

RockS²Net: 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 University 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 RockS²Net 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 RockS²Net 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).

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 Tahmasbi, 2019; Mlynarczuk 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 (RockS²Net) for rock image classification. The RockS²Net 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

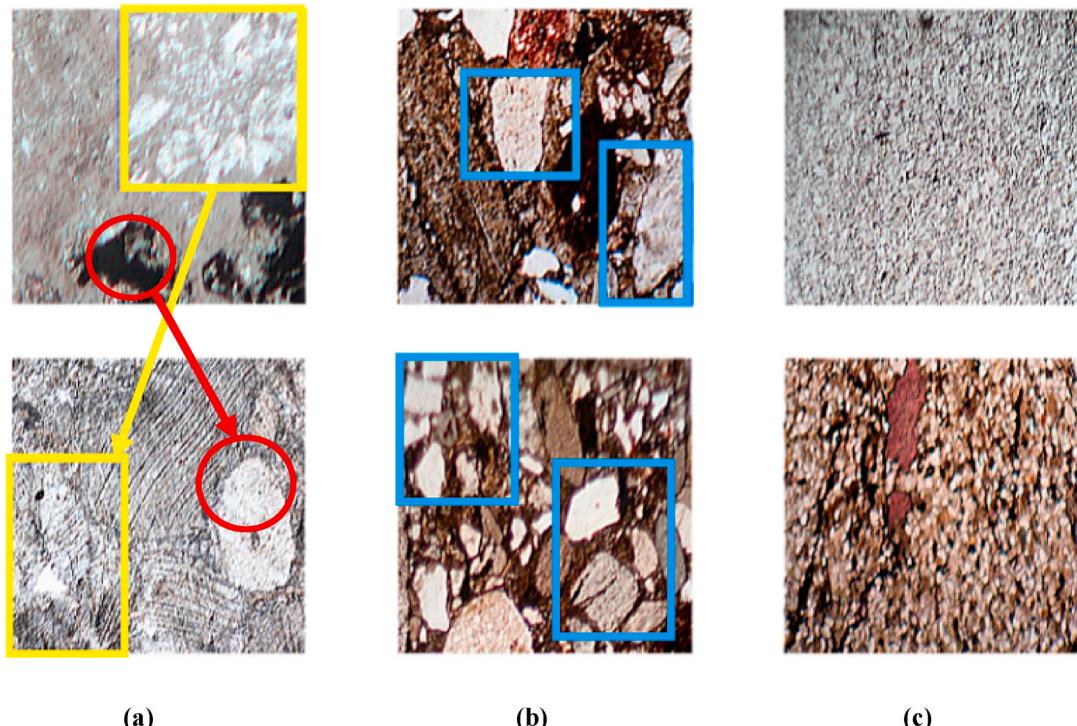


Fig. 1. Challenges in rock image classification for microscopic images of rock sections. (a) Critical areas which are difficult to capture. (b) Complex global spatial distribution features. (c) Similar spatial distribution features but quite different spectral features. (The scale is 500 × 500).

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. (Mlynarczuk 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. Karimoupoli 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.

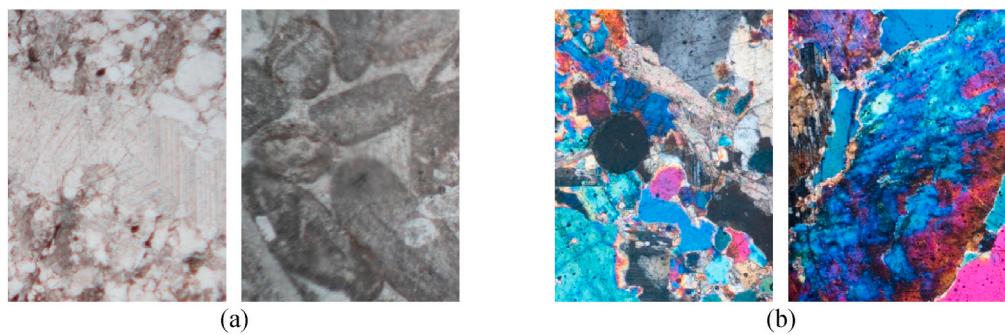


Fig. 2. (a) Represents images captured through PPL; (b) represents images captured through XLP.

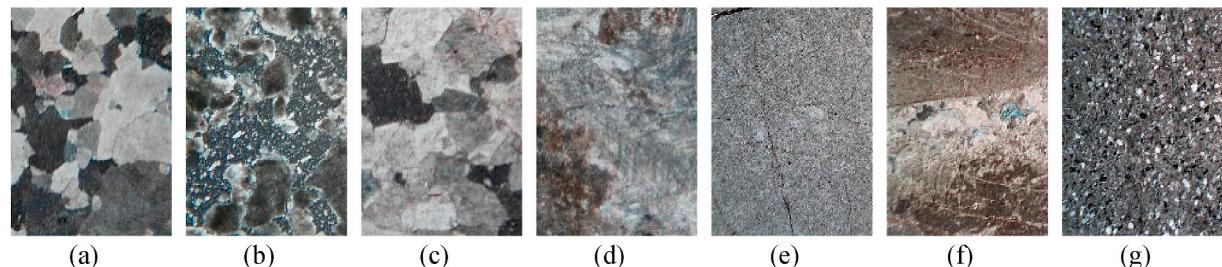


Fig. 3. Example images of grain property. (a) Coarse crystal. (b) Medium crystal. (c) Fine crystal. (d) Powder crystal. (e) Micro crystal. (f) Mud crystal. (g) Others. (Scale from 300 to 500).

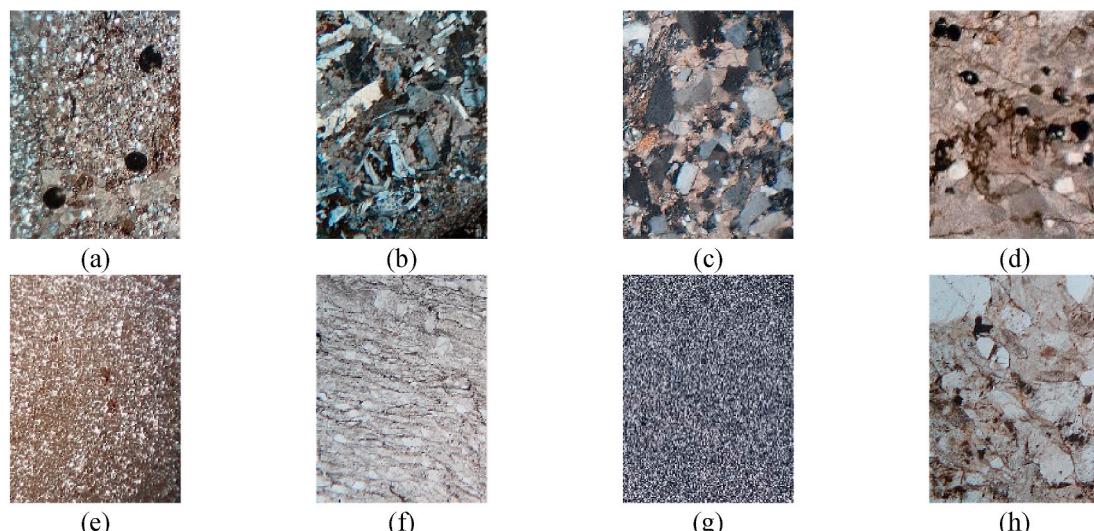


Fig. 4. Example images of clastic property. (a) Fine gravel. (b) Medium gravel. (c) Medium gravel sand. (d) Coarse sand. (e) Fine sand. (f) Coarse silt. (g) Fine silt. (h) Others. (Scale from 300 to 500).

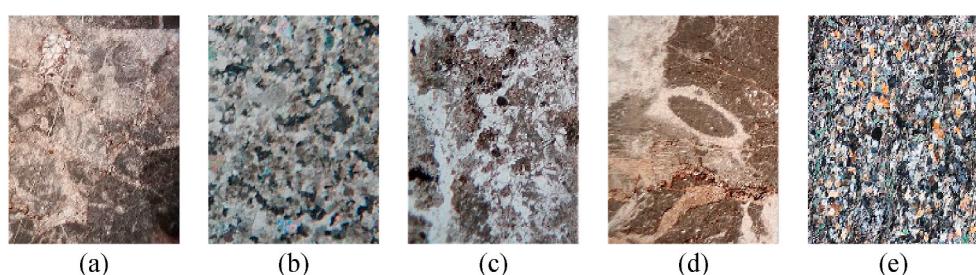


Fig. 5. Example images of mechanical genesis property. (a) Gravel. (b) Sand. (c) Powder. (d) Aggregate. (e) Others. (Scale from 300 to 500).

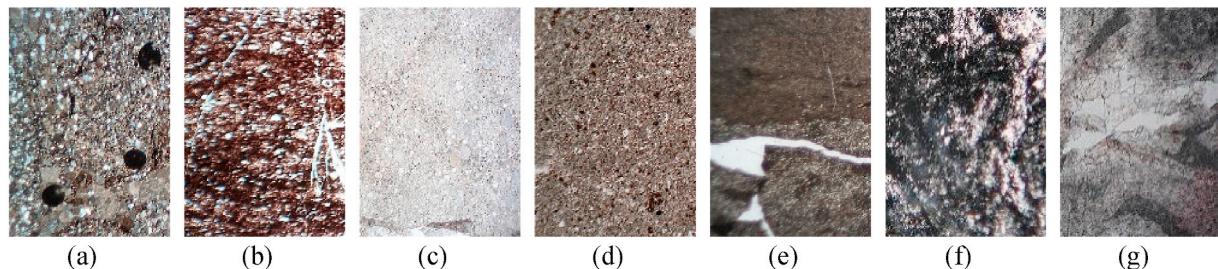


Fig. 6. Example images of mixture property. (a) Calcareous. (b) Iron. (c) Siliceous. (d) Containing silty. (e) Silty. (f) Carbonaceous. (g) Others. (Scale from 300 to 500).

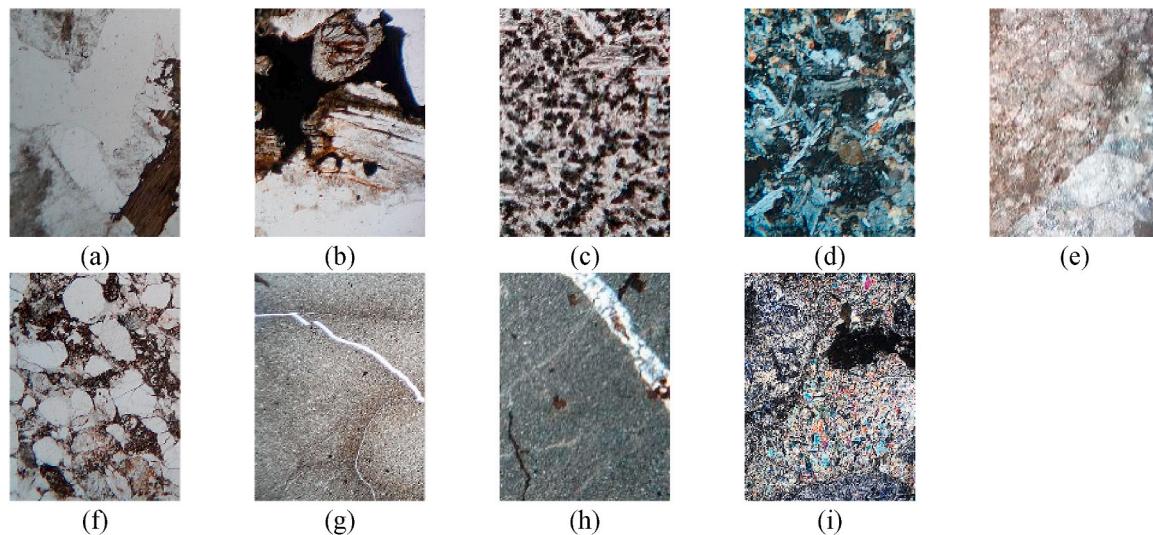


Fig. 7. Example images of basic category property. (a) Granite. (b) Diorite. (c) Basalt. (d) Tuff. (e) Lithic sandstone. (f) Quartz sandstone. (g) Siltstone. (h) Slate. (i) Others. (Scale from 300 to 500).

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 RockS²Net framework

to effectively solve the above problems. Fig. 8 illustrates the complete flowchart of the RockS²Net framework. A succinct introduction to the RockS²Net framework is given in the subsequent section.

Fig. 8 shows the overall architecture of RockS²Net. 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, RockS²Net 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. RockS²Net 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.

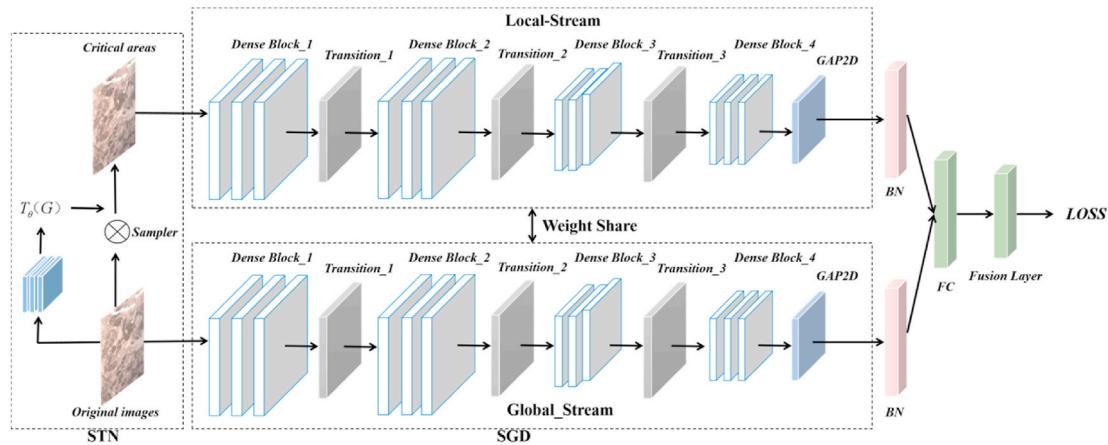


Fig. 8. Overview of the proposed framework for rock image classification.

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, b'_i) are the regularized coordinates of the input space and the output space, respectively.

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} = T_\theta \begin{pmatrix} a'_i \\ b'_i \\ 1 \end{pmatrix} = \begin{pmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{pmatrix} \begin{pmatrix} a'_i \\ b'_i \\ 1 \end{pmatrix} \quad (1)$$

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

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.

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

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

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

The affine transformation of STN used by RockS²Net 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.

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

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 RockS²Net, 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

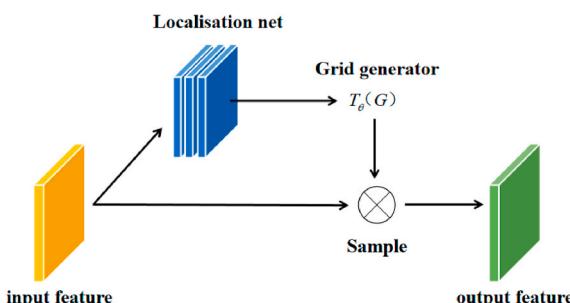


Fig. 9. The architecture of the spatial transformer network (STN).

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.

$$\hat{x}_i = \frac{x_i - E(x_i)}{\sqrt{Var(x_i) + \epsilon}} \quad (5)$$

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 RockS²Net framework

The RockS²Net 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.

$$K = D \times (k - 1) + 1 \quad (6)$$

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 RockS²Net, 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

$$F_g = GD_g(I_g) \quad (7)$$

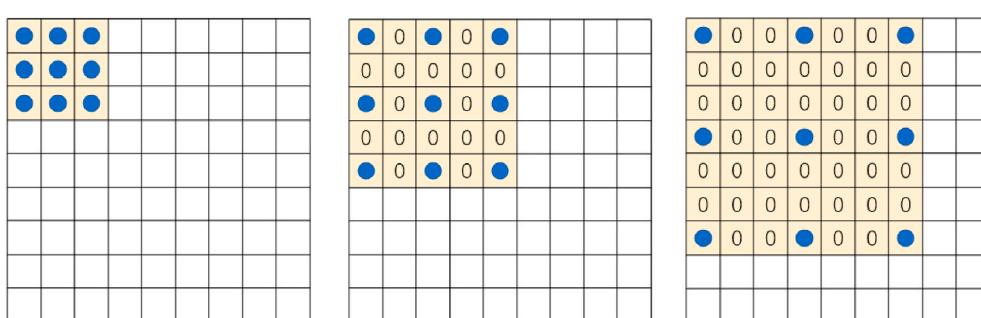


Fig. 10. Conceptual illustration of traditional and dilated convolution; (a) traditional convolution (b) dilated convolution with dilation rate 2; (c) dilated convolution with dilation rate 3.

$$F_l = GD_l(I_l) \quad (8)$$

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

$$F_s = \frac{F_g + F_l}{2} \quad (9)$$

5. Experiment and evaluation

5.1. Application of the RockS²Net

5.1.1. Model preprocessing

In this paper, VGG-16 (Simonyan and Zisserman, 2014) is chosen as the baseline to compare with the proposed RockS²Net, 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 (S²DNet), 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, S²DNet and RockS²Net were trained with a batch size of 4. In S²DNet and RockS²Net, 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 RockS²Net 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 RockS²Net 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 RockS²Net contains two components, STN and SGD block. The classification performance of GDNet, DCNet and S²DNet 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 RockS²Net to obtain the best rock image classification accuracy.

Fig. 10 shows some example images correctly classified by DenseNet-121 and RockS²Net. 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 RockS²Net 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 RockS²Net 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 DenseNet-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%
RockS²Net	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%
RockS²Net	87.14%	91.92%	97.75%	92.14%	91.85%	92.16%

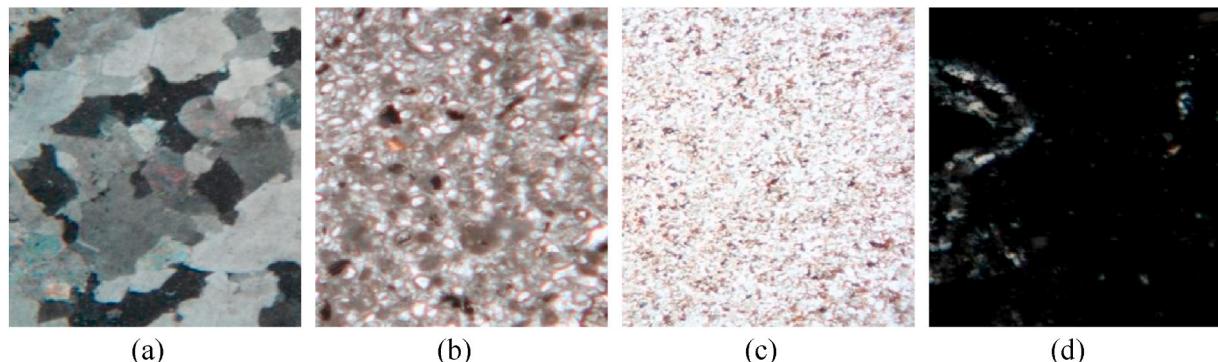


Fig. 11. Example images correctly classified by both DenseNet-121 and RockS²Net. (a) Coarse crystal. (b) Aggregate. (c) Calcareous. (d) Carbonaceous.

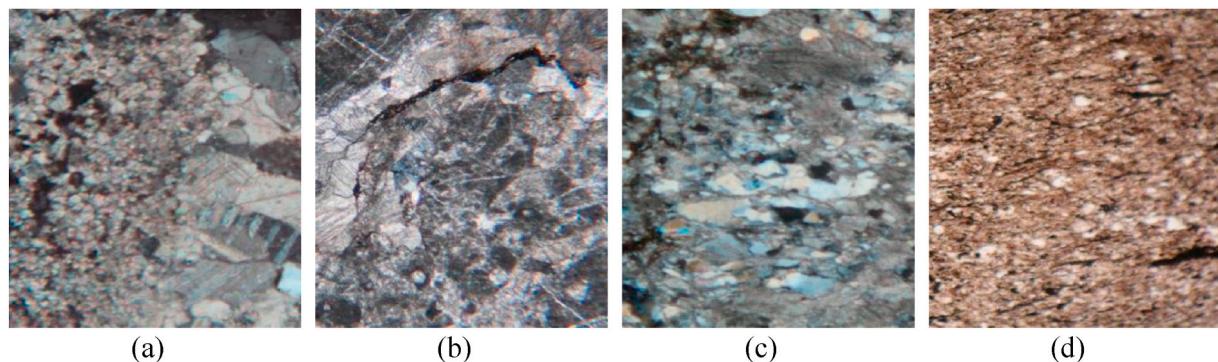


Fig. 12. Example images correctly classified by RockS²Net but incorrectly classified by DenseNet-121. (a) Coarse crystal. (b) Aggregate. (c) Calcareous. (d) Carbonaceous.

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.

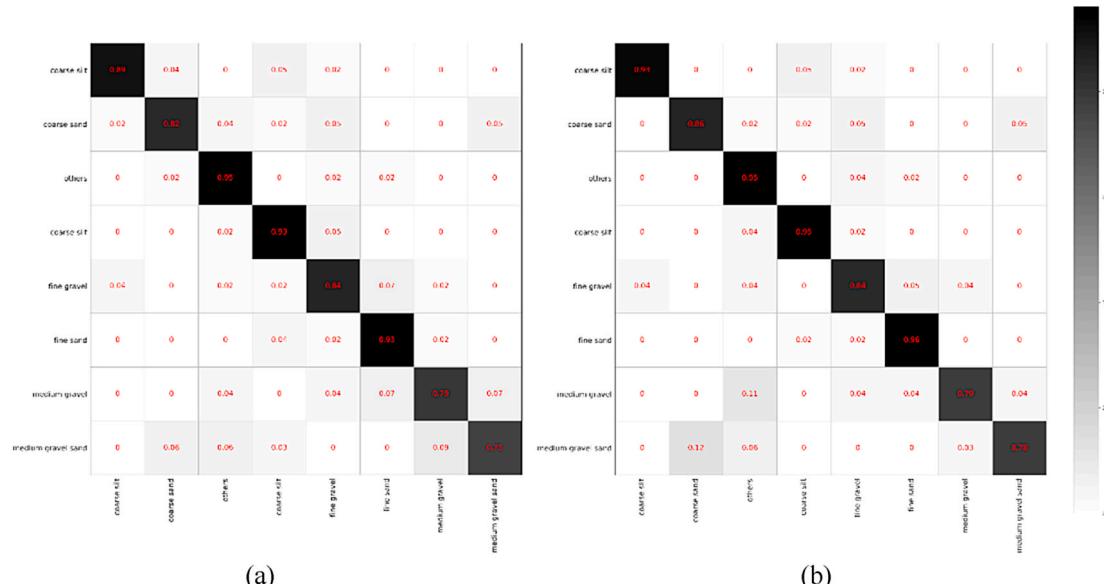


Fig. 13. Confusion matrixes on the clastic property classification. (a) VGG-16. (b) DenseNet-121.

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 of extracting 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 RockS²Net are compared. Four images which are incorrectly classified by GDNet and correctly classified by RockS²Net 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, RockS²Net 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 RockS²Net for rock image classification. RockS²Net 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 RockS²Net. The experimental results demonstrate that the RockS²Net 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).

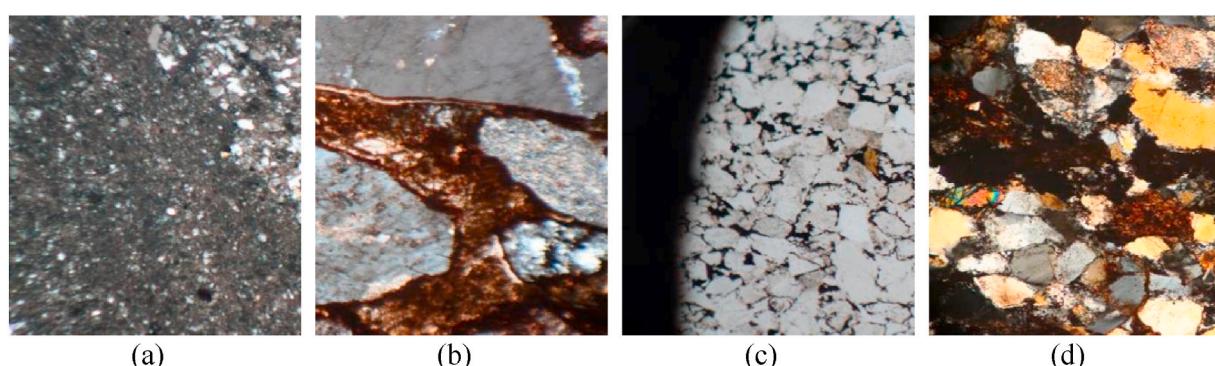


Fig. 14. Example images that are misclassified by DenseNet-121 and correctly classified by GDNet. (a) Silty. (b) Calcareous. (c) Carbonaceous. (d) Iron.

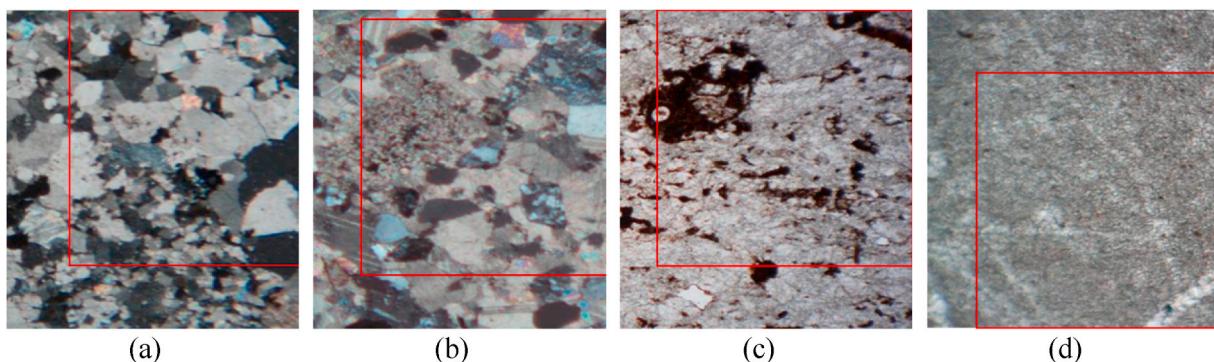


Fig. 15. Example images that are misclassified by GDNet and correctly classified by RockS²Net. (a), (b) Coarse crystal. (c), (d) Micro crystal. Note: The critical areas of images transformed by STN were in the red boxes.

Code availability section

Name of the code/library: RockS²Net
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., Koroteev, 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., 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, 124097–124095.
- 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.
- Mlynarczuk, 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. Icm.
- 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.