Mask R-CNN for rock-forming minerals identification on petrography, case study at Monterado, West Kalimantan

Muhammad Ridwan Iyas, Nugroho Imam Setiawan*, and I Wayan Warmada

Department of Geological Engineering, Universitas Gadjah Mada, Indonesia

Abstract. This paper explores the experiment of Deep Learning method using Mask Region-Convolutional Neural Network (Mask R-CNN) to identify rock-forming minerals on thin section images from petrographic observation in igneous rocks, which are plagioclase, quartz, K-feldspar, pyroxene, and hornblende. Train and validation dataset consisted of 2 quartz diorites and 1 granodiorite from Monterado, West Kalimantan, 1 quartz diorite and 1 granite from Nangapinoh, West Kalimantan, and 7 andesite and 2 basalts from Bangli, Bali, while test dataset consisted of 3 quartz diorites from Monterado, West Kalimantan. This study uses 4 Mask R-CNN models, which is influenced by the lighting on polarizing microscope and using ResNet-50 architecture (Model A) or ResNet-101 (Model B), and the models that is not affected by the lighting on polarizing microscope and using ResNet-50 architecture (Model C) or ResNet-101 (Model D). From Average Precision scores, it was found that Model B has the highest score (58.0%), followed by Model A (57.8%), Model C (45.8%), and Model D (43.6%). In conclusion, the lighting of polarizing microscope is a major factor to give a better performances of Mask R-CNN models by 12%-14.4%, while the type of backbone architecture on Mask R-CNN models was not too consequential.

1 Introduction

Rock classification is an important work for a geologist, and could be achieved through field observations of hand-specimen samples, as well as observations in the laboratory using a petrographic microscope, X-Ray Diffraction (XRD), or X-Ray Fluorescence (XRF). Petrographic observations is the most commonly used research method because it considered more effective and efficient. However, this method often takes a lot of time and sometimes has a high error rate, thus an artificial intelligence system under human supervision would be required to automate petrographic mineral identification in rock classification [1].

1.1. Related works

There have been a number of studies that implemented Computer Vision method for petrographic work. Baykan and Yilmaz [2] performed semantic segmentation on quartz, muscovite, biotite, chlorite, and opaque minerals in igneous rocks using Artificial Neural Network with a Fully Connected Network structure with three hidden layers. In the similar way, Izadi *et al.* [1] performs semantic segmentation on biotite, apatite, andalusite, muscovite, orthoclase, aegirine, quartz, actinolite, olivine, glass, talc, topaz, kyanite, sanidine, epidote, garnet, nepheline, nosean, hornblende, analcime, augite, and hypersthene in igneous rocks. Samples for dataset were taken at a microscope rotation angle of 0°, 45°, 90°, 135°,

180°, 225°, 270°, and 315°, with a resolution of 300×250 pixels and trained using machine learning on mineral colour and texture components. Bukharev *et al.* [3] conducted instance segmentation work on thin section images of sandstone using LinkNet and FCNN. In this study, 9000 individual grains in sandstone samples manually segmented, and then trained using FCNN to produce a model that localize object and predict binary mask. More recently, Maitre *et al.* [4] performs the work of identifying and calculating individual sand-sized minerals using Classification and Regression Trees, k-Nearest Neighbors, and Random Forest.

The goals of this study is to determine a geological factor; represented by the type of lights on petrographic microscope and non-geological factor; represented by the architecture of CNN, on Deep Learning models that built based on Mask R-CNN algorithm to identify rock forming mineral on thin section images. Plagioclase, quartz, and K-feldspar were chosen based on QAPF classification for plutonic rocks [5] and volcanic rocks [6] to determine igneous rocks that consisted of <90% mafic minerals, and hornblende with pyroxene (as well as plagioclase) were chosen based on classification of gabbroic rocks [5].

1.2. Mask R-CNN

Mask R-CNN [7] is an instance segmentation algorithm that extends Faster R-CNN [8] by adding a branch for

* Corresponding author: nugroho.setiawan@ugm.ac.id

predicting an object mask in parallel with the existing branch for bounding box recognition [7]. Mask R-CNN was developed in 2017 by Facebook AI Research team and by that time was the state of the art of instance segmentation algorithm. Mask R-CNN was built on backbone of convolutional neural network with ResNet architecture [9], Feature Pyramid Network [10], Region Proposal Network [8], RoI Align, and head network of Mask R-CNN.

2 Dataset collection

To support this research, a petrographic thin section image dataset (Table. 1) had compiled consisting of training dataset, validation dataset, and test dataset. The thin section images were taken using a polarizing microscope on Plane Polarized Light (PPL) and Cross-Polarized Light (XPL) with a 5× magnification lens with rotational angle of microscope of 0°, 15°, 30°, and 45°, and dimensions of 960×540 pixels in the Laboratory of Optical Geology, Department of Geological Engineering, UGM. As a light wave, PPL has an electrical component that travels in vector of one-direction of the polarization plane as a sine wave towards the direction of propagation and generated by a polarizer, while XPL was generated by an analyzer in polarizing microscope [11].

Training and validation dataset composed of 1 granodiorite and 2 quartz diorite samples of Ilmawan [12] in Monterado, West Kalimantan, 1 quartz diorite and 1 granite samples of Setiawan *et al.* [13] in Nangapinoh, West Kalimantan, and 7 andesite and 2 basalt samples of Nadirah [14] in Bangli, Bali. As a case study, the test dataset consisted of 3 quartz diorite samples of Ilmawan [12] that composed of plagioclase, quartz, opaque mineral, hornblende, pyroxene, and biotite as primary minerals, and clay minerals, chlorite, epidote, sericite, and K-feldspar as secondary minerals. Quartz diorite samples in test dataset is part of Quartz Diorite Unit [12] and Mensibau Granodiorite [15] (Figure 1).

Fig. 1. The location of case study [12] (red box) in Monterado, West Kalimantan on the Geological Map of Singkawang [15] (with modification).

3 Methodology

Fourteen samples of training dataset and validation dataset have annotated to the boundaries of plagioclase, pyroxene, K-feldspar, hornblende, and quartz using VGG Image Annotator [16]. The label of annotations based on the effect of lighting on polarizing microscope are:

1. 'pl', 'px', 'kfs', 'hb', and 'qz', for dataset that are affected by the lighting on polarizing microscope, performed by combining mineral labels on PPL and XPL appearance ('Dataset 1').
2. 'pl_ppl', 'pl_xpl', 'px_ppl', 'px_xpl', 'kfs_ppl', 'kfs_xpl', 'hb_ppl', 'hb_xpl', 'qz_ppl', and 'qz_xpl', for dataset that are not affected by the lighting on polarizing microscope, performed by separating mineral labels on PPL and XPL ('Dataset 2').

Table. 1. Number of annotated minerals on train and validation dataset

Sample	Name	Location	Reference	Train Dataset					Validation Dataset				
				pl	px	kfs	hb	qz	pl	px	kfs	hb	qz
II.153	Quartz diorite	Monterado, West Kalimantan	Ilmawan [12]	22	-	-	-	58	8	-	-	-	22
II.114				22	-	9	-	64	4	-	-	-	22
II.182	Grano-diorite			2	-	25	-	37	2	-	8	-	12
120313 TO3H(1)	Quartz diorite	Nangapinoh, West Kalimantan	Setiawan <i>et al.</i> [13]	1,171	278	156	534	500	250	78	35	146	118
120313 TO3H(2)	Granite			480	-	579	-	912	136	-	56	-	152
L.1849	Andesite	Bangli, Bali	Nadirah [14]	160	60	-	-	-	50	18	-	-	-
L.1888				340	80	-	-	-	122	18	-	-	-
L.1905				414	96	-	-	-	118	28	-	-	-
L.1926				218	50	-	-	-	72	22	-	-	-
L.1963A				112	10	-	-	-	34	2	-	-	-
L.1963B				68	6	-	-	-	20	2	-	-	-
L.1974				192	60	-	-	-	32	22	-	-	-
B.01A				24	66	-	-	-	8	20	-	-	-
B.01B	48			18	-	-	-	18	6	-	-	-	
Total				3,273	724	769	534	1,571	874	216	99	146	326

Both types of training dataset and validation dataset are trained using Mask R-CNN source code [17] with two types of backbone architectures, namely ResNet-50 and ResNet-101 [9]. In addition, ResNet-50 has 50 layers of Residual Network while ResNet-101 has 101 layers of Residual Network to perform feature extraction on Mask R-CNN. Four Deep Learning models that generated from the training process would be inferred through the test dataset. The value of Average Precision from each models calculated using this equation [3]:

$$AP = \frac{1}{N} \sum_N \frac{TP}{TP+FN+FP} \quad (1)$$

AP = Average Precision
 N = Number of samples
 TP = True Positive
 FN = False Negative
 FP = False Positive

3.1. Training preparation

In order to train Deep Learning model, we need first to adjust the hyperparameter of Mask R-CNN. Backbone architecture uses ResNet-50 and ResNet-101 with Feature Pyramid Network. We use Stochastic Gradient Descent and Momentum algorithm of 0.9 for backpropagation.

We train for total 100 epochs with learning rate of 10^{-3} for the first of 40 epochs, and ending with learning rate of 10^{-4} for the rest of 60 epochs later. Likewise, we adjust to train head layers for the first of 40 epochs and then for the next 60 epochs we train all layers. We use an anchor in the Region Proposal Network for 32, 64, 128, 256, and 512 pixels as well as Non-Maximum Suppression threshold of 0.7. Gradients are clipped to 5.0 and weights are decayed by 10^{-4} . Furthermore, we initialized the model using weights obtained from pre-training on MSCOCO dataset [18]. The dataset was augmented using horizontal and vertical flip, random rotation, random translation, random shear, random crop, Gaussian blurring, and grayscale filter.

3.2. Training stage

The training stages are carried out using Google Colaboratory, which is supported by NVIDIA K80 GPU with 8-10 hours runtime per 1 Deep Learning model. From the training stage, 4 Deep Learning Models are generated, namely Model A, Model B, Model C, and Model D. Model A and Model B are models that contains 'Dataset 1', which Model A was trained using ResNet-50 backbone while Model B using ResNet-101. On the other hand, Model C and Model D are models that contains 'Dataset 2', which Model C were trained using ResNet-50 backbone while Model D using ResNet-101.

The performances of four models illustrated on Figure 2. Model A has a validation accuracy score of 72.3%, while Model B, Model C and Model D each have an accuracy value of 73.1%, 68.2%, and 69.6%. From the results of the training stages, it could be seen that Dataset 1 has the highest validation accuracy score compared to Dataset 2, because Model A and Model B are the top two models that have the highest accuracy score. Furthermore,

the uses of ResNet-101 backbone has a slightly impact for rising validation accuracy score of the model, compared with the ResNet-50 one despite the fact that ResNet-101 has a higher computational load.

Fig. 2. Diagram of validation loss and validation accuracy, based on training result on Mask R-CNN. Model B has a highest validation accuracy (73.1%), followed by Model A (72.3%), Model D (69.6%), and Model C (68.2%).

3.3. Inferencing stage

The four of Mask R-CNN models from the training stages then conducted to inferencing stages on test dataset. To give an illustration on inferencing stage of Model B, look at Figure 3. We calculated each of predicted label that come from Mask R-CNN models, and then match it with its true label. To demonstrate that meanings, we use Model B as an example.

On the PPL appearance, Model B were correctly identify 5 plagioclase, 6 pyroxene, and 1 quartz of sample II.186, although 1 plagioclase identified as quartz and 1 pyroxene identified as hornblende. Not to mention on the XPL appearance, model B can identify 9 plagioclase, 5 pyroxene, and 1 quartz properly, yet 1 plagioclase was identified as pyroxene and 1 pyroxene was identified as hornblende. For sample II.143, Model B have predicted 3 plagioclase, 1 hornblende, and 1 quartz as their labels on PPL appearance, however 2 quartz were identified as plagioclase, and 3 hornblende along with 1 quartz were identified as pyroxene. On the XPL appearance, 2 plagioclase, 1 hornblende, and 5 quartz were precisely predicted, yet 1 plagioclase was identified as pyroxene, 3 quartz were identified as plagioclase, and 2 hornblende were identified as pyroxene. Furthermore, inferencing stage on sample II.176 on PPL appearance generate result as 2 plagioclase, 1 K-feldspar, and 6 quartz were correctly identified, although 1 K-feldspar and 1 quartz were identified as plagioclase. On the XPL appearance, Model B was identified properly 1 plagioclase, 1 K-feldspar, and 7 quartz. To summarize all of the inferencing stage, we suggest to the reader to look at the confusion matrix on Table. 2.

Fig. 3. Result of inferencing on Model B in test dataset within PPL and XPL appearances: II.186 (upper), II.143 (middle), and II.176 (lower) from Ilmawan [12]. Average Precision of Model B is 58.0%, which is the highest score among other models.

4 Discussion

After calculating predicted label for all four models on confusion matrix, we use Average Precision (in %) to determine which models that come up with the best score. As a result, it was found that Model B has the highest

Average Precision score compared to other Mask R-CNN models that is 58%, followed by Model A, Model C, and Model D, which their Average Precision values are 57.8%, 45.8%, and 43.6%, respectively. As shown above at diagram of validation loss/accuracy (Figure 2), and bar chart of Average Precision (Figure 4), we reveal that Model A and Model B are the top two models, which

Table. 2. Confusion matrix of Model A, B, C, and D on PPL and XPL appearances

Model A

PPL		True label					XPL	True label					
		p1	px	kfs	hb	qz		p1	px	kfs	hb	qz	
Predicted label	p1	5	-	2	-	2	Predicted label	p1	6	-	-	-	-
	px	-	4	-	1	-		px	1	4	-	1	-
	kfs	-	-	1	-	-		kfs	-	-	1	-	-
	hb	-	1	-	3	-		hb	-	1	-	1	-
	qz	-	-	-	-	10		qz	1	-	2	-	12

Model C

PPL		True label					XPL	True label					
		p1	px	kfs	hb	qz		p1	px	kfs	hb	qz	
Predicted label	p1	3	1	1	-	5	Predicted label	p1	9	-	-	-	-
	px	-	1	-	-	-		px	3	1	1	2	2
	kfs	-	-	1	-	-		kfs	1	-	1	-	-
	hb	-	-	-	1	-		hb	-	-	-	-	-
	qz	-	-	-	-	5		qz	2	-	-	-	11

Model B

PPL		True label					XPL	True label					
		p1	px	kfs	hb	qz		p1	px	kfs	hb	qz	
Predicted label	p1	10	-	1	-	3	Predicted label	p1	12	-	-	-	3
	px	-	6	-	3	1		px	2	5	-	2	-
	kfs	-	-	1	-	-		kfs	-	-	1	-	-
	hb	-	1	-	1	-		hb	-	1	-	1	-
	qz	1	-	-	-	8		qz	-	-	-	-	13

Model D

PPL		True label					XPL	True label					
		p1	px	kfs	hb	qz		p1	px	kfs	hb	qz	
Predicted label	p1	6	3	-	1	-	Predicted label	p1	7	-	1	-	2
	px	-	2	-	3	-		px	-	2	-	1	-
	kfs	-	-	-	-	-		kfs	-	1	1	1	-
	hb	-	-	-	3	-		hb	-	-	-	-	-
	qz	-	-	1	-	8		qz	1	-	-	-	13

Fig. 4. Bar chart of Average Precision of Model A, B, C, and D

outperforms another models. Model A and Model B have approximately 12%-14.4% difference in comparison with Model C and Model D. Hence, we agree that the lighting of polarizing microscope is a substantial variable to generate robust models on petrographic samples with Mask R-CNN. In contrast with the first variable, backbone architecture of Mask R-CNN that we studied did not give any huge impact to Average Precision score. ResNet-101 on Model B has a tiny impact to surpass ResNet-50 on Model A by 0.2% and conversely, ResNet-50 on Model C has a 2.2% better than ResNet-101 on Model D.

Moreover, we put all of Average Precision score together from each models to generate mean Average Precision diagram based on minerals object and the lighting of polarizing microscope (Figure 5). As can be seen on the bar chart, quartz, plagioclase, and K-feldspar have a higher Average Precision scores on the XPL appearance (78.9%, 67.3%, and 47.9% respectively) than on the PPL appearance (70.9%, 53%, and 33.3% respectively). On the other hand, pyroxene and hornblende have a higher Average Precision scores on the PPL appearance (49% and 55.7% respectively) than on the XPL appearance (42% and 14.5% respectively).

Based on a bar chart on Figure 5, the mean of Average Precision from highest to lowest are quartz, plagioclase, pyroxene, K-feldspar, and hornblende. Quartz has the highest Average Precision scores compared to other minerals with 74.9%. On the PPL appearance, quartz has a clearer colourless looks than other colourless minerals, such as plagioclase and K-feldspar, hence this factor

makes quartz more easily identified on PPL. On the XPL appearance, quartz has a wavy extinction that distinguished to other minerals, although quartz, plagioclase, and K-feldspar have more or less the same interference colour.

Fig. 5. Mean Average Precision based on minerals object and the lighting of polarizing microscope.

Plagioclase comes at the second place on Average Precision score with 60.1%. Plagioclase often predicted as quartz or K-feldspar, since plagioclase sometimes has a clear colourless or cloudy colourless appearance on PPL. On the XPL appearance, plagioclase has a relatively high Average Precision score that may come through its polysynthetic twin factor in plagioclase.

Pyroxene comes at third place on Average Precision score with 45.5%. On the PPL appearance, pyroxene in training dataset and test dataset have a similar green

colour, so when it comes to inferencing stage, pyroxene is more easily identified than on the XPL appearance, which it has a variety of interference colours. Furthermore, K-feldspar comes at fourth place on Average Precision score with 40.6%. On the PPL appearances, K-feldspar that has a cloudy colourless looks are often identified as plagioclase, as well as on the XPL appearance which K-feldspar are sometimes identified as quartz and plagioclase due to its first-order grey interference colour. Lastly, hornblende comes at the fifth place on Average Precision score with 35.1%. Seven out of 15 prediction of hornblende were identified as pyroxene, while 6 out of 8 wrong prediction occurred on the XPL appearance. This is probably due to the characteristics of hornblende such as mineral colour, cleavage, and interference colour of hornblende are similar with pyroxene. In addition, hornblende have relatively small amount of samples in the training dataset with 7.77% of total annotated minerals, thereby it may affecting inferencing stage on hornblende.

5 Conclusions

We have presented the uses of Mask R-CNN to identify rock-forming minerals in igneous rocks using thin section samples. We have discovered that the lighting of polarizing microscope is a substantial factor that give better performances for Mask R-CNN models about 12%-14.4%. Based on the calculation of Average Precision, quartz, plagioclase, and K-feldspar have a higher score on the XPL appearance, while pyroxene and hornblende have a higher score on the PPL appearance. In addition, we have determined minerals with the highest Average Precision is quartz, followed by plagioclase, pyroxene, K-feldspar, and hornblende. Furthermore, the choices of ResNet-50 or ResNet-101 as a backbone architecture of Mask R-CNN models was not too consequential, with reference to their Average Precision scores.

In future, we would like to include igneous rock texture such as crystal size, and intensity of alteration as factors to consider the forthcoming work. Balancing the number of mineral based on the labels in train and validation dataset and hyperparameter tuning are important to note for the future investigation.

This research was financially supported by Ministry of Research and Higher Education of Indonesia under RTA program with contract number 2488/UN1.P.III/DIT-LIT/PT/2020.

References

1. H. Izadi, J. Sadric, M. Bayatie. *An intelligent system for mineral identification in thin sections based on a cascade approach*. Comput. Geosci. **99**, 37–49 (2017)
2. N.A. Baykan, N. Yilmaz. *Mineral identification using color spaces and artificial neural networks*. Comput. Geosci. **36**, 91–97 (2010)
3. A. Bukharev, S. Budenny, O. Lokhanova, B. Belozarov, E. Zhukovskaya, *The Task of instance segmentation of mineral grains in digital images of rock samples*. IC-AIAI, 18–23 (2018)
4. J. Maitre, K. Bouchard, L.P. Bédard. *Mineral grains recognition using computer vision and machine learning*. Comput. Geosci. **130**, 84–93 (2019)
5. A. Streckeisen. *To each plutonic rock its proper name*. Earth Sci. Rev. **12**, 1–33 (1976)
6. A. Streckeisen. *Classification and Nomenclature of Volcanic Rocks, Lamprophyres, Carbonatites and Melilitic Rocks IUGS Subcommission on the Systematics of Igneous Rocks*. Geologische Rundschau. **69**, 194–207 (1980)
7. K. He, G. Gkioxari, P. Doll'ar, R. Girshick. *Mask R-CNN*. IEEE. ICCV. 2961–2969 (2017)
8. S. Ren, K. He, R. Girshick, J. Sun. *Faster R-CNN: Towards real-time object detection with region proposal networks*. NIPS, 91–99 (2015)
9. K. He, X. Zhang, S. Ren, J. Sun. *Deep residual learning for image recognition*. ICCV. 770–778 (2016)
10. T.Y. Lin, P. Doll'ar, R. Girshick, K. He, B. Hariharan, S. Belongie. *Feature pyramid networks for object detection*. CVPR (2017)
11. W.D. Nesse. *Introduction to Optical Mineralogy*. 11–19 (2004)
12. I. Ilmawan. *Geologi, alterasi hidrotermal dan mineralisasi pada endapan epitermal sulfidasi rendah–menengah di Desa Monterado, Kecamatan Monterado, Kabupaten Bengkayang, Provinsi Kalimantan Barat*. Thesis (2019). In Bahasa
13. N.I. Setiawan, Y. Osanai, N. Nakano, T. Adachi, L.D. Setiadji, J. Wahyudiono. *Late triassic metatonalite from the Schwaner Mountains in West Kalimantan and its contribution to sedimentary provenance in the Sundaland*. Berita Sedimentologi. **28**, 4–12 (2013)
14. Z. Nadirah. *Karakteristik dan evolusi magma syn-kaldera dan post-kaldera Batur, Kabupaten Bangli, Bali*. Thesis (2019)
15. N. Suwarna, R.P. Langford. *Peta geologi regional lembar Singkawang skala 1:250.000* (1993)
16. A. Dutta, A. Gupta, A. Zissermann. *VGG Image Annotator*. <https://robots.ox.ac.uk/~vgg/software/via/>, Version: 2.0.5 (2016)
17. W. Abdulla. *Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow*. https://github.com/matterport/Mask_RCNN (2017)
18. T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Doll'ar, C.L. Zitnick. *Microsoft COCO: Common objects in context*. ECCV (2014)