Lithology scene classification based on channel grouping fusion and adaptive feature filtering network

Zhiyuan Sui, Haoyi Wang, and Xianju Li*

School of Computer Science, China University of Geosciences, Wuhan 430074, China
ddwhlxj@cug.edu.cn

Abstract. Lithology classification is an important research direction in geological remote sensing. Lithology exhibits discernible textural features at a certain spatial scale, which require representation at the scene scale. Lithology is a high-level semantic information and its features are easily masked by vegetation, posing challenges in remote sensing feature extraction. In this paper, we constructed a lithology scene classification dataset named MSRS-LSC based on multi-source remote sensing data. Subsequently, we propose a lithology scene classification model called channel grouping fusion and adaptive feature filtering network (CGFAFFNet) to solve this problem. This model consists of two modules: 1) Channel grouping fusion (CGF) module: this module performs channel grouping learning, infor-mation interaction, and weighted fusion on the features extracted from multi-source remote sensing data, fully utilizing the complementary information in the channel dimension of the multi-source remote sensing data to extract key litholo-gy features; 2) Adaptive feature filtering module: this module cascades the fused features from different CGF modules and performs weighted calculations in both the channel and spatial dimensions. It filters out redundant feature information caused by multi-source remote sensing data, enhancing the model's ability to ex-tract key contextual information. The proposed model achieved an Overall accu-racy (OA), F1-score, and Kappa of 80.99% ± 0.4%, 81.26% ± 0.38%, and 78.85% ± 0.44%, respectively, outperforming mainstream scene classification models.

Keywords: Lithology classification, Remote sensing, Scene classification.

1 Introduction

Lithology classification is the basis of earth science research and is of great significance to geological survey, mineral exploration, environmental protection and resource management [1]. With the development of the type and quantity of remote sensing image data, the information provided is more and more abundant, which makes it possible to automatically classify lithology at regional scale. Therefore, it is of great theoretical and practical significance to carry out lithology classification by remote sensing [2].

*Corresponding author

Remote sensing lithology classification refers to the process of distinguishing rock and soil types of shallow surface based on remote sensing data [3]. Different from the research targets in conventional remote sensing images, lithology is a high-level abstract semantic feature, which is presented as a spatial aggregation feature at a certain scale in remote sensing images, and the spatial distribution has a certain continuity, so it needs to be characterized and learned at the scene scale. Scene scale can contain more information and better adapt to complex landforms and surface object types caused by different lithology. Moreover, some zonal texture structures can be observed on remote sensing images. Therefore, the spatial context information in image scene blocks can help to learn more lithology characteristics [4]. Fig**.** 1 shows the comparison of multispectral images of Calcium Magnesium Silicate Rocks at different scales. Fig**.** 1 (a) is the scene-level scale, where some banded texture structures can be observed, and Fig**.** 1 (b) is some information features that are difficult to characterize lithology at pixel scale.

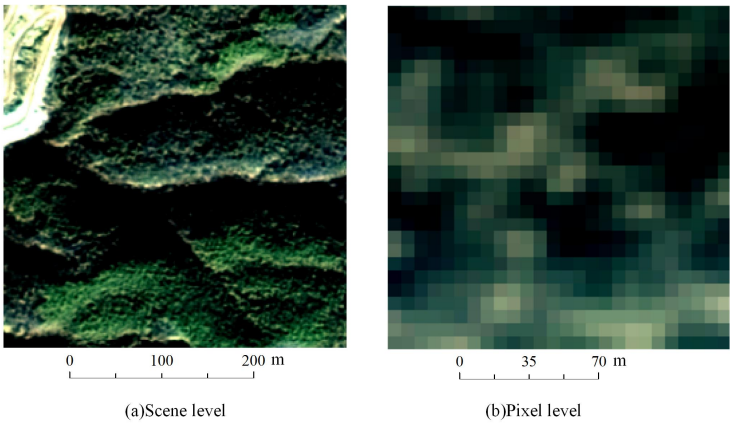

Fig. 1. Comparison of multispectral images of Calcium Magnesium Silicate Rocks at different scales

Lithology classification methods can be divided into three categories: artificial feature-based, machine learning, and deep learning. In the early stages, researchers used methods based on artificial features for lithology classification [5], [6], [7]. For example, Hunt et al. [8] studied the reflection spectra of rocks and minerals, measured the spectra of rocks and minerals in the visible-near-infrared range, and discussed and summarized the causes. Gaffey et al. [9] conducted spectral analysis of anhydrous carbonate minerals and proposed the concept that absorption peaks are diagnostic identification features for minerals. However, artificial feature-based methods have low efficiency and limited feature representation capability due to their shallow feature extraction levels.

Subsequently, the emergence of machine learning methods has partly alleviated the problem of low efficiency in manually extracting features. Previous researchers have employed machine learning algorithms for lithology classification. For example, Perez et al. [10] extracted spectral and texture information of lithology and employed Support Vector Machine (SVM) for the lithology classification. Rezaei et al. [11] utilized SVM to enhance lithology mapping in the Sangan region of northern Iran.

Machine learning methods have improved the efficiency of lithology classification. However, features based on shallow learning in machine learning methods have significant limitations in lithology information representation, especially in scenes where lithology is obscured. Additionally, machine learning methods cannot be trained in an end-to-end manner, which is no longer suitable in the era of remote sensing big data [12].

Recently, deep learning techniques have demonstrated powerful feature representation capabilities. Drawing on the advanced progress of deep learning techniques, various deep learning methods have dominated the field of lithology classification in remote sensing and achieved significant breakthroughs in classification performance [13]. Compared to traditional methods, deep neural network architectures can extract high-level semantic features and obtain more robust object feature representations [14]. With the widespread adoption of deep learning techniques, there have been numerous surveys published in recent years that utilize deep learning methods for lithology classification [15]. For example, Ye et al. [16] used the Gaofen-5 satellite data equipped with advanced hyperspectral collector, combined with deep learning to classify, and achieved good results. However, the lithology is mostly covered by vegetation in remote sensing images, and its feature information is weak. Conventional deep learning methods are difficult to extract the key features of lithology, resulting in poor classification performance.

In addition, at the data level, a single data source can only provide partial information about the lithology [17]. For example, multispectral data can capture the spectral and spatial features of lithology, SAR data has strong penetration capability and can effectively reflect differences in surface morphology and roughness, while DEM can represent information about surface topography [18]. In areas with vegetation cover, the spectral information about lithology is weak, and the surface topography and morphology caused by lithology can be complex. Relying solely on a single data source cannot provide sufficient information about the lithology. In such cases, the fusion of multi-source remote sensing data sources is necessary to complement each other's information and provide sufficient information about the lithology. Currently, researchers have used multi-source data for lithology classification [19], [20], [21]. For example, Chen et al. [22], [23] constructed a multi-source remote sensing dataset based on multispectral, SAR, and DEM data, and designed a network that performs multi-source data fusion at the feature level for lithology classification. Qasim et al. [24] utilized Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) and Sentinel-2B data, and applied techniques such as decorrelation stretch (DS), band indices (BI), principal component analysis (PCA), and minimum noise fraction (MNF) to analyze spectral features of rock lithology and achieve lithology mapping. Indeed, the previous mentioned dataset primarily focused on fusing the source data without fully leveraging some of the information contained within the source data. Therefore, in this study, in addition to utilizing multispectral, SAR, and DEM data, we also extract additional information from the DEM such as aspect, slope, and hillshade. This inclusion of additional data derived from the DEM enhances the richness of the lithology information present in the dataset.

To address the issues of insufficient lithology information in single-source data and underutilization of information from multi-source data, a multi-source remote sensing lithology scene classification dataset (MSRS-LSC) has been constructed. In addition, in order to extract lithology key features and context information, a lithology scene classification model called channel group fusion and adaptive feature filtering network (CGFAFFNet) was proposed and tested. Our model has the following two key structures: (1) Channel grouping fusion (CGF) module. This module can enhance channel information, extract lithology key features and improve classification accuracy by grouping and interacting channel dimensions. (2) Adaptive feature filtering (AFF) module. This module is used to filter out some redundant information when the multi-scale feature cascade, improve the ability of the model to extract context information, and improve the classification accuracy.

2 MSRS-LSC

2.1 Overview of the study area and remote sensing data

The study area is located in the southeast of Hubei Province, with a longitude of 113°59’ - 115°52’ and a latitude of 29°03’ - 30°27’. The study area is located in a low mountain and hilly area, with scattered lakes, more vegetation coverage, and a subtropical monsoon climate with more rainfall and mild climate. The wide distribution of vegetation in the area makes it challenging to classify lithology scenes. In this study, the multispectral image was captured by Gaofen-6 (GF-6) satellite in 2021 with a resolution of 2m; the SAR data was captured by Gaofen-3 (GF-3) satellite in 2021 with a resolution of 5m; the 10m DEM data was captured by Ziyuan-3 (ZY-3) satellite. The aspect, slope, and hillshade was extracted from DEM. The images of the study area are shown in Fig. 2.

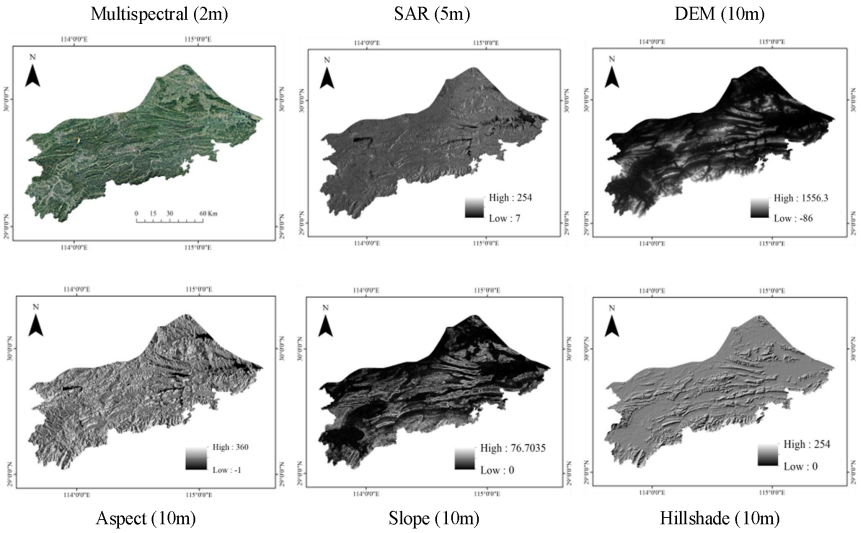

Fig. 2. Study area images

2.2 Label data

As shown in the Fig. 3, the label data of the study area contains a total of 15 types of lithology: Water, Metaquartzite-Quartz Conglomerate, Ultrabasic Intrusive Rocks, Quaternary, Calcium Magnesium Silicate Rocks, Siliceous Rocks, Basic Volcanic Lava, Basic Intrusive Rocks, Terrigenous Clastic Rocks, Acid Volcanic Lava, Acid Intrusive Rocks, Carbonate Rocks, Ferric Rocks, Intermediate Volcanic Lava, and Intermediate Intrusive Rocks.

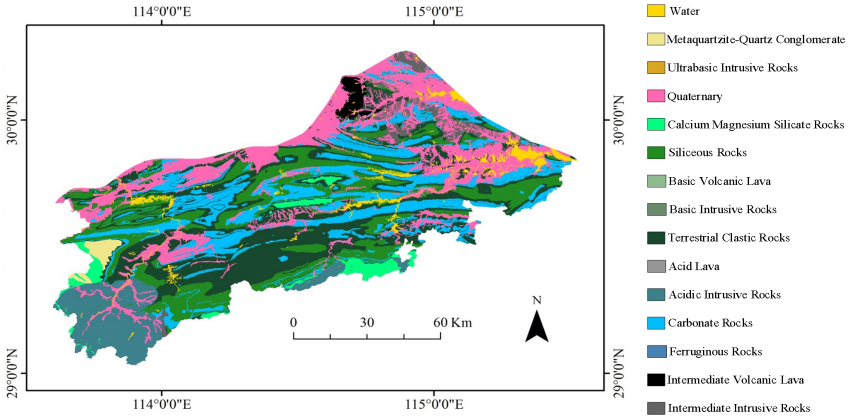

Fig. 3. Study area label

2.3 Introduction of the MSRS-LSC

The image cutting size was set to 256 × 256. A size analysis was carried out. The class ratio of 256 × 256 is more suitable, and the prior knowledge of lithology symbiosis probability can be obtained better than other sizes. Considering the lithology characteristics and prior knowledge of the vegetation area, 256 × 256 was selected.

After tailoring, 600 samples were randomly selected from each category of sample, and the training set, validation set and test set were divided according to 6:2:2, and the MSRS-LSC as shown in the following Table 1 was finally obtained.

Table 1. MSRS-LSC

Category	Training set	Validation set	Test set	Total
Water	360	120	120	600
Metaquartzite-Quartz Conglomerate	242	82	80	404
Quaternary	360	120	120	600
Calcium Magnesium Silicate Rocks	360	120	120	600
Siliceous Rocks	360	120	120	600
Terrestrial Clastic Rocks	360	120	120	600
Acidic Intrusive Rocks	360	120	120	600
Carbonate Rocks	360	120	120	600
Intermediate Volcanic Lava	293	99	97	489
Intermediate Intrusive Rocks	360	120	120	600

In the cutting samples, Ultrabasic Intrusive Rocks, Basic Intrusive Rocks and Ferric Rocks samples account for a relatively small proportion, and the number of samples is 0. The Basic Volcanic Lava and Acid Volcanic Lava samples are less than 10, and the scene label category is also discarded. Therefore, the dataset of scene classification produced only has 10 categories.

2.4 MSRS-LSC features

The samples of MSRS-LSC have the following features:

1) Serious vegetation coverage. In the remote sensing image, the lithology is heavily covered by vegetation, and the bare lithology is less, resulting in weak lithology information and difficult to extract key features.

2) Inter-class similarity and intra-class difference. The lithology in remote sensing image has some similarity among different categories and some difference in the same category, so the model is easy to misclassify, which makes it difficult to improve the classification accuracy.

3 Methods

3.1 Overall structure

CGFAFFNet is shown in the Fig. 4. Using DenseNet121 as the backbone. This network improves the efficiency of information transmission by designing dense connection blocks, and this structure enables the network to better reuse features and avoid information loss. In the Fig. 4, DenseBlock is a dense connection block of the backbone network DenseNet121, which is mainly used for feature extraction. The subsampling is the Transition of DenseNet, which mainly reduces the dimensionality of features. CGF is used to enhance channel information, and AFF is used to avoid false information caused by the fusion of different scale features, and fuses multi-scale features.

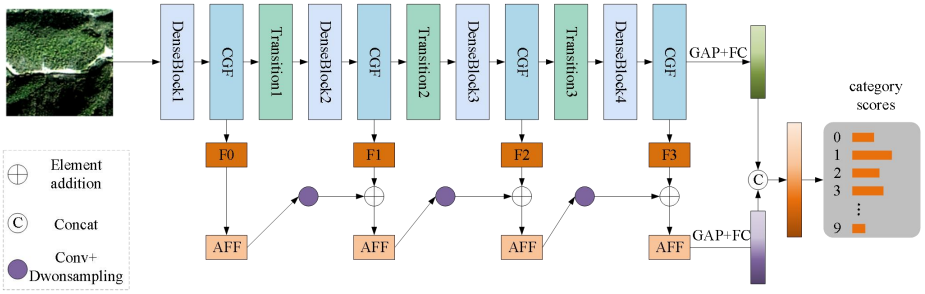

Fig. 4. CGFAFFNet

3.2 Channel grouping fusion

By performing convolution operations on the channel dimension, the network can learn the correlation and importance between different channels, so as to extract richer and more abstract feature representations. However, due to the complex information in multi-source remote sensing lithology data, the model cannot focus on key channel

information, and it may not be able to make full use of different channel information brought by multi-source data only through ordinary convolution. Inspired by SENet [40] and ECA-MSDWNet [41], a CGF was proposed. The key difference of CGF from SENet and ECA-MSDWNet lies in the combination of ideas from both models. In CGF, we integrate the concept of channel weighting from SENet and the idea of channel grouping and information interaction from ECA-MSDWNet. By facilitating information exchange among channels, CGF fully utilizes the channel information between multiple data sources and weights the channel information to extract crucial channel features. This combination enables CGF to effectively capture and leverage the essential channel-level information for better feature representation and classification in multi-source data. The CGF is shown in the Fig. 5.

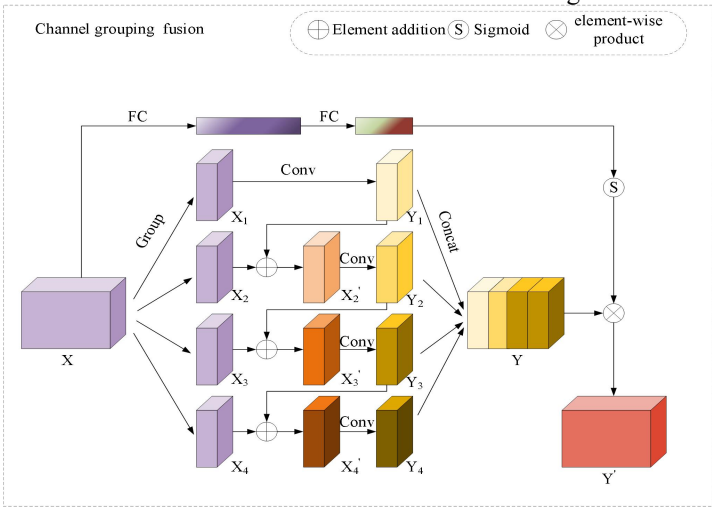

Fig. 5. Channel grouping fusion

The input feature X is divided to 4 parts along the channel dimension. X_1, X_2, X_3, X_4 will have the $Split(\bullet)$ function shown in formula 1, assuming number of X channels is C, then number of X_1, X_2, X_3, X_4 channels are C/4, so X is divied into 4 parts along the channel dimension.

$$X_1, X_2, X_3, X_4 = Split(X) \tag{1}$$

After a convolution operation, X_1 gets Y_1, and then Y_1 and X_2 are added to get the feature reuse feature $X_2^{'}$, and Y_2 is obtained by convolution, and Y_1, Y_2, Y_3, Y_4 is obtained according to formula 2 and formula 3. $Conv(\bullet)$ in formula 3 refers to the convolution operation of 1 × 1.

$$X_i^{'} = Y_{i-1} + X_i \tag{2}$$

$$Y_i = Conv(X_i^{'}) \tag{3}$$

Y_1, Y_2, Y_3, Y_4 connect channels, as shown in formula 4. $C(\bullet)$ represents the connection function of channel dimension, and Y feature map is obtained after connection.

$$Y = C(Y_1, Y_2, Y_3, Y_4) \tag{4}$$

While the channel information is reused in the grouping, the original feature X is calculated through two fully connected layers and the Sigmoid activation function to obtain the channel weight W. In formula 5, FC represents the fully connected operation.

$$W = Sigmoid(FC(FC(X))) \tag{5}$$

Finally, the channel weight W and the feature map Y are multiplied to obtain the feature map Y' which enhances the channel information.

$$Y' = W * Y \tag{6}$$

By grouping channels and mixing information, we can make full use of the information of each channel in the multi-source data, improve the network representation ability through channel information interaction, and then select more important channel information and extract key features based on the channel weight W obtained by the original feature X.

3.3 Adaptive feature filtering

Multi-scale features cascade is used in conventional feature fusion methods. However, these features contain redundant information, which increases the difficulty of classification. Therefore, it is necessary to filter these features before the cascade of multi-scale features. Inspired by attention [42], [38] and MSFT [43], an AFF was proposed. As Fig. 6 shown, the AFF consists of three branches: trunk branch, channel branch, and spatial branch.

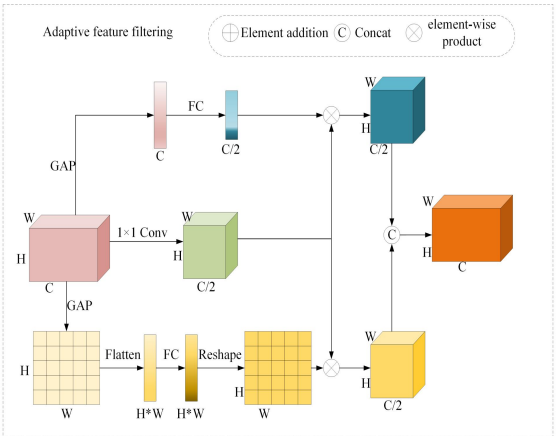

Fig. 6. Adaptive feature filtering

(1) Trunk branch: for the input feature diagram F, its length is H, width is W, and the number of channels is C, that is, its size is H × W × C, and after a 1 × 1 convolution operation, the size becomes H × W × C/2, and the feature is called $F^{'}$, as shown in formula 7.

$$F^{'} = Conv_{1\times1}(F) \tag{7}$$

(2) Channel branch: this branch learns the relationship between channels through global embedding. Specifically, global average pooling in spatial dimensions is used to generate the global features of each channel. Pooling operations can be expressed as follows:

$$V_C(C) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W}F(i,j) \tag{8}$$

F represents the input feature map, and since V_C does not agree with the number of feature map channels generated by the channel branch, a full connect operation is required for alignment. As shown in formula 9, $V_C^{'}$ is obtained through full connection operation.

$$V_C^{'} = FC(V_C) \tag{9}$$

$V_C^{'}$ is a vector with correlation between channels, and R_C can be reweighted in channel dimensions as follows:

$$R_C = V_C^{'} * F^{'} \tag{10}$$

R_C represents the feature map after channel weighting, that is, the blue feature block in the top half of the figure. $*$ represents element-wise product, which filters the feature information of channel dimensions by giving less weight to irrelevant channel information.

(3) Spatial branch: this branch can be regarded as the computation of the label relation from the spatial dimension, where semantic information can be further extracted from the spatial dimension.

Firstly, global average pooling is used in channel dimension to generate spatial tag features, as follows:

$$V_T(i,j) = \frac{1}{C}\sum_{k=1}^{C}P_{i,j}(k) \tag{11}$$

V_T represents an H × W × 1 feature map, and $P_{i,j}$ represents the (i,j) position of the feature map. Then, it is flattened by a Flatten operation, and then reshaped to the size of H × W × 1 to obtain a relation information between different semantics in the

spatial dimension, which can also be called the weight of the spatial dimension. The specific operations are as follows:

$$V_T^{'} = RSP(FC(FLA(V_T))) \tag{12}$$

$V_T^{'}$ is a two-dimensional vector that represents the semantic information of the spatial dimension and is used as the weight of the spatial dimension $F^{'}$. By giving less weight to irrelevant spatial information, the feature information of the spatial dimension is filtered. In addition, the output of the spatial branch is shown as follows:

$$R_S = V_T^{'} * F^{'} \tag{13}$$

R_S represents the features weighted by the spatial information, that is, the yellow feature block in the figure, the size of which is H × W × C/2, and $*$ represents element-wise product. The final output of the AFF is obtained from the output of two branches, and the specific operation is as follows:

$$R = Concat(R_S, R_C) \tag{14}$$

$Concat(\bullet)$ represents the channel dimension connection. Through this operation, R_S and R_C are connected in the channel dimension. Finally, the feature map R with the size of H × W × C is obtained, that is, the orange feature block in the figure.

The AFF performs weighted calculations in channel and spatial dimensions, giving higher weights to important features and less weights to irrelevant features, so as to filter feature information in both channel and spatial dimensions and avoid redundant information caused by the cascade of multi-scale features.

4 Experimental results

4.1 Experimental configuration

The hardware and software environment of the experiment are shown in the Table 2. Parameter settings are shown in the Table 3.

Table 2. Experimental environment

Experimental environment		Specific parameter
Hardware environment	CPU	AMD EPYC Processor
	GPU	NVIDIA RTX A5000(24G)
	Internal memory	32GB
Software environment	Operating system	Windows10
	Deep learning framework	Pytorch 2.1.0

The model will dynamically adjust the learning rate during training, and the initial learning rate is set to 0.0001. At the same time, the learning rate attenuation strategy is introduced, that is, the learning rate will be adjusted and the size of the learning rate

will be adjusted when a certain condition in the training meets the condition set. The learning rate is reduced after the verifica-tion set loss does not decrease for more than 5 epochs, and the multiple is 0.5. For other parameters involved in the experiment, the default coefficients in the experiment framework are used.

Table 3. Hyperparameters

Parameter	Value
Batch Size	32
Epoch	100
Lr	0.0001
Optimizer	Adam
Loss function	Cross entropy loss function

4.2 Accuracy evaluation results

Table 4 shows the evaluation index results of different network structures on the MSRS-LSC. On the MSRS-LSC, the proposed model (CGFAFFNet) achieved the best results in various indexes, OA reached 80.99% ± 0.4%, F1-score reached 81.26% ± 0.38%, and Kappa reached 78.85% ± 0.44%. Compared with DenseNet121, which had the best effect in the comparison model, OA increased by 2.66%, F1-score by 2.64% and Kappa by 2.96%.

Table 4. Experimental results on the MSRS-LSC

Model	Year	OA(%)	F1-score(%)	Kappa(%)
AlexNet [25]	2012	66.00 ± 0.90	66.02 ± 1.15	62.18 ± 0.98
Vgg16 [26]	2014	47.10 ± 3.53	45.45 ± 4.02	41.17 ± 3.98
GooleNet [27]	2014	66.00 ± 1.70	66.53 ± 1.81	62.17 ± 1.89
ResNet101 [29]	2015	63.46 ± 0.90	63.43 ± 1.07	59.36 ± 0.98
DenseNet121 [31]	2016	78.33 ± 0.44	78.62 ± 0.37	75.89 ± 0.49
ShuffleNet [28]	2017	71.51 ± 0.83	71.65 ± 0.93	68.31 ± 0.92
EffcientNet_b7 [35]	2019	62.67 ± 2.47	62.53 ± 2.79	58.53 ± 2.70
MobileNetV2 [30]	2019	57.72 ± 0.68	57.66 ± 0.61	52.98 ± 0.76
Vit [36]	2020	68.90 ± 0.74	69.09 ± 0.69	65.40 ± 0.83
Swintransformer [37]	2021	63.80 ± 1.77	64.04 ± 1.79	59.72 ± 1.97
ResNetMvt [39]	2022	74.53 ± 1.52	74.78 ± 1.60	71.67 ± 1.68
WaveMix [32]	2022	73.97 ± 1.09	74.22 ± 1.03	71.06 ± 1.22
DGBANet [33]	2023	77.10 ± 0.86	77.59 ± 0.85	74.52 ± 0.95
GhostNetV3 [34]	2024	52.01 ± 1.27	61.79 ± 3.58	46.59 ± 1.41
CGFAFFNet(ours)	2024	**80.99 ± 0.40**	**81.26 ± 0.38**	**78.85 ± 0.44**

Among the comparison models, VGG16 has the worst effect, while other models have the best effect, among which DensNet121, Vit and ResNetMvt have significantly better effects than other models, or close to 70% or greater than 70%, indicating that for the research objective of lithology, more deep features are needed. Features reuse and the acquisition of context information are also important. Therefore,

DenseNet121 is used as the backbone network, and CGF module is added to reuse and interact channel features to enhance channel information. AFF module is added for multi-scale feature cascade to enhance the ability of the model to extract context information, so as to obtain better classification effect.

5 Discussions

This section presents and discusses the ablation experimental results of different modules in the proposed CGFAFFNet model to verify the validity of these modules.

5.1 Effectiveness of different modules

In order to verify the effectiveness of each module, we conducted ablation experiments for CGF module, AFF module and multi-scale features cascade (MSFC). Due to the requirement of combining the AFF with MSFC, no combined experiment of CGF and AFF was conducted.

Table 5 shows the results of the ablation experiments. On top of the Backbone, the addition of individual modules has resulted in improvements, with CGF showing the highest enhancement. OA, F1-score, and Kappa have been improved by 2.25%, 2.22%, and 2.51%, respectively. This indicates that CGF facilitates channel grouping learning, information interaction, and weighted fusion of features, extract crucial lithology features and improve the classification accuracy of the model. Additionally, MSFC can extract contextual information and improve classification accuracy. Furthermore, incorporating AFF during MSFC, which filters out redundant information in both the channel and spatial dimensions, can further enhance the classification accuracy.

In addition, the experimental results of combining two modules together are superior to the results of adding individual modules alone. Compared to using a single module alone, the combination of CGF and MSFC as well as the combination of AFF and MSFC both show improvements in OA. This indicates that there is a certain complementary relationship between the modules, and combining different modules can further enhance the classification performance of the model.

Table 5. Results of ablation experiments on the MSRS-LSC

Model	CGF	AFF	MSFC	OA(%)	F1-score(%)	Kappa(%)
CGFAFFNet				78.33 ± 0.44	78.62 ± 0.37	75.89 ± 0.49
CGFAFFNet	√			80.58 ± 0.57	80.84 ± 0.53	78.40 ± 0.63
CGFAFFNet	√		√	80.74 ± 0.19	80.97 ± 0.16	78.58 ± 0.21
CGFAFFNet			√	79.77 ± 0.33	80.17 ± 0.31	77.49 ± 0.36
CGFAFFNet		√	√	80.33 ± 0.58	80.66 ± 0.50	78.13 ± 0.65
CGFAFFNet	√	√	√	80.99 ± 0.40	81.26 ± 0.38	78.85 ± 0.44

Combining the three modules together resulted in the highest OA of 80.99% ± 0.4%. This indicates that the three modules complement each other in terms of information extraction, enabling the extraction of critical lithology features and enhancing the

model's performance in lithology classification, particularly in areas with vegetation coverage.

6 Conclusion

To address the issues of insufficient lithology information in single-source data and underutilization of lithological information from multiple data sources, a multi-source remote sensing lithology scene classification dataset named MSRS-LSC was constructed. The MSRS-LSC incorporates multi-spectral, SAR, and DEM data, along with derived data such as slope direction, slope gradient, and hillshade. To address the challenge of extracting remote sensing features due to the high-level semantic nature of lithology and its susceptibility to vegetation coverage, we propose a method called CGFAFFNet. CGFAFFNet's effectiveness can be attributed to: 1) the Channel Grouping Fusion module is capable of grouping feature maps in the channel dimension, facilitating feature interaction and weighted fusion. This process enables the extraction of crucial fused features across channels; 2) the Adaptive Filtering module can filter out redundant information from both the channel and spatial dimensions when cascading multi-scale features, thereby extracting more accurate contextual information.

The experimental results show that the OA, F1-score and Kappa of CGFAFFNet on MSRS-LSC reached 80.99% ± 0.4%, 81.26% ± 0.38% and 78.85% ± 0.44%, respectively, which is superior to other models. The experimental results demonstrate the effectiveness of CGFAFFNet. In future research, we will focus on improving the dataset and further investigate lithology scene classification methods.

Acknowledgments. This study was jointly supported by the Natural Science Foundation of China under Grants 42071430 and U21A2013, the Opening Fund of Key Laboratory of Geological Survey and Eval-uation of Ministry of Education under Grants GLAB2022ZR02 and Grant GLAB2020ZR14. Computation of this study was performed by the High performance GPU Server (TX321203) Computing Center.

References

1. Bouwafoud A, Mouflih M, Benbouziane A. Lithological mapping using landsat 8 OLI in the meso-cenozoic Tarfaya Laayoune basin (south of Morocco): comparison between ANN and SID classification. Open Journal of Geology 11(12), 658-681 (2021).
2. Serbouti I, Raji M, Hakdaoui M, et al. Pixel and object-based machine learning classification schemes for lithology mapping enhancement of semi-arid regions using sentinel-2A imagery: a case study of the southern Moroccan meseta. IEEE Access 9, 119262-119278 (2021).
3. Lu J, et al. Lithology classification in semi-arid area combining multi-source remote sensing images using support vector machine optimized by improved particle swarm algorithm. International Journal of Applied Earth Observation and Geoinformation 119, 103318 (2023).

4. Li Z, et al. Interpretable semisupervised classification method under multiple smoothness assumptions with application to lithology identification. IEEE Geoscience and Remote Sensing Letters 18(3), 386-390 (2020).
5. Li X, Tang Z, Chen W, et al. Multimodal and multi-model deep fusion for fine classification of regional complex landscape areas using ZiYuan-3 imagery. Remote Sensing 11(22), 2716 (2019).
6. Wu C, Li X, Chen W, et al. A review of geological applications of high-spatial-resolution remote sensing data. Journal of Circuits, Systems and Computers 29(6), 2030006 (2020).
7. Yang G, Wang Z, He J, et al. Development and Allometry Patterns of Fine Scale Fish Larvae at Low Temperature. In Journal of Physics: Conference Series, vol. 1575, pp. 012202 (2020).
8. Hunt G, Ashley R. Spectra of altered rocks in the visible and near infrared. Economic Geology 74(7), 1613-1629 (1979).
9. Gaffey S. Spectral reflectance of carbonate minerals in the visible and near infrared (0.35-2.55 microns): calcite, aragonite, and dolomite. American Mineralogist 71(1-2), 151-162 (1986).
10. Perez C, Estévez P, Vera P, et al. Ore grade estimation by feature selection and voting using boundary detection in digital image analysis. International Journal of Mineral Processing 101(1–4), 28-36 (2011).
11. Rezaei A, Hassani H, Moarefvand P, et al. Lithological mapping in Sangan region in Northeast Iran using ASTER satellite data and image processing methods. Geology, Ecology, and Landscapes 4(1), 59-70 (2020).
12. Zhang J, and Li C. Remote Sensing Object Detection Meets Deep Learning: A metareview of challenges and advances. IEEE Geoscience and Remote Sensing Magazine (2023).
13. Gu Y, Zhang Z, Zhang D, et al. Complex lithology prediction using mean impact value, particle swarm optimization, and probabilistic neural network techniques. Acta Geophysica 68, 1727-1752 (2020).
14. Latifovic R, Pouliot D, Campbell J. Assessment of convolution neural networks for surficial geology mapping in the South Rae geological region, Northwest Territories, Canada. Remote sensing 10(2), 307(2018).
15. Liu H, Wu K, Xu H, et al. Lithology classification using TASI thermal infrared hyperspectral data with convolutional neural networks. Remote Sensing 13(16), 3117 (2021).
16. Ye B, Tian S, Cheng Q, et al. Application of Lithological Mapping Based on Advanced Hyperspectral Imager (AHSI) Imagery Onboard Gaofen-5 (GF-5) Satellite. Remote Sensing 12(23), 3990 (2020).
17. He D, Wang L. Recognition of lithological units in airborne SAR images using new texture features. TitleREMOTE SENSING 11(12), 2337-2344 (1990).
18. He D, Wang L. Recognition of lithological units in airborne SAR images using new texture features. TitleREMOTE SENSING 11(12), 2337-2344 (1990).
19. Seid A, T S. Identification of Lithology and Structures in Serdo, Afar, Ethiopia Using Remote Sensing and Gis Techniques. International Journal of Geoinformatics and Geological Science 8(1), 27-41 (2021).
20. Wang Z, Zuo R, Jing L. Fusion of Geochemical and Remote-Sensing Data for Lithological Mapping Using Random Forest Metric Learning. Mathematical Geosciences 53(6), 1125-1145 (2020).
21. Pal M, Rasmussen T, Porwal A. Optimized Lithological Mapping from Multispectral and Hyperspectral Remote Sensing Images Using Fused Multi-Classifiers. Remote Sensing 12(1), 177 (2020).

22. Chen W, Li X, Qin X, et al. Lithological Remote Sensing Scene Classification Based on Multi-view Data. Remote Sensing Intelligent Interpretation for Geology: From Perspective of Geological Exploration. Singapore: Springer Nature Singapore, 75-100 (2024).
23. Chen, W, Li X, Qin X, et al. Remote Sensing Lithology Intelligent Segmentation Based on Multi-source Data. Remote Sensing Intelligent Interpretation for Geology: From Perspective of Geological Exploration. Singapore: Springer Nature Singapore, 117-163 (2024).
24. Qasim M, Khan S, et al. Integration of multispectral and hyperspectral remote sensing data for lithological map in Zhob Ophiolite, Western Pakistan. Arabian Journal of Geosciences 15(7), 599 (2022).
25. Krizhevsky A, Sutskever I, Hinton E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25 (2012).
26. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. Computer Science (2014).
27. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition, 1-9 (2015).
28. Zhang X, Zhou X, Lin M, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices. Proceedings of the IEEE conference on computer vision and pattern recognition, 6848-6856 (2018).
29. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778 (2016).
30. Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks. Proceedings of the IEEE conference on computer vision and pattern recognition, 4510-4520 (2018).
31. Huang G, Liu Z, et al. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition, 4700-4708 (2017).
32. Jeevan P, Viswanathan K, Sethi A. WaveMix: A resource-efficient neural network for image analysis. arxiv preprint arxiv:2205.14375, (2022).
33. Xia J, Zhou Y, Tan L. DBGA-Net: Dual-Branch Global–Local Attention Network for Remote Sensing Scene Classification. IEEE Geoscience and Remote Sensing Letters, vol. 20, pp. 1-5 (2023).
34. Liu Z, Hao Z, Han K, et al. GhostNetV3: Exploring the Training Strategies for Compact Models. arxiv preprint arxiv:2404.11202, (2024).
35. Tan M, Le Q. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. International conference on machine learning. PMLR, 6105-6114 (2019)
36. Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929, (2020).
37. Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF international conference on computer vision, 10012-10022 (2021).
38. Hu T, Shen L, Wu D, et al. Research on transmission line ice-cover segmentation based on improved U-Net and GAN. Electric Power Systems Research, 221: 109405 (2023).
39. Tang X, Li M, Ma J, et al. EMTCAL: Efficient multiscale transformer and cross-level attention learning for remote sensing scene classification. IEEE Transactions on Geoscience and Remote Sensing, 60: 1-15 (2022).
40. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition, 7132-7141 (2018).

41. Ye Z, Zhang Y, Zhang J, et al. A Multiscale Incremental Learning Network for Remote Sensing Scene Classification. IEEE Transactions on Geoscience and Remote Sensing (2024).
42. Woo S, Park J, Lee J, et al. Cbam: Convolutional block attention module. Proceedings of the European conference on computer vision (ECCV), 3-19 (2018).
43. Wang G, Zhang N, Liu W, et al. MFST: A multi-level fusion network for remote sensing scene classification. IEEE Geoscience and Remote Sensing Letters 19, 1-5 (2022).