

Supplemental Materials for

**The application of “transfer learning” in optical microscopy: the petrographic
classification of metallic minerals**

**Yi-Wei Cai¹, Kun-Feng Qiu^{1*}, Maurizio Petrelli², Zhao-Liang Hou³, M. Santosh^{1,4}, Hao-
Cheng Yu¹, Ryan T. Armstrong⁵, Jun Deng^{1,6}**

¹ Frontiers Science Center for Deep-time Digital Earth, State Key Laboratory of Geological Processes and Mineral Resources, School of Earth Sciences and Resources, China University of Geosciences, Beijing 100083, China

² Department of Physics and Geology, University of Perugia, Perugia 06100, Italy

³ Department of Geology, University of Vienna, Vienna 1090, Austria

⁴ Department of Earth Science, University of Adelaide, Adelaide SA 5005, Australia

⁵ School of Minerals and Energy Resources Engineering, University of New South Wales, Sydney, NSW 2052, Australia

⁶ Geological Research Institute of Shandong Gold Group Co., Ltd., Jinan 250013, China

Contents of this file

Method comparisons

Reference cited in supplemental materials

Supplemental Figures 1 to 7, Supplemental Tables 1 and 2

The comparison between the model with augmentation and without augmentation

To improve the model's generalization ability and reduce overfitting, we increased the size and the robustness of the training set and introduced variations that could be found in "real world data" using six typical offline data augmentation methods (Shorten and Khoshgoftaar 2019; Feng et al. 2021): the random erasing, the flipping, the brightness adjust, the random zoom, the random contrast and the random saturation. Firstly, the random erasing consists of randomly selecting a rectangular region in an image and replacing its pixels with random values (Supplementary Figure 1b). This procedure generates new training images with various levels of occlusion, which, when used for training, reduces the risk of over-fitting and also makes the model less sensitive to occlusions (i.e., missing portions). Then, the flipping consists of mirroring the images both horizontally and vertically, along the vertical and horizontal axes, respectively (Supplementary Figure 1c). The third augmentation method is brightness adjust (Supplementary Figure 1d). To note, the coloring of a picture can be set using three parameters: hue (H), saturation (S) and value (V), with the latter mainly governing the brightness. By using the Auto Gamma Correction method, a non-linear operation $S=T(R)=R^{\gamma}$ (where S and R are the values of brightness in output and original image, respectively, that are mapped to [0 1]) is to automatically lighten or darken the image. And the random zoom is zooming into an image at a random location within the image (Supplementary Figure 1e). The random contrast adjusts the contrast of the images by a random factor (Supplementary Figure 1f). And the last is random saturation, which can adjust the saturation of images by a random factor (Supplementary Figure 1g). We compared the Swin Transformer models that with and without augmentation methods of random zoom, random contrast and

random saturation. We found that the model performance with these augmentation methods was improved (Supplementary Figure 2a) and the accuracy increased from 0.89 to 0.92. And the loss was also lower than the method without augmentation methods (0.004 for the methods with augmentation methods and 0.01 for the method without augmentation methods; Supplementary Figure 2b). Simultaneously, the abilities to classify between arsenopyrite and stibnite and between chalcopyrite and pyrite were improved (Supplementary Figure 3). In addition, we found that the precision and F1-score of the arsenopyrite and chalcopyrite were improved using the methods of random zoom, random contrast and random saturation (Supplementary Table 1). And the recall of pyrite and stibnite was also improved (0.81 and 0.84, respectively; (Supplementary Table 1). We can infer that the techniques of random zoom, random contrast and random saturation can improve the model ability to classify minerals based on color.

The comparison between adam gradient descent algorithm and stochastic gradient descent algorithm

There are two optimization algorithms that are always used in deep learning tasks named adam gradient descent algorithm and stochastic gradient descent algorithm, respectively. The stochastic gradient descent maintains a single learning rate for all weight updates and the learning rate does not change during training (Amari 1993). The adam gradient descent stands for Adaptive Moment Estimation, is an adaptive learning rate algorithm designed to improve training speeds in deep neural networks and reach convergence quickly (Kingma and Ba 2017). It customizes each parameter's learning rate based on its gradient history, and this adjustment helps the neural network learn efficiently as a whole. And we can find that on the sixteenth epoch, the accuracy of the model

with adam gradient descent achieved the highest of 0.92 (Supplementary Figure 4a). And on the twentieth epoch, the accuracy of the model with stochastic gradient descent achieved at 0.61 (Supplementary Figure 4a). As for the training loss, the model with adam gradient descent achieved the lowest value of 0.01 on the twentieth epoch, while the model with stochastic gradient descent achieved the lowest value of 1.35 on the twentieth epoch (Supplementary Figure 4b). The adam gradient descent algorithm highlighted superior performances than the stochastic gradient descent algorithm. In the manuscript, we reported the adam gradient descent algorithm.

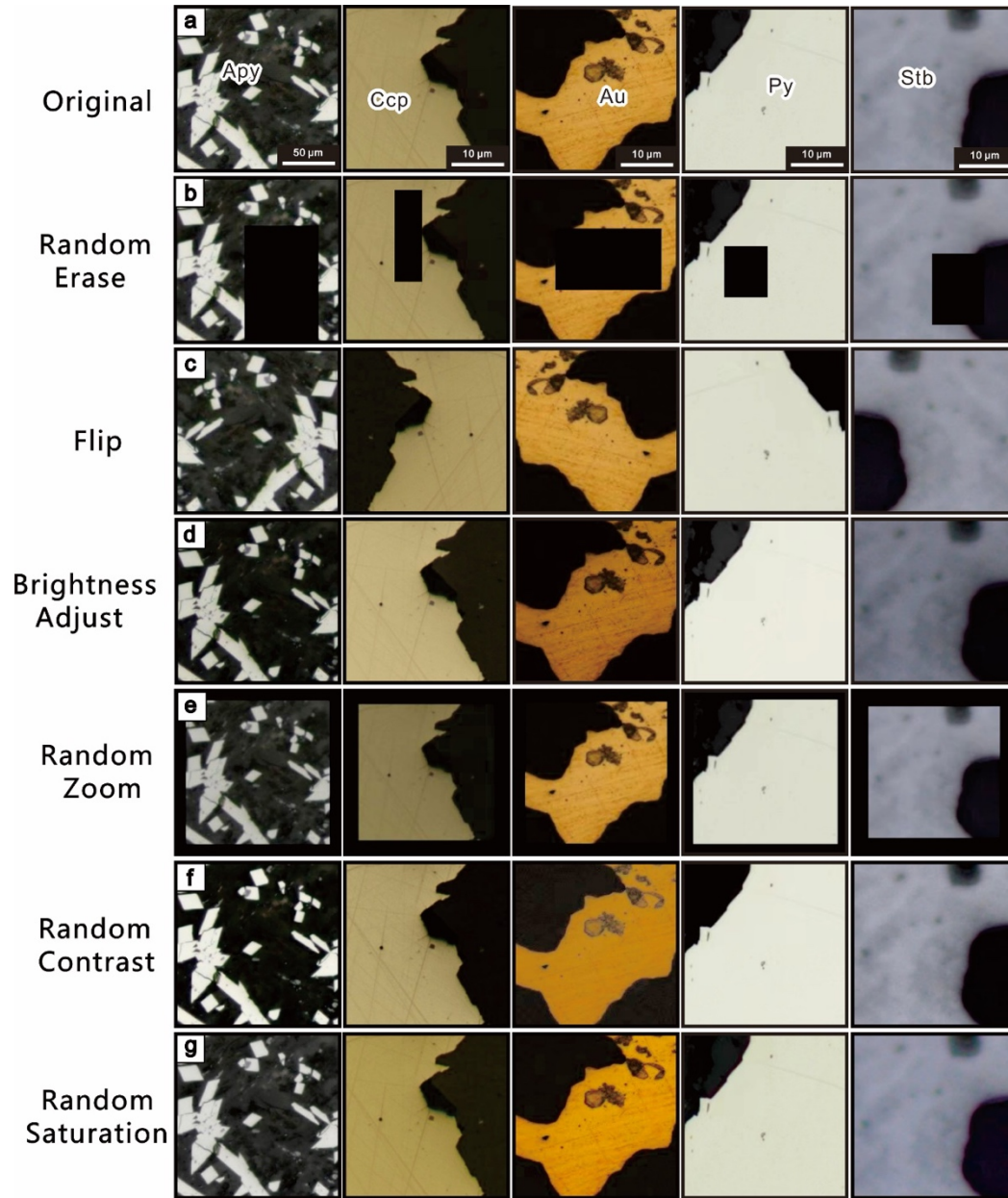
The comparison among transfer learning, freeze model and transfer learning plus freeze model

The “transfer learning” and “freeze model” are two common methods to improve model’s performance. The transfer learning is an ability of a system to recognize and apply knowledge and skills learned in previous tasks to novel tasks (Torrey and Shavlik 2010; Z. Liu et al. 2021). It’s currently very popular in deep learning because it can train deep neural networks with comparatively little data. This is very useful in the data science field since most real-world problems typically do not have millions of labeled data points to train such complex models. In addition, the freeze model is to keep the classification head unfrozen and trains the model with a high learning rate(Kumar et al. 2019; Y. Liu et al. 2021). And then to unfreeze the top convolution layers and fine-tune using a lower learning rate. We tried transfer learning, freeze model, and transfer learning plus freeze model, respectively to train the classification model. We found that the performance of Swin Transformer using transfer learning is better with an accuracy of 0.92 (Supplementary Table 2 and Figure 6), followed by using the method of freeze model with an

accuracy of 0.90, and with transfer learning and freeze model with an accuracy of 0.86 (Supplementary Figure 5 and Supplementary Figure 6). As for the ResNet50 and MobileNetv2, the model with transfer learning performs best, the second is the model with transfer learning and freeze model, and the last is the model with freeze model (Supplementary Figure 5 and Supplementary Figure 6). In the manuscript, we chose the method of transfer learning.

References

- Amari, S. (1993) Backpropagation and stochastic gradient descent method. *Neurocomputing*, 5, 185–196, [https://doi.org/10.1016/0925-2312\(93\)90006-O](https://doi.org/10.1016/0925-2312(93)90006-O).
- Feng, S.Y., Gangal, V., Wei, J., Chandar, S., Vosoughi, S., Mitamura, T., and Hovy, E. (2021, May 7) A Survey of Data Augmentation Approaches for NLP. *arXiv.org*, <https://arxiv.org/abs/2105.03075v5>.
- Kingma, D.P., and Ba, J. (2017, January 29) Adam: A Method for Stochastic Optimization. *arXiv*, <https://arxiv.org/abs/1412.6980v9>.
- Kumar, A., Balasubramanian, A., Venkataraman, S., and Akella, A. (2019) Accelerating Deep Learning Inference via Freezing. Presented at the 11th USENIX Workshop on Hot Topics in Cloud Computing (HotCloud 19).
- Liu, Y., Agarwal, S., and Venkataraman, S. (2021, April 3) AutoFreeze: Automatically Freezing Model Blocks to Accelerate Fine-tuning. *arXiv*, <http://arxiv.org/abs/2102.01386>.
- Liu, Z., Jin, L., Chen, J., Fang, Q., Ablameyko, S., Yin, Z., and Xu, Y. (2021) A survey on applications of deep learning in microscopy image analysis. *Computers in Biology and Medicine*, 134, 104523, <https://doi.org/10.1016/j.compbiomed.2021.104523>.
- Shorten, C., and Khoshgoftaar, T.M. (2019) A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6, 60, <https://doi.org/10.1186/s40537-019-0197-0>.
- Torrey, L., and Shavlik, J. (2010) Transfer Learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques* pp. 242–264. IGI Global.



Supplementary Figure 1. Data augmentation of five types of metal minerals on the training set.

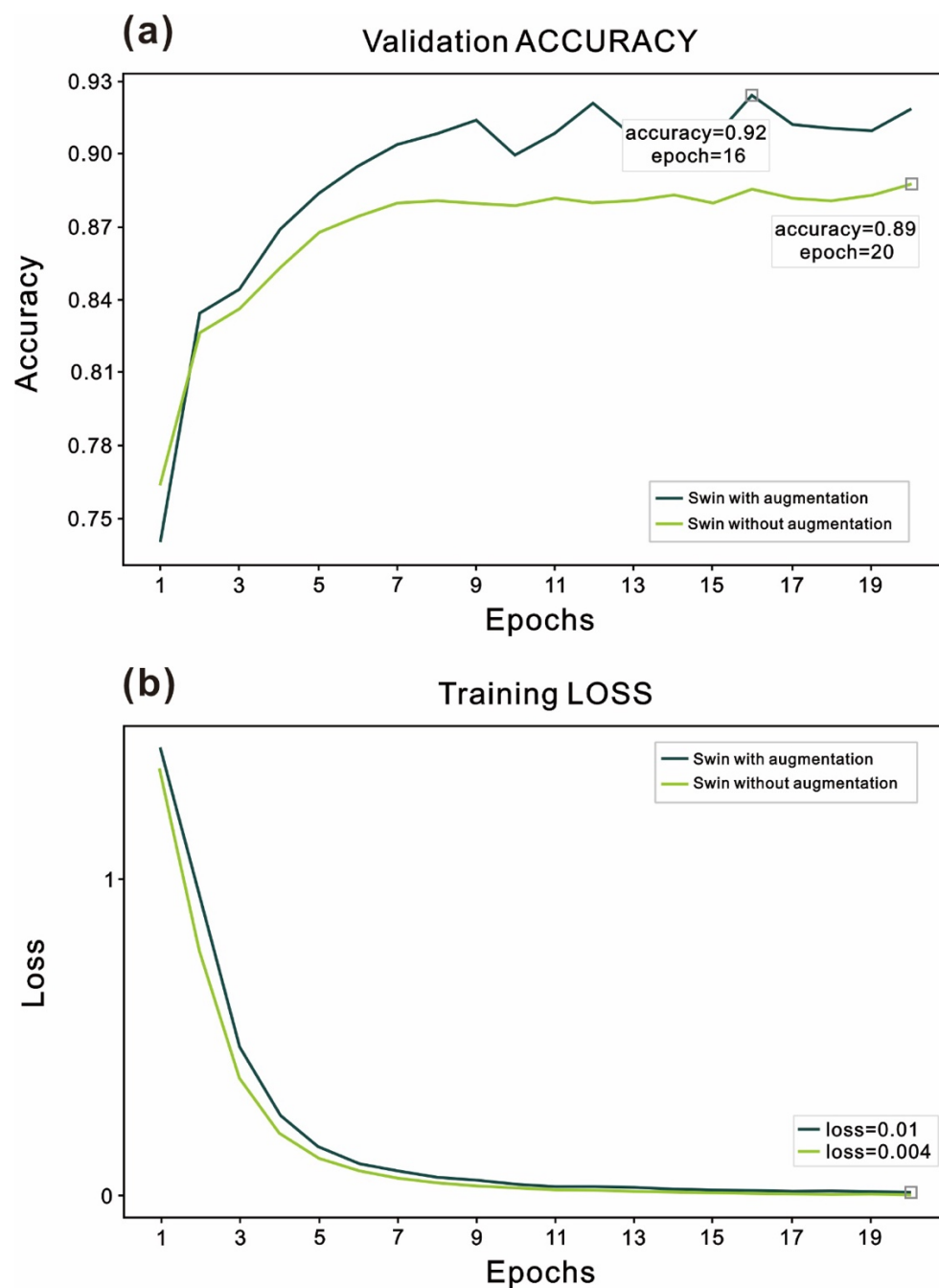
(a) The original image of five metal minerals. (b) The data augmentation method of Random Erase.

(c) The data augmentation method of Flip. (d) The data augmentation method of Brightness Adjust.

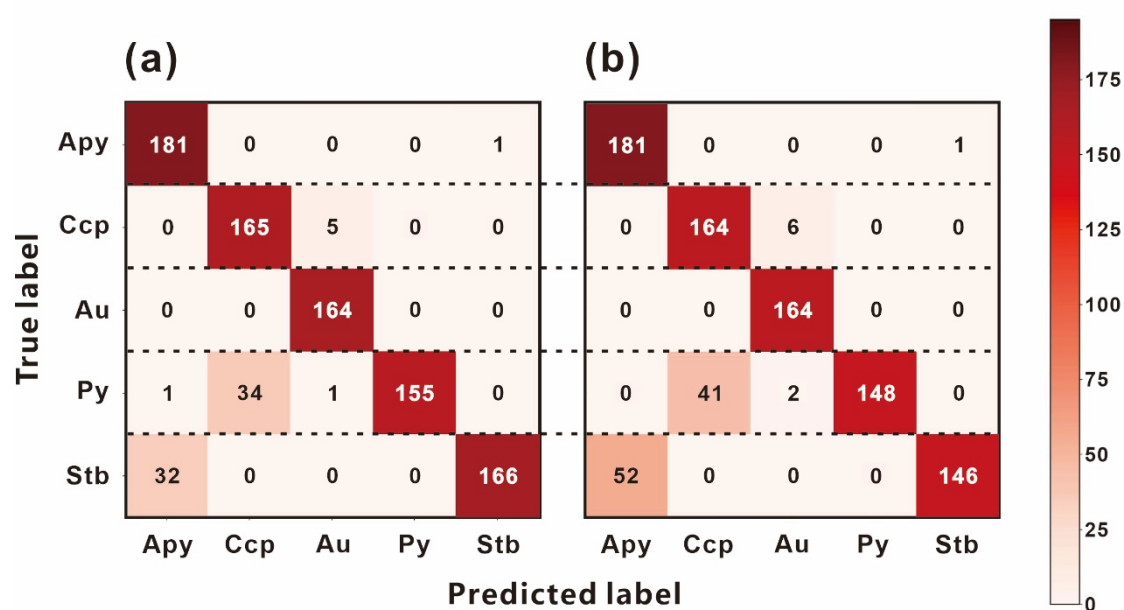
(e) The data augmentation method of Random Zoom. (f) The data augmentation method of

Random Contrast. (g) The data augmentation method of Random Saturation. Apy: arsenopyrite;

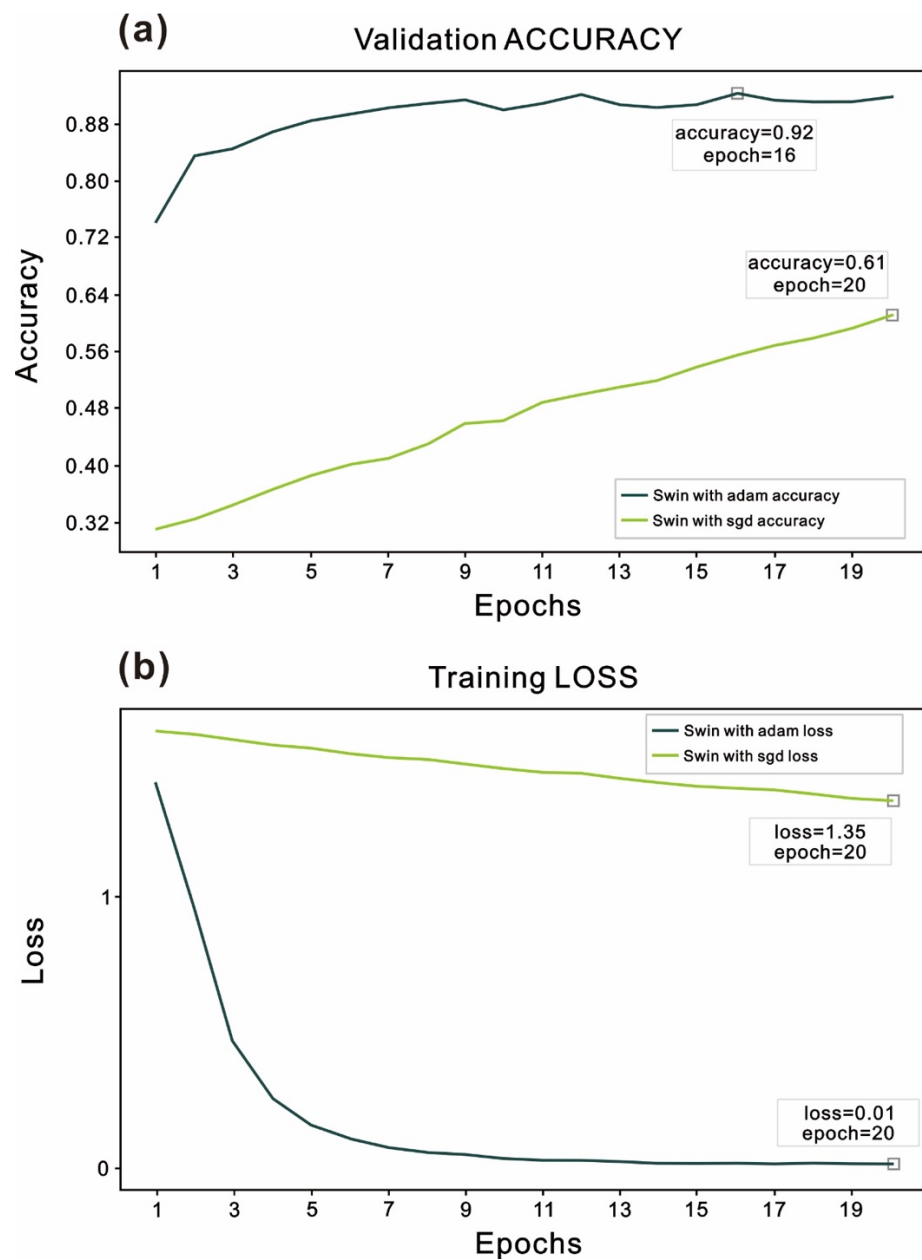
Ccp: chalcopyrite; Au: gold; Py: pyrite; Stb: stibnite.



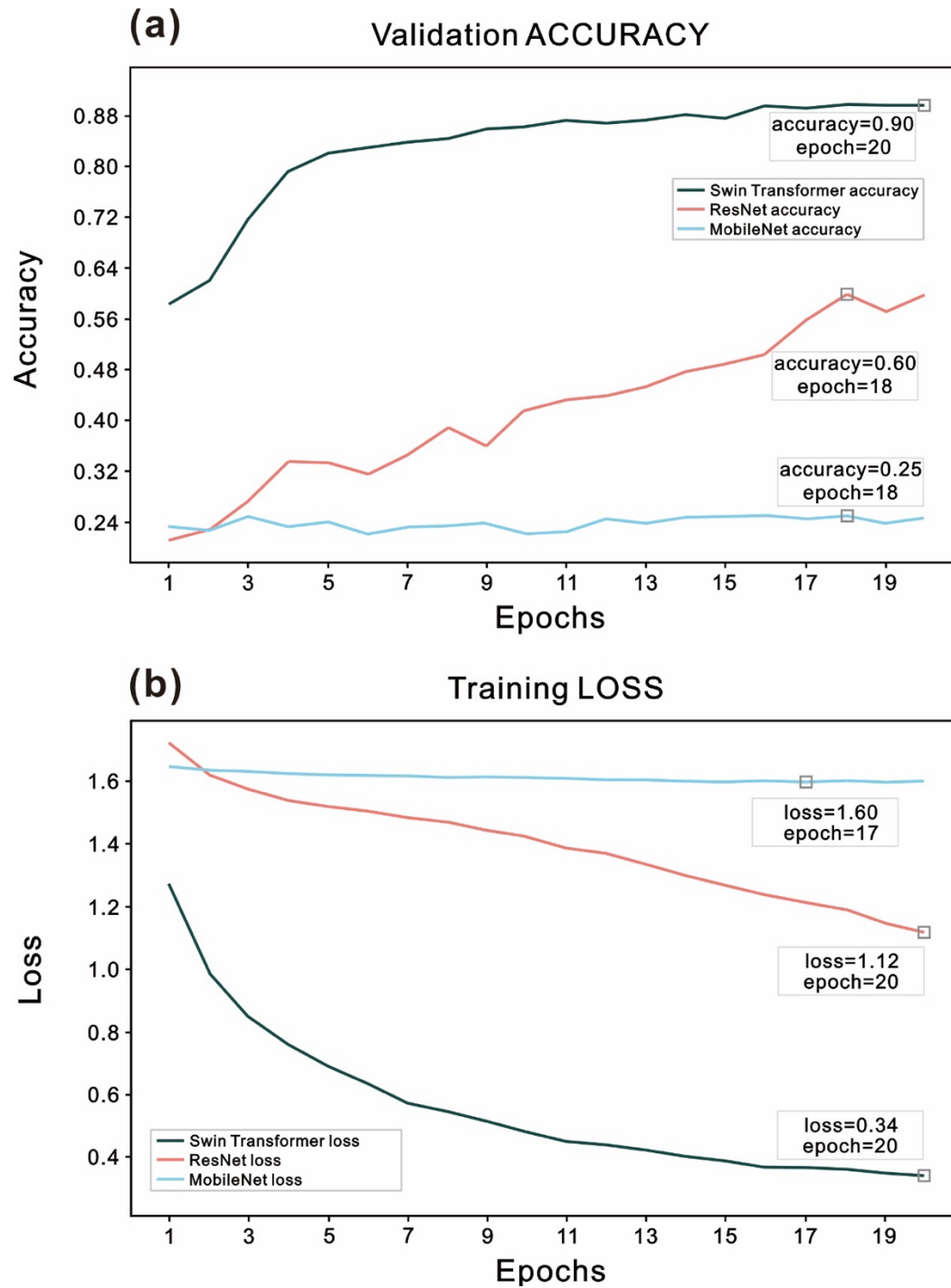
Supplementary Figure 2. Changes of (a) validation accuracy and (b) training loss of Swin Transformer with/without data augmentation methods using the method of transfer learning. The data augmentation methods include random contrast, random saturation and random zoom. The lines reflect the changes of different methods' performance within 20 epochs (dark green: Swin Transformer with data augmentation; light green: Swin Transformer without data augmentation).



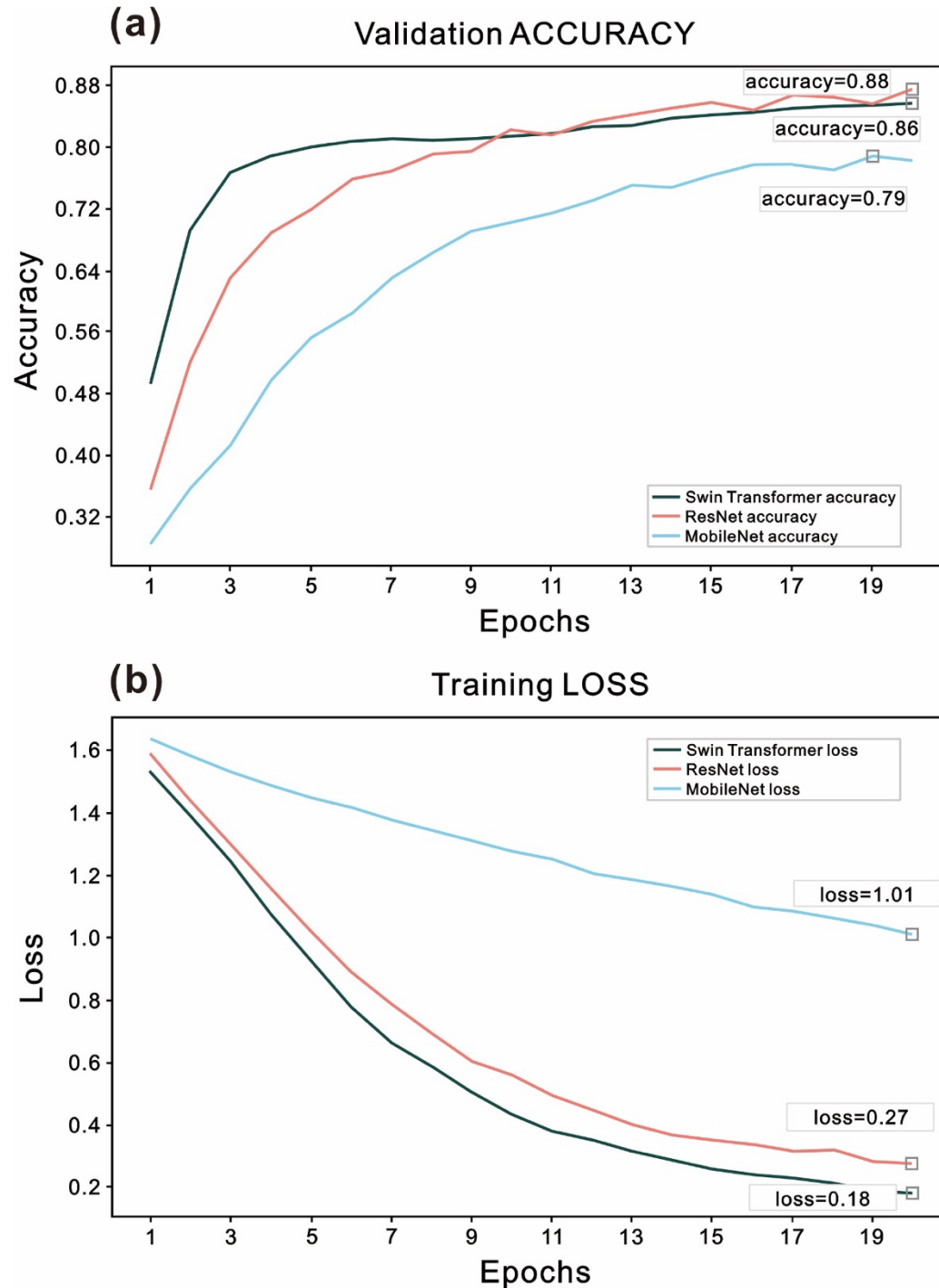
Supplementary Figure 3. Confusion matrix of the test set used to evaluate the method of data augmentation methods. (a) Swin Transformer with data augmentation methods; (b) Swin Transformer without data augmentation methods. The data augmentation methods include random contrast, random saturation and random zoom. Indicated values are the number of images. The horizontal axis is the predicted label, while the vertical axis is the true label. Apy: arsenopyrite; Ccp: chalcopyrite; Au: gold; Py: pyrite; Stb: stibnite.



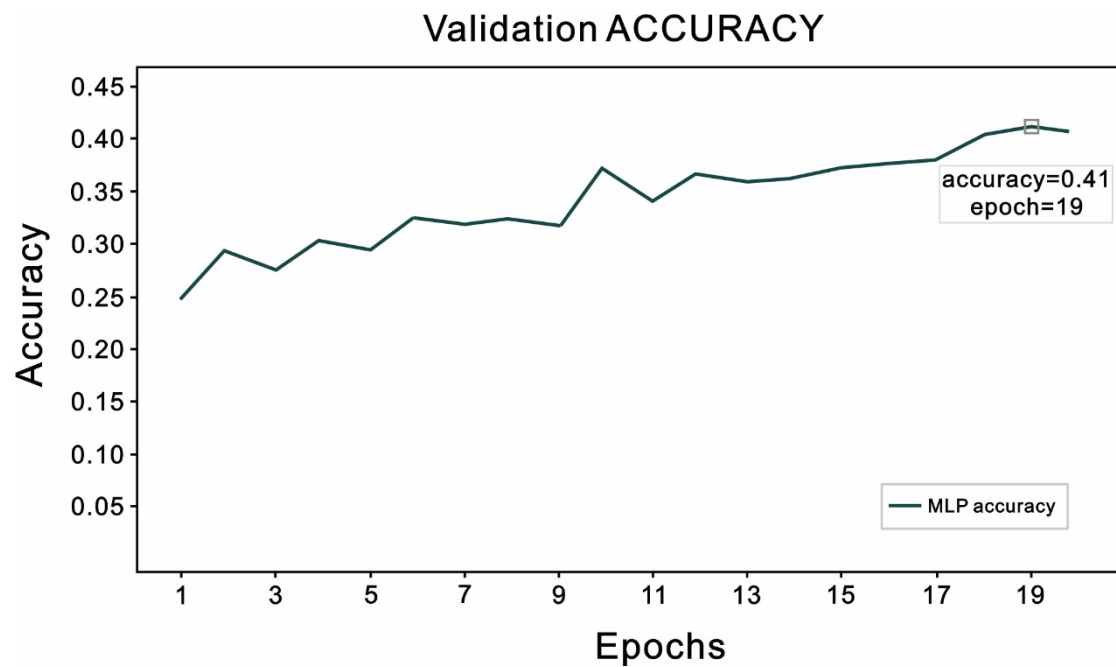
Supplementary Figure 4. Changes of (a) validation accuracy and (b) training loss of Swin Transformer with two common gradient descent algorithms using the method of transfer learning. The lines reflect the changes of different algorithms' performance within 20 epochs (dark green: adam gradient descent algorithm. The first-order momentum parameter is set to 0.9 and the second-order momentum parameter is set to 0.999; light green: stochastic gradient descent algorithm. The momentum parameter is set to 0.9).



Supplementary Figure 5. Changes of (a) validation accuracy and (b) training loss of three algorithms using the method of freeze model. The lines reflect the changes of different algorithms' performance within 20 epochs (green: Swin Transformer algorithm; red: ResNet50 algorithm; blue: MobileNetv2 algorithm).



Supplementary Figure 6. Changes of (a) validation accuracy and (b) training loss of three algorithms using the method of transfer learning and freeze model. The lines reflect the changes of different algorithms' performance within 20 epochs (green: Swin Transformer algorithm; red: ResNet50 algorithm; blue: MobileNetv2 algorithm).



Supplementary Figure 7. Changes of validation accuracy of the Multilayer Perceptron algorithm using the method of transfer learning. The lines reflect the model performance within 20 epochs. The Multilayer Perceptron algorithm is set as the baseline to evaluate whether the dataset is class imbalanced.

Table

Supplementary Table 1 Mineral Classification Performances of Swin Transformer with Data

Augmentation methods and without Data Augmentation methods on the Test Set

Method	Metric	Arsenopyrite	Chalcopyrite	Gold	Pyrite	Stibnite	Metric Value
augmentation	Acc						0.92
	Pre	0.85	0.83	0.96	1.00	0.99	0.93
	Rec	0.99	0.97	1.00	0.81	0.84	0.92
	F1	0.91	0.89	0.98	0.90	0.91	0.92
No augmentation	Acc						0.89
	Pre	0.78	0.80	0.95	1.00	0.99	0.90
	Rec	0.99	0.96	1.00	0.77	0.74	0.89
	F1	0.87	0.87	0.98	0.87	0.85	0.89

Note: The data augmentation methods include random contrast, random saturation and random zoom. Acc: abbreviation for model

evaluation indicator accuracy; Pre: abbreviation for model evaluation indicator precision; Rec: abbreviation for model evaluation

indicator recall; F1: abbreviation for model evaluation indicator F1-score.

Supplementary Table 2 Mineral Classification Performances of Swin Transformer with Transfer

Learning, Freeze Model, and Transfer Learning Plus Freeze Model on the Test Set

Method	Metric	Arsenopyrite	Chalcopyrite	Gold	Pyrite	Stibnite	Metric Value
Transfer learning	Acc						0.92
	Pre	0.85	0.83	0.96	1.00	0.99	0.93
	Rec	0.99	0.97	1.00	0.81	0.84	0.92
	F1	0.91	0.89	0.98	0.90	0.91	0.92
Freeze model	Acc						0.86
	Pre	0.73	0.79	0.92	0.99	0.96	0.88
	Rec	0.99	0.92	0.99	0.73	0.69	0.87
	F1	0.84	0.85	0.96	0.84	0.80	0.86
Transfer learning + freeze model	Acc						0.90
	Pre	0.81	0.98	0.90	0.85	0.99	0.91
	Rec	0.87	0.86	0.97	0.96	0.84	0.90
	F1	0.84	0.92	0.93	0.90	0.91	0.90

Note: Acc: abbreviation for model evaluation indicator accuracy; Pre: abbreviation for model evaluation indicator precision; Rec:

abbreviation for model evaluation indicator recall; F1: abbreviation for model evaluation indicator F1-score.