Computers & Geosciences 185 (2024) 105560

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

RockS^{2}Net: Rock image classification via a spatial localization siamese network

Zhu Qiqi [a], Wang Sai [a], Tong Shun [b], Yin Liangbin [a], Qi Kunlun [a], Guan Qingfeng [a,c,*]

[a] School of Geography and Information Engineering, China University of Geosciences, Wuhan, Hubei, China
[b] Xi'an Surveying and Mapping Station, Xi'an, 710054, China
[c] State Key Laboratory of Geological Processes and Mineral Resources, China Univesity of Geosciences, Wuhan, Hubei, China

ARTICLE INFO

Keywords:
Rock property
Convolutional neural networks (CNNs)
Spatial localization
Siamese network

ABSTRACT

The acquisition of rock property information is at the core of regional geological survey and mineral exploration, but hand-crafted feature-based methods are heavily influenced by human prior knowledge and have limited transferability. End-to-end deep learning techniques, exemplified by convolutional neural networks (CNNs), have attained significant accomplishments in the domain of image classification. However, previous end-to-end CNN-based methods are hard to focus on image critical areas, and they also cannot make full use of global alignment dependency of the rocks. In this paper, rock image classification via a RockS^{2}Net is proposed. The RockS2Net framework adopts a Siamese architecture, characterized by two branches that share parameters, enabling the efficient extraction of both global and local features. This approach facilitates the extraction of global features from entire images and focuses on critical areas to extract local features. The architecture of spatial transformer network (STN) is introduced to transform microscopic images of rock sections to their critical areas. By fusing local features and global features, the properties obtained from microscopic images of rock sections can be more accurately predicted. To test the robust generalizability of the proposed method, the constructed CHN-Rock images dataset is used for experiments and evaluation. Experimental results show that the accuracy of the proposed RockS^{2}Net on the CHN-Rock image dataset is 2–3% higher than that of other rock image classification networks.

1. Introduction

Rock formation is a testament to the dynamic evolution and transformation of Earth, encapsulating the intricate cycles of material processes. (Kuiper et al., 2008). As an important object of research in the field of geology, the rock is the carrier of the deposits (Cherkashina et al., 2014; Rollinson, 2014). Rock identification holds paramount significance in geological surveys, engineering explorations, and mineral prospecting endeavors. Accurate rock identification holds paramount significance across various domains, including geological surveys, engineering explorations, and mineral prospecting, as it represents a fundamental undertaking in these fields. (Chatterjee, 2013). Rocks can be named and classified according to the rock properties, such as mineral composition, content, and structure. These properties are reflected in the shape, color, and texture features of the rock section image, which are the basis for image classification (Zhu et al., 2018). For mining engineers and geologists, the rock property information is very significant for the successful mining of deposits (Patel et al., 2017; Perez et al., 2011). However, the structure of rocks is in-homogeneous and strongly directional (Lepistö et al., 2005a; Zhu et al., 2018). Hence, the precise extraction of rock property information stands as a pivotal task within the domain of geology. (Shang and Barnes, 2012).

The existing rock image classification methods can be categorized into two primary groups: handcrafted feature-based methods and deep learning methods. Handcrafted feature-based methods refer to the use of rock image color, texture, protrusion, particle shape and other characteristic parameters, using the classification method to predict the type of rock. Lepistö et al. (2005b) employed band-pass filtering techniques to process images in various color spaces, enabling the analysis of color texture images at multiple scales. Seng and Chen. (2009) applied the RS theory and the Support Vector Machine (SVM) model of machine learning to solve the problem of classifying rocks. Depending on how

* Corresponding author. School of Geography and Information Engineering, China University of Geosciences, Wuhan, Hubei, China.
E-mail addresses: zhuqq@cug.edu.cn (Z. Qiqi), 20171003123@cug.edu.cn (W. Sai), 572375405@qq.com (T. Shun), 1179660637@qq.com (Y. Liangbin), qikunlun@cug.edu.cn (Q. Kunlun), guanqf@cug.edu.cn (G. Qingfeng).

https://doi.org/10.1016/j.cageo.2024.105560
Received 5 June 2023; Received in revised form 22 November 2023; Accepted 5 February 2024
Available online 12 February 2024

Computers & Geosciences 185 (2024) 105560

Contents lists available at ScienceDirect

Computers and Geosciences

journal homepage: www.elsevier.com/locate/cageo

RockS ^{2} Net：基于空间定位孪生网络的岩石图像分类方法

朱琪琪 [a]、王赛 [a]、童顺 [b]、尹良斌 [a]、齐昆仑 [a]、关庆丰 [a,c,*]

[a] 中国地质大学（武汉）地理与信息工程学院，湖北武汉
[b] 西安测绘站，西安，710054
[c] 中国地质大学地质过程与矿产资源国家重点实验室，湖北武汉

文章信息

关键词：
岩石属性
卷积神经网络（CNNs）
空间定位
孪生网络

摘要

岩石属性信息的获取是区域地质调查与矿产勘查的核心环节，但基于人工特征的方法受先验知识影响显著且迁移性有限。以卷积神经网络(CNN)为代表的端到端深度学习技术在图像分类领域取得重大突破，但现有CNN方法难以聚焦图像关键区域，且无法充分利用岩石的全局对齐依赖性。本文提出基于RockS ^{2} Net的岩石图像分类方法，该框架采用参数共享的双分支孪生结构，可同步提取全局与局部特征：通过整图提取全局特征，同时聚焦关键区域获取局部特征。引入空间变换网络(STN)架构将岩石薄片显微图像转换至关键区域，通过融合局部与全局特征实现更精准的岩石薄片属性预测。为验证方法鲁棒性，采用自建CHN-Rock数据集进行实验评估，结果表明RockS ^{2} Net在CHN-Rock数据集上的分类准确率较其他岩石图像分类网络提升 $2-3\%$ 。

1. 引言

岩石形成是地球动态演化与物质转化过程的见证，蕴含着复杂的物质循环机制（Kuiper等，2008）。作为地质学研究的重要对象，岩石是矿床物质的载体（Cherkashina等，2014；Rollinson，2014）。岩石鉴定在地质调查、工程勘探和矿产勘查中具有首要意义，其精确识别是这些领域的基础性工作（Chatterjee，2013）。根据矿物成分、含量及结构等岩石属性，可对岩石进行命名与分类。这些属性通过岩石剖面图像的形态、色泽与纹理特征呈现，构成图像分类的依据（Zhu等，2018）。对于采矿对工程师和地质学家而言，岩性信息对矿床的成功开采至关重要（Patel等，2017；Perez等，2011）。然而岩石结构具有非均质性和强方向性特征（Lepistö等，2005a；Zhu等，2018）。因此，岩性信息的精确提取是地质学领域的关键任务（Shang和Barnes，2012）。

现有岩石图像分类方法可分为两大类：基于手工特征的方法与深度学习方法。基于手工特征的方法指利用岩石图像的颜色、纹理、凸起、颗粒形状等特征参数，通过分类方法预测岩石类型。Lepistö等（2005b）采用带通滤波技术处理不同色彩空间的图像，实现多尺度彩色纹理图像分析。Seng和Chen（2009）运用RS理论与机器学习中的支持向量机（SVM）模型解决岩石分类问题。

* 通讯作者。中国地质大学（武汉）地理与信息工程学院，湖北武汉。
电子邮箱：zhuqq@cug.edu.cn（朱琪琪），20171003123@cug.edu.cn（王赛），572375405@qq.com（唐顺），1179660637@qq.com（杨良斌），qikunlun@cug.edu.cn（齐昆仑），guanqf@cug.edu.cn（关清风）。

https://doi.org/10.1016/j.cageo.2024.105560
Received 5 June 2023; Received in revised form 22 November 2023; Accepted 5 February 2024
Available online 12 February 2024
0098-3004/© 2024 Elsevier Ltd. All rights reserved.

these problems are handled by pre-processing, existing hand-crafted feature-based methods have shown variable performance. However, the type of methods relied hand-crafted feature descriptors to design rock image prior feature and is difficult to capture deep semantic information of complex rock images.

In recent years, a plethora of deep learning methods have been widely used in the rock images classification. With features such as local connections, shared weights, pooling etc., deep learning methods offer the capacity to significantly mitigate network complexity and reduce the number of training parameters. Deep learning methods can effectively reduce the complexity of networks and the number of trainable parameters, enabling models to exhibit a certain degree of invariance to translation, distortion, and scaling. They possess strong robustness and fault tolerance, making them easy to train and optimize network structures. (LeCun et al., 2015). Therefore, using deep learning to build an automated model for the recognition and classification of rock images is a more effective way. Pascual et al. (2019) employed a three-layer CNN network method to enhance the classification accuracy of rock images. Y. Zhang et al. (2018) employed the Inception-v3 CNN architecture to training and testing rock type recognition models for granite, phyllite and breccia images using transfer learning. Su et al. (2020) introduced a novel approach called the concatenated convolutional neural network (Con-CNN) for geologic rock type classification using petrographic thin sections. Liang et al. (2021) introduced a fine-grained image classification framework that integrates the image cropping technique and the SBV algorithm, aiming to improve the classification accuracy of a limited set of fine-grained rock samples.

Previous state-of-the-art methods have shown the substantial efficacy of deep neural network methods in enhancing the efficiency of rock image recognition. (Baraboshkin et al., 2020; Karimpouli and Tahmasebi, 2019; Młynarczuk et al., 2013; Shu et al., 2017). However, there are still some shortcomings in their work: (1) Previous works usually extract features directly from the whole images, without focusing on the critical areas related to rock images. However, the background of rock image scenes is complex, which has a redundant effect on the classification. As shown in Fig. 1(a), both images are coarse crystal, the yellow and red boxes highlight critical regions within rock images that are challenging to capture but essential for rock classification. These regions serve as pivotal elements for the classification task. However, it is hard for the traditional models to ignore the differences of non-critical areas in both images. (2) Traditional end-to-end CNN-based models focus more on local information because of their important relational inductive bias, i.e., locality (Battaglia et al., 2018). Thus, they cannot make full use of spatial distribution features and spectral features of rock images. As shown in Fig. 1(b), two calcareous images in blue boxes are surrounded by other minerals, indicating that complex global spatial distribution information is also important. The presence of diverse spatial characteristics within the same rock type poses a challenge to our classification task. There are similar spatial distribution features but quite different spectral features in the two carbonaceous images as shown in Fig. 1(c). The variability of spectral features further increases the complexity of rock classification and the potential impact on its accuracy. (3) The datasets for microscopic images of rock sections used in existing studies are usually hundreds to thousands of images (Chatterjee, 2013; Shang and Barnes, 2012). They are not able to meet the training requirements of end-to-end CNN-based methods which are based on data-driven. In addition, there is only one type of rock property in each existing dataset, without a comprehensive description of different rock properties, leading to the “application gap".

Therefore, the purpose of this study is to construct a deep learning model and combine the spatial structure characteristics of rocks to carry out rock image classification. This study proposed a spatial localization Siamese network (RockS^{2}Net) for rock image classification. The RockS^{2}Net is designed to extract local features in critical areas and global characteristics from microscopic images of rock sections, while reducing the loss of features through the introduction of a specially designed Siamese Global DenseNet (SGD) block. Simultaneously, the SGD blocks can integrate global and local features and improve the accuracy of rock distinguishing features. The SGD blocks adeptly integrate both overall and localized features from two streams, producing distinctive features

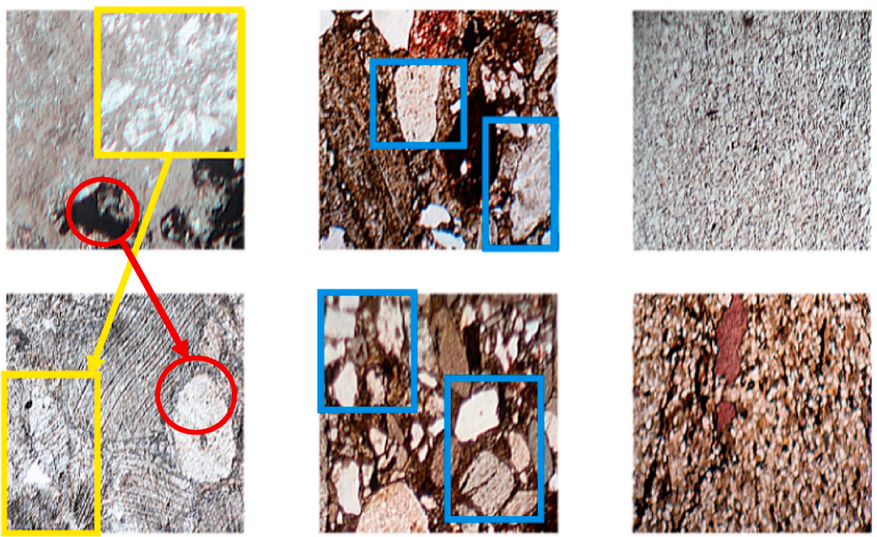

Fig. 1. Challenges in rock image classification for microscopic images of rock sections. (a) Critical areas which are difficult to capture. (b) Complex global spatial distribution features. (c) Similar spatial distribution features but quite different spectral features. (The scale is 500 × 500).

这些问题通过预处理解决，现有的基于手工特征的方法表现参差不齐。然而，这类方法依赖手工设计的特征描述符来构建岩石图像先验特征，难以捕捉复杂岩石图像的深层语义信息。

近年来，大量深度学习方法被广泛应用于岩石图像分类领域。凭借局部连接、权重共享、池化等特性，深度学习方法能显著降低网络复杂度并减少训练参数量。这类方法可有效减轻网络复杂性及可训练参数数量，使模型对平移、畸变和缩放具有一定的不变性，同时具备强鲁棒性和容错能力，便于网络结构的训练与优化(LeCun等, 2015)。因此，采用深度学习构建岩石图像识别与分类的自动化模型是更高效的途径。Pascual等(2019)采用三层CNN网络方法提升岩石图像分类精度；Y. Zhang等(2018)运用Inception-v3 CNN架构，通过迁移学习对花岗岩、千枚岩和角砾岩图像进行岩石类型识别模型的训练与测试；Su等(2020)提出串联卷积神经网络(Con-CNN)新方法用于岩相薄片的岩石类型分类；Liang等(2021)则引入结合图像裁剪技术与SBV算法的细粒度图像分类框架，旨在提升有限细粒度岩石样本的分类准确率。

先前最先进的方法已证明深度神经网络技术在提升岩石图像识别效率方面具有显著成效（Baraboshkin等，2020；Karimpouli与Tahmasebi，2019；Mlynarczuk等，2013；Shu等，2017）。然而这些研究仍存在不足：（1）既往工作通常直接从整幅图像提取特征，未聚焦与岩石图像相关的关键区域。但岩石图像场景背景复杂，会对分类。如图1(a)所示，两幅图像均为粗晶结构，黄色与红色框标注的是岩石图像中难以捕捉但对分类至关重要的关键区域，这些区域构成分类任务的核心要素。然而传统模型难以忽略两幅图像非关键区域的差异。(2)传统基于CNN的端到端模型因其重要的关系归纳偏置（即局部性，Battaglia等人，2018）更关注局部信息，无法充分利用岩石图像的空间分布特征与光谱特征。图1(b)中蓝色框内两幅钙质图像被其他矿物包围，表明复杂的全局空间分布信息同样重要。同一岩石类型内部呈现的多样化空间特征为分类任务带来挑战。图1(c)所示两幅碳质图像具有相似的空间分布特征但光谱特征差异显著，这种光谱特征的变异性进一步增加了岩石分类的复杂性及对准确率的潜在影响。(3)现有研究使用的岩石薄片显微图像数据集通常仅含数百至数千幅图像（Chatterjee，2013；Shang和Barnes，2012），无法满足数据驱动型CNN端到端方法的训练需求。此外，现有数据集每套仅包含单一岩石属性，缺乏对不同岩石属性的综合描述，导致"应用鸿沟"。

因此，本研究旨在构建深度学习模型，结合岩石空间结构特征进行岩石图像分类。研究提出了一种用于岩石图像分类的空间定位孪生网络（RockS ^{2} Net）。Rock- S^2 Net通过设计特殊的孪生全局密集网络（SGD）模块，既能提取岩石切片显微图像关键区域的局部特征和全局特征，又能减少特征损失。SGD模块可融合全局与局部特征，提升岩石鉴别特征的准确性。该模块能巧妙整合双流网络的整体与局部特征，生成具有区分度的特征

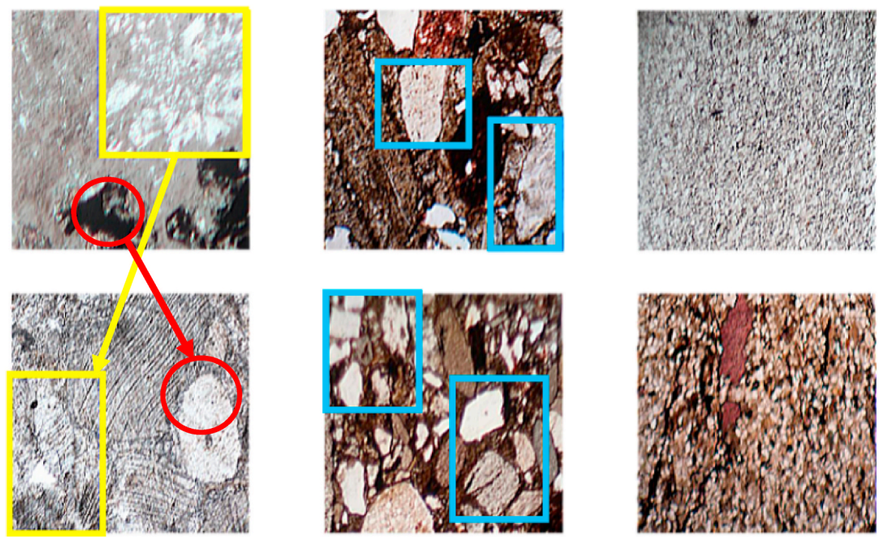

图1. 岩石切片显微图像分类面临的挑战：(a)难以捕捉的关键区域；(b)复杂的全局空间分布特征；(c)空间分布特征相似但光谱特征差异显著（比例尺为 500 × 500 ）

for the rock cross-section. To address the issue of affect of irrelevant features on rock image classification, a Spatial Transformer Network (STN) is introduced to transform microscopic images of rock sections. By localizing critical regions, the STN reduces the influence of irrelevant features on the rock image classification. Finally, a large-scale dataset, the CHN-Rock images dataset, is constructed for rock image classification. The CHN-Rock images dataset includes a more comprehensive description of rock properties, including more types of rock properties than the previous datasets. Our dataset is not uniform, as the images were captured by using different types of cameras, and the lighting conditions under which these images were taken vary considerably. The dataset proposed aligns more closely with practical application requirements and poses significant challenges, serving as a robust testbed for demonstrating our model's generalization capabilities.

The subsequent sections of this paper are structured as follows. Section II introduces the related works of rock image classification approaches based on handcrafted feature-based methods and CNN. In Section III, the proposed CHN-Rock images dataset for rock image classification is introduced. Section IV presented the proposed rock image classification method in detail. In Section V, the experiments and the evaluation of the model are introduced. Section VI introduces the performance evaluation and analysis. Finally, the conclusion is given that the proposed model is effective as well as exhibits robust generalization capabilities.

2. Related work

2.1. Handcrafted feature-based methods

Handcrafted feature-based methods extract rock property features through traditional mathematical statistical models and computational analysis. The extracted features are added to various classifiers, such as naïve Bayes (NB) (Zhang, 2004), decision tree (DT) (Swain and Hauska, 1977) to obtain classification results of rock properties. (Mkwelo, 2004) used the least square polynomial method to fit the irregular edges of ore, and then obtained the edge contour information of ore. Zhang (2004) applied the naive Bayes k-neighbor algorithm for the task of image classification. (Młynarczuk et al., 2013) utilized both the nearest neighbor method and k-nearest neighbor method to classify rock images for nine different rock categories. Sharif et al. (2015) designed a classifier that used 13 Haralick textural parameters to characterize rock images, automatically cataloged them, computed Bayesian probabilities for rock image classification. A new method of rock typing classification based on geometric features of grains instead of local features is proposed, but it has some limitations (Wang and Sun, 2021).

In recent years, Support Vector Machine (SVM) has been recognized acknowledged as one of the widely adopted approaches to pattern recognition (Qin and He, 2005; Sun et al., 2002). There have been many studies based on SVM classify (Chatterjee, 2013; Dunlop, 2006; Lepistö et al., 2005a; Patel et al., 2017, 2019). N. Li et al. (2017) proposed a transfer learning method and classified microscopic images of sandstone sections by four classifiers, i.e., NB, DT, LR and SVM to demonstrate the effectiveness. Momma et al. (2006) utilized color representation to characterize the extent of rock weathering, employing SVM and neural network algorithms to classify the degree of rock weathering. Li and Wang (2019) will apply PSO-SVM technology to classify and predict the rocks around the tunnel. *End-to-End CNN-Based Methods*.

CNN is a very successful and important part of the current deep learning system. It possesses a broad spectrum of applications and are adept at directly operating on input images, rendering them a versatile and impactful tool in the field of image analysis and processing. Karimpouli and Tahmasebi (2019) used deep convolutional autoencoder networks to segment digital rock images. Baraboshkin et al. (2020) designed a new method based on CNN models to reduce the time needed for the accurate description of rocks. Su et al. (2020) designed a concatenated CNN model, which can effectively extract features from microscopic images of rock sections and achieve automatic rock classification. Zhu et al. (2018) proposed a depth-wise separable convolution method for rock classification by microscopic images of rock sections. Ran et al. (2019) proposed a deep convolutional neural network-based method for wild rock type recognition, which can identify six common rock types. Guojian and Peisong, (2021) utilized a residual network to construct a classifier for rock image classification and demonstrated the efficiency and accuracy of the residual network. The Siamese network was proposed, which can skillfully integrate the characteristics from two types of polarized images, and eliminate the domain related information in the feature encoding through adversarial training, effectively reducing the negative impact of domain features on rock classification and recognition. (Hao et al., 2022). A new rock core classification method based on deep learning was established, which improves the overall interpretation accuracy and reduces the subjectivity and interpretation time (Dawson et al., 2023).

3. CHN-rock images dataset: a dataset for rock image classification

3.1. Construction of CHN-rock image dataset

In this study, the dataset employed spans 54 regions across China, comprising a total of 13,112 rock slices. These rock slices were derived from microscopes with varying resolutions, diverse slice orientations, and distinct lighting conditions. By capturing each slice from multiple angles, we acquired a collection of 264,648 microscopic images of rock slices. As shown in Fig. 2, our dataset contains both Plane-Polarized Light (PPL)- captured images as well as Cross-Polarized Light (XLP)-captured images. Considering the disparities in color, lighting, and resolution among the rock slices, the rock classification dataset we have proposed exhibits a considerable level of complexity, posing significant challenges for rock classification tasks. These microscopic images of rock slices encompass numerous subtypes of rocks falling within the three major categories of igneous, sedimentary, and metamorphic rocks. All these rock slice images were meticulously annotated by seasoned experts in geological image interpretation, based on the rock slice images and the fundamental information about the rocks. Following a comprehensive statistical analysis and rigorous image selection, we have constructed a dataset titled the CHN-Rock comprising five categories of rock attributes, which include grain properties, clastic properties, mechanical genesis attributes, mixture characteristics, and basic category attributes.

Among them, the grain property is consisted of seven types, including coarse crystal, medium crystal, fine crystal, powder crystal, micro crystal, mud crystal, and others. The clastic property has eight types, including fine gravel, medium gravel, medium gravel sand, coarse sand, fine sand, coarse silt, fine silt and others. The mechanical genesis property has five types, which are gravel, sand, powder, aggregate, and others. The mixture property includes seven types, which are calcareous, iron, siliceous, containing silty, silty, carbonaceous, and others. The basic category property is consisted of nine types, including granite, diorite, basalt, tuff, lithic sandstone, quartz sandstone, siltstone, slate, and others. In each of the above types, several images have been randomly selected from the remaining categories to construct the above 'other' category. Figs. 3–7 show some examples of these five types of rock properties.

In order to construct the CHN-Rock images dataset, data cleaning and pre-processing was performed on all microscopic images of rock sections. For the five types of rock properties, i.e., grain property, clastic property, mechanical genesis property, mixture property, and basic category property, 400, 250, 400, 400 and 300 images were selected for each property type, respectively. The ratio of the training set and test set for each category of properties was set to be 4:1, and all microscopic images of rock sections were uniformly resized to 512 × 512 pixels by nearest neighbor interpolate algorithm.

针对岩石切面图像分类中无关特征影响的问题，本研究引入空间变换网络(STN)对岩石显微图像进行形变处理。该网络通过定位关键区域，有效降低了无关特征对岩石图像分类的干扰。最终构建了大规模岩石图像数据集CHN-Rock，其特色在于：1)较既往数据集包含更丰富的岩石属性类型，提供更全面的岩性特征描述；2) 数据集具有非均匀特性，图像采集采用多类型相机设备且光照条件差异显著。该数据集更贴近实际应用需求，其复杂性为验证模型泛化能力提供了 rigorous 的测试平台。

本文后续章节结构如下：第二节介绍基于手工特征方法和CNN的岩石图像分类相关研究；第三节提出用于岩石图像分类的CHN-Rock数据集；第四节详细阐述所提出的岩石图像分类方法；第五节介绍实验及模型评估；第六节进行性能评估与分析；最终结论表明，所提模型不仅有效且具备强泛化能力。

2. 相关工作

2.1. 基于手工特征的方法

基于手工特征的方法通过传统数理统计模型与计算分析提取岩石属性特征，并将这些特征输入朴素贝叶斯(NB)(Zhang, 2004)、决策树(DT)(Swain and Hauska,1977)等分类器获取岩性分类结果。Mkwelo(2004)采用最小二乘多项式拟合矿石不规则边缘，进而获取矿石边缘轮廓信息。Zhang(2004)将朴素贝叶斯k近邻算法应用于图像分类任务。Młynarczuk等(2013)综合运用最近邻法与k近邻法对九类岩石图像进行分类。Sharif等(2015)设计了一种利用13个Haralick纹理参数表征岩石图像的分类器，实现自动编目并计算贝叶斯概率完成岩像分类。Wang与Sun(2021)提出了一种基于颗粒几何特征而非局部特征的岩相分类新方法，但该方法存在一定局限性。

近年来，支持向量机（SVM）已被公认为模式识别领域广泛采用的方法之一（Qin and He, 2005; Sun et al., 2002）。基于SVM分类的研究层出不穷（Chatterjee, 2013; Dunlop, 2006; Lepistö et al., 2005a; Patel et al., 2017, 2019）。N. Li等人（2017）提出迁移学习方法，通过朴素贝叶斯（NB）、决策树（DT）、逻辑回归（LR）和SVM四种分类器对砂岩切片显微图像进行分类以验证其有效性。Momma等（2006）利用颜色表征来刻画岩石风化程度，采用SVM和神经网络算法对岩石风化等级进行分类。Li与Wang（2019）将应用PSO-SVM技术对隧道围岩进行分类预测。端到端基于CNN的方法。

CNN是当前深度学习体系中非常成功且重要的组成部分。它拥有广泛的应用领域，擅长直接对输入图像进行操作，使其成为图像分析与处理领域多功能且极具影响力的工具。Karimpouli和Tahmasebi（2019）利用深度卷积自编码网络对数字岩石图像进行分割。Baraboshkin等人（2020）设计了一种基于CNN模型的新方法，可大幅缩短岩石精确描述所需时间。Su等（2020）提出的级联CNN模型能有效提取岩石切片的显微图像并实现岩石自动分类。Zhu等(2018)提出了一种基于深度可分离卷积的岩石切片显微图像分类方法。Ran等(2019)提出基于深度卷积神经网络的野外岩石类型识别方法，可识别六种常见岩石类型。郭建与裴松(2021)利用残差网络构建岩石图像分类器，验证了残差网络的效率与准确性。提出孪生网络模型，能巧妙融合两种偏振图像特征，通过对抗训练消除特征编码中的域相关信息，有效降低域特征对岩石分类识别的负面影响(Hao等，2022)。建立了基于深度学习的新型岩心分类方法，提高了整体解释精度，降低了主观性并缩短了解释时间(Dawson等，2023)。

3. CHN-rock岩石图像数据集：用于岩石图像分类的数据集

3.1. CHN-rock图像数据集的构建

本研究所采用数据集覆盖中国54个地区，共包含13,112片岩石薄片。这些薄片源自不同分辨率的显微镜、多样切片方位及差异化的光照条件。通过多角度采集每张薄片，我们获得了264,648张岩石显微图像集。如图2所示，数据集同时包含平面偏振光(PPL)成像与交叉偏振光(XPL)成像。鉴于岩石薄片在色彩、光照及分辨率方面存在的显著差异，我们所构建的岩石分类数据集具有较高复杂度，对岩石分类任务形成重大挑战。这些岩石显微图像涵盖火成岩、沉积岩和变质岩三大类下的众多亚型。所有薄片图像均由经验丰富的地质影像解译专家，基于薄片图像及岩石基础信息进行精细标注。经过全面统计分析及严格图像筛选后，我们最终构建了名为CHN-Rock的数据集，包含颗粒特征、碎屑特征、力学成因属性、混合特征及基本类别属性五类岩石属性。

其中，粒度属性包含粗晶、中晶、细晶、粉晶、微晶、泥晶及其他七类；碎屑属性分为细砾、中砾、中砾砂、粗砂、细砂、粗粉砂、细粉砂及其他八类；机械成因属性包括砾、砂、粉、集合体及其他五类；混合属性涵盖钙质、铁质、硅质、含粉砂质、粉砂质、碳质及其他七类；基础类别属性由花岗岩、闪长岩、玄武岩、凝灰岩、岩屑砂岩、石英砂岩、粉砂岩、板岩及其他九类构成。上述每类中，均从其余类别随机选取若干图像构建"其他"类别。图3-7展示了这五类岩石属性的部分示例。

为构建CHN-Rock图像数据集，对全部岩石切片显微图像进行了数据清洗与预处理。针对颗粒属性、碎屑属性、力学成因属性、混合属性及基本类别属性这五类岩石性质，分别选取了400、250、400、400和300张图像。各属性类别的训练集与测试集比例设定为 $4:1$，所有岩石切片显微图像均通过最近邻插值算法统一调整为 512×512 像素。

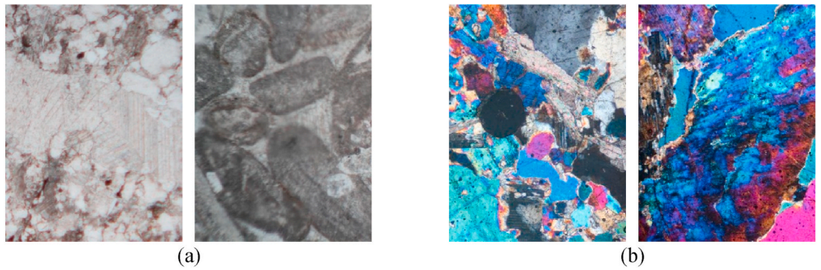

Fig. 2. (a) Represents images captured through PPL; (b) represents images captured through XLP.

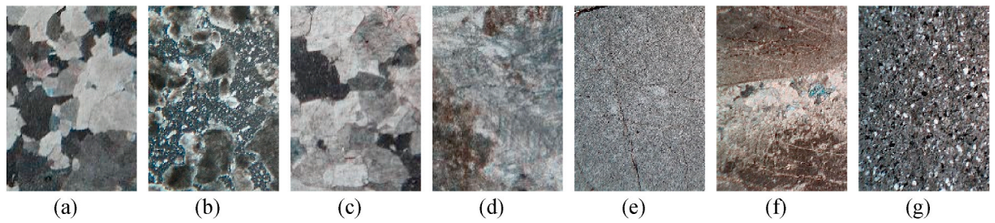

Fig. 3. Example images of grain property. (a) Coarse crystal. (b) Medium crystal. (c) Fine crystal. (d) Powder crystal. (e) Micro crystal. (f) Mud crystal. (g) Others. (Scale from 300 to 500).

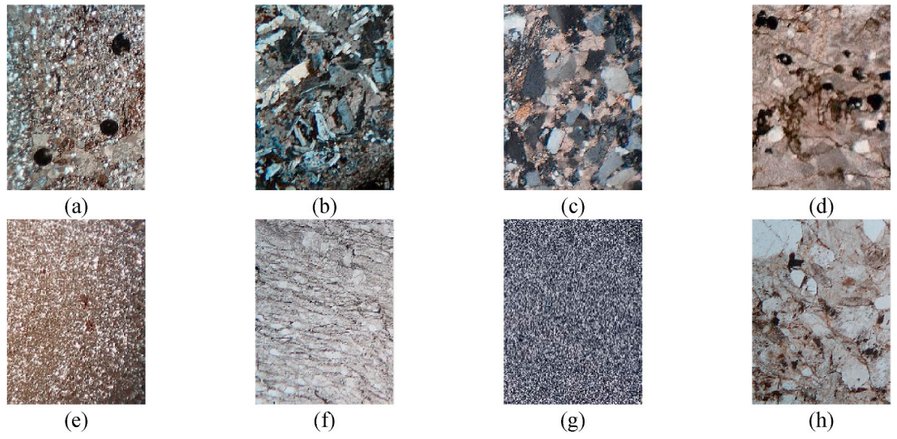

Fig. 4. Example images of clastic property. (a) Fine gravel. (b) Medium gravel. (c) Medium gravel sand. (d) Coarse sand. (e) Fine sand. (f) Coarse silt. (g) Fine silt. (h) Others. (Scale from 300 to 500).

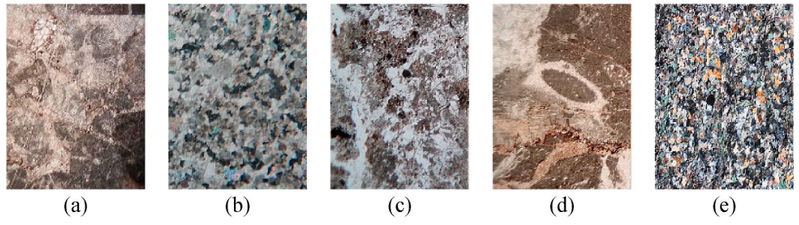

Fig. 5. Example images of mechanical genesis property. (a) Gravel. (b) Sand. (c) Powder. (d) Aggregate. (e) Others. (Scale from 300 to 500).

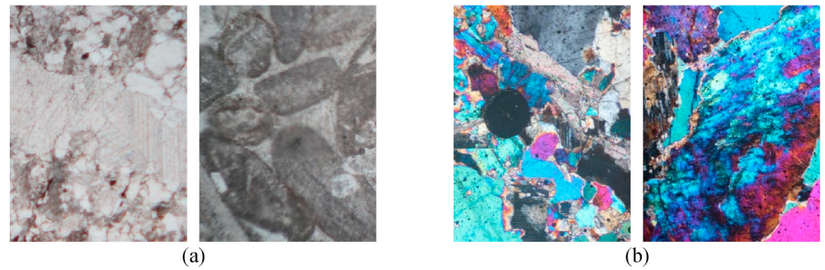

图2. (a)为通过PPL拍摄的图像；(b)为通过XLP拍摄的图像。

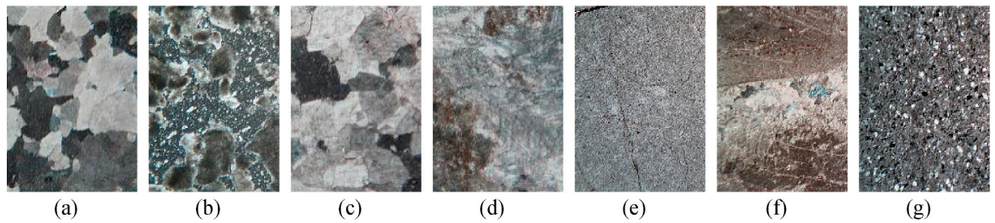

图3. 颗粒性质示例图像。(a)粗晶 (b)中晶 (c)细晶 (d)粉晶 (e)微晶 (f)泥晶 (g)其他（比例尺300-500）

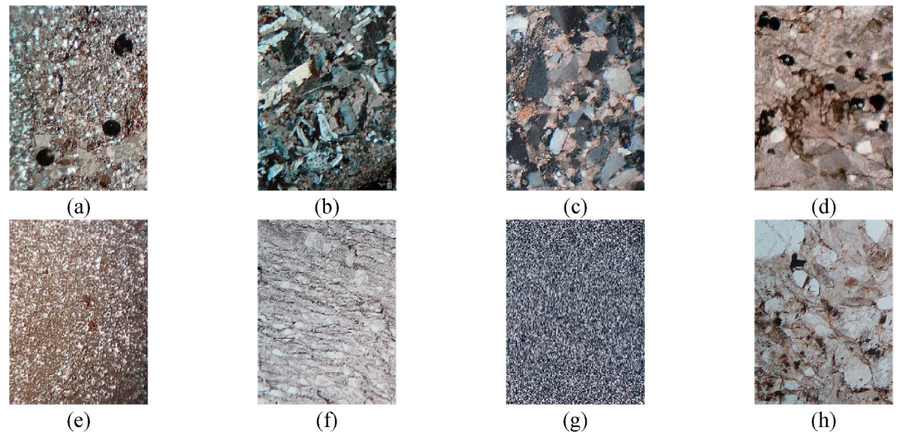

图4. 碎屑性质示例图像。(a)细砾 (b)中砾 (c)中砾砂 (d)粗砂 (e)细砂 (f)粗粉砂 (g)细粉砂 (h)其他（比例尺300-500）

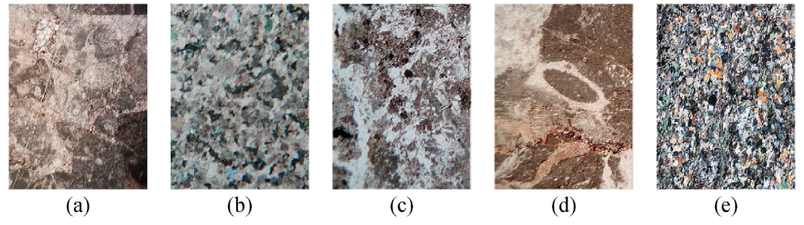

图5. 机械成因性质示例图像。(a)砾石 (b)砂 (c)粉砂 (d)集合体 (e)其他（比例尺300-500）

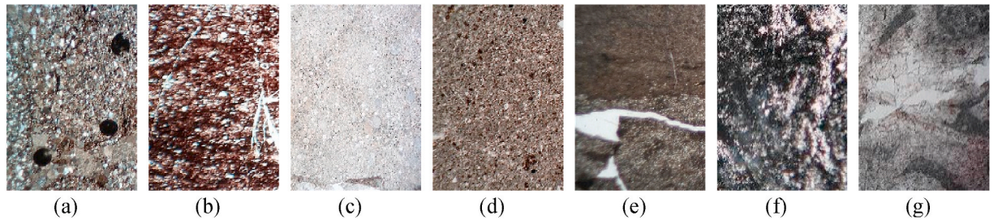

Fig. 6. Example images of mixture property. (a) Calcareous. (b) Iron. (c) Siliceous. (d) Containing silty. (e) Silty. (f) Carbonaceous. (g) Others. (Scale from 300 to 500).

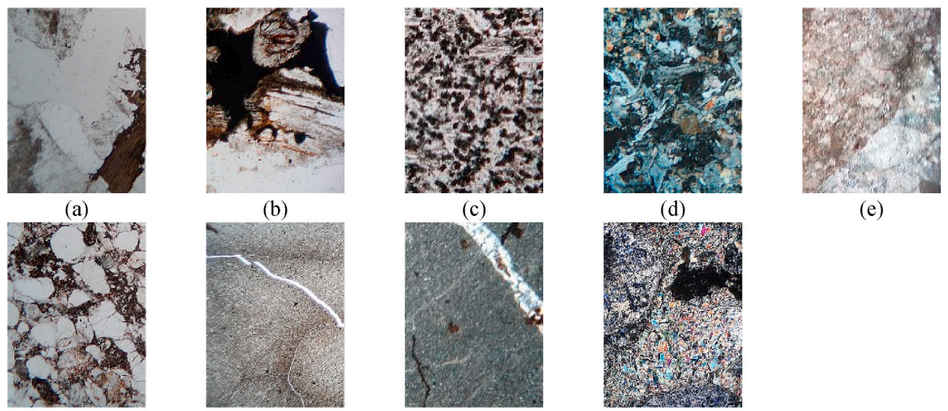

Fig. 7. Example images of basic category property. (a) Granite. (b) Diorite. (c) Basalt. (d) Tuff. (e) Lithic sandstone. (f) Quartz sandstone. (g) Siltstone. (h) Slate. (i) Others. (Scale from 300 to 500).

3.2. Characteristics and challenges of CHN-ROCK images dataset

As can be seen from Figs. 3–7, the microscopic images of different types of rock sections belonging to the same rock properties vary greatly. It is very difficult for non-professionals to distinguish between them with naked eyes. Among the five types of properties, the grain property is a small irregular crystal which consists of poly crystals, and each grain is sometimes composed of several sub-grains with slightly different orientations. Grains are also generally referred to as particles of crystalline minerals in rocks. The type of grain properties is classified according to the grain size and the average grain diameter is usually in the range of 0.015–0.25 mm. Different from the grain properties, the clastic property is an integral component of sedimentary rocks or sediment. It is a product of mechanical weathering of the parent rock and generally refers to the small pieces after the whole fracture, which is also a property related to the size of the fragments. It refers to the debris, excreta and their common decomposition products, which are always mixed with microorganisms. The mechanical diagenetic properties of rocks are related to their shape and texture, and involve a series of physical and chemical changes. The nature of rock mixture is related to the mineral composition of rocks. For many mineral components, the mineral property appears in the identification name of the rock only when the proportion of minerals reaches a certain threshold. The last property is the basic category property of rocks, which is the basic name of rocks.

4. Methodology

In order to improve the rock image classification performance of microscopic rock section images, we propose the RockS^{2}Net framework to effectively solve the above problems. Fig. 8 illustrates the complete flowchart of the RockS^{2}Net framework. A succinct introduction to the RockS^{2}Net framework is given in the subsequent section.

Fig. 8 shows the overall architecture of RockS^{2}Net. Firstly, the original image is transformed to produce the image representation with richer features. Through image transforming, the network attempts to extract more informative features from the original image. After image transforming, there is a risk of information loss. In order to prevent information loss after image transformation, RockS^{2}Net adopts a strategy that the transformed image and the original image are passed into two SGD blocks respectively for feature extraction. In the SGD block, three × three convolution kernels with a dilation rate of 2 are employed to expand the receptive field of the network. To avoid overfitting, fully connected (FC) layers in CNNs are replaced with global average pooling (GAP) (Lin et al., 2013) layers which have fewer training parameters. The features extracted from original images and transformed images are regularized and fused in batch normalization (BN) layers (Ioffe and Szegedy, 2015).

4.1. RockS^{2}Net framework of rock image classification

The proposed framework adopts a Siamese network as its backbone, consisting of two sub-networks with identical architectures and weights, which transform whole images and critical areas into feature maps. The critical regions of images are transformed from whole images by the introduced architecture of STN. To aggregate multi-scale contextual information, each convolution kernel of DenseNet is replaced dilated convolution kernel.

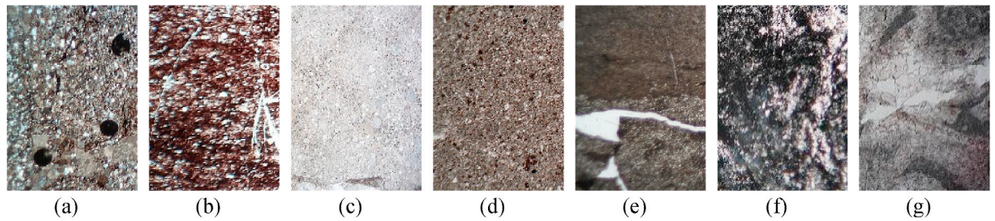

图6. 混合性质示例图像。(a) 钙质。(b) 铁质。(c) 硅质。(d) 含粉砂。(e) 粉砂质。(f) 碳质。(g) 其他。(比例尺300至500)。

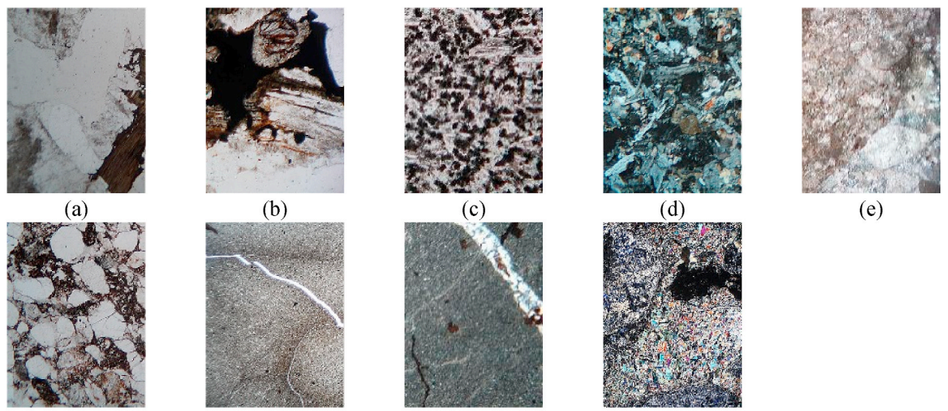

图7. 基本类别性质示例图像。(a) 花岗岩。(b) 闪长岩。(c) 玄武岩。(d) 凝灰岩。(e) 岩屑砂岩。(f) 石英砂岩。(g) 粉砂岩。(h) 板岩。(i) 其他。(比例尺300至500)。

3.2. CHN-ROCK图像数据集特性与挑战

如图3-7所示，同种岩性下不同类型岩石切片的显微图像差异显著，非专业人员难以通过肉眼区分。五种岩性中，粒状岩性指由多晶组成的细小不规则晶体，单个晶粒常由数个取向略异的亚晶粒构成。地质学中，晶粒通常泛指岩石中结晶矿物的颗粒，其分类依据粒径大小，平均粒径范围通常为 $0.015-0.25\ \mathrm{mm}$ 。与粒状岩性不同，碎屑岩性是沉积岩或沉积物的基本组分，为母岩机械风化产物，特指整体破裂后形成的碎块，其分类亦与碎屑尺寸相关。这类岩性包含岩屑、排泄物及其共同分解产物，常与微生物混杂。岩石的机械成岩特性与其形态和结构相关，涉及一系列物理化学变化。混合岩性则与岩石的矿物组成有关——只有当矿物比例达到特定阈值时，该矿物属性才会出现在岩石鉴定名称中。最后一种为岩石的基础类别属性，即岩石的基本名称。

4. 方法论

为提升岩石薄片图像的分类性能，我们提出RockS^2 Net框架以有效解决上述问题。图8展示了RockS^2 Net框架的完整流程图。下一节将简要介绍该框架。

图8展示了 RockS^2 网络的整体架构。首先对原始图像进行变换，生成具有更丰富特征的图像表示。通过图像变换，网络试图从原始图像中提取更具信息量的特征。图像变换后存在信息丢失的风险。为防止图像变换后的信息损失，RockS^2 网络采用将变换图像与原始图像分别输入两个SGD模块进行特征提取的策略。在SGD模块中，使用三个膨胀率为2的 × 卷积核来扩大网络感受野。为避免过拟合，将CNN中的全连接层替换为训练参数更少的全局平均池化层（Lin等人，2013）。原始图像与变换图像提取的特征在批归一化层（Ioffe和Szegedy，2015）中进行正则化与融合。

4.1 岩石图像分类的Rock S^2 网络框架

该框架采用孪生网络作为主干结构，由两个架构和权重相同的子网络组成，能够将完整图像和关键区域转换为特征图。通过引入的空间变换网络(STN)架构，图像关键区域由完整图像转换而来。为聚合多尺度上下文信息，DenseNet的每个卷积核均被替换为空洞卷积核。

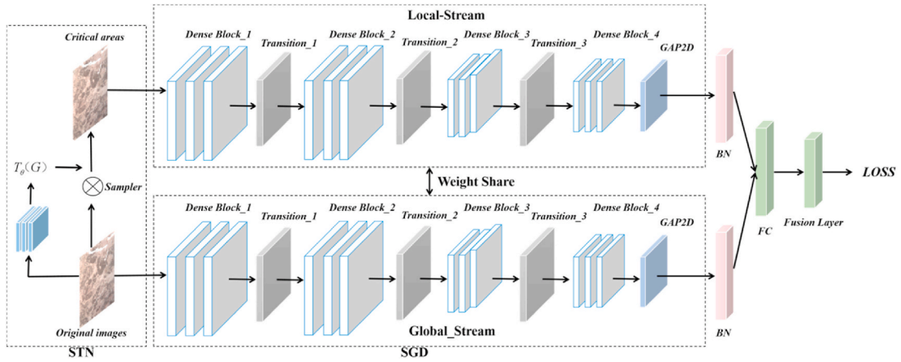

Fig. 8. Overview of the proposed framework for rock image classification.

4.2. Critical region spatial localization

Before the features are obtained from the contiguous structural encoding, active spatial transformation of the image (or feature map) through a mechanism such as the spatial transformer module that generates an appropriate transformation of the input rock image. The transformation is applied to the entire feature map (non-local) and includes scaling, cropping and rotation. The advantage of this mechanism is that the sample images can be transformed dynamically, whereas the samples received in the pooling layer are fixed and localized. Therefore, we use the spatial transformer module to select the most relevant regions in the image and transform these regions to a canonical, expected pose to improve the classification ability of the network.

The STN consists of three components: the localization network, the grid generator, and the sampler. This localization network performs the function of parameter prediction. The grid generator performs the function of coordinate mapping. The sampler provides the function of pixel acquisition. The architecture of STN is shown in Fig. 9.

Because the affine transformation is differentiable, the gradient can flow through this layer during back propagation. The parameters of transformer are provided by localization network, then back propagation allows learning parameter transformation. The affine transformation formula is shown in Eq. (1), where (a_i, b_i) and (a_i^t, b_i^t) are the regularized coordinates of the input space and the output space, respectively.

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} = T_\theta \begin{pmatrix} a_i^t \\ b_i^t \\ 1 \end{pmatrix} = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{pmatrix} \begin{pmatrix} a_i^t \\ b_i^t \\ 1 \end{pmatrix} \tag{1}$$

Eq. (2) shows the principle of bilinear interpolation sampling. The coordinate (a_i, b_i) represents a position in the original feature map U, and the pixel values of the output feature map V are obtained by the

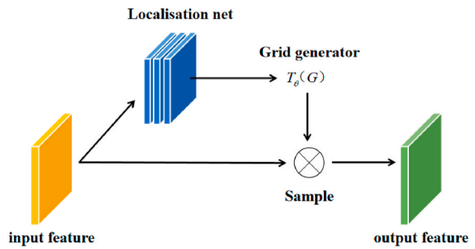

Fig. 9. The architecture of the spatial transformer network (STN).

sampling kernel. V_i^c is the output value of pixel i on channel c and U_{nm}^c is the input pixels at coordinate (n, m). The two maximum functions determine the relative weights of each pixel.

$$V_i^c = \sum_n^H \sum_m^W U_{nm}^c \cdot \max(0, 1 - |a_i - m|) \cdot \max(0, 1 - |b_i - n|) \tag{2}$$

For the bilinear interpolation in Eq. (2), the partial derivatives can be calculated according to Eq. (3).

$$\frac{\delta V_i^c}{\delta U_{nm}^c} = \sum_n^H \sum_m^W \max(0, 1 - |a_i - m|) \cdot \max(0, 1 - |b_i - n|) \tag{3}$$

The affine transformation of STN used by RockS^{2}Net is transformed into a cropping transformation. The affine transformation formula of cropping operation can be formulated as Eq. (4), where the scale, rotation, x-translation, y-translation transformation parameter is denoted as s, 0, t_x and t_y, respectively.

$$T_\theta = \begin{pmatrix} s & 0 & t_x \\ 0 & s & t_y \end{pmatrix} \tag{4}$$

If the transformed images are obtained directly from the original images, then each pixel point in the original image is not useful. Furthermore, the grid coordinates of the pixel points of the transformed images obtained from the grid coordinates of the pixel points of the input image are non-integer. The pixel values of the pixel points of the transformed images are not available. As the affine transformation is invertible, the inverse affine transformation is used here, which is the process of finding each pixel point of the transformed images which correspond to the grid position of the original image. The grid position of each coordinate point of the transformed image corresponding to the original image is also non-integer, but its pixel value can be obtained by interpolation of the pixels in the surrounding grid. The matrix of the inverse affine transformation is also an affine matrix, which is the inverse of the affine transformation matrix.

The function of STN is to extract regions of the attention from the microscopic images of rock sections, which can effectively distinguish between different categories. The images are cropped and translated with the values of scale transformation parameters varying between 0 and 1. In this way, STN can be regarded as a sliding window with a scale of to find the attention area from microscopic images of rock sections. The sliding distances in the horizontal and vertical directions are t_x and t_y, respectively.

The architecture of STN in RockS^{2}Net, especially the architecture of the localization network, includes three convolution layers, i.e., three pooling layers, a GAP layer, a BN layer, and a FC layer. The GAP layer is a kind of architecture which can replace FC layer in CNNs. The FC layer is prone to overfitting because of large parameters, which is a fatal

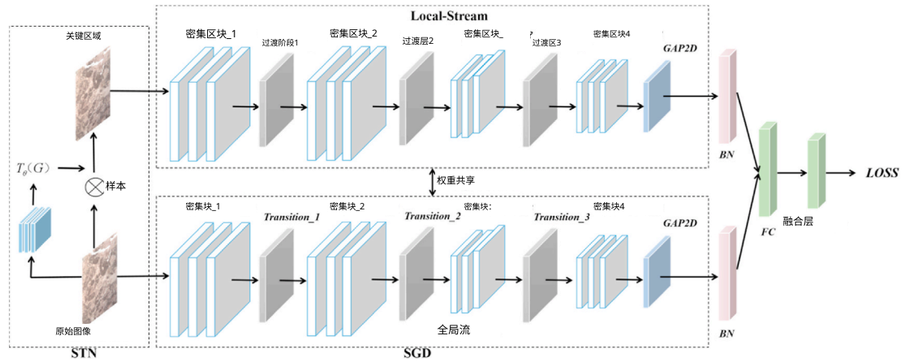

图8. 岩石图像分类提出的框架概览。

4.2 关键区域空间定位

在从连续结构编码中获取特征之前，需通过空间变换器模块等机制对图像（或特征图）进行主动空间变换，该模块会对输入的岩石图像生成适当的变换。该变换作用于整个特征图（非局部），包括缩放、裁剪和旋转。此机制的优点在于可动态变换样本图像，而池化层接收的样本是固定且局部化的。因此，我们利用空间变换器模块选择图像中最相关的区域，并将这些区域转换至标准预期位置，以提升网络的分类能力。

空间变换网络由三个组件构成：定位网络、网格生成器和采样器。定位网络负责参数预测功能，网格生成器实现坐标映射功能，采样器提供像素采集功能。空间变换网络架构如图9所示。

由于仿射变换是可微的，在反向传播过程中梯度能够流经该层。变换器的参数由定位网络提供，通过反向传播学习参数变换。仿射变换公式如式(1)所示，其中 (a_i, b_i) 和 (a_i^t, b_i^t) 分别表示输入空间与输出空间的归一化坐标。

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} = T_\theta \begin{pmatrix} a_i^t \\ b_i^t \\ 1 \end{pmatrix} = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{pmatrix} \begin{pmatrix} a_i^t \\ b_i^t \\ 1 \end{pmatrix} \tag{1}$$

式(2)展示了双线性插值采样的原理。坐标 (a_i, b_i) 代表原始特征图 U 中的位置，输出特征图 V 的像素值通过

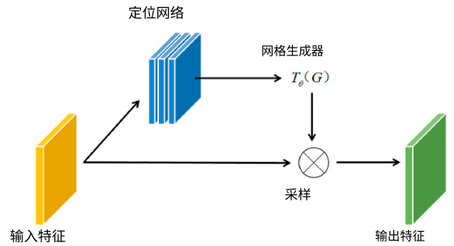

图9. 空间变换网络(STN)的架构。

采样核。V_i^c 表示通道 c 上像素i的输出值，U_{nm}^c 则是坐标(n,m)处的输入像素。两个最大值函数决定了各像素的相对权重。

$$V_i^c = \sum_n^H \sum_m^W U_{nm}^c \cdot \max(0, 1 - |a_i - m|) \cdot \max(0, 1 - |b_i - n|) \tag{2}$$

对于公式(2)中的双线性插值，其偏导数可根据公式(3)计算得出。

$$\frac{\delta V_i^c}{\delta U_{nm}^c} = \sum_n^H \sum_m^W \max(0, 1 - |a_i - m|) \cdot \max(0, 1 - |b_i - n|) \tag{3}$$

RockS^2 网络将STN的仿射变换转换为裁剪变换。该裁剪操作的仿射变换公式如式(4)所示，其中缩放、旋转、x轴平移、y轴平移参数分别用 $s, 0, t_x$ 和 t_y 表示。

$$T_\theta = \begin{pmatrix} s & 0 & t_x \\ 0 & s & t_y \end{pmatrix} \tag{4}$$

若变换后的图像直接源自原始图像，则原始图像中各像素点均无实用价值。此外，根据输入图像像素点网格坐标所获变换图像像素点的网格坐标均为非整数值，导致无法直接获取变换图像像素点的像素值。由于仿射变换具有可逆性，此处采用逆仿射变换——即寻找与原始图像网格位置相对应的变换图像各像素点的过程。虽然变换图像各坐标点对应原始图像的网格位置同样为非整数，但可通过周围网格像素插值获取其像素值。逆仿射变换矩阵亦为仿射矩阵，即原仿射变换矩阵的逆矩阵。

STN的功能是从岩石切片的显微图像中提取关注区域，能有效区分不同类别。图像通过尺度变换参数在0到1之间变化进行裁剪和平移。因此，STN可视为一个滑动窗口，以特定尺度从岩石切片显微图像中寻找关注区域。其水平和垂直方向的滑动距离分别为 t_x 与 t_y。

RockS^2 网络中STN的架构（特别是定位网络）包含三个卷积层、三个池化层、一个GAP层、一个BN层和一个FC层。GAP层是一种可替代CNN中FC层的架构。由于参数量大易导致过拟合，FC层存在致命缺陷

weakness. This GAP layer mainly pools the average values of feature maps of the last convolution layer, to speed up computation and reduce training parameters. The BN layer is also a kind of neural network architecture proposed in existing study. With the network's increasing depth, the eigenvalue distribution in each layer tends to approach the upper and lower bounds of the activation function's output interval, leading to the vanishing gradient problem. The BN layers pull the eigenvalue distribution of the layer towards a standard normal distribution, and the activation function becomes more responsive to the input, which accelerates the convergence by avoiding the disappearance of gradient. This architecture normalizes the output values of each batch from the previous layer, ensuring that the mean values of its output data approach 0 and standard deviations approach 1, which also reduces the insensitivity of network to initialization weights. These changes caused by BN layers make it possible to use larger learning rates in the networks. The BN layer can be written as Eq. (5). x_i is the feature map in a neural network layer with the index i. $E(x_i)$ and $Var(x_i)$ are the expectation and variance of the feature map x_i, respectively. ε is a small positive constant, such as 10–5, to prevent division by zero errors.

$$\widehat{x_i} = \frac{x_i = E(x_i)}{\sqrt{Var(x_i) + \varepsilon}} \tag{5}$$

The initial convolution layer has a depth of 16 and a convolution kernel size of 7×7. A rectified linear unit (ReLU) (Nair and Hinton, 2010) activation function is used and then followed by a 2×2 pooling layer. The convolution depth of the second convolution layer is 32 and the convolution kernel size is 5×5, using ReLU activation function and followed by a 2×2 pooling layer. The convolution depth of the third convolution layer is 64 and the convolution kernel size is 3×3. It is followed by a ReLU activation layer and a 2×2 pooling layer. The two translation parameters of affine transformations are obtained through several convolution layers, GAP layer, BN layer, and FC layer. The initialization parameters for the weights of the FC layer are manually specified and the activation function of tanh is used to ensure that the translation parameters are initialized to 0. The rotation parameters are set to 0 and specifying appropriate scaling parameters through the lambda layer. The lambda layer is to add two zero rotation parameters and two tuned scale parameters based on two translation parameters. These six parameters constitute the affine transformation matrix.

4.3. Dilated convolution in RockS^{2}Net framework

The RockS^{2}Net framework model is formed by using the dilated convolution kernels to instead the traditional convolution kernels. Dilated convolutional layers have been shown to improve the accuracy of classification tasks and it is a good alternative of pooling layers (Lei et al., 2019; Liu et al., 2020; Zhang, 2022). By introducing the Dilation Rate parameter, the null convolution results in a larger field of perception for the same size of convolution kernel. Accordingly, it is also possible to make the null convolution have a smaller number of parameters than the normal convolution with the same field size (Kudo and Aoki, 2017). Therefore, we introduce the convolution idea of dilated convolution for solving the information loss caused by image resolution reduction and down sampling in image classification problems.

Fig. 10 shows a conventional convolution kernel and a dilation convolution kernel on an image of size 9×9, where (a) is a conventional 3×3 convolution kernel and (b) is a null convolution with a dilation rate of 2. It is formed by inserting a hole (weight of 0) between each point in (a); similarly, (c) is a kernel with a dilation rate of 3. As shown in Fig. 10, the perceptual field of the convolution kernel is 3×3 in (a), 7×7 in (b), and 15×15 in (c). The size of the perceptual field increases with the addition of insertion holes, however, the parameter counts in (a), (b), and (c) remains the same. Therefore, using such an expanded convolution kernel to process the image allows the convolution kernel to obtain more information without increasing the computational effort.

Eq. (6) shows the principle of dilated convolution. Among them, k is the size of the input convolution kernel. D denotes the dilution coefficient employed. K signifies the resultant equivalent convolution kernel size after the application of dilation.

$$K = D \times (k-1) + 1 \tag{6}$$

To extract global contextual features of rock images, the global dense (GD) block is constructed with the dense block in which all the 3×3 convolution kernels are added a dilation rate of 2.

4.4. Global-local siamese architecture

The global feature of an image describes the overall features of the image, including its shape, color, texture and so on. With the development of deep networks, powerful methods for extracting global features are based on deep learning. The features extracted by deep learning are embedded in a higher dimension that contains more contextual information, enabling more powerful image classification. The local features of an image are mainly descriptions of specific regions or edge points. They are a guarantee for good classification results in case some regions of the query image are masked.

In recent years, image classification methods based on global and local features have become a trend, as combining the two features can improve the accuracy of image retrieval. We adopted DenseNet as the backbone, used SGD blocks to extract features, and used the weight share method to integrate global and local features. This method combines the two characteristics in one network for end-to-end training to achieve the greatest accuracy.

In the structure of RockS^{2}Net, original images I_g and transformed critical areas I_l by STN get their respective features by GD of global stream (GD_g) block and GD of local stream (GD_l) block. After regularization in the BN layer respectively, the global features and local features of images are concatenated and classification scores obtained in the FC layer are denoted as F_g, F_l , which are formulated as

$$F_g = GD_g(I_g) \tag{7}$$

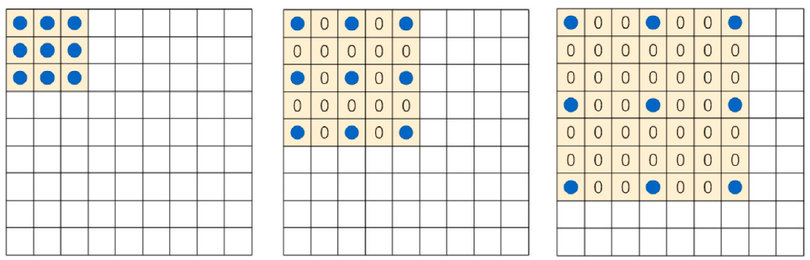

Fig. 10. Conceptual illustration of traditional and dilated convolution; (a) traditional convolution (b) dilated convolution with dilation rate 2; (c) dilated convolution with dilation rate 3.

弱点。该GAP层主要对最后一个卷积层的特征图进行平均值池化，以加速计算并减少训练参数。BN层也是现有研究中提出的一种神经网络架构。随着网络深度增加，每层的特征值分布会趋近激活函数输出区间的上下界，导致梯度消失问题。BN层将各层特征值分布拉向标准正态分布，使激活函数对输入更敏感，通过避免梯度消失来加速收敛。该架构对前层每批输出值进行归一化，确保输出数据均值趋近0、标准差趋近1，同时降低网络对初始化权重的敏感性。BN层带来的这些改变使得网络可采用更大学习率。BN层可表示为式(5)。 x_i 表示神经网络第i层的特征图，$\mathrm{E}(x_i)$ 和 $\mathrm{Var}(x_i)$ 分别代表特征图 x_i 的期望和方差， ε 是为防止除零错误设置的极小正常数（如10^-5）。

$$\widehat{x_i} = \frac{x_i = E(x_i)}{\sqrt{Var(x_i) + \varepsilon}} \tag{5}$$

初始卷积层深度为16，卷积核尺寸为 7×7 ，采用线性整流单元(ReLU)(Nair与Hinton，2010)激活函数后接 2×2 池化层。第二卷积层深度32，卷积核尺寸 5×5 ，使用ReLU激活函数并接 2×2 池化层。第三卷积层深度64，卷积核尺寸 3×3 ，后接ReLU激活层与 2×2 池化层。仿射变换的两个平移参数通过若干卷积层、全局平均池化层、批归一化层及全连接层获得。全连接层权重采用人工指定初始化参数，并选用tanh激活函数确保平移参数初始化为0。旋转参数设为0，通过lambda层设定适当缩放参数。该lambda层基于两个平移参数补充两个零值旋转参数与两个调谐后的缩放参数，这六个参数共同构成仿射变换矩阵。

4.3 Rock S^2 Net框架中的空洞卷积

RockS ^{2} Net框架模型通过采用空洞卷积核替代传统卷积核构建而成。研究表明，空洞卷积层能提升分类任务准确率，是池化层的有效替代方案（Lei等人，2019；Liu等人，2020；Zhang，2022）。通过引入扩张率参数，相同尺寸的卷积核可获得更大的感受野。相应地，在保持相同感受野时，空洞卷积的参数量也可少于标准卷积（Kudo与Aoki，2017）。

为此，我们引入空洞卷积思想，以解决图像分类中因分辨率降低和下采样导致的信息丢失问题。

图10展示了尺寸为 9×9 的图像上常规卷积核与空洞卷积核的对比：(a)为常规 3×3 卷积核，(b)是膨胀率为2的空洞卷积核，通过在(a)中每个点之间插入零值孔洞形成；(c)则是膨胀率为3的卷积核。如图所示，卷积核的感受野在(a)中为 3×3 ，(b)中增至7×7，(c)中达到15×15。随着孔洞插入，感受野不断扩大，但(a)、(b)、(c)三者的参数量保持不变。因此，使用这种扩展卷积核处理图像，可在不增加计算量的情况下获取更多信息。

式(6)展示了空洞卷积原理。其中 k 表示输入卷积核尺寸， D 代表采用的稀释系数， K 指代应用膨胀操作后得到的等效卷积核大小。

$$K = D \times (k-1) + 1 \tag{6}$$

为提取岩石图像全局上下文特征，全局密集(GD)模块通过密集块构建，其中所有 3×3 卷积核均添加了2的膨胀率。

4.4 全局-局部孪生架构

图像全局特征描述其整体特性，包括形状、颜色、纹理等。随着深度网络发展，基于深度学习的强大方法能提取蕴含更高维度上下文信息的特征，实现更强大的图像分类。局部特征主要描述特定区域或边缘点，当查询图像部分区域被遮挡时，这些特征是获得良好分类结果的保障。

近年来，基于全局与局部特征的图像分类方法已成为趋势，因为结合这两种特征能提升图像检索精度。我们采用DenseNet作为主干网络，利用SGD模块提取特征，并通过权重共享方法整合全局与局部特征。该方法将两种特征融合在单一网络中进行端到端训练，以实现最优准确率。

在 RockS^{2}Net 结构中，原始图像 I_g 和经STN变换的关键区域 I_l 分别通过全局流 (GD_g) 模块的GD与局部流 (GD_l) 模块的GD获取特征。经BN层分别正则化后，图像的全局特征与局部特征被拼接，最终在FC层获得的分类得分记为 F_g, F_l ，其表达式为

$$F_g = GD_g(I_g) \tag{7}$$

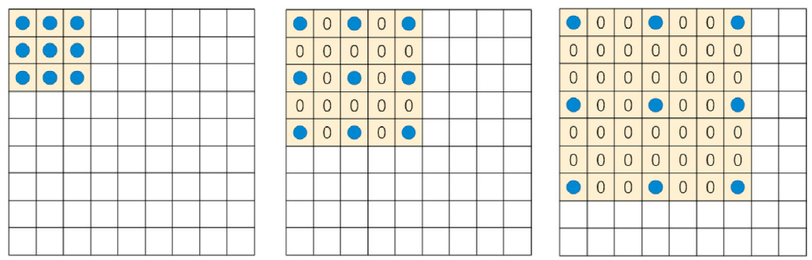

图10. 传统卷积与空洞卷积概念示意图：(a)传统卷积；(b)扩张率为2的空洞卷积；(c)扩张率为3的空洞卷积。

$$F_l = GD_l(I_l) \tag{8}$$

Both of these streams possess identical architectures and utilize shared parameters within the SGD block to minimize parameter computation. The combined classification score is obtained by averaging the values of F_g and F_l as

$$F_s = \frac{F_g + F_l}{2} \tag{9}$$

5. Experiment and evaluation

5.1. Application of the RocksS^{2}Net

5.1.1. Model preprocessing

In this paper, VGG-16 (Simonyan and Zisserman, 2014) is chosen as the baseline to compare with the proposed RocksS^{2}Net, according to the five rock properties. To validate the efficacy of each key components designed in the developed model, DenseNet-121 (Huang et al., 2017), GD block network (GDNet), Siamese DenseNet with Dilated Convolution (DCNet), and Siamese DenseNet with STN (S^{2}DNet), which refers to STN guided double-branch DenseNet-121 with shared parameters, are also compared.

5.1.2. Parameter setting

TensorFlow (Abadi et al., 2016) was used to deploy our model on NVIDIA RTX2080 GPUs. The adaptive moment estimation (Adam) (Kingma and Ba, 2014) was adopted as an optimizer with the learning rate of 10–5, and the cross-entropy loss was the loss function. All the deep learning models were trained with a batch size of 8. As the images were transformed by STN, S^{2}DNet and RocksS^{2}Net were trained with a batch size of 4. In S^{2}DNet and RocksS^{2}Net, a dropout layer (Srivastava et al., 2014) with a rate of 0.5 was added after the BN layer of STN, and the scale parameter of 0.8 was selected in STN after parameter tuning. In addition, the pre-trained weight on the ImageNet (Deng et al., 2009) was applied to the initialization weight of networks in all deep learning models.

6. Performance evaluation and analysis

6.1. Performance evaluation

Rock image classification employs Overall Accuracy (OA) as the accuracy evaluation criterion, utilizing five categories: grain property, clastic property, mechanical genesis property, mixture property, and basic category property, as classification classes, and selected four classical or state-of-the-art computer vision (CV) models, namely DarkNet-53 (Redmon and Farhadi, 2018), EfficientNet (Tan and Le, 2019), ShuffleNet (X. Zhang et al., 2018), and Deep Subdomain Adaptation Network (DSAN) (Zhu et al., 2020). These models encompass not only traditional image classification models but also the latest domain adaptation models.

The outcomes of the experiments are presented in Table 1. Its classification OA of grain property, clastic property, mechanical genesis property, mixture property, and basic category property can reach to 87.14%, 91.92%, 97.75%, 92.14%, and 91.85%, respectively, which are all above 85%. Comparing the proposed model with state-of-the-art models reveals that the RockS^{2}Net model exhibits satisfactory performance. While its accuracy in the 'Basic category' is lower than that of ShuffleNet, its overall accuracy still surpasses that of ShuffleNet.

The outcomes of the experiments are presented in Table 2. The proposed RockS^{2}Net present satisfying classification performance. Its classification OA of grain property, clastic property, mechanical genesis property, mixture property, and basic category property can reach to 87.14%, 91.92%, 97.75%, 92.14%, and 91.85%, respectively, which are all above 85%.

Based on the experimental outcomes presented in Table 1, the classification effect of DenseNet-121 is better than that of VGG-16. DenseNet-121 is used as a model, and then each model is gradually added. The proposed RockS^{2}Net contains two components, STN and SGD block. The classification performance of GDNet, DCNet and S^{2}DNet are better than that of DenseNet-121, as shown in Table 1. Finally, the combination of STN and SGD block achieve the best performance. This implies the necessity of each component for RockS^{2}Net to obtain the best rock image classification accuracy.

Fig. 10 shows some example images correctly classified by DenseNet-121 and RockS^{2}Net. Among them, the spatial distribution features and spectral features are relatively simple images in Fig. 10. In Fig. 11(a)(b), there are uniform spatial distribution features in the coarse crystal rock image and the aggregate rock image. In Fig. 11(c)(d), there are not obvious spatial distribution features in the calcareous rock image and the carbonaceous rock image, but the spectral features are relatively simple. Therefore, the above two models correctly classified such microscopic images of rock sections easily. Fig. 12 shows some example images correctly classified by RockS^{2}Net but wrongly classified by DenseNet-121. In Fig. 12(a)(b), there are complex spatial distribution features in the coarse crystal rock image and the aggregate rock image. In Fig. 12(c)(d), there are complex spectral features in the calcareous rock image and the carbonaceous rock image. Therefore, DenseNet-121 misclassified the four microscopic images of rock sections in Fig. 12. It implies that RockS^{2}Net can extract more sufficient spatial spectrum features from microscopic images of rock sections.

6.2. Analysis on different backbones

From the experimental results, the classification effect of Densent-121 is better than that of VGG-16 in classifying properties of microscopic images of rock sections. The core architecture of DenseNet is the dense block. The input of each layer within the dense block is derived from the output of all preceding layers, which enhances the transmission of features and optimizes the utilization of features of each layer in the

Table 2
Classification results of various rock properties under different models.

Model	Grain	Clastic	Mechanical genesis	Mixture	Basic category
VGG-16	78.93%	86.62%	92.75%	87.75%	86.85%
DenseNet-121	82.32%	88.38%	93.50%	88.75%	89.26%
GDNet	85.71%	90.40%	95.00%	89.64%	90.00%
DCNet	83.29%	88.78%	94.34%	90.32%	90.20%
S^{2}DNet	83.93%	89.65%	95.50%	90.54%	91.48%
RockS^{2}Net	**87.14%**	**91.92%**	**97.75%**	**92.14%**	**91.85%**

Table 1
Classification results of various rock properties under different models.

Model	Grain	Clastic	Mechanical genesis	Mixture	Basic category	AVG
DarkNet-53	80.92%	87.67%	89.88%	86.17%	88.73%	86.67%
EfficientNet	87.71%	90.21%	93.87%	91.53%	92.98%	91.26%
ShuffleNet	85.03%	91.72%	92.86%	89.21%	96.49%	91.06%
DSAN	82.79%	89.65%	96.25%	87.32%	90.37%	89.28%
RockS^{2}Net	**87.14%**	**91.92%**	**97.75%**	**92.14%**	**91.85%**	**92.16%**

$$F_l = GD_l(I_l) \tag{8}$$

这两条流具有相同的架构，并在SGD块内共享参数以减少参数量计算。通过平均 F_g 和 F_l 的值获得综合分类分数

$$F_s = \frac{F_g + F_l}{2} \tag{9}$$

5. 实验与评估

5.1. Rock S^2 网络的应用

5.1.1. 模型预处理

本文选择VGG-16（Simonyan和Zisserman，2014）作为基线模型，与提出的RockS^{2}网络根据五大岩石特性进行对比。为验证所开发模型中各关键组件的有效性，同时比较了DenseNet-121（Huang等，2017）、GD块网络（GDNet）、采用扩张卷积的连体DenseNet（DCNet）以及STN $(\mathrm{S}^2\mathrm{DNet})$ 引导的双分支参数共享DenseNet-121（即STN连体DenseNet）。

5.1.2. 参数设置

采用TensorFlow（Abadi等人，2016年）在NVIDIA RTX2080 GPU上部署模型，使用自适应矩估计优化器Adam（Kingma和Ba，2014年），学习率设为10^-5，并以交叉熵损失作为损失函数。所有深度学习模型训练批次大小为8。经STN转换的图像中，S^2 DNet和Rock S^2 Net采用4的批次规模。在 S^2 DNet和Rock S^2 Net中，STN的BN层后添加了丢弃率为0.5的Dropout层（Srivastava等人，2014年），参数调优后选定STN的缩放参数为0.8。此外，所有深度学习模型的网络初始化权重均采用ImageNet（Deng等人，2009年）预训练参数。

6. 性能评估与分析

6.1 性能评估

岩石图像分类采用总体准确率(OA)作为评估标准，选取颗粒属性、碎屑属性、机械成因属性、混合属性和基本类别属性五类作为分类类别，并选用DarkNet-53(Redmon与Farhadi，2018)、EfficientNet(Tan和Le，2019)、ShuffleNet(X. Zhang等，2018)以及深度子域自适应网络(DSAN)(Zhu等，2020)四种经典或前沿计算机视觉模型。这些模型既涵盖传统图像分类模型，也包含最新的域自适应模型。

实验结果如表1所示。其在颗粒属性、碎屑属性、机械成因属性、混合属性和基本类别属性上的分类OA分别可达 $87.14\%, 91.92\%, 97.75\%, 92.14\%$ 和91.85%，均高于85%。将所提模型与前沿模型对比

表1
不同模型下各类岩石属性的分类结果。

模型	颗粒	碎屑	机械成因	混合物	基本类别	AVG
暗网-53	80.92%	87.67%	89.88%	86.17%	88.73%	86.67%
高效网络	87.71%	90.21%	93.87%	91.53%	92.98%	91.26%
混洗网络	85.03%	91.72%	92.86%	89.21%	96.49%	91.06%
DSAN	82.79%	89.65%	96.25%	87.32%	90.37%	89.28%
岩石 ^{2} 网络	**87.14%**	**91.92%**	**97.75%**	**92.14%**	**91.85%**	**92.16%**

模型分析表明，RockS ^{2} Net模型展现出令人满意的性能。尽管其在"基础类别"上的准确率低于ShuffleNet，但总体准确率仍超越ShuffleNet。

实验结果如表2所示。提出的RockS ^{2} Net具有优异的分类性能，其颗粒属性、碎屑属性、机械成因属性、混合属性及基础类别属性的分类总体准确率分别可达 $87.14\%, 91.92\%, 97.75\%, 92.14\%$ 和91.85%，均超过85%。

根据表1所示的实验结果，DenseNet-121的分类效果优于VGG-16。本研究以DenseNet-121为基础模型，逐步添加各组件。提出的RockS ^{2} Net包含STN和SGD模块两个核心部分。如表1所示，GDNet、DCNet及 S^2 DNet的分类性能均超越DenseNet-121。最终，STN与SGD模块的组合实现了最优性能，这表明RockS ^{2} Net中每个组件对于获得最佳岩石图像分类精度都是必要的。

图10展示了DenseNet-121和RockS ^{2} 网络正确分类的部分示例图像。其中，图10中空间分布特征和光谱特征相对简单的图像被准确识别。图11(a)(b)中，粗晶岩图像和集合岩图像具有均匀的空间分布特征；图11(c)(d)中，钙质岩图像和碳质岩图像虽无明显空间分布特征，但光谱特征较为简单。因此上述两种模型都能轻松正确分类这类岩石薄片显微图像。图12呈现了RockS ^{2} 网络正确分类而DenseNet-121误判的示例图像：图12(a)(b)中粗晶岩与集合岩图像具有复杂的空间分布特征，图12(c)(d)中钙质岩与碳质岩图像则存在复杂的光谱特征。这表明 RockS^2 网络能从岩石薄片显微图像中提取更充分的空间-光谱联合特征。

6.2 不同骨干网络分析

实验结果表明，在岩石切片显微图像属性分类任务中，DenseNet-121的分类效果优于VGG-16。DenseNet的核心架构是密集连接块，其内部每一层的输入均来自前面所有层的输出，这种设计增强了特征传递能力，并优化了各层特征的利用率。

表2
不同模型下各类岩石属性的分类结果

模型	颗粒	碎屑	机械成因	混合物	基础类别
VGG-16	78.93%	86.62%	92.75%	87.75%	86.85%
密集网络-121	82.32%	88.38%	93.50%	88.75%	89.26%
GD网络	85.71%	90.40%	95.00%	89.64%	90.00%
DC网络	83.29%	88.78%	94.34%	90.32%	90.20%
S^2 D网络	83.93%	89.65%	95.50%	90.54%	91.48%
岩石 ^{2} 网络	**87.14%**	**91.92%**	**97.75%**	**92.14%**	**91.85%**

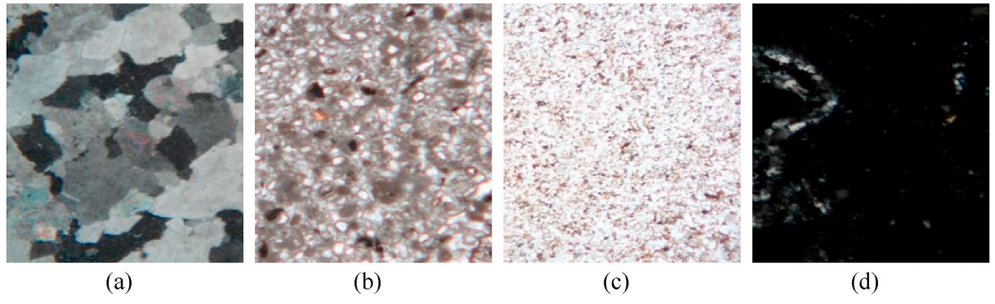

Fig. 11. Example images correctly classified by both DenseNet-121 and RockS^{2}Net. (a) Coarse crystal. (b) Aggregate. (c) Calcareous. (d) Carbonaceous.

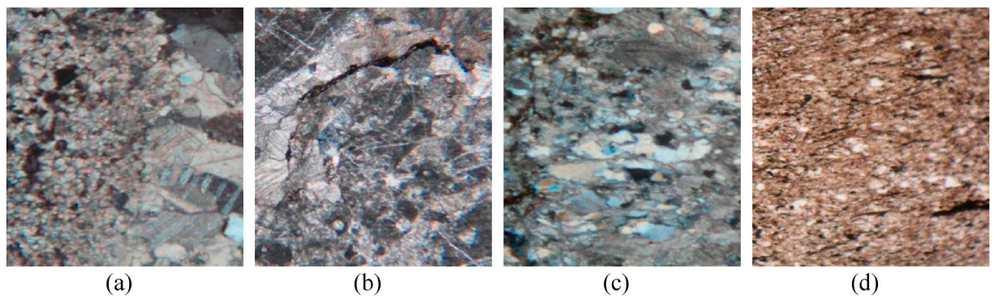

Fig. 12. Example images correctly classified by RockS^{2}Net but incorrectly classified by DenseNet-121. (a) Coarse crystal. (b) Aggregate. (c) Calcareous. (d) Carbonaceous.

network. VGG is a layer-by-layer network architecture, and the input of each layer can only come from the output of the preceding layer, which tends to lead to the loss of features in the middle layers. In microscopic images of rock sections, DenseNet usually loses fewer attribute features than VGG. Therefore, it is better to use DenseNet for feature extraction.

Fig. 13 shows the confusion matrixes of VGG-16 and DenseNet-121 for the classification of clastic rock properties. In addition to the type of properties as fine sand and others, the classification effect of DenseNet-121 is better than VGG-16. The result demonstrates that the architecture of DenseNet can effectively extract the classification features related to the clastic property features from microscopic images of rock sections, and can improve the classification effect by its shorter connections between layers.

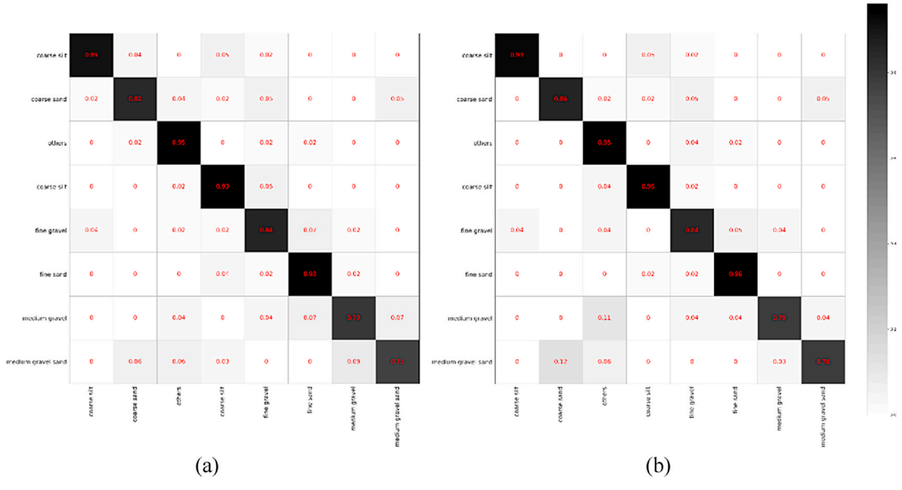

Fig. 13. Confusion matrixes on the clastic property classification. (a) VGG-16. (b) DenseNet-121.

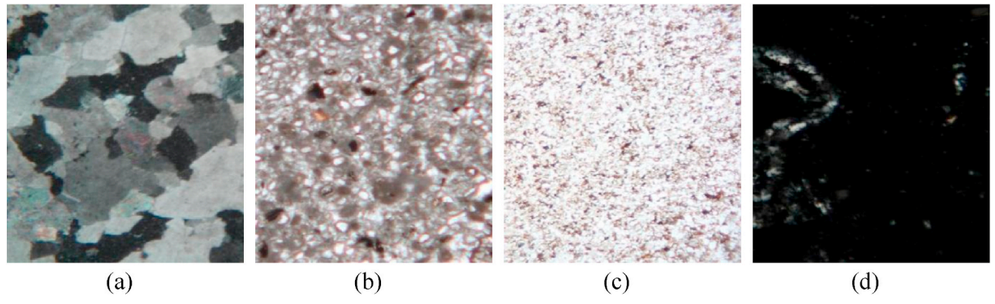

图11. DenseNet-121与RockS^{2}网络均正确分类的示例图像。(a)粗晶结构 (b)集合体 (c)钙质 (d)碳质

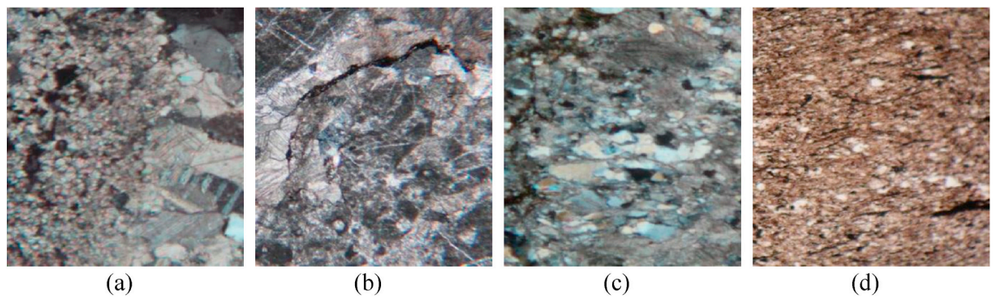

图12. RockS^{2}网络正确分类但DenseNet-121误分类的示例图像。(a)粗晶结构 (b)集合体 (c)钙质 (d)碳质

网络架构。VGG采用逐层连接结构，每层输入仅能来自前一层输出，易导致中间层特征丢失。在岩石薄片显微图像中，DenseNet通常比VGG能保留更多属性特征，因此更适合用于特征提取。

图13展示了VGG-16与DenseNet-121在碎屑岩属性分类中的混淆矩阵。除细砂岩等属性类型外，其分类效果DenseNet-121优于VGG-16。结果表明，DenseNet的架构能有效从岩石切片显微图像中提取与碎屑属性特征相关的分类特征，并通过更短的层间连接提升分类效果。

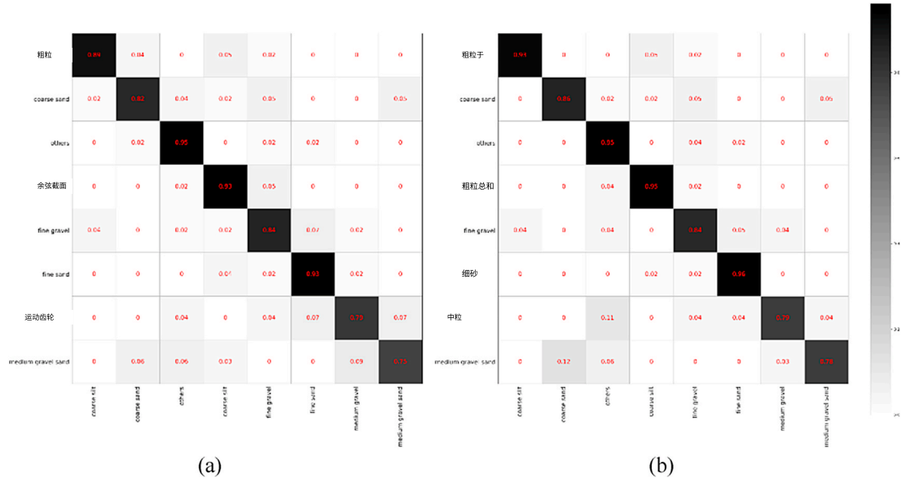

图13. 碎屑属性分类的混淆矩阵。(a)VGG-16网络。(b)DenseNet-121网络。

6.3. Analysis on global contextual acquisition

Conventional image classification algorithms typically use pooling and convolutional layers to increase the receptive area. However, this diminishes the dimensionality of feature map, which results in the loss of accuracy. The larger the receptive area is, the larger the range of the original image it can touch, which implies more global contextual features can be captured. The distribution of the various property features in microscopic images of rock sections is often inhomogeneous, while GDNet is capable to extract more global property features.

In the experiments on the mixture properties, the results of DenseNet-121 and GDNet are compared. Four rock images which are incorrectly classified by DenseNet-121 and correctly classified by GDNet are selected, as shown in Fig. 14. Among them, Fig. 14(a) is silty property, but DenseNet-121 incorrectly classifies it as containing silty. The difference between images of silty rocks and containing silty rocks is the amount of silt content. Fig. 14(a) is silty property overall but there is silt at the top right corner of the image. However, other white mixture conceals silty property of the upper right corner, which makes it difficult for the DenseNet-121 to extract the silt property features in this part well. When using GDNet, the receptive field is increased, and the range of original images contacted is expanded. The silt property features in the upper right corner can be extracted by GDNet. Similarly, Fig. 14(b) is calcareous property, but the DenseNet-121 classifies it as iron property. In this figure, the red iron area occupies the most central position, but the surrounding calcium area is larger overall. For these non-uniform distributed features, the architecture of GDNet can extract more global attribute features and classify it correctly. Fig. 14(c) is carbonaceous, and Fig. 14(d) is iron. Because the receptive field of the convolution kernel of DenseNet-121 is not large enough and the global property features cannot be extracted, they are incorrectly classified as calcareous and carbonaceous, respectively. The result demonstrates that GDNet can extract more contextual information from microscopic images of rock sections, while SGD block can extract more global and local contextual information.

6.4. Analysis on critical areas capture

To validate the effectiveness of STN, the experimental results of grain properties were analyzed. The experimental results of GDNet and RockS^{2}Net are compared. Four images which are incorrectly classified by GDNet and correctly classified by RockS^{2}Net are shown in Fig. 15. In addition, their critical areas transformed by STN are also visualized in the red boxes. As shown in Fig. 15(a) and (b), the coarse crystal property was misclassified as medium crystal property by the GDNet. The grain size of coarse crystal is larger than that of medium crystal. After the cropping operation by STN, a part of the smaller grain size area under the image is cropped out and the transformed image is more concentrated in the larger grain size areas. In the subsequent process of feature extraction using DenseNet-121, global contextual features of larger grain size areas can be extracted, and thus coarser crystal features can be obtained in the final feature average layer. Fig. 15(c) and (d) are micro crystal property, which misclassified as mud-crystal features by GDNet. The grain size of micro crystal and mud crystal are very small and have very small difference. After transforming and enlarging the micro crystal image by STN, grain size features are amplified to a certain extent to be correctly classified. If GDNet is used to classify these two images directly, it will be easier to misclassify them.

As can be seen, RockS^{2}Net using STN to crop and enlarge rock images can crop out the areas which have less relationship with rock image classification. It reduces the influence of the features in such areas on rock image classification, extracting corresponding features. At the same time, some rock features are enhanced to some extent, making it easier to distinguish some of the more difficult features.

7. Conclusions

This study proposes a RockS^{2}Net for rock image classification. RockS^{2}Net extracts and fuses both local features and global features as well as extending the attribute features of the region of interest to identify the rock property features accurately. It introduces a novel method for deep learning-based classification of rock images. In addition, CHN-Rock images dataset was constructed, which provided the research community with a large-scale rock image benchmark. We designed a case study to demonstrate the performance of the RockS^{2}Net. The experimental results demonstrate that the RockS^{2}Net framework achieved superior performance on the CHN-Rock image dataset compared to several existing methods, proving that integrating both structural characteristic and attribute features of the region helps identify the rock property features. Future work aims to combine self-attention methods to improve the classification effect in microscopic images of rock sections.

Authorship contribution statement

Qiqi Zhu: Conceptualization, Methodology, Writing – review & editing. Sai Wang: Software, Visualization, Data curation, Writing – original draft. Shun Tong: Data curation. Liangbin Yin: Validation. Kunlun Qi: Investigation, Resources, Project administration. Qingfeng Guan: Supervision, Funding acquisition, Writing – review & editing.

Funding

The authors would like to thank the editor, the associate editor, and the anonymous reviewers for their helpful comments and advice. This work was supported by the National Key Research and Development Program of China (Grant No. 2022YFB3903402) and the National Natural Science Foundation of China (Grant No. 42271413).

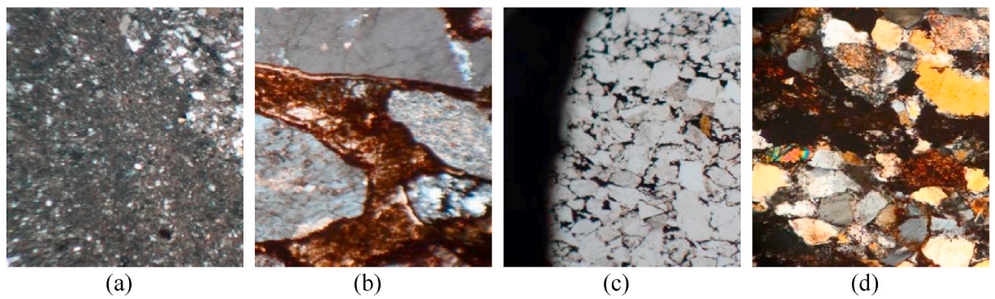

Fig. 14. Example images that are misclassified by DenseNet-121 and correctly classified by GDNet. (a) Silty. (b) Calcareous. (c) Carbonaceous. (d) Iron.

6.3 全局上下文特征获取分析

传统图像分类算法通常通过池化层和卷积层扩大感受野，但这会降低特征图维度并导致精度损失。感受野越大，其覆盖的原始图像范围越广，意味着能捕获更多全局上下文特征。岩石薄片显微图像中各类物性特征的分布往往不均匀，而GDNet能够提取更具全局性的物性特征。

在混合属性实验中，对比了DenseNet-121与GDNet的分类结果。图14展示了四幅被DenseNet-121误判但被GDNet正确识别的岩石图像。其中，图14(a)实际为粉砂质属性，但DenseNet-121误判为含粉砂质——二者的区别在于粉砂含量。该图像整体呈粉砂质特征，但右上角存在粉砂区域，其他白色混合物掩盖了该区域的粉砂属性，导致DenseNet-121难以有效提取该部分特征。GDNet通过扩大感受野和原始图像接触范围，成功捕捉到右上角的粉砂特征。同理，图14(b)实际为钙质属性，但被误判为铁质属性：虽然红色铁质区域占据中心位置，但周围钙质区域整体占比更大。GDNet的网络结构能提取这类非均匀分布特征的全局属性。图14(c)碳质属性和图14(d)铁质属性图像，因DenseNet-121卷积核感受野不足而分别被误判为钙质和碳质。实验证明GDNet能从岩石切片显微图像中提取更丰富的上下文信息，其SGD模块可同步捕获全局与局部特征。

6.4 关键区域捕捉分析

为验证STN的有效性，对晶粒特性实验结果进行分析。将GDNet与RockS^{2}Net的实验结果进行对比。图15展示了四幅被GDNet错误分类但被RockS^{2}Net正确分类的图像，经STN转换后的关键区域在红色方框中可视化呈现。如图15(a)(b)所示，GDNet将粗晶特性误判为中晶特性。粗晶的晶粒尺寸大于中晶，经过STN裁剪操作后，图像下方较小晶粒尺寸区域被裁切，转换后的图像更集中于较大晶粒尺寸区域。在使用DenseNet-121进行特征提取的后续过程中，较大晶粒的全局上下文特征通过提取晶粒尺寸区域，可在最终特征平均层中获得更粗大的晶体特征。图15(c)和(d)展示了微晶特性，这些特性被GDNet误判为泥晶特征。微晶与泥晶的晶粒尺寸均非常微小且差异极小。经STN变换放大后，微晶图像的晶粒特征被一定程度放大，从而得以正确分类。若直接使用GDNet对这两类图像进行分类，则更易产生误判。

可见，RockS^{2}Net利用STN裁剪放大岩石图像时，能剔除与岩石分类关联性较弱的区域。这降低了此类区域特征对岩石图像分类的干扰，同时提取出对应特征。部分岩石特征还得到一定程度的增强，使得某些较难区分的特征更易辨识。

7. 结论

本研究提出了一种用于岩石图像分类的RockS^{2}Net网络。该网络通过提取并融合局部特征与全局特征，同时扩展感兴趣区域的属性特征，实现了对岩石特性特征的精准识别。该研究为基于深度学习的岩石图像分类提供了新方法。此外，研究团队构建了CHN-Rock图像数据集，为学界提供了大规模岩石图像基准。通过案例研究验证了RockS^{2}Net的性能，实验结果表明该框架在CHN-Rock数据集上优于现有多种方法，证实了整合区域结构特征与属性特征对识别岩石特性的有效性。未来工作将结合自注意力方法以提升岩石薄片显微图像分类效果。

作者贡献声明

朱琪琪：概念设计、方法论、文稿审阅与编辑。王赛：软件开发、可视化、数据整理、初稿撰写。童顺：数据整理。尹良斌：验证。祁昆仑：调研、资源整合、项目管理。关清风：监督指导、资金获取、文稿审阅与编辑。

基金资助

作者谨感谢编辑、副编辑及匿名评审专家提出的宝贵意见与建议。本研究得到国家重点研发计划（项目编号：2022YFB3903402）和国家自然科学基金（项目编号：42271413）的资助。

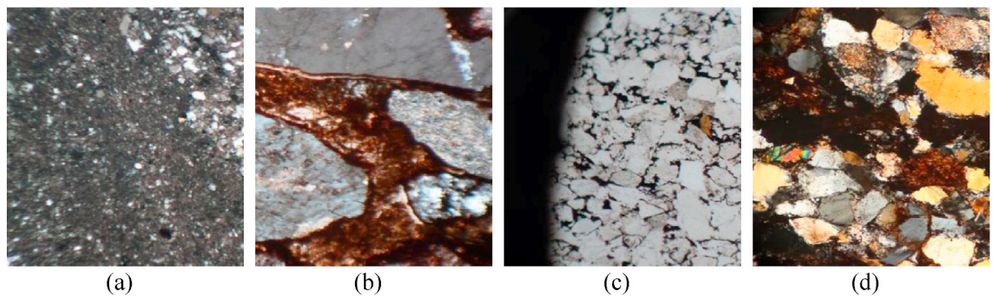

图14. DenseNet-121误分类而GDNet正确分类的示例图像。(a)粉砂质 (b)钙质 (c)碳质 (d)铁质

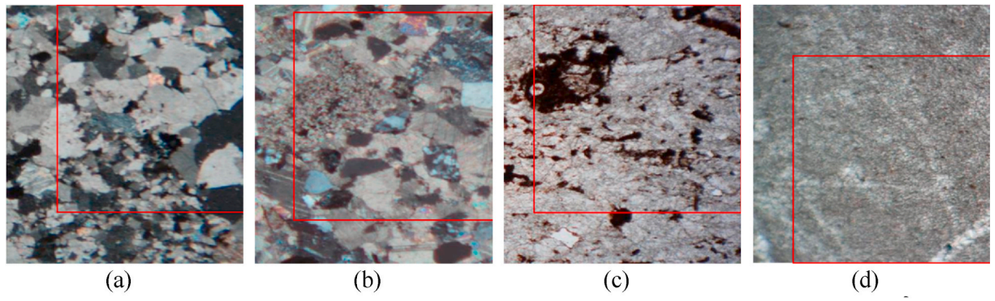

Fig. 15. Example images that are misclassified by GDNet and correctly classified by RockS^{2}Net. (a), (b) Coarse crystal. (c), (d) Micro crystal. Note: The critical areas of images transformed by STN were in the red boxes.

Code availability section

Name of the code/library: RockS^{2}Net
Contact: e-mail and phone number:20171003123 @cug.edu.cn; 15827250895.
Hardware requirements: NVIDIA RTX 2080 Ti GPU、PyTorch 1.13.1.
Program language: python.
Software required: PyCharm、MobaXterm、WinSCP.
Program size: 29 KB.
The source codes are available for downloading at the link: https://github.com/sara084/RockSlice.

Declaration of competing interest

All authors declare that No conflict of interest exists. I would like to declare on behalf of my co-authors that the work described was original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part. All the authors listed have approved the manuscript that is enclosed.

Data availability

The authors do not have permission to share data.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., 2016. TensorFlow: a system for Large-Scale machine learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), pp. 265–283.
Baraboshkin, E.E., Ismailova, L.S., Orlov, D.M., Zhukovskaya, E.A., Kalmykov, G.A., Khotylev, O.V., Baraboshkin, E.Y., Korotеev, D.A., 2020. Deep convolutions for in-depth automated rock typing. Comput. Geosci. 135, 104330.
Battaglia, P.W., Hamrick, J.B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., 2018. Relational Inductive Biases, Deep Learning, and Graph Networks arXiv preprint arXiv: 1806.01261.
Chatterjee, S., 2013. Vision-based rock-type classification of limestone using multi-class support vector machine. Appl. Intell. 39, 14–27.
Cherkashina, T.Y., Panteeva, S.V., Pashkova, G.V., 2014. Applicability of direct total reflection X-ray fluorescence spectrometry for multielement analysis of geological and environmental objects. Spectrochim. Acta B Atom Spectrosc. 99, 59–66.
Dawson, H.L., Dubrule, O., John, C.M., 2023. Impact of dataset size and convolutional neural network architecture on transfer learning for carbonate rock classification. Comput. Geosci. 171, 105284.
Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Ieee, pp. 248–255.
Dunlop, H., 2006. Automatic Rock Detection and Classification in Natural Scenes. Masters Thesis. Carnegie Mellon University.
Guojian, C., Peisong, L., 2021. Rock thin-section image classification based on residual neural network. In: 2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP). IEEE, pp. 521–524.
Hao, H., Jiang, Z., Ge, S., Wang, C., Gu, Q., 2022. Siamese Adversarial Network for image classification of heavy mineral grains. Comput. Geosci. 159, 105016.
Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.
Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning. PMLR, pp. 448–456.
Karimpouli, S., Tahmasebi, P., 2019. Segmentation of digital rock images using deep convolutional autoencoder networks. Comput. Geosci. 126, 142–150.
Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization arXiv preprint arXiv:1412.6980.
Kudo, Y., Aoki, Y., 2017. Dilated convolutions for image classification and object localization. In: 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA). IEEE, pp. 452–455.
Kuiper, K.F., Deino, A., Hilgen, F.J., Krijgsman, W., Renne, P.R., Wijbrans, J., 2008. Synchronizing rock clocks of Earth history. Science 320, 500–504.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
Lei, X., Pan, H., Huang, X., 2019. A dilated CNN model for image classification. IEEE Access 7, 124087–124095.
Lepistö, L., Kunttu, I., Visa, A., 2005a. Color-based classification of natural rock images using classifier combinations. In: Scandinavian Conference on Image Analysis. Springer, pp. 901–909.
Lepistö, L., Kunttu, I., Visa, A.J., 2005b. Rock image classification using color features in Gabor space. J. Electron. Imag. 14, 040503.
Li, N., Hao, H., Gu, Q., Wang, D., Hu, X., 2017. A transfer learning method for automatic identification of sandstone microscopic images. Comput. Geosci. 103, 111–121.
Li, X., Wang, Q., 2019. Prediction of surrounding rock classification of highway tunnel based on PSO-SVM. In: 2019 International Conference on Robots & Intelligent System (ICRIS). IEEE, pp. 443–446.
Liang, Y., Cui, Q., Luo, X., Xie, Z., 2021. Research on classification of fine-grained rock images based on deep learning. Comput. Intell. Neurosci. 2021.
Lin, M., Chen, Q., Yan, S., 2013. Network in Network arXiv preprint arXiv:1312.4400.
Liu, Q., Kampffmeyer, M., Jenssen, R., Salberg, A.-B., 2020. Dense dilated convolutions' merging network for land cover classification. IEEE Trans. Geosci. Rem. Sens. 58, 6309–6320.
Mkwelo, S., 2004. A Machine Vision-Based Approach to Measuring the Size Distribution of Rocks on a Conveyor Belt. University of Cape, Town.
Młynarczuk, M., Górszczyk, A., Ślipek, B., 2013. The application of pattern recognition in the automatic classification of microscopic rock images. Comput. Geosci. 60, 126–133.
Momma, E., Ono, T., Ishii, H., 2006. Rock classification by types and degrees of weathering. In: 2006 SICE-ICASE International Joint Conference. IEEE, pp. 149–152.
Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted Boltzmann machines. Icml.
Pascual, A.D.P., Shu, L., Szoke-Sieswerda, J., McIsaac, K., Osinski, G., 2019. Towards natural scene rock image classification with convolutional neural networks. In: 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE). IEEE, pp. 1–4.
Patel, A.K., Chatterjee, S., Gorai, A.K., 2019. Effect on the performance of a support vector machine based machine vision system with dry and wet ore sample images in classification and grade prediction. Pattern Recogn. Image Anal. 29, 309–324.
Patel, A.K., Chatterjee, S., Gorai, A.K., 2017. Development of machine vision-based ore classification model using support vector machine (SVM) algorithm. Arabian J. Geosci. 10, 1–16.
Perez, C.A., Estévez, P.A., Vera, P.A., Castillo, L.E., Aravena, C.M., Schulz, D.A., Medina, L.E., 2011. Ore grade estimation by feature selection and voting using boundary detection in digital image analysis. Int. J. Miner. Process. 101, 28–36.
Qin, J., He, Z.-S., 2005. A SVM face recognition method based on Gabor-featured key points. In: 2005 International Conference on Machine Learning and Cybernetics. IEEE, pp. 5144–5149.
Ran, X., Xue, L., Zhang, Y., Liu, Z., Sang, X., He, J., 2019. Rock classification from field image patches analyzed using a deep convolutional neural network. Mathematics 7, 755.

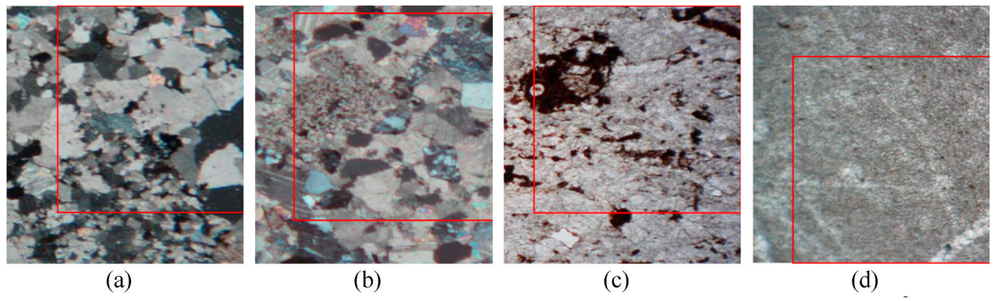

图15. GDNet误分类而RockS ^{2} Net正确分类的示例图像。(a)、(b)粗晶；(c)、(d)微晶。注：经STN变换的图像关键区域已用红框标出。

代码可用性声明

代码库/工具名称： RockS^2 Net
联系方式：电子邮箱及电话号码：20171003123@cug.edu.cn；15827250895
硬件要求：NVIDIA RTX 2080 Ti显卡、PyTorch1.13.1框架

编程语言：Python
所需软件：PyCharm、MobaXterm、WinSCP。
程序大小： 29KB 。
源代码可通过以下链接下载：https://github.com/sara084/RockSlice。

利益冲突声明

所有作者声明不存在利益冲突。我谨代表合著者声明，所述工作为原创研究，此前未曾发表，且未以任何形式考虑在其他地方发表。所有列出的作者均已认可随附的稿件。

数据可用性

作者未获得数据共享授权。

参考文献

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., 2016. TensorFlow: a system for Large-Scale machine learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), pp. 265–283.
Baraboshkin, E.E., Ismailova, L.S., Orlov, D.M., Zhukovskaya, E.A., Kalmykov, G.A., Khotylev, O.V., Baraboshkin, E.Y., Korotеev, D.A., 2020. Deep convolutions for in-depth automated rock typing. Comput. Geosci. 135, 104330.
Battaglia, P.W., Hamrick, J.B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., 2018. Relational Inductive Biases, Deep Learning, and Graph Networks arXiv preprint arXiv: 1806.01261.
Chatterjee, S., 2013. Vision-based rock-type classification of limestone using multi-class support vector machine. Appl. Intell. 39, 14–27.
Cherkashina, T.Y., Panteeva, S.V., Pashkova, G.V., 2014. Applicability of direct total reflection X-ray fluorescence spectrometry for multielement analysis of geological and environmental objects. Spectrochim. Acta B Atom Spectrosc. 99, 59–66.
Dawson, H.L., Dubrule, O., John, C.M., 2023. Impact of dataset size and convolutional neural network architecture on transfer learning for carbonate rock classification. Comput. Geosci. 171, 105284.
Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Ieee, pp. 248–255.
Dunlop, H., 2006. Automatic Rock Detection and Classification in Natural Scenes. Masters Thesis. Carnegie Mellon University.
Guojian, C., Peisong, L., 2021. Rock thin-section image classification based on residual neural network. In: 2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP). IEEE, pp. 521–524.
Hao, H., Jiang, Z., Ge, S., Wang, C., Gu, Q., 2022. Siamese Adversarial Network for image classification of heavy mineral grains. Comput. Geosci. 159, 105016.
Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.
Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning. PMLR, pp. 448–456.
Karimpouli, S., Tahmasebi, P., 2019. Segmentation of digital rock images using deep convolutional autoencoder networks. Comput. Geosci. 126, 142–150.
Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization arXiv preprint arXiv:1412.6980.
Kudo, Y., Aoki, Y., 2017. Dilated convolutions for image classification and object localization. In: 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA). IEEE, pp. 452–455.
Kuiper, K.F., Deino, A., Hilgen, F.J., Krijgsman, W., Renne, P.R., Wijbrans, J., 2008. Synchronizing rock clocks of Earth history. Science 320, 500–504.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
Lei, X., Pan, H., Huang, X., 2019. A dilated CNN model for image classification. IEEE Access 7, 124087–124095.
Lepistö, L., Kunttu, I., Visa, A., 2005a. Color-based classification of natural rock images using classifier combinations. In: Scandinavian Conference on Image Analysis. Springer, pp. 901–909.
Lepistö, L., Kunttu, I., Visa, A.J., 2005b. Rock image classification using color features in Gabor space. J. Electron. Imag. 14, 040503.
Li, N., Hao, H., Gu, Q., Wang, D., Hu, X., 2017. A transfer learning method for automatic identification of sandstone microscopic images. Comput. Geosci. 103, 111–121.
Li, X., Wang, Q., 2019. Prediction of surrounding rock classification of highway tunnel based on PSO-SVM. In: 2019 International Conference on Robots & Intelligent System (ICRIS). IEEE, pp. 443–446.
Liang, Y., Cui, Q., Luo, X., Xie, Z., 2021. Research on classification of fine-grained rock images based on deep learning. Comput. Intell. Neurosci. 2021.
Lin, M., Chen, Q., Yan, S., 2013. Network in Network arXiv preprint arXiv:1312.4400.
Liu, Q., Kampffmeyer, M., Jenssen, R., Salberg, A.-B., 2020. Dense dilated convolutions' merging network for land cover classification. IEEE Trans. Geosci. Rem. Sens. 58, 6309–6320.
Mkwelo, S., 2004. A Machine Vision-Based Approach to Measuring the Size Distribution of Rocks on a Conveyor Belt. University of Cape, Town.
Młynarczuk, M., Górszczyk, A., Ślipek, B., 2013. The application of pattern recognition in the automatic classification of microscopic rock images. Comput. Geosci. 60, 126–133.
Momma, E., Ono, T., Ishii, H., 2006. Rock classification by types and degrees of weathering. In: 2006 SICE-ICASE International Joint Conference. IEEE, pp. 149–152.
Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted Boltzmann machines. Icml.
Pascual, A.D.P., Shu, L., Szoke-Sieswerda, J., McIsaac, K., Osinski, G., 2019. Towards natural scene rock image classification with convolutional neural networks. In: 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE). IEEE, pp. 1–4.
Patel, A.K., Chatterjee, S., Gorai, A.K., 2019. Effect on the performance of a support vector machine based machine vision system with dry and wet ore sample images in classification and grade prediction. Pattern Recogn. Image Anal. 29, 309–324.
Patel, A.K., Chatterjee, S., Gorai, A.K., 2017. Development of machine vision-based ore classification model using support vector machine (SVM) algorithm. Arabian J. Geosci. 10, 1–16.
Perez, C.A., Estévez, P.A., Vera, P.A., Castillo, L.E., Aravena, C.M., Schulz, D.A., Medina, L.E., 2011. Ore grade estimation by feature selection and voting using boundary detection in digital image analysis. Int. J. Miner. Process. 101, 28–36.
Qin, J., He, Z.-S., 2005. A SVM face recognition method based on Gabor-featured key points. In: 2005 International Conference on Machine Learning and Cybernetics. IEEE, pp. 5144–5149.
Ran, X., Xue, L., Zhang, Y., Liu, Z., Sang, X., He, J., 2019. Rock classification from field image patches analyzed using a deep convolutional neural network. Mathematics 7, 755.

Redmon, J., Farhadi, A., 2018. Yolov3: an Incremental Improvement arXiv preprint arXiv:1804.02767.
Rollinson, H.R., 2014. Using Geochemical Data: Evaluation, Presentation, Interpretation. Routledge.
Seng, D., Chen, W., 2009. Application of RS theory and SVM in the ore-rock classification. In: 2009 International Conference on Computational Intelligence and Software Engineering. IEEE, pp. 1–4.
Shang, C., Barnes, D., 2012. Support vector machine-based classification of rock texture images aided by efficient feature selection. In: The 2012 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 1–8.
Sharif, H., Ralchenko, M., Samson, C., Ellery, A., 2015. Autonomous rock classification using Bayesian image analysis for rover-based planetary exploration. Comput. Geosci. 83, 153–167.
Shu, L., McIsaac, K., Osinski, G.R., Francis, R., 2017. Unsupervised feature learning for autonomous rock image classification. Comput. Geosci. 106, 10–17.
Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition arXiv preprint arXiv:1409.1556.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.
Su, C., Xu, S., Zhu, K., Zhang, X., 2020. Rock classification in petrographic thin section images based on concatenated convolutional neural networks. Earth Science Informatics 13, 1477–1484.
Sun, A., Lim, E.-P., Ng, W.-K., 2002. Web classification using support vector machine. In: Proceedings of the 4th International Workshop on Web Information and Data Management, pp. 96–99.
Swain, P.H., Hauska, H., 1977. The decision tree classifier: design and potential. IEEE Trans. Geosci. Electron. 15, 142–147.
Tan, M., Le, Q., 2019. Efficientnet: rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. PMLR, pp. 6105–6114.
Wang, Y., Sun, S., 2021. Image-based rock typing using grain geometry features. Comput. Geosci. 149, 104703.
Zhang, H., 2004. The Optimality of Naive Bayes. Aa, vol. 1, p. 3.
Zhang, Q., 2022. A novel ResNet101 model based on dense dilated convolution for image classification. SN Appl. Sci. 4, 1–13.
Zhang, X., Zhou, X., Lin, M., Sun, J., 2018. Shufflenet: an extremely efficient convolutional neural network for mobile devices. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6848–6856.
Zhang, Y., Li, M., Han, S., 2018. Automatic identification and classification in lithology based on deep learning in rock images. Yanshi Xuebao/Acta Petrologica Sinica 34, 333–342.
Zhu, Y., Bai, L., Peng, W., Zhang, X., Luo, X., 2018. Depthwise separable convolution feature learning for homogeneous rock image classification. In: International Conference on Cognitive Systems and Signal Processing. Springer, pp. 165–176.
Zhu, Y., Zhuang, F., Wang, J., Ke, G., Chen, J., Bian, J., Xiong, H., He, Q., 2020. Deep subdomain adaptation network for image classification. IEEE Transact. Neural Networks Learn. Syst. 32, 1713–1722.

Z. Qiqi et al. Computers and Geosciences 185 (2024) 105560

Redmon, J., Farhadi, A., 2018. Yolov3: an Incremental Improvement arXiv preprint arXiv:1804.02767.
Rollinson, H.R., 2014. Using Geochemical Data: Evaluation, Presentation, Interpretation. Routledge.
Seng, D., Chen, W., 2009. Application of RS theory and SVM in the ore-rock classification. In: 2009 International Conference on Computational Intelligence and Software Engineering. IEEE, pp. 1–4.
Shang, C., Barnes, D., 2012. Support vector machine-based classification of rock texture images aided by efficient feature selection. In: The 2012 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 1–8.
Sharif, H., Ralchenko, M., Samson, C., Ellery, A., 2015. Autonomous rock classification using Bayesian image analysis for rover-based planetary exploration. Comput. Geosci. 83, 153–167.
Shu, L., McIsaac, K., Osinski, G.R., Francis, R., 2017. Unsupervised feature learning for autonomous rock image classification. Comput. Geosci. 106, 10–17.
Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition arXiv preprint arXiv:1409.1556.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.
Su, C., Xu, S., Zhu, K., Zhang, X., 2020. Rock classification in petrographic thin section images based on concatenated convolutional neural networks. Earth Science Informatics 13, 1477–1484.
Sun, A., Lim, E.-P., Ng, W.-K., 2002. Web classification using support vector machine. In: Proceedings of the 4th International Workshop on Web Information and Data Management, pp. 96–99.
Swain, P.H., Hauska, H., 1977. The decision tree classifier: design and potential. IEEE Trans. Geosci. Electron. 15, 142–147.
Tan, M., Le, Q., 2019. Efficientnet: rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. PMLR, pp. 6105–6114.
Wang, Y., Sun, S., 2021. Image-based rock typing using grain geometry features. Comput. Geosci. 149, 104703.
Zhang, H., 2004. The Optimality of Naive Bayes. Aa, vol. 1, p. 3.
Zhang, Q., 2022. A novel ResNet101 model based on dense dilated convolution for image classification. SN Appl. Sci. 4, 1–13.
Zhang, X., Zhou, X., Lin, M., Sun, J., 2018. Shufflenet: an extremely efficient convolutional neural network for mobile devices. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6848–6856.
Zhang, Y., Li, M., Han, S., 2018. Automatic identification and classification in lithology based on deep learning in rock images. Yanshi Xuebao/Acta Petrologica Sinica 34, 333–342.
Zhu, Y., Bai, L., Peng, W., Zhang, X., Luo, X., 2018. Depthwise separable convolution feature learning for homogeneous rock image classification. In: International Conference on Cognitive Systems and Signal Processing. Springer, pp. 165–176.
Zhu, Y., Zhuang, F., Wang, J., Ke, G., Chen, J., Bian, J., Xiong, H., He, Q., 2020. Deep subdomain adaptation network for image classification. IEEE Transact. Neural Networks Learn. Syst. 32, 1713–1722.

12