Multilayer Convolutional Parameter Tuning based Classification for Geological Igneous Rocks

Agus Nursikuwagus School of Electrical Engineering and Informatics Bandung Institute of Technology agusnursikuwagus@email.unikom.ac.id Rinaldi Munir School of Electrical Engineering and Informatics Bandung Institute of Technology rinaldi@informatika.org Masayu Leylia Khodra School of Electrical Engineering and Informatics Bandung Institute of Technology masayu@stei.itb.ac.id

Abstract— A framework different CNN has been proposed to solve image classification. The power of CNN and the ability to extract demanding features has to be a target for proposed the new ideas. In the geology domain, issues in ascertaining igneous rock from volcanic eruptions often contrast in classification when explored from the location of the rocks. These domain problems must be resolved, contemplating to have consistency and accelerate rock classification. CNN has used to figure out the problem by expanding in multilayer convolution. Besides, parameter tuning has anointed to get the high accuracy to enhance the CNN model. This study has exploited many parameters tuning such as rescaling, cropping, size of inaccurate filter prediction. The exploration has shown that CNN(64,5) achieves a high accuracy of 98.9% and validation carries out accuracy of 81.1%. This study has confirmed that enumerating the tuning parameter on rescaling and cropping does not boost accuracy, even modifying the filter size and stride. Some results have shown still have an inaccuracy class, specifically in the diorite and limestone. The error forecast is 31.7% of 41 predicted diorite images and 30% of 50 predicted limestone images, respectively. (Abstract)

Keywords—multilayer, convolutional, tuning, parameters, classification (keywords)

I. INTRODUCTION

Detection of objects in exact conditions has completed with the help of computer vision and the knowledge acquired by an expert. This effort associates image processing and domain knowledge maintained by experts in a particular domain. Nevertheless, it is not accessible to develop because of the narrow ability achievement of an expert to classify rocks. The estimate is very diverse, specifically when involving the color and structure of rocks around the same but have distinctive classifications. What is more, it has supported by other regions and lighting on where the geological rocks have located, which regularly causes problems in classifying the types of geological rocks. For example, ascertaining igneous rock from volcanic eruptions regularly differs in classification when viewed from the location of the rock. These problems are issues that must be resolved, intending to have coherence and expedite rocks classification. Geologists need an artificial annotation method to identify rocks more quickly with the help of rock samples observed with a microscope [1].

The accessibility of digital cameras, handheld equipment, and the expansion of computer-aided image study afford technical backing for a different range of applications. This peripheral allows some rock characteristics to be stored and evaluated digitally. Photos can represent the features of the color, grain size, and texture of the rock. Despite rock images do not represent homogeneous format, textures, or colors, computer image analysis can divide several categorize of rock images [2]. Rock is the essential component of the earth and contains the basic material for almost all modern construction and manufacturing requiring rock. In extension to the conduct utilize of rock, mining, drilling, and quarrying afford substantial provenances for metals, plastics, and fuels. Types of ordinary rock have a diversity of source and purpose. The three main classes of rocks (igneous, sedimentary, and metamorphic) are subdivided into subtypes according to distinct of peculiarities [2].

Rocks can be identified in assorted forms, such as visibly with a helped magnifying glass, microscope, or by chemical analysis. Taking rocks sample on the field is a difficult identified when using a visual looking, especially for finegrained rock. Inspection visual assess the rocks has identified by nature as color, composition, grain size, and structure. Rock properties that can reflect genesis, formation environment, mineral, and chemical composition. The color of the rock can reflect the age of the rock and its chemical composition [2]. Fan and his friends succeeded in making rock image recognition models with CNN in detecting color by making thin slices of rocks and looking at these pieces to make a classification model. Fan method has successfully identified and classified 28 rock types. Through an exhaustive observation of the two designs, the accuracy of SqueezeNet and MobileNets is 94.55% and 93.27% on the dataset test, respectively. The average detection time of a single rock image was 557 and 836 milliseconds through these two models. Rock images with a detection accuracy of more than 96% regard for 95% and 93% of all test data [1]. The CNN model identifies six prevailing rock types with an overall classification accuracy of 97.96%, thus exceeding other enact learning models and linear models [2]. Another CNN to classify the rocks achieves 98.5% accuracy for identifying rocks [3]. CNN with models ResNet 50, ResNet 101, Inception-V4, and Inception-Resnet-V2 are also used to classify rocks based on the type of structure, along with Mosaic Structure (MS), Granular Structure (GS), Layered Structure (LS), Block Structure (BS), Fragmentation Structure (FS) obtained an accuracy of 95.93%, 93.90%, 97.05%, 98.12%, and 94.80%, respectively [4]. A probabilistic neural network (PNN) model in which histogram features of colors are leveraged as input was used to classifying limestone gets a classification error below 6% [5]. CNN, VGG16 (standardization only), VGG16 (migration learning), and InceptionV3 models, using image input from preprocessing results such as random flipping and standardization of feature values get 85.38% accuracy, 95.16%, and 87.50%, respectively. CNN, which was used to classify images from aerial images, was successfully carried out and obtained an accuracy of 65.03% [6]. CNN for the identification of glacier rocks that were dip The earth's surface from aerial photography has succeeded in classifying glacier rocks with a

total of 6 layers and 1x1, 3x3, and 7x7 filters and obtained an accuracy of 87% [7]. Research on the analysis of geological structures using geological images with CNN is overfitting with an accuracy below 40%. However, the use of transfer learning techniques by utilizing Inception-V3 for image extraction obtains an accuracy of 92.6% with RGB. CNN has been successfully applied in various image classification tasks to obtain feature extraction to obtain high accuracy [8].

CNN can operate by using several advantages, especially in rock classification problems. First, the detection method can directly use pixels, then the reading pattern for each pixel can be limited by a bounding box in the form of a filter that can be 3x3, 5x5, and 7x7. Second, the number of convolution layers becomes very influential on the success of an image extraction. Third, direct touch on the color intensity causes the level of color analysis to be more powerful; the detection pattern by relying on a particular feature in the image becomes more accurate [7]. However, CNN also sometimes gets low accuracy as geological images show its geological structure with constraints such as tiny image sizes and using a small number of convolutions [4]. Therefore, every proposed model CNN has advantages and disadvantages. This study offers a CNN-based igneous rock classification method, namely by tuning parameters related to image preprocessing, number of convolution layers, filters, pooling strategies, ReLU, and activation functions. The main contribution of this research is: 1) provides a method of classification of rock to igneous rocks of various photo rocks obtained, 2) the object image is a photo of igneous rocks taken from the image field, 3) treatment of modifications classification methods are directed from various aspects of the CNN architecture such as image input size, number of convolutions, number of filters, pooling methods, activation functions, and logistic regression functions.

To convenient and straightforward read, this paper writes on systematic sections. This paper constructs a separable section such as the introduction. Data collection includes image acquisition and data preprocessing, a proposed method like the CNN model and evaluation metric, experiment and result, discussion, and conclusion.

II. DATA COLLECTION

A. Image Acquisition

The models are made using digital photos of various sizes, and some are taken from websites. The Photographs is taken from the front of the object. Photographs were taken only of igneous rock types, namely andesite, basalt, pumice, siltstone, mudstone, sandstone, breccia, diabase, diorite, gabbro, limestone, granite, conglomerate, marl, and chert. Fig. 1 is an example of a rock photo. The Photographs were taken in Indonesia in an area containing igneous rocks.

B. Data Processing

In this study, a dataset of 1296 images was used. The size of this image is varied then resized to an altitude of 224x224 pixels. The image used is divided into 15 labels which are the names of the igneous rocks. Table I shows the number of impressions divided into the training dataset and the validation dataset. The number of training datasets is 1001 samples divided into 15 igneous rock label classes, while the validation dataset is 296 samples divided into 15 label classes. A validation dataset is used with 296 images with 15 label classes as validation on the finished model. The dataset is worked out by directly dividing the number of pictures for the model is worked on.



Fig. 1. Image acquisition an example photo

TABLE I. D	ATASET COI	LECTION
Rock label	Training	Validation
andesite	65	14
basaltic	84	25
pumice	63	19
siltstone	38	4
mudstone	11	4
sandstone	51	10
breccia	127	36
diabase	103	38
diorite	93	33
gabbro	97	29
limestone	140	35
granite	101	36
conglomerate	10	5
marl	10	4
chert	8	4
Total	1001	296

III. METHOD

A. CNN Methods

The pipeline classification process carried out in this research can be shown in Fig. 2. The pipeline describes the step-by-step process is done in making the classification models. In the image acquisition section, the actual image is first used with varying sizes, then preprocessing is carried out by rescaling it to 224x224 pixels. The next preprocess is to crop the image with cropping2D size ((30,2),(30,2)), which means that the picture will be taken in a specific part so that the image becomes smaller. In this study, research is designed by many models, as shown in Table II and Table III. Choosing models is fundamentally based on research [1].

1) Convolutional Layer

The CNN architecture consists of different unique layers: convolution, activation, pooling, dropout, and Softmax layers with diverse functions. As a baseline model, this study uses the CNN architecture developed by [2]. Ran has used two convolutional layers with a 5x5 filter. In Fig.2, the first convolutional layer uses a 5x5 filter with stride one and padding = same. Then a pooling operation is performed using MaxPooling(3,2), namely the 3x3 filter and stride two. The first convolution result then each value of the map feature will be activated with the ReLu function. The convolution operation process itself uses (1), and the ReLu function (3).



Fig. 1. Pipeline classification igneous rocks baseline model

Layer Name	Function	Filter Sizes / Kernel	Padding	Stride	Output Tensor
Rescaling	-	-	-	-	224x224
Cropped Image	random_crop	-	-	-	192 x 192 x 3
Conv1	conv2d	5 x 5 x 3 / 64	SAME	1	192 x 192 x 3
Pool1	max_pool	3 x 3	SAME	2	96 x 96 x 3
Conv2	conv2d	5 x 5 x 64 / 64	SAME	1	96 x 96 x 3
Pool2	max_pool	3 x 3	SAME	2	48 x 48 x 64
Output	softmax	-	-	-	15 x 1

TABLE III. TUNING PARAMETER ARCHITECTU	JRES
--	------

Model	Input	Rescaling	Cropping	Conv2D	Pooling	Dense
Δ	224x224	1 /255	(30.2)(30.2)	(3,5), Padding = SAME, ReLu	MaxPooling	Dense (64)
л	2277227	1./255	(30,2),(30,2)	(64,5), Padding = SAME, ReLu	(3,3)	Dense (15)
D	224-224	1 /255		(3,5), Padding = SAME, ReLu	MaxPooling	Dense (64)
D	В 224X224	1./233	-	(64,5), Padding = SAME, ReLu	(3,3)	Dense (15)
C	224-224	1 /255	(20.2)(20.2)	(3,7), Padding = SAME, ReLu	MaxPooling	Dense (64)
C 224X224	1./255	(30,2),(30,2)	(64,7), Padding = SAME, ReLu	(3,3)	Dense (15)	
D	224-224	1 /255		(3,7), Padding = SAME, ReLu	MaxPooling	Dense (64)
D	224X224	1./255	-	(64,7), Padding = SAME, ReLu	(3,3)	Dense (15)

$$h_{ij}^k = \sum_{i \in M_j} \left((w^k \times x)_{ij} + b_k \right) \tag{1}$$

At the letter of k perform of the kth layer, h_{ij}^k show of the value of the feature, (i, j) organizes as a point of pixels, w^k means of the convolution kernel of the current layer, and b_k is the bias. The parameters of CNNs, such as the bias (b_k) and convolution kernel (w^k), are customarily trained without supervision [9]. In addition to the baseline model, this study also develops parameter tuning by eliminating cropping operations and trying 3x3 and 7x7 filters.

2) Activation Layer

The ReLU activation function is a nonlinear function that maps the convolution layer output graph to operate neurons while ward off overfitting and exceed of learning. This function has proposed by AlexNet [10]. The model leverages the ReLU activation function (2) for the output feature maps of every convolutional layer:

$$f(x) = \max(0, x) \tag{2}$$

3) Pooling Layer

The pooling layer used is to perform nonlinear downsampling, feature map size reduction, convergence acceleration, and boost computational performance [11]. Max-pooling is chosen over mean-pooling because it aims to increase the texture features [12]. Calculation of the maxpooling operation using the following formulation:

$$\mathbf{h}_{j} = \underbrace{\max}_{i \in \mathbf{R}_{j}} \alpha_{i} \tag{3}$$

where R_j is an area j in feature map for pooling α , i is the indicator of each element inward of the area, and h is the pooled feature map.

4) Fully Connected Layer

A fully connected layer has a task to connect to entire the nodes of the uppermost layer. The fully connected layers are operated to fusion the features extracted from the image, and another task is to reconstruct the two-dimensional feature map into one-dimensional feature vectors [11]. The fully connected layers have a task to map the distributed feature portrayal to the fragment of label space. The fully corresponding operation is codified by (4):

$$a_i = \sum_{j=0}^{m * n * d - 1} w_{ij} * x_i + b_i \tag{4}$$

5) Softmax Layer

As a second fully connected layer, the Softmax layer has an output probability distribution over the fifteen classes. The preeminent value of the output vector of the Softmax is recognized as the precise indicator type for the igneous rock images. The Softmax operation has supported by logistic regression. The Softmax is formulated by (5):

$$f_j(z) = \frac{e^{z_j}}{\sum_k e^{z_k}} \tag{5}$$

B. Evaluation Metric

The study uses the accuracy measure to see the model's performance. As a measure of the evaluation of the model that has done. The accuracy metric can be written as follows:

$$accuracy = \frac{tp+tn}{tp+tn+fp+fn}$$
(6)

IV. EXPERIMENT AND RESULTS

The model training is carried out using software and hardware configurations, as shown in Table IV.

TABLE IV. SYSTEM CONFIGURATIONS					
Configuration Item	Value				
CPU	Processor Intel(R) Core(TM) i7-				
	10700K CPU @ 3.80GHz, 3792 Mhz, 8				
	Core(s), 16 Logical Processor(s)				
Graphic Processor Unit	NVDIA GeForce 950Ti, 640 CUDA				
-	cores, 2 GB				
Memory	8 GB				
Harddisk	1 TB				
Solid Stated Disk	512 GB				
Pyhton	3.8.5				
Tensorflow	2.5.0				

A. Training, and Validation Result

The training has carried out using the model in Fig. 1 and models in Table III. All input images are resized to the same size, which is 224x224. The model was compiled with Adaptive Moment Estimation (ADAM) gradient descent [13], and sparse categorical entropy as a determinant of loss and model accuracy, while fitting has performed for 200 epochs. In Fig. 3, it has shown that when epoch-60, training reaches a high accuracy of 98.9%, and validation reaches an accuracy of 81.1%. As for loss, training is close to the value starting from epoch-60, which is 0.04, while validation accuracy reaches 0.94 loss. The model shows convergence values for training and validation and achieves the highest accuracy and lowest loss at epochs above 150. At a final curation at epoch-200, training accuracy reaches 97.9%, while validation accuracy reaches 83.11%, loss value for training accuracy at epoch-200 is 0.049, and validation loss is 0.94.



Fig. 3. Average accuracy (a) and Loss curves (b) for the training and validation dataset using samples of 224x224.

The evaluation model in Table VI uses an accuracy measure. Meanwhile, Fig. 4 shows the validation results using model Fig.1, which is presented in a confusion matrix. In Fig. 4, row and column labels are igneous rock class labels, namely andesite, basalt, pumice, siltstone, mudstone, sandstone, breccia, diabase, diorite, gabbro, limestone, granite, conglomerate, marl, and chert, respectively, and presented in section 2 on image acquisition. The number of validation datasets is 296 images spread over 15 rock classes.

[14,	0,	0,	0,	0,	0,	0,	0,	0,	1,	0,	0,	0,	0,	0],
[0,	12,	0,	0,	0,	0,	1,	1,	0,	0,	0,	0,	0,	0,	0],
[0,	0,	13,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0],
[0,	0,	0,	з,	0,	0,	0,	1,	0,	0,	0,	0,	0,	0,	0],
[0,	0,	0,	0,	4,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0],
[0,	1,	0,	1,	ø,	9,	0,	2,	0,	0,	0,	0,	0,	0,	0],
[0,	ø,	1,	0,	ø,	0,	32,	1,	0,	ø,	0,	0,	0,	ø,	0],
[0,	2,	0,	0,	ø,	0,	1,	20,	1,	ø,	0,	0,	0,	ø,	0],
[0,	4,	2,	0,	ø,	0,	0,	2,	28,	ø,	0,	0,	0,	ø,	0],
[0,	6,	1,	0,	0,	0,	0,	4,	2,	28,	0,	0,	0,	0,	0],
[0,	0,	2,	0,	0,	1,	2,	7,	2,	0,	35,	1,	0,	0,	0],
[0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	35,	0,	0,	0],
[0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	5,	0,	0],
[0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	4,	0],
[0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	4]]

Fig. 4. Confusion Matrix model Fig. 1

Considering the values in Fig. 4, model A managed to predict precisely as many as 246 images, and there were still errors of 50 images. In Table V, there are still erroneous prediction classes, especially in andesite, basalt, siltstone, mudstone, sandstone, breccia, diabase, diorite, gabbro, limestone, and granite classes. In the diorite class, the error that occurred was 31.7% of the 41 predicted images. Then the limestone class occurs 30% wrong predictions from a total of 50 predicted images. If the total prediction errors are calculated, 16.7% of the 296 images are predicted to be an error.

TABLE V. TRUE AND FALSE PREDICTION

	andesite	bas al t	pumice	siltstone	muds tone	sands tone	breccia	diabase	diorite	gabbro	limestone	granite	conglomerate	marl	chert
True Prediction	14	12	13	3	4	9	32	20	28	28	35	35	5	4	4
False Prediction	1	2	0	1	0	4	2	4	8	13	15	0	0	0	0
Total	15	14	13	4	4	13	34	24	36	41	50	35	5	4	4

The comparison between the models in Table VI shows that model A has better validation accuracy than models B, C, or D. The parameter tuning performed on B, C, and D still cannot exceed model A. The cropping process on the image in the specified area is still a strength in the classification of igneous rocks [2]. When viewed, the accuracy of model D is almost close to model A. However, model D still has a more significant loss than model A, and the difference in loss is 0.1676.

FABLE VI. ACCURACY AND LOSS VALIDATION DATASET

		Nidel							
Metric	Α	В	С	D					
Acc.	0.831	0.814	0.8068	0.8271					
Loss	0.942	1.122	1.1057	1.1098					

B. Evaluation Execution Time

Evaluation of time is also a consideration in determining results. Time is often used as a backup for applications to be efficient. In convolution-based models, the number of convolution layers, rescaling, cropping, flatten, and Softmax layers will affect a model's time and predictive output. In Table VII, it can see that there is an execution time until the prediction results have obtained. The time shown in Table VII is for training and validation time with a total dataset of 1297 images.

CNN architecture changes will affect the execution time to get the training model with the best accuracy [21]. The addition of a convolutional layer also does not always get good accuracy for the working domain. Likewise, tuning filter size does not necessarily get high accuracy. This accuracy is evidenced by the experiment, results in the accuracy in Table VI, and execution time in Table VII.

TABLE VII. EXECUTION TIME COMPARISON							
Time	Model ^a						
Time	Α	В	С	D			
Execution Time (minute)	10.8	14.56	11.02	107.67			
		^a based on c	onfiguration sys	stem in Table IV			

V. DISCUSSION

In Fig. 3, it has noted that the model still has difficulty getting fittings; the difference in accuracy at the training and validation stages is still significant, which is around 14.49%. In the graph Fig. 3, it has seen that the training reaches convergent accuracy when the epoch-30. In contrast, the loss reaches a convergent value at epoch-40. This evidence is described by research from which makes a model with several convolutions as many as two layers and uses rescaling and cropping techniques [2]. Another study that supports this research uses many layers of convolution to get high accuracy [14]. Model B, C, and D obtain high accuracy, although not yet exceeded model A. The model without rescaling and cropping are also in the study [15], which uses VGG16 [16] as a model classification. In this case, VGG16 takes a long time to get good accuracy. The convergence of accuracy during training at the beginning of the epoch has demonstrated by research [17], which used the same image model to predict mineral images in rocks. Comparing with Fig. 5, Fig. 6, and Fig. 7, the accurate measurement is almost the same; the convergence reaches less than epoch-50. For models A, B, C, and D still have difficulty achieving high accuracy. If we used a naked eye to differentiate on geological rocks, it is hard to identify and detect the object. It caused the rocks almost to have the same color, texture, and The geologist tries to identify the rocks by pattern. microscope aid and make a thin section of rocks [15]. It is possible to identify the rock because the thin section has a sharp surface about texture, color, and pattern.



Fig. 5. Average accuracy (a) and Loss curves (b) for the training and validation dataset using samples of 224x224 for model B.

The model also still got prediction failures of 16.7% and only reached 83.3% correct; this is evidenced in Table VI. Some classes are still challenging to predict, such as andesite, basalt, siltstone, mudstone, sandstone, breccia, diabase, diorite, gabbro, limestone, and granite [2]. The size of the



Fig. 6. Average accuracy (a) and Loss curves (b) for the training and validation dataset using samples of 224x224 for model C.



Fig. 7. Average accuracy (a) and Loss curves (b) for the training and validation dataset using samples of 224x224 for model D.

image pixels, the similarity of colors, and the image structure make the model that is being worked on the need to be improved. Especially in precise image sizes, many layers of convolution, considerations using regularization, size filters, and pooling. Another consideration is using the dropout function to avoid overfitting too high and far from the training accuracy [8], [19], [22]. Same as models B, C, and D, the difficult to identify, the class has a higher error like gabbro and limestone. The high error on class gabbro and limestone considers the number of pictures, pixel size, and cropping technique. It can be proved by the experiment in Fig. 3, Fig. 5, Fig. 6, and Fig. 7 that use model in Table 3. Using tune parameters like rescaling, cropping, many layers of convolutional did not guarantee to reach a high accuracy, especially in this domain. Some experiment to change the parameter has the same problem that is at the CNN part especially on layer function. The investigation tries to delete the cropping and change the filter, see Table 3; the result has the same that there is no influence on cropping, rescaling, and filtering to high accuracy [16]. Convolutional layer, regularization, and normalization are the best impact on the CNN process [16].

If we see Table VIII, every model has different number of trainable parameters or feature extraction. Model B and model D are the models that construct without cropping.

Table VIII shows that even though the process does not use cropping method, the model still has difficulty making a high accuracy. The number of trainable parameters bigger than models A and C is not guaranteed to reach high accuracy. Indeed, the error has made more increase than with the cropping process, see Table VI. Color prediction using CNN has proven the ability to increase accuracy. Several things can improve accuracy, as shown in Table VI, namely by making parameter tuning involving many layers of convolution, resizing, rescaling, cropping, pooling technique, ReLU function, and Softmax Function [18] [19] [20]. The use of ADAM as an optimizer for gradient descent helps to focus more on feature search to represent specific features when used in the prediction model [13].

TABEL VIII. EXECUTION TIME COMPARISON							
Model	Number of Trainable Parameter						
А	9,443,315						
В	12,851,187						
С	9,448,139						
D	12,856,011						

VI. CONCLUSION

Fashioning CNN with contrasting layers has influenced to achieve high accuracy. Encouragement of filter, stride, and cropping has helped make a diverse feature, then alteration of a constituent has trained to classify for this domain. Model A composed of rescaling, copping, conv(3,5), and Maxpooling have an accuracy of 0.831 and loss is 0.942. This outcome is more noteworthy than models B, C, and dan D, which have an accuracy of 0.814, 0.8068, and 0.8271. Nevertheless, the figure still has difficulty attaining high accuracy, especially identifying diorite class and limestone class. Error prediction to this confirmation around 31.7% and 30%, respectively. If the total prediction errors are learned, then there is 16.7% of the error predicted class. This fact has caused by the failure to identify objects when conducting the CNN framework. The diverse of color, shape, texture, and pattern makes it difficult to decipher the feature.

Several challenges are employed on CNN framework, like the length of convolution, filter, stride, dropout, regularization, and normalization to improve CNN's identification of igneous rocks. Generally, the architecture CNN to be dominant reconstruct to get better accuracy.

ACKNOWLEDGMENT

We would like to Dr. Djoko and Mrs. Fitriyani thanks; they are a geologist rocks interpreter. They have provided photos of igneous rocks and help to identify and give a label class for every image.

References

- G. Fan, F. Chen, D. Chen, and Y. Dong, "Recognizing multiple types of rocks quickly and accurately based on lightweight CNNs model," *IEEE Access*, vol. 8, pp. 55269–55278, 2020, doi: 10.1109/ACCESS.2020.2982017.
- [2] X. Ran, L. Xue, Y. Zhang, Z. Liu, X. Sang, and J. He, "Rock classification from field image patches analyzed using a deep convolutional neural network," *Mathematics*, vol. 7, no. 8, pp. 1–16, 2019, doi: 10.3390/math7080755.
- [3] G. Cheng and W. Guo, "Rock images classification by using deep convolution neural network," *J. Phys. Conf. Ser.*, vol. 887, no. 1, 2017, doi: 10.1088/1742-6596/887/1/012089.
- [4] J. Chen, T. Yang, D. Zhang, H. Huang, and Y. Tian, "Deep learning

based classification of rock structure of tunnel face," *Geosci. Front.*, vol. 12, no. 1, pp. 395–404, 2021, doi: 10.1016/j.gsf.2020.04.003.

- [5] A. K. Patel and S. Chatterjee, "Computer vision-based limestone rocktype classification using probabilistic neural network," *Geosci. Front.*, vol. 7, no. 1, pp. 53–60, 2016, doi: 10.1016/j.gsf.2014.10.005.
- [6] A. Sharma, X. Liu, X. Yang, and D. Shi, "A patch-based convolutional neural network for remote sensing image classification," *Neural Networks*, vol. 95, pp. 19–28, 2017, doi: 10.1016/j.neunet.2017.07.017.
- [7] B. A. Robson, T. Bolch, S. MacDonell, D. Hölbling, P. Rastner, and N. Schaffer, "Automated detection of rock glaciers using deep learning and object-based image analysis," *Remote Sens. Environ.*, vol. 250, no. August, 2020, doi: 10.1016/j.rse.2020.112033.
- [8] J. Feng, G. Qing, H. Huizhen, and L. Na, "Feature Extraction and Segmentation Processing of Images Based on Convolutional Neural Networks," *Opt. Mem. Neural Networks (Information Opt.*, vol. 30, no. 1, pp. 67–73, 2021, doi: 10.3103/S1060992X21010069.
- [9] B. Liu, Y. Zhang, D. J. He, and Y. Li, "Identification of apple leaf diseases based on deep convolutional neural networks," *Symmetry* (*Basel*)., vol. 10, no. 1, 2017, doi: 10.3390/sym10010011.
- [10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst. 25 (NIPS 2012)*, vol. 25, pp. 1–9, 2012, doi: 10.1201/9781420010749.
- [11] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, 2016, doi: 10.1016/j.neucom.2015.09.116.
- [12] Y. Boureau, J. Ponce, J. P. Fr, and Y. Lecun, "A Theoretical Analysis of Feature Pooling in Visual Recognition," in *International Conference on Machine Learning*, 2010, pp. 111–118, [Online]. Available: https://www.di.ens.fr/sierra/pdfs/icml2010b.pdf.
- [13] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.
- [14] G. Fan, F. Chen, D. Chen, Y. Li, and Y. Dong, "A Deep Learning Model for Quick and Accurate Rock Recognition with Smartphones," *Mob. Inf. Syst.*, vol. 2020, 2020, doi: 10.1155/2020/7462524.
- [15] W. Ren, M. Zhang, S. Zhang, J. Qiao, and J. Huang, "Identifying rock thin section based on convolutional neural networks," *Proc. 2019 9th Int. Work. Comput. Sci. Eng. WCSE 2019*, vol. 052, pp. 345–351, 2020, doi: 10.18178/wcse.2019.06.052.
- [16] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arxiv*, pp. 1–14, 2015, [Online]. Available: http://arxiv.org/abs/1409.1556.
- [17] L. Kocarev, J. Makraduli, and P. Amato, "rock classification in petrographic thin section images based on concatenated convolutional neural networks," vol. 9, no. 1, pp. 497–517, 2005.
- [18] S. Niekum, "Reliable rock detection and classification for autonomous science," C. Thesis, 2005, [Online]. Available: http://people.cs.umass.edu/~sniekum/pubs/SeniorThesis.pdf.
- [19] J. Maitre, K. Bouchard, and L. P. Bédard, "Mineral grains recognition using computer vision and machine learning," *Comput. Geosci.*, vol. 130, no. February, pp. 84–93, 2019, doi: 10.1016/j.cageo.2019.05.009.
- [20] Y. C. Zhang and A. C. Kagen, "Machine Learning Interface for Medical Image Analysis," *J. Digit. Imaging*, vol. 30, no. 5, pp. 615– 621, 2017, doi: 10.1007/s10278-016-9910-0.
- [21] Y. Zhang, G. Wang, M. Li, and S. Han, "Automated classification analysis of geological structures based on images data and deep learning model," *Appl. Sci.*, vol. 8, no. 12, 2018, doi: 10.3390/app8122493.
- [22] Y. Zhang, M. Li, S. Han, Q. Ren, and J. Shi, "Intelligent identification for rock-mineral microscopic images using ensemble machine learning algorithms," *Sensors (Switzerland)*, vol. 19, no. 18, 2019, doi: 10.3390/s19183914.