STA326 Project Interim Report

I Dataset

1. Data Source and Information

Our data, consisting of 132213 labeled rock images, is acquired directly from geology researchers. The dataset is made up of images of 3 main types of rocks: metamorphic rock, sedimentary rock, and igneous rock. The images are split into 3 subsets: training set with 112213 images, validation set and testing set with 15000 images respectively.

2. Exploratory Analysis

- Category distributions:

Category	Metamorphic rock	Sedimentary rock	Igneous rock
Training set	24985	30896	46332
Validation set	5000	5000	5000

- Means for each image channel:

Channel	1	2	3
Training set	0.46020824	0.4554496	0.45052096
Validation set	0.46107383	0.45589945	0.45033285

- Standard deviation values for each image channel:

Channel	1	2	3
Training set	0.28402117	0.28318824	0.28876383
Validation set	0.28876383	0.28181302	0.28723977

3. Data Preprocessing

Images in both the training and validation sets are preprocessed. We randomly flipped, rotated, and implemented color jitter on training images, and all of the images are resized to 224×224 and normalized with the mean/standard deviation values from each split:

```
transform_train = transforms.Compose(
    [transforms.Resize((224, 224)),
    transforms.RandomHorizontalFlip(p=0.2),
    transforms.ColorJitter(brightness=0.2, contrast=0.2, saturation=0.2, hue=0.1),
    transforms.RandomRotation(15),
    transforms.ToTensor(),
    transforms.Normalize((0.46020824, 0.4554496,  0.45052096),
                         (0.28402117, 0.28318824, 0.28876383))])

```

```
transform_test = transforms.Compose(
    [transforms.Resize((224, 224)),
    transforms.ToTensor(),
    transforms.Normalize((0.46107383, 0.45589945, 0.45033285),
                         (0.28263338, 0.28181302, 0.28723977))])
```

II Preliminary Results

1. Model Selection

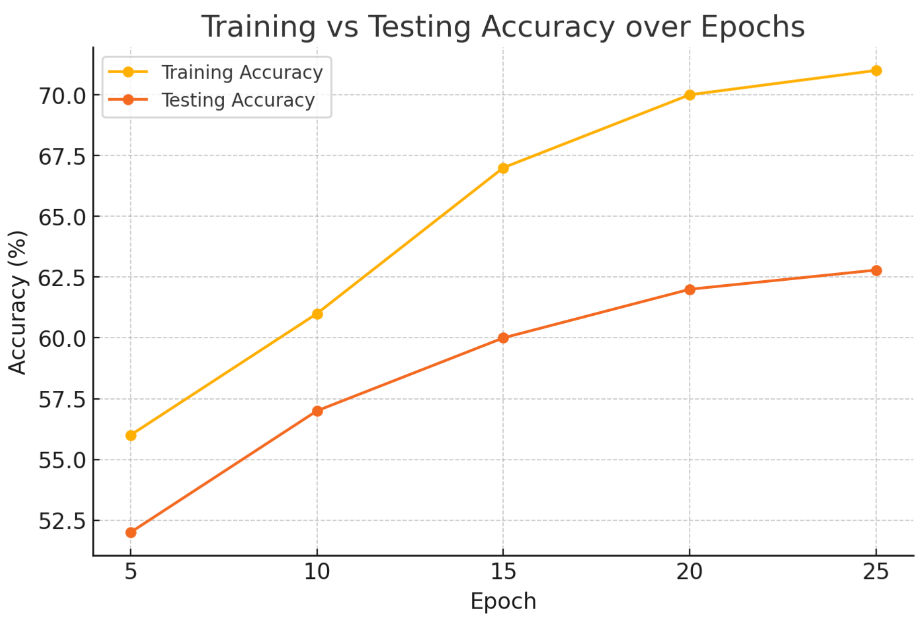
We chose **ViT-B/16** to leverage its global self-attention for capturing long-range dependencies in stone surface textures, aiming to improve the discrimination of subtle class differences that standard CNNs may miss. Its proven scalability on large image datasets made it a strong candidate for our ~102 K-image corpus.

2. Experimental Set-ups

Batch Size	Learning rate	Learning rate scheduler	Optimizer
64	1×10^{-5}	StepLR, step = 5	Adam

3. Result Analysis

- **Best accuracy**: **71.55%** for training set, **62.79%** for validation set.
- **Loss and accuracy plots**:



Apparently, our model's training loss has reached its convergence at around epoch 20 with such a setting, and the model's fitting capability still needs improving.

III Future Plans

To tackle the problem that we have encountered, we plan to make these attempts in the future (in priority order):

- Use the CosineAnnealing Warmup scheduler, which is suitable for ViT.
- Add `autocast` and `GradScaler` to help accelerate training.

- Add `RandomResizedCrop`, `RandomRotation` to transform to help model be robust.
- Resize the training pictures to remove unnecessary parts (use OpenCV to detect the main part of the picture).
- Try different hyperparameters (number of training epoch, learning rate, batch size, seed, etc.).

IV Summary

Our ViT-B/16 reached 71.55% training and 62.79% accuracy at epoch 20. To boost generalization by ~10–20%, we will adopt a CosineAnnealing Warmup scheduler for cyclic learning rates, enable PyTorch AMP mixed precision with `torch.autocast` and `GradScaler`, integrate RandomResizedCrop and RandomRotation augmentations, perform ROI cropping via OpenCV, plus hyperparameter optimization.

Reference

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *An image is worth 16×16 words: Transformers for image recognition at scale*. In *International Conference on Learning Representations*. https://openreview.net/forum?id=YicbFdNTTy

Wightman, R. (2019). *PyTorch Image Models (timm)* [Computer software]. GitHub repository. https://github.com/rwightman/pytorch-image-models (doi: 10.5281/zenodo.4414861)