

**SELF-SUPERVISED DEEP LEARNING FOR
GEOSCIENTIFIC IMAGE ANALYSIS**

BY

IVAN EDUARDO FERREIRA CHACUA

A Thesis Presented to the
COLLEGE OF GRADUATE AND INTERDISCIPLINARY STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

GEOLOGY

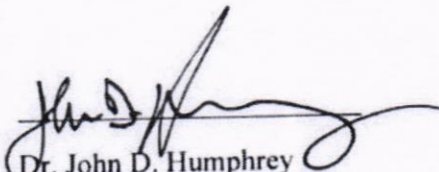
DECEMBER 2024

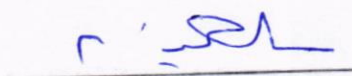
KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **IVAN EDUARDO FERREIRA CHACUA** under the direction of his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN GEOLOGY**


Dr. John D. Humphrey
(Chairman)

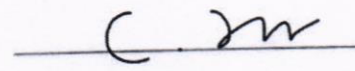

Prof. Suliman Saleh Al-Homidan
Dean of Graduate Studies

26/12/2024

Date


Dr. Ardiansyah Koeshidayatullah
(Advisor)


Prof. Khalid Al-Ramadan
(Member)


Prof. Cedric John
(Member)



© Ivan Ferreira

2024

To my mother, my grandmother and my father

ACKNOWLEDGMENTS

This thesis is dedicated to the incredible people who shaped and supported me throughout my journey. To my mother, Maria Luisa Chacua, whose unwavering love and encouragement have been a constant source of strength and inspiration, and to my father, José Iván Ferreira, a steadfast beacon of support and guidance. To my grandmother, Socorro del Carmen Chacua, who, with her warmth and nurturing care, has always been there to welcome me with a comforting embrace and a bowl of soup.

I also extend my heartfelt gratitude to Dr. Ardi, my mentor and thesis advisor, for his invaluable guidance, patience, and encouragement, even during the most challenging moments of this journey. I also extend my sincere gratitude to the committee members, Prof. Cedric John and Prof. Khalid Al Ramadan, for their invaluable support and insightful corrections. Additionally, I am grateful for the support provided by the KFUPM International Scholarship.

To my friends at KFUPM, thank you for your unwavering camaraderie and for sharing in both the triumphs and challenges of this academic journey. A special thanks to Julio Zevallos, whose steadfast companionship have been with me since the very beginning of this adventure. To Sara Garcia Uribe, for being my guiding light in moments of uncertainty, I am profoundly grateful—from the depths of my heart and soul. Your encouragement turned what seemed insurmountable into a rewarding experience.

To each of you, thank you for your love, faith, and support. This thesis is not just my accomplishment; it is a testament to the collective strength and inspiration you have given me throughout this journey.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	V
LIST OF TABLES	IX
LIST OF FIGURES	X
LIST OF ABBREVIATIONS	XIII
ABSTRACT.....	XIII
ملخص الرسالة	XV
CHAPTER 1 INTRODUCTION.....	1
1.1. Research Problems	5
1.2. Objectives	6
CHAPTER 2 PREVIOUS WORK	8
2.1. Convolutional Neural Networks for Image Classification in the Geosciences.....	8
2.2. Unsupervised Image Generation Using GANs	10
2.3. Super-Resolution for Data Augmentation.....	11
2.4. Semantic Segmentation Using Pre-Trained GANs	13
2.5. Semantic Segmentation Using Self-Supervised Learning.....	15
2.6. Foundational Models for segmentation.....	16
2.7. Semantic Segmentation using SSL	18
CHAPTER 3 DATASETS	26
3.1. Image Classification.....	26
3.2. Object Detection	28
3.3. Semantic Segmentation.....	29

CHAPTER 4 METHODOLOGY	32
4.1. Self-Supervised Learning	32
4.2. SSL Architectures	33
4.3. Inner workings of an SSL Architecture, SimCLR.....	34
4.4. You Only Look Once (YOLOV8).....	44
4.5. DINOv2: Self-Supervised Transformer, Focused on Getting Robust Visual Features	44
4.6. STEGO Self-Supervised Transformer, Energy-Based Graph Optimization	44
4.7. Computing Requirements.....	44
CHAPTER 5 RESULTS.....	46
5.1. Image Classification on LithoFacies using SSL	46
5.2. Image Similarity using SSL on Carbonate Thin Sections	47
5.4. Few-Shot Segmentation using pretrained GANs.....	50
CHAPTER 6 DISCUSSIONS AND FURTHER IMPROVEMENTS.....	57
CONCLUSIONS	62
APPENDIX.....	64
Glossary of Terms	64
REFERENCES	67
VITAE.....	74

LIST OF TABLES

Table 1. A selection of some Categories and Sub-categories of Self-Supervised Learning. Taken from Wang et al, 2022.	20
Table 2. Results for Image classification using both SSL and a custom CNN. The datasets were taken from (Shoji et al., 2018) for the volcanic ash particles and from (Baraboshkin et al., 2020; Koeshidayatullah, Al-Azani, et al., 2022) for the lithofacies dataset.	48
Table 3. Object Detection Performance Metrics for the Sandstone clast dataset.	52

LIST OF FIGURES

Figure 1. Number of publications containing Self-Supervised Learning in the paper. Data collected from Google Scholar (with a cut-off on November, 2024)	4
Figure 2. Workflow for Image Classification using Deep Convolutional Neural Networks, Taken from (Koeshidayatullah et al., 2020)	9
Figure 3. Examples of Image generation in the goesciences (A) Petrographical (B) Foraminifera specimens (C) Pollen Samples. (Ferreira et al., 2022)	11
Figure 4. Superresolution and Image Generation as a tool for data augmentation of Micropaleontological images. Taken from (Ferreira-Chacua & Koeshidayatullah, 2023)	13
Figure 5. (A) Five-Shot labels of foraminifera samples (B) Inference results with the Few-shot model trained.....	14
Figure 6. SSL Semantic Segmentation for Pyrite Analysis (a) Original images, (b) Segmentation Masks and (c) Predictions using STEGO.	16
Figure 7. Results of fine-tuning Segment Everything Model (SAM), leveraging pretrained foundational models and applied to an (A) Ore and a (B) Microfossil segmentation task.	18
Figure 8. The family of Self-Supervised Learning Algorithms.Taken from Wang et al, 2022.....	20
Figure 9. Schematic of self-supervised based learning, the two stage traning divided in pseudolabel pretraining and a downstream task (Image Classification or Image segmentation). Taken from Wang et al, 2022	24
Figure 10. Example of SSL augmentations, applied on a seismic dataset. Taken from Birnie et al, 2021.....	25
Figure 11. Example of PetroGAN training dataset, consisting of (a) Plutonic Rocks (b) Metamorphic rocks (c) Volcanic rocks and (d) Sedimentary rocks.....	27
Figure 12. Example of the Lithofacies Dataset of Core Images (a) Argillaceous Sandstone (B) Granite (C) Sandstone Massive (D) Limestone (E) Sandstone Laminated (F) Siltstone.	28
Figure 13. Example of the Sandstone component dataset, with the bounding box for each detected and labeled clast.....	29
Figure 14. Example of Images from the Carbonate Microfossil Dataset.....	30
Figure 15. Sample of the dataset used from Chuquicamata (a) Original Images (b) Segmentation Masks	31
Figure 16. SimCLR from the original images, the augmentations and representations obtained and how using contrastive loss (A) different classes of images representations will repel and (B) same class images will attract. Modified from (Chen et al., 2020)	38

Figure 17. Simplified Schematic of the STEGO Architecture, CRF: Conditional. Random Field, KNN: K-Nearest Neighbors (Hamilton et al., 2022)	43
Figure 18. Downstream accuracy of training an Image Classifier using only 20% of the Original Dataset and Validating on the remaining 80%.	46
Figure 19. Clustered Image embeddings of Microfossils using SimCLR	47
Figure 20. Accuracy and Loss for the 6-Class classification of the Lithofacies dataset (Baraboshkin et al., 2020; Koeshidayatullah et al., 2022).....	49
Figure 21. Accuracy and Loss for the 4-Class classification of the Volcanic Ash dataset (Shoji et al., 2018).....	50
Figure 22. Results of Accuracy and mIoU of SegFormer applied to paleontological image segmentation, using a one-epoch and five-shot approach.....	51
Figure 23. Example of a five-shot labels of synthetically generated foraminifera parts as input to the model (b) segmentations made by the model	51
Figure 24. Results of Object detection of classes in the Ordos Sandstone Dataset.	53
Figure 25. Loss, Precision and Recall graphs during the YoloV8 model training	54
Figure 26. Training graphs from Segformer trained on the dataset with a mIoU reaching 83%	55
Figure 27. Sample of (a) Original images in the dataset, (b) Ground truth pyrite masks (c) predictions made with the SSL trained model (d) SegFormer predictions.....	56
Figure 28. Comparison of STEGO for ore vein segmentation and comparison with a Binary thresholding operation.....	59
Figure 29. Improvements of predictions by an SSL model over human labels (a) original image (b) ground truth mask (c) model prediction.	61

LIST OF ABBREVIATIONS

SSL: Self-Supervised Learning

GAN: Generative Adversarial Network

FID: Fréchet Inception Distance

CNN: Convolutional Neural Network

ViT: Vision Transformer

ML: Machine Learning

XPL: Cross-Polarized Light

PPL: Plane-Polarized Light

mIoU: Mean Intersection over Union

SAM: Segment Anything Model

SEG: Segment Every Grain

SimCLR: Simple Framework for Contrastive Learning of Visual Representations

PSNR: Peak Signal-to-Noise Ratio

BYOL: Bootstrap Your Own Latent

MoCo: Momentum Contrast

SwAV: Swapping Assignments between Views

TiCo: Transformation Invariance and Covariance Contrast

ABSTRACT

Full Name : Ivan Eduardo Ferreira Chacua
Thesis Title : Self-Supervised Learning for Geoscientific Image Analysis
Major Field : Geology
Date of Degree : December 2024

Traditional deep learning approaches in the geosciences are often hindered by the limited availability and quality of labeled datasets, which are crucial for supervised learning methods. This scarcity of annotated data impedes the development of accurate and generalizable models for classification, segmentation, and clustering in geological and geophysical applications. To address these challenges, this study proposes novel methodologies leveraging self-supervised learning (SSL), a paradigm shift in machine learning enabling models to learn meaningful abstractions, or representations, from unlabeled data, thereby reducing reliance on extensive annotated datasets. In this research, we provide a comprehensive introduction to self-supervised learning within the context of the geosciences, emphasizing its critical importance in fields where pre-labeled datasets are scarce. We discuss the advantages of initiating neural network training with pre-trained models and explore the viability of foundational models in enhancing the effectiveness of SSL across various domains, including petrographic, micropaleontological, and seismic image analysis.

The study delves into a range of SSL tasks applicable to geoscientific data, such as classification, segmentation, and clustering, which are essential for unraveling intrinsic characteristics and facilitating downstream analytical processes. A significant focus is placed on implementing SSL techniques to solve concrete problems within geoscience, with particular emphasis on petrography. We review recent advancements in SSL as applied to computer vision and geosciences, address the challenges associated with data annotation, and investigate how SSL can effectively leverage unlabeled and sparsely labeled data to overcome these obstacles. Our research involves developing and critically evaluating SSL methods against established benchmarks, demonstrating their potential to improve model performance while reducing dependency on labeled data. We were able to achieve competitive accuracy using self-supervised algorithms such as STEGO for image segmentation and DinoV2 and SimCLR for image classification. These methods were applied to a variety of tasks, including the segmentation of ore mineral thin sections and microfossils, as well as the classification of volcanic ash particles and lithofacies core images. For instance, in our experiments with volcanic ash particle classification, DinoV2 achieved an impressive accuracy of 94.9% when using 781 images (80% of the dataset) and 91.8% when using 194 images (20% of the dataset), significantly outperforming a traditional CNN-based approach, which achieved 86.9%. Similarly, the STEGO model demonstrated its effectiveness in semantic segmentation tasks, achieving a mean IoU score

of 0.8, indicating a strong correlation between the ground truth masks and the model's predictions. Furthermore, in the few-shot generative autoshot approach, the segmentation maps for synthetic foraminifera images were highly satisfactory, achieving an overall IoU score of 84%.

These findings highlight SSL's ability to significantly contribute to more efficient and less label-intensive methods in geoscientific image analysis. We conclude by proposing future research directions in self-supervised learning, aiming to further advance computer vision applications in the earth sciences. By fostering the development of SSL techniques tailored to geoscientific challenges, this study contributes to the progression of machine learning methodologies that can handle the unique complexities of geoscientific data, ultimately enhancing our ability to interpret and analyze the Earth's subsurface structures and processes.

ملخص الرسالة

الاسم الكامل: إيفان إدواردو فيريرا تشاكوا
عنوان الأطروحة: التعلم الذاتي الإشراف لتحليل الصور الجيو علمية
مجال التخصص: الجيولوجيا
تاريخ الدرجة: ديسمبر 2024

يُعد التعلم العميق في مجال الجيو علم العديد من العقبات بسبب قلة توفر وجودة البيانات تواجه الطرق التقليدية الموصوفة، والتي تُعتبر ضرورية لأساليب التعلم تحت الإشراف. يؤدي هذا النقص في البيانات المشروحة إلى إعاقة تطوير نماذج دقيقة وعامة للمهام مثل تصنيف الصور، وتقسيمها، وعقدتها في التطبيقات الجيولوجية والجيوفيزيائية. وهو تحول (SSL) لمعالجة هذه التحديات، تقترح هذه الدراسة منهجيات مبتكرة تعتمد على التعلم الذاتي الإشراف نوعي في التعلم الآلي يمكن النماذج من تعلم تمثيلات ذات معنى من البيانات غير الموصوفة، مما يقلل من الاعتماد على مجموعات البيانات المشروحة بشكل مكثف. في هذا البحث، نقدم مقدمة شاملة عن التعلم الذاتي الإشراف في سياق الجيو علم، مع التركيز على أهميته البالغة في المجالات التي تندر فيها مجموعات البيانات المشروحة مسبقاً.

نناقش الفوائد التي يوفرها البدء بتدريب الشبكات العصبية باستخدام نماذج مدربة مسبقاً، ونستكشف إمكانية النماذج الأساسية في تعزيز فعالية التعلم الذاتي الإشراف في مختلف المجالات، بما في ذلك التحليل البتروغرافي، والمجهريّة الأحفورية، وتحليل الصور الزلزالية. تتناول الدراسة مجموعة من المهام القائمة على التعلم الذاتي الإشراف والمطبقة على البيانات الجيو علمية، مثل التصنيف والتقسيم والعقدة، التي تُعد ضرورية لفهم الخصائص الجوهرية وتسهيل العمليات التحليلية اللاحقة. يُركز بشكل كبير على تطبيق تقنيات التعلم الذاتي الإشراف لحل مشكلات عملية في الجيو علم، مع إيلاء اهتمام خاص للبتروغرافيا.

نستعرض التقدم الحديث في تطبيق التعلم الذاتي الإشراف على رؤية الحاسوب والجيو علم، ونُعالج التحديات المرتبطة بتوصيف البيانات، ونستكشف كيفية استخدام التعلم الذاتي الإشراف للبيانات غير الموصوفة والمحدودة التغذية لتجاوز هذه العقبات. يشمل بحثنا تطوير وتقييم تقنيات التعلم الذاتي الإشراف مقارنة بالمعايير المعمول بها، مما يُظهر قدرتها على تحسين أداء النماذج وتقليل الاعتماد على البيانات المشروحة. تشير نتائج بحثنا إلى أننا تمكنا من تحقيق دقة SimCLR وDinoV2 لتقسيم الصور، و STEGO تنافسية باستخدام خوارزميات التعلم الذاتي الإشراف مثل لتصنيف الصور.

تم تطبيق هذه الأساليب على مجموعة متنوعة من المهام، بما في ذلك تقسيم مقاطع المعادن الخام الرقيقة والأحفورات الدقيقة، بالإضافة إلى تصنيف جزيئات الرماد البركاني وصور اللثوفيسيات الأساسية. على سبيل المثال، في تجاربنا دقة مذهلة بلغت 94.9%، متفوقة بشكل كبير على DinoV2 على تصنيف جزيئات الرماد البركاني، حققت ، التي حققت 86.9%. وبالمثل، أثبت نموذج (CNN) الأساليب التقليدية المعتمدة على الشبكات العصبية الالتفافية بلغ 0.8، مما يشير إلى ارتباط (IoU) فعاليته في مهام التقسيم الدلالي، حيث حقق متوسطاً لنقاط التقاطع STEGO قوي بين الأفتعة المرجعية الحقيقية وتنبؤات النموذج. علاوة على ذلك، في نهج التوليد القليل الأمثلة، كانت خرائط %التقسيم للصور الأحفورية التركيبية مرضية للغاية، حيث حققت إجمالي نقطة تقاطع بلغت 84.

تُبرز هذه النتائج قدرة التعلم الذاتي الإشراف على المساهمة بشكل كبير في طرق أكثر كفاءة وأقل اعتماداً على البيانات المشروحة لتحليل الصور الجيو علمية. نختم باقتراح اتجاهات بحثية مستقبلية في مجال التعلم الذاتي الإشراف، بهدف تعزيز تطبيقات رؤية الحاسوب في علوم الأرض بشكل أكبر. من خلال تعزيز تطوير تقنيات التعلم الذاتي الإشراف المصممة للتحديات الجيو علمية، تُسهم هذه الدراسة في تطور منهجيات التعلم الآلي التي يمكنها التعامل مع التعقيدات الفريدة للبيانات الجيو علمية، مما يعزز في نهاية المطاف قدرتنا على تفسير وتحليل الهياكل والعمليات تحت سطح الأرض.

CHAPTER 1

INTRODUCTION

Geosciences have entered a transformative era, marked by the confluence of traditional geological expertise and innovative machine learning paradigms, notably by supervised Learning (Koeshidayatullah et al., 2020; E. J. Y. Koh et al., 2021; Shoji et al., 2018) and unsupervised Learning (Izadi et al., 2015), however with the advent of Self-Supervised Learning (SSL) methodologies, the use of new paradigms that attempt to reduce the reliance on labeled samples poses itself as an upcoming shift in how we approach machine learning systems in the geosciences. This review aims to synthesize the multitude of ways in which SSL is reshaping methodologies, applications, and insights intrinsic to various realms of geosciences such as seismic data processing, remote sensing, and petrography.

SSL stands out due to its proficiency in leveraging unlabeled data, thereby enhancing model robustness and facilitating nuanced geological interpretations without the necessity of extensive labeled datasets, a prerequisite that often hampers the applicability of supervised learning paradigms in real-world geoscientific problems (Y. Wang et al., 2022; Z. Wang et al., 2023)

In the domain of remote sensing, SSL manifests its profound impact, being instrumental in amplifying the accuracy of land cover and scene classification. It has catalyzed a synergy of multispectral and synthetic aperture radar images, enhancing the utilization of spectral characteristics (Berg et al., 2022; Montanaro et al., 2021). The adoption of SSL in seismic

exploration showcases its pivotal role in automating and enhancing geological feature detection. Leveraging SSL has facilitated the delineation of geological features from 3D seismic data (Pratama & Latiff, 2022).

Focusing on petrography, its intersection with machine learning (ML) methodologies has cultivated a fertile ground for data-driven pattern recognition (Ho et al., 2022; Koeshidayatullah et al., 2020; E. J. Y. Koh et al., 2021; R. P. Pires de Lima & Duarte, 2021). Over the years, the integration of ML has observed a near-exponential growth, reflecting the escalating interest and the transformative impact it has on the field. ML methodologies, especially those enhanced by vast labeled datasets, or supervised learning, are increasingly intertwined with fundamental tasks of petrographical image analysis, like image classification (Shoji et al., 2018), object detection (Koeshidayatullah et al., 2020) and Semantic Segmentation (E. J. Y. Koh et al., 2021), projecting a trajectory toward in the methodologies applied so far in petrographical image analysis could be further enhanced, besides the traditional approach of larger datasets (Deng et al., 2009) and deeper models (He et al., 2016) by ways of reducing the amount of necessary data either partially or totally, examples of this are exploration in the use of a reduced amount of an existing labelled dataset size (Dawson et al., 2023), transfer learning as a way of avoid random weight initialization and improve learning (R. Pires de Lima et al., 2020), weak supervision for tasks such as pore segmentation (Z. Wang et al., 2023) and use of self-supervised Learning for mineral segmentation in restricted scenarios, where high difference in contrast between minerals, such as in ore mineralogy is tested.

During the past decade, a remarkable surge in research efforts aimed at tackling the challenges of data annotation in ML, especially within the domain of self-supervised

learning (SSL). The number of publications related to SSL has grown exponentially, from just 8 in 2015 to over 3,200 by November, 2024 (Figure 1), highlighting the escalating interest and advancements in this SSL-related publications. This significant increase underscores the scientific community's recognition of SSL's potential to overcome these limitations of the available traditional supervised learning algorithms available, which often require large, labeled datasets that are costly and time-consuming to produce. Among these, SSL has recently garnered considerable attention in computer vision, achieving significant milestones toward reducing human supervision (X. Liu et al., 2023; Mamode et al., 2025; Rani et al., 2023a). Using the distillation of representative features or embeddings from unlabeled data, SSL algorithms have begun to outperform supervised pretraining on many problems. This progress is particularly relevant in fields like geosciences, where the scarcity of labeled data poses a significant barrier to the application of deep learning techniques.

The rapid growth in SSL research not only reflects its theoretical advancements but also the applicability in real-world scenarios in general. The ability of SSL to make use of the vast amounts of unlabeled data aligns well with the needs of geoscientific image analysis, where obtaining labeled datasets is often impractical. Therefore, the expanding body of SSL research is pivotal for driving innovation and overcoming existing challenges in geosciences and other data-intensive disciplines.(X. Liu et al., 2023; Rani et al., 2023a)

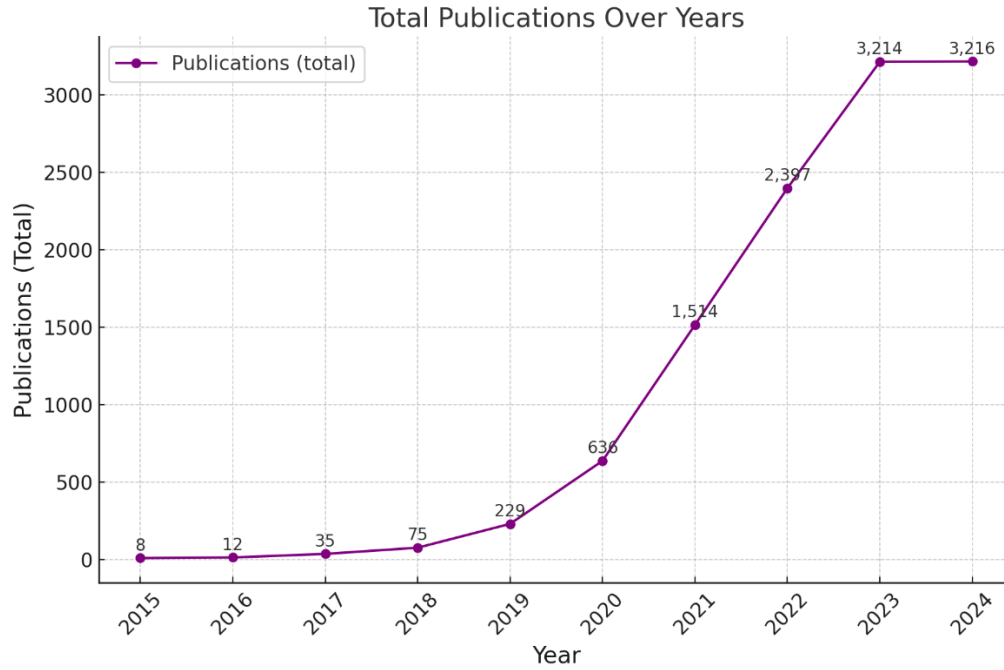


Figure 1. Number of publications containing Self-Supervised Learning in the Abstract.

Data collected from Google Scholar (with a cut-off on November, 2024)

In the realm of geoscientific analysis, the adoption of SSL methodologies, complemented by advances in deep learning models, presented in the literature and successful in many case scenarios such as convolutional neural networks (CNNs) and Vision Transformers (ViTs)(Dosovitskiy et al., n.d.; E. Koh et al., 2021; Lin et al., 2014), has fostered a renaissance in data-driven explorations. ViTs, equipped with architectures that prioritize attention mechanisms, has shown some promise in tasks such as image analysis, enhancing the quality and accuracy of geological interpretations in some sceneraios, although not completely outperforming CNNs (Koeshidayatullah, Al-Azani, et al., 2022; Liang et al., 2021).

While navigating through the landscape of SSL in geosciences, it is imperative to consider the ongoing evolution of methodologies and frameworks. Works such as those by (Sawarkar et al., 2022) have illuminated the paths toward implementing SSL, providing a beacon for the integration of SSL in various applications. Comprehensive reviews and explorations have further contributed to the discourse, offering critical insights into the concepts, challenges, and future trajectories of SSL in geosciences.

1.1. Research Problems

The primary research problem tackled in this thesis revolves around the limitations inherent in traditional deep learning methodologies, particularly in the geosciences. These limitations include (1) scarcity of Labeled Data: In geosciences, acquiring extensive labeled datasets for training deep learning models is often challenging, costly, and time-consuming. (2) quality of Available Data: The quality of available labeled datasets is frequently variable, which can impact the performance of supervised learning models. (3) Generalization of Models: Traditional models often struggle to generalize well to new, unseen data, especially in complex and varied geological contexts. These problems underscore the need for innovative approaches that can efficiently leverage unlabeled data, reduce reliance on large annotated datasets, and enhance the generalizability of models in geoscientific applications.

Data acquisition in the geosciences is often a costly and time-consuming endeavor, resulting in datasets that are limited in size—sometimes consisting of only tens of examples

(on the order of 10^1). This is significantly less than the dataset sizes typically required for most machine learning applications, which usually range from thousands to tens of thousands of examples (10^3 to 10^4). This discrepancy poses a substantial challenge for applying standard machine learning techniques (Dawson et al., 2023; Deng et al., 2009; E. J. Y. Koh et al., 2021). While self-supervised learning (SSL) reduces the dependence on labeled data, it still requires a considerable amount of unlabeled data to be effective. Therefore, this research thesis focuses on datasets comprising hundreds to a few thousand examples (on the order of 10^2 to 10^3), while using the available SSL models pretrained on 10^4 datasets such as ImageNet (Deng et al., 2009). For situations where even less data is available, alternative approaches such as traditional machine learning methods (e.g., Random Forests) or transfer learning techniques are recommended. (de Lima & Duarte, 2021; Ferreira et al., 2022; E. J. Y. Koh et al., 2021)

1.2.Objectives

The objectives of this research are designed to address the outlined problems through the application of Self-Supervised Learning (SSL). The specific goals include:

- Development of SSL Techniques: To develop and implement SSL methodologies tailored for geoscientific image analysis, focusing on extracting relevant features from unlabeled data.
- Benchmarking SSL Methods: To critically evaluate and benchmark the performance of SSL methods against traditional supervised learning paradigms in

various geoscientific domains, such as petrography, micropaleontology, and seismic image analysis.

- Application to Specific Geoscientific Tasks: To apply SSL techniques in solving concrete tasks within geoscience, particularly in petrography, and explore the potential of SSL in enhancing the accuracy and efficiency of geological interpretations.

Additionally, there is the necessity for integration with other machine learning paradigms like transfer learning and few-shot learning, aiming to create robust and versatile models for geoscientific applications. These objectives aim to contribute significantly to the advancement of computer vision applications in the earth sciences, offering new insights and methodologies for geoscientific image analysis.

CHAPTER 2

PREVIOUS WORK

2.1. Convolutional Neural Networks for Image Classification in the Geosciences

CNNs have revolutionized the field of image classification and segmentation, offering unparalleled accuracy and efficiency, this has been used extensively in the domain of the geosciences, Figure 2 (Koeshidayatullah et al., 2020; E. J. Y. Koh et al., 2021; Shoji et al., 2018; Wu et al., 2019). These architectures excel at handling the complexity and variety inherent in geoscientific imagery, such as satellite images, seismic data, and microscopic mineralogical structures. Unlike traditional neural networks, CNNs are designed to automatically and adaptively learn spatial hierarchies of features from input images. This capability stems from their unique architecture, which includes convolutional layers that reduce image dimensionality without losing critical information (Bridi Guazzelli et al., 2020; Koeshidayatullah et al., 2020; E. J. Y. Koh et al., 2021; Marchant et al., 2020; Ronneberger et al., 2015; Sevillano et al., 2020).

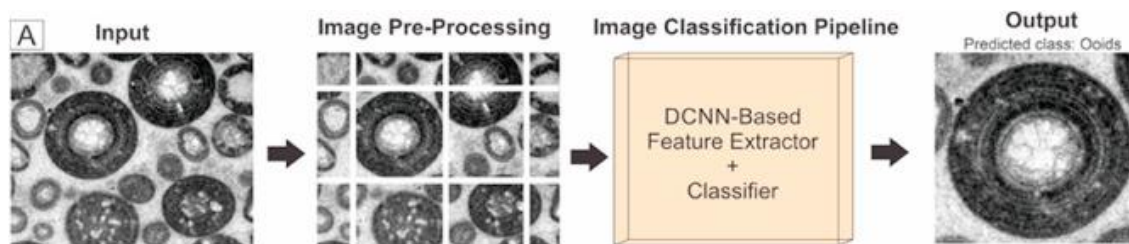


Figure 2. Workflow for Image Classification using Deep Convolutional Neural Networks, Taken from (Koeshidayatullah et al., 2020)

The implementation of CNNs in geosciences is not without challenges. One significant hurdle is the requirement for large volumes of labeled training data. While transfer learning has mitigated this issue to some extent, allowing models to be trained on smaller datasets, the need for extensive and accurate annotations remains a bottleneck (Dawson et al., 2023). This is particularly evident in tasks like carbonate petrography and modal mineralogy analysis, where detailed and precise labeling is crucial. Automated methods for labeling and annotating geoscientific images are still in their infancy, often requiring significant manual input from experts (E. J. Y. Koh et al., 2021).

Despite these challenges, the potential of CNNs in geosciences is immense. Their ability to classify and analyze images at both macro and micro scales opens up new avenues for exploration and research. Future advancements in CNN architectures, coupled with better data acquisition and labeling techniques, are likely to further enhance their applicability in geosciences. This could lead to more accurate geological interpretations, improved exploration techniques, and a deeper understanding of Earth's subsurface structures and processes. However, as technology progresses, CNNs will be progressively coupled with ViTs (Koeshidayatullah, Al-Azani, et al., 2022) and SSL architectures as they continue in their integration as tools in the geoscientist's arsenal, transforming the way we visualize and interpret our planet's geological fabric (Berg et al., 2022; Y. Wang et al., 2022).

2.2. Unsupervised Image Generation Using GANs

An innovative use of unlabeled data is the training of Generative Adversarial Networks, for this we trained a style-based Generative Adversarial Network, GAN, (Goodfellow et al., 2014b) training PetroGAN a generative model capable of generating high-quality thin section images, Figure 3 - A (Ferreira et al., 2022), this can be used for creating realistic synthetic petrographic datasets that can be further trained into several more specific domains. For this we leveraged the StyleGAN2 architecture (Karras, Aittala, et al., 2020; Karras, Laine, et al., 2020), incorporating the adaptive discriminator augmentation (ADA) to effectively replicate petrographic data's statistical and aesthetic characteristics. The model was trained on over 10,000 thin section images, achieving a state-of-the-art Fréchet Inception Distance (FID) (Heusel et al., 2017) score in the geosciences, This metric being a proxy for high-fidelity during the image generation. The resultant synthetic images were so realistic that even experts in geosciences could not distinguish them from actual images (Ferreira et al., 2022). This breakthrough demonstrates the significant potential of GANs in generating synthetic datasets, particularly in fields like petrography where data scarcity is a challenge.

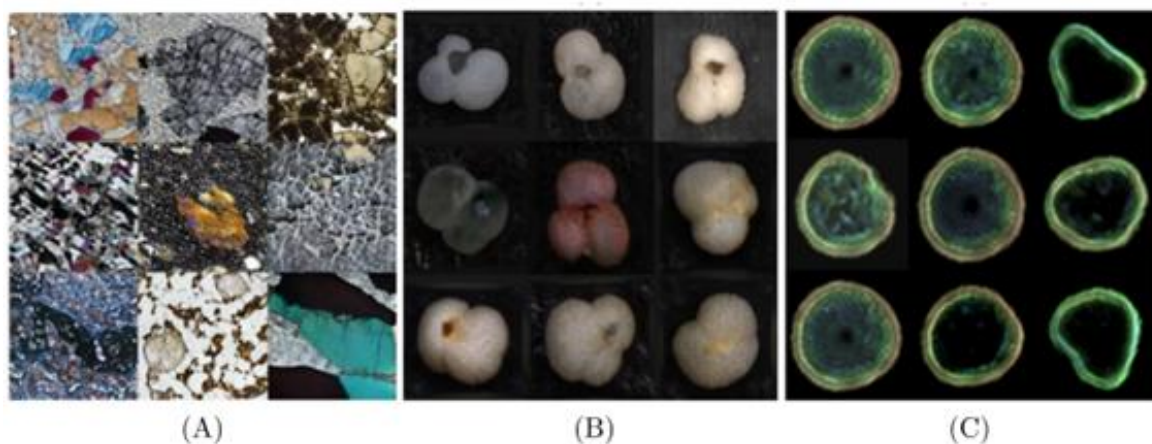


Figure 3. Examples of Image generation in the goesciences (A) Petrographical data (B) Foraminifera specimens (C) Pollen Samples. Modified from (Ferreira et al., 2022)

The research further delves into the broader context of deep learning and artificial intelligence in geosciences. Generative models like GANs, Variational Auto-Encoders, and Diffusion models have shown remarkable capabilities in generating realistic and varied images (Dhariwal & Nichol, 2021; Goodfellow et al., 2014b; Kingma & Welling, 2013), each of these approaches with several unique advantages to them. These models, especially GANs, have been successful in multiple domains in the earth sciences, including image super-resolution (Bizhani et al., 2022), data augmentation (Klyuchnikov et al., 2022) and image to image translation (Koeshidayatullah & Ferreira, 2022). In geosciences, the role of GANs in enhancing image processing and analysis images is a paradigm shift.

2.3. Super-Resolution for Data Augmentation

Another approach for dealing with data scarcity is enhancing existing imagery, for this we used available foraminifera datasets for training a micropaleontological image analysis workflow. In micropaleontology, crucial for reconstructing past environmental and climatic conditions, we rely heavily on the visual recognition of microfossils (Ferreira-Chacua & Koeshidayatullah, 2023; Sen, 2003). This field often grapples with the limitations of low-resolution, unlabeled datasets. Traditional methods, primarily using deep convolutional neural networks (CNNs) with transfer learning, and have limitations in data scarcity, specially of certain specimens and resolution, most of them dealing with low-resolution downsampled versions of an image (Marchant et al., 2020; Mitra et al., 2019). The ForamViT-GAN aims to overcome these challenges by integrating hierarchical vision transformers, using SwinIR (Dosovitskiy et al., 2020; Z. Liu et al., 2021) with generative adversarial network algorithms, using StyleGAN2 (Karras et al., 2019), enabling both the synthesis of and the super-resolution of realistic micropaleontological images.

The approach employs a deep learning workflow, Figure 4, to generate high-resolution images with high peak signal-to-noise ratios (PSNR) and Fréchet inception distance (FID) similarity scores. Key aspects include few-shot semantic segmentation for classifying various foraminifera chambers (Tritrong et al., 2021) and creating a large volume of self-labeled datasets for model benchmarking and downstream visual tasks. The methodology uniquely combines the strengths of GANs and vision transformers, setting a new path in image analysis and classification in micropaleontology.

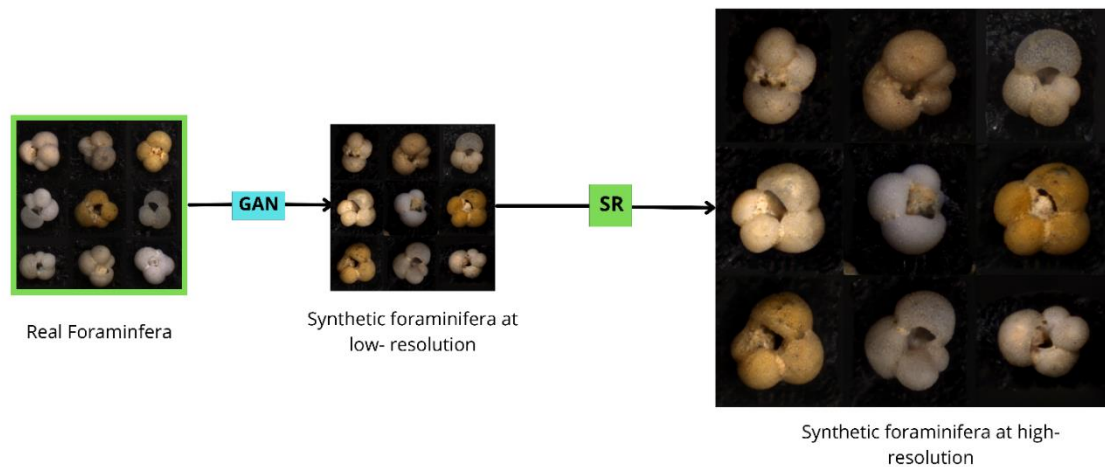


Figure 4. Superresolution and Image Generation as a tool for data augmentation of Micropaleontological images. Taken from (Ferreira-Chacua & Koeshidayatullah, 2023)

2.4. Semantic Segmentation Using Pre-Trained GANs

We can use the pre-trained weights of a Generative Adversarial Network (GAN), a set of networks that compete against each other and are able to generate realistic images by using an unlabeled dataset for training, by repurposing the learned representations during the training of this GAN (Goodfellow et al., 2014; Karras, Laine, et al., 2020) we then can label a few images and use the learned representations or embeddings with a CNN to train a FewShot CNN, this is harnessed via a one-shot or few-shot segmentation task (Tritrong et al., 2021) and is designed to address the challenges in the geosciences where data labels may not be available, particularly we apply it in the classification and recognition of microfossils such as foraminifera, of ooids and of heavy minerals. The research highlights the potential of using few-shot and one-shot learning techniques in computer vision, which are crucial for extracting meaningful labels from image datasets efficiently and effectively,

a critical need in geological workflows due to the time-consuming nature of manual labeling (E. J. Y. Koh et al., 2021).

The methodology involves using StyleGAN2 architecture to generate synthetic datasets of micropaleontological samples (Karras, Aittala, et al., 2020; Karras et al., 2019). These datasets are then used for semantic segmentation using a few-shot CNN segmentation architecture (Tritrong et al., 2021). The datasets used include high-resolution images of ooids and foraminifera (Koeshidayatullah, Trower, et al., 2022; Marchant et al., 2020), crucial in sedimentology and paleoclimatology (Sen, 2003). The StyleGAN2 models, trained on unlabeled images, produce synthetic samples that closely resemble real samples in both aesthetic and statistical properties of the datasets in question and the few-shot CNN segmentation architecture is applied to these synthetic images, using a limited number of labeled examples from the original datasets, to create self-labeled data for semantic segmentation, a technique also called, auto-shot (Tritrong et al., 2021). Successfully demonstrating the use of an unsupervised pretrained GAN in generating self-labeled geological datasets for semantic segmentation generating labels that can be used for a downstream task, Figure 5.

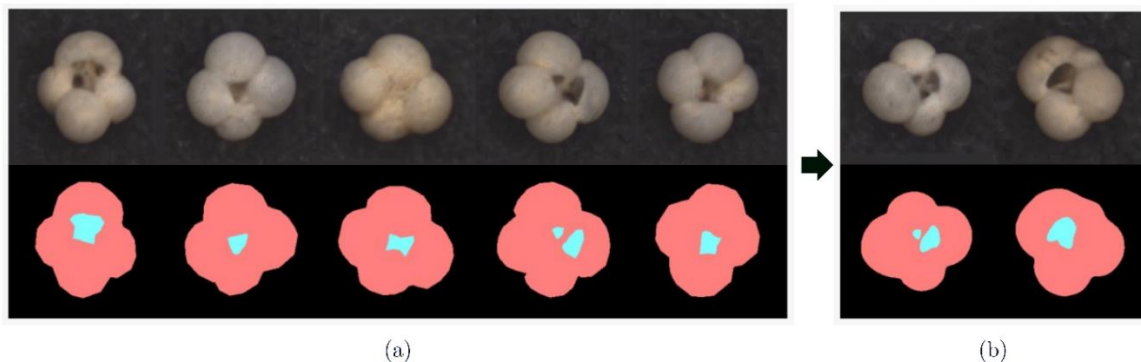


Figure 5. (A) Five-Shot labels of foraminifera samples (B) Inference results with the Few-shot model trained.

2.5. Semantic Segmentation Using Self-Supervised Learning

Focusing on quantitative mineralogical analysis, we also addressed the challenges of obtaining modal mineralogy (E. J. Y. Koh et al., 2021), traditionally hindered by the need for high-quality annotated datasets. These conventional datasets require extensive, error-prone human effort in labeling, making the process both time-consuming and susceptible to inaccuracies (Dawson et al., 2023; E. J. Y. Koh et al., 2021; R. Pires de Lima et al., 2020). For this purpose, we utilize the Self-Supervised Transformer with Energy-based Graph Optimization (STEGO) (Hamilton et al., 2022) architecture for clustering ore minerals. A model designed to identify categories of interest from images without relying on human annotations. The study specifically targets pyrite segmentation within thin sections from the Chuquicamata porphyry deposit and it demonstrates the model's impressive capability of accurately segmenting minerals, achieving a mean Intersection over Union (mIoU) score of 0.8 when tested against 42 human-labeled images of pyrite examples, Figure 5. This score not only validates the model's effectiveness but also its ability to correctly segment minerals that were not labeled during the annotation process.

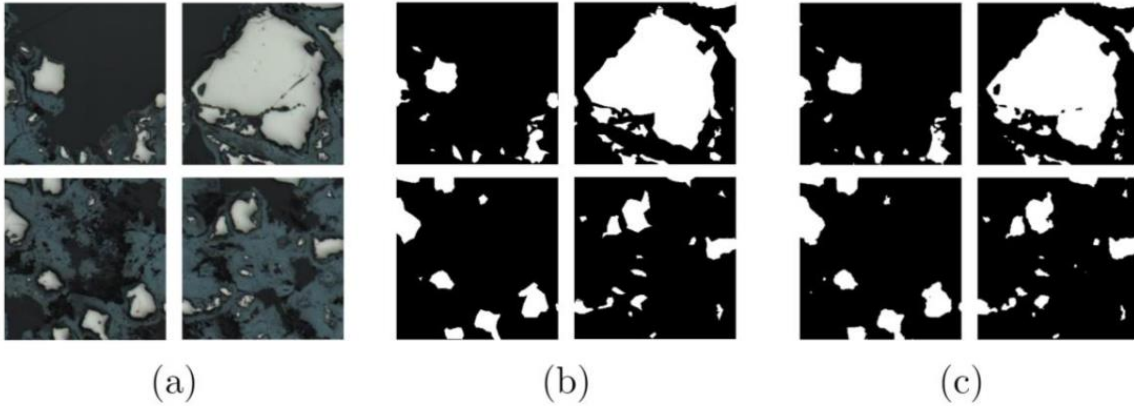


Figure 6. SSL Semantic Segmentation for Pyrite Analysis (a) Original images, (b) Segmentation Masks and (c) Predictions using STEGO.

This research represents a significant stride in the domain of petrography, showcasing the potential of SSL in automating and enhancing the accuracy of mineral identification and segmentation. The SSL approach, exemplified by the STEGO model, marks a departure from traditional methods reliant on extensive human annotation, offering a more efficient and potentially more accurate alternative. While the results are promising, particularly in high-contrast scenarios like ore minerals in reflected light where minerals are easily distinguishable, further research is needed to validate these findings across different lighting conditions and mineral types.

2.6. Foundational Models for segmentation

Currently computer vision models are based upon what are called foundational models, these are general purpose models that be applied across a wide range of applications one

of these applications being the "Segment Anything Model" (SAM) (Kirillov et al., 2023) and the subsequent "Segment Every Grain" (SEG) (Silvester, 2024), we repurpose these architectures for the automatic segmentation of microfossils in thin section images, Figure 7. Trained on a comprehensive dataset of over 8000 labeled microfossil images, this fossil segmentation model uses a custom U-Net architecture and a modified Attention U-Net architecture to enhance the precision and focus on crucial segmentation areas like fossil boundaries. The application of attention mechanisms in the latter architecture allows the model to concentrate on specific regions in the input images, thereby improving the segmentation accuracy. This sophisticated approach enables us to adeptly handle the intricate details and diversity of microfossil structures, which are often challenging for traditional computer vision methods such as thresholding or clustering. We achieved an accuracy of 0.91 and a corresponding loss of 0.93, demonstrating its high proficiency in microfossil segmentation tasks. The model's capacity for detailed morphological analysis of microfossils, such as diameter, perimeter, and axis measurements, as shown in the provided figures, highlights its potential to substantially contribute to paleontological and geological research, we also explored this model with other geological datasets such as modal mineralogy for ore thin sections, Figure 7-A.



Figure 7. Results of fine-tuning Segment Everything Model (SAM), leveraging pretrained foundational models and applied to an (A) Ore and a (B) Microfossil segmentation task.

2.7. Semantic Segmentation using SSL

Quantitative mineralogical analysis is critical for understanding the origins and characterization of mineral deposits. Recent advancements in supervised deep learning have transformed how qualitative and quantitative mineral analysis is performed in ore mineral exploration. However, the current deep learning-based approaches depend primarily on high-quality annotated mineral datasets. This is further hindered by the time-consuming and error-prone labeling processes. The emergence of self-supervised deep learning in many computer vision tasks provides a novel method to accurately perform quantitative mineral analysis without the need of manual labeling. Furthermore, the actual potential of self-supervised deep learning in geosciences remains underexplored. In this study, we adopted a Self-Supervised Transformer architecture (STEGO, Self-supervised Transformer with Energy-based Graph Optimization) to automatically segment out ore

minerals (e.g., pyrite) from other background gangue minerals in the thin sections collected from the Chuquicamata porphyry deposit (Chile). Our proposed method achieved 80% in the mean Intersection over Union, indicating the model's ability to accurately segment ore minerals without the need of extensive labeling. We further evaluated the model with human annotated minerals and observed the ability of the model to identify more ore minerals than manual labeling. Some other applications of this method would allow an efficient and fully automated quantitative mineralogical analysis for ore mineral exploration elsewhere. This work describes the first use of self-supervised deep learning in mineralogical analysis and also demonstrates how this approach can be implemented to perform advanced image-based deep learning geosciences analysis without large volume of labelled datasets.

2.8. Self-Supervised Learning in the Geosciences

Despite this promising progress, supervised deep learning methods come with significant limitation whereby a large number of labelled samples during the training of these models are required. In geosciences, such datasets (in the order of 10³) are generally hard to acquire, particularly for semantic segmentation task where the data labeling process is the most expensive, taking 100 times more human effort to label than other computer vision tasks (Zlateski et al., 2018). Self-Supervised methods are presented themselves as an label-effective alternative to these supervised algorithms, a brief taxonomy of the proposed methodologies presented in Table 1 and Figure 8; discriminating between three main classes Generative, Contrastive and Predictive; of which the former two are explored in this research.

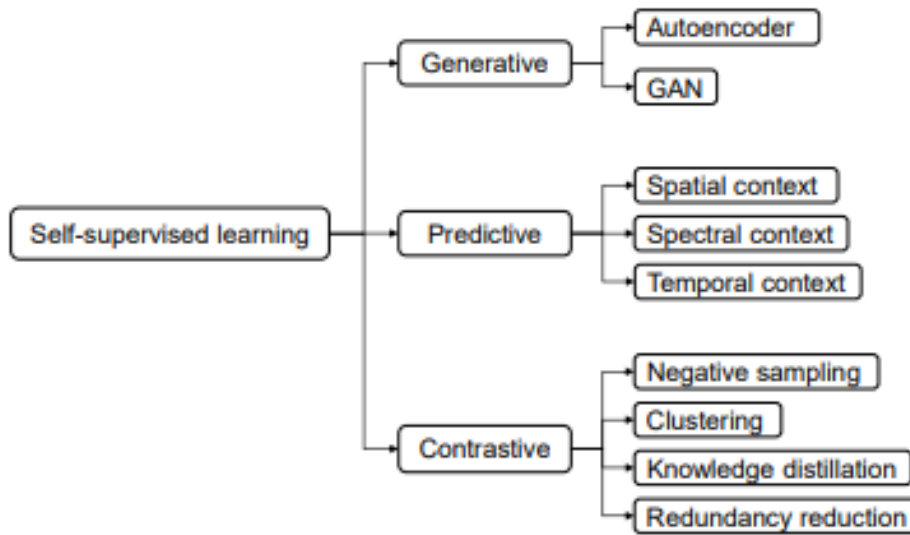


Figure 8. The family of Self-Supervised Learning Algorithms. Taken from Wang et al, 2022.

Category	Sub-category	Representative methods
Generative	Autoencoder	Autoencoder (AE) (Ballard, 1987): encode and reconstruct the input.
Generative	Autoencoder	Sparse AE (Ng et al., 2011): sparsity constraints on the hidden units.
Generative	Autoencoder	Denoising AE (Vincent et al., 2010): reconstruct clear image from noisy input.
Generative	Autoencoder	VAE (Kingma and Welling, 2013): encode the input to a normal distribution.
Generative	Autoencoder	MAE (He et al., 2021): reconstruct randomly masked patches with vision transformer.
Generative	GAN	GAN (Goodfellow et al., 2014): adversarial training with a generator and a discriminator.
Generative	GAN	AAE (Makhzani et al., 2015): adversarial variational autoencoder.
Generative	GAN	BiGAN (Donahue et al., 2016): an additional encoder to map samples to representations.

Predictive	Spatial	Relative position (Doersch et al., 2015): predict relative position of random patch pairs.
Predictive	Spatial	Jigsaw (Noroozi and Favaro, 2016): predict the correct order of a jigsaw puzzle.
Predictive	Spatial	Rotation (Gidaris et al., 2018): predict the rotation angle of rotated images.
Predictive	Spatial	Inpainting (Pathak et al., 2016): recover a missing patch of the input image.
Predictive	Spectral	Colorization (Zhang et al., 2016): predict colorful image from gray-scale input.
Predictive	Temporal	Frame order (Kim et al., 2019): predict the order of frame sequences.
Predictive	Others	Counting (Noroozi et al., 2017): count the visual primitives within the input image.
Predictive	Others	Artifact (Jenni and Favaro, 2018): spot and predict artifacts.
Predictive	Others	Audio (Owens et al., 2016): predict a statistical summary of the sound associated with a video.
Contrastive	Negative sampling	Triplet loss (Chopra et al., 2005): learn a similarity metric discriminatively with triplet loss.
Contrastive	Negative sampling	CPC (van den Oord et al., 2018): contrastive predictive coding with InfoNCE loss.
Contrastive	Negative sampling	DIM (Hjelm et al., 2018): maximize mutual information between global-local feature pairs.
Contrastive	Negative sampling	InstDisc (Wu et al., 2018): maximally scatter the features of different images over a unit sphere.
Contrastive	Negative sampling	PIRL (Misra and van der Maaten, 2020): pretext tasks as data augmentation.
Contrastive	Negative sampling	MoCo (He et al., 2020): store negative samples in a queue and momentum-update the key encoder.
Contrastive	Negative sampling	SimCLR (Chen et al., 2020): end-to-end contrastive structure with large batch size and strong data augmentation.

Contrastive	Negative sampling	MoCo-v3 (Chen et al., 2021): introduce vision transformer as the encoder backbone for MoCo-v2.
Clustering	Clustering	DeepCluster (Caron et al., 2018): iteratively leverage k-means clustering to yield pseudo labels for prediction.
Clustering	Clustering	LocalAgg (Zhuang et al., 2019): optimize a local soft-clustering metric.
Clustering	Clustering	SeLa (Asano et al., 2019): solve an optimal transport problem to obtain the pseudo-labels.
Clustering	Clustering	SwAV (Caron et al., 2020): contrastive learning with online clustering.
Clustering	Clustering	PCL (Li et al., 2020): perform iterative clustering and representation learning in an EM-based framework.
Knowledge distillation	Knowledge distillation	BYOL (Grill et al., 2020): mean-teacher network with a predictor on top of the teacher encoder.
Knowledge distillation	Knowledge distillation	SimSiam (Chen and He, 2021): explore the simplest design of contrastive self-supervised learning.
Knowledge distillation	Knowledge distillation	DINO (Caron et al., 2021): explores the self-distillation scheme with vision transformer backbones.
Knowledge distillation	Knowledge distillation	EsViT (Li et al., 2021): improve DINO with additional region-level contrastive task.
Knowledge distillation	Knowledge distillation	iBOT (Zhou et al., 2021): cross-view and in-view self-distillation with masked image modeling.
Redundancy reduction	Redundancy reduction	Barlow Twins (Zbontar et al., 2021): redundancy reduction on the cross-correlation matrix between features.
Redundancy reduction	Redundancy reduction	VICReg (Bardes et al., 2021): variance-invariance-covariance regularization.

Table 1.. A selection of some Categories and Sub-categories of Self-Supervised Learning.

Taken from Wang et al, 2022.

In addition, manual human annotation for some tasks, especially in geosciences, are not always well defined and/or require extensive domain expertise to achieve satisfactory result (de Lima & Duarte, 2021; Koeshidayatullah et al., 2020; Koh et al., 2021).

Therefore, the implementation of other deep learning approaches that can achieve similar results to supervised ones is a current active area of interest in computer vision. With the goal of using partial or no annotation at all in the dataset as a solution to the human annotation process, the recently emerges Self-Supervised Learning (SSL) approach have been proposed (Jaiswal et al., 2020; Kolesnikov et al., 2019; Zhai et al., 2019). SSL approach can be divided into two main stages. First, it aims to learn from unlabeled datasets, a goal that is achieved by generating pseudo-labels from the images and solving tasks during a pretraining phase, such as completing, colorizing, and solving jigsaw puzzles using the given images. Second, this pretraining without labels then is improved with a computer vision task such as image classification or segmentation with a small fraction of the annotations or even no annotations at all (Jaiswal et al., 2020; Kolesnikov et al., 2019; Zhai et al., 2019) a diagram of this process, the self-supervised training on pseudolabeled data and the downstream task to solve is shown in Figure 9. In image segmentation, recent progress has been made by accomplishing distillation of features via SSL that correlate with ground truth labels in datasets, with the mean Intersection over Union (mIoU) scores being competitive when compared to Supervised learning approaches in standard benchmarking datasets (Hamilton et al., 2022).

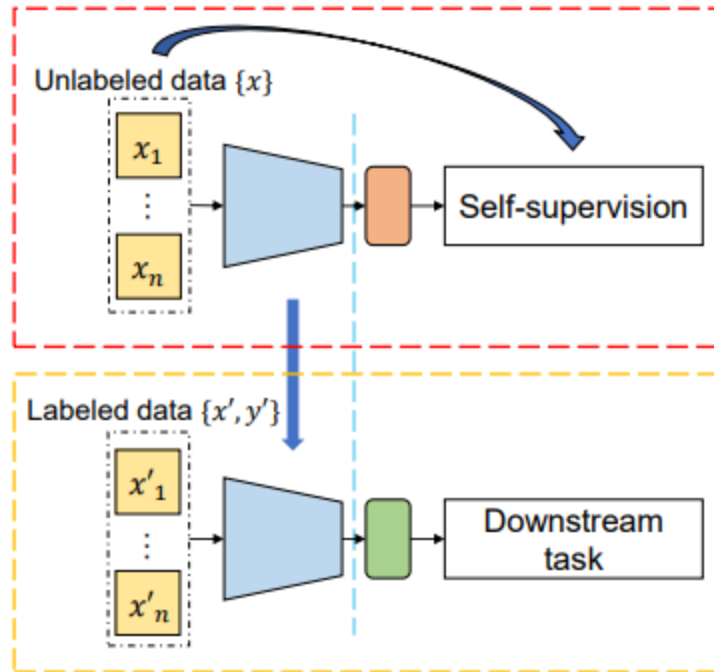


Figure 9. Schematic of self-supervised based learning, the two stage training divided in pseudolabel pretraining and a downstream task (e.g. Image Classification or Image segmentation). Taken from Wang et al, 2022

In geosciences, the use of SSL approach remains very. Most works have focused their efforts in applying SSL to seismic, with denoising and interpolation being tasks where SSL have been of use. In such a case, the reliance on human effort to create datasets for seismic segmentation being addressed, achieving the creation of annotated datasets of geological elements, such as horizons, faults, and salt layers in seismic images. However, and despite its application to seismic datasets, the potential of this approach to machine learning in petrography and other visual tasks in geosciences remaining ill-explored at the moment, an example applied to seismic images, with the self-supervised pseudo-labels shown in Figure 10 (Aribido et al., 2020; Mamode et al., 2025)

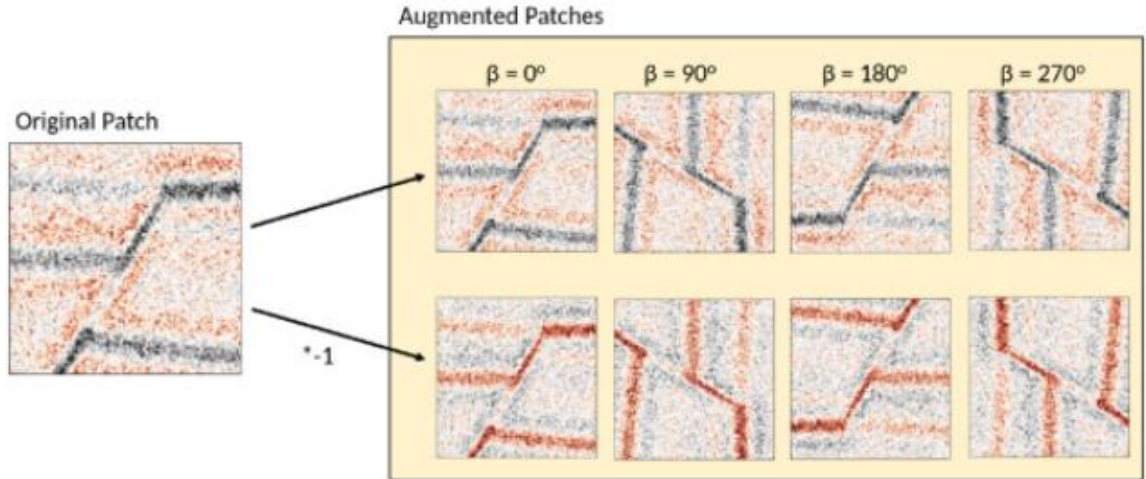


Figure 10. Example of SSL augmentations, applied on a seismic dataset. Taken from Birnie et al, 2021

Therefore, our study aims to expand the application of SSL methodologies into the domain of petrography and mineral segmentation, areas in which the utility of this algorithms has not been yet fully explored or leveraged. We aim to demonstrate the effectiveness of SSL for ore mineral segmentation without relying on labeled examples. The potential implications of successfully adopting SSL approaches to machine learning in the geosciences are to reduce the time and financial cost associated with the manual labelling of geological datasets and also to give alternative architecture to the current supervised ones.

CHAPTER 3

DATASETS

For training and testing all the major computer vision tasks: Classification, Object Detection and Semantic Segmentation we have compiled or created several datasets, addressing several geological problems such as mineral recognition, fault delineation and microfossil morphometry, Augmentations such as rotation, scaling, and intensity transformations were applied only after the initial dataset splits into training, testing and validation, to avoid introducing augmented variants into multiple subsets and attempting to prevent train-test leakage. Some of the techniques employed, such as similarity image search using SimCLR, can be utilized to identify existing augmentations within a dataset, thereby preventing the application of redundant or conflicting custom augmentations.

3.1. Image Classification

PetroGAN

Synthetic dataset obtained from the training of 10000 images of petrographical thin sections, the original dataset available upon request and at the repository in the original publication (Ferreira et al., 2022), fine-tuned to sedimentary, metamorphic and volcanic generators, capable of generating realistic representation of a thin section in XPL and PPL, the model is also useful as the first step of training, to initialize further specialized models

(such as carbonate microfacies, ooids, heavy minerals). This dataset is displayed with its four classes in Figure 11. Since the data is synthetically generated using a GAN we didn't need to do a train and test split during the training and generation of this dataset.

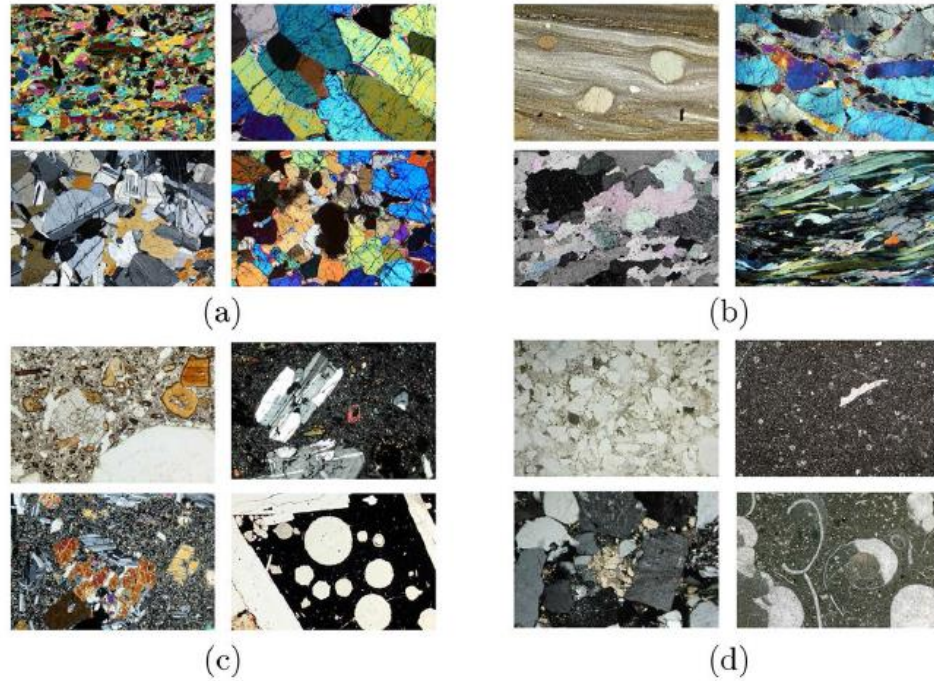


Figure 11. Example of PetroGAN training dataset, consisting of (a) Plutonic Rocks (b) Metamorphic rocks (c) Volcanic rocks and (d) Sedimentary rocks

LithoFacies

This dataset consists of 3558 core images classified into six classes of lithological facies, the dataset is class-balanced and divided into: Argillaceous Sandstone (662 images), Limestone (312 images), Laminated Sandstone (642 images), Massive Sandstone (675 images), Granite (661 images) and Siltstone (649 images). The dataset is at a varying degree of resolution. An example array of the images is displayed in Figure 12.

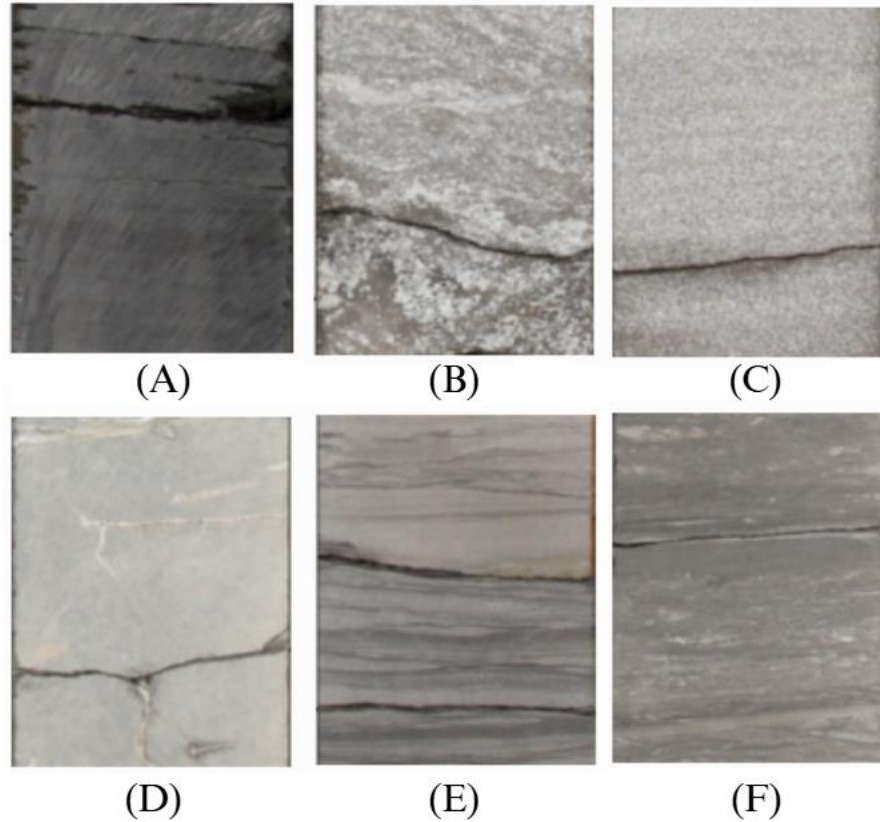


Figure 12. Example of the Lithofacies Dataset of Core Images (a) Argillaceous Sandstone (B) Granite (C) Sandstone Massive (D) Limestone (E) Sandstone Laminated (F) Siltstone.

3.2. Object Detection

Sandstone Components Dataset

This dataset comprises 1876 images, each labeled for object detection within the domain of sedimentary petrography, specifically focusing on sandstone components; with 9 classes having over 100 labeled objects (they being Quartz, Lithics, Feldspar and Plagioclase). The images are categorized into several distinct classes, each representing a different

component or mineral found in sandstone. The classes labelled are Quartz components (Monocrystalline, Polycrystalline), Feldspar components (Plagioclase, Potassium Feldspar) and Lithic Fragments (volcanic, metamorphic, carbonate, mudstones, sandstones, chert). The Images are at high resolution 4908x3264px. Collected from Xialong et al, 2020, using data from Yarlung Tsanpo in Tibet. The dataset is plotted with the corresponding bounding boxes and scale bar (Figure 13).

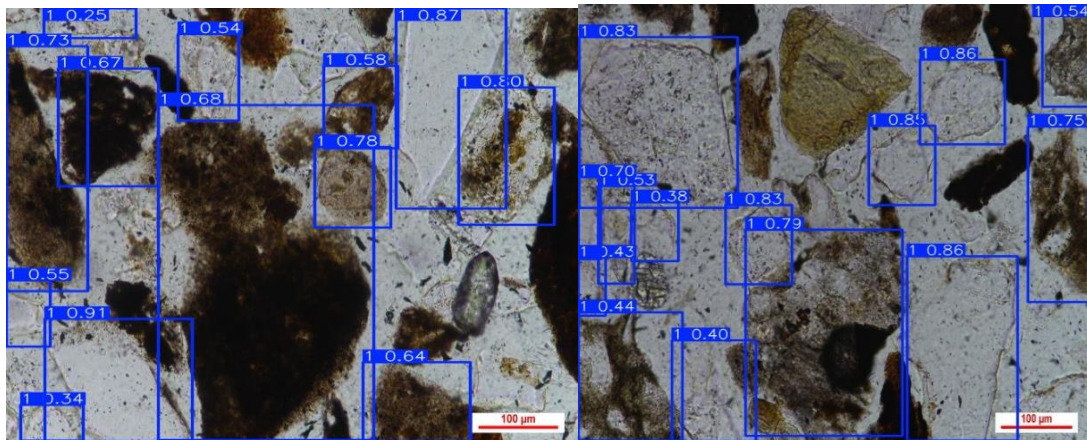


Figure 13. Example of the Sandstone component dataset, with the bounding box for each detected and labeled clast.

3.3. Semantic Segmentation

Carbonate Microfossils

Dataset consisting of over 8723 images of labeled carbonate thin section with corresponding microfossil; brachiopods, foraminifera, ostracod, algae, bivalve, echinoderms, gastropod, bioclasts, tubiphytes, calcimicrobe, ammonoid, radiolarian and

bryozoans. This labels for image classes are present for each image and a sample of the dataset is displayed in Figure 14.

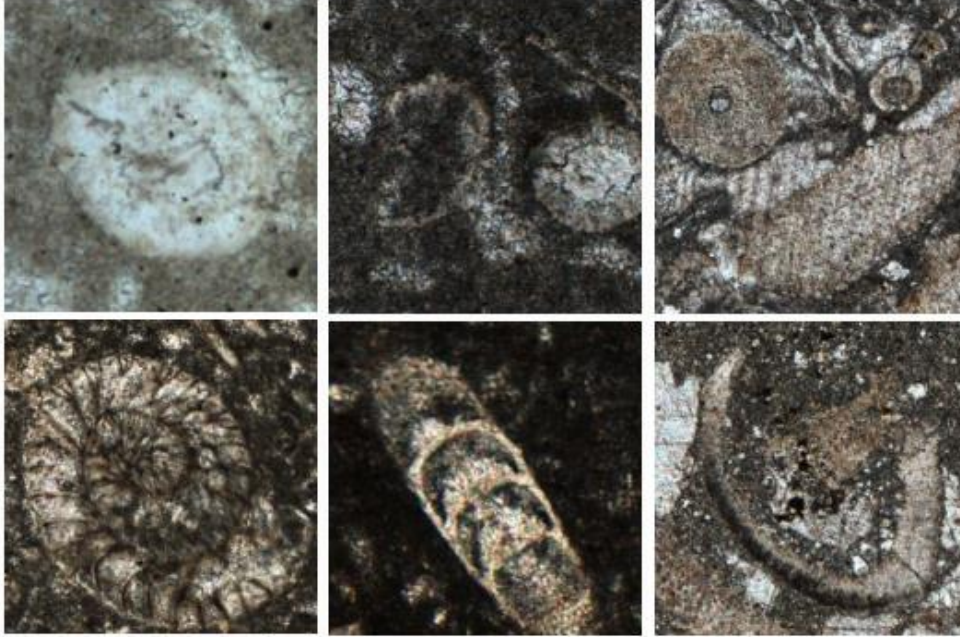


Figure 14. Example of Images from the Carbonate Microfossil Dataset

Ore Minerals

we utilized a dataset comprising reflected light petrographic images of ore minerals, specifically focusing on pyrite, from the Chuquicamata porphyry deposit in Chile. The original images were sourced from the "Microscopía y Microtermometría" Laboratory at the National University of Colombia. These images were captured using a MetaSystem automated thin-section scanning equipment, producing high-resolution scans with dimensions of approximately $23,000 \times 23,000$ pixels.

To prepare the data for analysis, we processed the high-resolution thin-section scans to create a semantic segmentation dataset suitable for training deep learning models. The images were cropped into smaller sections and downsampled to 256×256 pixels using Lanczos interpolation. This resizing technique preserves the essential features of the images while making them computationally manageable for model training. The images and the corresponding labels are displayed respectively in figure 15-A and Figure 15-N.

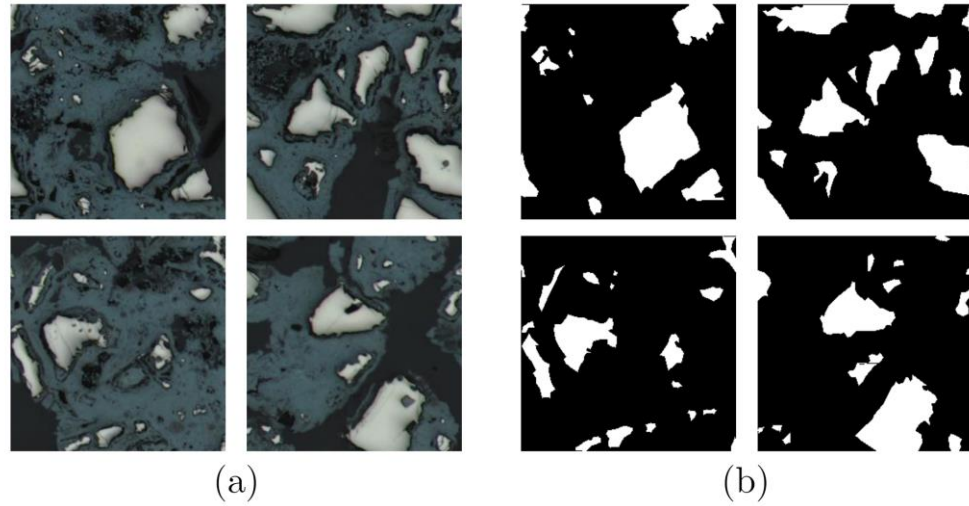


Figure 15. Sample of the dataset used from Chuquicamata (a) Original Images (b) Segmentation Masks

CHAPTER 4

METHODOLOGY

4.1. Self-Supervised Learning

SSL is a paradigm that leverages the volumes unlabeled data, images or text, for the pre-training of a model. This is predicated on the use of pretext tasks – which are specialized training objectives designed to enable a model to learn from data without having explicit annotations. These tasks are formulated so that the model learns to predict some parts of the data from other parts, this is also known as creating pseudolabels from the data, thereby extracting useful features and representations that can be beneficial for a wide range of downstream tasks in computer vision. Common pre-text tasks for computer vision include predicting the missing part in an image, solving jigsaw puzzles with the images, or colorizing black-and-white images (Jaiswal et al., 2020; X. Liu et al., 2023). When the model completes these tasks successfully, the model develops a structured latent space where similar data points are grouped together and can be then fine-tuned for specific downstream tasks, retuning the embeddings present in some pretraining SSL architectures for these tasks (Kirillov et al., 2023; Rani et al., 2023). These downstream tasks are the actual applications for which the model is intended, this can be image classification, object detection, or semantic segmentation in the context of geoscientific image analysis. This two-step process, involving pre-training on pretext tasks followed by fine-tuning of the model to a specific task.

4.2. SSL Architectures

The seminal and state-of-the-art SSL architectures tested are (1) Barlow Twins, an architecture focuses on learning embeddings that are invariant to input sample distortions by minimizing redundancy between the components of these vectors. It uses a cross-correlation matrix between the outputs of two networks fed with distorted versions of a sample (Zbontar et al., 2021). (2) Bootstrap Your Own Latent (BYOL) which uses a pair of neural networks (online and target) that learn from each other. It trains the online network to predict the target network representation of the same image but under a different view (Grill et al., 2020). (3) SimCLR, A simple framework for contrastive learning of visual representations, focusing on the composition of data augmentations, a learnable nonlinear transformation, and benefiting from larger batch sizes and extended training steps (Chen et al., 2020). (4) Momentum Contrast (MoCo): MoCo builds a dynamic dictionary for contrastive unsupervised learning, offering competitive results and excellent transferability to downstream tasks (He et al., 2019). (5) SwAV (Swapping Assignments between Views) an online algorithm that clusters data while maintaining consistency between cluster assignments of different image views, using a swapped prediction mechanism (Caron et al., 2020). (6) TiCo (Transformation Invariance and Covariance Contrast) which maximizes agreement among embeddings of distorted versions of the same image, avoiding trivial constant solutions and producing transformation invariant representations (Zhu et al., 2022).

4.3. Inner workings of an SSL Architecture, SimCLR

An illustrative example of a self-supervised learning (SSL) architecture is the **Simple Framework for Contrastive Learning of Visual Representations (SimCLR)** (Chen et al., 2020). SimCLR is designed to learn meaningful visual representations from unlabeled data by maximizing agreement between differently augmented views of the same image. This approach enhances the weights of a model before it is applied to specific computer vision tasks such as image classification, object detection, or segmentation.

The SimCLR framework operates through the following key components:

Data Augmentation Module:

The first step involves applying a series of stochastic data augmentation transformations to each input image to generate two correlated views, known as a positive pair. These transformations include:

- **Random Cropping and Resizing:** Cropping a random portion of the image and resizing it back to the original size.
- **Horizontal Flipping:** Flipping the image horizontally with a certain probability.
- **Color Jittering:** Randomly changing the brightness, contrast, saturation, and hue.

- **Color Dropping (Grayscale):** Converting the image to grayscale randomly.
- **Gaussian Blurring:** Applying Gaussian blur to the image with a specified probability.

The goal of these augmentations is to create diverse yet semantically consistent variations of the same image. By generating different views, the model is encouraged to learn invariant features that are robust to such transformations. The composition of these augmentations is crucial, as it defines the contrastive prediction task and significantly impacts the quality of the learned representations.

Base Encoder:

Each augmented image is then passed through a base encoder network to extract feature representations. Common choices for the encoder architecture include convolutional neural networks (CNNs) like **ResNet-50**, but **Vision Transformers (ViTs)** can also be employed. The base encoder processes each augmented image independently, producing a representation vector (also known as an embedding) $\mathbf{h}$ that captures high-level features of the input. These representations are intended to retain essential information needed for downstream tasks while being invariant to the applied augmentations (Chen et al., 2020).

Projection Head:

Following the base encoder, a projection head $g(\cdot)$ maps the extracted feature representations into a latent space where the contrastive loss is applied. The projection head is typically a multilayer perceptron (MLP) with one hidden layer and a nonlinear activation function (e.g., ReLU). It transforms the encoder's output h into a new representation $z=g(h)$:

$$\mathbf{z} = W^{(2)}\sigma(W^{(1)}\mathbf{h})$$

where $W^{(1)}$ and $W^{(2)}$ are weight matrices of the MLP, and σ is the activation function. The inclusion of the projection head helps in improving the quality of the learned representations by separating the information useful for the contrastive loss from that needed for downstream tasks. It allows the encoder to preserve more information in h , as the projection head can focus on aspects relevant for the contrastive learning objective (Chen et al., 2020).

Contrastive Loss Function:

The core of the SimCLR framework is its contrastive loss function, specifically the **Normalized Temperature-scaled Cross Entropy Loss (NT-Xent)**. This loss encourages the model to bring representations of positive pairs (augmented views of the same image) closer together in the latent space while pushing apart representations of negative pairs (augmented views of different images within the same batch) (Chen et al., 2020).

For a positive pair of examples i and j , the loss is defined as:

$$\ell_{i,j} = -\log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_j)/\tau)}{\sum_{k=1}^{2N} 1_{[k \neq i]} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_k)/\tau)}$$

where:

- z_i and z_j are the normalized projection vectors of the augmented images.
- $\text{sim}(u, v) = u^T v$ denotes the cosine similarity between two vectors.
- τ is a temperature parameter that scales the logits.
- N is the number of original images in the batch (resulting in $2N$ augmented examples).
- $1_{[k \neq i]}$ is an indicator function equal to 1 if $k \neq i$ and 0 otherwise.

This loss function ensures that the representations of augmented images originating from the same original image are close to each other in the latent space, while those from different images are far apart. By utilizing a large batch size, SimCLR effectively samples a significant number of negative examples without the need for a memory bank or specialized architectures.

As an example of the workflow followed by an SSL architecture, the Simple Framework for Contrastive Learning of Visual Representations (SimCLR, Figure 16) is a strategy designed to improve the weights of a model before it's applied to specific computer vision tasks like image classification or object detection (Chen et al., 2020).

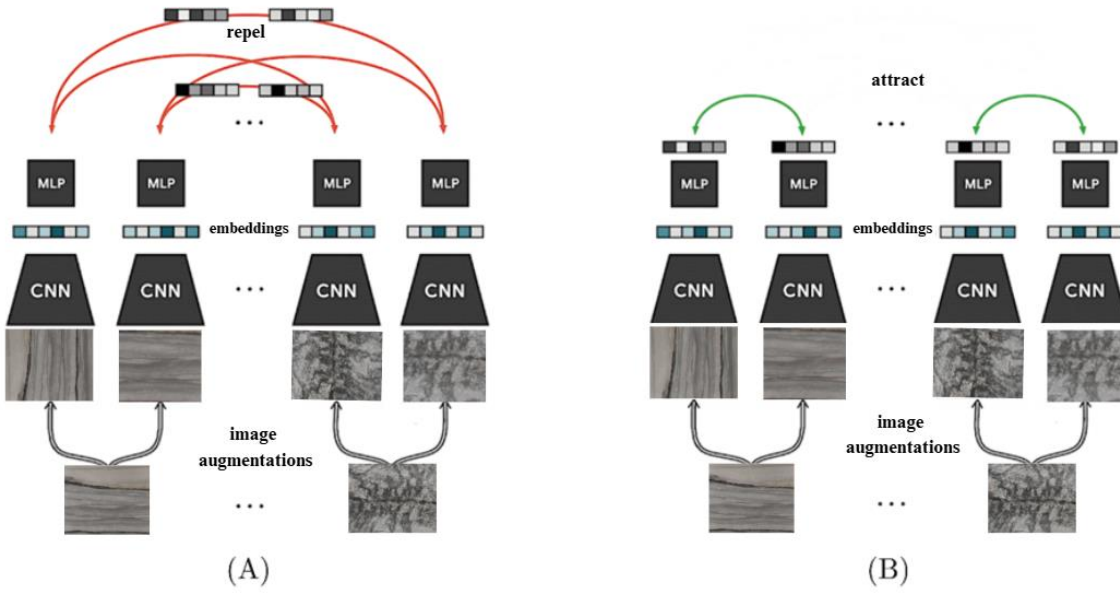


Figure 16. SimCLR from the original images, the augmentations and representations obtained and how using contrastive loss (A) different classes of images representations will repel and (B) same class images will attract. Modified from (Chen et al., 2020)

4.4. You Only Look Once (YoloV8)

For testing a workflow where SSL can be used we use a workflow using YoloV8 for object detection, YOLOv8 (You Only Look Once version 8) is the latest iteration in the YOLO family of real-time object detection algorithms. Building upon the advancements of previous versions like YOLOv5, YOLOv6, and YOLOv7, YOLOv8 aims to enhance both the speed and accuracy of object detection tasks. It introduces several innovations, including improved network architecture, attention mechanisms, dynamic convolution, and novel loss functions, to address challenges such as small object detection and complex backgrounds. (Redmon et al., 2015; Reis et al., 2023)

Network Architecture

The architecture of YOLOv8 is divided into two main parts:

a. Backbone Network

YOLOv8 employs an enhanced version of the **EfficientNet** architecture, specifically **EfficientNet-B4**, as its backbone. EfficientNet is known for its compound scaling method that balances network depth, width, and resolution to optimize both accuracy and efficiency.(Reis et al., 2023)

- **Input Processing:**

The input image is resized to 512×512 pixels to match the network's expected input size.

- **Feature Extraction:**

The backbone network extracts multi-scale feature maps (**P3, P4, P5, P6, P7**) that capture different levels of detail from the input image.

b. Detection Head

The detection head is responsible for processing the features extracted by the backbone network to produce bounding box predictions.

- **NAS-FPN (Neural Architecture Search - Feature Pyramid Network):**

YOLOv8 utilizes NAS-FPN to automatically generate an optimal feature pyramid network. NAS-FPN combines features from different levels using fusion operations (e.g., addition, multiplication, concatenation) and connections in both top-down and bottom-up pathways.

- **Bounding Box Predictions:**

The detection head outputs bounding box coordinates, objectness scores, and class probabilities. It predicts bounding boxes at different scales to handle objects of varying sizes.

Attention Mechanism

YOLOv8 incorporates attention modules to enhance focus on important features, utilizing a self-attention module to compute the similarity between features across spatial regions, emphasizing relevant spatial information, and a channel attention module that assigns importance weights to different feature channels, prioritizing significant features. It also employs dynamic convolution, adapting convolutional filters based on input through a conditional convolution layer that predicts convolutional weights conditioned on the input image, and a dynamic routing layer that dynamically selects appropriate filters for each feature map, improving adaptability to diverse and complex scenes. Additionally, YOLOv8 introduces a new loss function, Focal Loss, which addresses class imbalance by concentrating on hard examples while down-weighting easy negatives, enhancing training effectiveness. YOLOv8 also introduces a new loss function to improve training, which they term Focal Loss, its goal being to address class imbalance by focusing on hard examples and down-weighting easy negatives and it is defined as:

$$\text{Focal Loss} = -\alpha(1 - p_t)^\gamma \log(p_t)$$

Where:

- p_t is the model's estimated probability for the true class.

- α balances the importance of positive/negative examples.
- γ controls the focusing parameter.

4.5. DINOv2: Self Supervised Transformed, focused on getting Robust Visual Features without Supervision

1. Data Processing and Curation

For the training of DinoV2, the authors did a process involved developing a curated dataset, LVD-142M, comprising 142 million images to ensure diversity and quality. This dataset was derived from both curated sources, such as ImageNet-22k, and uncurated web data. Advanced techniques, including PCA-based deduplication, were employed to remove redundant images, ensuring uniqueness across training and evaluation splits. Additionally, a robust retrieval pipeline was implemented, utilizing sample- and cluster-based retrieval methods powered by self-supervised embeddings to augment the curated datasets with relevant samples from uncurated sources (Oquab et al., 2023).

2. Model Architecture and Objectives

The training leveraged Vision Transformers (ViT) models, ranging from ViT-Small to ViT-Gigantic, with configurations optimized for both image- and patch-level feature extraction. Multi-objective losses were employed to enhance the training process, including an image-level loss (DINO), which used cross-entropy loss based on softmax-normalized prototype scores in a teacher-student network, leveraging diverse views of the same image. For patch-level features, the iBOT objective applied a masked prediction loss to patch embeddings, promoting dense and robust feature representation. Regularization

techniques included the KoLeo regularizer, which enhanced feature diversity within batches through a differential entropy estimator, and Sinkhorn-Knopp centering, which improved stability in feature distributions through effective normalization.

3. Training Optimization

The training process incorporated resolution adaptation, starting with low resolutions (224×224) and progressively scaling up to 518×518 in the final phase to optimize performance on pixel-level tasks such as segmentation. Efficient training techniques included FlashAttention, a custom implementation designed to reduce memory overhead in transformer attention mechanisms, and sequence packing, which consolidated sequences to improve processing efficiency while preserving attention integrity. Additionally, stochastic depth was employed to skip computation for a fraction of residual connections, saving resources. Fully Sharded Data Parallelism (FSDP) was utilized to distribute memory consumption across GPUs, enabling the training of large models with reduced communication costs.

4.6. STEGO: Self-supervised Transformer, Energy-Based Graph Optimization

The Self-supervised Transformer with Energy-based Graph Optimization (STEGO) framework is designed to tackle the task of unsupervised semantic segmentation by leveraging high-level features that correspond to discrete semantic features often labeled by humans. Unlike conventional approaches that rely on end-to-end frameworks for feature learning and cluster compactification, STEGO uniquely separates these stages, Figure 17. This separation is based on the observation that current unsupervised feature learning

frameworks already produce dense features with semantically consistent correlations (Hamilton et al., 2022).

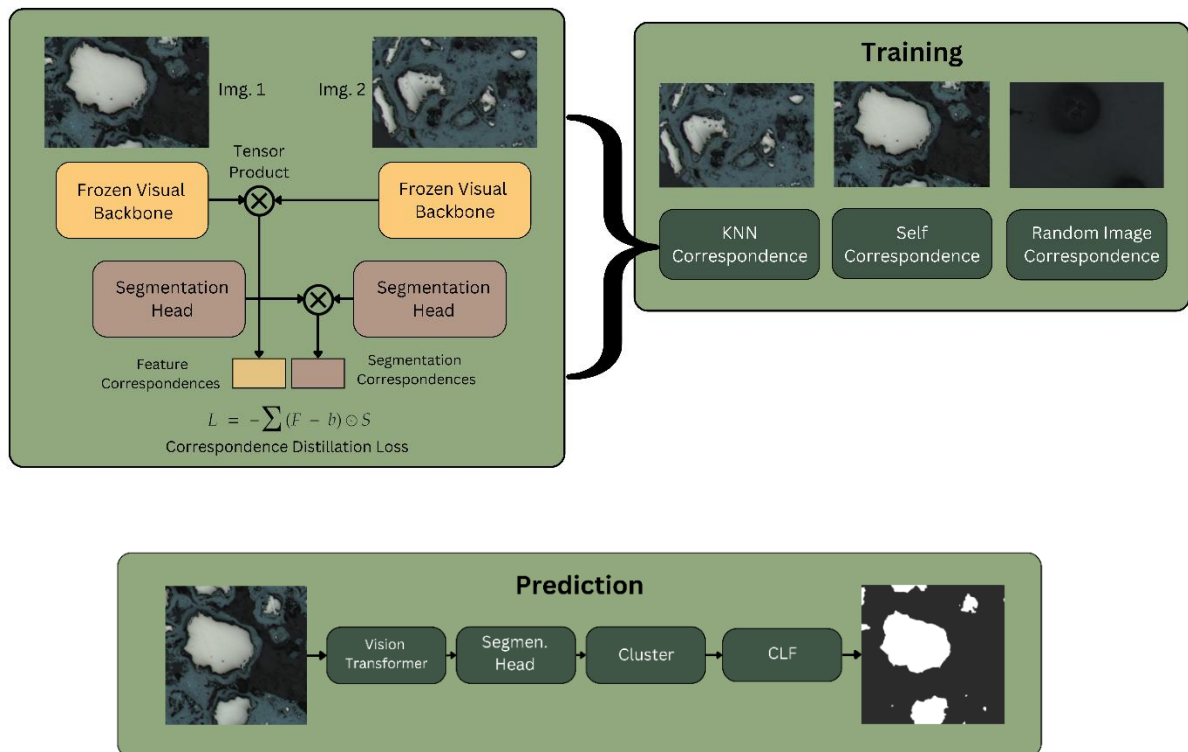


Figure 17. Simplified Schematic of the STEGO Architecture, CRF: Conditional Random Field, KNN: K-Nearest Neighbors. Modified from (Hamilton et al., 2022)

The architecture aims at grouping corresponding features both within an image and across a dataset, with these correspondences being highly likely to match a single class or object label. At the core of STEGO is a novel contrastive loss function designed to encourage feature formation into compact clusters while preserving their semantic relationships across the image corpus. For parameter tuning and optimization, STEGO utilizes a combination of self, K-Nearest Neighbors (KNN), and random image correspondences to train the segmentation head. This approach balances positive and negative forces, ensuring compact cluster formation and feature distillation. This mechanism dramatically improves

segmentation performance, setting a new state-of-the-art on standard benchmarks. The images are featurized using a Frozen Visual Transformer, an architecture that recently has been proven to provided matching results with comparison to CNNs (Dosovitskiy et al., 2020; Koeshidayatullah et al., 2023), these visual features are used further to extract segmentation correspondences, this is done by comparing images in the dataset, the segmentation correspondences are then clustered and refined using a Conditional Random Field (CRF), a simplified schematic of the architecture, modified from (Hamilton et al., 2022) is presented in Figure 17. The method requires minimal parameter tuning from the user, being the declaration of the number of classes the minimal requirement for the use of this algorithm, the original implementation being tested on the Microsoft COCO dataset.

For the evaluation of STEGO, we adopted a two-pronged approach to assess the model's performance. Firstly, we conducted a visual comparison of the segmentation maps obtained. Secondly, we evaluated the degree of overlap between the ground truth, as labelled by a human expert, and the predictions derived from the STEGO trained model. This overlap is measured through the widely-used metric called Intersection over Union (IoU), with higher values indicating better performance. A perfect match would result in an IoU score of 1, while complete non-overlap would yield a score of 0.

4.7. Computing Requirements

Our workflow begins with acquiring a substantial dataset—both labeled and unlabeled—which is then pre-processed to suit the specific requirements of the various SSL architectures and the tasks at hand. Training these models necessitates significant

computational resources due to the intensive demands of self-supervised learning and generative algorithms. To meet these requirements, we utilize a combination of local and cloud-based computing environments. Primarily, we operate on a Unix system equipped with four GPUs, each with 24 GB of VRAM, 256 GB of RAM, and a 48-core processor. This high-performance setup allows us to handle the substantial computations and memory requirements associated with training complex models. Additionally, we have access to supplementary machines with 16 GB of VRAM to further augment our computational capacity.

For tasks that benefit from scalable cloud resources, we leverage platforms like Google Colaboratory (Colab). The default Colab environment provides a NVIDIA Tesla K80 GPU with 12 GB of VRAM and an Intel Xeon CPU with 2 virtual CPUs (vCPUs) and 13 GB of RAM. However, Colab offers the flexibility to upgrade to virtual machines with significantly higher specifications—up to 96 vCPUs and 624 GB of RAM—for projects requiring more extensive computational power.

This adaptability allows us to tailor our computing resources to the specific needs of different stages in our workflow. Throughout the model development process, we engage in iterative training and performance monitoring, conducting hyperparameter optimization for factors such as model depth and the choice of loss functions. This iterative approach enables us to fine-tune our models effectively, ensuring optimal performance on the targeted geoscientific tasks. By combining robust local hardware with scalable cloud computing options, we can efficiently manage the computational demands of our research while maintaining flexibility and cost-effectiveness.

CHAPTER 5

RESULTS

5.1. Image Classification on LithoFacies using SSL

Through pretraining SimCLR using the LithoFacies dataset and removing the class labels, we are able to finetune the resulting model using only 20% of the labels for training the SSL model and achieve enough accuracy as to match with a validating with 80% of the data. The SSL methods is used in a simulated data scarcity scenario decimating the amount of dataset required for a downstream image classification task using a CNN architecture, Figure 18.

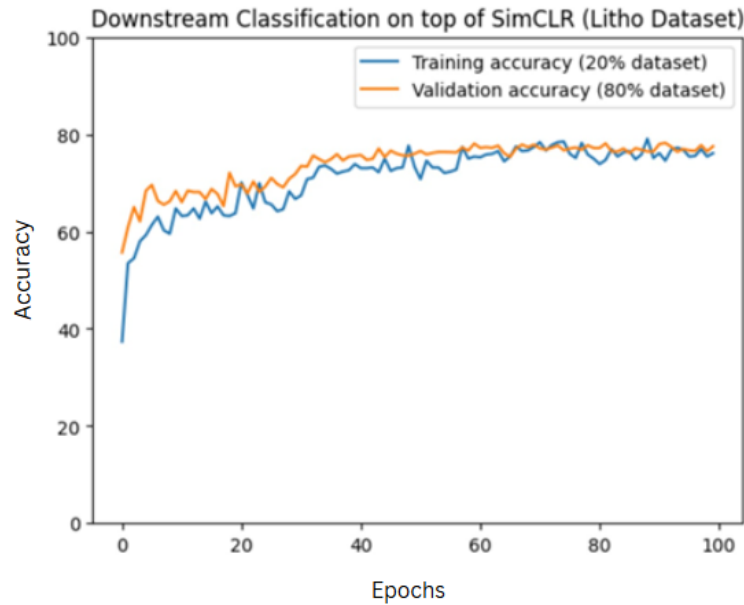


Figure 18. Downstream accuracy of training an Image Classifier using only 20% of the Original Dataset and Validating on the remaining 80%.

5.2. Image Similarity using SSL on Carbonate Thin Sections

During the training of SSL algorithms, SimCLR⁴, the embeddings, or high-level structures of the image data are produced, these image embeddings can be used to automatically cluster similar images in the data, in the case of the carbonate thin section dataset we are able to cluster similar structures that in turn can correspond to same microfossils, Figure 19, this can be further used as a first step to discriminate images containing certain structures, detect data augmentations in the dataset or distinguish between XPL and PPL images in the thin sections helping as a first step before an image classification task.

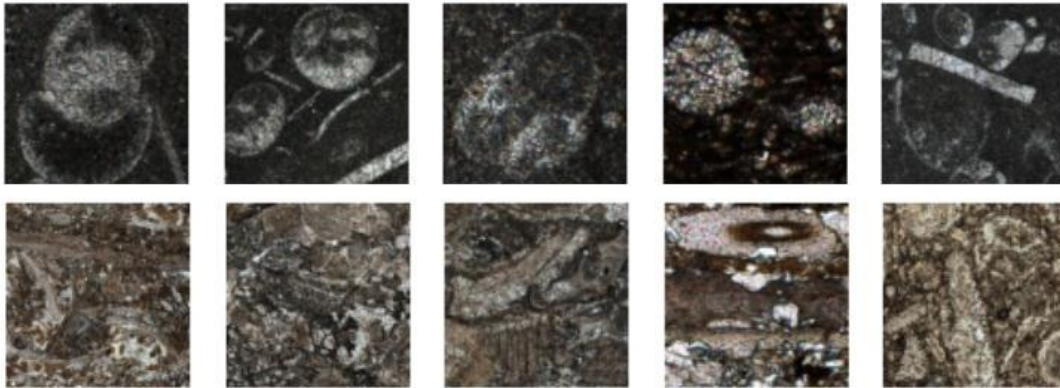


Figure 19. Clustered Image embeddings of Microfossils using SimCLR

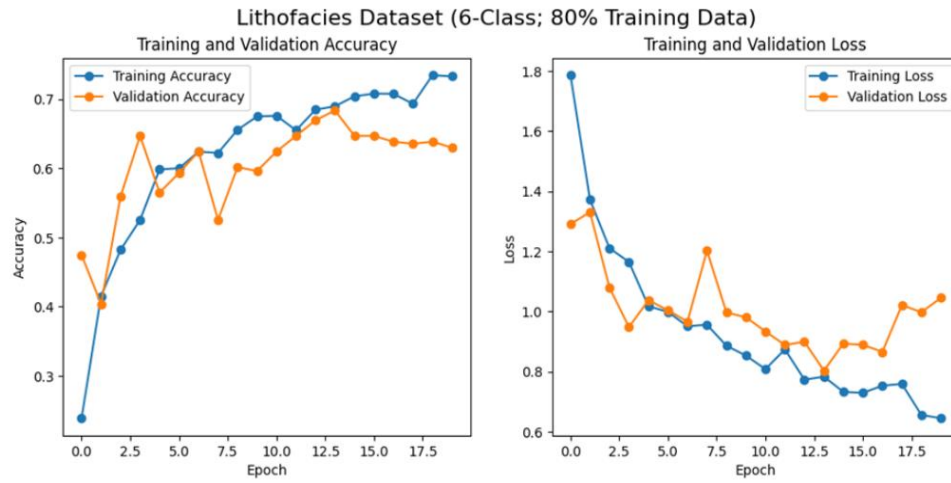
5.3. Self-Supervised Image Classification Using DinoV2.

The results of the image classification task using DINOv2 and conventional methodologies are summarized in Table 1 together with the image resolution and the number of images used.

Dataset	Architecture	% Dataset Used	Test Accuracy (%)	Image Resolution	# Images
Volcanic Ash Particles	DINOv2 + SVM	20%	91.8	(50, 50, 1)	194
Lithofacies	DINOv2 + SVM	20%	83.1	(224, 224, 3)	710
Volcanic Ash Particles	DINOv2 + SVM	80%	94.9	(50, 50, 1)	781
Lithofacies	DINOv2 + SVM	80%	89.4	(224, 224, 3)	2876
Volcanic Ash Particles	CNN	80%	86.9	(50, 50, 1)	781
Lithofacies	CNN	80%	65.3	(224, 224, 3)	2876

Table 2. Results for Image classification using both SSL and a custom CNN. The datasets were taken from (Shoji et al., 2018) for the volcanic ash particles and from (Baraboshkin et al., 2020; Koeshidayatullah, Al-Azani, et al., 2022) for the lithofacies dataset.

Training for the SVM (Cortes et al., 1995) was done after getting the embeddings from DinoV2 to get the final accuracy on the test dataset for each case scenario taken into account. The results of training both datasets under a custom CNN without transfer learning is shown in Figure 20 and Figure 21 for each of the two datasets used.



DinoV2 + SVM (20% Dataset)
710 Images (train) 2845 Images (test)
Accuracy: 0.83

Custom CNN (80% Dataset)
2845 Images (train) 710 Images (test)
Accuracy: 0.65

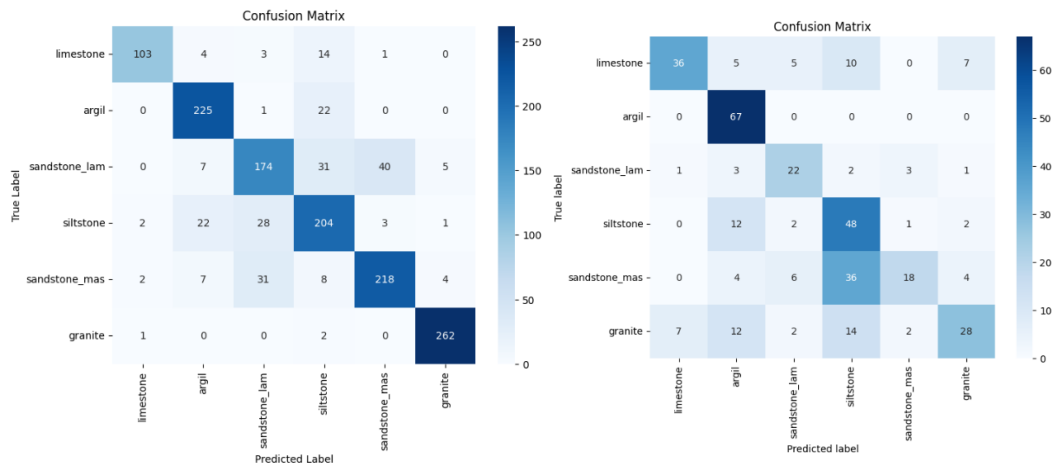


Figure 20. Accuracy and Loss for the 6-Class classification using the CNN of the Lithofacies dataset. And comparison of Confusion Matrix for both the DinoV2 + SVM (80% and 20% of the dataset used) training and the Custom CNN (80% of the dataset used) when calculated on the test set (Baraboshkin et al., 2020; Koeshidayatullah, Al-Azani, et al., 2022)

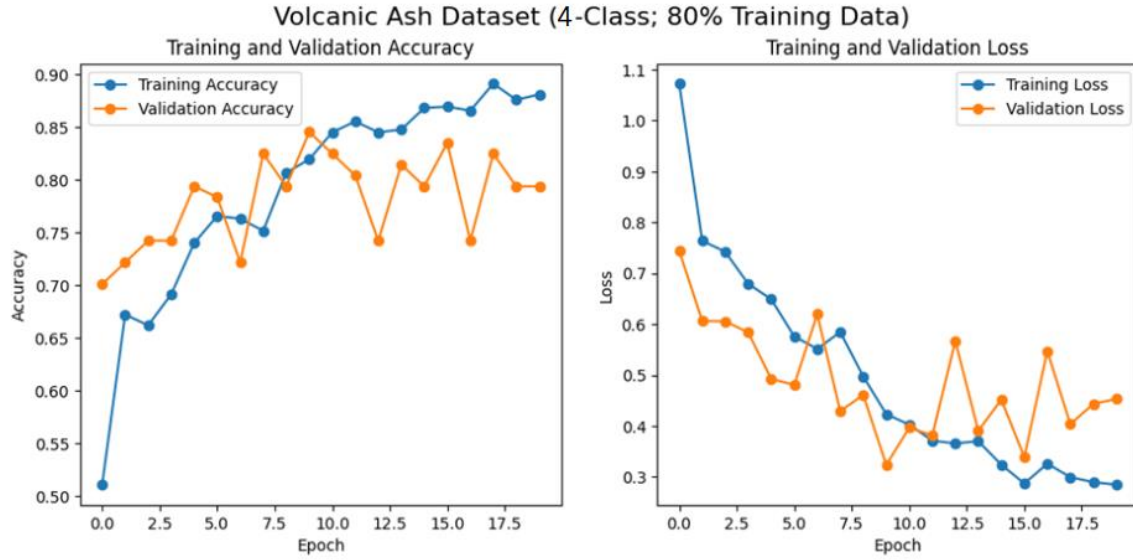


Figure 21. Accuracy and Loss for the 4-Class classification of the Volcanic Ash dataset (Shoji et al., 2018)

5.4. Few-Shot Segmentation using pretrained GANs

Through using the weights gained during the training of a GAN for generating Foraminifera, a further FewShot CNN (with less than 5 labeled examples, as shown in Figure 3) can be used to generate a labeled dataset through self-labeling a dataset, a technique known as autoshot, we use this generated dataset of 1000 images to serve as the training data for a semantic segmentation model, in this case we test SegFormer, Figure 22 (Xie et al., 2021), using transfer learning, allowing for a resulting One-Epoch and Few-Shot segmentation of foraminifera species, we test this data on human labeled examples at the end of the training, showcasing an Intersection over Union (IoU) matching the train

and validation IoU with synthetic labels, the resulting auto-shot dataset generated and the predictions done by SegFormer on real data is presented in Figure 23.

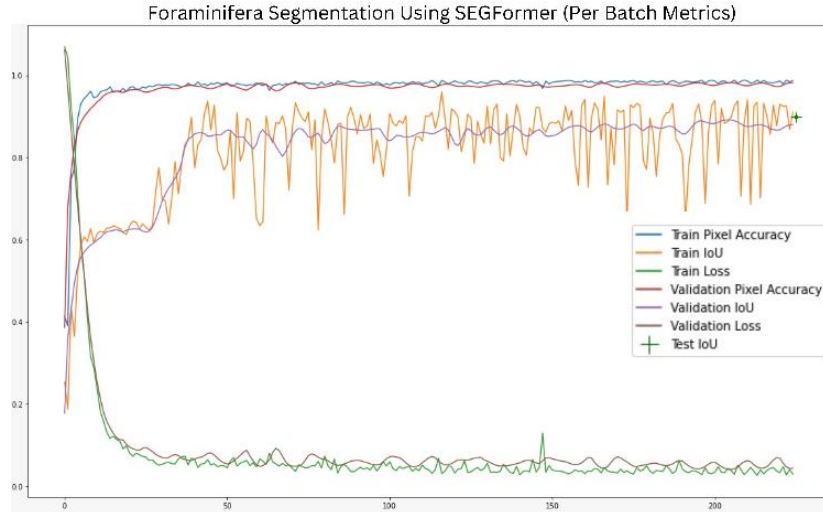


Figure 22. Results of Accuracy and mIoU of SegFormer applied to micropaleontological image segmentation, using a one-epoch and five-shot approach.

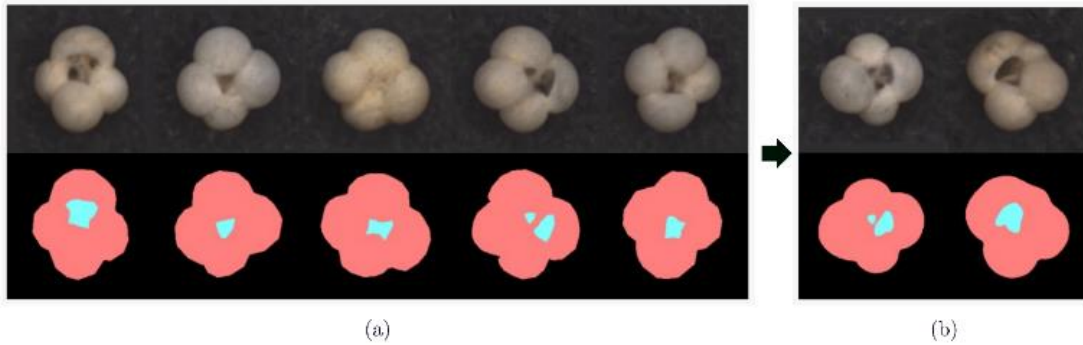


Figure 23. Example of a five-shot labels of synthetically generated foraminifera parts as input to the model (b) resulting segmentations made by the model

5.5. Object Detection and SSL Image Classification

The object classification task on clasts within the sandstone dataset was performed using a supervised learning approach. The performance metrics are summarized in Table 1 and discussed below.

Metric	Value
Total Images	94
Total Instances	1,275
Precision (P)	0.800
Recall (R)	0.798
mAP@50	0.864
mAP@50-95	0.668

Table 3. Object Detection Performance Metrics for the Sandstone clast dataset.

Precision and Recall: The model achieved a precision of 0.800 and a recall of 0.798.

These metrics indicate that the model effectively identified clasts within the images, with a minimal rate of false positives and false negatives, respectively. The near-equivalence of precision and recall suggests a balanced performance across the dataset.

1. Mean Average Precision (mAP):

- The mAP@50 score of 0.864 demonstrates that the model is highly capable of detecting objects with a bounding box intersection over union (IoU) threshold of 0.5.
- The mAP@50-95 score of 0.668, which averages precision across multiple IoU thresholds (from 0.5 to 0.95), indicates that the model maintains good performance even at stricter IoU thresholds. This result highlights the model's robustness in accurately localizing clasts in the sandstone images.

Interpretation

These results confirm that the object detection model is suitable for classifying and localizing clasts in sandstone datasets. The high mAP@50 score reflects the model's effectiveness for practical applications, while the moderate mAP@50-95 highlights areas for potential improvement in handling challenging localization scenarios. Future work could explore refining the model architecture, incorporating data augmentation, or utilizing self-supervised learning to further enhance performance on this task.

This performance forms a solid baseline for subsequent studies on automated clast analysis in sedimentology, an array of the predicted valeted images shown in Figure 24 and training metrics in Figure 25.

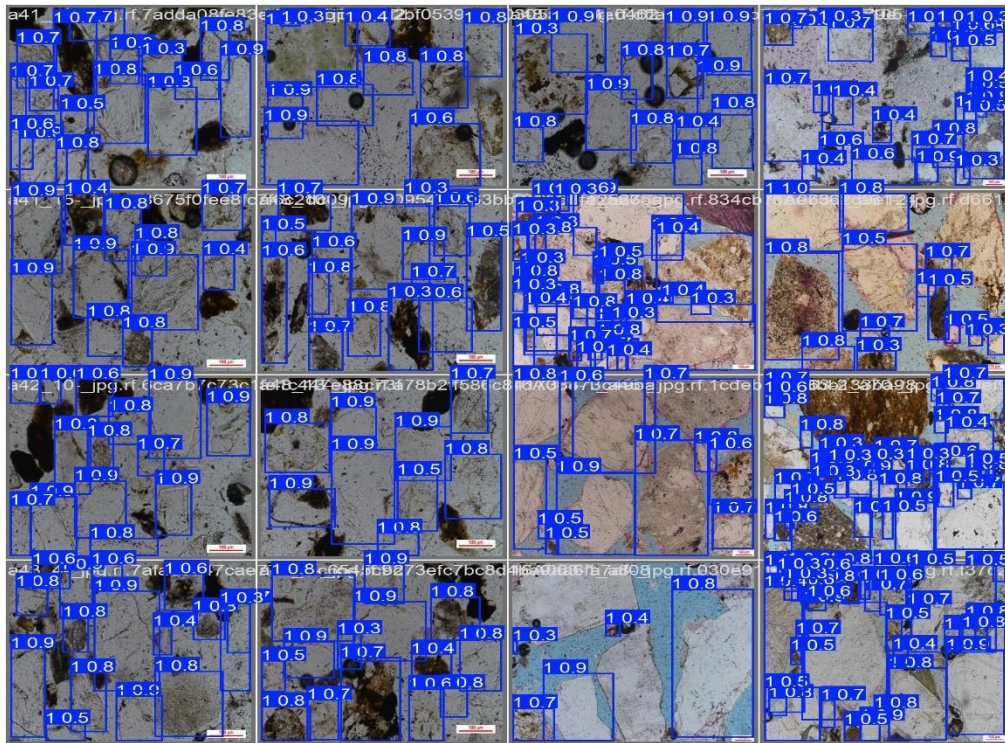


Figure 24. Results of Object detection of classes in the Ordos Sandstone Dataset.

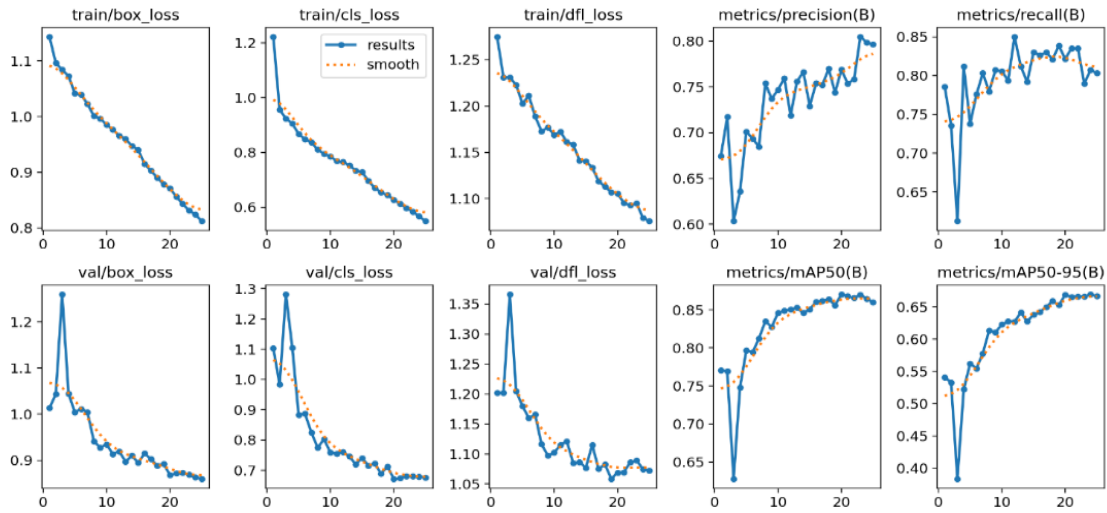


Figure 25. Loss, Precision and Recall graphs during the training of the YolV8 model

5.6. Semantic Segmentation Using Self-Supervised Learning

The implementation of Vision Transformers via the SegFormer architecture has demonstrated significant promise, the loss being provide as the training metric in Figure 26, given that this was a fully selfsupervised training and the metrics were calculated with a validation set, manually labeld post-training achieving a notable mean Intersection over Union (mIoU) of 83%; the training loss being presented as the main metric in this self-supervised learning. This achievement underscores the model's adeptness at performing precise semantic segmentation within petrography, particularly in distinguishing ore minerals from their matrix. The efficacy of SegFormer, grounded in its strategic processing of petrographic images into smaller patches and the sophisticated encoder-decoder structure, has been instrumental. This structure not only facilitates the nuanced analysis of mineral content but also ensures the generation of highly accurate segmentation masks.

Despite the overall success, visual inspections of the segmentation maps reveal areas where the model's adherence to the training labels restricts its capacity to perfectly outline mineral boundaries, highlighting the model's limitations and areas for refinement, Figure 27.

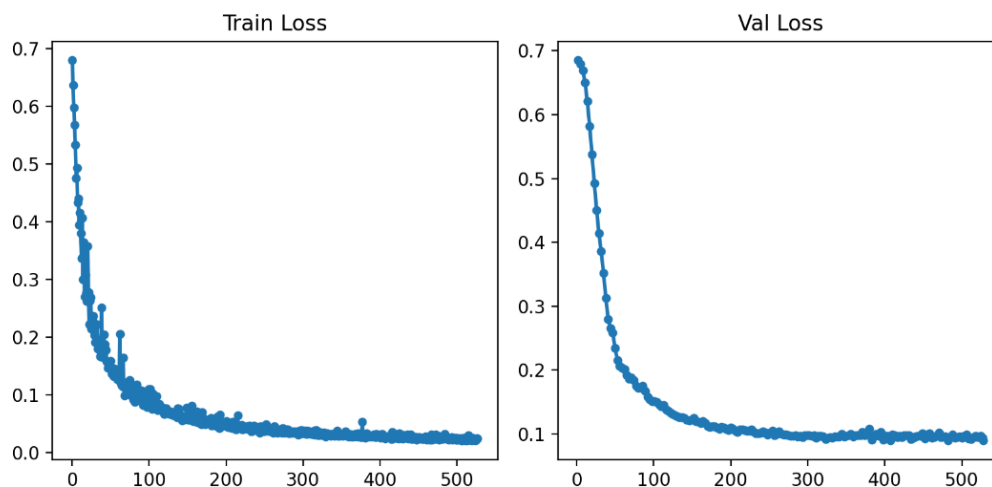


Figure 26. Training graphs from Segformer trained on the dataset with a mIoU reaching 83%

5.7. Self-supervised Transformer

A visual inspection of the segmentation maps in Figure 27 highlights the model's ability to accurately identify and delineate the regions of interest, however we see more accuracy, with respect to mIoU and after visual inspections we observe the trained model being supervised is constrained by the labels given, resulting in some parts of the image not following the correct outline of the mineral.

To quantify the model's performance, we computed the mean IoU across the 42 test images evaluated. In our experiments, the STEGO model achieved a mean IoU score of 0.8, a comparison with traditional methods and the ground truth being presented in Figure 27,

indicating a strong correlation between the ground truth masks and the model's predictions, demonstrating the model's effectiveness in performing semantic segmentation task within the context of geology and machine learning.

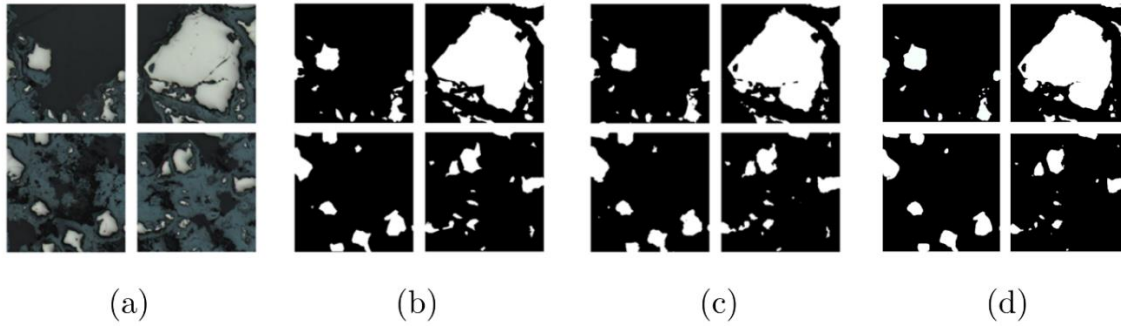


Figure 27. Sample of (a) Original images in the dataset, (b) Ground truth pyrite masks (c) predictions made with the SSL trained model (d) SegFormer Predictions.

CHAPTER 6

DISCUSSIONS AND FURTHER IMPROVEMENTS

SSL, by eliminating the need for extensive labeled datasets, accelerates the model training process, making it particularly advantageous for projects with limited resources for data annotation. The future direction of this research includes exploring ways to further reduce the reliance on labeled samples and assessing the impact of dataset size and pre-training in more complex application scenarios. For this it is necessary to constrain what pretext tasks are most useful according to each geological scenario, whether it involves seismic analysis, petrography, or remote sensing, presents unique challenges and data characteristics. Identifying and designing SSL tasks that align closely with the intrinsic properties of these specific domains will enhance the effectiveness and relevance of the learning process. For instance, tasks that emphasize spatial relationships and textural patterns might be more beneficial in petrographic image analysis, whereas those focusing on temporal coherence could better serve seismic data interpretation. Moreover, assessing the impact of dataset size and diversity in SSL is essential. The hypothesis is that SSL can maintain, or even surpass, the performance of supervised learning with significantly fewer labeled examples. However, this needs to be rigorously tested across a variety of dataset sizes and compositions, this work has already been tested with one SSL architecture in the recent literature (Mamode et al., 2025). It's crucial to understand how SSL models scale with the amount of available data and how they cope with diverse and complex geological imagery.

Another promising direction is the integration of SSL with other machine learning paradigms, such as transfer learning and few-shot learning. By leveraging pre-trained models on large, diverse datasets, we can potentially achieve a high-performance level with minimal fine-tuning on specific geological tasks. This hybrid approach could effectively combine the strengths of various learning methodologies, leading to more robust and versatile models. In the context of geological applications, where data labeling is often labor-intensive and requires expert knowledge, SSL and Few-Shot approaches can play a transformative role. By efficiently utilizing unlabeled data, with the potential to democratize access to recent analytical techniques in the geosciences, making them more accessible to areas in the geosciences with limited access to labeled and curated data.

Exploring and adapting these innovations to the specific needs of geoscientific image analysis will be key to unlocking the full potential of SSL in this field, the future applications of SSL in geosciences hinges on developing more sophisticated and tailored pretext tasks, understanding the dynamics of dataset size and composition, integrating SSL with other learning paradigms, and keeping pace with broader advancements in machine learning. With these focused efforts, SSL is poised to significantly advance the field of geoscientific image analysis, offering new insights and efficiencies in understanding our planet's complex geological processes.

In this study, we introduce the innovative application of Self-Supervised Learning (SSL) to semantic segmentation tasks within petrography, leveraging a dataset specific to this field. Addressing the challenge of developing reliable Machine Learning models in geosciences without relying on extensive datasets (Koeshidayatullah et al., 2020; Koh et al., 2021; Dawson et al., 2023), this work underscores the viability of an architecture

trained without the necessity for human-annotated data for effective mineral segmentation. An approach that currently is demonstrating significant promise for more complex segmentation tasks within computer vision (Oquab et al, 2023) and is yet to be tested on petrographic image analysis workflows, such as ore vein segmentation, offering a substantial improvement over simpler methods like binarization, as evidenced in Figure 28. The outcomes of our study reveal a remarkable consistency in the model's predictions when compared to human labeling, which typically involves broader, less precise delineations of minerals. These findings show that the model's predictions closely match the actual contours of the minerals, capturing intricate details such as small cracks and holes in the pyrite samples, as depicted in Figure 29. This level of detail and accuracy in segmentation aligns with the benchmarks set by the original dataset, as facilitated by the STEGO architecture (Hamilton et al., 2022), thus illustrating the potential for SSL in advancing the field of petrographic analysis.

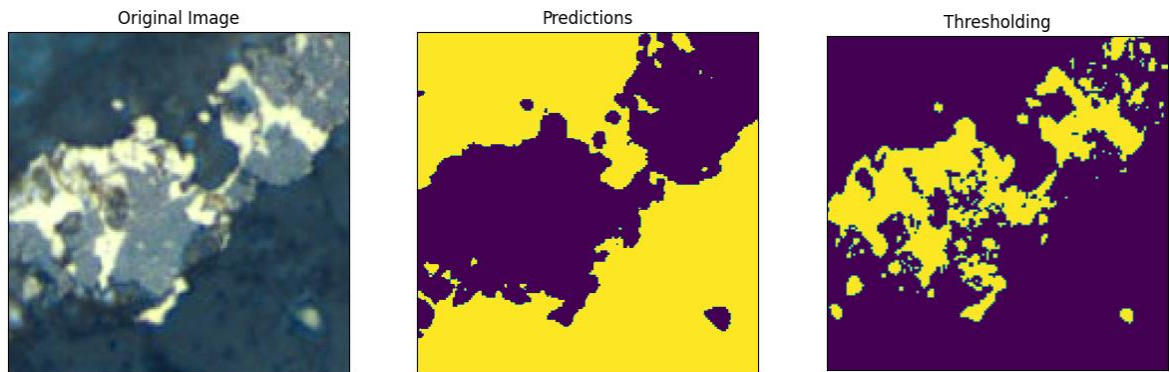


Figure 28. Comparison of STEGO for ore vein segmentation and comparison with a Binary thresholding operation.

This finding suggests a promising avenue for training semantic segmentation models without the necessity of annotating entire datasets, some other approaches such as data augmentation of existing labels and label generation are also two possible approaches for training this kind of models without extensive labeled datasets (Ferreira and Koeshidayatullah, 2023; Dawson et al, 2023). This approach could be particularly beneficial for leveraging vast repositories of unlabeled thin section images, for which comprehensive manual annotation might be challenging or impractical. However, further research is required to validate these results, as the metrics achieved are supported by the high contrast between ore minerals in reflected light. A comparable training approach with a transmitted light dataset may necessitate a larger dataset to achieve similar outcomes. SSL demonstrates significant potential in not only surpassing human annotation but also in assisting the segmentation of minerals of interest, a critical aspect for obtaining modal mineralogy in thin sections, a topic of interest in machine learning applied to ore mineralogy (Koh et al, 2021). In this work, we have successfully employed the STEGO architecture to cluster ore minerals, specifically focusing on a pyrite segmentation task using a dataset from the Chuquicamata porphyry deposit. The model's performance is evaluated using mean Intersection over Union (mIoU) on 42 human-labeled images, yielding a score of 0.8 mIoU. This evaluation further highlights the model's capability to accurately segment minerals that were not labeled during the annotation process, Figure 29. While this study focused on pyrite segmentation, the underlying principles of the STEGO architecture and similar approaches hold promise for classifying a wider array of ore minerals. Future work will explore its adaptability and potential for comprehensive ore mineral segmentation. This result not only validates the robustness of the STEGO model

but also highlights its potential for practical applications in geological settings. Such applications may include automated rock classification, mineral exploration, and resource estimation among other tasks

In conclusion, the application of SSL for semantic segmentation in ore mineral identification marks the first step towards mitigating the time-consuming and error-prone labeling processes associated with traditional methods, a problem in labeling that is highlighted in some supervised machine learning applications in the geosciences (de Lima et al, 2020). The potential of SSL has been previously demonstrated in other geological applications, such as denoising and interpolation of seismic data (Want et al, 2024; Fang et al 2023), but its full potential in petrography and other visual tasks within geosciences remains largely unexplored. This study sets the stage for further investigations into the capabilities of SSL in these areas, with the ultimate goal of developing more efficient and accurate techniques for mineral identification and segmentation in petrographic datasets.

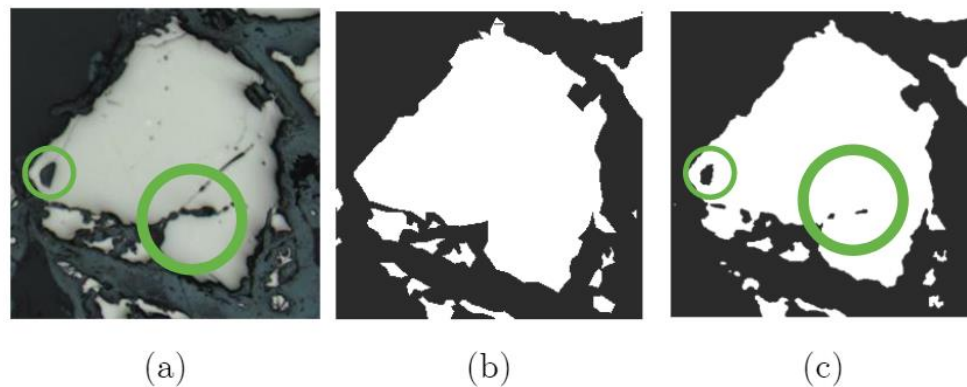


Figure 29. Improvements of predictions by an SSL model over human labels (a) original image (b) ground truth mask (c) model prediction.

CONCLUSIONS

- Self-Supervised Learning (SSL) Potential: This study demonstrates that SSL methods significantly reduce dependency on labeled datasets, addressing a critical bottleneck in geoscientific applications, including petrography, micropaleontology, and seismic image analysis.
- Competitive Performance: SSL-based models such as STEGO, DinoV2, and SimCLR achieved competitive performance metrics, with segmentation and classification accuracies surpassing traditional supervised methods, as evidenced by metrics like IoU (e.g., 0.8 for pyrite segmentation) and classification accuracy (e.g., 94.9% for volcanic ash particles).
- Efficiency in Data Utilization: By leveraging unlabeled data and few-shot approaches, this study demonstrates the capability of SSL techniques to perform effectively even in scenarios with limited annotated data, making it feasible to analyze under-resourced geological datasets. The integration of SSL methods, especially for semantic segmentation, highlights the potential to automate and significantly reduce the labor-intensive and error-prone process of manual labeling in geosciences.

- The integration of GAN-based synthetic data generation, Generative based SSL, provided enhanced segmentation capabilities, particularly for few-shot learning scenarios, proving effective for micropaleontological datasets.
- This work establishes a foundation for the integration of SSL with advanced machine learning paradigms like transfer learning and few-shot learning, providing a roadmap for future geoscientific research.
- While SSL showed great promise, its performance was influenced by dataset-specific factors like contrast and image quality. Future work should explore its adaptability to low-contrast datasets and evaluate its scalability for broader geological applications.

APPENDIX

Glossary of Terms

Self-Supervised Learning (SSL): A machine learning paradigm that enables the extraction of pertinent features from unlabeled data, reducing the dependence on extensively annotated datasets.

Petrography: The study and classification of rocks, especially through microscopic examination.

Generative Adversarial Networks (GANs): A machine learning framework used for unsupervised image generation, consisting of two networks, a generator and a discriminator, that compete against each other.

StyleGAN2: An advanced Generative Adversarial Network known for generating high-resolution and realistic images.

Fréchet Inception Distance (FID): A metric to assess the quality of generated images by comparing them to real images.

Convolutional Neural Networks (CNNs): Deep learning architectures used for image classification and segmentation, known for their capability to learn spatial hierarchies of features.

Vision Transformers (ViTs): Deep learning models that utilize attention mechanisms for tasks like image analysis.

Semantic Segmentation: A process in computer vision to classify each pixel of an image into a predefined category.

Foundational Models: General-purpose models in computer vision that can be applied across a wide range of applications.

PetroGAN: A generative model for creating synthetic petrographic datasets.

Few-Shot Learning: A machine learning approach where models are trained to make predictions with a limited amount of labeled data.

SimCLR: A framework for contrastive learning of visual representations.

Momentum Contrast (MoCo): A method for unsupervised visual representation learning that builds a dynamic dictionary for contrastive learning.

SwAV (Swapping Assignments between Views): An online algorithm that clusters data while maintaining consistency between cluster assignments of different image views.

Barlow Twins: An SSL architecture that focuses on learning embeddings invariant to input sample distortions by minimizing redundancy between the components of these vectors.

Bootstrap Your Own Latent (BYOL): A self-supervised learning method using a pair of neural networks that learn from each other.

TiCo (Transformation Invariance and Covariance Contrast): A method that maximizes agreement among embeddings of distorted versions of the same image.

Pretext Tasks: Specialized training objectives in SSL where a model learns to predict some parts of the data from other parts.

Segment Anything Model (SAM): A model used for the automatic segmentation of microfossils in thin section images.

SwinIR: An architecture used in hierarchical vision transformers for image super-resolution.

Ooids: Spherical grains, typically composed of calcium carbonate, used for geological studies.

Micropaleontology: The study of microscopic fossils, crucial for reconstructing past environmental and climatic conditions.

REFERENCES

- Aribido, O. J., Alregib, G., & Deriche, M. (2020). Self-Supervised Annotation of Seismic Images using Latent Space Factorization. *2020 IEEE International Conference on Image Processing*. <https://iee-dataport.org/open-access/landmass>
- Berg, P., Pham, M. T., & Courty, N. (2022). Self-Supervised Learning for Scene Classification in Remote Sensing: Current State of the Art and Perspectives. In *Remote Sensing* (Vol. 14, Issue 16). MDPI. <https://doi.org/10.3390/rs14163995>
- Bizhani, M., Ardakani, O. H., & Little, E. (2022). Reconstructing high fidelity digital rock images using deep convolutional neural networks. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-08170-8>
- Bridi Guazzelli, A., Roisenberg, M., & Rodrigues, B. B. (2020, July 1). Efficient 3D Semantic Segmentation of Seismic Images using Orthogonal Planes 2D Convolutional Neural Networks. *Proceedings of the International Joint Conference on Neural Networks*. <https://doi.org/10.1109/IJCNN48605.2020.9206661>
- Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., & Joulin, A. (2020). Unsupervised Learning of Visual Features by Contrasting Cluster Assignments. *ArXiv*. <http://arxiv.org/abs/2006.09882>
- Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). *A Simple Framework for Contrastive Learning of Visual Representations*. <http://arxiv.org/abs/2002.05709>
- Cortes, C., Vapnik, V., & Saitta, L. (1995). Support-Vector Networks Editor. In *Machine Learning* (Vol. 20). Kluwer Academic Publishers.
- Dawson, H. L., Dubrule, O., & John, C. M. (2023). Impact of dataset size and convolutional neural network architecture on transfer learning for carbonate rock classification. *Computers and Geosciences*, 171. <https://doi.org/10.1016/j.cageo.2022.105284>
- de Lima, R. P., & Duarte, D. (2021). Pretraining convolutional neural networks for mudstone petrographic thin-section image classification. *Geosciences (Switzerland)*, 11(8). <https://doi.org/10.3390/GEOSCIENCES11080336>
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. *Computer Vision and Pattern Recognition*.
- Dhariwal, P., & Nichol, A. (2021). Diffusion Models Beat GANs on Image Synthesis. *ArXiv, preprint arXiv(2105.05233v4)*. <http://arxiv.org/abs/2105.05233>

- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Hounsby, N. (n.d.). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale* (ArXiv, preprint). <https://github.com/>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Hounsby, N. (2020). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. *ArXiv, preprint*. <https://github.com/>
- Ferreira-Chacua, I., & Koeshidayatullah, A. I. (2023). ForamViT-GAN: Exploring New Paradigms in Deep Learning for Micropaleontological Image Analysis. *IEEE Access*, 11, 67298–67307. <https://doi.org/10.1109/ACCESS.2023.3291620>
- Ferreira, I., Ochoa, L., & Koeshidayatullah, A. (2022). On the generation of realistic synthetic petrographic datasets using a style-based GAN. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-16034-4>
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014a). Generative Adversarial Nets. *ArXiv, preprint*(1802.03065v3). <http://www.github.com/goodfeli/adversarial>
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014b). Generative Adversarial Networks. *Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2, 2672–2680. <http://arxiv.org/abs/1406.2661>
- Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P. H., Buchatskaya, E., Doersch, C., Pires, B. A., Guo, Z. D., Azar, M. G., Piot, B., Kavukcuoglu, K., Munos, R., & Valko, M. (2020). Bootstrap your own latent: A new approach to self-supervised Learning. *ArXiv*. <http://arxiv.org/abs/2006.07733>
- Hamilton, M., Zhang, Z., Hariharan, B., Snavely, N., & Freeman, W. T. (2022). Unsupervised Semantic Segmentation by Distilling Feature Correspondences. *ArXiv Preprint*. <http://arxiv.org/abs/2203.08414>
- He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2019). *Momentum Contrast for Unsupervised Visual Representation Learning*. <http://arxiv.org/abs/1911.05722>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-December*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>

- Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *ArXiv, preprint arXiv(1706.08500v6)*. <http://arxiv.org/abs/1706.08500>
- Ho, M., Idgunji, S., Payne, J. L., & Koeshidayatullah, A. (2022). Hierarchical multi-label taxonomic classification of carbonate skeletal grains with deep learning. *Sedimentary Geology*, 443, 106298. <https://doi.org/10.1016/j.sedgeo.2022.106298>
- Izadi, H., Sadri, J., & Mehran, N. A. (2015). A new intelligent method for minerals segmentation in thin sections based on a novel incremental color clustering. *Computers and Geosciences*, 81, 38–52. <https://doi.org/10.1016/j.cageo.2015.04.008>
- Jaiswal, A., Babu, A. R., Zadeh, M. Z., Banerjee, D., & Makedon, F. (2020). A Survey on Contrastive Self-Supervised Learning. *Technologies*, 9(1), 2. <https://doi.org/10.3390/technologies9010002>
- Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., & Aila, T. (2020). Training Generative Adversarial Networks with Limited Data. *ArXiv, preprint arXiv(2006.06676v2)*. <http://arxiv.org/abs/2006.06676>
- Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2019). Analyzing and Improving the Image Quality of StyleGAN. *ArXiv, preprint arXiv(1812.04948)*. <http://arxiv.org/abs/1912.04958>
- Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and Improving the Image Quality of StyleGAN. *ArXiv, preprint(1912.04958v2)*. <http://arxiv.org/abs/1912.04958>
- Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. *ArXiv, preprint arXiv(1312.6114v10)*. <http://arxiv.org/abs/1312.6114>
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment Anything. *ArXiv Preprint*. <https://segment-anything.com>.
- Klyuchnikov, N., Ismailova, L., Kovalev, D., Safonov, S., & Koroteev, D. (2022). Generative Adversarial Networks for synthetic wellbore data: Expert perception vs mathematical metrics. *Journal of Petroleum Science and Engineering*, 211. <https://doi.org/10.1016/j.petrol.2022.110106>
- Koeshidayatullah, A., Al-Azani, S., Baraboshkin, E. E., & Alfarraj, M. (2022). FaciesViT: Vision transformer for an improved core lithofacies prediction. *Frontiers in Earth Science*, 10. <https://doi.org/10.3389/feart.2022.992442>

- Koeshidayatullah, A., Ferreira-Chacua, I., & Li, W. (2023). Is attention all geosciences need? Advancing quantitative petrography with attention-based deep learning. *Computers and Geosciences*, 181. <https://doi.org/10.1016/j.cageo.2023.105466>
- Koeshidayatullah, A., & Ferreira, I. (2022). A novel deep learning-assisted reservoir fracture delineation with conditional generative adversarial networks. *Second International Meeting for Applied Geoscience & Energy*, 3234–3237. <https://doi.org/10.1190/image2022-3751545.1>
- Koeshidayatullah, A., Morsilli, M., Lehrmann, D. J., Al-Ramadan, K., & Payne, J. L. (2020). Fully automated carbonate petrography using deep convolutional neural networks. *Marine and Petroleum Geology*, 122. <https://doi.org/10.1016/j.marpetgeo.2020.104687>
- Koeshidayatullah, A., Trower, E. J., Li, X., Mukerji, T., Lehrmann, D. J., Morsilli, M., Al-Ramadan, K., & Payne, J. L. (2022). Quantitative evaluation of the roles of ocean chemistry and climate on ooid size across the Phanerozoic: Global versus local controls. *Sedimentology*, 69(6), 2486–2506. <https://doi.org/10.1111/sed.12998>
- Koh, E., Amini Eiman, McLachlan Geoffrey, & Beaton Nick. (2021). Utilising convolutional neural networks to perform fast automated modal mineralogy analysis for thin-section optical microscopy. *Minerals Engineering*, 173. <https://doi.org/https://doi.org/10.1016/j.mineng.2021.107230>
- Koh, E. J. Y., Amini, E., McLachlan, G. J., & Beaton, N. (2021). Utilising convolutional neural networks to perform fast automated modal mineralogy analysis for thin-section optical microscopy. *Minerals Engineering*, 173. <https://doi.org/10.1016/j.mineng.2021.107230>
- Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., & Timofte, R. (2021). SwinIR: Image Restoration Using Swin Transformer. *ArXiv*. <http://arxiv.org/abs/2108.10257>
- Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2014). Microsoft COCO: Common Objects in Context. *ArXiv, preprint*. <http://arxiv.org/abs/1405.0312>
- Liu, X., Zhang, F., Hou, Z., Mian, L., Wang, Z., Zhang, J., & Tang, J. (2023). Self-Supervised Learning: Generative or Contrastive. *IEEE Transactions on Knowledge and Data Engineering*, 35(1), 857–876. <https://doi.org/10.1109/TKDE.2021.3090866>
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In *arXiv (ArXiv, preprint)*. <http://arxiv.org/abs/2103.14030>

- Mamode, H. I., Hampson, G. J., & John, C. M. (2025). Do more with less: Exploring semi-supervised learning for geological image classification. *Applied Computing and Geosciences*, 25, 100216. <https://doi.org/10.1016/j.acags.2024.100216>
- Marchant, R., Tetard, M., Pratiwi, A., Adebayo, M., & De Garidel-Thoron, T. (2020). Automated analysis of foraminifera fossil records by image classification using a convolutional neural network. *Journal of Micropalaeontology*, 39(2), 183–202. <https://doi.org/10.5194/jm-39-183-2020>
- Mitra, R., Marchitto, T. M., Ge, Q., Zhong, B., Kanakiya, B., Cook, M. S., Fehrenbacher, J. S., Ortiz, J. D., Tripathi, A., & Lobaton, E. (2019). Automated species-level identification of planktic foraminifera using convolutional neural networks, with comparison to human performance. *Marine Micropaleontology*, 147, 16–24. <https://doi.org/10.1016/j.marmicro.2019.01.005>
- Montanaro, A., Valsesia, D., Fracastoro, G., & Magli, E. (2021). *Semi-supervised learning for joint SAR and multispectral land cover classification*. <https://doi.org/10.1109/LGRS.2022.3195259>
- Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba, W., Howes, R., Huang, P.-Y., Li, S.-W., Misra, I., Rabbat, M., Sharma, V., ... Bojanowski, P. (2023). *DINOv2: Learning Robust Visual Features without Supervision*. <http://arxiv.org/abs/2304.07193>
- Pires de Lima, R., Duarte, D., Nicholson, C., Slatt, R., & Marfurt, K. J. (2020). Petrographic microfacies classification with deep convolutional neural networks. *Computers and Geosciences*, 142. <https://doi.org/10.1016/j.cageo.2020.104481>
- Pires de Lima, R. P., & Duarte, D. (2021). Pretraining convolutional neural networks for mudstone petrographic thin-section image classification. *Geosciences (Switzerland)*, 11(8). <https://doi.org/10.3390/GEOSCIENCES11080336>
- Pratama, H., & Latiff, A. H. A. (2022). Automated Geological Features Detection in 3D Seismic Data Using Semi-Supervised Learning. *Applied Sciences (Switzerland)*, 12(13). <https://doi.org/10.3390/app12136723>
- Rani, V., Nabi, S. T., Kumar, M., Mittal, A., & Kumar, K. (2023a). Self-supervised Learning: A Succinct Review. In *Archives of Computational Methods in Engineering* (Vol. 30, Issue 4, pp. 2761–2775). Springer Science and Business Media B.V. <https://doi.org/10.1007/s11831-023-09884-2>
- Rani, V., Nabi, S. T., Kumar, M., Mittal, A., & Kumar, K. (2023b). Self-supervised Learning: A Succinct Review. In *Archives of Computational Methods in*

- Engineering* (Vol. 30, Issue 4, pp. 2761–2775). Springer Science and Business Media B.V. <https://doi.org/10.1007/s11831-023-09884-2>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). You Only Look Once: Unified, Real-Time Object Detection. *ArXiv*, *arXiv:1506.02640v5*. <http://arxiv.org/abs/1506.02640>
- Reis, D., Kupec, J., Hong, J., & Daoudi, A. (2023). Real-Time Flying Object Detection with YOLOv8. *ArXiv*, *arXiv:2305.09972v2*. <http://arxiv.org/abs/2305.09972>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *ArXiv*, *preprint*. <http://arxiv.org/abs/1505.04597>
- Sawarkar, K., Joglekar, A., & Solanki, S. R. (2022). *Deep learning with PyTorch Lightning: swiftly build high-performance Artificial Intelligence (AI) models using Python*.
- Sen, B. (2003). *Modern foraminifera* (1st ed.). Kluwer Academic Publishers.
- Sevillano, V., Holt, K., & Aznarte, J. L. (2020). Precise automatic classification of 46 different pollen types with convolutional neural networks. *PLoS ONE*, *15*(6). <https://doi.org/10.1371/journal.pone.0229751>
- Shoji, D., Noguchi, R., Otsuki, S., & Hino, H. (2018). Classification of volcanic ash particles using a convolutional neural network and probability. *Scientific Reports*, *8*(1). <https://doi.org/10.1038/s41598-018-26200-2>
- Tritrong, N., Rewatbowornwong, P., & Suwajanakorn, S. (2021). Repurposing GANs for One-shot Semantic Part Segmentation. *ArXiv*, *preprint*(2103.04379v5). <http://arxiv.org/abs/2103.04379>
- Wang, Y., Albrecht, C. M., Braham, N. A. A., Mou, L., & Zhu, X. X. (2022). Self-supervised Learning in Remote Sensing: A Review. *ArXiv*. <http://arxiv.org/abs/2206.13188>
- Wang, Z., Ma, G., Su, B., Du, X., Li, J., Sun, T., & Jiang, G. (2023). CSSRS: Pore segmentation method of sandstone cast thin section images based on weak supervised learning. *Geoenergy Science and Engineering*, *230*. <https://doi.org/10.1016/j.geoen.2023.212207>
- Wu, X., Liang, L., Shi, Y., & Fomel, S. (2019). FaultSeg3D: using synthetic datasets to train an end-to-end convolutional neural network for 3D seismic fault segmentation. *Geophysics*, *84*(3).

- Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., & Luo, P. (2021). SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. *ArXiv*. <http://arxiv.org/abs/2105.15203>
- Zbontar, J., Jing, L., Misra, I., LeCun, Y., & Deny, S. (2021). Barlow Twins: Self-Supervised Learning via Redundancy Reduction. *ArXiv*. <http://arxiv.org/abs/2103.03230>
- Zhu, J., Moraes, R. M., Karakulak, S., Sobol, V., Canziani, A., & LeCun, Y. (2022). TiCo: Transformation Invariance and Covariance Contrast for Self-Supervised Visual Representation Learning. *ArXiv*. <http://arxiv.org/abs/2206.10698>

VITAE

Name	Ivan Eduardo Ferreira Chacua
Nationality	Colombia
Date of Birth	22/05/1996
Email	ivaneduardoferreira@gmail.com
Address	Carrera 144 #13-35. Barrio Girasoles, Bogotá D.C. Postal Code: 111121
Academic Background	B.Sc. in Geology. Universidad Nacional de Colombia

As a graduate from the M.Sc. in Geology from the King Fahd University & Minerals (KFUPM) in Dhahran, Saudi Arabia, with a focus on the applications of Artificial Intelligence in geosciences. My academic journey is distinguished by the publication of three Q1 journal papers and presentations at four key conferences, including the EAGE Annual, SEG Image, EGU, and AASP-TPS, showcasing my research proficiency.

My expertise extends to programming, particularly in Python, and Geographic Information Systems (GIS) with tools like ArcGIS, QGIS, and Oasis Montaj. A native Spanish speaker and fluent in English, my communication skills are professional in both languages. Professionally, I contributed to the assessment of the exploratory potential of Colombia's offshore basins at Neoil for the National

Hydrocarbon Agency (ANH) of Colombia. This role underscored my analytical capabilities in a real-world context.

Further exploring the confluence of geology and technology, I undertook a Data Science and Machine Learning internship at Grupodot, subsequently advancing to a full-time Data Scientist role. My academic and professional pursuits in Germany were enriched by a DAAD Travel Scholarship, emphasizing sedimentological and paleontological studies.

Beyond formal education, I actively pursued training in GIS, petroleum geology, and was part of the first cohort of Mission TIC 2022, under the aegis of my country's Ministry of Technology. This was complemented by advanced Data Science training with Correlation ONE where I graduated with honors. I am keen to continue blending programming, data science, and deep learning with geology, aiming to contribute innovatively to the field. My multidisciplinary background positions me uniquely for roles that demand a blend of geological expertise and technological acumen.

Publications:

1. Ferreira, I., Ochoa, L., & Koeshidayatullah, A. (2022). On the generation of realistic synthetic petrographic datasets using a style-based GAN. *Scientific Reports*, 12(1), 12845.

2. Ferreira-Chacua, I., & Koeshidayatullah, A. (2023). Foramvit-gan: Exploring new paradigms in deep learning for micropaleontological image analysis. *IEEE Access*, *11*, 67298-67307.
3. Koeshidayatullah, A., Ferreira-Chacua, I., & Li, W. (2023). Is attention all geosciences need? Advancing quantitative petrography with attention-based deep learning. *Computers & Geosciences*, *181*, 105466.

Conference Presentations:

1. Yusmananto, S., Fahmi, B., Ferreira, I., Ibrahim, S., & Koeshidayatullah, A. (2023). CO₂ storage and utilization potential in the Middle East: Lessons learned from the Cretaceous carbonate reservoirs. *84th EAGE Annual Conference & Exhibition 2023*(1), 1-5.
2. Ferreira, I., Koeshidayatullah, A., Baharudin, F., Yusmananto, S., & Allam, S. (2023). PoreFormer: A Novel Microporosity Characterization in Mudstone using a Deep Learning Vision Transformer Approach. *84th EAGE Annual Conference & Exhibition 2023*(1), 1-5.
3. Ferreira, I., & Koeshidayatullah, A. (2023). Generating self-labeled geological datasets for semantic segmentation using pretrained GANs. *EGU General Assembly Conference Abstracts*, EGU-888.
4. Koeshidayatullah, A., & Ferreira, I. (2022). A novel deep learning-assisted reservoir fracture delineation with conditional generative adversarial networks. *Second International Meeting for Applied Geoscience & Energy*, 3234-3237.

5. Ferreira-Chacua, I., & Koeshidayatullah, A. (2024). Segment Every Fossil—A Deep Learning Model Tailored for Automatic Segmentation of Microfossils in Thin Section Images. *EGU24*.
6. Allam, S., Ferreira, I., Fahmi, B., Yusmananto, S., & Koeshidayatullah, A. (2023). AI-assisted Oceanic Anoxic Events (OAEs) identification and prediction: A case from the Late Cretaceous OAE-2. *84th EAGE Annual Conference & Exhibition 2023*(1), 1-5.
7. Koeshidayatullah, A., & Ferreira, I. (2023). SelfMin: Self-Supervised Deep Learning for Advanced Mineralogical Analysis. *EGU General Assembly Conference Abstracts*, EGU-2189.
8. Ferreira, I., Ochoa, L., & Koeshidayatullah, A. (2023). This thin section doesn't exist: On the generation of synthetic petrographic datasets. *Second International Meeting for Applied Geoscience & Energy*, 1620-1623.