

---

# Towards Automated Petrography

---

**Isai Daniel Chacón\***  
Universidad de los Andes, Colombia

**Paola Ruiz Puentes\***  
Universidad de los Andes, Colombia

**Jillian Pearse**  
California State University, Long Beach

**Pablo Arbeláez**  
Universidad de los Andes, Colombia

## Abstract

Petrography is a branch of geology that analyzes the mineralogical composition of rocks from microscopical thin section samples. It is essential for understanding rock properties across geology, archaeology, engineering, mineral exploration, and the oil industry. However, petrography is a labor-intensive task requiring experts to conduct detailed visual examinations of thin section samples through optical polarization microscopes, thus hampering scalability and highlighting the need for automated techniques. To address this challenge, we introduce the Large-scale Imaging and Thin section Optical-polarization Set (LITHOS), the largest and most diverse publicly available experimental framework for automated petrography. LITHOS includes 211,604 high-resolution RGB patches of polarized light and 105,802 expert-annotated grains across 25 mineral categories. Each annotation consists of the mineral class, spatial coordinates, and expert-defined major and minor axes represented as intersecting vector paths, capturing grain geometry and orientation. We evaluate multiple deep learning techniques for mineral classification in LITHOS and propose a dual-encoder transformer architecture that integrates both polarization modalities as a strong baseline for future reference. Our method consistently outperforms single-polarization models, demonstrating the value of polarization synergy in mineral classification. We have made the LITHOS Benchmark publicly available, comprising our dataset, code, and pretrained models, to foster reproducibility and further research in automated petrographic analysis.

🔗 **Benchmark and code:** <https://github.com/BCV-Uniandes/LITHOS>

📄 **Dataset:** LITHOS Dataset

## 1 Introduction

Petrography is a branch of geology that describes and analyzes rocks and soils, focusing on their mineralogical composition, texture, and structure. It enables the identification of minerals and other constituents within a sample and the characterization of grain size, shape, and porosity distribution [4]. These attributes are critical in various fields, particularly in economic geology, where petrographic analysis informs resource exploration and extraction strategies [20, 21].

Understanding a rock’s petrographic properties is fundamental across multiple disciplines. In geology, petrography aids in reconstructing Earth’s geological history, revealing past environmental conditions and tectonic processes [6, 11]. In archaeology, it helps determine the provenance of artifacts and ancient trade routes by analyzing ceramic and lithic materials [9]. In engineering, petrographic examination is crucial for assessing the durability and quality of construction materials, such as concrete and aggregates [24, 28]. In mineral exploration, it provides essential insights into the

---

\*These authors contributed equally to this work.

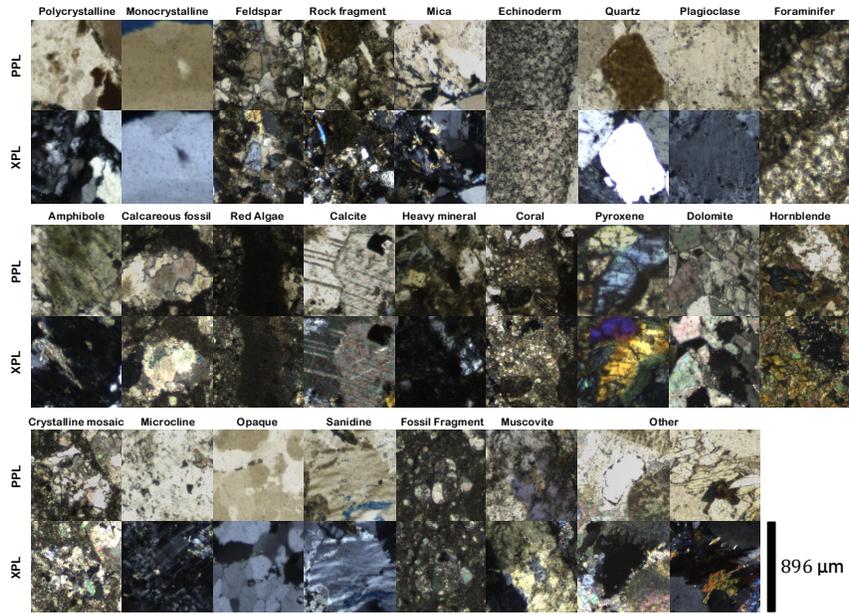


Figure 1: Example of  $256 \times 256$  high-resolution image patches extracted from the LITHOS Dataset, illustrating the 25 mineral classes under the two polarization conditions: plane-polarized light (PPL) at  $0^\circ$  and cross-polarized light (XPL) at  $0^\circ$ . Each patch represents an area of  $896\mu m^2$ . These paired images highlight the variation in color, texture, and birefringence patterns, which are critical for mineral identification in thin section petrography.

characterization of the ore, directly influencing the beneficiation and metallurgical processes [3, 11]. Additionally, in the oil industry, petrography plays a key role in evaluating a rock's capacity to store and transmit hydrocarbons, guiding reservoir characterization and exploration efforts [11, 23].

To assess a rock's mineralogical and textural characteristics, petrographers examine thin sections, which are thin flat slices of material prepared for microscopic analysis. Specifically, an approach known as the point-counting method is used. This technique superimposes a grid of equidistant points on the thin section. The petrographer examines each point, identifying and recording the mineral or feature present. This method allows for the determination of the modal composition of the rock, including the percentages of various minerals, porosity, and other constituents [16].

Despite its significance, petrographic analysis is a labor-intensive and expertise-driven task. With the growth of high-resolution imaging and modern computational tools, there is an opportunity to automate parts of this process, helping reduce subjectivity and improving efficiency. However, building reliable automated methods requires large, well-annotated datasets, which are currently lacking. Most existing databases are limited in size and variety, making it challenging to train robust models for this type of analysis [8, 22, 30, 32]. Therefore, there is also a lack of publicly accessible methods that researchers can readily adopt and apply in this area.

This paper introduces the Large-scale Imaging and Thin section Optical-polarization Set (LITHOS), a novel experimental framework for automated petrography. LITHOS is the largest and most diverse publicly available resource for this task, surpassing existing frameworks by two orders of magnitude in the number of images and by one in the number of annotated mineral grains. Each mineral grain refers to an individual crystal or crystal fragment that can be visually distinguished in a thin section under a petrographic microscope based on its optical properties. The dataset encompasses 25 mineral categories, totaling 105,802 annotated grains and 211,604 high-resolution image patches. Each section was digitized using an automated petrographic microscope, capturing images under two polarization conditions. Expert annotations include mineral labels, grain size measurements, major and minor axes represented as intersecting vector paths, and spatial coordinates. Figure 1 presents representative samples from the 25 categories under both polarization conditions. These examples illustrate the dataset's diversity in color, texture, and birefringence patterns. Additionally, they underscore the inherent complexity of the task, where several mineral categories exhibit significant

visual similarities. This challenge reflects real-world petrographic analysis scenarios and makes LITHOS a robust benchmark for developing and evaluating automated mineral classification methods. This resource is intended to support the advancement of automated petrographic analysis, fostering further research in the field.

We train and evaluate a suite of standard deep learning models commonly used for image classification tasks on a single-polarization subset of LITHOS. Additionally, we introduce the LITHOS Baseline, a model specifically designed to leverage the dual-polarization imaging captured in our benchmark. The LITHOS Baseline is a transformer-based architecture that processes paired views of thin section patches under plane-polarized (PPL) and cross-polarized (XPL) light. This dual-encoder model exploits the complementary optical characteristics revealed by the two polarization modes, enabling more accurate mineral identification.

Our main contributions can be summarized as follows:

- (1) We collect the LITHOS Dataset, the largest publicly available image collection for automated petrography. We digitized thin sections under two polarization conditions, PPL and XPL, yielding 211,604 paired high-resolution image patches and 105,802 expert annotations across 25 mineral classes, representing the most comprehensive and diverse dataset of its kind. Expert annotations include mineral categories, grain coordinates, and geometric paths representing major and minor axes.
- (2) We introduce the LITHOS Benchmark, a robust framework designed to evaluate automated mineral identification capabilities. The benchmark includes both binary and multi-class classification tasks. Additionally, we evaluate a set of deep learning single-polarization methods on this Benchmark.
- (3) We propose the LITHOS Baseline, a dual-encoder transformer architecture specifically designed to leverage paired PPL and XPL views of mineral grains. Our baseline consistently improves mineral classification metrics over single-modality approaches, demonstrating the advantage of incorporating complementary polarization information in automated petrography.

The LITHOS Dataset and pretrained models will be publicly released under a Creative Commons Attribution Non Commercial ShareAlike 4.0 International (CC BY-NC-SA 4.0) license upon acceptance, to promote reproducibility, transparency, and enable further research.

## 2 Related Work

### 2.1 Traditional Petrography

Petrographic analysis involves arduous and time-consuming procedures by trained personnel using specialized equipment [16, 31]. The process begins with the preparation of thin sections that are essential for microscopic examination. To obtain them, a rock sample is selected and cut into a small block using a diamond saw. This block is then sliced to obtain a thin sliver, typically around 1 mm thick. The fragment is ground flat and polished using progressively finer abrasives to achieve a smooth, optically flat surface. Subsequently, the polished sample is fixated to a glass slide with an adhesive, such as epoxy resin or Canada balsam. The mounted sample is ground to a standard thickness of approximately 30  $\mu\text{m}$ . This process can take around 12 hours per thin section. Then, these thin sections are analyzed using a petrographic microscope equipped with polarizing filters. These filters produce polarized light that interacts with the properties of minerals, revealing critical optical characteristics.

The petrographer examines each thin section under both PPL and XPL [31]. By observing light intensity and color changes, the petrographer can identify minerals and assess their properties. For instance, a critical property observed during petrographic analysis is the extinction angle. This angle is the measure between a prominent crystallographic direction (such as cleavage planes or elongation) and the position at which the mineral goes dark (extinguishes) under XPL. Accurate measurement of the extinction angle provides insight into the mineral's internal structure and symmetry, which is essential for precise identification [1, 33]. The expert systematically repeats this process for numerous grains within each section using the point-counting method, which involves superimposing a grid of equidistant points onto the thin section. The petrographer examines each point, identifying and recording the mineral or feature present. The time required for analysis depends on the density of the grid and the petrographer's expertise. It is a common practice to analyze between 300 and 500 points

Table 1: Comparison between the LITHOS Dataset and existing petrographic datasets used for mineral identification. We report the polarizations available, the number of total images, the number of annotated grains, the number of mineral classes, and availability. A check mark (✓) indicates presence, a dash (–) indicates absence, and (NR) indicates information Not Reported.

Dataset	Polarizations	No. of Images	Annotated Grains	No. of Classes	Publicly available
Hoque et al. [18]	-	993	NR	4	-
Song et al. [25]	PPL, XPL	1,249	875	10	✓
Hongjue Li et al. [22]	<b>PPL, six XPLs</b>	140	NR	NR	-
Keshk et al. [12]	PPL, XPL	1,388	NR	16	-
Ma et al. [35]	PPL, XPL	1,790	19,463	10	Upon request
Şener et al. [2]	XPL	600	NR	3	Upon request
LITHOS Dataset (Ours)	PPL, XPL	<b>211,604</b>	<b>105,802</b>	<b>25</b>	✓

per thin section to achieve a statistically valid representation. Reducing the number of points can accelerate the process, but may compromise the statistical robustness and accuracy of the results [16].

## 2.2 Existing Machine Learning Benchmarks

Recent advances in high-resolution imaging and computational modeling have driven interest in automating petrographic analysis. Despite these developments, progress remains constrained by the lack of large-scale, well-annotated datasets. Existing resources often fall short in capturing the mineralogical complexity required to support modern machine learning methods [8, 22, 30, 32]. In particular, current datasets are often limited in the number of annotated samples, the diversity of mineral classes, the range of imaging modalities, or their public accessibility, all of which difficult training and evaluation of generalizable models.

Several efforts have attempted to fill this gap, but important limitations persist. Table 1 shows existing petrographic datasets for mineral identification. Datasets introduced by Hoque et al. [18], Hongjue Li et al. [22], and Keshk et al. [12] are not publicly available. Song et al. [25], the only publicly available dataset, lacks either PPL or XPL images for some sections, limiting its utility for comprehensive model development. Şener et al. [2] includes a very narrow set of mineral categories and images, limiting effective model training, the range of trainable architectures, and the ability to generalize to new mineral classes. Finally, although the proposal by Ma et al. [35] offers a valuable contribution to mineral semantic segmentation, the access to their full dataset requires direct request, limiting immediate usability for broader research and benchmarking. To overcome these drawbacks, we present the LITHOS Dataset: a publicly accessible resource featuring paired polarization images for all annotated grains, the broadest mineral class coverage to date, and a substantial volume of images and annotations suitable for training with state-of-the-art computer vision techniques.

As summarized in Table 1, LITHOS comprises 211,604 high-resolution RGB patches, and includes 105,802 mineral grain annotations spanning 25 classes. Compared to prior datasets, LITHOS provides two orders of magnitude more high-resolution images, one order of magnitude more labeled grains and broader class diversity. LITHOS includes the geometric paths of the major and minor axes for each grain. These axes provide a form of weak supervision that supports instance-level learning tasks. This representation offers a meaningful compromise between coarse annotations and full segmentation, enabling models to leverage spatial and structural representations without requiring exhaustive manual labeling. These characteristics make LITHOS the most comprehensive publicly available dataset for supervised learning in automated petrography.

Beyond dataset limitations, recent efforts have explored the use of deep learning for petrographic image analysis, particularly for tasks such as mineral classification and grain segmentation. Early studies employed conventional architectures such as CNNs [17, 19], Faster RCNN [29], U-Net [12, 14, 17, 18], ResNet [5], Mask2Former [2], DeeplabV3+ [2] and Segformer [2] for segmentation and classification tasks, yet often lacked adaptation to the dual-polarization modality. Furthermore, the Multi-channel Attention Transformer proposed by [22] for rock segmentation, which takes into account multiple polarizations, lacks public code or pretrained models. Similarly, Trans-SedNet [35], a dual-modal Vision Transformer, cannot be fully reproduced using the publicly released

implementation. As a result, there remains a gap in the development of reproducible, polarization-aware deep learning models that are both open and extensible for broader applications in automated petrography. To address these limitations, we propose the LITHOS Baseline, an open source deep learning model adapted for dual-polarization processing and release the pretrained weights of our models so it can serve as a basis for multiple case studies in petrography.

### 3 The Large-scale Imaging and Thin section Optical-polarization Set

This section describes the construction of The Large-scale Imaging and Thin section Optical-polarization Set (LITHOS). First, we describe the thin sections collection, digitization and annotation process (Section 3.1). Then, we define our benchmark for mineral classification and evaluation framework (Section 3.2).

#### 3.1 LITHOS Dataset

We compiled and annotated a total of 580 thin sections from Colombian soils, predominantly composed of formations and wells of interest for hydrocarbons exploration.

**Thin Sections Digitization.** We digitized each thin section using an automated petrographic microscope, which captures high-resolution images under different polarization angles [10]. During imaging, the thin section is placed on the microscope stage, with a polarizer positioned below the sample and an analyzer above it. By rotating either the stage or the polarizing filters, the microscope can capture images in PPL (plane-polarized light) and XPL (cross-polarized light). Specifically, images were acquired in PPL and XPL at 40X magnification. Each pixel represents an area of approximately  $3.5\mu\text{m} \times 3.5\mu\text{m}$ . For our dataset, we obtained two images per thin section: one PPL at  $0^\circ$  and one XPL at  $0^\circ$ , as shown in Fig. 2. The image acquisition process involves an average of 143 acquisition points per thin section, with 20 images captured at each point, resulting in approximately 2,860 raw images per section. These images are processed using image alignment algorithms that automatically register and stitch them into coherent high-resolution mosaics ( $24,786 \times 16,259$  pixels) representing the entire thin sections. On average, the acquisition and initial processing requires 3 hours per sample. This pipeline, enables accurate reconstruction of large petrographic samples and supports downstream annotation on the OneGeo software [10]. The dataset reflects a significant technological and human effort, corresponding to more than 1,700 hours of high-resolution image capture and processing.

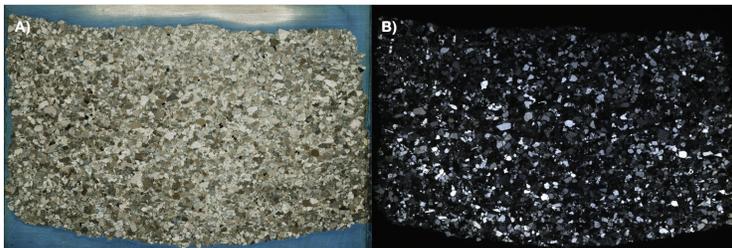


Figure 2: Example of a digitized thin section under polarized light. (A) Plane-polarized light (PPL) at  $0^\circ$ . (B) Cross-polarized light (XPL) at  $0^\circ$ .

**Annotation Process.** OneGeo [10] was used to annotate all thin sections. This software functions as a petrographic microscope, utilizing the previously captured images to enable multi-magnification analysis. It also allows users to switch between polarization angles and visualize parallel and cross nicols simultaneously. These capabilities ensure a precise mineral identification. OneGeo generates an annotation grid of 300 to 600 equidistant points over each sample, and each point is individually analyzed. The annotation process for each 600-point thin section took around 10 hours. It included assigning a mineral category per grain, recording the x and y coordinates, measuring their minor and major axes, and representing its geometric paths as intersecting vectors through an HTML element. Each annotation undergoes a review by a second annotator, which typically requires approximately half the time of the initial annotation process. Detailed explanation of the annotation files can be found in the supplementary material (Listing 1).

**Dataset.** In total, our database consists of 1,164 high-resolution RGB images, 105,802 annotated grains, and 25 mineral categories. This effort represents the equivalent of a full-time expert dedicating an entire year exclusively to annotating our database, highlighting the substantial value and scale of our contribution. The mineral categories exhibit a long-tail distribution dominated by quartz. The dataset is highly imbalanced, with 49,8% of the annotated points representing quartz, categorized into three main classes: Quartz, Monocrystalline, and Polycrystalline. From each high-resolution image, we extracted patches of size  $256 \times 256$  pixels. Each patch was centered at the intersection of the major and minor axes of the mineral. The axes paths were stored in an HTML element and reconstructed using the `svgpathools` library in Python. This results in a total of 211,604 high-resolution RGB patches (Table 1).

### 3.2 LITHOS Benchmark

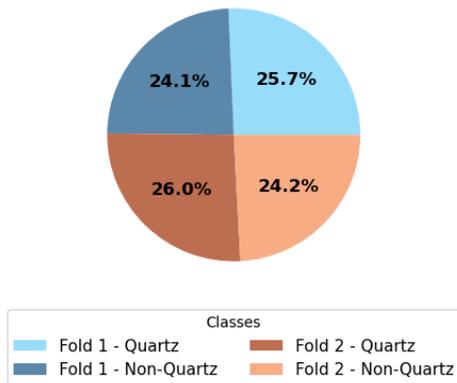


Figure 3: Distribution of annotated minerals across the two folds in the binary task.

The LITHOS Benchmark is designed to assess mineral classification performance in soil thin section images. Considering the long-tail distribution of our dataset dominated by quartz, we propose two classification tasks. In the binary task, the goal is to classify the central mineral grain in each  $256 \times 256$  image patch as either Quartz vs. Non-Quartz. The Quartz class aggregates several quartz varieties, including Monocrystalline, Polycrystalline, and generic Quartz annotations, while Non-Quartz aggregates the remaining classes. Fig 3 shows folds distribution for this setup.

In the multi-class task, the objective is to classify the central grain into one of 25 mineral categories. This ensures that the benchmark supports the development of more robust and generalizable AI models.

We implemented a 2-fold cross-validation setup, assigning each thin section exclusively to one of the folds to avoid data leakage. This strategy preserved the relative mineral distribution across folds. The distribution of mineral categories across folds is shown in Fig. 4. Model performance of both tasks is evaluated using standard classification metrics: accuracy, precision, recall, and F1-score.

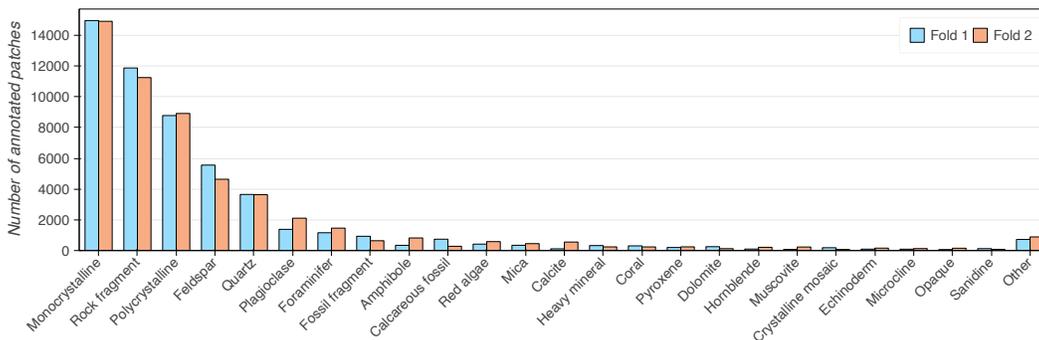


Figure 4: **Distribution of annotated minerals per class across the two folds.** The distribution is imbalanced, with a few dominant classes such as Monocrystalline, Rock Fragment, and Polycrystalline accounting for the majority of annotations. This long-tailed distribution reflects the natural occurrence of minerals in thin sections and poses a significant challenge for learning robust classification models, particularly for rare classes.

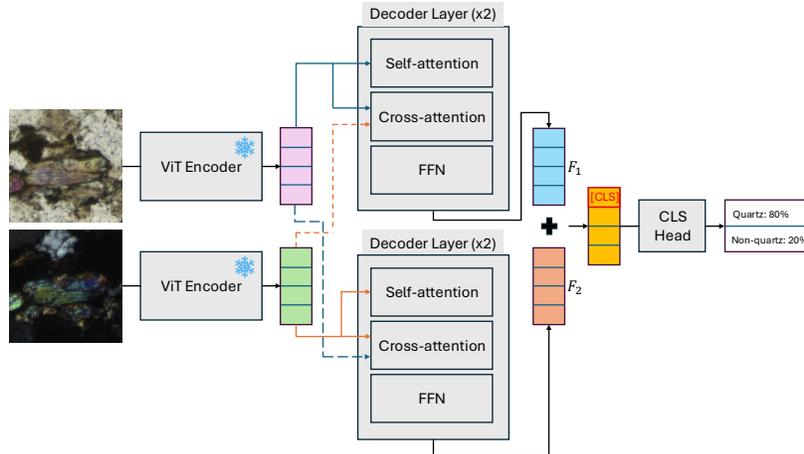


Figure 5: **Overview of the LITHOS Baseline.** Two frozen ViT encoders extract specific polarized representations from petrographic images at PPL and XPL. A dual-decoder module captures feature dependencies through self-attention and cross attention mechanisms. These features are then fused together via a learnable weighted sum. Lastly, the [CLS] token of the combined representation is passed through the classification head of the model. FFN stands for Feed-Forward Network.

## 4 LITHOS Baseline

For mineral classification in the LITHOS Benchmark, we propose the LITHOS Baseline (Fig. 5), a transformer-based model designed to take advantage of multipolarization petrographic images. This model takes as input the PPL and XPL images of the same thin section patch to capture complementary mineralogical features.

Firstly, we pretrained two single-polarization ViTs [7] on the LITHOS dataset, each specializing on either PPL or XPL images. Then, we extracted their encoders and froze them to preserve their learned polarization-specific features. This design encourages each encoder to focus on distinct, complementary representations. The frozen pretrained encoders independently transform each image into deep feature representations capturing mineral structures under distinct polarization conditions. Then, to combine the extracted representations, we introduce a dual-decoder architecture inspired by cross modality transformers [27, 34]. Each decoder comprises two consecutive layers that apply self-attention within their assigned feature set and cross-attention to features from the other polarization, followed by a feed-forward network (FFN).

We consider two different feature fusion strategies: (1) concatenating the outputs of both decoders along the feature dimension followed by a linear projection into a common space before classification, and (2) computing a weighted sum of the decoder outputs using a single learnable parameter  $\alpha \in [0, 1]$  such that the combined feature is  $F_{\text{combined}} = \alpha F_1 + (1 - \alpha) F_2$ , where  $F_1$  and  $F_2$  are the corresponding decoder outputs. We adopt the latter,  $\alpha$ -weighted summation as the fusion strategy, as it involves a single trainable parameter, directly combines the features, and maintains similar performance metrics compared to using the linear projection layer.

Finally, we pass the [CLS] token from  $F_{\text{combined}}$  through a fully connected classification head, which predicts the mineral type based on the integrated multi-polarization features. We optimize the model with the Cross Entropy loss, commonly employed for classification tasks.

**Implementation details.** We trained all variants of the model on a single Quadro RTX 8000 GPU, equipped with 48 GB of VRAM. For all experiments, we used a batch size of 64 and trained each model for 10 epochs. In the binary classification task (Quartz vs. Non-Quartz), we used the Adam optimizer and StepLR scheduler, a learning rate of  $1.3e - 4$ , momentum of 0.88, and weight decay factor ( $\gamma$ ) of 0.66. In the multi-class setting, we used the SGD optimizer with CosineAnnealingLR scheduling. The initial learning rate was set to  $8.9e - 4$ , momentum to 0.67, and  $\gamma$  to 0.89. The multi-class model was initialized from the pretrained weights of the binary classification task. Each single-polarization ViT experiment took around 10 hours, whereas the dual-polarization experiments

took approximately 18. Our LITHOS baseline has around 674M parameters, from which only 10% remain unfrozen during training.

## 5 Results and Discussion

Table 2 presents the results of our binary mineral classification task (Quartz vs. Non-quartz), comparing the performance of several deep learning architectures over our 2-fold cross-validation setup. We trained single-polarization baselines architectures, including ResNet, GoogLeNet, ViT, and Swin Transformer using either PPL or XPL, with our proposed LITHOS Baseline incorporating both modalities. Among the single-modality baseline architectures, ViT achieved the highest recall ( $0.894 \pm 0.021$ ), whereas Swin Transformer achieved the best F1-score ( $0.855 \pm 0.006$ ) when trained with XPL. However, the LITHOS Baseline consistently outperformed the single-modality methods on multiple metrics, including the F1 score ( $0.861 \pm 0.005$ ), precision ( $0.835 \pm 0.013$ ), and accuracy ( $0.851 \pm 0.002$ ).

Table 2: Binary classification results (Quartz vs. Non-Quartz). Models trained using either single or dual polarization modalities are evaluated using accuracy (Acc), recall, precision, and F1-score. Our LITHOS Baseline shows consistent improvements across various classification metrics compared to single-modality baselines. **Bold** indicates the best result, and underlined indicates the second best.

Binary Classification Task						
DL Architecture	PPL	XPL	Acc	Recall	Precision	F1-score
Resnet [13]	✓	✗	$0.819 \pm 0.017$	$0.882 \pm 0.037$	$0.792 \pm 0.003$	$0.834 \pm 0.018$
Googlenet [26]	✓	✗	$0.826 \pm 0.018$	$0.855 \pm 0.021$	$0.817 \pm 0.014$	$0.836 \pm 0.017$
ViT [7]	✓	✗	$0.834 \pm 0.018$	<u><math>0.893 \pm 0.012</math></u>	$0.807 \pm 0.017$	$0.848 \pm 0.016$
ViT [7]	✗	✓	$0.840 \pm 0.002$	<b><math>0.894 \pm 0.021</math></b>	$0.815 \pm 0.013$	$0.852 \pm 0.002$
Swin Transformer[15]	✓	✗	$0.835 \pm 0.014$	$0.872 \pm 0.016$	$0.821 \pm 0.011$	$0.845 \pm 0.013$
Swin Transformer[15]	✗	✓	<u><math>0.843 \pm 0.008</math></u>	$0.892 \pm 0.001$	<u><math>0.822 \pm 0.013</math></u>	<u><math>0.855 \pm 0.006</math></u>
LITHOS Baseline	✓	✓	<b><math>0.851 \pm 0.002</math></b>	$0.888 \pm 0.023$	<b><math>0.835 \pm 0.013</math></b>	<b><math>0.861 \pm 0.005</math></b>

Table 3: Multi-class mineral classification results across 25 categories. The models are evaluated using accuracy (Acc), and macro-averaged recall, precision, and F1-score. Our LITHOS Baseline achieves the highest classification scores across all metrics. **Bold** indicates the best result, and underlined indicates the second best.

Multi-Classification Task						
DL Architecture	PPL	XPL	Acc	Recall	Precision	F1-score
Resnet [13]	✓	✗	$0.573 \pm 0.025$	$0.342 \pm 0.024$	$0.412 \pm 0.018$	$0.343 \pm 0.013$
Googlenet [26]	✓	✗	$0.579 \pm 0.029$	$0.367 \pm 0.037$	$0.459 \pm 0.009$	$0.380 \pm 0.030$
ViT [7]	✓	✗	$0.573 \pm 0.019$	$0.353 \pm 0.018$	$0.436 \pm 0.032$	$0.360 \pm 0.018$
ViT [7]	✗	✓	$0.586 \pm 0.009$	$0.365 \pm 0.030$	$0.459 \pm 0.008$	$0.382 \pm 0.025$
Swin Transformer [15]	✓	✗	$0.593 \pm 0.025$	<b><math>0.421 \pm 0.016</math></b>	$0.470 \pm 0.006$	$0.422 \pm 0.010$
Swin Transformer [15]	✗	✓	<u><math>0.604 \pm 0.017</math></u>	<b><math>0.421 \pm 0.001</math></b>	<u><math>0.472 \pm 0.039</math></u>	<u><math>0.424 \pm 0.011</math></u>
LITHOS Baseline	✓	✓	<b><math>0.623 \pm 0.019</math></b>	<u><math>0.415 \pm 0.006</math></u>	<b><math>0.519 \pm 0.001</math></b>	<b><math>0.432 \pm 0.002</math></b>

Table 3 presents the results of our multi-class mineral classification task. Similar to the binary classification setup, we conducted all experiments using a 2-fold cross-validation strategy, ensuring a reliable assessment of each model’s generalization performance. Given the 25-class setting and the highly imbalanced nature of the dataset (as shown in Fig. 4), we report macro-averaged precision, recall, and F1-score to provide an unbiased evaluation across all classes. Among the single-polarization baselines, the Swin Transformer trained with XPL achieves the highest F1-score ( $0.424 \pm 0.011$ ), recall ( $0.421 \pm 0.001$ ), and precision ( $0.472 \pm 0.039$ ). Likewise to the binary task, our LITHOS Baseline achieves the best overall performance across key metrics, including an accuracy of  $0.623 \pm 0.019$ , a precision of  $0.519 \pm 0.001$ , and an F1-score of  $0.432 \pm 0.002$ , demonstrating the benefits of polarization-based feature extraction.

These results highlight the effectiveness of leveraging both polarization modalities, with the LITHOS Baseline consistently outperforming single-modality models across both proposed classification tasks. Particularly noteworthy is the cross-attention mechanism integrated within our dual-decoder fusion module (Fig. 5), which enables effective interaction between the feature maps extracted from each polarization modality. This design facilitates richer representations and leads to more accurate classifications in both tasks.

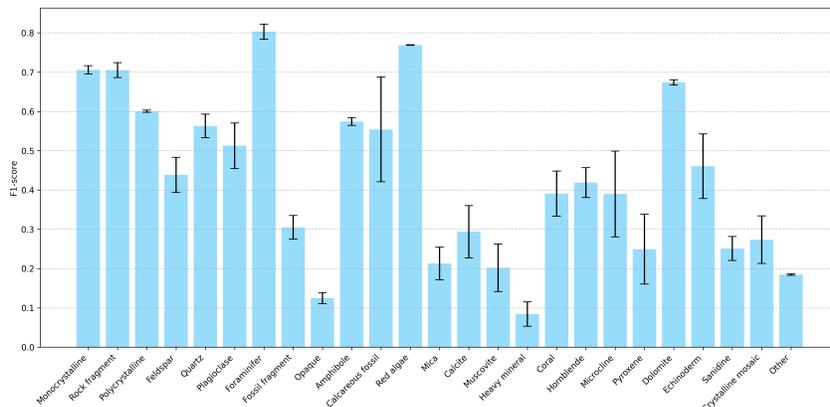


Figure 6: **Main Results of our proposed baseline model for the multi-classification task in the LITHOS Benchmark.** Bars represent the mean F1-score with standard deviation computed over the 2-folds for all 25 mineral categories.

Additional details on model complexity and inference performance are provided in the supplementary material (Table 1), where we show that the LITHOS Baseline, despite having a high total parameter count and number of FLOPs due to its dual ViT encoders, maintains a moderate training cost by keeping these encoders frozen. Precision-Recall curves for each class are also included in the supplementary material (Fig. 1 & 2) for both classification tasks.

Even though we observe an overall performance improvement with our model, a more detailed per-class analysis (Fig. 6) reveals a significant long-tail challenge related to class imbalance. Classes such as Monocrystalline, Rock Fragment, and Polycrystalline dominate the dataset, often achieving F1-scores above 0.60. However, low-frequency classes such as Opaque, Heavy mineral and Other, consistently exhibited F1-scores below 0.2. This imbalance reflects the natural distribution of minerals in geological samples, but also limits model performance on underrepresented classes.

We hypothesize that certain classes, such as Foraminifer and Red Algae, achieve notably high performance due to their distinctive biological structures, often characterized by porous textures or recognizable morphologies. Similarly, the model performs strongly on Dolomite, likely due to its unique rhombohedral shapes and cleavage or twinning patterns under XPL. In contrast, mid-performing classes often share overlapping visual characteristics under polarized light, especially in fine-grained or heterogeneous textures, making them harder to distinguish (eg. Feldspar, Quartz, and Plagioclase). Additionally, broader categories such as Heavy Mineral and Other encompass diverse and visually inconsistent instances, further contributing to classification challenges.

Further insight into the model’s behavior is provided by the confusion matrices in Supplementary Fig. 3. A considerable proportion of the model’s misclassifications arise from confusions with the two most represented classes: Monocrystalline and Rock fragment. The Other class is misclassified across nearly all categories, likely due to its intrinsic heterogeneity. Regarding the confusions between Sanidine, Microcline, and Feldspar, it is important to note that Sanidine and Microcline are specific potassium feldspars, whereas the Feldspar class refers to feldspars in general. All three minerals are colorless, low-relief, and weakly birefringent under polarized light, and are primarily distinguished by their twinning patterns, which can be subtle at the patch level.

Similarly, the misclassification among Fossil fragment, Foraminifer, Coral, and Echinoderm is reasonable, as these biogenic carbonates share similar optical behavior under both PPL and XPL light. Although Dolomite exhibits strong performance overall, its confusion with Crystalline mosaic likely results from their comparable textural and optical features: both display interlocking, equigranular carbonate crystals forming a mosaic-like texture and appear colorless under PPL. Despite these limitations, our baseline demonstrates robustness and generalization across both folds, validating the value of polarization fusion and setting a new benchmark in automated petrographic classification.

## 6 Limitations and Impact

Our work presents some limitations that offer avenues for future work. First, we acknowledge the geographical limitations of our dataset. However, Colombia exhibits a high degree of soil and geological diversity, which enhances the representativeness of our data to some extent. The country spans multiple climatic zones and varied topography, resulting in a broad range of soil types. Additionally, our collaborators collected these samples throughout the productive and potentially productive areas of the country. While this diversity does not fully compensate for the lack of data from other global regions, it provides a valuable starting point for broader applications. Additionally, the dataset exhibits a long-tail distribution, with nearly half of the annotations corresponding to quartz and its subtypes. This distribution reflects the natural abundance of these minerals in many geological contexts. We intentionally preserved this imbalance to ensure that models trained on the dataset are exposed to realistic mineral proportions encountered in practical applications, particularly in sustainable soil and rock exploration. While this long-tail distribution introduces challenges for multi-class classification, it also provides a more authentic benchmark for developing models capable of handling naturally imbalanced mineral compositions.

Regarding our proposed method, the supervision strategy of future methods could be improved by incorporating not only the intersection of the annotated grain-defining paths, but also the full extent of those paths, which represent the mineral’s minor and major axes. However, exploring such enhanced supervision was beyond this paper’s scope, as our primary goal was to establish a strong benchmark with baseline models based on single- and dual-modality polarization data using single-grain supervision. Additionally, the dataset exhibits class imbalance, a reflection of the natural distribution of minerals in soil, which suggests that further techniques could be applied to mitigate this and potentially improve performance metrics. Finally, our proposed LITHOS Baseline, is relatively large in terms of parameters because it combines two pretrained ViT encoders from single-modality baselines. Nevertheless, by freezing these encoders during training, we effectively reduce the computational burden.

The public release of LITHOS aims to democratize access to high-quality petrographic data, encourage reproducibility in mineral classification tasks, and stimulate interdisciplinary work between geoscientists and machine learning researchers. We believe that this work does not present ethical risks or negative societal impact and offers positive impact by enabling progress in automated petrography and geological analysis.

## 7 Conclusion

We introduced the Large-scale Imaging and Thin-section Optical-polarization Set (LITHOS) Benchmark, a comprehensive experimental framework that formalizes the problem of automated petrographic mineral classification by combining a large-scale, richly annotated dataset with a rigorous evaluation protocol and a set of baseline models. This benchmark defines a set of challenging binary and multi-class classification tasks grounded in real-world petrographic workflows, supported by expert annotations and dual-polarized imaging modalities. Our proposed LITHOS Baseline, which effectively integrates multi-modality polarization data, consistently outperformed single-modality baselines on both binary and multi-class setups, demonstrating the benefits of multi-view feature fusion. By publicly releasing the dataset, code, and pretrained models, we aim to establish a robust foundation that fosters reproducibility and stimulates interdisciplinary research towards automated petrography.

## 8 Acknowledgements

The authors gratefully acknowledge the financial support of the University of the Andes Foundation. This work was supported by Azure sponsorship credits granted by Microsoft’s AI for Good Research Lab. The authors further thank Cristhian Forigua, Natalia Valderrama, and Catalina Gómez Caballero for their valuable contributions in the initial stages of this project. Isai Daniel Chacón acknowledges the support of the 2023 UniAndes–DeepMind Scholarship.

## References

- [1] Saeed Aligholi, Reza Khajavi, and Morteza Razmara. Automated mineral identification algorithm using optical properties of crystals. *Computers and Geosciences*, 85:175–183, December 2015. ISSN 0098-3004. doi: 10.1016/j.cageo.2015.09.014. URL <http://dx.doi.org/10.1016/j.cageo.2015.09.014>.
- [2] İlhan Aydın, Ayşe Didem Kılıç, and Taha Kubilay Şener. Improving rock type identification through advanced deep learning-based segmentation models: A comparative study. *Applied Sciences*, 15(3):1630, February 2025. ISSN 2076-3417. doi: 10.3390/app15031630. URL <http://dx.doi.org/10.3390/app15031630>.
- [3] Edgar Berrezueta, Berta Ordóñez-Casado, Wilson Bonilla, Richard Banda, Ricardo Castroviejo, Paul Carrión, and Stalin Puglla. Ore petrography using optical image analysis: Application to zaruma-portovelo deposit (ecuador). *Geosciences*, 6(2):30, June 2016. ISSN 2076-3263. doi: 10.3390/geosciences6020030. URL <http://dx.doi.org/10.3390/geosciences6020030>.
- [4] Maria Heloisa Barros de Oliveira Frascá. *Petrographic Analysis*, page 705–713. Springer International Publishing, 2018. ISBN 9783319735689. doi: 10.1007/978-3-319-73568-9\_218. URL [http://dx.doi.org/10.1007/978-3-319-73568-9\\_218](http://dx.doi.org/10.1007/978-3-319-73568-9_218).
- [5] Paolo Dell’Aversana. An integrated deep learning framework for classification of mineral thin sections and other geo-data, a tutorial. *Minerals*, 13(5):584, April 2023. ISSN 2075-163X. doi: 10.3390/min13050584. URL <http://dx.doi.org/10.3390/min13050584>.
- [6] Beatriz M. Dias, Victor F. Velázquez, Rodrigo F. Lucena, and José M. Azevedo Sobrinho. Petrographic microscope digital image processing technique for texture and microstructure interpretation of earth materials. *Earth Science Research*, 9(1):58, January 2020. ISSN 1927-0542. doi: 10.5539/esr.v9n1p58. URL <http://dx.doi.org/10.5539/esr.v9n1p58>.
- [7] Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.
- [8] Ivan Ferreira, Luis Ochoa, and Ardiansyah Koeshidayatullah. On the generation of realistic synthetic petrographic datasets using a style-based gan. *Scientific Reports*, 12(1), July 2022. ISSN 2045-2322. doi: 10.1038/s41598-022-16034-4. URL <http://dx.doi.org/10.1038/s41598-022-16034-4>.
- [9] Ervan Garrison. *Petrography for Archaeological Geology*, page 145–178. Springer International Publishing, 2016. ISBN 9783319302324. doi: 10.1007/978-3-319-30232-4\_6. URL [http://dx.doi.org/10.1007/978-3-319-30232-4\\_6](http://dx.doi.org/10.1007/978-3-319-30232-4_6).
- [10] Gmas Geoscientific Solutions Laboratory. OneGeo, howpublished = <https://gmaslab.com/onegeo>.
- [11] David Haberland, Craig Strong, Duncan Pirrie, Gavyn K Rollinson, Paul Gottlieb, Pieter PWSK Botha, and Alan R Butcher. Automated petrography applications in quaternary science. *Quaternary Australasia*, 28(2):3–12, 2011.
- [12] Safaa M. Hassan, Noureldin Laban, Saif M. Abo Khashaba, N. H. El-Shibiny, Bashar Bashir, Mokhles K. Azer, Kirsten Drüppel, and Hatem M. Keshk. Semantic segmentation of some rock-forming mineral thin sections using deep learning algorithms: A case study from the nikeiba area, south eastern desert, egypt. *Remote Sensing*, 16(13):2276, June 2024. ISSN 2072-4292. doi: 10.3390/rs16132276. URL <http://dx.doi.org/10.3390/rs16132276>.
- [13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [14] Xin Liu, Viswasanthi Chandra, Ahmad Ihsan Ramdani, Rainer Zuhlke, and Volker Vahrenkamp. Using deep-learning to predict dunham textures and depositional facies of carbonate rocks from thin sections. *Geoenergy Science and Engineering*, 227:211906, August 2023. ISSN 2949-8910. doi: 10.1016/j.geoen.2023.211906. URL <http://dx.doi.org/10.1016/j.geoen.2023.211906>.

- [15] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.
- [16] Mezghani Mokhles, Anifowose Fatai, and Masrahy Mohammed. Advances in rock petrography: Image processing techniques for automated textural thin section analysis. In *SPE Middle East Oil and Gas Show and Conference*, 19MEOS. SPE, March 2019. doi: 10.2118/194835-ms. URL <http://dx.doi.org/10.2118/194835-MS>.
- [17] Soumitra B Nande and Samarth D Patwardhan. A deep learning approach for automating the thin-section petrography of the carbonate rocks for effective reservoir characterization. *Arabian Journal for Science and Engineering*, pages 1–33, 2024.
- [18] Fatick Nath, Sarker Asish, Shaon Sutradhar, Zhiyang Li, Nazmul Shahadat, Happy R. Debi, and S M Shamsul Hoque. Rock thin-section analysis and mineral detection utilizing deep learning approach. In *Proceedings of the 11th Unconventional Resources Technology Conference*, URTEC 2023. American Association of Petroleum Geologists, 2023. doi: 10.15530/urtec-2023-3865660. URL <http://dx.doi.org/10.15530/urtec-2023-3865660>.
- [19] Maria Gracia C. Padrique, Mark Jeremy G. Narag, Allan Gil S. Fernando, and Maricor N. Soriano. Enhancing geothermal petrography with convolutional neural networks. *Geothermics*, 127:103221, March 2025. ISSN 0375-6505. doi: 10.1016/j.geothermics.2024.103221. URL <http://dx.doi.org/10.1016/j.geothermics.2024.103221>.
- [20] H.M.T.G.A. Pitawala. Mineralogy, petrography, geochemistry and economic potential of carbonate rocks of sri lanka. *Journal of the Geological Society of Sri Lanka*, Feb 2019. doi: 10.4038/jgssl.v20i1.24.
- [21] Walter L. Pohl. *Economic Geology Principles and Practice: Metals, Minerals, Coal and Hydrocarbons – Introduction to Formation and Sustainable Exploitation of Mineral Deposits*. Wiley, April 2011. ISBN 9781444394870. doi: 10.1002/9781444394870. URL <http://dx.doi.org/10.1002/9781444394870>.
- [22] Yili Ren, Xin Li, Jianzhong Bi, Yunying Zhang, Qianxiao Su, Wenjie Wang, and Hongjue Li. Multi-channel attention transformer for rock thin-section image segmentation. *Journal of Engineering Research*, 2024. ISSN 2307-1877. doi: <https://doi.org/10.1016/j.jer.2024.04.009>. URL <https://www.sciencedirect.com/science/article/pii/S2307187724000993>.
- [23] S. A. Salman, E. A. Abou El-Anwar, W. A. Makled, K. H. Mahfouz, and Z. L. Belal. Hydrocarbon generation potential evaluation via petrographic and geochemical analyses of el-maghara coal in sinai, egypt. *Scientific Reports*, 14(1), January 2024. ISSN 2045-2322. doi: 10.1038/s41598-024-51291-5. URL <http://dx.doi.org/10.1038/s41598-024-51291-5>.
- [24] Yu Song, Robbie M. Damiani, Chuanyue Shen, Daniel I. Castaneda, and David A. Lange. A 3d petrographic analysis for concrete freeze-thaw protection. *Cement and Concrete Research*, 128:105952, February 2020. ISSN 0008-8846. doi: 10.1016/j.cemconres.2019.105952. URL <http://dx.doi.org/10.1016/j.cemconres.2019.105952>.
- [25] Zhu Shi Song, Yang Wen Yi, Lu Bi Bo, Huang Guang Yao, Hou Guang Shun, Wei Shi Peng, and Zhang Yan Lin. Micro image data set of some rock forming minerals, typical metamorphic minerals and oolitic thin sections, 2020. URL <https://www.scidb.cn/en/detail?dataSetId=684362351280914432>.
- [26] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [27] Hao Tan and Mohit Bansal. Lxmert: Learning cross-modality encoder representations from transformers. *arXiv preprint arXiv:1908.07490*, 2019.

- [28] Chiara Telloli, Alessandra Aprile, and Elena Marrocchino. Petrographic and physical-mechanical investigation of natural aggregates for concrete mixtures. *Materials*, 14(19):5763, October 2021. ISSN 1996-1944. doi: 10.3390/ma14195763. URL <http://dx.doi.org/10.3390/ma14195763>.
- [29] Hanyu Wang, Wei Cao, Yongzhang Zhou, Pengpeng Yu, and Wei Yang. Multitarget intelligent recognition of petrographic thin section images based on faster rcnn. *Minerals*, 13(7):872, June 2023. ISSN 2075-163X. doi: 10.3390/min13070872. URL <http://dx.doi.org/10.3390/min13070872>.
- [30] Jiaxin Yu, Florian Wellmann, Simon Virgo, Marven von Domarus, Mingze Jiang, Joyce Schmatz, and Bastian Leibe. Superpixel segmentations for thin sections: Evaluation of methods to enable the generation of machine learning training data sets. *Computers and Geosciences*, 170:105232, January 2023. ISSN 0098-3004. doi: 10.1016/j.cageo.2022.105232. URL <http://dx.doi.org/10.1016/j.cageo.2022.105232>.
- [31] Jiaxin Yu, Florian Wellmann, Simon Virgo, Marven von Domarus, Mingze Jiang, Joyce Schmatz, and Bastian Leibe. Superpixel segmentations for thin sections: Evaluation of methods to enable the generation of machine learning training data sets. *Computers and Geosciences*, 170:105232, January 2023. ISSN 0098-3004. doi: 10.1016/j.cageo.2022.105232. URL <http://dx.doi.org/10.1016/j.cageo.2022.105232>.
- [32] Ping Zhang, Jiazhou Zhou, Wen Zhao, Xuyang Li, and Liu Pu. The edge segmentation of grains in thin-section petrographic images utilising extinction consistency perception network. *Complex and Intelligent Systems*, 10(1):1231–1245, August 2023. ISSN 2198-6053. doi: 10.1007/s40747-023-01208-y. URL <http://dx.doi.org/10.1007/s40747-023-01208-y>.
- [33] Ping Zhang, Jiazhou Zhou, Wen Zhao, Xuyang Li, and Liu Pu. The edge segmentation of grains in thin-section petrographic images utilising extinction consistency perception network. *Complex and Intelligent Systems*, 10(1):1231–1245, August 2023. ISSN 2198-6053. doi: 10.1007/s40747-023-01208-y. URL <http://dx.doi.org/10.1007/s40747-023-01208-y>.
- [34] Chen Zheng, Quan Guo, and Parisa Kordjamshidi. Cross-modality relevance for reasoning on language and vision. *arXiv preprint arXiv:2005.06035*, 2020.
- [35] Dongyu Zheng, Li Hou, Xiumian Hu, Mingcai Hou, Kai Dong, Sihai Hu, Runlin Teng, and Chao Ma. Sediment grain segmentation in thin-section images using dual-modal vision transformer. *Computers and Geosciences*, 191:105664, September 2024. ISSN 0098-3004. doi: 10.1016/j.cageo.2024.105664. URL <http://dx.doi.org/10.1016/j.cageo.2024.105664>.

## NeurIPS Paper Checklist

### 1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper’s contributions and scope?

Answer: [Yes]

Justification: The main claims are justified because the abstract and introduction clearly present the LITHOS Benchmark as a comprehensive framework—including our novel public dataset, problem formulation, and baseline models (single- and double-modality). We explicitly state our key contributions at the end of the introduction, aligning fully with the paper’s scope.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

### 2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [Yes]

Justification: The paper discusses its limitations clearly in a dedicated “Limitations and Impact” section (Section 6). This section acknowledges key assumptions and areas for future improvement, such as the scope of supervision strategies, the natural class imbalance in the dataset, and the computational demands of the proposed LITHOS baseline model. It also reflects on the practical implications of these factors and the trade-offs made, demonstrating transparency about the robustness and scope of the approach.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate “Limitations” section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren’t acknowledged in the paper. The authors should use their best

judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

### 3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: There are no theoretical results, theorems, lemmas or proofs that should be stated or deduced in our paper.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

### 4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: We provide public access to our code and data, so that people can reproduce our results. Regardless of this, we fully describe our proposed architecture in the LITHOS Baseline (Section 4) and show a diagram of it (Fig. 5). In this section we also describe the resources we used and important implementation details for reproducibility purposes. Since a new dataset is also a contribution of our work, we fully describe the way in which we obtained our high resolution petrographic images and posterior RGB patches in the Section 3.1.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.

- (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
- (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
- (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

## 5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We provide full open access to our data via Kaggle () and code in github with clear and descriptive READMEs that allow any person interested in our work and petrography to faithfully reproduce the main experimental results.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

## 6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: Section 4 describes all the implementation details, hyperparameters and optimizer for training the LITHOS baseline. Also, our open access dataset in kaggle contains 2 csv folds regarding our 2-fold cross validation data split. In the Section 3.2 we also describe that we assigned each thin section exclusively to one of the folds to avoid data leakage. This allowed us to preserve the relative mineral distribution across folds.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.

- The full details can be provided either with the code, in appendix, or as supplemental material.

## 7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: We used a 2 fold-cross validation precisely for statistical significance and generalization assessment of our baseline models. All of the results we present in Table 2 and Table 3 present the mean and standard deviation between the two folds. Also, our Figure 6 shows the mean F1-Score and error bars for each mineral class in our dataset.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

## 8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: The paper provides information on computer resources in the Implementation Details paragraph from LITHOS baseline section 4.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

## 9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines?>

Answer: [Yes]

Justification: : We reviewed the NeurIPS Code of Ethics and assure that the research conforms in every respect to it.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

#### 10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: The paper discusses the potential positive societal impacts of the work in the "Limitations and Impact" section (Section 6), emphasizing how the public release of LITHOS aims to democratize access to high-quality petrographic data, promote reproducibility, and foster interdisciplinary collaboration between geoscientists and machine learning researchers. In this section we also describe that we do not foresee ethical risks or negative societal impacts associated with this research.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

#### 11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper poses no such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.

- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

## 12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We credit all people and institutions involved in our data collection and investigation. We use CC BY-NC-SA 4.0 for our Dataset. We are the owners of both Dataset and Code being released.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, [paperswithcode.com/datasets](https://paperswithcode.com/datasets) has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

## 13. New assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: All new assets introduced in the paper are thoroughly documented: the code is available on GitHub with detailed READMEs and inline comments, the dataset is hosted on Kaggle with complete metadata accessible on the webpage, and the annotation format is clearly explained in the supplementary material.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

## 14. Crowdsourcing and research with human subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

**15. Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

**16. Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [NA]

Justification: LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigor, or originality of the research.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.