位置指的是图片像素对应的坐标；由于变换时会遇到计算出的位置坐标是小数的情况，因此 STN 使用 Sampler 操作来解决这个问题，该操作具体采用的方法是双线性插值。ST 模块可以加到卷积神经网络的任意位置，且计算量较少。

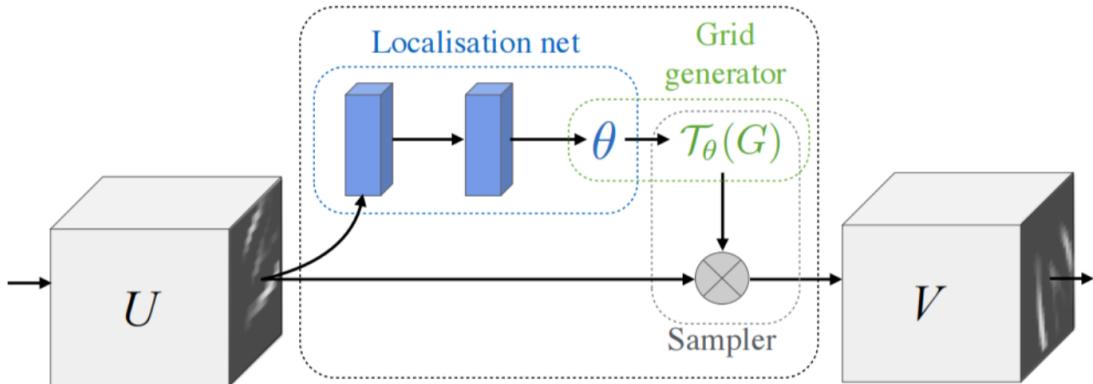


图 2.6 ST 模块结构图[108]

CBAM[109]是混合注意力机制的代表模型，它结合了空间注意力和通道注意力。一般情况下，相比于 SENet 只关注通道注意力，CBAM 可以取得相对更好的效果。图 2.7 展示了 CBAM 的结构，给定一个输入到网络中的特征图，然后沿着通道和空间两个维度依次推断出注意力权重，再将权重与原特征图相乘，从而实现特征的自适应调整。图 2.8 展示了 CBAM 中的通道注意力子模块和空间注

意力子模块。

在通道注意力模块中，输入尺寸大小为 $H \times W \times C$ 的特征图 F ，对特征图 F 进行基于特征图高、宽的全局最大池化（MaxPool）和全局平均池化（AvgPool），得到两个大小为 $1 \times 1 \times C$ 的特征图，再利用共享的全连接层（Shared MLP）对这两个特征图进行处理，然后将处理后的两个特征图相加，接着使用 Sigmoid 激活函数获得尺寸大小为 $1 \times 1 \times C$ 的通道注意力图 M_c ，也就是获得了输入特征图 F 每个通道的权重。利用该模块得到通道注意力图后，再将通道权重通过乘法运算逐通道加权到输入特征图 F 上，即可得到特征图 F' 。

在空间注意力模块中，输入特征是经过通道注意力模块处理后的特征图 F' ，尺寸为 $H \times W \times C$ 。为了计算空间注意力，首先要对 F' 做基于特征图通道的全局平均池化和全局最大池化，得到两个 $H \times W \times 1$ 的特征图，然后将这两个特征图进行通道拼接得到尺寸为 $H \times W \times 2$ 的特征图，再使用 7×7 卷积操作将拼接后的特征图压缩为一个通道，即此时的特征图尺寸仍为 $H \times W \times 1$ ，接着使用 Sigmoid 函数获得输入特征图 F' 中每个特征点的权重值，即空间注意力图 M_s 。最后将 M_s 与空间注意力模块的输入特征图 F' 做乘法，即可得到处理后的新特征。

在通道注意力模块和空间注意力模块中，维度压缩的过程均由全局最大池化和全局平均池化实现，选用这两种池化方式的原因在于平均池化能有效学习目标对象的范围，最大池化能获得独特目标特征的重要信息，两种池化方式结合后能推断出更好的注意力图。

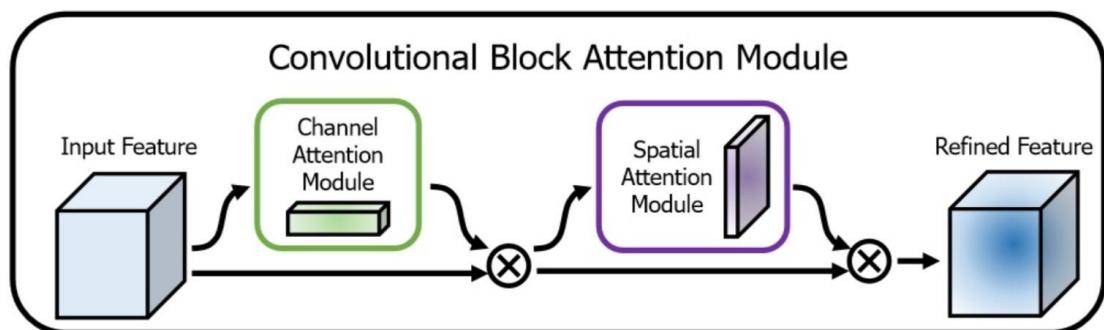


图 2.7 CBAM 网络结构图[109]

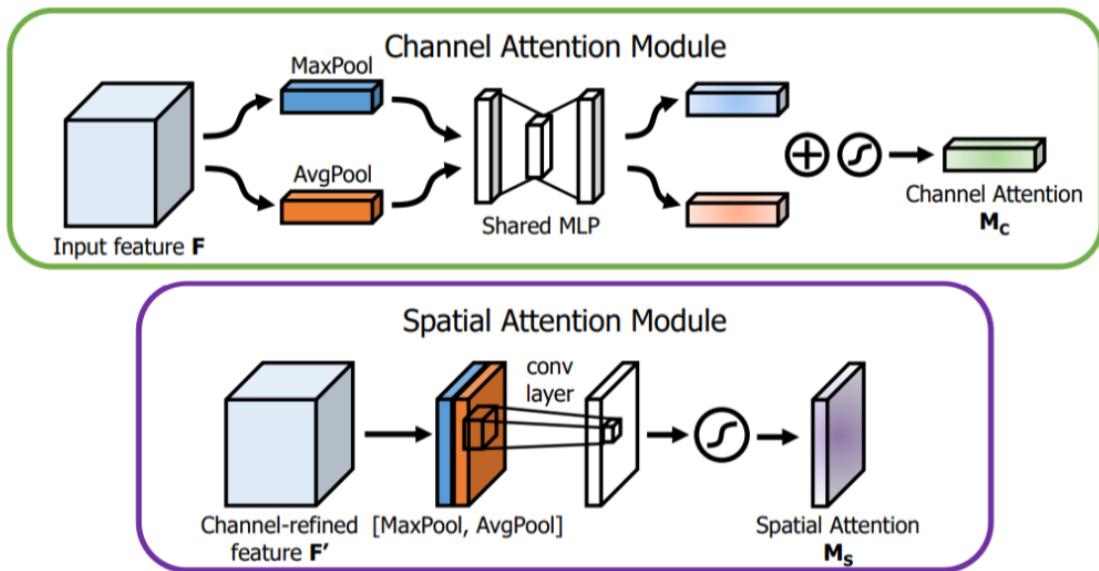


图 2.8 CBAM 中的通道注意力子模块和空间注意力子模块[109]

2.4 本章小结

本章首先介绍了图像滤波和数学形态学方法，对本文中用于图像增强的中值滤波和闭运算做了简要说明，并详细分析了无监督图像生成模型中GAN和VAE的基本原理及数学推导，最后介绍了计算机视觉中的注意力机制，浅析了通道注意力、空间注意力及混合注意力中的代表模型。

第三章 基于图像增强的热障涂层形貌特征识别

热障涂层因其极佳的耐高温性被广泛使用在各种高精尖仪器和设备上，如航空发动机以及轮船、汽车、能源等领域的热端部件，因此科研人员十分重视对热障涂层隔热性能的研究。等离子喷涂作为热障涂层陶瓷层最常用的制备方法之一，具备较高的灵活性和多功能化，但在制备过程中常会导致热障涂层形成独特的孔隙和裂纹结构。对孔隙和裂纹这两种形貌特征的识别及相关参数的计算是研究热障涂层隔热性能的重要因素，而本章的研究重点是识别热障涂层中的裂纹。

人工识别热障涂层中的裂纹并计算其形貌参数，速度慢、效率低、误差大，因此本章提出基于机器视觉的方法，自动化识别热障涂层中的裂纹，并计算裂纹长度。由于数据集包含的图像样本极少，仅有 5 张图像，因此难以使用基于深度学习的视觉算法处理图像。此外，热障涂层图像中的裂纹数量很多，且大部分裂纹存在像素点缺失的情况，这导致直接识别裂纹的结果很差，需要提前对断裂的裂纹进行修复，故借助传统图像增强方法提高图像质量，降低裂纹识别的难度。

3.1 热障涂层形貌特征识别概述

材料微观形貌学和图像处理技术是材料结构分析的重要组成部分和关键环节。对热障涂层形貌特征的识别与形貌参数计算主要包含两部分：一是识别并去除涂层中的孔隙，使图像中只包含裂纹；二是识别出涂层中的裂纹并计算其形貌参数，即裂纹的长度。

图 3.1 展示了由扫描电镜拍摄出的热障涂层图像及图中孔隙和裂纹的局部放大图，图 3.1(a)所示的热障涂层图像清晰地展示出孔隙和裂纹的数量之多，从图 3.1(b)也能看出孔隙轮廓非常不规则。因此在进行裂纹识别任务之前，要首先去除众多孔隙。其次，在喷涂热障涂层时，飞溅的涂层粒子落在材料表面形成了许多毫无意义的痕迹，即干扰噪声，这些噪声点会干扰裂纹的检测识别，因此要在保证不影响裂纹的情况下去除这些噪声。此外，由于很多裂纹存在像素点的缺失，计算机会将视觉感知为一个的裂纹线条识别为多条断线，因此需要先对这些断掉的裂纹进行连接修补，再将其识别为正确的整条裂纹。再者，裂纹的形状不规则，

既有曲线也有较平缓的直线段，因此采取仅能识别直线段的方法并不可取，需要设计一种能识别各种形态裂纹的新方法。

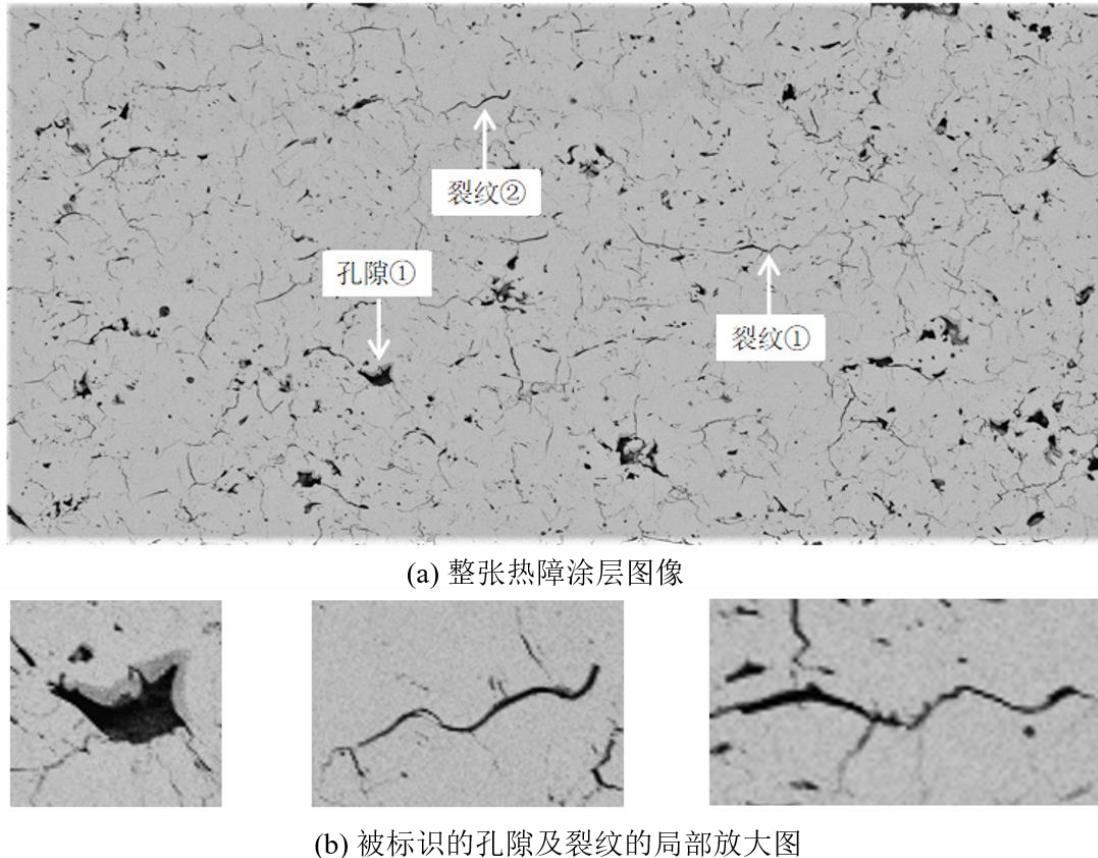


图 3.1 扫描电镜拍摄出的热障涂层图像及局部放大图

针对上述问题，本章提出了基于图像增强的热障涂层形貌特征识别方法，整体流程图如图 3.2 所示。首先，对原始图像进行颜色通道转换，将其转成单通道灰度图；然后使用阈值分割、形态学和图像滤波等操作为图片去噪，同时加强图像细节和孔隙轮廓的边缘，对图像进行初步增强；接着将二值化后的图像与增强后的图像做差，以去除孔隙；再使用基于闭运算的断线修复连接像素点缺失的裂纹断线；最后结合图像细化、图像反二值化和遍历搜索像素点等方法完成对热障涂层中裂纹的自动识别与长度计算，最终输出裂纹识别的结果图及裂纹长度列表。该方法对热障涂层图像上进行了大量增强，修复了像素点缺失的裂纹，降低了裂纹识别的难度，相比于人工识别，该方法更加便捷高效。

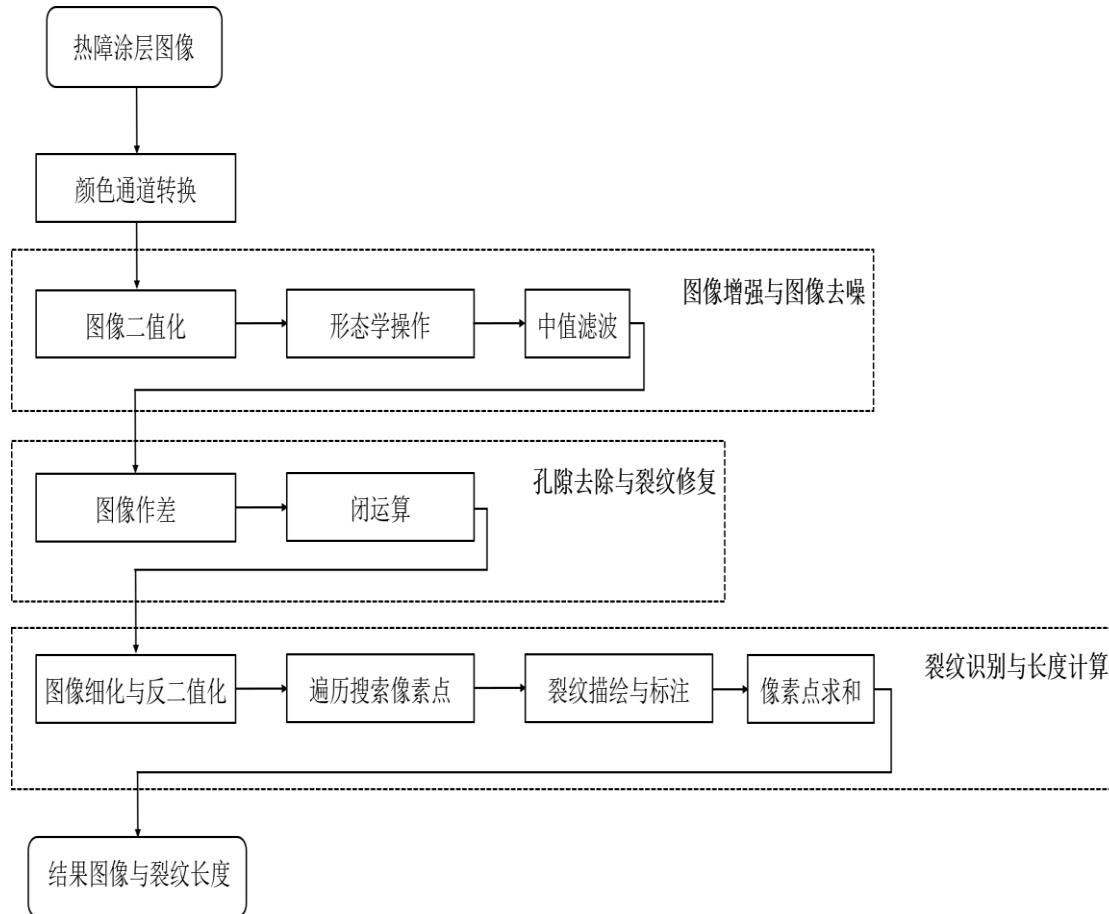


图 3.2 裂纹识别和长度计算流程图

3.2 基于图像增强的热障涂层裂纹识别

3.2.1 孔隙轮廓增强与图像去噪

在识别热障涂层裂纹之前，需要先识别涂层中的孔隙并去除。为了突显出热障涂层中孔隙的轮廓，需要二值化待处理的热障涂层图像。Otsu 阈值分割方法能够计算一幅图像的最佳阈值，而不需要输入任何附加参数。与同样不需要输入附加参数的迭代均值和均值阈值比较，Otsu 给出的阈值能使像素分类更加均匀，故采用 Otsu 阈值分割法。为了更好的消除图像中的噪声，并使孔隙去除的更干净，使用 Otsu 阈值处理的同时还结合了反二值化阈值处理。但在此之前需要使用色彩空间转换将原始的 RGB 图转换为单通道灰度图，这是因为 Otsu 算法建立的基础是图像的灰度直方图。

Otsu 阈值的计算流程为：(1) 计算图像的直方图，将 0-255 依次当作阈值，

把直方图强度大于阈值的像素分成一组，把小于阈值的像素分成另一组；(2) 分别计算这两组的组内方差，并把两个组内方差相加；(3) 比较每个阈值下的组内方差和，总和最大的就是结果阈值；(4) 以结果阈值为标准进行图像二值化操作，将灰度值大于此阈值的像素点赋为 255，将灰度值小于此阈值的像素点赋为 0。由于 Otsu 算法是依据类间距离极大准则来确定区域分割阈值，因此这种方法能在最大限度内降低误分割的概率，有效增强了热障涂层中孔隙的轮廓。

由于热障涂层图像中存在大量噪声，因此还需使用数学形态学操作进行图像去噪。去噪时若直接使用腐蚀操作，部分孔隙和裂纹也会被腐蚀，其完整性会遭到破坏。由于基本的形态学操作是非线性的，后续想要完整地恢复孔隙轮廓和裂纹形态也十分困难。因此，为了防止腐蚀操作使孔隙和轮廓出现缺失，首先需要对图像进行形态学膨胀后再对图像腐蚀，以去除干扰噪声，然后再次使用膨胀操作，恢复去除噪声后孔隙和裂纹的大致形态。

经过上述去噪操作后，部分孔隙和裂纹的形态细节会丢失，因此需要将上述结果图与二值化后的原图进行逻辑与运算，从而在去噪后的图像中重新引入孔隙和轮廓丢失的信息。图像的逻辑与运算是将两张二值图像的对应像素进行逻辑与。逻辑与运算的具体操作如公式(3.1)所示：

$$d(x, y) = s1(x, y) \text{ and } s2(x, y) \quad (3.1)$$

其中 $s1(x, y)$ 表示第一张二值图像的像素矩阵， $s2(x, y)$ 表示第二张图像的像素矩阵， $d(x, y)$ 为输出图像的像素矩阵，且输出图像和输入图像具有相同的大小和通道数。

接下来还需再次进行膨胀操作，以填补孔隙内部的空白，使后续识别孔隙时更容易找到孔隙的外轮廓，而不提取出内轮廓。经过上述步骤后，图像中仍然存在部分较小的孤立噪声，为了去除这些噪声并使孔隙边缘变得光滑，需要使用中值滤波算法。中值滤波的基本原理是：取当前像素点邻域内奇数个像素值的中值作为当前像素点的像素值。如计算某张图像中第 5 行第 5 列像素点的中值滤波值，具体做法为：设该像素点邻域大小为 3×3 ，对邻域内像素点的像素值升序排列或降序排列，假设升序排序后的序列为 [67, 79, 91, 92, 94, 95, 96, 98, 102]，该序列的中值是 94，于是将该值作为第 5 行第 5 列像素点的新像素值。中值滤波作为一种基于排序理论的非线性平滑技术，运算过程简单，且能有效消除杂散噪声。

而不会或较小程度地造成边缘模糊。

以上就是对孔隙轮廓增强与图像去噪的处理流程，经过上述操作，增强了图像质量，减轻了后续图像处理的压力。

3.2.2 孔隙去除与裂纹修复

热障涂层图像经过 3.2.1 节的处理后，图像中只包含了增强后的涂层孔隙，通过将二值化后的原图与只包含孔隙的图像作差，以去除热障涂层中的孔隙，从而得到只包含裂纹的图像。图像作差的形式化表示如公式(3.2)所示：

$$g(x, y) = f1(x, y) - f2(x, y) \quad (3.2)$$

其中 $g(x, y)$ 为输出图像的像素矩阵， $f1(x, y)$ 和 $f2(x, y)$ 为要作差的两张二值图像的像素矩阵。

经上述操作后仍不能直接识别裂纹，因为部分裂纹像素点的缺失易导致识别结果出现偏差。由于像素点的缺失，计算机会将视觉感知为整条曲线或线段的裂纹识别为多条断线，因此需要对此类裂纹进行基于闭运算的裂纹修复。

数学形态学中的闭运算通过填充图像的凹角来滤波图像，该操作能够填平前景物体内的小孔并弥合小裂缝，而总的位置和形状不变。闭运算是先膨胀、后腐蚀的运算，其形式化表示如公式(3.3)所示：

$$X \cdot B = (X \oplus B) \ominus B \quad (3.3)$$

其中 X 是待处理的图像， B 是结构元素， \oplus 表示膨胀操作， \ominus 表示腐蚀操作。

图 3.3 展示了使用闭运算将图 3.3(a) 中非闭合区域的缺口连至闭合的过程，结果图如图 3.3(b) 所示，黄色框圈出了连接前后的缺口位置。

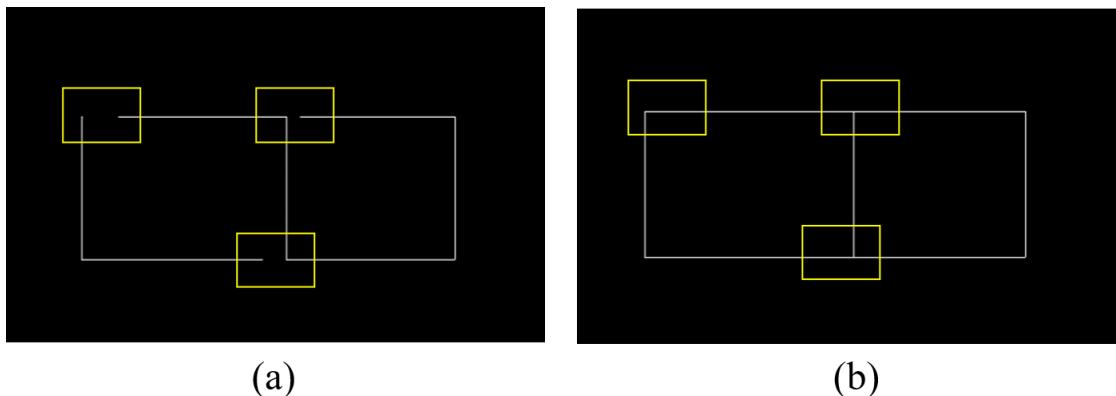


图 3.3 闭运算连接断线的效果图

对热障涂层中像素点缺失的裂纹而言，闭运算是一种非常简单有效的修复和增强手段，使用闭运算对断线连接修补的同时还能进一步去除图像上的残留噪声。

3.2.3 裂纹识别与长度计算

对裂纹修复完毕后，需要细化裂纹。图像细化是将图像的线条从多像素宽度减少到单像素宽度的过程，也称骨架提取或骨骼化。“骨架”是指一幅图像的骨骼部分，能够描述物体的几何形状和拓扑结构，是重要的图像描绘子之一。计算骨架的过程就是对图像骨骼化，即图像细化，对目标物体进行细化有助于突出目标的形状特点和拓扑结构，减少冗余信息。图像细化后，裂纹变成单像素宽度，又因为在3.2.1节中使用了反二值化阈值处理，这导致细化后的裂纹在黑色背景中显示的并不清晰。因此，在图像细化后，再次使用反二值化阈值处理，将图像背景变为白色、裂纹变为黑色，从而能够更清晰的显示裂纹。

裂纹识别使用的是遍历搜索像素点的方式，步骤如下：（1）遍历寻找细化后图像中的黑色裂纹像素点；（2）由于找到的点不一定是裂纹的端点，因此需要从该点开始分别向两端寻找相邻的裂纹像素点；（3）重复搜索裂纹像素点，延伸寻找裂纹曲线；（4）搜索至没有相邻点的裂纹像素点，则表示单条裂纹搜索完成；（5）重复步骤（1）至步骤（4），直至遍历完图像中所有的裂纹像素点。详细流程如算法3.1所示。

最后对识别出的裂纹线条用随机颜色描绘，并标注序号、计算长度，具体步骤为：（1）对识别出的每个裂纹线条随机上色、标注序号；（2）按照序号依次计算每条裂纹的长度，长度由单像素裂纹线条的像素点求和计算得到；（3）大于长度阈值的裂纹仍保持步骤（1）中的随机颜色，小于等于长度阈值的裂纹不予显示，将其变为图像背景色，即白色，形式化表示如公式(3.4)所示。

$$c(x, y) = \begin{cases} rand_val, & len > thresh \\ 255, & otherwise \end{cases} \quad (3.4)$$

其中， $c(x, y)$ 是裂纹线条 (x, y) 处的像素值， $rand_val$ 是不包含0和255的随机像素值， len 是裂纹长度， $thresh$ 是可灵活设定的裂纹长度阈值。

算法 3.1: 裂纹的遍历搜索算法

输入: 仅包含单像素宽度裂纹的二值图像 Z (裂纹为黑色, 背景为白色)

- 1 初始化一个用来存放图像中所有裂纹的容器 V ;
- 2 搜索图像中黑色的裂纹像素点 (x, y) ;
- 3 **while** 存在 (x, y) **do**
- 4 初始化一个用来存放单条裂纹像素点的队列 D ;
- 5 将 (x, y) 像素点置为白色, 防止重复寻找, 并将其存入队列 D 中;
- 6 在 (x, y) 八邻域内搜索一端相邻的裂纹像素点 (x_m, y_m) ;
- 7 **while** 存在 (x_m, y_m) **do**
- 8 将 (x_m, y_m) 像素点置为白色, 防止重复寻找, 并将其存入队列 D 中;
- 9 在 (x_m, y_m) 八邻域内搜索相邻的裂纹像素点 (x'_m, y'_m) ;
- 10 **if** 存在 (x'_m, y'_m) **then**
- 11 $(x_m, y_m) \leftarrow (x'_m, y'_m)$;
- 12 **else**
- 13 跳出循环;
- 14 **end if**
- 15 **end while**
- 16 在 (x, y) 八邻域内搜索另一端相邻的裂纹像素点 (x_n, y_n) ;
- 17 **while** 存在 (x_n, y_n) **do**
- 18 将 (x_n, y_n) 像素点置为白色, 防止重复寻找, 并将其存入队列 D 中;
- 19 在 (x_n, y_n) 八邻域内搜索相邻的裂纹像素点 (x'_n, y'_n) ;
- 20 **if** 存在 (x'_n, y'_n) **then**
- 21 $(x_n, y_n) \leftarrow (x'_n, y'_n)$;
- 22 **else**
- 23 跳出循环;
- 24 **end if**
- 25 **end while**
- 26 将队列 D 存放至容器 V 中;
- 27 **end while**

输出: 存放所有裂纹的容器 V

3.3 实验验证

本节以图 3.4 中的 5 张热障涂层图像为实验对象，验证所提方法的可行性和有效性。这 5 张图像都包含很多像素点缺失的裂纹，且噪声很多，并存在较多孔隙。实验分为两个步骤进行，步骤一包括图像增强、图像去噪、孔隙去除、裂纹修复及裂纹细化；步骤二包括裂纹识别与长度计算。

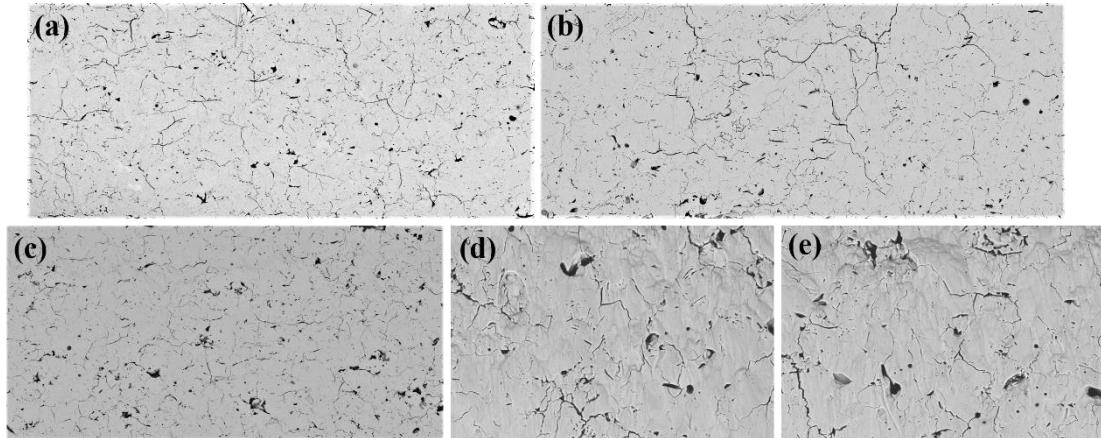


图 3.4 5 张热障涂层图像

针对图 3.4 中的 5 张图像，经过步骤一后，会得到图 3.5 所示的中间结果图。此时图中的裂纹已经被细化成一个像素的宽度，并且图像中只包含裂纹骨架，孔隙及部分噪声点已经去除。

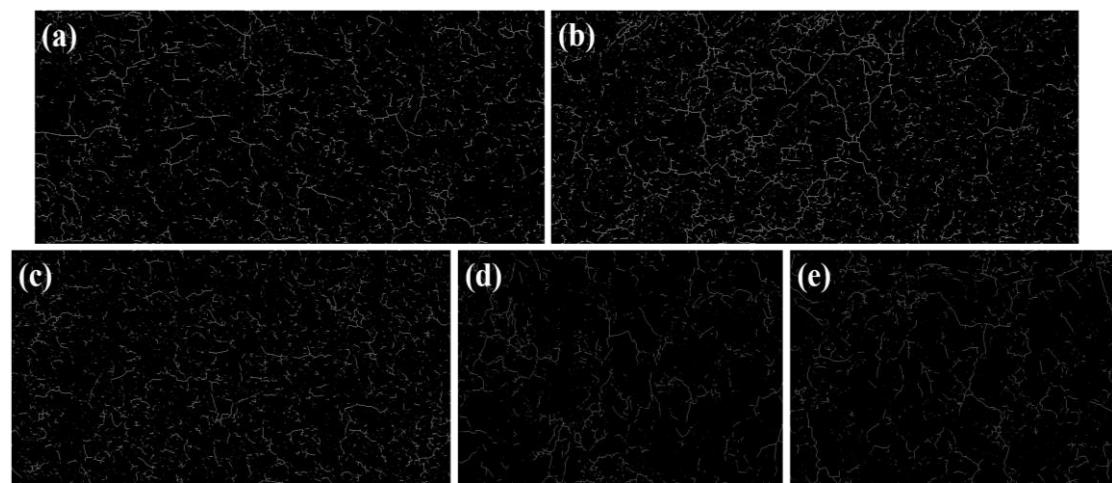


图 3.5 5 张热障涂层图像处理后的中间结果图

图 3.6 对比了热障涂层图像两处不同区域放大后的原图与中间结果图，图 3.6(a)和图 3.6(c)是原图，图 3.6(b)和图 3.6(d)分别是图 3.6(a)和图 3.6(c)处理后的图像。红色框圈出的部分比较了同一条裂纹使用闭运算修复前后的形态，可以看

出, 像素点缺失的裂纹得到了修补, 蓝色框圈出的部分也显示孔隙去除效果较好。

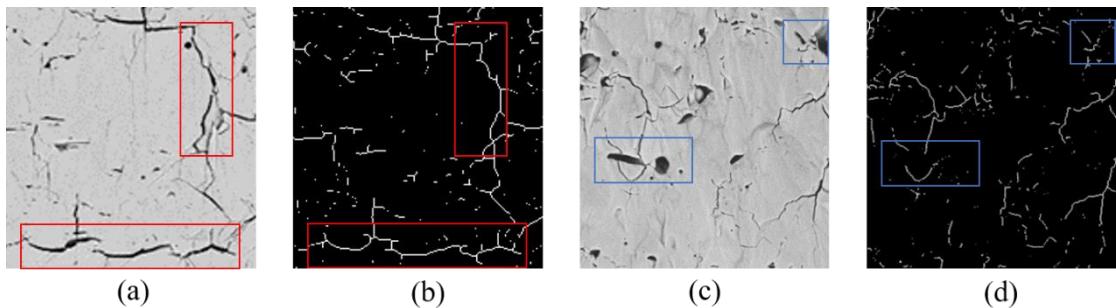


图 3.6 热障涂层图像不同区域放大后的原图与中间结果图

对细化后的裂纹图像, 再次进行反二值化处理, 以清晰的显示裂纹。图 3.7 展示了反二值化阈值处理后的 5 张热障涂层图像, 此时图像中的裂纹变为黑色, 图像背景为白色。

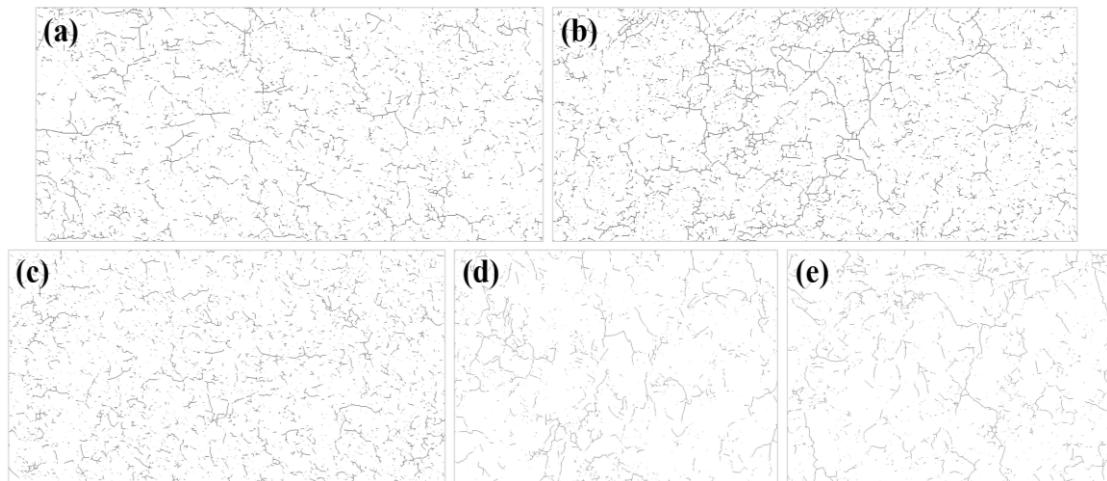


图 3.7 反二值化处理后的热障涂层图像

进行上述处理后, 再使用 3.2.3 节中遍历搜索像素点的方法识别图像中的裂纹线条, 并计算长度。裂纹长度的计算使用的是像素点求和的方式, 对于图像中大于设定阈值的裂纹进行描绘和长度计算, 小于设定阈值的裂纹将不予显示。将裂纹的长度阈值设置为不同的值, 则显示的裂纹数量也不相同, 阈值越小识别出的裂纹数量越多, 阈值越大识别出的裂纹数量越少。

图 3.8 展示了当阈值设置为 20pt、30pt 和 40pt 时, 图 3.4(d)这张热障涂层图像中的裂纹识别结果, pt 是裂纹的长度单位。从图 3.8 可以看出, 长度阈值越大, 识别出的裂纹数量越少。

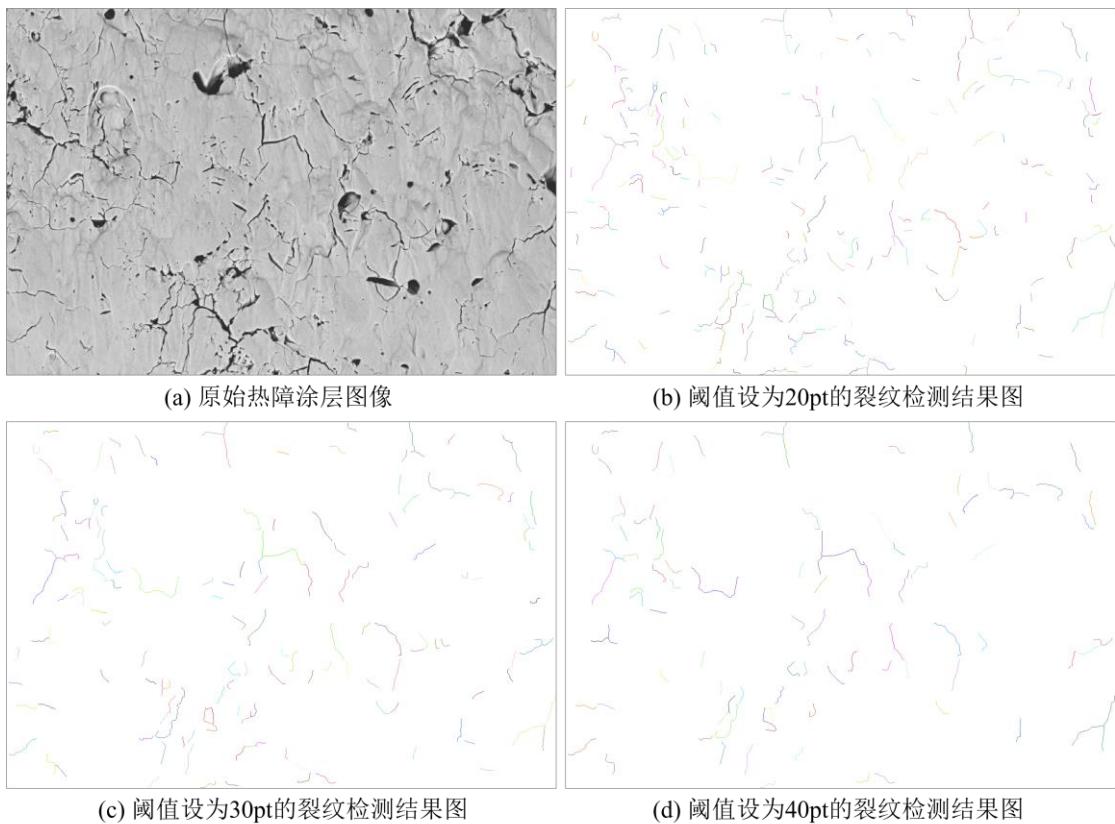


图 3.8 不同长度阈值下热障涂层图像中裂纹的识别结果

表 3.1 也展示了不同长度阈值下识别出的裂纹数量的对比，随着长度阈值的增大，识别出的裂纹数量也在逐渐减少，这也验证了图 3.8 中的识别结果，表中裂纹的数量单位为条。

表 3.1 不同长度阈值下识别出的裂纹数量对比

| | 阈值=20pt | 阈值=30pt | 阈值=40pt |
|-------|---------|---------|---------|
| 图像(a) | 183 | 73 | 40 |
| 图像(b) | 246 | 116 | 69 |
| 图像(c) | 200 | 97 | 51 |
| 图像(d) | 238 | 145 | 91 |
| 图像(e) | 240 | 150 | 110 |

本文为保证结果的准确性，在征求材料领域专家的意见后，将长度阈值设为 20pt，图 3.9 至图 3.13 展示了在阈值等于 20pt 的条件下，5 张热障涂层图像的裂纹识别与标注结果。

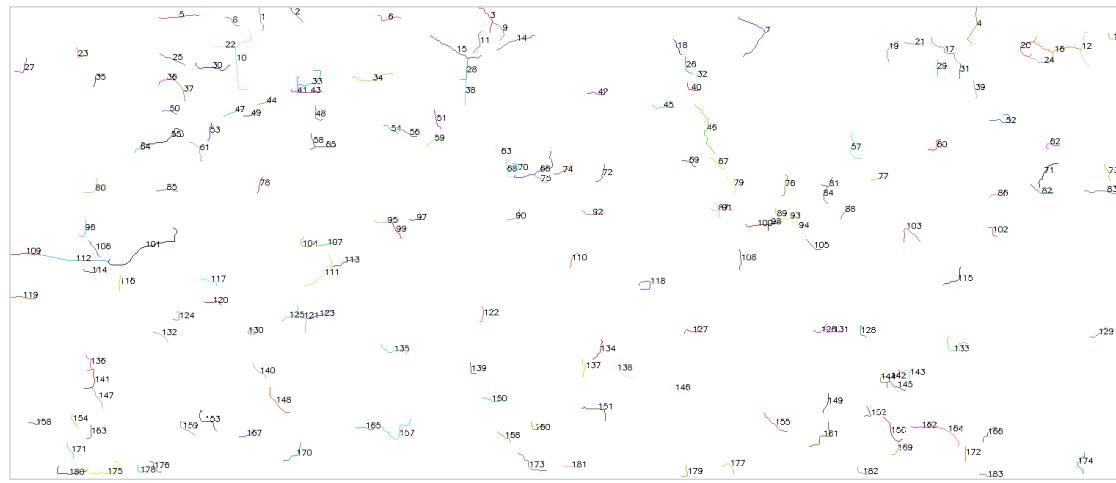


图 3.9 热障涂层图像(a)的裂纹识别与标注结果图

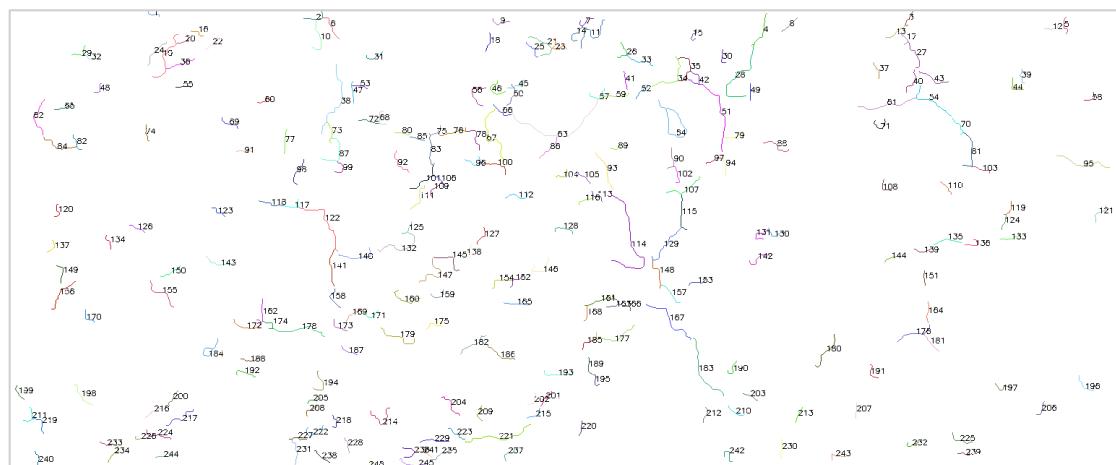


图 3.10 热障涂层图像(b)的裂纹识别与标注结果图

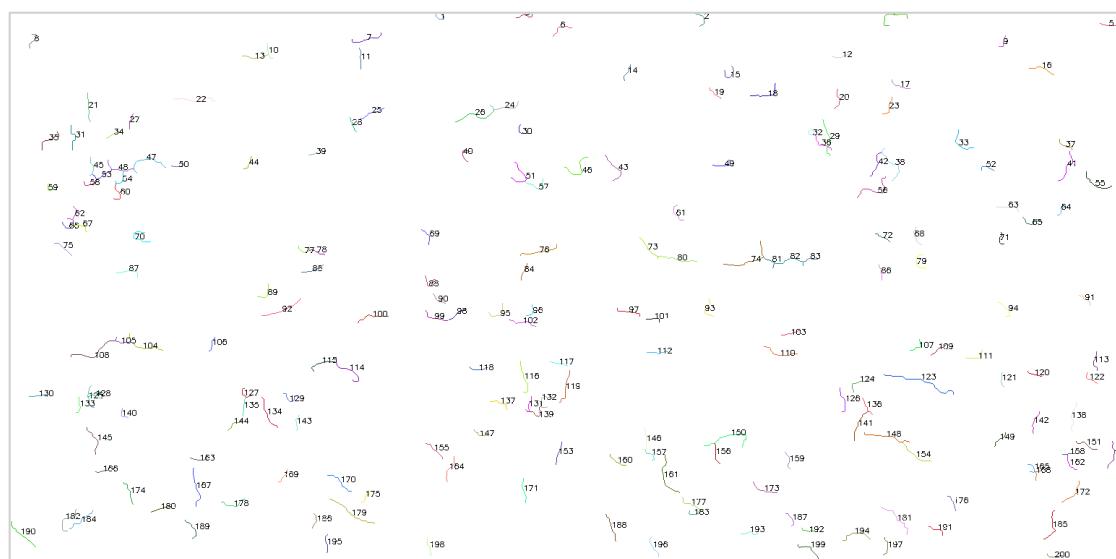


图 3.11 热障涂层图像(c)的裂纹识别与标注结果图

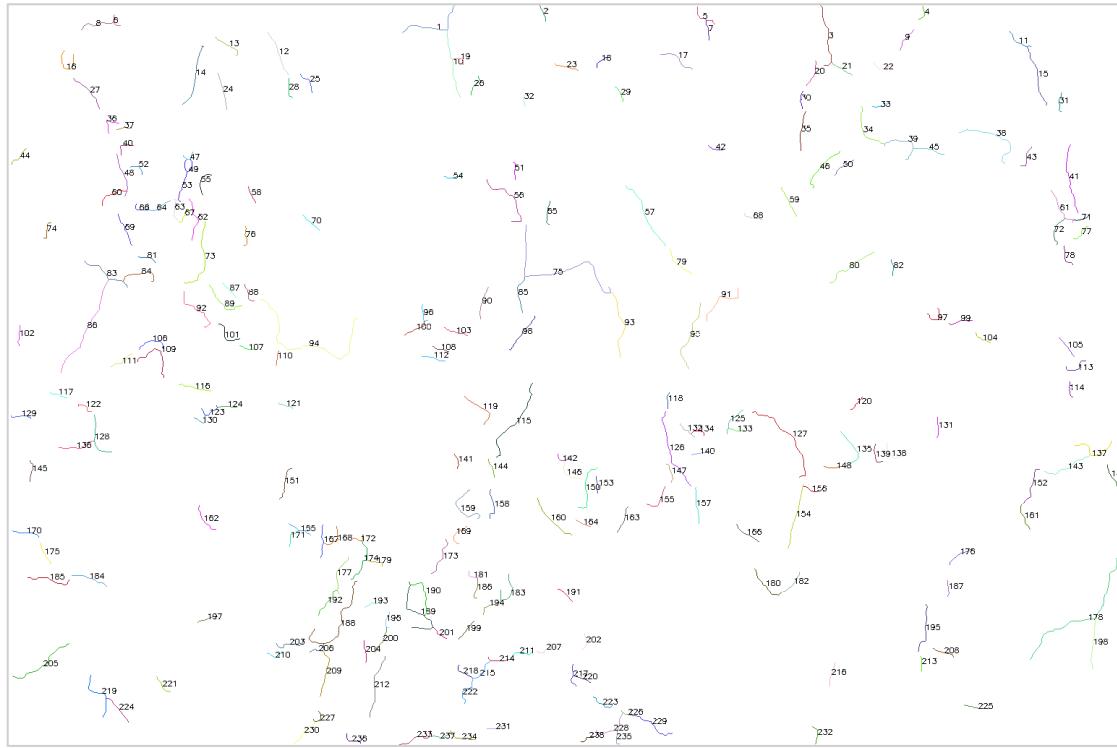


图 3.12 热障涂层图像(d)的裂纹识别与标注结果图

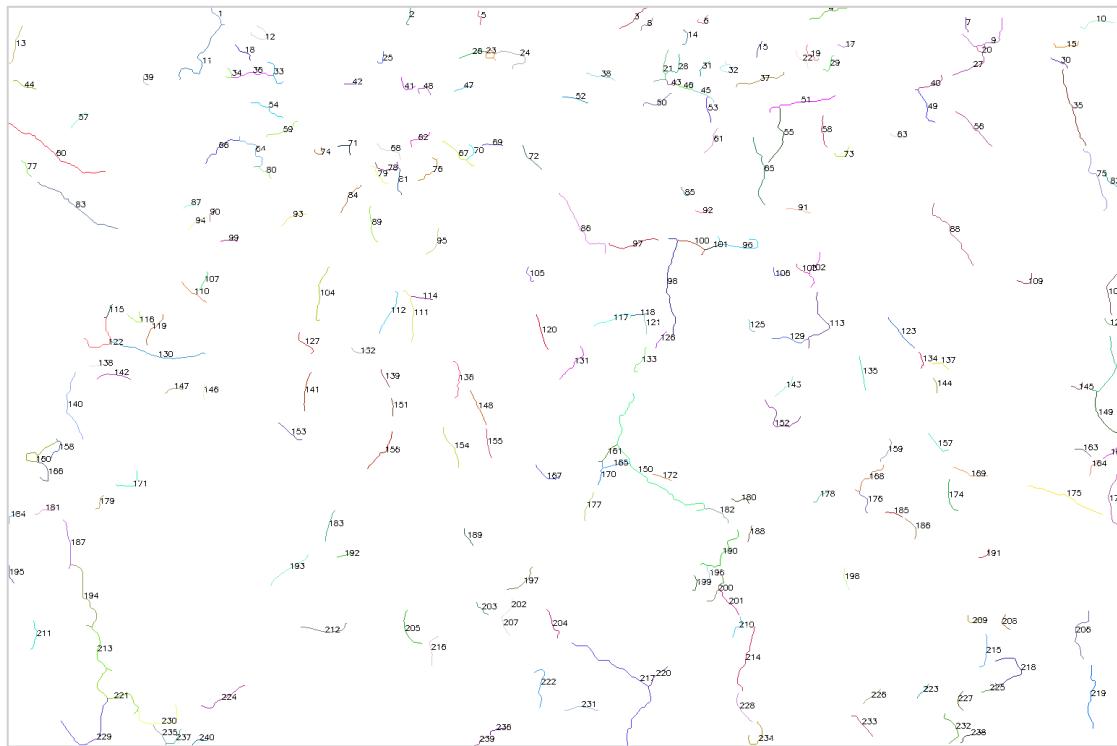


图 3.13 热障涂层图像(e)的裂纹识别与标注结果图

最后是计算识别出的裂纹长度，由于长度阈值设为 20pt 时，识别出的裂纹数量多达上百条，故只展示 5 张热障涂层图像中前 6 条裂纹的长度，具体结果如表

3.2 所示，单位为 pt。

表 3.2 5 张热障涂层图像中前 6 条裂纹长度的计算结果

| | 裂纹 1 | 裂纹 2 | 裂纹 3 | 裂纹 4 | 裂纹 5 | 裂纹 6 |
|-------|------|------|------|------|------|------|
| 图像(a) | 33 | 24 | 42 | 51 | 55 | 36 |
| 图像(b) | 21 | 36 | 25 | 50 | 27 | 37 |
| 图像(c) | 21 | 27 | 26 | 56 | 24 | 33 |
| 图像(d) | 100 | 23 | 90 | 22 | 39 | 22 |
| 图像(e) | 31 | 27 | 45 | 55 | 22 | 23 |

3.4 热障涂层裂纹识别软件的研发

由于前文所提方法能简单高效地识别出热障涂层中的裂纹，并自动计算裂纹长度，提高了材料领域研究者的工作效率。大多数科研人员希望将该方法集成到可直接使用的系统或软件中，以方便使用。因此本节设计了两款软件完成裂纹的识别与长度计算，第一款软件主要用于热障涂层图像增强和图像中裂纹的细化，第二款软件用于识别裂纹并计算其长度。未将两款软件集成于一个软件中，出于两方面的考虑：一是软件使用场景的多样性，即对于需求是图像细化的材料图像，只需使用第一款软件即可，而对于需要识别并计算图像中线条长度的任务，只需使用第二款软件；二是软件后续的迭代开发，对于第一款软件中集成的图像增强方法来说，在未来工作中大概率要进一步改进，因此无需将该软件的功能融合进第二款软件。本节对这两款软件分别做出详细说明，首先介绍热障涂层图像增强与裂纹细化软件的设计与实现，再介绍热障涂层裂纹识别与长度计算软件的设计与实现。

3.4.1 图像增强与裂纹细化软件的设计与实现

图像增强与裂纹细化软件的整体流程如图 3.14 所示，共包括四个功能模块：

(1) 图像导入。该模块主要负责图像导入功能，用户从本地文件夹中选择任意一张热障涂层图像并打开，打开的图像将展示在软件中。

(2) 图像预览。该模块主要负责展示图像，预览功能包括原始图像的展示和结果图像的展示，两种图像分别展示在不同的显示框里。

(3) 图像增强。该模块按照 3.2 节中介绍的方法完成初步的图像增强、裂纹修复以及裂纹细化，具体操作包括对图像颜色通道的转换、图像二值化、图像滤波、形态学操作、基于闭运算的裂纹修复以及图像细化，这些操作都已经被集成于图像增强模块中。

(4) 结果保存。增强后的图像被自动保存至原始图像所在的路径，结果图像的格式也与原始图像的格式一致，结果图像的名称为“result”。

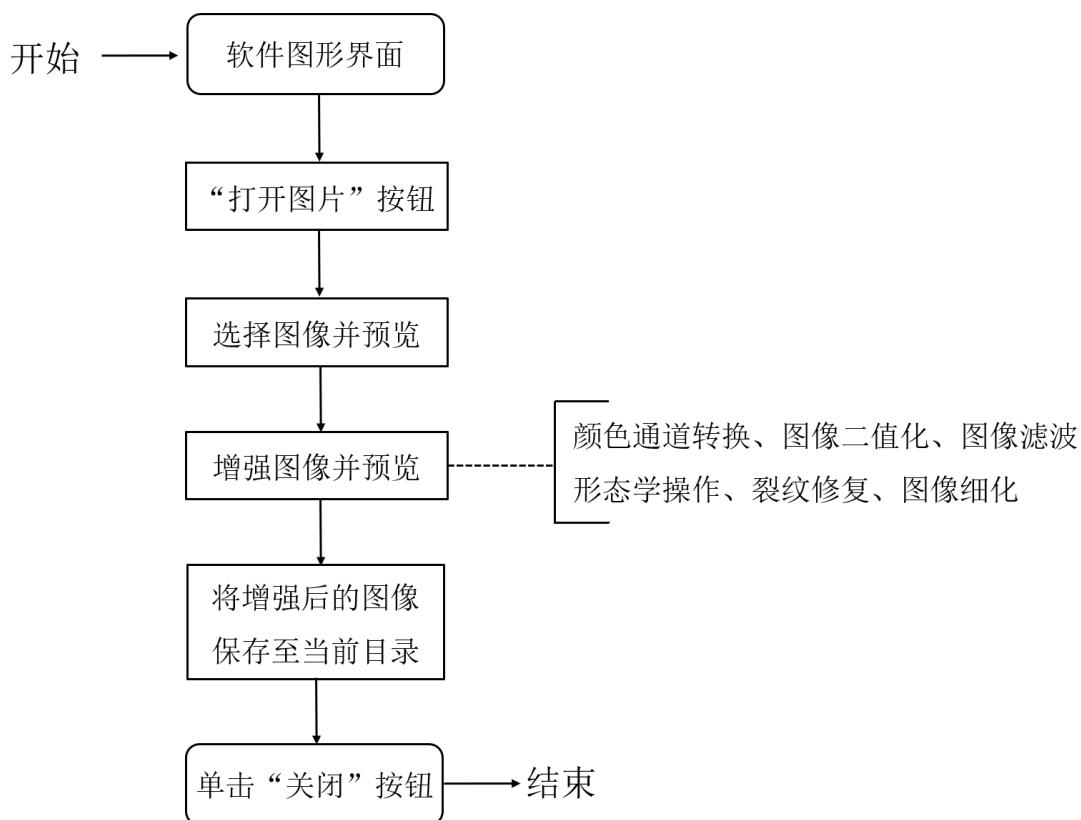


图 3.14 图像增强与裂纹细化软件的整体流程

设计图像增强与骨架提取软件所使用的开发语言是 Python，开发工具为 PyCharm，图形用户界面（Graphical User Interface，GUI）的设计工具选用的是 PyQt5。PyQt5 可视为 Python 版本的 Qt 框架，相较于 Python 自带的标准图形用户界面库 Tkinter，PyQt 功能更加强大、界面更加美观、跨平台效果好，且简单易用。而 Tkinter 虽是 python 的内置图形界面开发框架，但是没有 QT Designer 和 QSS 的支持，常导致软件开发的工作量十分庞大。

以下是软件的具体实现。图 3.15 展示了软件的主体界面，主要包含一个“打开图片”按钮和两个图像展示窗口，一个窗口显示打开的原始热障涂层图像，令一个显示处理后的图像。

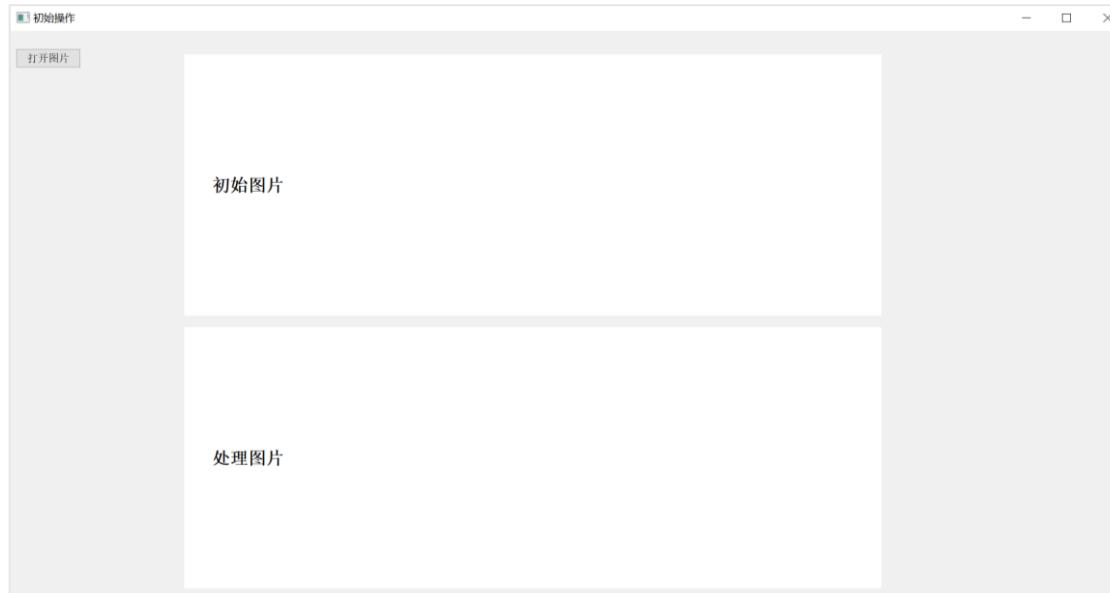


图 3.15 软件界面展示

单击“打开图片”按钮即可实现图像导入，图 3.16 展示了该功能。打开本地文件夹，文件格式选择“All Files”即可对路径下所有格式的图像文件进行选择，图像导入后，即可在“初始图片”的显示窗口预览选择的图像。

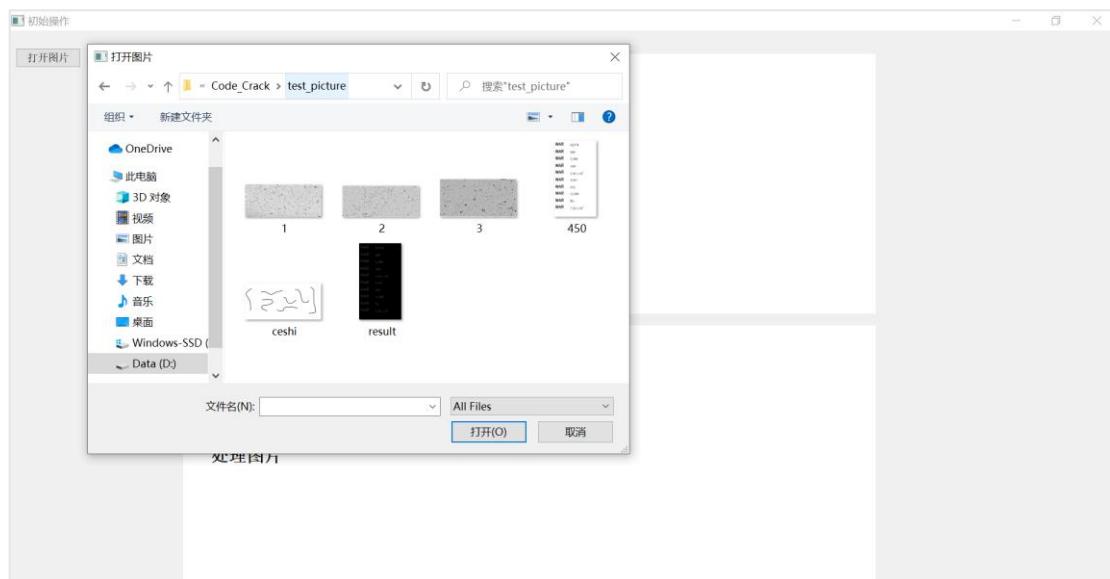


图 3.16 图像导入功能展示

接下来是对图像的增强和裂纹骨架的提取，具体操作被全部封装在软件中，

一旦选择好需要增强的热障涂层图像，处理后的结果将与原始图像一起显示在对应的图像显示框中，如图 3.17 所示。结果图像去除了原始热障涂层图像中的孔隙和噪声，并完成了裂纹修复及裂纹细化。

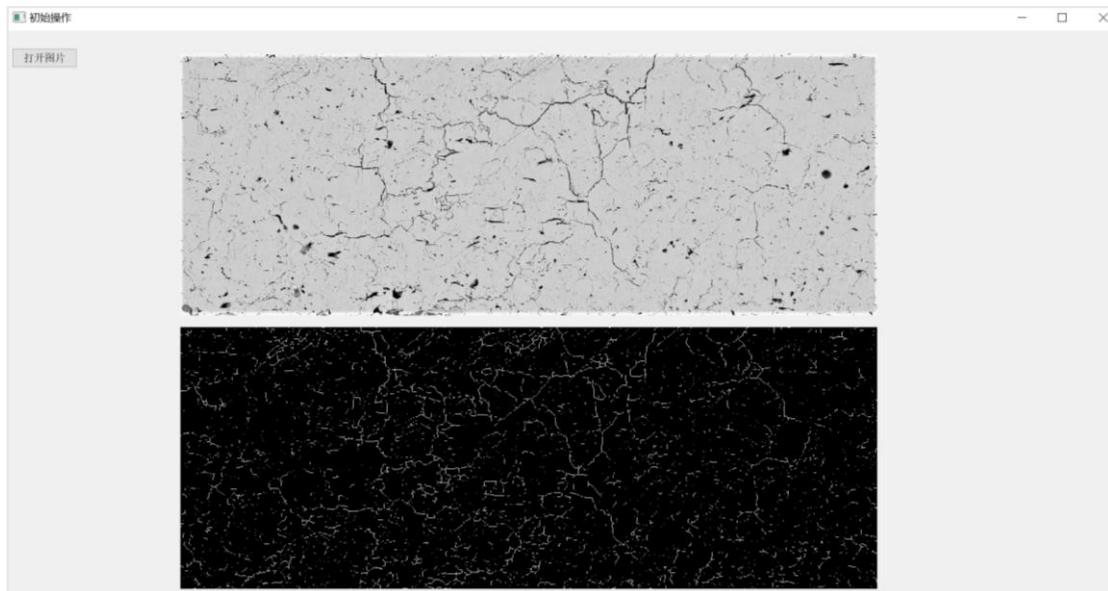


图 3.17 原始热障涂层图像和处理后图像的预览

最后，增强后的图像被保存至原始图像所在的路径下，并命名为“result”，且结果图的图像格式与原图像格式一致。由于图像分辨率过高，且受软件窗口大小的限制，不能十分精确地展示处理结果，只需在文件夹中打开保存后的结果图像即可正常预览。

此外，用户直接运行软件程序必须要在本机安装 Python 解释器及第三方库，这对非计算机专业人员来说是不被接受的。因此本程序会打包为.exe 可执行文件，直接供材料科研工作者使用，打包好的软件在 Windows 或 Linux 系统中均可以成功运行，省去环境配置的麻烦，实现了用户友好的目的。

3.4.2 裂纹识别与长度计算软件的设计与实现

裂纹识别与长度计算软件的整体流程如图 3.18 所示，也包括四个功能模块：

(1) 图像导入。该模块负责导入要处理的图像，用户从本地文件夹中选择任意一张使用 3.4.1 节中的软件处理后的图像，被选择的图像展示在图像显示框中。导入功能通过点击“OpenImage”按钮实现，导入图像所在的目录也被显示

在“文件路径”文本框中。

(2) 图像预览。该模块主要负责展示图像，预览功能包括导入图像的展示和裂纹识别结果图的展示，两种图像展示在同一个显示框里。

(3) 裂纹识别与长度计算。该模块按照 3.2 节中的裂纹识别方法完成对裂纹的识别标注，并计算每条裂纹的长度。该功能通过点击“calculate length”按钮实现，经计算得出的裂纹长度显示在“calculate length”按钮下方的文本框中。

(4) 结果保存。裂纹识别后的结果图像将被保存至导入图像所在的路径下，结果图像被命名为“Result”，结果图像的格式也与导入的原始图像格式一致。

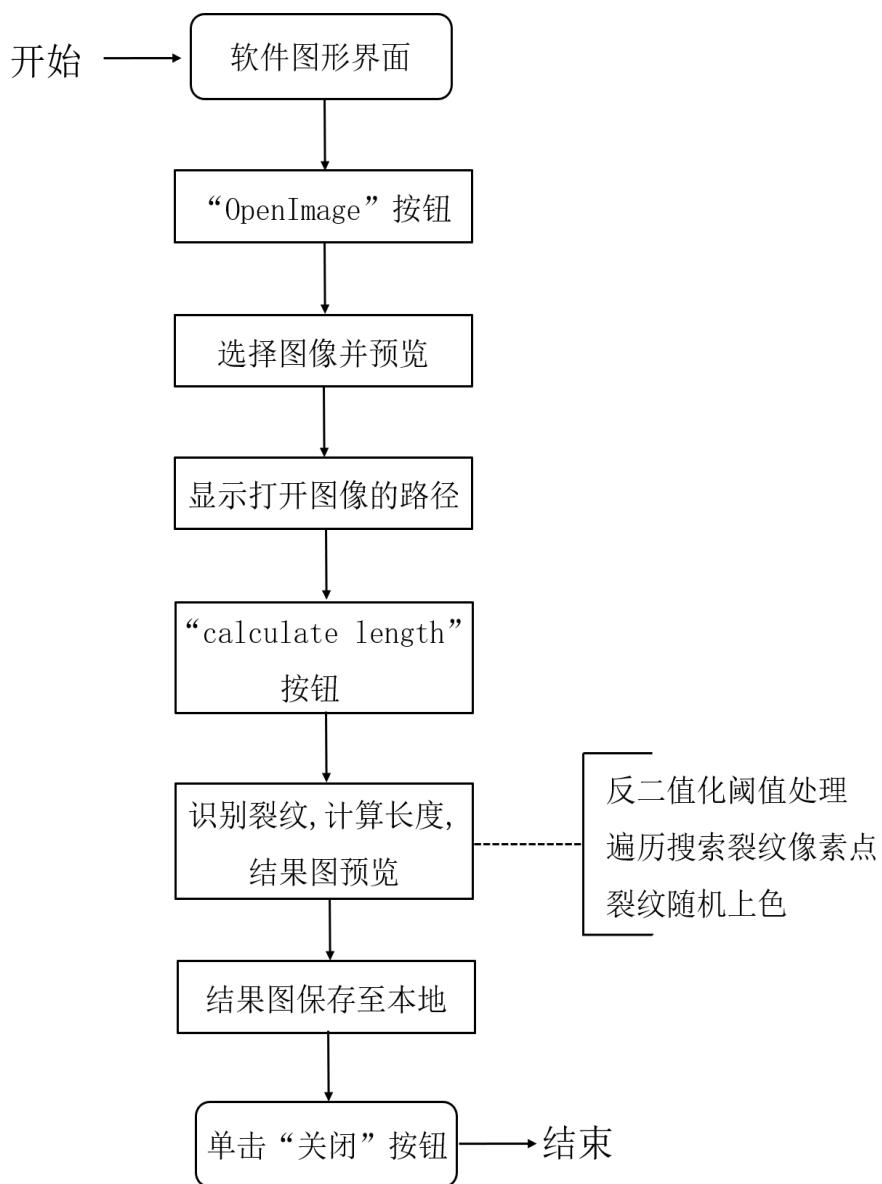


图 3.18 裂纹识别与长度计算软件的整体流程

设计裂纹识别与长度计算软件使用的开发语言是 C++，开发工具是 QT Creator，软件界面的设计工具为 Qt 框架。使用 C++语言开发该软件的原因在于 C++的运行速度非常快，且本软件涉及的算法需要遍历搜索大量的裂纹像素点。相比于 Python 这种解释型语言，C++这种编译型语言执行算法的效率更高。而选用 Qt 设计软件界面的原因在于，Qt 具有跨平台特性。Windows 下基于 C++的 GUI 库有很多，如 Qt、MFC、wxWidgets 和 DirectUI；而 Linux 下基于 C++的 GUI 库较少，常用的有 Qt、GTK+和 wxWidgets。Qt 不像其他 GUI 库，只能应用于单一平台，Qt 作为一种跨平台的 GUI 库，能够一次编写，到处运行。由于本软件需要运行在多个平台下，故选用可以跨平台的 Qt 作为开发框架。并且 Qt 还有很多其他优势，如：封装性好，几行代码就可以开发出一个简单的客户端；容易做出漂亮的界面和炫酷的动画，而 MFC、WTL、wxWidgets 比较麻烦；独立安装，Qt 程序最终会编译为本地代码，无需其他库的支撑。

以下是软件的具体实现。图 3.19 展示了软件界面，界面主要包含“OpenImage”和“calculate length”两个按钮、一个图像显示框和两个文本框，一个文本框用来显示导入图像的路径，另一个文本框用于输出裂纹的长度。软件的主要功能有：导入图像、图像预览、裂纹识别标注、裂纹长度计算、结果图预览、裂纹长度输出、图像文件路径获取及展示。

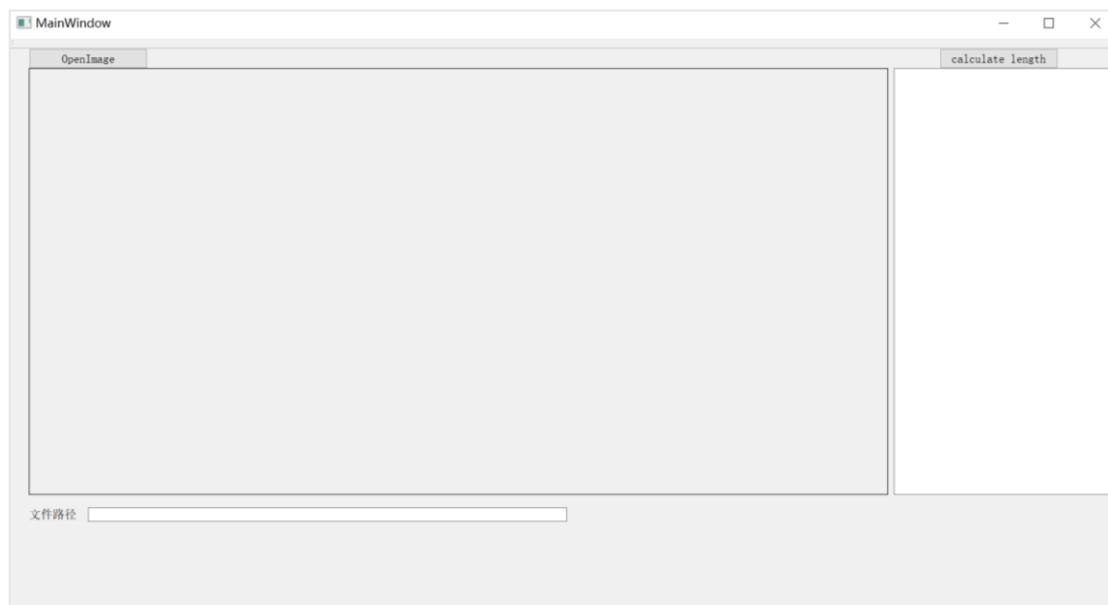


图 3.19 软件界面展示图

点击“OpenImage”按钮即可弹出文件选择窗，选择并打开使用3.4.1节中的软件处理并保存的结果图像，即可实现图像导入。导入后的图像可在图像显示窗口预览，图3.20和图3.21分别展示了图像导入功能和图像预览功能。此外，图像显示框下方的文本框能够展示导入图像的路径。

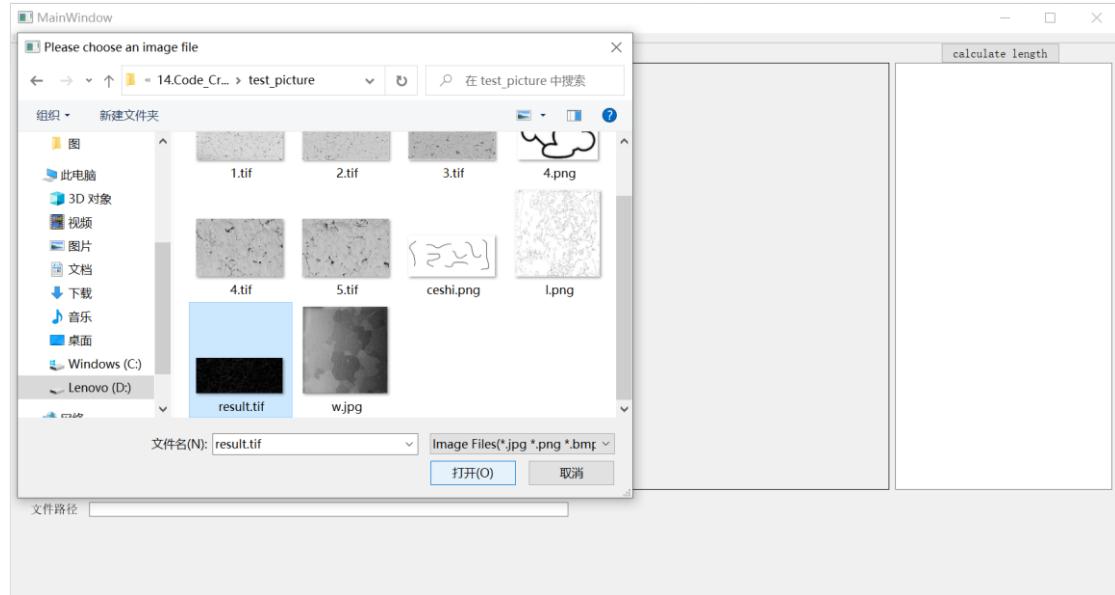


图3.20 图像导入功能展示

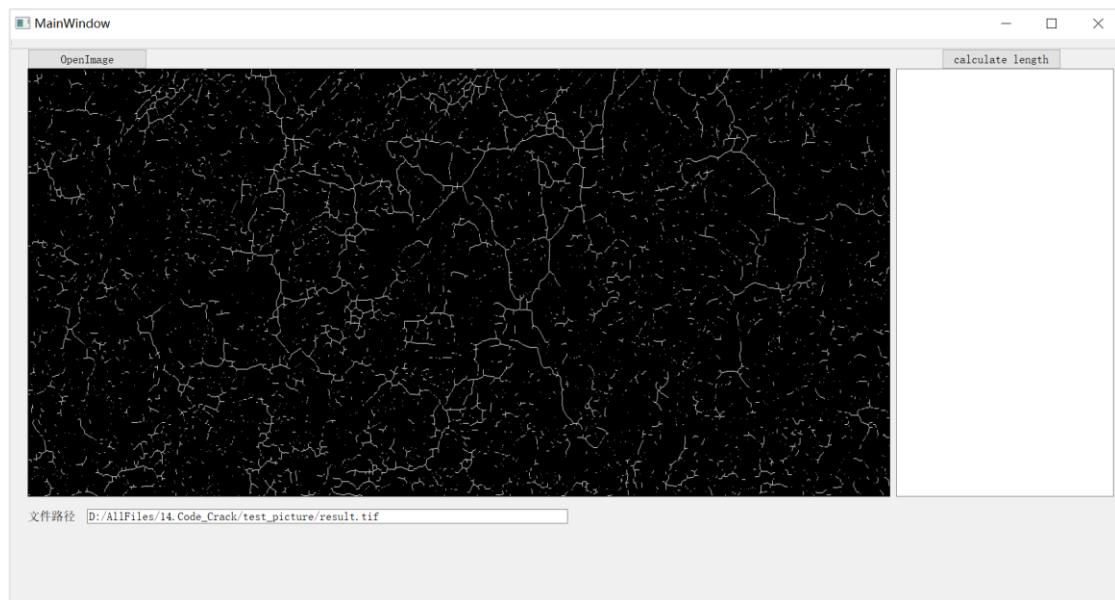


图3.21 图像预览功能展示

接下来是对裂纹的识别标注及长度计算，只需点击“calculate length”按钮即可实现该功能。裂纹识别完毕后会被随机上色，每条裂纹都会被标注序号，同时，

每个裂纹的长度也会显示在右侧的文本框中，图 3.22 展示了软件的处理结果。

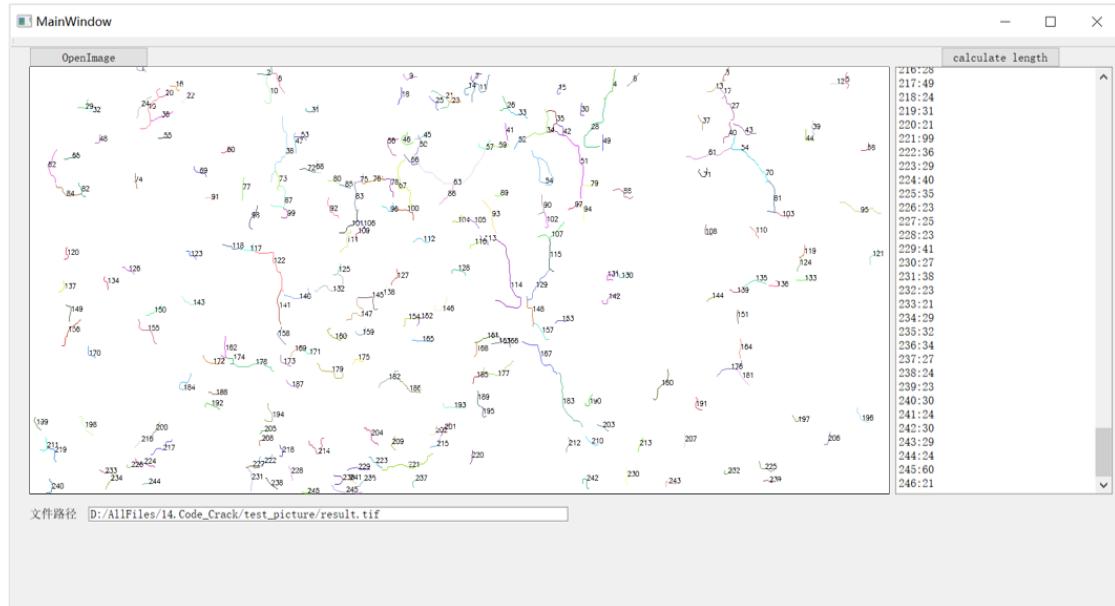


图 3.22 结果展示

最后，将裂纹识别的结果图像保存至导入图像所在的路径，命名为“Result”，其格式与导入图像的格式一致。同时，该软件也会被打包为.exe 可执行文件，直接供用户使用。

3.5 本章小结

针对小样本热障涂层形貌特征识别任务，本章提出了基于图像增强方法的裂纹识别框架。通过对原始图像进行图像增强，去除了无关孔隙和干扰噪声，进而使用断线修复方法增强了裂纹特征，弥补了缺失的像素点，最后基于骨架提取算法和遍历搜索算法识别出热障涂层中的裂纹，并计算其长度。在 5 张热障涂层图像上的实验结果显示，本章所提方法在裂纹识别任务上取得了较好的结果，增强方法的使用也大大降低了裂纹识别的难度。

此外，针对裂纹识别问题，本章还开发了两款软件，一款用于图像增强与裂纹细化，另一款用于裂纹识别与长度计算。对这两款软件的开发环境、整体架构、功能模块及软件实现，本章都做出了说明。开发出的软件为材料科研工作者提供了一个可视化的操作工具，并且软件的识别结果误差小、速度快，极大节省了热障涂层研究中的时间成本和人力成本，提高了材料微观结构分析的效率。

第四章 基于单样本生成模型的材料图像数据增强

上一章取得的成果得益于对热障涂层图像进行了大量的增强，此类增强方法均是对原始图像的组合变换。只有当图像分布不具备视觉手性时，才能在不改变原始图像分布或标签的前提下，使用翻转等传统图像增强操作来增强数据集，所以传统的图像增强技术具有较大的应用局限性。而多样本生成模型需要足够的数据来训练网络，对样本量极少的材料图像而言，使用多样本生成模型进行数据增强也不现实。因此，本章提出基于单样本图像生成的数据增强方法，通过在 HP-VAE-GAN 模型中加入 CBAM 注意力机制，提高模型的生成效果，使用生成图像扩充训练数据集后，提升分类模型的性能，缓解材料图像分类中的小样本问题。

4.1 材料图像数据增强概述

为保证在样本缺少的场景下深度学习模型依然能够精准识别材料微观图像，本章提出结合卷积注意力模块（Convolutional Block Attention Module，CBAM）的 HP-VAE-GAN 生成材料微观图像，并用生成图像进行数据增强，再进行图像分类，整体流程如图 4.1 所示。

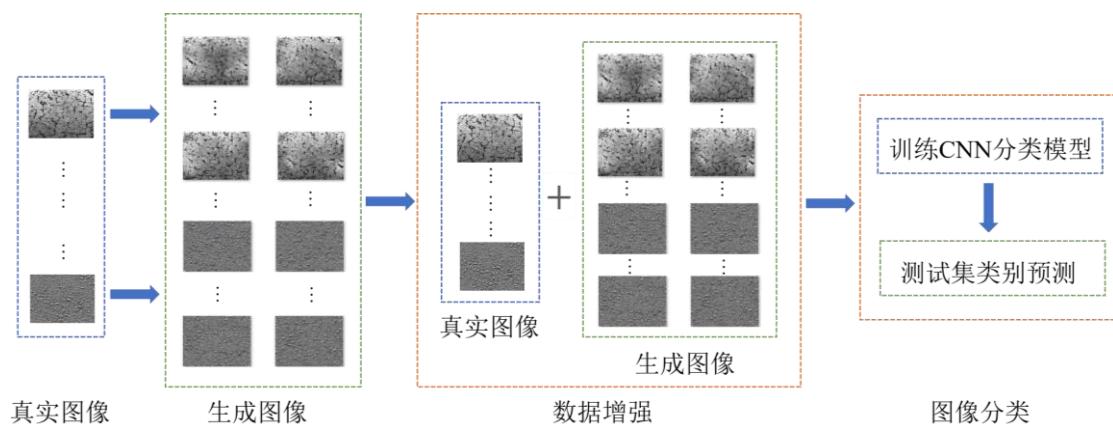


图 4.1 数据增强辅助材料图像分类的整体流程图

不同于常见的多样本图像生成算法，改进后的 HP-VAE-GAN 和原始的 HP-VAE-GAN 一样，都是对单张图像进行训练并生成，因此不需要大量的训练样本。改进后的 HP-VAE-GAN 在编码器中加入 CBAM，特征图获取到通道维度和空间

维度上的注意力权重，从而加深各个特征在通道和空间上的联系，提高网络的特征表达能力，使得到的特征图更加精细，进而生成质量更高的图像。生成的数据作为真实数据的补充，增加训练集的样本数，以训练出更好的 CNN 模型来提高材料微观图像分类的准确率。

4.2 融合 CBAM 的 HP-VAE-GAN

4.2.1 模型整体架构

图 4.2 展示了改进后的 HP-VAE-GAN，该模型共包含 $N + 1$ 个生成器，按照生成结果从粗糙到精细分别记为 $G^0 \dots G^N$ ，模型学习从 0 到 N 不同尺度下图像块的分布，并从粗糙到细致、从低分辨率到高分辨率逐步生成接近真实样本的图像。对每个尺度 $n = 0, \dots, N$ ，训练样本 x 都进行了下采样，下采样得到的图像记为 x^n 。与原模型不同的是，本文在 PatchVAE[110] 的编码器中插入了 CBAM 和 1 个卷积块，修改后的编码器为 CE。在编码器中加入 CBAM 是为了使网络更加关注图像中的重点信息，进一步提取到图像的有效特征，而增加的 1 个卷积块是为了使 CBAM 插入位置位于所有卷积块中间，避免位置因素影响模型效果，改进后的编码器结构如图 4.4 所示。

在 PatchVAE 中，输入的单张样本 x^0 被裁剪成大小为 $r * r$ 的图像块，经过包含 CBAM 的编码器 CE 后得到尺寸为 $H \times W \times C$ 的特征图 $CE(x^0)$ ， H, W, C 分别代表特征图的高、宽及通道数，图像块 $\rho_j (j = 1 \dots T)$ 的个数为 $T = H \times W$ 个。CBAM 的具体结构和基本原理如 2.3 节中的图 2.7 和图 2.8 所示。PatchVAE 与 VAE 最大的不同在于，对于那些对图像识别任务有用的特征表示，PatchVAE 只捕捉图像中的感兴趣区域，而不捕捉图像中的所有内容。从生成模型的角度出发，如果能轻松拟合出 T 个图像块的分布，那么直接从这个分布中采样，就能产生许多新的样本，然而这个分布很难直接拟合。于是模型从 VAE 的理论入手，希望从先验分布 $P(z_j) = N(0, I)$ 中采样隐变量 z_j ，根据 z_j 来重构 ρ_j 。

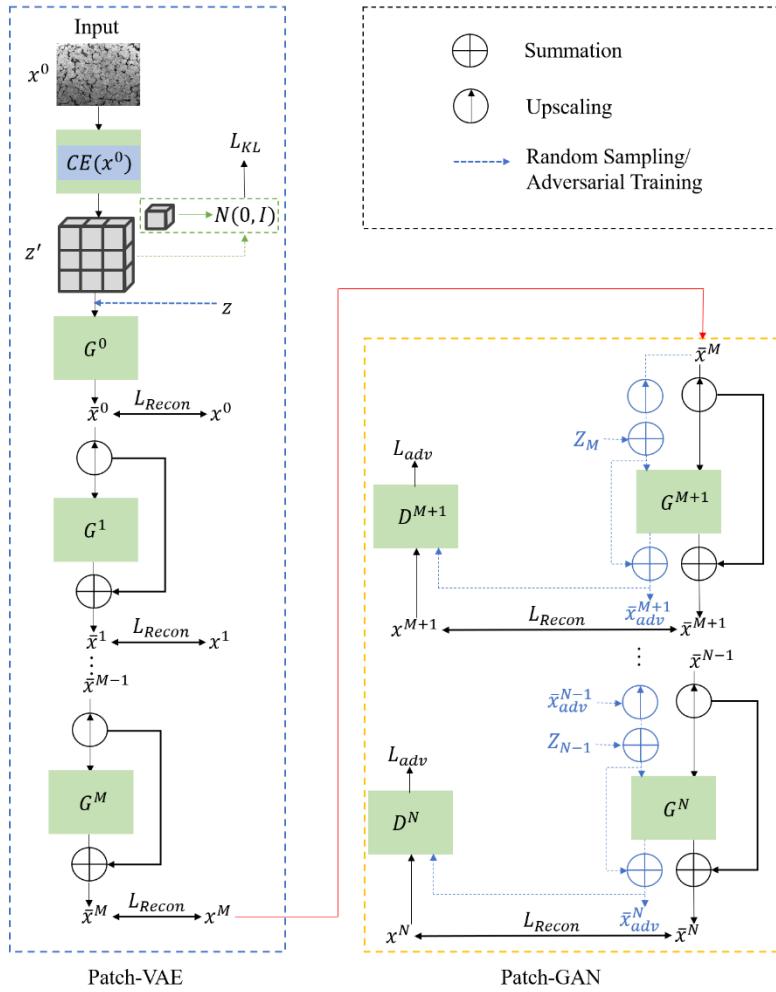


图 4.2 改进的 HP-VAE-GAN 整体结构，PatchVAE 中包含 CBAM 的编码器为 CE。

VAE 模型假设后验分布 $P(z_j|\rho_j)$ 是正态分布，经过神经网络的拟合，得出 $P(z_j|\rho_j)$ 的均值 μ_j 和方差 σ_j^2 。然后使用重参数技巧得到 $z'_j = \epsilon\sigma_j + \mu_j$, $\epsilon \sim N(0, I)$ ，使模型从 $N(\mu_j, \sigma_j^2)$ 中采样变成从 $N(0, I)$ 中采样。为了让 $N(\mu_j, \sigma_j^2)$ 向标准正态分布看齐，VAE 通过计算两者的 KL 损失来实现，KL 损失的计算公式如下所示：

$$L_{KL}(x) = \sum_{j=1}^T KL[N(\mu_j, \sigma_j^2) || N(0, I)] \quad (4.1)$$

图 4.2 中的 z' 正是经过重参数后采样得到的 z'_j 组成，且形状与特征图 $CE(x^0)$ 相同。 z' 经过解码器 G^0 后得到图像 \bar{x}^0 ，解码器 G^0 通过 \bar{x}^0 和 x^0 的重构损失进行更新。然后对 \bar{x}^0 上采样，将上采样后的图像送入生成器 G^1 中，此时将 G^1 输出的图像与上采样后的 \bar{x}^0 叠加，得到 \bar{x}^1 ，再次计算 \bar{x}^1 和 x^1 的重构损失来更新编码器 G^1 。以此类推，

直至经过编码器 G^M 得到 \bar{x}^M , 形式化表示如公式(4.2)所示, 其中 \bar{x}^{n-1} 是上一尺度的输出, $\uparrow \bar{x}^{n-1}$ 是 \bar{x}^{n-1} 向上采样到尺度 n 的结果。

$$\bar{x}^n = \uparrow \bar{x}^{n-1} + G^n(\uparrow \bar{x}^{n-1}) \quad (4.2)$$

在 M 尺度之前, PatchVAE 保证了生成图像的多样性, 使其产生具有高度多样性的样本, 且不易陷入模式崩溃。

从 G^{M+1} 开始, 使用 PatchGAN[111][112] 训练生成器和判别器, 由于感受野变小, 因此生成器引入了更加精细的纹理细节, 保证了生成样本的高质量。PatchGAN 与原始 GAN 的不同之处在于, 原始 GAN 的判别器输出的值是对整张生成图像的评价, 而 PatchGAN 中判别器的输出不再是一个用以评价真假的标量值, 而是一个 $N * N$ 的矩阵 X_{ij} , 其中 X_{ij} 表示的是图像块 $Patch_{ij}$ 的真假性, 即 X_{ij} 对应了判别器对输入图像每个 patch 的判别输出。PatchGAN 的网络结构设计成了全卷积形式, 因此也被称为全卷积对抗生成网络 (Fully Convolutional GAN)。图像经过各个卷积层后, 不再被送入激活函数或全连接层, 而是再使用卷积操作将输入映射为 $N * N$ 的矩阵 X_{ij} , 矩阵 X_{ij} 就是对整幅图的评价。图 4.3 通过一副马氏体不锈钢的微观图像直观展示了 PatchGAN 与原始 GAN 的区别。

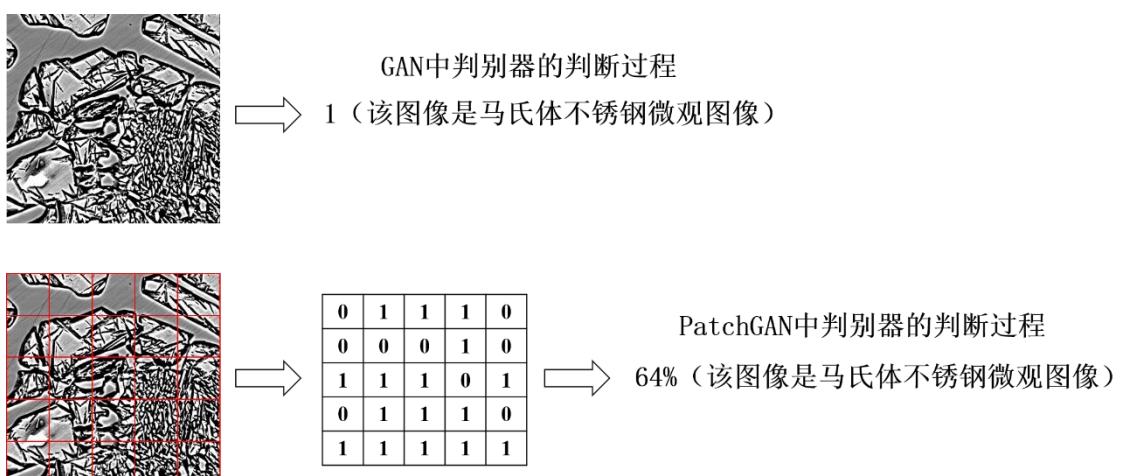


图 4.3 PatchGAN 与原始 GAN 中判别器的区别

PatchGAN 秉承对图像 patch 进行真假判断这一理念, 为此, 一种直观的处理方法是将图像裁剪成多个重叠的 patch, 然后对每个 patch 使用判别器判断, 最后平均得到的结果, 但是该方法的运算消耗极大。而全卷积网络利用卷积的感受野特性恰好能够完成对图像 patch 的真假判断, 这也是对卷积神经网络的巧妙利用。

卷积神经网络在 PatchGAN 中的优势在于，它能以同一方式独立处理每个图像 patch，因此在实际运算中，每个神经元都等同于在执行对每个 patch 的单独判断，这也避免了上述那种直观处理方法的低效性。所以，在模型中使用全卷积网络具备十分高效的运算效率，并且评价矩阵的使用也让模型更加关注图像的细节，这也使图像的局部特征与整体特征更好地融合，增强了图像生成效果。

图 4.2 中的 PatchGAN 在尺度 $n > M$ 时开始发挥作用，首先对与 $CE(x^0)$ 形状相同的 $z \in N(0, I)$ 进行采样，然后按照公式(4.2)得到 \bar{x}^M ， \bar{x}^M 上采样后被送入生成器 G^{M+1} ， G^{M+1} 输出的图像与上采样后的 \bar{x}^M 叠加，得到 \bar{x}^{M+1} ；同时输入噪声 z_M ，与上采样后的 \bar{x}^M 叠加后送入 G^{M+1} ，然后把 G^{M+1} 输出的图像与送入 G^{M+1} 的图像再次叠加，得到样本 \bar{x}_{adv}^{M+1} ，将该样本送入判别器 D^{M+1} 中。通过计算对抗损失 L_{adv} 和重构损失 L_{Recon} ，完成对 G^{M+1} 和 D^{M+1} 的更新。以此类推，直至经过生成器 G^N 得到拥有细节更丰富的图像，形式化表示如公式(4.3)所示：

$$\bar{x}_{rand}^n = \begin{cases} \uparrow \bar{x}_{rand}^{n-1} + G^n(\uparrow \bar{x}_{rand}^{n-1} + z_n) & n > M \\ \uparrow \bar{x}_{rand}^{n-1} + G^n(\uparrow \bar{x}_{rand}^{n-1}) & 0 < n \leq M \\ G^0(z') & n = 0 \end{cases} \quad (4.3)$$

其中 z_n 是与 $\uparrow \bar{x}_{rand}^{n-1}$ 维度相同的随机噪声， \bar{x}_{rand}^n 是在尺度 n 下随机生成的样本，该样本符合 $\bar{x}_{rand}^n = \gamma x^n + (1 - \gamma) \bar{x}_{rand}^n$ ，其中 γ 在 0 到 1 之间均匀采样， z' 如前文所述是从重参数后的分布中采样的随机噪声。

4.2.2 编码器和解码器

HP-VAE-GAN 的编码器共包含 3 个卷积块，每个卷积块均由 1 个 Conv2d 卷积层、BN 层及 Leaky_relu 激活函数组成。改进后的 Patch-VAE 中编码器的结构如图 4.4 所示，即在原始的编码器中增加了 1 个卷积块，并在 4 个卷积块的中间位置插入了 CBAM，CBAM 能够对卷积块提取的特征进一步加强。网络中使用的 function 和参数也如图 4.4 所示：Conv2d 是卷积运算，用于提取二维输入数据的不同特征，kernel size 设置为 3×3 ，stride 为 1；BN（Batch Normalization）层用以避免梯度消失现象；激活函数选用的是 $\alpha=0.2$ 的 Leaky Relu 函数。

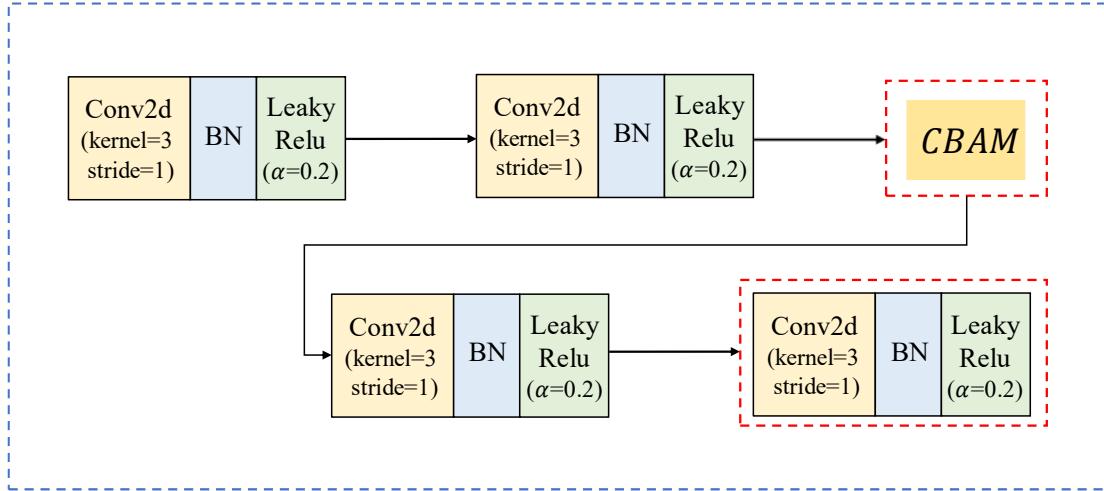


图 4.4 编码器结构，红色虚线框出来的是增加的 CBAM 模块和卷积块。

解码器部分仍沿用了 HP-VAE-GAN 中解码器的结构，共包含 7 个卷积块，各个卷积块使用的 function 和参数与编码器是一致的，其中最深和最浅的卷积块不使用 BN 层和 Leaky Relu 激活函数，具体的解码器结构如图 4.5 所示。需要说明的是，模型中 PatchVAE 的解码器就是构成 PatchGAN 的生成器。

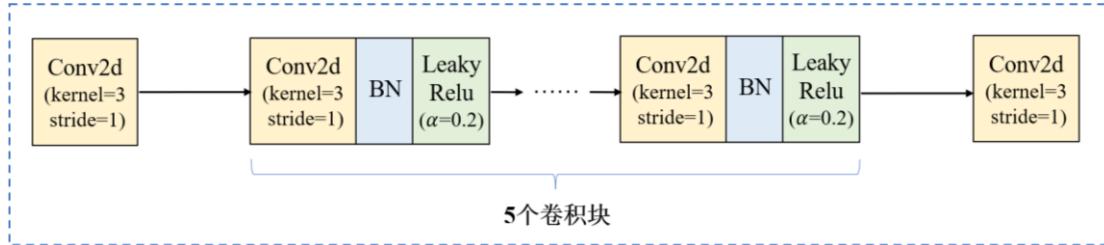


图 4.5 解码器结构

4.2.3 损失函数

损失函数沿用了原始 HP-VAE-GAN 的损失函数，在尺度 $0 < n \leq M$ 时，没有判别器的加入，此时编码器 CE 通过 x^n 的重构损失 $L_{Recon}(\bar{x}^n, x^n)$ 进行更新。 n 尺度下的损失函数如公式(4.4)所示，使用该损失仅用于更新编码器 CE、生成器 G^0 及 G^n 。损失函数中的 KL 项只和编码器 CE 有关，与 $\beta - VAE$ [113] 一样，KL 项损失也设置了平衡因子 β_{vae} 作为超参数。

$$L_{vae}(x^0, \bar{x}^n, x^n) = L_{Recon}(\bar{x}^n, x^n) + L_{Recon}(\bar{x}^0, x^0) + \beta_{vae} L_{KL}(x^0) \quad (4.4)$$

在引入判别器后，损失由对抗损失和重构损失两部分组成，其中重构损失仍然为 $L_{Recon}(\bar{x}^n, x^n)$ ，对抗损失选用了 WGAN-GP 的损失函数 $L_{adv}(z, x^n)$ ， $z \in$

$N(0, I)$, 该损失函数具体内容可参考文献[50]。因此在 $n > M$ 时总体的损失函数如公式(4.5)所示:

$$L_{adv}(z, \bar{x}^n, x^n) = L_{Recon}(\bar{x}^n, x^n) + \beta_{adv} L_{adv}(z, x^n) \quad (4.5)$$

此外, 在 Patch-GAN 的训练中, 对于 $n > M$, 只训练了 G^n 和 D^n , CE 和 $G^0, \dots G^{n-1}$ 被冻结。

4.3 小样本材料图像数据增强实验

4.3.1 数据集和实验设置

数据集 实验在超高碳钢显微照片数据库 UHCSDB (UltraHigh Carbon Steel Micrograph DataBase) [114]的子集上进行。该子集的具体构成为: 从碳化物网络 (Carbide Network)、珠光体 (Pearlite)、珠光体+球状体 (Pearlite + Spheroidite)、球状体 (Spheroidite) 以及马氏体 (Martensite) 这 5 个类别的超高碳素钢 SEM 显微照片中各随机选择 10 张图像, 共计 50 张图像, 构成实验用的子集。其中训练集与测试集的比例为 8:2, 训练集用于生成模型和分类模型的训练, 测试集用于评估分类任务的结果。除此之外, 还选取了 Kylberg 纹理数据集[115]中的 4 张图像、STex-512 纹理图像数据库[116][117]中的 4 张图像以及第三章中的 5 张热障涂层图像进行生成, 以观察模型效果, 并验证模型是否具有广泛性。

实验细节 为了公平的比较模型效果, 实验采用了与 HP-VAE-GAN 基本一致的训练设置, 即 $N = 9, M = 3, r = 11$ 。数据集中的所有图像在训练时均使用原始尺寸, 并对图像水平翻转以进行数据增强。在各数据集的实验上, 模型均训练 5000 个 iteration, batch_size 设为 2, 在每个尺度上都使用学习率为 5×10^{-4} 的 Adam (Adaptive Moment Estimation) 优化器[118]。模型中改进后的编码器共包含 4 个卷积块和 1 个 CBAM 模块, 编码器中共 7 个卷积块, 具体参数设置可以从图 4.4 和图 4.5 中得到。使用 HP-VAE-GAN 作为 Baseline, 将改进后的模型与其进行比较。

4.3.2 实验分析与讨论

定性分析

对于 UHCSDB 子集中的图像，对训练样本逐一训练，并使用训练好的模型生成图像，生成图像的数量可以手动设置。图 4.6 展示了使用 UHCSDB 中部分图像训练后的生成结果，第一行是用于训练的真实图像，第二行是随机挑选的几张使用改进后的 HP-VAE-GAN 生成的图像，第三行是使用原始 HP-VAE-GAN 生成的图像，红色框圈出了生成图像里的瑕疵。

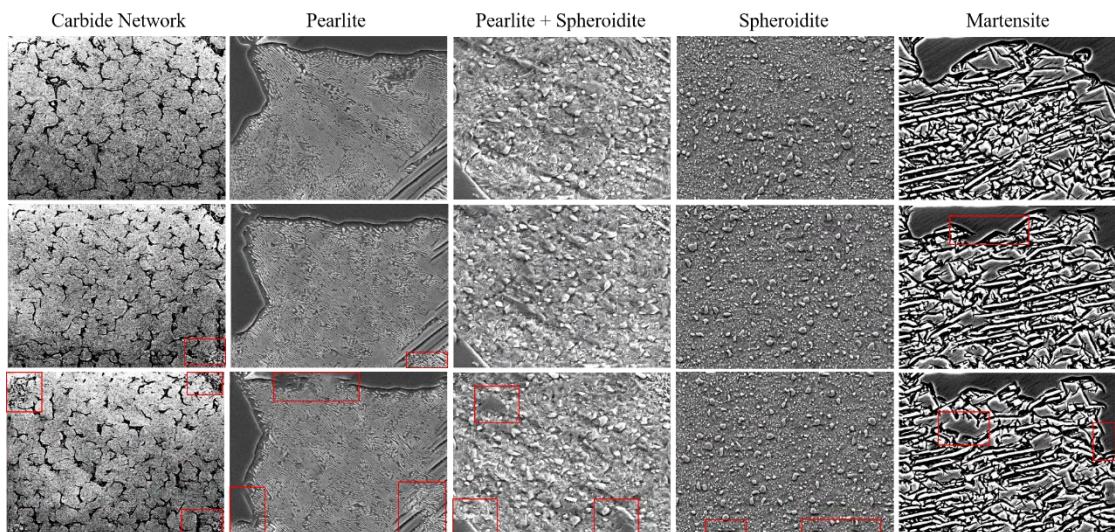


图 4.6 使用 UHCSDB 中部分图像训练后的生成结果。第一行是真实图像，第二行是使用改进后的 HP-VAE-GAN 生成的图像，第三行是使用原始 HP-VAE-GAN 生成的图像。

图 4.6 中的实验结果显示，改进后的 HP-VAE-GAN 生成的图像能够清晰地显示纹理细节，使用原始的 HP-VAE-GAN 生成的图像虽然纹理也比较清晰，但存在较多瑕疵，而改进后的模型生成图像中的瑕疵要少得多。图 4.7 展示了 3 张真实图像各自对应的 20 张生成图像，第 1 行第 1 列、第 4 行第 1 列以及第 7 行第 1 列位置上的均是真实图像（训练样本），其他位置上的是每个真实图像对应的 20 个生成样本。从图 4.7 中可以看出，生成图像并未完全仿照真实图像，每张生成图像各不相同，但风格都与真实图像一致。这也说明，使用改进的 HP-VAE-GAN 生成的图像在兼顾高质量的同时，也具备多样性。

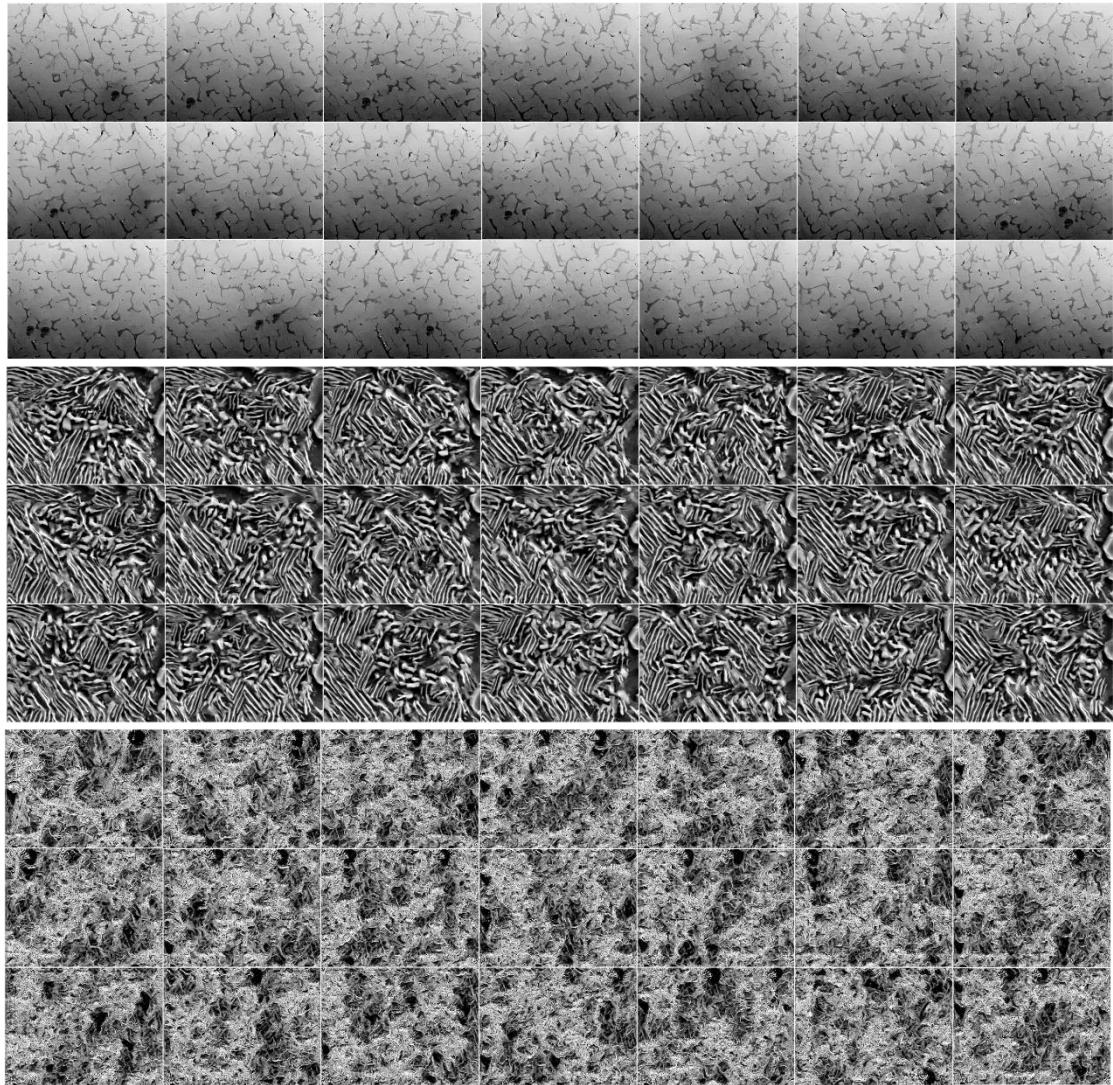


图 4.7 3 张真实图像及各自对应的 20 张生成图像。第 1 行第 1 列、第 4 行第 1 列以及第 7 行第 1 列位置上的均是真实图像，其他位置上的是每个真实图像对应的 20 个生成样本。

此外，还使用散点图对原始图像和生成图像的特征进行可视化，以观察生成样本的分布。具体做法是用图像灰度共生矩阵的 Correlation 和 Dissimilarity 特征来表征图像在特征空间中的分布，散点图的 x 轴和 y 轴分别代表了 Correlation 特征和 Dissimilarity 特征。图 4.8(a)表示的是 40 张真实图像的特征散点图，真实图像共 5 个类别，每类 8 张。图 4.8(b)、图 4.8(c)及图 4.8(d)分别代表训练 1 张真实图像对应生成 20 张、100 张及 500 张图像的特征散点图，也即图 4.8(b)、图 4.8(c)和图 4.8(d)各自的特征散点图中共包含了 800 个、4000 个、20000 个生成样本。从图中可以看出，生成样本数量较多时，偶尔会有少量样本发生偏移，但整体来看生成样本基本都分布于原始样本附近。

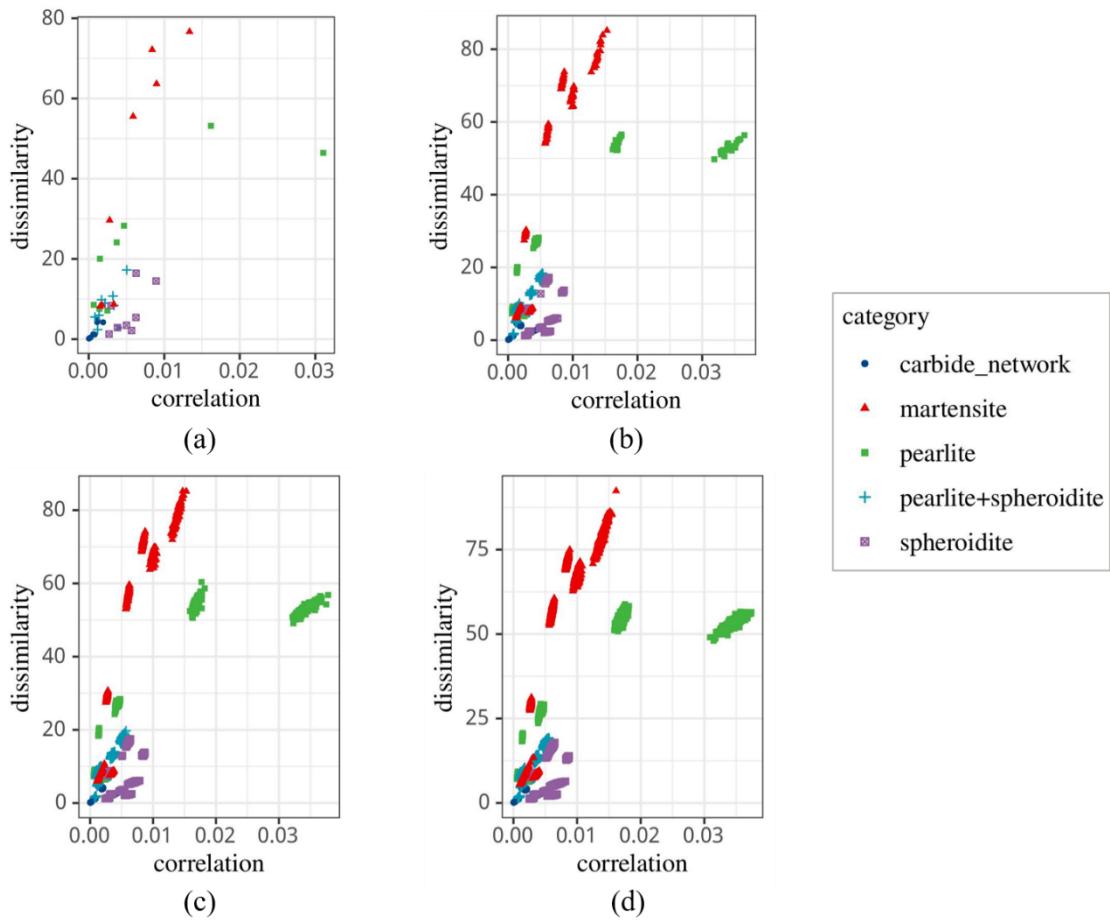


图 4.8 原始图像和生成图像的特征散点图。(a)包含 40 个真实样本; (b)包含 800 个生成样本; (c)包含 4000 个生成样本; (d)包含 20000 个生成样本。

为了进一步验证改进的 HP-VAE-GAN 这一生成模型的有效性，除了对钢铁微观图像进行生成，还从 Kylberg 材料纹理数据集和 STex-512 材料纹理图像数据库中分别选取了 4 个类别中的各 1 张图像进行实验。在 Kylberg 数据集中选取了 canvas、cushion、linseeds 及 stone 这 4 种类别中的各 1 张图像，在 STex-512 数据集里选择的是 bark、fabric、floor 及 gravel 这 4 种类别中的各 1 张图像。图 4.9 和图 4.10 展示了使用改进的 HP-VAE-GAN 进行生成实验的结果，第一行是训练样本，第二行是训练样本对应的 1 张生成图像。从图 4.9 和图 4.10 可以看出，生成图像并不是对真实图像的简单复制，在纹理细节上两者仍存在一定差异。除了图 4.10 中 fabric 对应的生成图像以外，其他生成图像的色彩与其对应的训练样本的色彩基本保持一致。

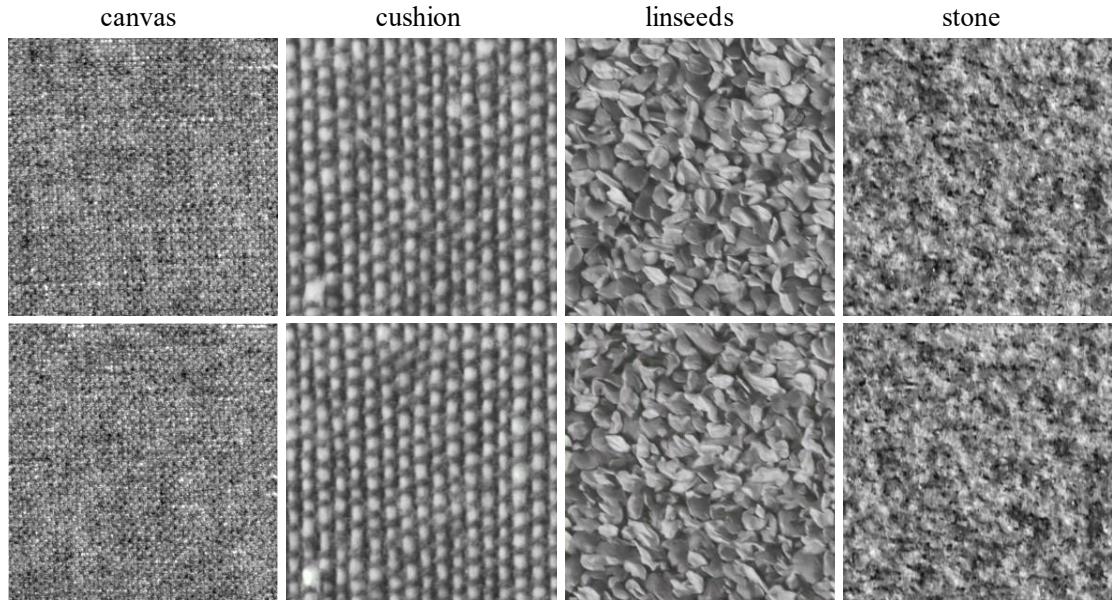


图 4.9 在 Kylberg 数据集上使用改进的 HP-VAE-GAN 进行生成实验的结果，第一行是训练样本（真实图像），第二行是随机挑选的训练样本对应的 1 张生成图像。

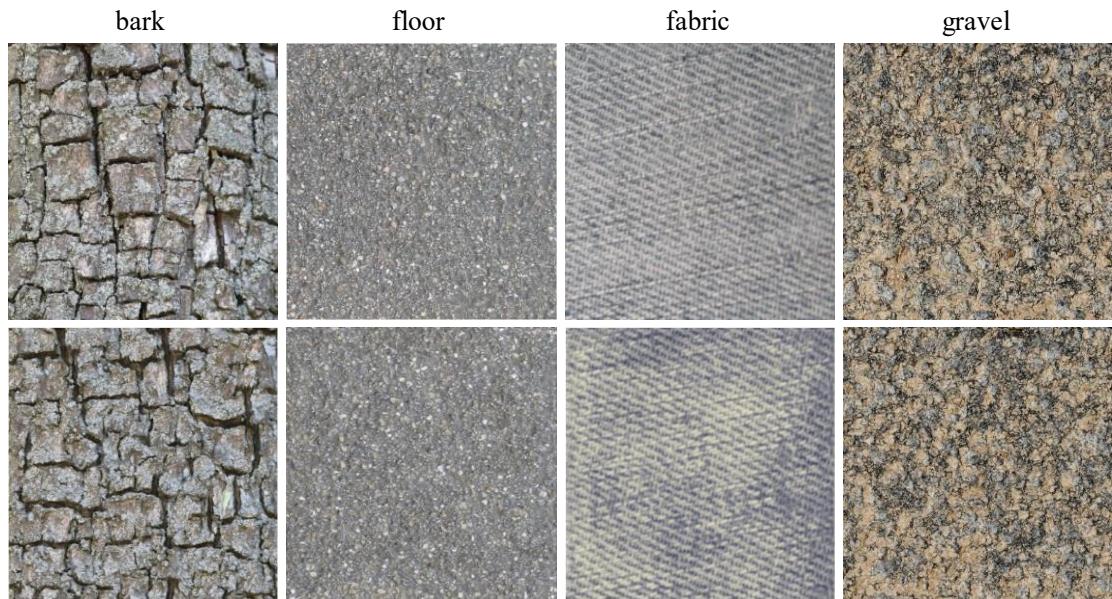


图 4.10 在 STex-512 数据集上使用改进的 HP-VAE-GAN 进行生成实验的结果，第一行是训练样本（真实图像），第二行是随机挑选的训练样本对应的 1 张生成图像。

图 4.11 展示了 Kylberg 数据集中 canvas 和 stone 这两张真实图像各自对应的 20 张生成图像，图 4.12 展示了 STex-512 数据集中 bark 和 gravel 这两张真实图像各自对应的 20 张生成图像。图 4.11 和图 4.12 中的第 1 行第 1 列、第 4 行第 1 列位置上的均是真实图像，其他位置上的是使用改进的 HP-VAE-GAN 生成的图像。综合图 4.11、图 4.12 展示的实验结果，可以看出生成图像已经具有“以假乱真”的效果。

真”的效果，改进的 HP-VAE-GAN 生成的图像不仅自然而接近真实，纹理细节也比较清晰，并且每张生成图像也各不相同。

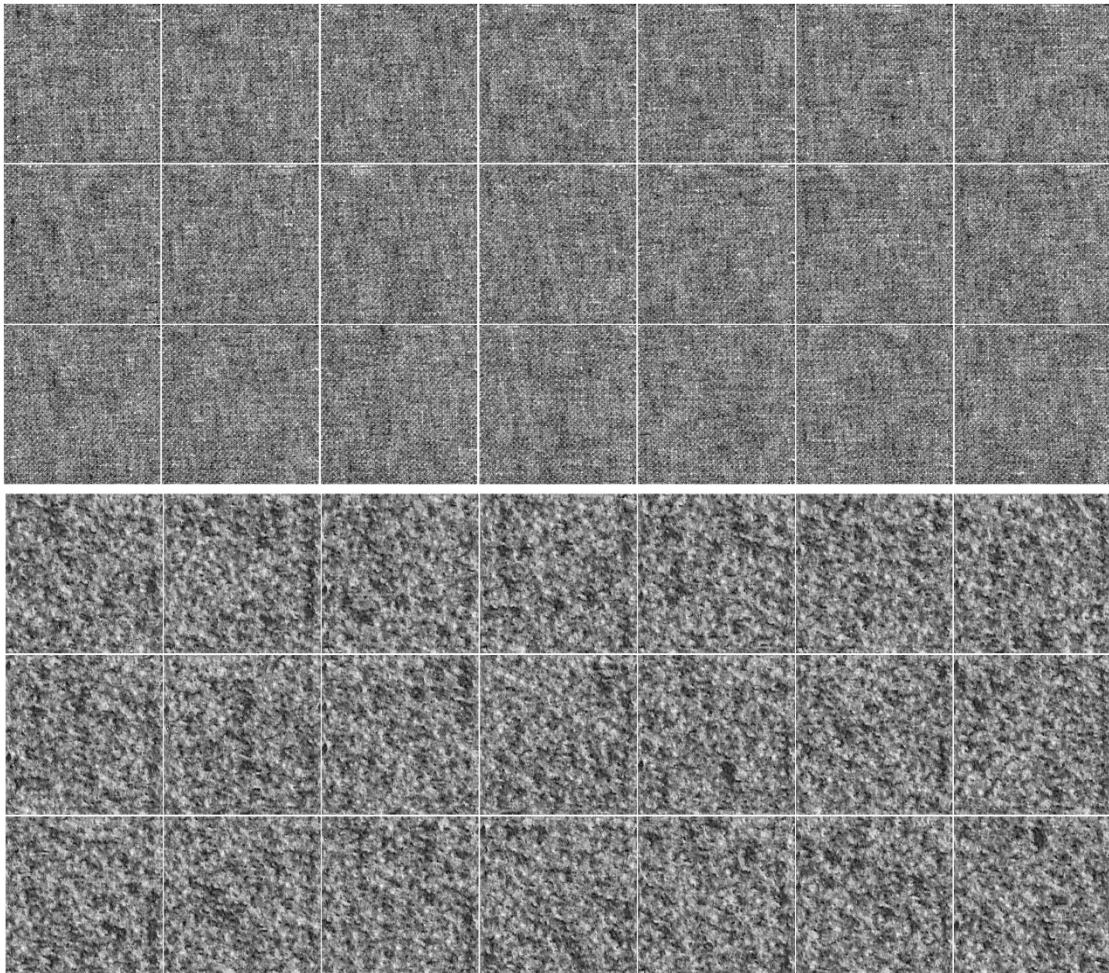


图 4.11 Kylberg 数据集中的两张真实图像以及各自对应的 20 张生成图像。第 1 行第 1 列和第 4 行第 1 列位置上的是真实图像，其他位置上的是真实图像对应的 20 个生成样本。

上述生成实验都是在公共数据集中的部分图像上进行的，对于第三章中的 5 张热障涂层图像，也使用改进的 HP-VAE-GAN 对其进行生成。图 4.13 展示了使用改进的 HP-VAE-GAN 对热障涂层图像(a)和热障涂层图像(b)的生成结果，图 4.14 展示了使用改进的 HP-VAE-GAN 对热障涂层图像(c)、热障涂层图像(d)和热障涂层图像(e)的生成结果。图 4.13 和图 4.14 的第一行均是真实图像，第二行是真实图像对应的生成图像。从图 4.13 和图 4.14 中可以看出，生成的热障涂层图像也具有和真实图像类似的众多细微孔隙和细小裂纹。尽管热障涂层图像具有十分繁杂的微观形貌，但改进的 HP-VAE-GAN 仍具有比较优异的生成表现，即生成图像的纹理较为清晰，并且生成图像的整体风格与真实图像十分接近。



图 4.12 Kylberg 数据集中的两张真实图像以及各自对应的 20 张生成图像。第 1 行第 1 列和第 4 行第 1 列位置上的是真实图像，其他位置上的是真实图像对应的 20 个生成样本。

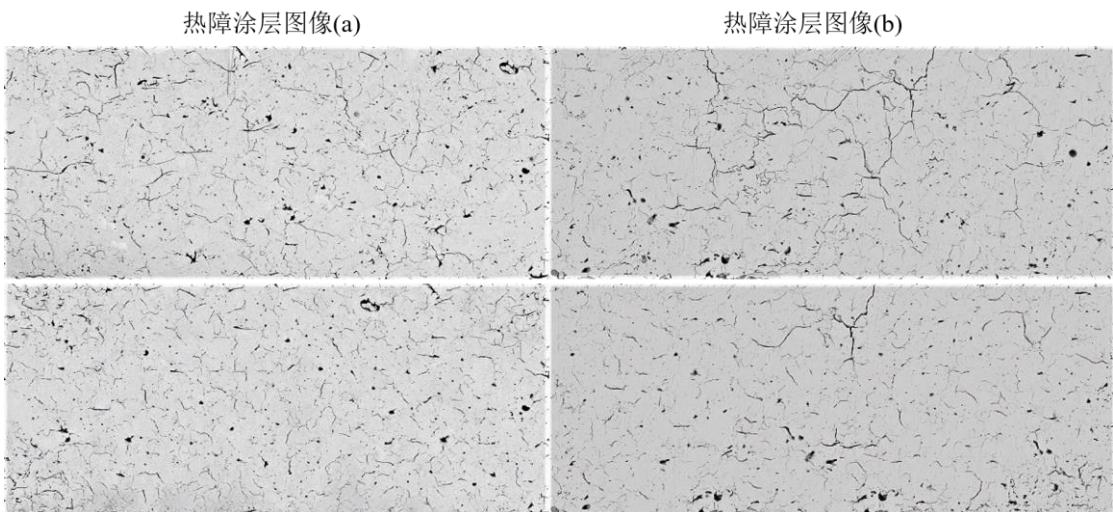


图 4.13 在热障涂层图像(a)和热障涂层图像(b)上使用改进的 HP-VAE-GAN 进行生成实验的结果。第一行是真实图像，第二行是随机挑选的每张真实图像对应的 1 张生成图像。

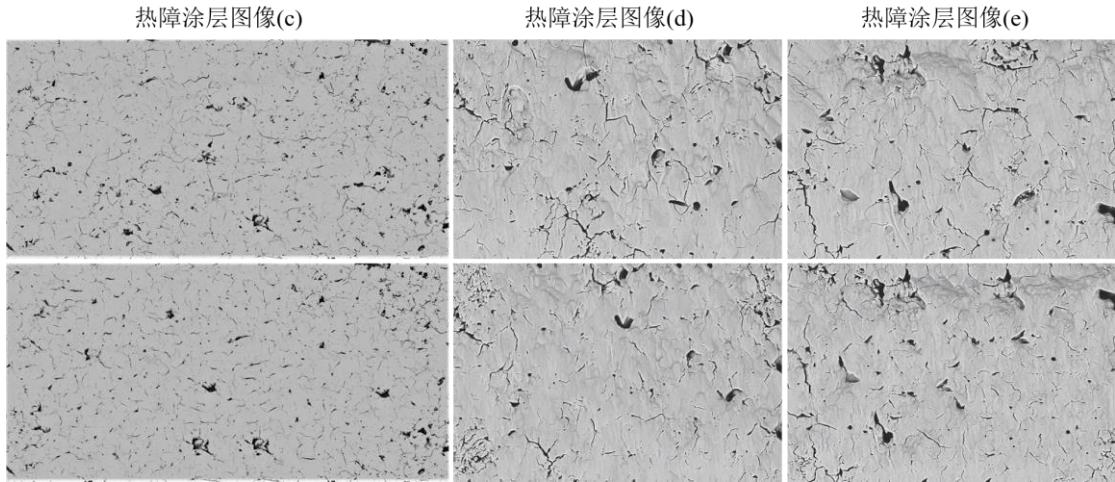


图 4.14 在热障涂层图像(c)至热障涂层图像(e)上使用改进的 HP-VAE-GAN 进行生成实验的结果。第一行是真实图像，第二行是随机挑选的每张真实图像对应的 1 张生成图像。

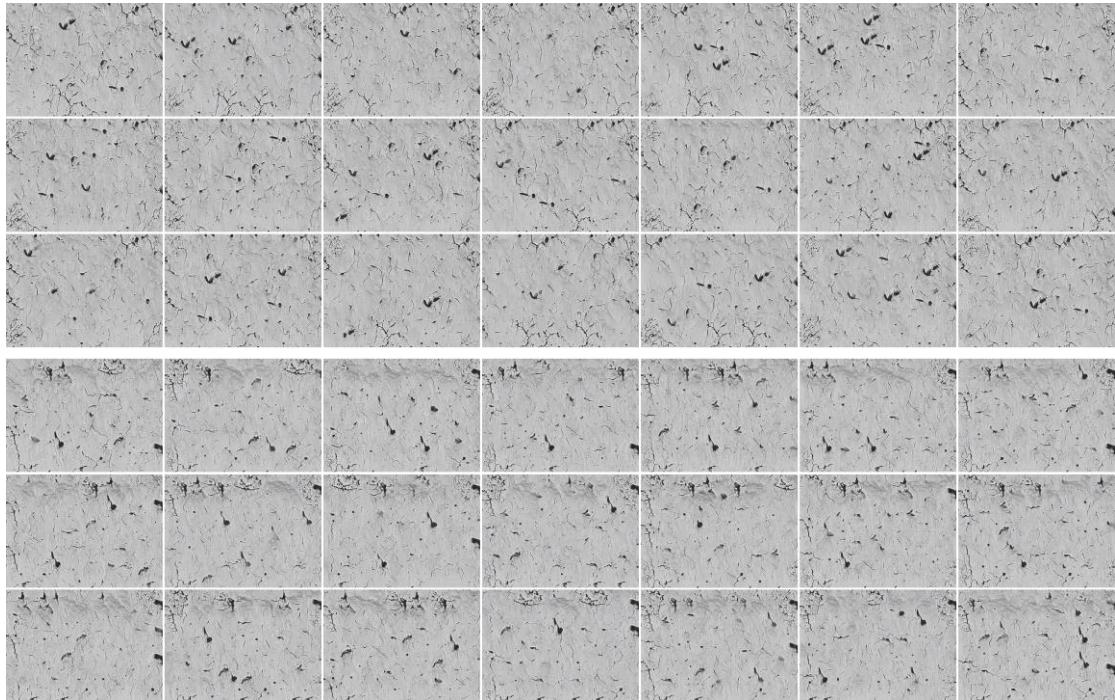


图 4.15 热障涂层图像(d)和热障涂层图像(e)以及各自对应的 20 张生成图像。第 1 行第 1 列和第 4 行第 1 列位置上的是真实图像，其他位置上的是真实图像对应的 20 个生成样本。

图 4.15 展示了热障涂层图像(d)和热障涂层图像(e)这两张真实图像各自对应的 20 张生成图像，第 1 行第 1 列位置上的是真实的热障涂层图像(d)，第 4 行第 1 列位置上的是真实的热障涂层图像(e)，其他位置上的是这两张真实图像对应的生成图像。从图 4.15 中可以看出，生成的热障涂层图像在保留真实性的同时存在较强的多样性，生成图像偶尔也存在纹理重复的现象，但并不影响整体的生成效果。以上实验结果也验证了改进后的 HP-VAE-GAN 在使用场景上的广泛性。

定量分析

对生成图像，使用几个常用的评价指标如结构相似性 SSIM (Structural Similarity) [119]、余弦相似度[120]、KL 散度 (Kullback-Leibler Divergence) [121] 及 JS 散度 (Jensen-Shannon Divergence) [122]对生成结果进行定量分析。其中 SSIM 是评价两幅图像相似性的一个重要指标，SSIM 从图像构成的角度出发，反映了图像中纹理结构的属性，取值范围为 $[-1, 1]$ ，值越大表示图像间结构相似度越高。而余弦相似度是先把图像表示成一个向量，通过计算向量间的余弦距离来表征两张图片的相似度，值越接近 1，表明图像间越相似。KL 散度用于衡量两个概率分布之间的差异性，二者越相似，KL 散度越小。而 JS 散度是 KL 散度的一种变体，两个概率分布之间越相似，JS 散度越小。KL 散度和 JS 散度在图像生成领域是十分常用的评价指标。

表 4.1 展示了使用改进后的 HP-VAE-GAN 在 UHCSDB 子集的训练数据上进行图像生成后的评分结果。几种指标的评分结果表明，生成图像的分布与真实图像的分布十分接近。

表 4.1 生成的钢铁微观图像的评分结果

| | 结构 相似性 | 余弦 相似度 | KL 散度 | JS 散度 |
|----------------------|-----------|-----------|----------|----------|
| carbide network | 0.999899 | 0.954812 | 0.068004 | 0.01548 |
| pearlite | 0.999866 | 0.914536 | 0.14235 | 0.029988 |
| pearlite+spheroidite | 0.999823 | 0.89313 | 0.160105 | 0.035283 |
| spheroidite | 0.99981 | 0.963096 | 0.041842 | 0.010241 |
| martensite | 0.999875 | 0.861535 | 0.235227 | 0.049621 |

除了计算上述评价指标，还进行了分类实验，以验证使用生成图像扩充训练样本是否能提升分类准确率。数据集仍是 UHCSDB 的子集，在生成实验和分类实验中，数据集的划分策略如前文 4.3.1 节中所述，即 40 张图像作为训练集，10 张图像作为测试集。在生成实验中，40 张图像依次训练，每张真实图像对应生成 20 张与真实图像尺寸相同的样本，所以总的生成样本数目共 800 张。在分类实验中，40 张图像作为训练集，10 张图像作为测试集。由于钢铁微观图像尺寸太

大(645×484)，故分类实验中把真实图像从中间部分裁剪成 448×448 大小，再裁剪成 4 个 224×224 的 patch，生成样本也会进行同样的处理。从而得到，未扩充前的训练集有 $160(40 \times 4)$ 个样本，记为 Train_A；扩充后的训练集在 160 个训练样本的基础上增加了 $3200(800 \times 4)$ 个生成样本，共计 3360 个样本，记为 Train_B；测试集为 Test_C，有 $40(10 \times 4)$ 个样本。为了避免结果的偶然性，图像分类任务中选取了三种常见的 CNN 分类模型，分别是 MobileNet[123]、ResNet50[124]和 VGG16。表 4.2 的结果显示，使用生成图像进行训练数据扩充后，测试集上的 Top-1 accuracy 都有所提升，其中使用 VGG16 作为分类网络的 Top-1 accuracy 提升了 55%。这在一定程度上缓解了训练样本过少导致的分类任务中 CNN 模型过拟合及目标任务欠拟合的问题，同时证明了使用改进的 HP-VAE-GAN 对材料图像进行数据增强这一方法是可行的。

表 4.2 不同分类模型（MobileNet、ResNet50、VGG16）在测试集上的分类结果

| 训练集 | 测试集 | Top-1 Accuracy | | |
|---------|--------|----------------|------------|------------|
| | | MobileNet | ResNet50 | VGG16 |
| Train_A | Test_C | 75% | 72.5% | 40% |
| Train_B | Test_C | 82.5% | 90% | 95% |

为了消除由于测试集数量过少导致分类结果的偶发性，在上述实验的基础上，又扩充了测试集的数量，测试集的图像数量由 10 张增加至 110 张。同样，需要把每张测试图像都裁剪成 4 个 224×224 的 patch，因此新的测试集有 $440(110 \times 4)$ 个样本，新的测试集记为 Test_D。表 4.3 展示了更改测试集后的分类结果，显然，对训练数据扩充后的 Top-1 准确率也都得到了提升，但相比于在测试集 Test_C 上的结果，准确率都有所下降，这是由测试集数量增多造成的。尽管如此，与没有增强的训练集相比，增强训练集后仍提高了分类模型的性能。

表 4.3 更改测试集后不同模型的分类结果

| 训练集 | 测试集 | Top-1 Accuracy | | |
|---------|--------|----------------|------------|---------------|
| | | MobileNet | ResNet50 | VGG16 |
| Train_A | Test_D | 73.41% | 67.05% | 33.41% |
| Train_B | Test_D | 81.36% | 80% | 77.95% |

4.3.3 消融实验

本文主要在 HP-VAE-GAN 中 Patch-VAE 的编码器中加入了 CBAM 和 1 个卷积块，为了验证 CBAM 及卷积块对网络模型的影响，还做了消融实验。数据集仍选用 UHCSDB 的子集，在生成实验和分类实验中，数据集的划分策略也如前文所述，本次实验中 1 张训练图像生成的样本数目仍是 20，分类实验的训练集仍由真实图像的 patch 和生成图像的 patch 组成，即 3360 张训练图像，测试集仍是 40 个真实图像的 patch。

表 4.4 展示了 3 个模型生成的样本对训练数据集扩充后的分类结果，模型 A 使用的是原始的 HP-VAE-GAN，模型 B 在 HP-VAE-GAN 的基础上结合了 CBAM，模型 C 在 HP-VAE-GAN 的基础上既结合了 CBAM，又增加了 1 个卷积块。表 4.4 中的实验结果显示，HP-VAE-GAN 在添加了 CBAM 和 1 个卷积块后，使用 VGG16 进行分类实验时 Top-1 accuracy 的提升最为显著，达到了 95%，在 40 张测试集中，仅预测错了两张样本的类别。只添加了 CBAM 的模型 B 在 ResNet50 和 VGG16 上也都有所提升，这也说明使用 CBAM 和卷积块对模型进行修改，可以进一步提升模型的性能。

表 4.4 每种模型的 Top-1 accuracy

| 模型 | 结构 | Top-1 Accuracy | | |
|----|---------------------------|----------------|----------|--------------|
| | | MobileNet | ResNet50 | VGG16 |
| A | HP-VAE-GAN | 80% | 82.5% | 90% |
| B | HP-VAE-GAN + CBAM | 80% | 87.5% | 92.5% |
| C | HP-VAE-GAN + CBAM + 卷积块*1 | 82.5% | 90% | 95% |

4.3.4 CBAM 与其他注意力机制的对比

为了比较不同注意力机制对模型的影响，本文还做了 SENet、ECANet 和 CBAM 三种注意力机制的对比试验。表 4.5 展示了对比实验的结果，具体的数据集划分策略也和 4.3.3 节中一致。

表 4.5 中，模型 B 是在 HP-VAE-GAN 的基础上结合了 CBAM，模型 D 是在 HP-VAE-GAN 的基础上结合了 SENet，模型 E 是在 HP-VAE-GAN 的基础上结合

了 ECANet。需要说明的是，注意力模块的插入位置相同，都位于编码器第二个卷积块后。表 4.5 中使用 VGG16 进行分类的实验结果表明，相比于在 HP-VAE-GAN 中添加 SENet 和 ECANet，添加 CBAM 的效果更好，Top-1 accuracy 的准确率更高，达到了 92.5%，相比于另外两种注意力机制，CBAM 的优势更明显。

表 4.5 不同注意力机制的对比

| 模型 | 方法 | Top-1 Accuracy | | |
|----|---------------------|----------------|----------|--------------|
| | | MobileNet | ResNet50 | VGG16 |
| B | HP-VAE-GAN + CBAM | 80% | 87.5% | 92.5% |
| D | HP-VAE-GAN + SENet | 82.5% | 85% | 87.5% |
| E | HP-VAE-GAN + ECANet | 85% | 90% | 87.5% |

4.4 本章小结

本章所提方法的三个主要贡献总结如下：

(1) 改进了 HP-VAE-GAN 模型，通过在模型中加入 CBAM，学习到图像多尺度特征的同时，又引入通道域和空间域的特征信息，提高了网络的特征表达能力，使模型的生成效果进一步提升。

(2) 从单样本生成模型的角度出发，仅使用单张样本进行训练，避免了因图像数量稀少导致常见的多样本生成模型（GAN 和 VAE）难以拟合真实数据分布的问题。

(3) 为小样本材料图像数据集提供一种新的数据增强思路，避免分类任务中模型训练产生过拟合问题。4.3 节展示了 UHCSDB 子集上的分类结果，使用改进的 HP-VAE-GAN 进行数据增强后，测试集 Test_C 上的 TOP-1 准确率最高达到了 95%，分类准确率得到大幅度提升。

第五章 总结与展望

5.1 工作总结

本文以小样本材料图像为研究对象,对小样本材料图像增强的相关内容进行了研究,在热障涂层图像和钢铁显微图像等多个材料组织图像中开展了一些探索性的工作,主要取得了以下成果:

(1)提出了一种基于图像增强的热障涂层形貌特征识别方法。该方法包括三个步骤:孔隙轮廓增强与图像去噪、孔隙去除与裂纹修复以及裂纹识别与长度计算,能够顺利完成对热障涂层中裂纹的识别。同时,图像滤波和数学形态学等增强方法的合理使用,使热障涂层中裂纹识别的完整性高、偏差小。所提方法能够自动化识别热障涂层中的裂纹,并计算裂纹长度。相比于人工检测,该方法对裂纹的识别更精确,裂纹长度的计算速度也更快,有效辅助了材料科学领域的研究者便捷高效地分析热障涂层微观结构。

(2)设计了两款用于识别热障涂层形貌特征的软件。一款软件实现了热障涂层图像增强与裂纹的骨架提取,另一款软件实现了裂纹识别与长度计算。两款软件的结合使用,进一步提高了热障涂层微结构分析的速度,降低了时间成本和人力成本。并且,该软件也可处理与热障涂层图像类似的其他材料图像,促进了材料科学的研究与发展。

(3)提出了一种使用改进的 HP-VAE-GAN 进行材料图像数据增强的方法。改进的 HP-VAE-GAN 使用注意力机制 CBAM 模块来细化特征映射,从而提高网络的特征表示能力。与此同时,还在编码器网络中增加了 1 个卷积块,以进一步提高网络的特征提取能力,并消除 CBAM 插入位置对模型性能的影响。生成结果表明,本文提出的结合 CBAM 注意力机制的 HP-VAE-GAN 能够有效地提高生成图像的质量。分类实验的结果表明,该方法比使用 HP-VAE-GAN 进行数据扩充的方法取得了更好的效果,这也为小样本材料图像数据集提供了一种新的数据增强思路。

5.2 未来展望

通过对小样本材料图像的研究与分析,可以发现图像增强和数据增强技术为受小样本问题困扰的材料科学领域提供了强有力的帮助,为了更好地推动材料科学领域的发展,还存在以下问题值得进一步扩充和深入:

(1) 目前不同的目标识别方法都是针对特定任务来设计的特定处理方案,本文所提出的基于图像增强的热障涂层形貌特征识别方法也仅能检测固定光源、相机位置的同一类图像,鉴于此,文中识别方法仍存在提升空间。虽然受光源和相机位置拍摄影响而导致每一张热障涂层图像存在亮度或其他形式上的差别,但热障涂层图像中的基本特征不会有较大改变。因此,处理其他类似的热障涂层图像,仅需要在本文方法中的图像增强部分进行细微调整。下一步的工作重点是针对各类热障涂层图像数据,充分利用特征识别提取出的信息,构建形貌特征参数与涂层热导率之间的定量关系。

(2) 针对第三章中开发的软件,还需进一步设计并开发软件的交互功能,如在裂纹识别与长度计算软件中添加能够调整裂纹长度阈值的功能,增加材料科研工作者与软件的交互,进一步提高材料研究的效率,节省研究时间。

(3) 虽然使用融合 CBAM 的 HP-VAE-GAN 在小样本材料图像生成方面取得了一定效果,但模型仍有待于加强和改进。例如某些生成结果与训练样本存在色彩不一致的情况,在以后的研究中就要消除颜色失真的现象。目前,本文只将提出的数据增强方法用在了材料图像分类中,未来可以将该方法延伸到材料图像分割领域,并进一步探究该方法在其他下游任务中的应用。除此之外,由于本文的研究对象是材料图像,后续可以考虑对医学图像等其他小样本数据集进行单样本图像生成,以解决这些特定领域中的小样本问题。

参考文献

- [1] 张统一. 材料信息学——数据驱动的材料研发[J]. 上海大学学报 (自然科学版), 2022, 28(3): 357-360.
- [2] 范晓丽. 材料基因组计划与第一性原理高通量计算[J]. 中国材料进展, 2015, 34(9): 689-695.
- [3] 谢建新, 宿彦京, 薛德祯, 等. 机器学习在材料研发中的应用[J]. 金属学报, 2021, 57(11): 1343-1361.
- [4] Wang J, Perez L. The effectiveness of data augmentation in image classification using deep learning[J]. Convolutional Neural Networks Vis. Recognit, 2017, 11(2017): 1-8.
- [5] Khosla C, Saini B S. Enhancing performance of deep learning models with different data augmentation techniques: A survey[C]//2020 International Conference on Intelligent Engineering and Management. IEEE, 2020: 79-85.
- [6] Shorten C, Khoshgoftaar T M. A survey on image data augmentation for deep learning[J]. Journal of Big Data, 2019, 6(1): 1-48.
- [7] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of artificial intelligence research, 2002, 16: 321-357.
- [8] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE conference on computer vision and pattern recognition. IEEE, 2009: 248-255.
- [9] Lake B, Salakhutdinov R, Gross J, et al. One shot learning of simple visual concepts[C]//Proceedings of the annual meeting of the cognitive science society. Cognitive Science Society, 2011: 2568-2573.
- [10] Everingham M, Eslami S M A, Van Gool L, et al. The pascal visual object classes challenge: A retrospective[J]. International journal of computer vision, 2015, 111: 98-136.
- [11] Coates A, Ng A, Lee H. An analysis of single-layer networks in unsupervised feature learning[C]//Proceedings of the fourteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings, 2011: 215-223.
- [12] Han Y, Liu Y, Wang B, et al. A novel transfer learning for recognition of overlapping nano

- object[J]. Neural Computing and Applications, 2022, 34(7): 5729-5741.
- [13] Cun Y L, Boser B, Denker J S, et al. Handwritten digit recognition with a back-propagation network[J]. Advances in neural information processing systems, 1990, 2(2): 396-404.
- [14] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [15] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2015-04-10) [2023-05-09]. <https://arxiv.org/abs/1409.1556>.
- [16] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
- [17] Deng L. The mnist database of handwritten digit images for machine learning research [best of the web][J]. IEEE signal processing magazine, 2012, 29(6): 141-142.
- [18] 朱晓慧, 钱丽萍, 傅伟. 图像数据增强技术研究综述[J]. 软件导刊, 2021, 20(05): 230-236.
- [19] Moreno-Barea F J, Strazzeri F, Jerez J M, et al. Forward noise adjustment scheme for data augmentation[C]//2018 IEEE Symposium Series on Computational Intelligence. IEEE, 2018: 728-734.
- [20] Asuncion A, Newman D. UCI machine learning repository[DB/OL]. (2007-06-25) [2023-05-09]. <http://archive.ics.uci.edu/ml/>.
- [21] 邢静, 张艺驰, 高俊钗. 基于 HSV 色彩空间的改进红外与可见光图像融合算法研究[J]. 机械与电子, 2022, 40(12): 15-19.
- [22] Mikołajczyk A, Grochowski M. Data augmentation for improving deep learning in image classification problem[C]//2018 international interdisciplinary PhD workshop. IEEE, 2018: 117-122.
- [23] 王舒雅, 刘强春, 陈云芳, 等. 颜色模型扰动的语义对抗样本生成方法[J]. 计算机工程与应用, 2021, 57(15): 163-170.
- [24] Zhong Z, Zheng L, Kang G, et al. Random erasing data augmentation[C]//Proceedings of the AAAI conference on artificial intelligence. AAAI, 2020: 13001-13008.
- [25] DeVries T, Taylor G W. Improved regularization of convolutional neural networks with cutout[EB/OL]. (2017-11-29) [2023-05-09]. <https://arxiv.org/abs/1708.04552>.

- [26] He K, Chen X, Xie S, et al. Masked autoencoders are scalable vision learners[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2022: 16000-16009.
- [27] 罗丹. 基于模糊均差的低照度图像平滑去噪方法[J]. 科学技术与工程, 2022, 22(27): 12061-12067.
- [28] He K, Sun J, Tang X. Guided image filtering[J]. IEEE transactions on pattern analysis and machine intelligence, 2012, 35(6): 1397-1409.
- [29] He K, Sun J, Tang X. Single image haze removal using dark channel prior[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 33(12): 2341-2353.
- [30] Simard P Y, Steinkraus D, Platt J C. Best practices for convolutional neural networks applied to visual document analysis[C]// Seventh International Conference on Document Analysis and Recognition. IEEE, 2003: 958-963
- [31] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention. Springer, 2015: 234-241.
- [32] 李少荣. 基于改进直方图均衡化的红外图像增强技术的研究[J]. 工业控制计算机, 2022, 35(12): 52-53+56.
- [33] Otsu N. A threshold selection method from gray-level histograms[J]. IEEE transactions on systems, man, and cybernetics, 1979, 9(1): 62-66.
- [34] 唐守锋, 翟少奇, 全光明, 等. 改进 Canny 算子与形态学融合的边缘检测[J]. 计算机工程与设计, 2023, 44(01): 224-231.
- [35] Inoue H. Data augmentation by pairing samples for images classification[EB/OL]. (2018-04-11) [2023-05-09]. <https://arxiv.org/abs/1801.02929>.
- [36] Zhang H, Cisse M, Dauphin Y N, et al. mixup: Beyond empirical risk minimization[EB/OL]. (2018-04-27) [2023-05-09]. <https://arxiv.org/abs/1710.09412>
- [37] Yun S, Han D, Oh S J, et al. Cutmix: Regularization strategy to train strong classifiers with localizable features[C]//Proceedings of the IEEE/CVF international conference on computer vision. IEEE, 2019: 6023-6032.

- [38] Bunkhumpornpat C, Sinapiromsaran K, Lursinsap C. DBSMOTE: density-based synthetic minority over-sampling technique[J]. *Applied Intelligence*, 2012, 36: 664-684.
- [39] Han H, Wang W Y, Mao B H. Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning[C]//*Advances in Intelligent Computing: International Conference on Intelligent Computing*. Springer, 2005: 878-887.
- [40] Bunkhumpornpat C, Sinapiromsaran K, Lursinsap C. Safe-level-smote: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem[C]//*Advances in Knowledge Discovery and Data Mining: 13th Pacific-Asia Conference*. Springer, 2009: 475-482.
- [41] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. *Communications of the ACM*, 2020, 63(11): 139-144.
- [42] Kingma D P, Welling M. Auto-encoding variational bayes[EB/OL]. (2013-12-20) [2023-05-09]. <https://arxiv.org/abs/1312.6114>.
- [43] Bond-Taylor S , Leach A , Long Y , et al. Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 44(11): 7327-7347.
- [44] Croitoru F A, Hondu V, Ionescu R T, et al. Diffusion models in vision: A survey[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 14(8): 1-20.
- [45] Mirza M, Osindero S. Conditional generative adversarial nets[EB/OL]. (2014-11-06) [2023-05-09]. <https://arxiv.org/abs/1411.1784>.
- [46] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[EB/OL]. (2015-11-19) [2023-05-09]. <https://arxiv.org/abs/1511.06434>.
- [47] Chrysos G G, Kossaifi J, Zafeiriou S. Rocgan: Robust conditional gan[J]. *International Journal of Computer Vision*, 2020, 128: 2665-2683.
- [48] Donahue J, Krähenbühl P, Darrell T. Adversarial feature learning[EB/OL]. (2016-05-31) [2023-05-09]. <https://arxiv.org/abs/1605.09782>.
- [49] Zhao J, Mathieu M, LeCun Y. Energy-based generative adversarial network[EB/OL]. (2016-

- 09-11) [2023-05-09]. <https://arxiv.org/abs/1609.03126>.
- [50] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of wasserstein gans[C]//Advances in neural information processing systems. MIT Press, 2017: 5767-5777.
- [51] Zhang H, Goodfellow I, Metaxas D, et al. Self-attention generative adversarial networks[C]//International conference on machine learning. PMLR, 2019: 7354-7363.
- [52] Allahyani M, Alsulami R, Alwafi T, et al. DivGAN: A Diversity Enforcing Generative Adversarial Network for Mode Collapse Reduction[J]. Artificial Intelligence, 2023: 103863.
- [53] Li C, Xu T, Zhu J, et al. Triple generative adversarial nets[C]//Advances in neural information processing systems, MIT Press, 2017: 4088-4098.
- [54] Zhang H, Xu T, Li H, et al. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. IEEE, 2017: 5907-5915.
- [55] Jain N, Olmo A, Sengupta S, et al. Imperfect ImaGANation: Implications of GANs exacerbating biases on facial data augmentation and snapchat face lenses[J]. Artificial Intelligence, 2022, 304: 103652.
- [56] Kollias D, Cheng S, Ververas E, et al. Deep neural network augmentation: Generating faces for affect analysis[J]. International Journal of Computer Vision, 2020, 128: 1455-1484.
- [57] Chen Z, Ouyang W, Liu T, et al. A shape transformation-based dataset augmentation framework for pedestrian detection[J]. International Journal of Computer Vision, 2021, 129(4): 1121-1138.
- [58] Anirudh R, Thiagarajan J J, Kailkhura B, et al. Mimicgan: Robust projection onto image manifolds with corruption mimicking[J]. International Journal of Computer Vision, 2020, 128: 2459-2477.
- [59] Frid-Adar M, Klang E, Amitai M, et al. Synthetic data augmentation using GAN for improved liver lesion classification[C]//2018 IEEE 15th international symposium on biomedical imaging. IEEE, 2018: 289-293.
- [60] Liu Y, Zhou Y, Liu X, et al. Wasserstein GAN-based small-sample augmentation for new-generation artificial intelligence: a case study of cancer-staging data in biology[J]. Engineering, 2019, 5(1): 156-163.

- [61] Gao H, Pei J, Huang H. Progan: Network embedding via proximity generative adversarial network[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2019: 1308-1316.
- [62] Choi Y, Choi M, Kim M, et al. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, 2018: 8789-8797.
- [63] Sun X, Ding X L. Data augmentation method based on generative adversarial networks for facial expression recognition sets [J]. Computer Engineering and Applications, 2020, 56(4): 115-121.
- [64] Zhong Z, Zheng L, Zheng Z, et al. Camera style adaptation for person re-identification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, 2018: 5157-5166.
- [65] 孔祥斌, 沈忠志, 陈树骏. 基于生成对抗网络的铁路周界行人样本生成算法[J]. 铁道通信信号, 2019, 55(07): 57-61.
- [66] Gatys L A, Ecker A S, Bethge M. A neural algorithm of artistic style[EB/OL]. (2015-08-26) [2023-05-09]. <https://arxiv.org/abs/1508.06576>.
- [67] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. IEEE, 2017: 2223-2232.
- [68] Mo S, Cho M, Shin J. Instagan: Instance-aware image-to-image translation[EB/OL]. (2018-12-28) [2023-05-09]. <https://arxiv.org/abs/1812.10889>.
- [69] Perarnau G, Van De Weijer J, Raducanu B, et al. Invertible conditional gans for image editing[EB/OL]. (2016-11-19) [2023-05-09]. <https://arxiv.org/abs/1611.06355>.
- [70] Zhao B, Chang B, Jie Z, et al. Modular generative adversarial networks[C]//Proceedings of the European conference on computer vision. Springer, 2018: 150-165.
- [71] 马征, 褚钧正, 武鹏飞. 一种基于对抗学习的仿真遥感图像生成方法[J]. 计算机工程与科学, 2023, 45(03): 489-494.
- [72] 朱仲贤, 毛语实, 蔡科伟, 等. 面向工业巡检的图像风格迁移方法[J/OL]. 计算机工程与

- 应用: 1-10. (2022-09-27) [2023-05-09]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20220926.1924.016.html>.
- [73] Bao J, Chen D, Wen F, et al. CVAE-GAN: fine-grained image generation through asymmetric training[C]//Proceedings of the IEEE international conference on computer vision. IEEE, 2017: 2745-2754.
- [74] Huang H, He R, Sun Z, et al. Introvae: Introspective variational autoencoders for photographic image synthesis[C]//Advances in neural information processing systems. MIT Press, 2018: 52-63.
- [75] Kingma D P, Salimans T, Jozefowicz R, et al. Improved variational inference with inverse autoregressive flow[C]//Advances in neural information processing systems. MIT Press, 2016: 4743-4751.
- [76] Su J, Wu G. f-VAEs: Improve VAEs with conditional flows[EB/OL]. (2018-09-16) [2023-05-09]. <https://arxiv.org/abs/1809.05861>.
- [77] Vahdat A, Kautz J. NVAE: A deep hierarchical variational autoencoder[C]//Advances in neural information processing systems. MIT Press, 2020: 19667-19679.
- [78] Razavi A, Oord A, Poole B, et al. Preventing posterior collapse with delta-VAEs[EB/OL]. (2019-01-10) [2023-05-09]. <https://arxiv.org/abs/1901.03416>.
- [79] Zhu Q, Su J, Bi W, et al. A batch normalized inference network keeps the KL vanishing away[EB/OL]. (2020-04-27) [2023-05-09].<https://arxiv.org/abs/2004.12585>.
- [80] Miyazawa K, Nagai T. Concept formation through multimodal integration using multimodal BERT and VQ-VAE[J]. Advanced Robotics, 2023, 37(4): 281-296.
- [81] Zhang T, Yang Y, Zhang A. 3D reconstruction of porous media using a batch normalized variational auto-encoder[J]. Computational Geosciences, 2022, 26(5): 1261-1278.
- [82] Zhang F, Teng Q, Chen H, et al. Slice-to-voxel stochastic reconstructions on porous media with hybrid deep generative model[J]. Computational Materials Science, 2021, 186: 110018.
- [83] 杜娟, 胡静. 基于变分自编码器的现代服饰局部中国风格迁移[J]. 毛纺科技, 2021, 49(09): 72-77.
- [84] Huang Y, Mei W, Liu S, et al. Asymmetric Training of Generative Adversarial Network for

- High Fidelity SAR Image Generation[C]//IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2022: 1576-1579.
- [85] Shaham T R, Dekel T, Michaeli T. Singan: Learning a generative model from a single natural image[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, 2019: 4570-4580.
- [86] Hinz T, Fisher M, Wang O, et al. Improved techniques for training single-image gans[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. IEEE, 2021: 1300-1309.
- [87] Gur S, Benaim S, Wolf L. Hierarchical patch vae-gan: Generating diverse videos from a single sample[C]//Advances in Neural Information Processing Systems. MIT Press, 2020: 16761-16772.
- [88] 程国建, 张福临. 基于 SinGAN 的岩石薄片图像超分辨率重建[J]. 西安石油大学学报(自然科学版), 2021, 36(02): 116-121.
- [89] 彭晏飞, 邓佳楠, 王刚. 基于改进 SinGAN 的遥感图像数据增强方法[J]. 液晶与显示, 2023, 38(03): 387-396.
- [90] Liu Z, Mao H, Wu C Y, et al. A convnet for the 2020s[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2022: 11976-11986.
- [91] 胡文杰, 吴晓波, 李波, 等. 基于 Self-Attention 的单样本 ConSinGAN 模型的工业缺陷样本图像生成[J]. 中南民族大学学报(自然科学版), 2022, 41(03): 356-364.
- [92] Wang Y, Han Y, Lin C, et al. Effect of spraying power on the morphology of YSZ splat and micro-structure of thermal barrier coating[J]. Ceramics International, 2021, 47(13): 18956-18963.
- [93] Han Y, Lai C, Wang B, et al. Segmenting Images with Complex Textures by Using Hybrid Algorithm [J]. Journal of Electronic Imaging, 2019, 28 (1): 13-30.
- [94] Ye Y, Li Y, Ouyang R, et al. Improving machine learning based phase and hardness prediction of high-entropy alloys by using Gaussian noise augmented data[J]. Computational Materials Science, 2023, 223: 112140.
- [95] Nawa K, Hagiwara K, Nakamura K. Prediction-accuracy improvement of neural network to

- ferromagnetic multilayers by Gaussian data augmentation and ensemble learning[J]. Computational Materials Science, 2023, 219: 112032.
- [96] Yang Z, Li S, Li S, et al. A two-step data augmentation method based on generative adversarial network for hardness prediction of high entropy alloy[J]. Computational Materials Science, 2023, 220: 112064.
- [97] Ma B, Wei X, Liu C, et al. Data augmentation in microscopic images for material data mining[J]. npj Computational Materials, 2020, 6(1): 1-9.
- [98] Mathematical morphology and its applications to image and signal processing[M]. Springer Science & Business Media, 2006.
- [99] Mathematical morphology: from theory to applications[M]. John Wiley & Sons, 2013.
- [100] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.
- [101] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in neural information processing systems. MIT Press, 2017: 6000–6010.
- [102] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, 2018: 7132-7141.
- [103] Gao Z, Xie J, Wang Q, et al. Global second-order pooling convolutional networks[C]//Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition. IEEE, 2019: 3024-3033.
- [104] Qin Z, Zhang P, Wu F, et al. Fcanet: Frequency channel attention networks[C]//Proceedings of the IEEE/CVF international conference on computer vision. IEEE, 2021: 783-792.
- [105] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. IEEE, 2020: 11534-11542.
- [106] Lee H J, Kim H E, Nam H. Srm: A style-based recalibration module for convolutional neural networks[C]//Proceedings of the IEEE/CVF International conference on computer vision. IEEE, 2019: 1854-1862.
- [107] Yang Z, Zhu L, Wu Y, et al. Gated channel transformation for visual recognition[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern

- recognition. IEEE, 2020: 11794-11803.
- [108]Jaderberg M, Simonyan K, Zisserman A. Spatial transformer networks[C]//Advances in neural information processing systems. MIT Press, 2015: 875-878.
- [109]Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European conference on computer vision. Springer, 2018: 3-19.
- [110]Gupta K, Singh S, Shrivastava A. Patchvae: Learning local latent codes for recognition[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2020: 4746-4755.
- [111]Demir U, Unal G. Patch-based image inpainting with generative adversarial networks[EB/OL]. (2018-03-20) [2023-05-09]. <https://arxiv.org/abs/1803.07422>.
- [112]Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, 2017: 1125-1134.
- [113]Higgins I, Matthey L, Pal A, et al. beta-vae: Learning basic visual concepts with a constrained variational framework[C]//International conference on learning representations. Ithaca, 2017.
- [114]DeCost B L, Hecht M D, Francis T, et al. UHCSDB: ultrahigh carbon steel micrograph database: tools for exploring large heterogeneous microstructure datasets[J]. Integrating Materials and Manufacturing Innovation, 2017, 6: 197-205.
- [115]Kylberg G. Kylberg texture dataset v. 1.0[M]. Centre for Image Analysis, Swedish University of Agricultural Sciences and Uppsala University, 2011.
- [116]Kwitt R, Meerwald P. Salzburg Texture Image Database (STex) [DB/OL]. (2009-05-15) [2023-05-09]. <https://wavelab.at/sources/STex/>.
- [117]Bianconi F, Fernández A. An appendix to “Texture databases—A comprehensive survey” [J]. Pattern recognition letters, 2014, 45: 33-38.
- [118]Kingma D P, Ba J. Adam: A method for stochastic optimization[EB/OL]. (2014-12-2) [2023-05-09]. <https://arxiv.org/abs/1412.6980>.
- [119]Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE transactions on image processing, 2004, 13(4): 600-612.

- [120]Lahitani A R, Permanasari A E, Setiawan N A. Cosine similarity to determine similarity measure: Study case in online essay assessment[C]//2016 4th International Conference on Cyber and IT Service Management. IEEE, 2016: 1-6.
- [121]Kullback S, Leibler R A. On information and sufficiency[J]. The annals of mathematical statistics, 1951, 22(1): 79-86.
- [122]Fuglede B, Topsøe F. Jensen-Shannon divergence and Hilbert space embedding[C]// Proceedings of the International symposium on Information theory. IEEE, 2004: 31-36.
- [123]Howard A G, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17) [2023-05-09]. <https://arxiv.org/abs/1704.04861>.
- [124]He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, 2016: 770-778.

作者在攻读硕士学位期间取得的成果

论文:

- 【1】 刘宇虹, 韩越兴, 汪语嫣等. 基于数字图像处理技术的热障涂层形貌特征识别方法[J]. 上海大学学报(自然科学版), 2022, 28(03): 523-533. (已发表, 中文核心, 第一作者)
- 【2】 Han Y, Liu Y, Chen Q. Data augmentation in material images using the improved HP-VAE-GAN. (已录用, 中科院 3 区, 导师第一作者, 本人第二作者)

专利:

- 【3】 韩越兴,刘宇虹,王冰等. 热障涂层裂纹的几何信息提取方法[P]. 上海市: CN112581433B,2022-10-11. (已授权, 导师排名第一, 本人排名第二)

软著:

- 【4】 软件名称: 热障涂层裂纹的几何信息提取软件 V1.0; 登记号: 2021SR0492853; 登记日期: 2021.04.02. (导师排名第一, 本人排名第二)

作者在攻读硕士学位期间参与的项目

- 【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，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；编号：2021PE0AC02。

致 谢

本文是在导师韩越兴老师的悉心指导下完成的。即使工作繁忙，韩老师仍抽出时间给予我学术上的帮助，为我指点迷津、提供灵感、明确方向，使我受益匪浅。在韩老师的指导下，我的科研能力、学术写作能力、交流能力都得到了一定的提升，这为我以后的工作打下了坚实的基础，也是我人生中一笔宝贵的财富。韩老师严谨的科研方法、敏锐的学术洞察力、勤勉的工作作风以及勇于创新的精神都令我十分敬佩。生活上，韩老师平易近人、关心学生、谦和友善，在老师的组织和带领下，我们组形成了和谐欢乐、互帮互助的良好氛围。在此，谨向韩老师致以深深的敬意和由衷的感谢！

同时，还要感谢陈老师，在平时的组会中，陈老师总是能为大家指出问题并提供思路，陈老师在我写论文时也提供了很多宝贵意见。在此，也向陈老师致以深深的敬意和由衷的感谢！

感谢实验室的每一位小伙伴，是你们使我的研究生生活丰富多彩。王璐学姐的热情开朗、胡干学长的专心致志、夏锦桦同学的乐于助人、李睿祺同学的严谨细致、池洳婷学妹的温文尔雅、万冠新学弟的踏实幽默……从他们身上我学习到了很多为人处事的道理和积极上进的生活态度。祝愿每位小伙伴都前途似锦，万事顺遂！

最后，感谢所有关心我、支持我和帮助过我的同学、朋友、老师和亲人，谢谢你们！