

# 一种基于卷积神经网络的砂岩显微图像特征表示方法\*

李娜<sup>1</sup>, 顾庆<sup>1</sup>, 姜枫<sup>1,2</sup>, 郝慧珍<sup>1,3</sup>, 于华<sup>1</sup>, 倪超<sup>1</sup>



<sup>1</sup>(计算机软件新技术国家重点实验室(南京大学), 江苏 南京 210023)

<sup>2</sup>(南京理工大学 泰州科技学院 移动互联网学院, 江苏 泰州 225300)

<sup>3</sup>(南京工程学院 通信工程学院, 江苏 南京 211167)

通讯作者: 顾庆, E-mail: guq@nju.edu.cn

**摘要:** 砂岩显微图像分类是地质学研究中一项基本工作,在油气储集层评估等方面有重要意义.在实现自动分类时,由于砂岩显微图像具有复杂多变的显微结构,人工定义特征对砂岩显微图像的表达能力有限.此外,由于样本采集和标注成本高昂,带标记的砂岩显微图像很少.提出一种面向小规模数据集的基于卷积神经网络的特征表示方法 FeRNet,以便有效地捕获砂岩显微图像的语义信息,提高对砂岩显微图像的特征表示能力.FeRNet 网络结构简单,可降低网络对带标记图像数据量的要求,防止参数过拟合.针对带标记砂岩显微图像数量不足的问题,提出了图像扩增预处理方法及基于卷积自编码网络的权重初始化策略,降低了因数据不足造成的过拟合风险.基于采自西藏地区的砂岩显微图像数据集设计并进行实验,实验结果表明,在带标记砂岩显微图像数据不足的情况下,图像扩增和卷积自编码网络可以有效地改善 FeRNet 网络的训练效果,通过 FeRNet 网络提取的特征对砂岩显微图像的表达能力优于人工定义特征.

**关键词:** 特征表示;砂岩显微图像;卷积神经网络;图像扩增;卷积自编码

**中图法分类号:** TP391

中文引用格式: 李娜,顾庆,姜枫,郝慧珍,于华,倪超.一种基于卷积神经网络的砂岩显微图像特征表示方法.软件学报,2020,31(11):3621-3639. <http://www.jos.org.cn/1000-9825/5836.htm>

英文引用格式: Li N, Gu Q, Jiang F, Hao HZ, Yu H, Ni C. Feature representation method of microscopic sandstone images based on convolutional neural network. Ruan Jian Xue Bao/Journal of Software, 2020,31(11):3621-3639 (in Chinese). <http://www.jos.org.cn/1000-9825/5836.htm>

## Feature Representation Method of Microscopic Sandstone Images Based on Convolutional Neural Network

LI Na<sup>1</sup>, GU Qing<sup>1</sup>, JIANG Feng<sup>1,2</sup>, HAO Hui-Zhen<sup>1,3</sup>, YU Hua<sup>1</sup>, NI Chao<sup>1</sup>

<sup>1</sup>(State Key Laboratory for Novel Software Technology (Nanjing University), Nanjing 210023, China)

<sup>2</sup>(College of Mobile Internet, Taizhou Institute of Science and Technology, Nanjing University of Science and Technology, Taizhou 225300, China)

<sup>3</sup>(School of Communication Engineering, Nanjing Institute of Technology, Nanjing 211167, China)

**Abstract:** The classification of microscopic sandstone images is a basic work in geological research, and it has an important significance in the evaluation of oil and gas reservoirs. In the automatic classification of microscopic sandstone images, due to their complex and variable micro-structures, the hand-crafted features have limited abilities to represent them. In addition, since the collection

\* 基金项目: 国家自然科学基金(61373012, 61321491, 91218302); 国家重点研发计划(2018YFB1003800); 软件新技术与产业化协同创新中心

Foundation item: National Natural Science Foundation of China (61373012, 61321491, 91218302); National Key Research and Development Program of China (2018YFB1003800); Collaborative Innovation Center of Novel Software Technology and Industrialization

收稿时间: 2018-08-27; 修改时间: 2018-11-29; 采用时间: 2019-03-11

and labeling of sandstone samples are costly, labeled microscopic sandstone images are usually few. In this study, a convolutional neural network based feature representation method for small-scale data sets, called FeRNet, is proposed to effectively capture the semantic information of microscopic sandstone images and enhance their feature representation. The FeRNet has a simple structure, which reduces the quantity requirements for labeled images, and prevents the overfitting. Aiming at the problem of insufficient labeled microscopic sandstone image, the image augmentation preprocessing and a CAE network-based weight initialization strategy are proposed, to reduce the risk of overfitting. Based on the microscopic sandstone images collected from Tibet, the experiments are designed and conducted. The results show that both image augmentation and CAE network can effectively improve the training of FeRNet network, when the labeled microscopic sandstone images are few; and the FeRNet features are more capable of the representations of microscopic sandstone images than the hand-crafted features.

**Key words:** feature representation; microscopic sandstone image; convolutional neural network; image augmentation; convolutional autoencoder

在地质学中,砂岩鉴别是一项非常重要的任务.砂岩是一种沉积岩,是地壳的重要组成部分,在自然界分布广泛.按照其组成成分,砂岩可分为长石砂岩、岩屑砂岩和石英砂岩这 3 大类.砂岩和砂层是石油、天然气和地下水的主要储集层,世界上 40%的油气都储集在该类储层中.我国的大庆、大港、克拉玛依等油田的主要储集层均为砂岩类.因此,有效的砂岩鉴别在油气储层评估、资源勘探、水利勘测等方面有重要意义.

人工砂岩鉴别主要是通过地质工作者在光学显微镜下观察砂岩薄片的光学性质,研究其组成成分,确定砂岩类型.人工鉴别方法耗时费力,且鉴别结果有很强的个人主观性,不同的鉴别人员可能会给出不同的鉴别结果,鉴别效率低.随着计算机技术的发展,砂岩薄片通过显微成像技术被制备成图像,并结合图像处理和机器学习等方法<sup>[1-5]</sup>,实现砂岩薄片的自动分类,提高砂岩鉴别效率.

在基于机器学习的砂岩自动分类方法中,砂岩显微图像的特征表示是分类模型训练和预测过程中的重要依据.特征表示的优劣,在很大程度上会对分类模型的预测性能产生影响.因此,合理的特征表示是实现砂岩自动分类的基础.

目前已有的砂岩分类方法主要采用人工定义特征来描述砂岩显微图像,包括图像的像素颜色<sup>[3,6]</sup>、纹理<sup>[7-11]</sup>和形状<sup>[5,12]</sup>等.人工定义特征主要描述砂岩显微图像的低级结构信息,未能捕获其高级语义<sup>[13]</sup>,对结构复杂的砂岩显微图像特征表示能力受到限制.卷积神经网络(CNN)<sup>[14]</sup>是一种可用于提取图像高级语义的方法.近年来,卷积神经网络已被成功应用于自然场景图像分类中,并取得了显著成果.但是砂岩显微图像分类与自然场景图像分类存在一定区别:自然场景图像有长时间积累的数据支持,易满足深度卷积神经网络在训练过程中对数据量的要求;而砂岩显微图像在薄片制备、显微图像采集过程中需要非常精密的仪器,且砂岩显微图像类别需要领域专家标注,时间和经济成本高,因此带标记砂岩显微图像的数量往往很少,难以满足网络训练的要求.

针对以上问题,本文提出一种面向小规模数据集的卷积神经网络 FeRNet,用于提取砂岩显微图像的语义特征,提高砂岩显微图像的分类准确性.本文的主要贡献包括:

- (1) 针对砂岩显微图像的特征表示问题,本文提出一种基于卷积神经网络 FeRNet 的特征表示方法,提取砂岩显微图像中隐含的高级语义特征,为分类模型提供更具判别性的信息.
- (2) 根据砂岩显微图像裁剪语义不变特性,本文提出一种基于滑动窗口的图像扩增预处理方法,通过滑动窗口对原图像进行裁剪,倍数级增加带标记砂岩显微图像的数量,改善 FeRNet 网络的训练效果.
- (3) 针对带标记砂岩显微图像数量不足的问题,本文提出一种卷积自编码网络 FeRCAE,预编码砂岩显微图像的语义信息,并使用编码器权重初始化 FeRNet 网络,改善 FeRNet 网络的训练效果.

本文第 1 节介绍图像特征表示的相关工作,主要包括人工定义特征、基于深度学习的自学习特征和面向小规模数据集的深度自学习特征.第 2 节介绍本文提出的用于砂岩显微图像特征表示的卷积神经网络 FeRNet,主要描述 FeRNet 的结构及训练过程等.第 3 节详细描述实验设计方案,包括数据集、评价指标、实验设置等.第 4 节分析实验结果,探究采用 FeRNet 网络提取的语义特征对砂岩显微图像自动分类预测性能的影响.第 5 节对方法进行讨论,主要探究图像扩增预处理、卷积自编码网络和激活函数的选择对 FeRNet 网络训练的影响.最后总结全文.

## 1 相关工作

### 1.1 人工定义图像特征

在图像分类应用中,需要根据图像的特征表示进行分析.特别是在基于机器学习的图像分类方法中,作为分类模型训练和预测的重要依据,图像的特征表示尤为重要.在深度学习被用于图像特征表示之前,人工定义的图像特征是使用最广泛的特征表示方法.按特征描述的图像范围,人工定义特征可分为全局特征和局部特征两种.

全局特征主要描述图像的整体属性,为像素级的低层可视特征.常见的全局特征包括图像的颜色、纹理和形状等.颜色特征主要基于直方图、累积直方图、平均灰度级等计算均值、方差等统计特征<sup>[3,6]</sup>,具有计算简单、性能稳定等优点.纹理特征对粗细和疏密等方面存在较大差别的图像具有不错的表示能力.经典的全局纹理特征包括小波变换<sup>[7-9]</sup>、灰度共生矩阵<sup>[10,11]</sup>等.相比颜色和纹理特征,形状特征带有一些语义信息,尽可能提取图像中目标对象的轮廓或区域特征.但是由于形状特征的表示通常以图像分割为基础,而大部分现有图像分割算法不完全准确,给形状特征的提取带来一定难度.现有的形状特征有二值化阈值、Canny 边界像素数目、周长像素数目、白色区域数目以及白色区域面积等<sup>[15,15]</sup>.全局特征能够表示图像整体信息,但是难以描述图像的细节变化,图像信息利用率较差.

局部特征主要对图像的局部区域进行描述,更关注图像的细节属性,包括边缘、角点、曲线等.经典的局部特征包括局部二值模式(LBP)<sup>[12]</sup>、Gabor 小波变换<sup>[16]</sup>、尺度不变特征变换(SIFT)<sup>[17]</sup>和加速鲁棒特征(SURF)<sup>[18]</sup>等.针对图像的局部特征表示,大量学者进行了深入研究,并提出许多有效的方法.周书仁等人<sup>[19]</sup>提出一种具有 Haar 型特性局部二值模式(HLBP)的图像局部纹理特征表示方法,在确保较低计算复杂度的前提下,有效提取深层次的图像纹理信息.孙俊等人<sup>[20]</sup>基于 LBP 和 Damper-Shafter 证据理论,提出一种结合二者优势的多尺度纹理识别方法,实验表明,作者提出的方法优于基于 LBP 纹理的方法.

全局图像特征和局部图像特征描述的图像属性不同,仅基于单一模态的图像特征不能全面地描述图像属性.因此,一些研究者融合不同类型的特征,实现特征之间的优势互补,提高图像的特征表示能力<sup>[21-23]</sup>.

### 1.2 基于深度学习的自学习特征

人工定义图像特征可以描述图像的结构、纹理等属性,这些特征主要表示图像的低级语义<sup>[13]</sup>.对于简单结构图像,人工定义特征可以较好地描述其属性;但对于复杂结构图像,如砂岩显微图像,人工定义特征不具有很好的表示能力.

卷积神经网络<sup>[14]</sup>可以有效地提取图像的高级语义,常被用于图像的自动特征表示.LeCun 等人<sup>[14]</sup>最早提出一种有监督的反向传播网络 LeNet,自动学习图像特征,实现数字图像的识别.LeNet 网络可认为是卷积神经网络的雏形.随后,AlexNet<sup>[24]</sup>网络在 2012 年的 ImageNet 图像分类竞赛中以绝对优势获得冠军,卷积神经网络的发展取得历史性突破.

近年来,网络层次更深、特征表示能力更强的 VGG<sup>[25]</sup>、GoogleNet<sup>[26]</sup>、ResNet<sup>[27]</sup>等网络相继出现,卷积神经网络受到广泛关注和研究.许多研究者对已有的卷积神经网络进行改进,应用于不同图像的特征表示.Amrouch 等人<sup>[28]</sup>基于 LeNet-5 网络提取手写图像的语义特征,并结合隐马尔可夫模型(HMM)识别阿拉伯字母构成的单词.白琮等人<sup>[29]</sup>提出一种可应用于大规模图像分类的深度学习框架,从网络框架和内部结构两方面对 AlexNet 网络进行优化和改进,提高网络对图像的特征表示能力.

### 1.3 面向小规模数据集的深度学习特征

随着网络层次的加深,需训练的网络参数也大量增加,网络训练对带标记样本的数量要求也急剧上升.但是由于数据标注的高昂成本,在很多实际应用中这种要求往往难以满足.目前,已有研究者致力于面向小规模数据集的图像特征表示的研究.

为尽可能利用有限的带标记样本,研究者通过一系列随机变换对图像进行预处理,提升数据量.常见的方法包括旋转(rotation)、缩放(zoom)、反射(reflection)、翻转(flipping)、错切(shear mapping)和尺度变换(scaling)

等<sup>[24,30,31]</sup>对图像进行随机变换,可保证用于训练深度网络的图像不会完全相同,有利于提升有效的训练样本量,防止过拟合现象产生。

利用深度学习网络的可重用性,一些研究者将在大规模数据集(如 ImageNet)上训练好的模型(如 VGG, GoogleNet 等)应用于小规模数据集图像的特征表示或分类.对与原数据集相近的小规模数据集,一些方法使用预训练模型的卷积层或全连接层的输出作为图像的特征表示,并结合支持向量机等分类模型实现分类<sup>[32-34]</sup>. Donahue 等人<sup>[34]</sup>提取 Krizhevsky 等人<sup>[24]</sup>预训练的卷积神经网络不同层次的输出作为图像特征,实现目标检测、场景识别等计算机视觉相关任务.另一些研究者基于预训练的深度卷积模型,在新数据集上对网络的部分或全部参数进行微调(fine-tuning),提高图像的特征表示或分类效果<sup>[35,36]</sup>.Yosinski 等人<sup>[36]</sup>采用层冻结方法在小规模数据集上微调卷积网络,实现图像分类。

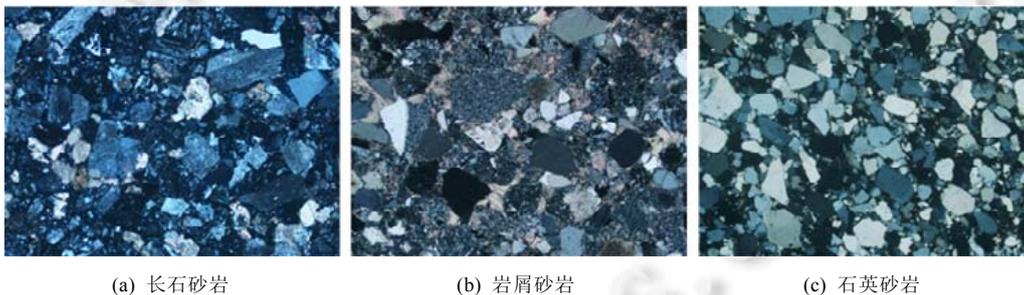
此外,一些研究者也通过减少网络的参数数目,如调整网络的层数及各网络层的规模等,或对参数进行正则化约束来降低对带标记样本的数量要求,防止过拟合。

## 2 特征表示方法

为提高砂岩显微图像的特征表示,本文提出一种面向小规模数据集的卷积神经网络 FeRNet,提取砂岩显微图像的语义信息,提高砂岩显微图像的自动分类性能.针对带标记砂岩显微图像数量不足的问题,本文分别提出了基于滑动窗口的图像扩增预处理方法和基于卷积自编码网络的权重初始化策略,改善 FeRNet 网络的训练效果,防止过拟合。

### 2.1 方法动机

砂岩由各种不同的砂粒胶结而成,因此砂岩显微图像具有复杂的显微结构.图 1 所示为 3 种类型砂岩显微图像示例.如图 1 所示,长石砂岩显微图像的矿物颗粒多表现出双晶、解理和裂理等特性,部分颗粒形状近似不规则矩形,颗粒边界不清晰;岩屑砂岩显微图像包含大量细碎的颗粒,颗粒边界难以分辨;石英砂岩显微图像表面较为平滑,颗粒形状近似椭圆形,颗粒边界较为清晰.复杂多变的显微结构导致砂岩显微图像的特征表示难度增加。



(a) 长石砂岩

(b) 岩屑砂岩

(c) 石英砂岩

Fig.1 Different types of microscopic sandstone images

图 1 不同类型砂岩显微图像

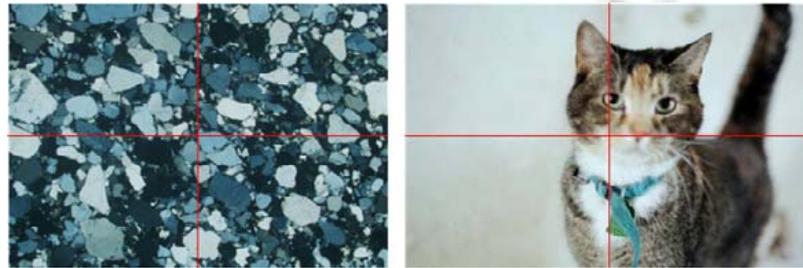
在采用机器学习技术实现自动分类时,砂岩显微图像的特征表示对分类模型的训练非常重要,但现有方法对砂岩显微图像的特征表示能力受到限制.目前,多数自动分类方法主要采用人工定义特征来描述砂岩显微图像<sup>[5]</sup>,包括像素颜色、纹理和形状等.然而,人工定义特征只能描述砂岩显微图像较低级的结构信息,不能捕获其隐含的高级语义信息,可能削弱砂岩显微图像的预测性能.近年来,基于深度卷积神经网络的自动特征表示方法受到广泛关注.经典的卷积神经网络,如 VGG<sup>[25]</sup>、GoogleNet<sup>[26]</sup>等,包含大量的网络层,其训练需要大量的带标记数据支持.但是砂岩显微图像的采集和标注成本高昂,带标记的砂岩显微图像数量往往很少,难以满足这些网络对数据量的要求。

基于以上原因,本文提出一种面向小规模数据集的卷积神经网络 FeRNet,用于砂岩显微图像的特征表示,

旨在挖掘图像隐含的高级语义信息,提高砂岩显微图像的预测性能.针对砂岩显微图像带标记数据不足的问题,本文提出一种基于滑动窗口的图像扩增策略,并采用预训练的卷积自编码网络 FeRCAE 的编码器权重来初始化 FeRNet 网络,改善 FeRNet 网络训练效果,防止过拟合.

## 2.2 图像扩增

将卷积神经网络应用于自然场景图像的特征表示,可使其分类性能取得很大程度的提升<sup>[24]</sup>.砂岩显微图像与自然场景图像存在一定差别.如图 2(b)所示,自然场景图像区分前景和背景,前景部分为感兴趣的目标对象(图中的猫),一般只占整张图像的一部分.若对原图像进行裁剪,裁剪后的图像可能只包含目标对象的一小部分,原图的语义信息将被严重破坏.如图 2(a)所示,与自然场景图像不同,砂岩显微图像由不同的矿物颗粒构成,整张图像均为感兴趣的目标区域.对砂岩显微图像进行适当地裁剪,裁剪后的图像能保留原图绝大部分的语义信息,不会对图像分类结果造成影响.



(a) 砂岩显微图像示例

(b) 自然场景图像示例

Fig.2 Comparison between the microscopic sandstone image and natural scene image

图 2 砂岩显微图像与自然场景图像对比

基于以上分析,砂岩显微图像允许对原图像进行合理的裁剪,且不改变原图像的语义信息.本文将砂岩显微图像这一特性称为裁剪语义不变性.根据此特性,本文提出一种基于滑动窗口的图像扩增方法,对砂岩显微图像以裁剪的方式进行预处理,实现图像数量的倍数级扩增.具体处理方式如下.

- (1) 对原始砂岩显微图像  $I$ , 设置滑动窗口  $w$ .
- (2) 滑动窗口以步长  $h$  沿水平方向扫描图像  $I$ , 直至到达图像当前“行”最右端, 本轮扫描结束.
- (3) 滑动窗口返回至本轮扫描起始位置.
- (4) 将滑动窗口沿垂直方向滑动步长  $v$ , 从新位置开始, 执行步骤(2)和步骤(3).
- (5) 重复步骤(2)~步骤(4), 直至扫描整张砂岩显微图像.

每次扫描取滑动窗口内的图像为一幅新的砂岩显微子图, 以图像  $I$  的类标对其进行标记并保存. 图 3 所示为基于滑动窗口的图像扩增示意图, 其中, 线框表示滑动窗口, 虚线和实线框分别表示滑动窗口先前位置和当前位置, 实心箭头表示窗口滑动方向.

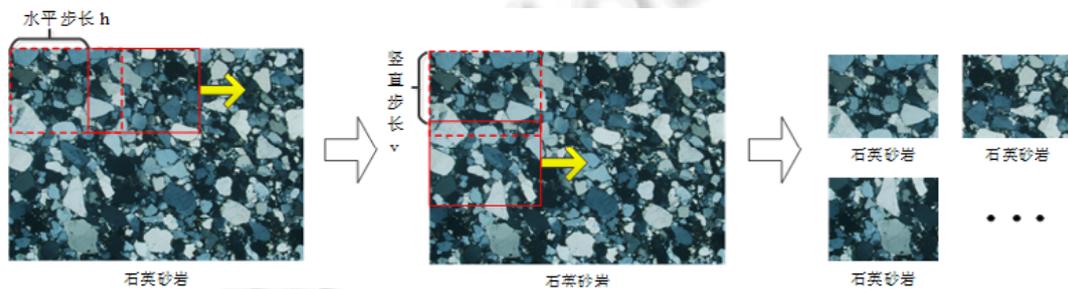


Fig.3 Image augmentation based on sliding window

图 3 基于滑动窗口的图像扩增

采用上述图像扩增方式,相邻子图像间会存在部分重合,但不会完全相同.通过对砂岩显微图像进行裁剪预处理,可有效扩增带标记砂岩显微图像数量,有助于防止过拟合,提高模型的泛化性能.

## 2.3 FeRNet特征表示

### 2.3.1 FeRNet 网络结构

人工定义图像特征对具有复杂显微结构的砂岩显微图像表示能力有限,不能捕获图像中隐含的高级语义信息,削弱砂岩显微图像的自动分类性能.此外,由于砂岩样本采集和标注的困难性,带标记的砂岩显微图像数量远远难以达到深度神经网络训练的要求.

针对以上问题,本文提出一种卷积神经网络 FeRNet,用于表示砂岩显微图像的语义信息.FeRNet 网络具有相对简单的结构,在保证对砂岩显微图像特征表示能力的前提下,尽可能地降低网络对带标记样本数据量的要求,避免出现过拟合风险.图 4 所示为本文提出的 FeRNet 网络结构示意图,表 1 列出了网络的详细设置.

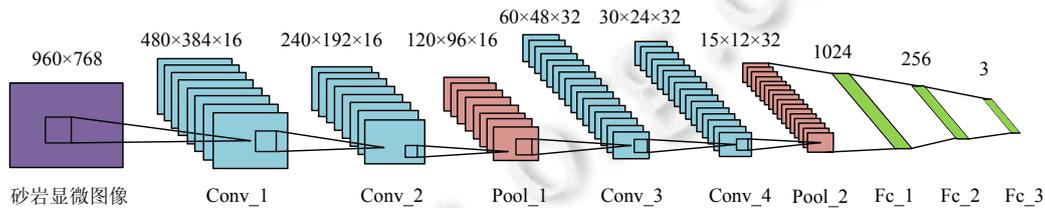


Fig.4 Architecture of FeRNet network

图 4 FeRNet 网络结构

Table 1 Detailed settings of FeRNet network

表 1 FeRNet 网络详细设置

网络层	层类型	过滤器/步长	通道数	激活函数	输出尺寸	参数数量
Conv_1	卷积层	5x5/2	16	ReLU <sup>[24]</sup>	480x384x16	1 216
Conv_2	卷积层	5x5/2	16	ReLU <sup>[24]</sup>	240x192x16	6 416
Pool_1	最大池化层	2x2/1	16	-	120x96x16	0
Conv_3	卷积层	5x5/2	32	ReLU <sup>[24]</sup>	60x48x32	12 832
Conv_4	卷积层	5x5/2	32	ReLU <sup>[24]</sup>	30x24x32	25 632
Pool_2	最大池化层	2x2/1	32	-	15x12x32	0
Fc_1	全连接层	-	1	ReLU <sup>[24]</sup>	1 024	5 899 864
Dropout_1	Dropout 层	-	1	-	1 024	0
Fc_2	全连接层	-	1	ReLU <sup>[24]</sup>	256	262 400
Dropout_2	Dropout 层	-	1	-	256	0
Fc_3	全连接层	-	1	Softmax <sup>[37]</sup>	3	771

如表 1 所示,本文提出的 FeRNet 网络主要由 11 个网络层构成,其中包括 4 个卷积层(Conv)、2 个最大池化层(Pool)、3 个全连接层(Fc)以及 2 个 Dropout 层.FeRNet 网络的具体架构详细介绍如下.

#### (1) 输入层

为避免因图像缩放等原因造成的数据损失,预处理后的砂岩显微图像不再进行任何缩放处理,直接输入 FeRNet 网络.因此,网络输入层图像的尺寸为 960x768.但在输入网络之前,需要将砂岩显微图像的像素值归一化至[0,1]范围,方便计算.

#### (2) 卷积层

为捕获砂岩显微图像的语义信息,同时避免因数据不足造成过拟合,FeRNet 网络只包含 4 个卷积层.相比深度卷积网络,FeRNet 网络在保证对砂岩显微图像特征表示能力的前提下,尽可能减少网络的参数数量,降低网络训练对数据量的要求.所有卷积层均采用尺寸为 5x5 的卷积核(过滤器),步长均设为 2.前两个卷积层通道数为 16,后两个为 32.

常用的激活函数包括 Sigmoid<sup>[38]</sup>、Tanh<sup>[37]</sup>和 ReLu<sup>[24]</sup>.与 Sigmoid 和 Tanh 函数相比,ReLU 函数计算简单,在

网络训练时具有更快的收敛速度,且不存在梯度饱和问题.因此,FeRNet 网络卷积层均采用 ReLu 函数作为激活函数,其定义为

$$ReLU(x)=\max(0,x).$$

### (3) 最大池化层

为降低数据的特征维度,防止参数过拟合,在每两个卷积层之后加入一个最大池化层对特征图进行下采样.最大池化层的卷积核(过滤器)采用  $2 \times 2$  大小,步长设为 1.每次对特征图进行最大池化操作后,输出特征图变为输入大小的  $1/2$ .

### (4) 全连接层

全连接层的目的是将卷积层和池化层输出的特征图映射至样本标记空间.FeRNet 共包含 3 个全连接层.前两个全连接层使用 ReLu 作为激活函数,输出维度分别为 1 024 和 256.最后一个全连接层为 FeRNet 网络的输出层,属于分类层,采用 Softmax<sup>[37]</sup>作为激活函数,其定义为

$$Softmax(x) = \frac{e^{x_j}}{\sum_{n=1}^N e^{x_n}},$$

其中,  $N$  为输出神经元的数目.本文中  $N=3$ ,对应于 3 种砂岩显微图像类型,即长石砂岩、岩屑砂岩和石英砂岩.

### (5) Dropout 层

为了进一步防止参数过拟合,在前两个全连接层之后各增加一个 Dropout 层<sup>[24]</sup>,在训练过程中更新参数时以 50%的概率随机断开输入神经元.

FeRNet 网络的目标是对砂岩显微图像进行特征表示,提取其语义信息.因此在网络训练过程中,所有网络层均参与网络参数的训练;在特征提取阶段,则丢弃 Dropout\_2 和 Fc\_3 层,将第 2 个全连接层 Fc\_2 作为特征提取器,取其输出作为砂岩显微图像的特征表示.因此,本文共提取 256 维语义特征来描述砂岩显微图像.

## 2.3.2 FeRNet 网络训练

如前所述,带标记砂岩显微图像数量不足可能会削弱砂岩显微图像的预测准确性.针对此问题,上文分别从图像扩增和 FeRNet 网络结构两个方面提出解决方案.本节将从 FeRNet 网络训练的角度进行优化.在 FeRNet 网络训练之前,预先训练一个卷积自编码网络<sup>[39]</sup>(简称 FeRCAE),用于初始化 FeRNet 网络的权重,强化 FeRNet 网络的学习.FeRCAE 网络的详细设置在表 2 中列出.

Table 2 Detailed settings of FeRCAE network

表 2 FeRCAE 网络详细设置

网络层	层类型	过滤器/步长	通道数	激活函数	输出尺寸	参数数量	
编码器	Conv_1	卷积层	5×5/2	16	ReLU <sup>[24]</sup>	480×384×16	1 216
	Conv_2	卷积层	5×5/2	16	ReLU <sup>[24]</sup>	240×192×16	6 416
	Pool_1	最大池化层	2×2/1	16	—	120×96×16	0
	Conv_3	卷积层	5×5/2	32	ReLU <sup>[24]</sup>	60×48×32	12 832
	Conv_4	卷积层	5×5/2	32	ReLU <sup>[24]</sup>	30×24×32	25 632
	Pool_2	最大池化层	2×2/1	32	—	15×12×32	0
解码器	UpSampling_1	上采样	2×2/1	32	—	15×12×32	5 899 864
	DeConv_1	反卷积层	5×5/2	32	ReLU <sup>[24]</sup>	30×24×32	25 632
	DeConv_2	反卷积层	5×5/2	32	ReLU <sup>[24]</sup>	60×48×32	25 632
	UpSampling_2	上采样	2×2/1	16	—	120×96×16	0
	DeConv_3	反卷积层	5×5/2	16	ReLU <sup>[24]</sup>	240×192×16	1 2816
	DeConv_4	反卷积层	5×5/2	16	Sigmoid <sup>[38]</sup>	480×384×16	1 203

FeRCAE 网络的输入和输出均为砂岩显微图像,大小为  $960 \times 768$ .为了便于计算,输入砂岩显微图像像素值归一化至  $[0,1]$  范围.如表 2 所示,FeRCAE 网络由编码器和解码器两部分组成.编码器的目的是将输入的砂岩显微图像压缩至一个潜在的特征表示空间,编码砂岩显微图像隐含的语义信息.编码器包含 4 个卷积层和 2 个最大池化层,网络层的设置与 FeRNet 网络相同.解码器的目的是根据编码器输出的语义特征重构出砂岩显微图

像.解码器包含 2 个上采样层(UpSampling)<sup>[40]</sup>和 4 个反卷积层(DeConv)<sup>[41]</sup>.

上采样层和反卷积层介绍如下.

### (1) 上采样层

上采样层的目的是将特征图还原为池化操作之前的大小.FeRCAE 网络中,上采样层采用最近邻插值法计算目标特征图的像素值.如图 5 所示,最近邻插值法将目标特征图中各点的像素值设为原特征图中与其最邻近点的像素值.上采样层的卷积核(过滤器)采用  $2 \times 2$  大小,步长设为 1.

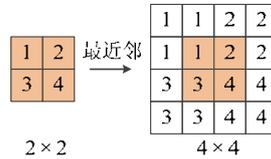


Fig.5 An example for nearest neighbor upsampling

图 5 最近邻插值上采样示例

### (2) 反卷积层

反卷积操作是卷积操作的逆过程,目的是尽可能还原特征图执行卷积操作之前的信息.FeRCAE 网络中,所有反卷积层均采用尺寸为  $5 \times 5$  的卷积核(过滤器),步长均设为 2.前 3 个反卷积层采用 ReLu 激活函数.最后一个反卷积层采用 Sigmoid<sup>[38]</sup>激活函数,目的是将输出图像的像素值缩放至  $[0,1]$  区间,与输入图像对应.

FeRCAE 网络的优化目标是 minimized 重构的砂岩显微图像与输入图像之间的差异.本文采用均方误差作为 FeRCAE 网络的损失函数,其目标函数形式化为

$$\text{minimize } \frac{1}{2} \|I - O\|_2^2,$$

其中,  $I$  和  $O$  分别表示 FeRCAE 网络的输入和输出砂岩显微图像.在 FeRCAE 网络优化过程中,当达到最大迭代次数或损失函数值小于规定阈值时停止训练.本文设置最大迭代次数为 100.

如表 2 所示,FeRCAE 网络的编码器部分与 FeRNet 网络的卷积部分结构一致.故在 FeRCAE 网络训练完成后,将编码器部分的参数用于 FeRNet 网络卷积部分的权重初始化,缓解带标记样本数据量不足对网络训练的影响,改善 FeRNet 网络的训练效果.

## 3 实验设计

本节详细说明实验设计方案,包括数据集和评测指标等.本文实证研究主要探究 3 个问题.

- RQ1:与人工定义特征相比,本文 FeRNet 网络提取的特征在砂岩显微图像自动分类问题中是否具有更好的预测性能?
- RQ2:与其他面向小规模数据集的图像特征表示方法相比,本文 FeRNet 网络提取的砂岩显微图像特征是否具有更好的预测性能?
- RQ3:本文 FeRNet 网络提取的砂岩显微图像特征与哪种分类模型结合,能够获得更好的预测性能?

### 3.1 砂岩显微图像数据集

本文选用来自西藏 4 个不同地区的砂岩显微图像构成数据集进行实验研究.实验共采集 344 张砂岩显微图像,其中,长石砂岩 33 张,岩屑砂岩 206 张,石英砂岩 105 张.每张显微图像都由一个独立的砂岩薄片采集而来,没有重复.所有砂岩显微图像的规格为  $1920 \times 1536$  像素.

如上所述,该数据集中带标记砂岩显微图像数量较少,特别是长石砂岩,确实难以满足 FeRNet 网络训练的数据量要求.但本文采集的砂岩显微图像分辨率较高,且具有裁剪语义不变特征,支持第 2.2 节所述的图像扩增操作.在通过滑动窗口对原图像进行裁剪时,窗口大小的设置非常重要:若窗口设置太大,则可能减小数据的扩增空间,且容易造成相邻子图像间存在大面积重合;若窗口设置太小,则不能保留原图的语义信息,可能导致扩

增后的砂岩显微图像数据集类别标注不准确.综合以上考虑,本文选取原图像 1/4 大小的滑动窗口对原砂岩显微图像进行裁剪,裁剪后子图像大小为 960×768 像素,扩增图像数据量的同时,尽可能保证裁剪前后图像的语义信息保持不变.

为保证实验的合理性,本文将原砂岩显微图像数据集均分为两部分处理:一部分砂岩显微图像用于 FeRNet 网络的训练,称为 CNN 训练集;另一部分用于不同方法的实验对比,称为验证数据集.对于 CNN 训练集中的每张砂岩显微图像,滑动窗口沿水平和竖直方向各扫描 10 次,即每张原始砂岩显微图像被裁剪为 100 张新的子图像.采用图像扩增预处理后,CNN 训练集的图像数量被扩增为原来的 100 倍.对于验证数据集中的砂岩显微图像,滑动窗口在每张原图像中沿水平和竖直方向各扫描 2 次,故每张原始砂岩显微图像被裁剪为 4 张完全无重合的子图像,保证实验中任意两张砂岩显微子图均不相同.表 3 列出不同设置下砂岩显微图像数据集的详细信息.

**Table 3** Number of microscopic sandstone images in different cases

**表 3** 不同情况下各类砂岩显微图像的数量

	图像数量			总计
	长石砂岩	岩屑砂岩	石英砂岩	
原始数据集	33	206	105	344
CNN 训练集	1700 (17×100)	10300 (103×100)	5300 (53×100)	17 300
验证数据集	64 (16×4)	412 (103×4)	208 (52×4)	684

### 3.2 评测指标

如表 3 所示,本文使用的砂岩显微图像数据集存在类别不平衡问题.为此,本文采用类别比例加权的精确度和  $F1$  值作为评测指标,客观地评测不同方法的预测表现.两种评测指标都适用于对不平衡数据预测准确性的评估.加权的精确度(*accuracy*)指标计算公式为

$$Accuracy = \frac{n_c}{\sum_{c=1}^N n_c} \times acc_c,$$

其中, $n_c$  和  $acc_c$  分别表示第  $c$  类砂岩显微图像的样本数与分类精度; $N$  表示砂岩类别总数,本文中  $N=3$ .

加权的  $F1$  指标的计算公式为

$$F1 = \frac{n_c}{\sum_{c=1}^N n_c} \times \frac{2TP_c}{2TP_c + FN_c + FP_c},$$

其中, $TP_c$ 、 $FP_c$  和  $FN_c$  分别表示第  $c$  类砂岩显微图像视为正例时的真正例、假正例和假反例样本数.加权的精确度指标和  $F1$  指标的取值范围均为  $[0,1]$ ,其值越大,表示分类越准确,特征表示能力越好.

### 3.3 实验设计

为验证本文 FeRNet 网络对砂岩显微图像的特征表示能力,本文基于第 3.1 节所述的验证数据集进行实验和分析.针对目标研究问题,本文设计 3 个实验进行验证.本节将对各个实验的详细设置进行描述.

#### 3.3.1 人工定义特征

针对研究问题 1,本文选取 3 类常用的人工定义特征作为对比:颜色、纹理和基于颗粒形状的特征<sup>[5]</sup>.

##### (1) 颜色特征(RGB、HSV 和 Gray)

颜色特征包括 RGB、HSV 及灰度空间下提取的特征.特征指标包含均值、方差、中位数、众数、值域、平均绝对偏差、四分位差、平滑度、均一性、熵、偏度、峰值和分形维数等 13 个统计量,所有特征基于砂岩显微图像的色阶分布直方图计算得到.不同颜色空间下的特征维度分别为 39,39 和 13.

##### (2) 纹理特征(GLCM 和 Wavelet)

纹理特征包括基于灰度共生矩阵(GLCM)和基于小波变换(Wavelet)的特征,均根据砂岩显微图像的灰度图像计算得到.基于灰度共生矩阵的图像特征首先计算位移矢量分别为(1,0),(1,1),(0,1)和(-1,1)的 4 个灰度共生矩阵,并基于归一化的矩阵分别计算能量、熵、对比度和相关性 4 个统计指标.基于小波变换的纹理特征首先将

灰度图像进行小波分解,得到图像水平、竖直、对角线方向分量及其低频分量的小波系数矩阵,再分别统计 4 个系数矩阵的均值、方差、中位数等 13 个特征指标.两种纹理特征的维度分别为 16 和 52.

### (3) 基于颗粒形状的特征(Grain-based)

基于颗粒形状的特征包括基于图像的和基于矿物颗粒的特征两部分.基于图像的特征提取自整张砂岩显微图像,包括二值化阈值、Canny 边界像素数目、周长像素数目、白色区域数目以及白色区域面积等.基于矿物颗粒的图像特征在图像分割基础上计算得到.本文选取最大的 3 个矿物颗粒分别计算颗粒面积、凸包差、紧致度和外接特性 4 个指标.基于颗粒形状的图像特征维度为 17.

为探究不同类型的特征在砂岩显微图像自动分类中的预测性能,本文基于不同特征,结合支持向量机<sup>[42]</sup>和逻辑回归<sup>[43]</sup>分类模型进行实验,对比实验结果.

### 3.3.2 面向小规模数据集的深度自学习特征

针对研究问题 2,本文选取 3 种常用的面向小规模数据集的、基于卷积神经网络的特征表示方法提取砂岩图像特征,结合支持向量机<sup>[42]</sup>和逻辑回归<sup>[43]</sup>分类模型进行实验,比较实验结果.

- (1) Gen-aug:通过对砂岩显微图像进行尺度变换、缩放、水平和垂直翻转等通用的数据扩增操作,扩充数据集.基于扩充后的数据集,训练 FeRNet 网络,提取砂岩显微图像特征.
- (2) VGG-ft:以 ImageNet 数据集上预训练的 VGG16 网络为基础,将原 VGG16 模型中的全连接层替换为本文 FeRNet 网络的全连接层,采用层冻结方法,微调(fine-tuning)最后一个卷积模块(convolution block)及全连接层的参数,取第 2 个全连接层的输出作为砂岩显微图像的特征表示.
- (3) VGG-ft-gen:在 VGG-ft 基础上,采用尺度变换、缩放、水平和垂直翻转等通用数据扩增操作,扩充数据集.

### 3.3.3 分类模型

针对研究问题 3,本文选取机器学习中经典的分类算法来构建分类模型,探究 FeRNet 特征与不同分类模型结合的预测性能.实验中采用的分类模型包括线性判别分析(LDA)<sup>[44]</sup>、决策树(DT)<sup>[45]</sup>、朴素贝叶斯(NB)<sup>[46]</sup>、K 近邻(KNN)<sup>[47]</sup>、逻辑回归(LR)<sup>[43]</sup>、多层感知机(MLP)<sup>[48]</sup>、随机森林(RF)<sup>[49]</sup>、AdaBoost<sup>[50]</sup>和支持向量机(SVM)<sup>[42]</sup>.其中,线性判别分析、逻辑回归是线性模型代表,决策树是决策模型代表,朴素贝叶斯是贝叶斯理论模型代表,K 近邻是懒惰学习模型代表,多层感知机是人工神经网络的代表,随机森林和 AdaBoost 是集成学习模型代表,支持向量机是基于核函数模型代表.以上模型几乎涵盖机器学习中监督学习模型的各个类型,尽可能保证实验结果的有效性和全面性.

上述实验中,只有在 FeRNet 网络训练时采用第 3.1 节所述 CNN 训练集,其余实验均采用验证数据集,两种数据集不交叉使用.实验 1 和实验 2 分别计算单个类型砂岩显微图像的精确度和 F1 评测指标,实验 3 计算类别比例加权的指标.所有实验采用 10 折交叉验证计算结果.为减小随机性,每个实验重复 10 次,评测指标计算均值得到.

## 4 实验结果

本节整理并分析实验结果以回答实验设计问题.首先,采用 FeRNet 网络提取砂岩显微图像的特征,基于支持向量机和逻辑回归分类模型进行实验,并与人工定义特征和常见的面向小规模数据集的深度自学习特征进行对比,探究 FeRNet 特征对砂岩显微图像的代表能力.然后,将砂岩显微图像的特征与不同分类模型结合进行实验,比较实验结果,探究哪种分类模型能获得更好的预测性能.

### 4.1 与人工定义特征对比

为探究与人工定义特征相比,FeRNet 网络提取的特征在砂岩显微图像自动分类中的预测性能,本文提取不同类型的人工定义特征,包括颜色、纹理和颗粒形状等,结合支持向量机和逻辑回归分类模型分别进行实验,比较实验结果.表 4 和表 5 分别列出使用支持向量机和逻辑回归作为分类模型时,不同特征对 3 种类型砂岩显微图像的预测性能.

**Table 4** Prediction performance (Mean±Std) of different features based on SVM classification model**表 4** 基于支持向量机分类模型,不同特征的预测性能(均值±标准差)

特征	精确度			F1		
	长石砂岩	岩屑砂岩	石英砂岩	长石砂岩	岩屑砂岩	石英砂岩
RGB	0.783±0.080	0.907±0.006	0.825±0.008	0.602±0.139	0.921±0.029	0.886±0.052
HSV	0.866±0.172	0.905±0.052	0.845±0.053	0.653±0.126	0.927±0.023	0.897±0.025
Gray	0.616±0.169	0.880±0.055	0.630±0.125	0.451±0.069	0.889±0.032	0.687±0.084
GLCM	0.550±0.158	0.742±0.073	0.540±0.115	0.370±0.091	0.791±0.048	0.544±0.067
Wavelet	0.716±0.158	0.922±0.032	0.765±0.108	0.683±0.142	0.911±0.017	0.793±0.056
Grain-based	0.166±0.223	0.825±0.048	0.700±0.094	0.142±0.167	0.868±0.030	0.667±0.091
FeRNet	<b>0.916±0.088</b>	<b>0.938±0.035</b>	<b>0.870±0.075</b>	<b>0.805±0.135</b>	<b>0.948±0.015</b>	<b>0.899±0.042</b>

**Table 5** Prediction performance (Mean±Std) of different features based on LR classification model**表 5** 基于逻辑回归分类模型,不同特征的预测性能(均值±标准差)

特征	精确度			F1		
	长石砂岩	岩屑砂岩	石英砂岩	长石砂岩	岩屑砂岩	石英砂岩
RGB	0.850±0.183	0.912±0.024	0.810±0.077	0.678±0.146	0.926±0.032	0.847±0.061
HSV	0.850±0.145	0.885±0.057	0.830±0.054	0.615±0.102	0.915±0.035	0.870±0.044
Gray	0.666±0.176	0.882±0.029	0.610±0.074	0.435±0.071	0.901±0.027	0.687±0.053
GLCM	0.633±0.153	0.710±0.105	0.470±0.132	0.349±0.052	0.775±0.075	0.497±0.111
Wavelet	0.733±0.179	0.857±0.057	0.770±0.082	0.585±0.091	0.879±0.033	0.788±0.071
Grain-based	0.600±0.196	0.825±0.060	0.560±0.107	0.390±0.125	0.863±0.037	0.601±0.095
FeRNet	<b>0.867±0.131</b>	<b>0.933±0.044</b>	<b>0.880±0.042</b>	<b>0.781±0.077</b>	<b>0.940±0.031</b>	<b>0.881±0.038</b>

从表 4 可见,当使用支持向量机作为分类模型时,相比人工定义图像特征,本文 FeRNet 网络提取的特征在所有砂岩类型及评测指标中均取得最好的预测表现。在人工定义特征中,HSV 颜色特征和基于小波变换的纹理特征预测表现最佳。具体而言,当使用精确度作为评测指标时,HSV 颜色特征在长石砂岩和石英砂岩类型预测表现最好,基于小波变换的纹理特征在岩屑砂岩类型获得最高预测性能;当使用 F1 值作为评测指标时,基于小波变换的纹理特征在长石砂岩类型预测表现最好,HSV 颜色特征在岩屑砂岩和石英砂岩类型获得最高预测性能。然而,无论使用哪种评测指标,FeRNet 网络提取的图像特征在 3 种类型砂岩的预测表现均优于上述最佳人工定义特征。特别是在长石砂岩类型,预测性能获得显著提升。相对于人工定义特征获得的最佳性能指标,FeRNet 网络提取的图像特征在精确度和 F1 评测指标中分别获得 5%和 12.2%的性能提升。

从表 5 可见,当使用逻辑回归作为分类模型时,相比人工定义图像特征,本文 FeRNet 网络提取的特征在所有砂岩类型及评测指标中同样取得最好的预测表现。在人工定义特征中,RGB 和 HSV 两种颜色特征预测表现最佳。具体而言,当使用精确度作为评测指标时,RGB 颜色特征在长石砂岩和岩屑砂岩类型中预测表现最好,HSV 颜色特征在长石砂岩和石英砂岩类型获得最高预测性能;当使用 F1 值作为评测指标时,RGB 颜色特征仍在长石砂岩和岩屑砂岩类型中预测表现最好,HSV 颜色特征在石英砂岩类型获得最高预测性能。然而,与基于 SVM 分类模型的结果一致,FeRNet 网络提取的图像特征在 3 种类型砂岩的预测表现均明显优于最佳人工定义特征。值得说明的是,在使用其他分类模型时,相比人工定义图像特征,本文 FeRNet 网络提取的特征始终表现出最佳预测性能。

基于以上分析可得:与人工定义图像特征相比,FeRNet 网络提取的图像特征在砂岩显微图像自动分类中能够获得最好的预测性能。由此表明,在带标记砂岩显微图像数量不足的情况下,本文提出的 FeRNet 网络能够有效地提取砂岩显微图像的高级语义特征,为分类模型的训练和预测提供更具判别性的信息,提高模型的预测性能。

#### 4.2 与面向小规模数据集的深度自学习特征对比

为了探究与面向小规模数据集的深度自学习特征相比,FeRNet 网络提取的特征在砂岩显微图像自动分类中的预测性能,本文基于 3 种常见的小规模数据集图像特征表示方法,结合支持向量机和逻辑回归分类模型分别进行实验,比较实验结果。表 6 和表 7 分别列出使用支持向量机和逻辑回归作为分类模型时,不同特征对 3 种

类型砂岩显微图像的预测性能.

**Table 6** Prediction performance (Mean±Std) of different deep self-taught features based on SVM classification model

**表 6** 基于支持向量机分类模型,不同深度自学习特征的预测性能(均值±标准差)

特征	精确度			F1		
	长石砂岩	岩屑砂岩	石英砂岩	长石砂岩	岩屑砂岩	石英砂岩
Gen-aug	0.633±0.221	0.925±0.033	0.845±0.116	0.605±0.179	0.918±0.029	0.877±0.089
VGG-ft	0.351±0.148	0.834±0.051	0.658±0.058	0.254±0.097	0.858±0.013	0.680±0.034
VGG-ft-gen	0.348±0.197	0.848±0.058	0.682±0.070	0.255±0.125	0.865±0.015	0.701±0.051
FeRNet	<b>0.916±0.088</b>	<b>0.938±0.035</b>	<b>0.870±0.075</b>	<b>0.805±0.135</b>	<b>0.948±0.015</b>	<b>0.899±0.042</b>

**Table 7** Prediction performance (Mean±Std) of different deep self-taught features based on LR classification model

**表 7** 基于逻辑回归分类模型,不同深度自学习特征的预测性能(均值±标准差)

特征	精确度			F1		
	长石砂岩	岩屑砂岩	石英砂岩	长石砂岩	岩屑砂岩	石英砂岩
Gen-aug	0.600±0.211	0.893±0.041	0.860±0.052	0.531±0.184	0.904±0.027	0.869±0.032
VGG-ft	0.416±0.072	0.808±0.042	0.668±0.072	0.256±0.078	0.845±0.031	0.679±0.049
VGG-ft-gen	0.393±0.088	0.796±0.047	0.694±0.079	0.301±0.073	0.842±0.022	0.688±0.058
FeRNet	<b>0.867±0.131</b>	<b>0.933±0.044</b>	<b>0.880±0.042</b>	<b>0.781±0.077</b>	<b>0.940±0.031</b>	<b>0.881±0.038</b>

从表 6 可见,当使用支持向量机作为分类模型时,相比其他面向小规模数据集的深度自学习网络提取的图像特征,本文 FeRNet 网络提取的砂岩显微图像特征在所有砂岩类型及评测指标中均取得最好的预测表现.在所有深度自学习特征中,VGG-ft 方法提取的图像特征预测表现最差,远低于本文 FeRNet 网络提取的特征,说明微调预训练网络参数的方法并不一定适用于所有图像.结合了平移、旋转等通用数据扩增方法后,VGG-ft-gen 方法提取的特征预测性能提高 1%~2%.采用平移、旋转变换等通用数据扩增处理的 FeRNet 网络(Gen-aug)提取的特征在其他深度自学习特征中预测表现最佳.然而,无论使用哪种评测指标,采用本文图像扩增处理的 FeRNet 网络提取的图像特征在 3 种类型砂岩的预测表现均优于通用数据扩增方法(Gen-aug).由此表明,相比平移、旋转等通用数据扩增操作,本文提出的图像扩增方法更能保留图像原始的信息,有助于砂岩显微图像的特征表示.

从表 7 可见,当使用逻辑回归作为分类模型时,相比其他面向小规模数据集的深度自学习网络提取的图像特征,本文 FeRNet 网络提取的砂岩显微图像特征在所有砂岩类型及评测指标中同样取得最好的预测表现.在其他深度自学习特征中,Gen-aug 方法提取的特征取得最优预测性能.VGG-ft 和 VGG-ft-gen 方法提取的特征预测表现相近,分别在不同砂岩类型获得高于对方的评测指标.与基于 SVM 分类模型的结果一致,FeRNet 网络提取的图像特征在 3 种类型砂岩的预测表现均明显优于其他深度自学习特征.

基于以上分析可得:与其他面向小规模数据集的深度自学习图像特征相比,FeRNet 网络提取的图像特征在砂岩显微图像自动分类中能够获得最好的预测性能.由此表明,(1) 由于与 ImageNet 数据集中图像存在差异,微调预训练深度网络的部分参数不能有效增强其提取砂岩显微图像语义特征的性能;(2) 与平移、旋转等通用数据扩增方法相比,本文提出的基于滑动窗口的图像扩增预处理方法,能够保留更多的原始图像信息,有助于砂岩显微图像的特征表示.

#### 4.3 不同分类模型预测性能

为了探究 FeRNet 网络提取的特征在使用何种分类模型时能够取得更好的预测性能,本文将 FeRNet 特征分别与线性判别分析、决策树等经典分类模型结合进行实验,比较实验结果.图 6 列出基于不同分类模型,FeRNet 特征的预测性能,图中实心菱形表示均值.

如图 6(a)所示,当使用精确度作为评测指标时,FeRNet 特征具有一定的普适性,与各分类模型结合均可获得理想的预测表现.相比其他分类模型,朴素贝叶斯和随机森林两种模型在交叉检验中评测指标的方差较小,说明其具有更稳定的预测表现.但是,从评测指标的中值和均值角度来看,K 近邻和支持向量机模型比其他模型具有

更高的预测表现.实验一中,逻辑回归模型的预测性能在所有模型中表现并不突出,但相比人工定义图像特征,FeRNet 特征仍取得更高的预测性能,进一步证明 FeRNet 特征对砂岩显微图像的表达力.

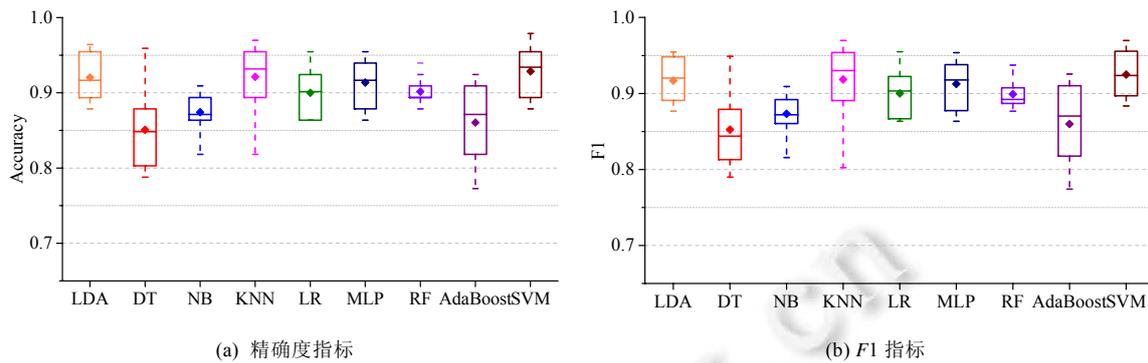


Fig.6 Prediction performance of FeRNet features based on different classification models

图 6 基于不同分类模型,FeRNet 特征的预测性能

如图 6(b)所示,当使用  $F1$  值作为评测指标时,FeRNet 特征与各分类模型结合仍可获得理想的预测表现.相比其他分类模型,朴素贝叶斯和随机森林模型同样表现出更稳定的预测性能,但随机森林模型的稳定性有轻微下降.从预测指标的中值和均值角度来看, $K$  近邻和支持向量机模型相比其他模型仍具有更高的预测表现.但是相对于精确度指标,两种分类模型的  $F1$  指标均表现出一定程度的下降.

基于以上分析,FeRNet 特征与各种分类模型结合均可获得理想的砂岩显微图像预测表现,表明 FeRNet 特征具有良好的普适性.在使用朴素贝叶斯和随机森林分类模型时,FeRNet 网络提取的砂岩显微图像特征可取得更稳定的预测性能;在使用  $K$  近邻和支持向量机作为分类模型时,FeRNet 特征在砂岩显微图像自动分类中可取得更高的平均预测表现.

## 5 方法讨论

如第 2 节所述,针对带标记砂岩显微图像数量不足问题,本文提出基于滑动窗口的图像扩增预处理方法和基于卷积自编码网络的权重初始化策略,改善 FeRNet 网络的训练效果,防止过拟合.本节探究两种策略及不同激活函数的选择对 FeRNet 网络训练的影响.

### 5.1 图像扩增对 FeRNet 网络训练的影响

为了探究图像扩增预处理对 FeRNet 网络训练的影响,本文基于未进行图像扩增预处理的 CNN 训练集训练 FeRNet 网络(简记为 FeRNet-),采用该网络提取验证数据集中砂岩显微图像的特征,结合支持向量机和逻辑回归模型进行实验,与使用图像扩增预处理的 FeRNet 网络的预测性能比较.图 7 列出两种网络提取的特征在不同类型砂岩的预测表现.

如图 7 所示,与基于未进行图像扩增预处理的 CNN 训练集训练得到的 FeRNet-网络相比,使用图像扩增预处理的 FeRNet 网络提取的特征在砂岩显微图像自动分类中始终具有更好的预测性能.如图 7(a)和图 7(b)所示,当使用支持向量机作为分类模型时,相比未使用图像扩增预处理的 FeRNet-网络,使用图像扩增预处理后,FeRNet 网络提取的特征在 3 种类型砂岩的预测性能均获得提升.特别是在原始数据量较少的长石砂岩和石英砂岩类型,预测性能提升显著.如图 7(c)和图 7(d)所示,当使用逻辑回归作为分类模型时,与 FeRNet-网络相比,FeRNet 网络提取的特征在 3 种类型砂岩的预测性能同样获得显著提升.

基于以上分析,对带标记数据较少的砂岩显微图像数据集,本文提出的图像扩增预处理能够显著改善 FeRNet 网络的训练效果.在数据量较少的砂岩类型中,图像扩增处理的作用更加明显.图像扩增预处理在不改变原始砂岩显微图像语义信息的基础上,通过对原图像的裁剪操作,可实现数据量的倍数级增长,改善网络的训

练效果,防止过拟合.值得注意的是,本文提出的图像扩增预处理方法可扩展至具有相似特性的显微图像,如其他类型岩石显微图像、医学显微图像等的特征表示中.

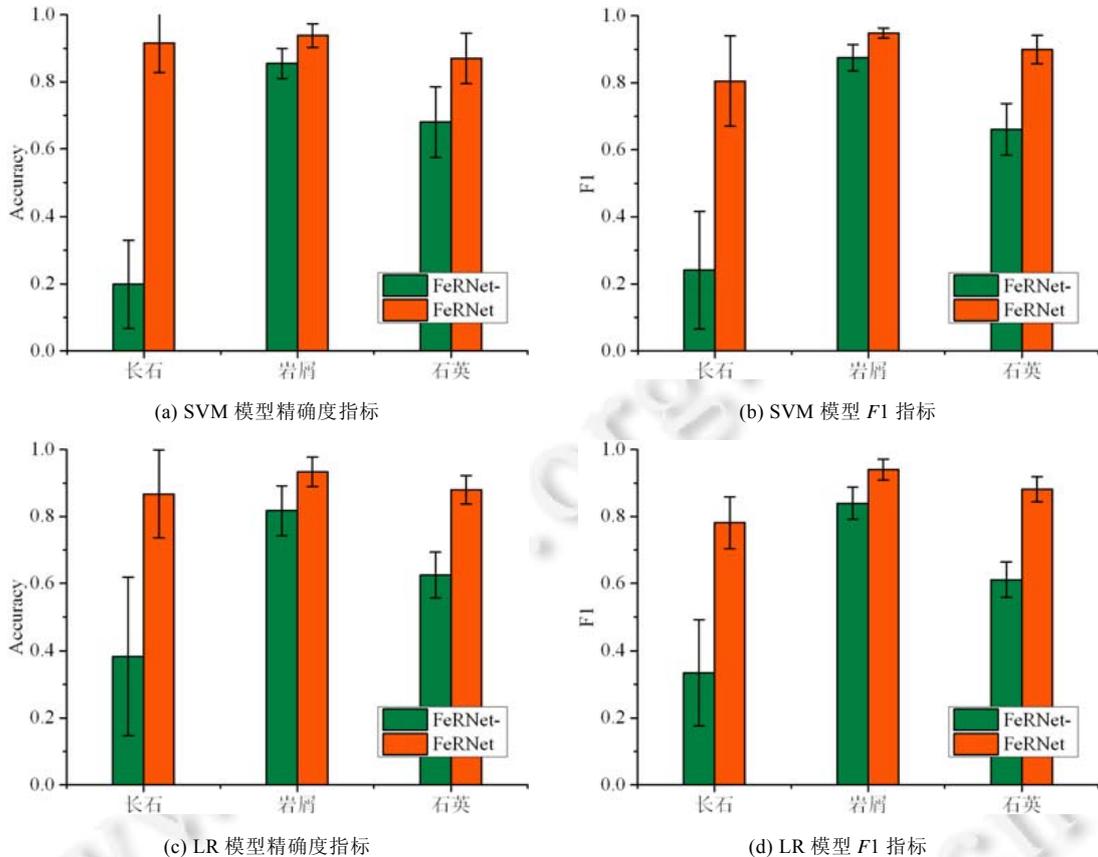


Fig.7 Comparison of prediction performance of FeRNet features with and without image augmentation

图7 使用和不使用图像扩增情况下,FeRNet 特征预测性能对比

## 5.2 卷积自编码网络对FeRNet网络训练的影响

为了探究卷积自编码网络对 FeRNet 网络训练的影响,本文基于随机初始化的权重训练 FeRNet 网络(简称 FeRNet-nCAE),采用该网络提取验证数据集中砂岩显微图像的特征,结合第 3.3.3 节所述的分类模型进行实验,与基于卷积自编码网络编码器权重训练得到的 FeRNet 网络的预测性能比较.图 8 列出两种网络提取的特征在不同分类模型的预测表现.

如图 8 所示,与基于随机初始化的权重训练得到的 FeRNet-nCAE 网络相比,基于卷积自编码网络编码器权重训练得到的 FeRNet 网络提取的特征在砂岩显微图像自动分类中始终具有更好的预测性能.如图 8(a)所示,当使用精确度作为评测指标时,与随机初始化得到的 FeRNet-nCAE 网络相比,基于卷积自编码网络权重得到的 FeRNet 网络提取的特征在所有分类模型的预测性能均获得提升.特别是在朴素贝叶斯模型中,相比 FeRNet-nCAE 特征,FeRNet 特征的预测性能提高 8.64%.如图 8(b)所示,当使用 F1 值作为评测指标时,与使用精确度指标结果一致.与随机初始化得到的 FeRNet-nCAE 网络相比,基于卷积自编码网络权重得到的 FeRNet 网络提取的特征预测性能同样获得显著提升.

基于以上分析,卷积自编码网络可有效改善 FeRNet 网络的训练效果.卷积自编码网络可初步编码砂岩显微图像的语义信息,为 FeRNet 网络的训练提供基础.与随机初始化权重方式相比,使用卷积自编码网络编码器权

重初始化的 FeRNet 网络在训练开始时已具有一定的砂岩显微图像语义信息,有助于 FeRNet 网络快速准确地找到最优的权重,提高 FeRNet 网络对砂岩显微图像的特征表示能力.基于卷积自编码网络的权重初始化策略亦可扩展至带标记图像较少的图像分类应用中.

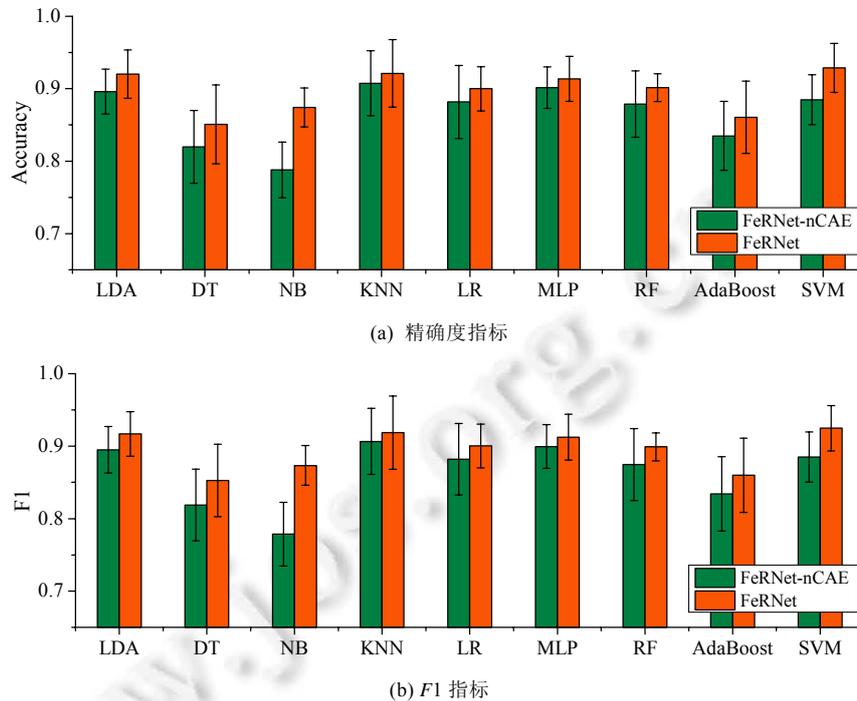


Fig.8 Comparison of prediction performance of FeRNet features with and without FeRCAE

图 8 使用和不使用 FeRCAE 情况下,FeRNet 特征预测性能对比

### 5.3 不同激活函数对 FeRNet 网络训练的影响

为了探究不同激活函数的选择对 FeRNet 网络训练的影响,本文分别选取 Sigmoid、Tanh 和 ReLu 函数作为 FeRCAE 和 FeRNet 网络的激活函数,采用小批量梯度下降(mini-batch gradient descent,简称 MBGC)算法对网络进行训练.基于训练得到的 3 种 FeRNet 网络,分别从对砂岩显微图像的特征表示能力和网络收敛速度两方面对比 Sigmoid、Tanh 和 ReLu 函数对网络训练的影响.表 8 和图 9 分别列出采用不同激活函数时,FeRNet 网络特征的预测性能及收敛速度.

**Table 8** Prediction performance (Mean±Std) of FeRNet features based on different activation functions

**表 8** 基于不同激活函数,FeRNet 特征的预测性能(均值±标准差)

激活函数	SVM		LR	
	精确度	F1	精确度	F1
Sigmoid	0.822±0.019	0.834±0.018	0.810±0.019	0.823±0.020
Tanh	0.731±0.024	0.788±0.034	0.717±0.026	0.789±0.044
ReLu	<b>0.929±0.034</b>	<b>0.925±0.031</b>	<b>0.900±0.031</b>	<b>0.905±0.030</b>

如表 8 所示,无论使用何种分类模型和评测指标,与 Sigmoid 函数和 Tanh 函数相比,使用 ReLu 函数作为激活函数的 FeRNet 网络提取的特征在砂岩显微图像自动分类中始终具有更好的预测性能.如图 9 所示,在使用小批量梯度下降算法对 FeRNet 网络进行优化时,与 Sigmoid 函数和 Tanh 函数相比,采用 ReLu 函数具有更快的收敛速度和更优质的解.从图中可见,在使用 Sigmoid 函数和 Tanh 函数时,FeRNet 网络收敛速度相对缓慢,且损失值在 0.4~0.5 左右时即趋于稳定,不再降低;而使用 ReLu 作为激活函数时,FeRNet 网络的训练损失在开始优化时

即快速减小,最终在 0.2 左右时趋于稳定.

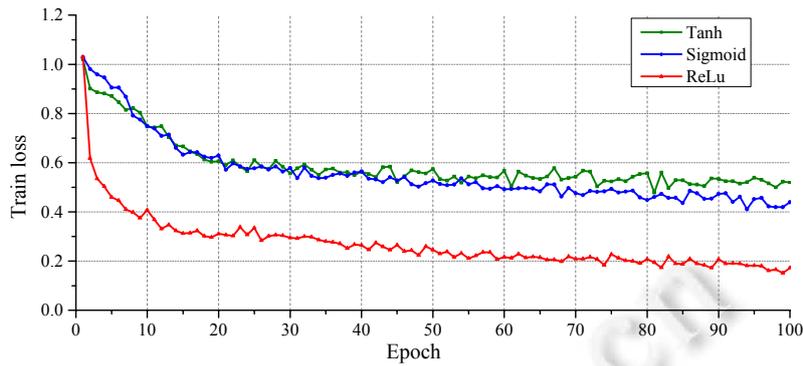


Fig.9 Comparison of coverage speed of FeRNet network using different activation functions

图 9 使用不同激活函数,FeRNet 网络收敛速度对比

基于以上分析,与 Sigmoid 函数和 Tanh 函数相比,采用 ReLu 函数作为 FeRNet 网络的激活函数更有助于网络的训练:一方面,采用 ReLu 函数可提高 FeRNet 网络对砂岩显微图像的特征表示能力,获得更优的预测性能;另一方面,在使用小批量梯度下降算法对 FeRNet 网络进行优化时,ReLu 函数具有更快的收敛速度,有助于提高网络的训练效果.因此,本文 FeRNet 网络选择 ReLu 函数作为激活函数.

## 6 总结和展望

本文提出一种面向小规模数据集的卷积神经网络 FeRNet,用于砂岩显微图像的特征表示,提高砂岩显微图像的自动分类准确性.FeRNet 网络具有相对简单的网络结构,在保证对砂岩显微图像表示能力的前提下,尽可能地减少网络参数,降低对带标记砂岩显微图像数据量的要求,防止参数过拟合.另外,针对带标记砂岩显微图像数量不足的问题,本文分别从图像预处理和 FeRNet 网络训练两方面提出基于滑动窗口的图像扩增和基于卷积自编码网络的权重初始化解策略.一方面,根据砂岩显微图像裁剪语义不变特性,采用滑动窗口方式对原砂岩显微图像进行大量裁剪,使带标记图像数量倍数级增加;另一方面,使用预训练的卷积自编码网络 FeRCAE 的编码器权重来初始化 FeRNet 网络,改善 FeRNet 网络的训练效果,防止过拟合.基于采自西藏地区的砂岩显微图像数据集,本文设计并进行了实验.实验结果表明,图像扩增和卷积自编码网络能够有效改善 FeRNet 网络的训练效果;与人工定义图像特征及其他面向小规模数据集的深度自学习特征相比,本文的 FeRNet 网络能够提取砂岩显微图像的语义信息,对砂岩显微图像具有更好的表示能力.

值得注意的是,本文提出的框架对其他相似显微图像的特征表示具有可扩展性,如其他类型岩石显微图像、医学显微图像等.一方面,此类显微图像也很难采集到大量带标记样本,不能满足模型训练的数据需求;另一方面,与本文砂岩显微图像相似,此类图像中感兴趣的目标对象遍布整张显微图像,也具有裁剪语义不变性.对于这些显微图像的特征表示,同样可通过图像扩增及预编码语义信息的方式改善特征表示神经网络的训练效果,防止过拟合.

另外,本文方法还有一些工作值得扩展:首先,采集更多的砂岩显微图像,扩充数据集,进一步提高 FeRNet 网络对砂岩显微图像的特征表示能力;其次,与更多人工定义特征及深度自学习特征对比,探究 FeRNet 网络在砂岩显微图像特征表示中的稳定性;最后,将 FeRNet 网络应用于其他图像的特征表示,探究其是否仍具有良好的特征表示能力.

**References:**

- [1] Albar A, Osman MH, Abdullah MS, Ismail BN. Classification of intrusive igneous rocks using digital image processing: A binary approach. *Journal of Engineering Science*, 2013,9:11–19.
- [2] Młynarczyk M, G Rszczyk A, Ślipek B. The application of pattern recognition in the automatic classification of microscopic rock images. *Computers & Geosciences*, 2013,60:126–133.
- [3] Gonzalez RC, Woods RE, Eddins SL. *Digital Image Processing Using MATLAB*. Publishing House of Elec, 2013.
- [4] Chatterjee S. Vision-based rock-type classification of limestone using multi-class support vector machine. *Applied Intelligence*, 2013,39(1):14–27.
- [5] Li N, Hao H, Gu Q, Wang D, Hu X. A transfer learning method for automatic identification of sandstone microscopic images. *Computers & Geosciences*, 2017,103:111–121.
- [6] Chatterjee S, Bhattacharjee A. Genetic algorithms for feature selection of image analysis-based quality monitoring model: An application to an iron mine. *Engineering Applications of Artificial Intelligence*, 2011,24(5):786–795.
- [7] Arivazhagan S, Ganesan L. Texture classification using wavelet transform. *Pattern Recognition Letters*, 2003,24(9):1513–1521.
- [8] Choy SK, Tong CS. Statistical wavelet subband characterization based on generalized gamma density and its application in texture retrieval. *IEEE Trans. on Image Processing*, 2010,19(2):281–289.
- [9] Sun CN, Yi Q, Cui L. Application of wavelet transform and fuzzy clustering in thermopaint color image segmentation. *Ruan Jian Xue Bao/Journal of Software*, 2012,23:64–68 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/12025.htm>
- [10] Vargas JF, Ferrer MA, Travieso CM, Alonso JB. Off-line signature verification based on grey level information using texture features. *Pattern Recognition*, 2011,44(2):375–385.
- [11] Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. *IEEE Trans. on Systems, Man and Cybernetics*, 1973,SMC-3(6):610–621.
- [12] Ojala T, Valkealahti K, Oja E, Pietik Inen M. Texture discrimination with multidimensional distributions of signed gray level differences. *Pattern Recognition*, 2001,34(3):727–739.
- [13] Feng J, Qing G, Huizhen H, Na L. Feature extraction and grain segmentation of sandstone images based on convolutional neural networks. In: *Proc. of the Int'l Conf. on Pattern Recognition*. 2018. 2636–2641.
- [14] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc. of the IEEE*, 1998,86(11):2278–2324.
- [15] Singh N, Singh T, Tiwary A, Sarkar KM. Textural identification of basaltic rock mass using image processing and neural network. *Computational Geosciences*, 2010,14(2):301–310.
- [16] Jiang W, Lam KM, Shen TZ. Efficient edge detection using simplified Gabor wavelets. *IEEE Trans. Syst. Man Cybern. Part B*, 2009,39(4):1036–1047.
- [17] Lowe DG. Object recognition from local scale-invariant features. In: *Proc. of the 7th IEEE Int'l Conf. on Computer Vision*. 1999. 1150–1157.
- [18] Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up robust features. In: *Proc. of the European Conf. on Computer Vision*. Berlin: Springer-Verlag, 2006. 404–417.
- [19] Zhou SR, Yin JP. LBP texture feature based on Haar characteristics. *Ruan Jian Xue Bao/Journal of Software*, 2013,24(8):1909–1926 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4277.htm> [doi: 10.3724/SP.J.1001.2013.04277]
- [20] Sun J, He FZ, Chen X, Yuan ZY. Efficient multi-scale texture recognition algorithm. *Ruan Jian Xue Bao/Journal of Software*, 2014, 25:278–289 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4277.htm> [doi: 10.3724/SP.J.1001.2013.04277]
- [21] Wang H, Fu G, Cai Y, Wang S. Multiple feature fusion based image classification using a non-biased multi-scale kernel machine. In: *Proc. of the Int'l Conf. on Fuzzy Systems and Knowledge Discovery*. IEEE, 2016. 700–704.
- [22] Liu T, Zhou XC, Yan XJ. Texture similarity calculation with LBP features fused by multi-channel and multi-mode. *Application Research of Computers*, 2019,36(1):1–8 (in Chinese with English abstract).
- [23] Isaza C, Anaya K, De Paz JZ, Vasco-Leal JF, Hernandez-Rios I, Mosquera-Artamonov JD. Image analysis and data mining techniques for classification of morphological and color features for seeds of the wild castor oil plant (*Ricinus communis L.*). *Multimedia Tools and Applications*, 2018,77(2):2593–2610.

- [24] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Proc. of the Int'l Conf. on Neural Information Processing Systems. 2012. 1097–1105.
- [25] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 779–788.
- [26] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. 1–9.
- [27] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 770–778.
- [28] Amrouch M, Rabi M, Es-Saady Y. Convolutional feature learning and cnn based hmm for Arabic handwriting recognition. In: Proc. of the Int'l Conf. on Image and Signal Processing. Cham: Springer, 2018. 265–274.
- [29] Bai C, Huang L, Chen JN, Pan X, Chen SY. Optimization of deep convolutional neural network for large scale image classification. Ruan Jian Xue Bao/Journal of Software, 2018,29(4):1029–1038 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5404.htm> [doi: 10.13328/j.cnki.jos.005404]
- [30] Lin M, Chen Q, Yan S. Network in network. arXiv preprint arXiv:13124400, 2013.
- [31] Perez L, Wang J. The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621, 2017.
- [32] Sharif Razavian A, Azizpour H, Sullivan J, Carlsson S. CNN features off-the-shelf: an astounding baseline for recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops. 2014. 806–813.
- [33] Oquab M, Bottou L, Laptev I, Sivic J. Learning and transferring mid-level image representations using convolutional neural networks. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2014. 1717–1724.
- [34] Donahue J, Jia Y, Vinyals O, Hoffman J, Zhang N, Tzeng E, Darrell T. Decaf: A deep convolutional activation feature for generic visual recognition. In: Proc. of the Int'l Conf. on Machine Learning. 2014. 647–655.
- [35] Duan M, Wang GP, Niu CY. Method of small sample size image recognition based on convolution neural network. Computer Engineering and Design, 2018,39(1):224–229 (in Chinese with English abstract).
- [36] Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? In: Proc. of the Advances in Neural Information Processing Systems. 2014. 3320–3328.
- [37] Bishop CM, Nasrabadi NM. Pattern Recognition and Machine Learning. Cambridge: Academic Press, 2006.
- [38] Shewhart WA, Wilks SS. Applied Logistic Regression. 2nd ed., New York: John Wiley & Sons, 2005.
- [39] Masci J, Meier U, Cireşan D, Schmidhuber J. Stacked convolutional auto-encoders for hierarchical feature extraction. In: Proc. of the Int'l Conf. on Artificial Neural Networks. Berlin, Heidelberg: Springer-Verlag, 2011. 52–59.
- [40] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: Proc. of the European Conf. on Computer Vision. Cham: Springer-Verlag, 2014. 818–833.
- [41] Zeiler MD, Krishnan D, Taylor GW, Fergus R. Deconvolutional networks. In: Proc. of the Computer Vision and Pattern Recognition. San Francisco, 2010. 2528–2535.
- [42] Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intelligent Systems and Their Applications, 1998,13(4):18–28.
- [43] Walker SH, Duncan DB. Estimation of the probability of an event as a function of several independent variables. Biometrika, 1967, 54(1-2):167–179.
- [44] Fisher RA. The use of multiple measurements in taxonomic problems. Annals of Human Genetics, 2012,7(2):179–188.
- [45] Swain PH, Hauska H. The decision tree classifier: Design and potential. IEEE Trans. on Geoscience Electronics, 1977,15(3): 142–147.
- [46] Murphy KP. Naive Bayes Classifiers. University of British Columbia, 2006. 1–8.
- [47] Cover TM, Hart PE. Nearest neighbor pattern classification. IEEE Trans. on Information Theory, 1967,13(1):21–27.
- [48] Zhou ZH. Machine Learning. Beijing: Tsinghua University Press, 2016. 98–101 (in Chinese).
- [49] Ho TK. Random decision forests. In: Proc. of 3rd Int'l Conf. on Document Analysis and Recognition. Montreal: IEEE, 1995. 278–282.

- [50] Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 1997,55(1):119–139.

#### 附中文参考文献:

- [9] 孙超男,易芹,崔丽.小波变换结合模糊聚类在示温漆彩色图像分割中的应用. *软件学报*,2012,23:64–68. <http://www.jos.org.cn/1000-9825/12025.htm>
- [19] 周书仁,殷建平.基于 Haar 特性的 LBP 纹理特征. *软件学报*,2013,24(8):1909–1926. <http://www.jos.org.cn/1000-9825/4277.htm> [doi: 10.3724/SP.J.1001.2013.04277]
- [20] 孙俊,何发智,陈晓,袁志勇.一种有效的多尺度纹理识别方法. *软件学报*,2014,25:278–289. <http://www.jos.org.cn/1000-9825/4277.htm> [doi: 10.3724/SP.J.1001.2013.04277]
- [22] 刘涛,周先春,严锡君.多通道多模式融合 LBP 特征的纹理相似度计算. *计算机应用研究*,2019,36(1):1–8.
- [29] 白琮,黄玲,陈佳楠,潘翔,陈胜勇.面向大规模图像分类的深度卷积神经网络优化. *软件学报*,2018,29(4):1029–1038. <http://www.jos.org.cn/1000-9825/5404.htm> [doi: 10.13328/j.cnki.jos.005404]
- [35] 段萌,王功鹏,牛常勇.基于卷积神经网络的小样本图像识别方法. *计算机工程与设计*,2018,39(1):224–229.
- [48] 周志华. *机器学习*.北京:清华大学出版社,2016.98–101.



李娜(1991—),女,博士生,主要研究领域为机器学习.



郝慧珍(1974—),女,博士生,讲师,主要研究领域为模式识别.



顾庆(1972—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为软件质量保障,分布式计算.



于华(1995—),男,硕士,主要研究领域为机器学习.



姜枫(1980—),男,博士,副教授,CCF 专业会员,主要研究领域为机器学习,计算机视觉.



倪超(1990—),男,博士,CCF 专业会员,主要研究领域为机器学习,软件缺陷预测.