IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 4 June 2023, accepted 27 June 2023, date of publication 3 July 2023, date of current version 10 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3291620

RESEARCH ARTICLE

ForamViT-GAN: Exploring New Paradigms in Deep Learning for Micropaleontological Image Analysis

IVAN FERREIRA-CHACUA[1] AND ARDIANSYAH IBNU KOESHIDAYATULLAH[1,2]

[1]Department of Geosciences, College of Petroleum Engineering and Geosciences, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia
[2]Center for Integrated Petroleum Research, Dhahran 34464, Saudi Arabia

Corresponding authors: Ivan Ferreira-Chacua (g202211300@kfupm.edu.sa) and Ardiansyah Ibnu Koeshidayatullah (a.koeshidayatullah@kfupm.edu.sa)

This work was supported in part by the King Fahd University of Petroleum and Minerals Startup Grant SF21011.

ABSTRACT In geosciences, micropaleontology studies the evolution of microfossils (e.g., foraminifera) throughout geological records and utilizes such information to reconstruct past environmental and climatic conditions. This field depends primarily on the visual recognition of various features in microfossils, which makes it ideal for applying computer vision technology, specifically deep convolutional neural networks (CNNs), to automate and optimize different microfossil identification and classification. In addition, the unlabeled, low-resolution micropaleontological dataset is often available in a large volume compared to another geosciences dataset. While the application of deep learning in micropaleontology is rapidly growing, these efforts have been severely hampered by (i) the limited availability of high-quality and high-resolution labeled fossil images and (ii) significant effort in manually labeling various fossils by subject matter experts. Furthermore, previous works primarily exploited CNN with transfer learning to obtain high-accuracy prediction, which may reduce the explainability and reproducibility of the model. To overcome this issue, we propose a novel deep learning workflow that couples hierarchical vision transformers with style-based generative adversarial network algorithms to efficiently acquire and synthetically generate realistic high-resolution labeled datasets of micropaleontology in a large volume. Our study demonstrates that the proposed workflow could generate high-resolution images with a high signal-to-noise ratio, achieving 39.1 dB, and realistic synthetic images with a Fréchet inception distance similarity score of 14.88. In addition, our proposed workflow could provide a considerable volume of self-labeled datasets that can be used for model benchmarking and various downstream visual tasks, including fossil classification and segmentation. We further performed, for the first time, a few-shot semantic segmentation of different foraminifera chambers on both the generated and synthetic images with high accuracy. This novel meta-learning approach is only possible when a high-resolution and high-volume labeled dataset is available. Therefore, our proposed deep learning-based workflow is promising and shows a potential to advance and optimize micropaleontological research and other visual-dependent geological analysis.

INDEX TERMS Image analysis, foraminifera, deep learning, GAN, transformer, semantic segmentation, foraminifera, few-shot learning.

I. INTRODUCTION

Throughout geological history, foraminifera represents one of the most exceptionally diverse groups of marine microfossils, with an estimated number of current species between 8,966 and an estimated number of 40,888 fossil species in the geological record [1]. This accounts for approximately 2% of all animal species from the Cambrian to the present [1]. The size of fossil foraminifera is very diverse, ranging from less than 100 microns to 20 centimeters,

The associate editor coordinating the review of this manuscript and approving it for publication was Wenbing Zhao.

and their shell can be made up of diverse compositions, such as calcite, aragonite, agglutinated particles, and other organic compounds. In foraminifera, factors such as their cosmopolitan nature and evolutionary diversification make them of particular interest to provide a paleontological and stratigraphic record, which is of significant value in carrying out biostratigraphic correlations and paleoenvironmental interpretations [1], [2].

In both ancient and modern environments, the relative abundance of specific species and their corresponding morphometric characteristics are used as a proxy for (paleo) temperature, (paleo) oxygen concentration, and (paleo) oceanic salinity and paleoproductivity. In addition, foraminifera is often used as the building block to define biofacies for paleobathymetric studies as an aid in characterizing sedimentary sub-environments. With the progressive change of different macro- and microfossils throughout the history of the Earth, especially planktonic foraminifera being utilized as markers on the geological time scale and the occurrence of specific events in the stratigraphic record [1], [2], [3]

Detailed identification of both species and morphotypes and producing high-quality photomicrographs of foraminifera have been primarily dependent on the availability of high-end equipment, such as advanced stereomicroscopes and high-resolution scanning systems. Although some techniques have become a staple in research institutes, some high-end equipment for digitizing specimens at high resolution is not widely accessible to the geoscientific community. This is exacerbated by the expertise needed to perform species and genera classification. All of this leads to an issue with standardization across various laboratories and institutions, severely limiting the reproducibility of such classification and its accessibility to non-experts. As a result, there is an urgent need to develop an efficient, automated approach or workflow to improve the resolution of microfossil images and obtain labeled datasets without requiring high-end equipment. Furthermore, the widespread implementation of robust deep learning models for foraminifera classification and morphological diversity distillation using advanced computer vision technology, particularly deep convolutional neural networks, has yet to be thoroughly investigated.

In the current literature, numerous works leverage computer vision technology to classify and segment microfossil specimens [3], [4]. From seminal implementations in which the algorithms used did not achieve human accuracy [4], [5], [6] to recent works in which the algorithms exceed human accuracy and speed when classifying microfossils [4], [7]. Recently, the use of Deep Convolutional Neural Networks (CNNs) has been notable in this research corpus, with CNNs having several advantages for these tasks given the reduced need for feature engineering, the scalability to larger datasets, and the exceptional ability to process grid-like data [4], [8], [9], [10], [11], [12]. This is further supported by the accessibility to powerful pre-trained architectures as backbones for a starting point during the training of new models [10], [12], [13], [14].

In geosciences, recent advances in Generative Adversarial Networks (GANs) enable geoscientists to generate additional synthetic data as a novel augmentation technique to conventional ones, assisting in the improvement of machine learning model performance [15], [16]. In addition, GANs are effective at balancing the distribution of data within a particular geological dataset, ensuring a better representation, and reducing the risk of bias during training [15], [17], [18].

In recent research trends, Vision Transformers (ViT) have emerged from the Natural Language Processing (NLP) corpus as a powerful technique for tackling visual tasks [19], [20]. Unlike traditional CNNs, ViTs employ a transformer architecture to address global and local relationships in an image, resulting in more effective feature extraction and representation. Image Super-Resolution (SR) is a field of Computer Vision that focuses on enhancing low-resolution images, making them more visually appealing and informative. Recent improvements in CNN and ViT-based architectures have become state-of-the-art for upscaling and restoring images [21], [22]. This application could be helpful in preserving fine-grained details and textural information for micropaleontological image analysis.

Additionally, we explored Few-shot Learning, a machine learning subfield that focuses on training models to recognize new classes with very limited labeled data, typically only a few examples per class. This approach is particularly relevant in micropaleontological research, where obtaining a large volume of labeled samples can be time-consuming and labor-intensive, as in many other areas of geosciences. Few-shot learning techniques can be broadly categorized into two main approaches: Non-Meta-Learning (Transfer Learning) and Meta-Learning (Siamese Networks, CAN, Meta-Networks) [23] [24]. On the other hand, approaches like transfer learning leverage knowledge learned from one task and apply it to a related but different task. Meta-Learning aims to approximate a function with optimal performance on a task randomly sampled from a distribution. This approach is divided into metric-based, model-based, and parameter-based techniques [24]. In our study, we explore the potential of few-shot learning techniques to improve a segmentation task of foraminifera images, even when only a small number of labeled examples are available (i.e., less than five labeled samples). By incorporating few-shot learning into our deep learning workflow, we aim to enhance the efficiency and generalizability of our models, making them more applicable to real-world micropaleontological scenarios [3], [4], [5], [6].

The primary goals of this research were to create and suggest new approaches to traditional microfossil image scaling methods and to investigate the application of end-to-end deep learning for enhancing the quality and accurately representing the morphological diversity of foraminifera images [15] successfully showcased this in their study using petrographic datasets (Fig. 1a and b). Such tools can potentially expand the variety of micropaleontological datasets and provide synthetic digital counterparts for confidential data.

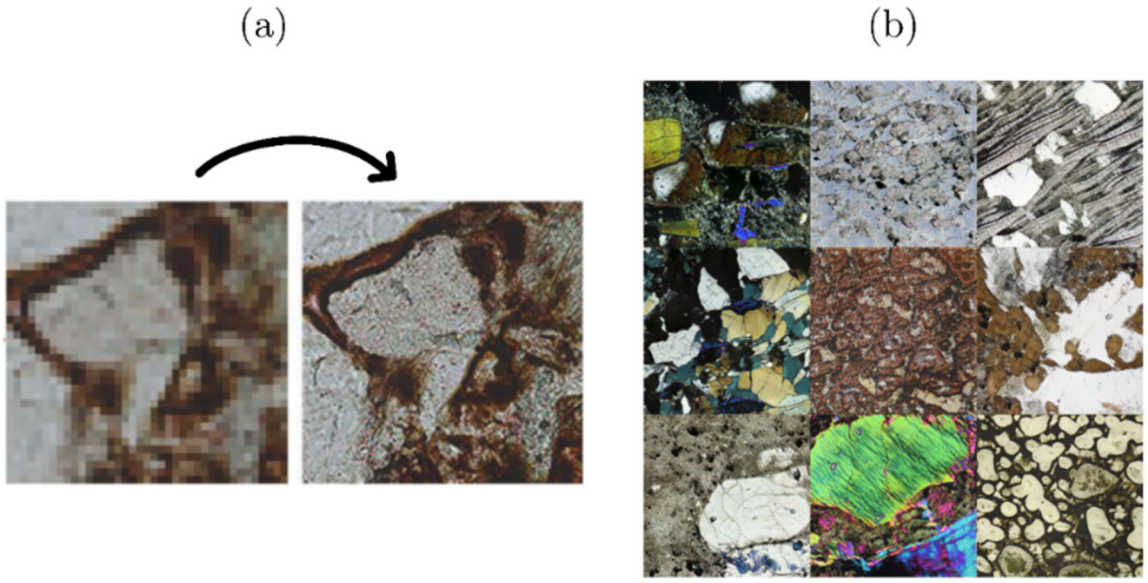

FIGURE 1. Examples of the use of deep learning architectures in petrography (a) super-resolution imaging of sandstones [35] (b) generation of thin sections [15].

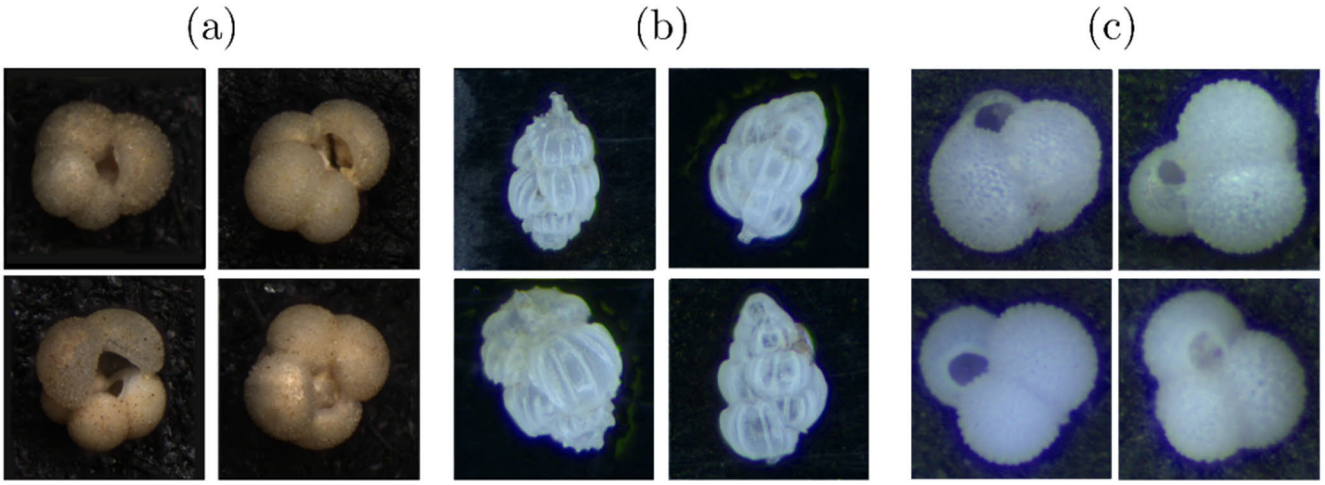

FIGURE 2. Image samples from the three databases used (a) Globigerina bulloides from Endless Forams, (b) Uvigerina peregrina from MD02-2508 (c) Globigerinoides ruber from MD97-2138. Selected from the compilation by [3].

II. METHODOLOGY

A. DATASET

In this study, three open-source machine learning-ready databases of foraminifera are used. These datasets were curated and preprocessed before being used in classification tasks using CNNs [3]. All three sources are being preprocessed further in preparation for implementing the workflow proposed in this study to generate a high-resolution and realistic synthetic micropaleontological dataset.

The first source is the Endless database Forams which is a compilation of more than 34000 foraminifera [25] (Fig. 2a). The second dataset is obtained from the set MD02-2508 compiled by the RV Marion Dufresne oceanography mission MD126 MONA during 2002 in the Northeast Pacific Ocean [3] (Fig. 2b). Lastly, the MD97-2138 dataset, these images were collected from the IHPIS mission of the RV Marion Dufresne to analyze the last climatic cycle in sediment cores [3] (Fig. 2c).

These databases are selected for their previous application in training a CNN for classification and for having numerous foraminifera specimens and samples. Furthermore, the datasets vary in terms of backgrounds, illumination, fragmentation of the specimen, and number of foraminifera per species. These features are useful for robustly training deep learning models by presenting different potential variations and conditions of foraminifera images acquired from real-world datasets. This would help deep learning

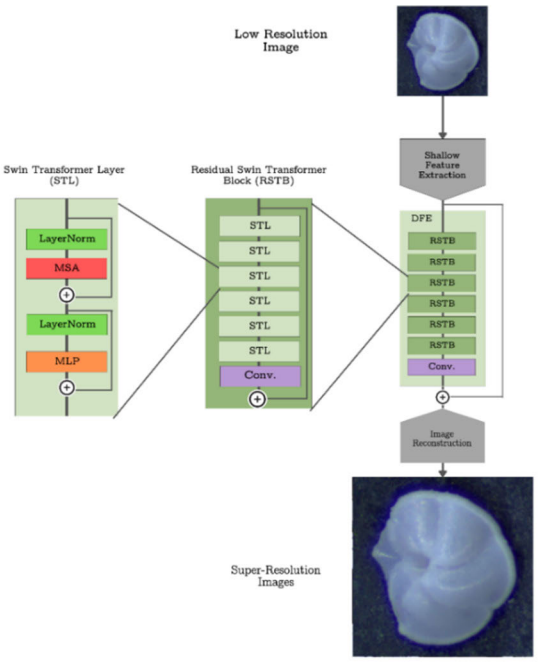

FIGURE 3. Schematic of the adopted SwinIR architecture used to extract features and upscale images (a) Low-resolution image input to the model (b) Generalized schematic of SwinIR (c) Image upscaled using super-resolution (d) RSTB block, Residual Swin Transformer Block (e) Transformer Layer Swin, Swin transformer Layer. MSA: Multi-head Self-Attention. MLP: Multi-Layer Perceptron, LayerNorm: Layer Normalization. Modified from [21].

models realistically and aesthetically restore and generate the foraminifera dataset.

B. TRANSFORMER-BASED IMAGE SUPER-RESOLUTION

For the multiscale image restoration task (i.e., image super-resolution), 3,563 specimens are selected from the three databases that meet the condition of having a dimension greater than 800 × 800 px. This threshold was applied to ensure that the datasets are above high-definition quality (> 720px) and have enough samples to train the models successfully. The collected images are divided into a training set with 3,263 images and a test set with 300 randomly selected images. The images were reduced to half, a quarter, and an eighth of their original resolution of 800px2, using a Lanczos interpolation kernel to preserve the best possible features of the image to be reconstructed. In such a case, the dimensions to be restored are from (i) 400px2 to 800px2 (2x time super-resolution task); (ii) 200px2 to 800px2 (4x time super-resolution task); and 100px2 to 800px2 (8x time super-resolution task). The algorithms were trained until reaching 50000 iterations for comparison. The training was done using 25 hours of GPU with an NVIDIA RTX 5000 of 16Gb of video RAM in a Cloud Linux environment, with eight CPU cores and 30 Gb of RAM.

For the foraminifera microphotograph image super-resolution task, our study adopted the SwinIR (Shifted Windows Transformers for Image Restoration) [21] (Fig. 3) architecture which is an extension of the Swin transformer algorithm [26]. This architecture is based on the hierarchical neural attention mechanism concept. It has positioned itself as the state-of-the-art deep learning algorithm across standardized datasets used in computer vision tasks. This trend, combined with their combination with CNNs for architectural improvements, has been seen since the original implementation of the transformers for vision tasks, Vision Transformers (ViT) [20].

After serving as the backbone of several architectures due to its SoTA results, ViT was implemented in the Shifted Window Transformer (Swin) [26] and applied to image restoration and super-resolution. This architecture uses stacked residual Swin transformer blocks that are windows using a shifted window attention coupled with convolutional layers for further feature extraction (Fig. 3). These architectural improvements allow the model to achieve better results than current methods by both reducing the number of parameters needed and capturing long-range relationships in the image, allowing in contrast with approaches solely based in CNNs [21], [26].

In geosciences, the implementation of ViTs, a recent development in computer vision, has shown promising results in various image analysis tasks [19], [20], [27], [28]. Unlike traditional Convolutional Neural Networks (CNNs), ViTs use a transformer architecture based on self-attention mechanisms to capture global and local relationships in an image. This allows ViTs to effectively extract and represent features from images, even when dealing with large datasets or complex visual patterns [19], [20]. In our study, we employ ViTs to enhance the quality of image foraminifera by leveraging the transformer architecture. With this approach, we can capture fine-grained details critical for accurately identifying foraminifera species and morphotypes. Additionally, ViTs have the advantage of being more interpretable than traditional CNNs, as the self-attention mechanism allows us to visualize which parts of the image are most important for classification [27]. This makes ViTs a valuable tool for micropaleontological research, where accurate and interpretable image analysis is crucial for understanding the paleoenvironmental and paleoclimatic conditions of the past. Overall, the use of ViTs in our study represents a significant advancement in micropaleontological image analysis and has the potential to significantly improve the accuracy and efficiency of foraminifera image analysis [27].

C. STYLE-BASED IMAGE GENERATION

The generative model was trained by selecting the nine species with the most images (*Globigerina bulloides, Globigerinita glutinata, Globigerinoides ruber, Globigerinoides sacculifer, Globorotalia inflata, Globorotalia menardii, Globorotalia truncatulinoides, Neogloboquadrina incompta, and Neogloboquadrina pachyderma*). In addition, these

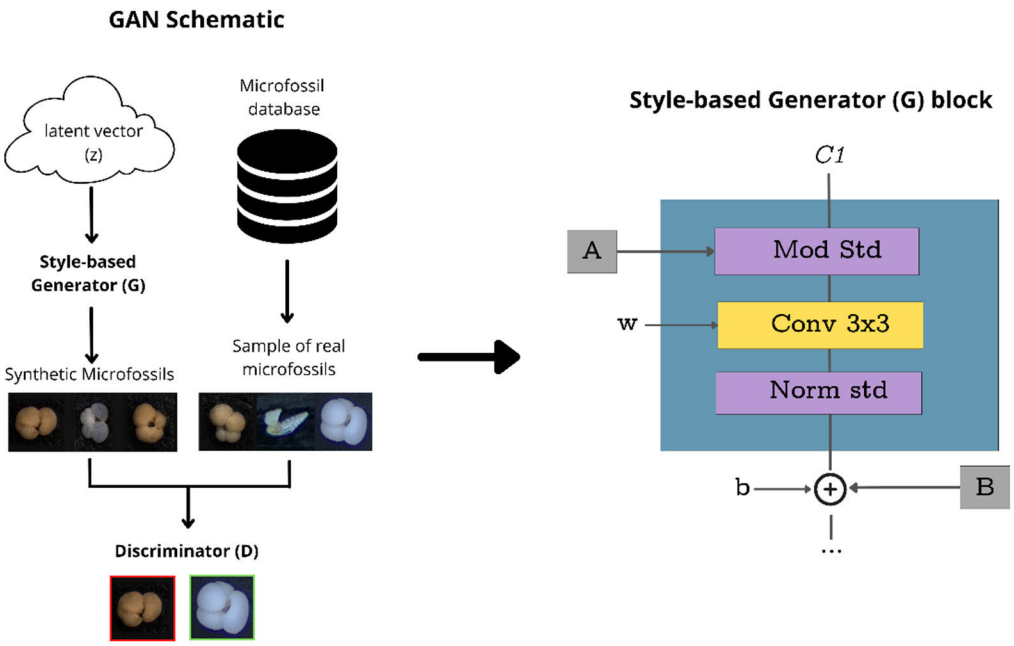

FIGURE 4. On the left, diagram of a GAN and how it works by taking a latent vector (z), that can be random or conditioned, and a sample from the real dataset. It then learns to replicate this image distribution using a style-based generator (G). These generated are then discriminate from the real ones using a discriminator (D). The goal of the two neural networks competing in a minimax game in which the goal of G is to deceive D and the goal of D is to discriminate the real samples from the synthetic ones generated by D. On the right, the main Style block used for style extraction on given image datasets by StyleGAN2, with modulation, convolution, normalization, A – Affine transforms, B – Noise broadcast operation, w - learned weights and b – biases [31], [32].

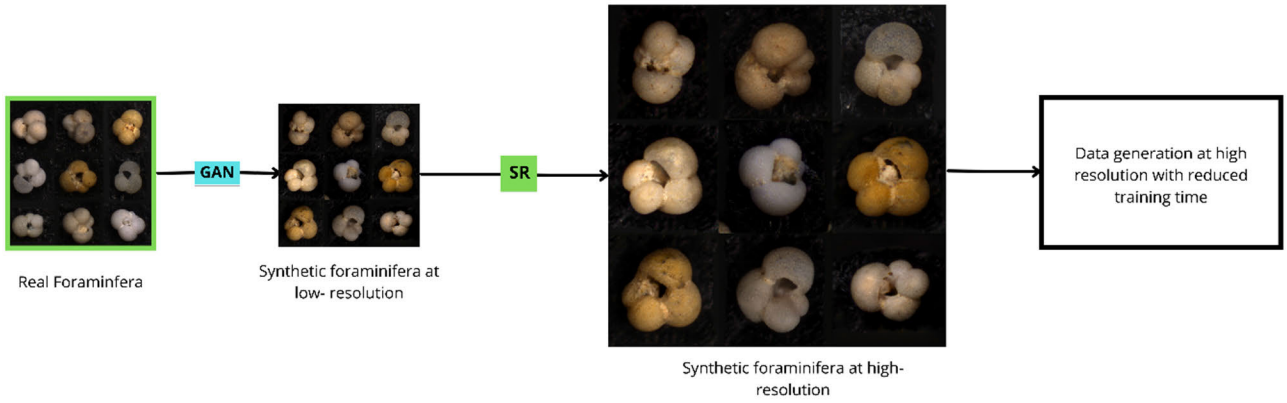

FIGURE 5. The proposed workflow for coupling GANs with Super-Resolution algorithms to train GANs at lower resolution and upscale the generated images.

species were selected because they share some similarities and distinct characteristics across species simultaneously. A total of 18166 images were taken at a resolution of 256px2 to obtain the largest number of images and satisfy the requirement of the minimum dataset to properly train generative adversarial network (10^4 - 10^5 images) [26], [29], [30], [31], [32].

Another training was conducted using a $512px^2$ upsampling of this same dataset to further experiment with the associated latent space at that scale and evaluate the convergence of the model during training with different dataset resolutions. The idea for this is to combine the training of GANs and Super-resolution for image generation at higher resolutions with limited data and reduced training time (Fig. 5). Various architectures take advantage of the zero-sum game between two neural networks, constituting the core of Generative Adversarial Nets [29]. Recently, these architectures have been able to deceive the human eye for both face recognition [34], [35] and even for the recognition of specific datasets in geology [15]. In this study, the style-based GAN architecture (StyleGAN2) was adopted and implemented to reconstruct and generate realistic synthetic foraminifera images that are realistic to the human eye [26], [29], [30], [31], [32]. (Fig. 4).

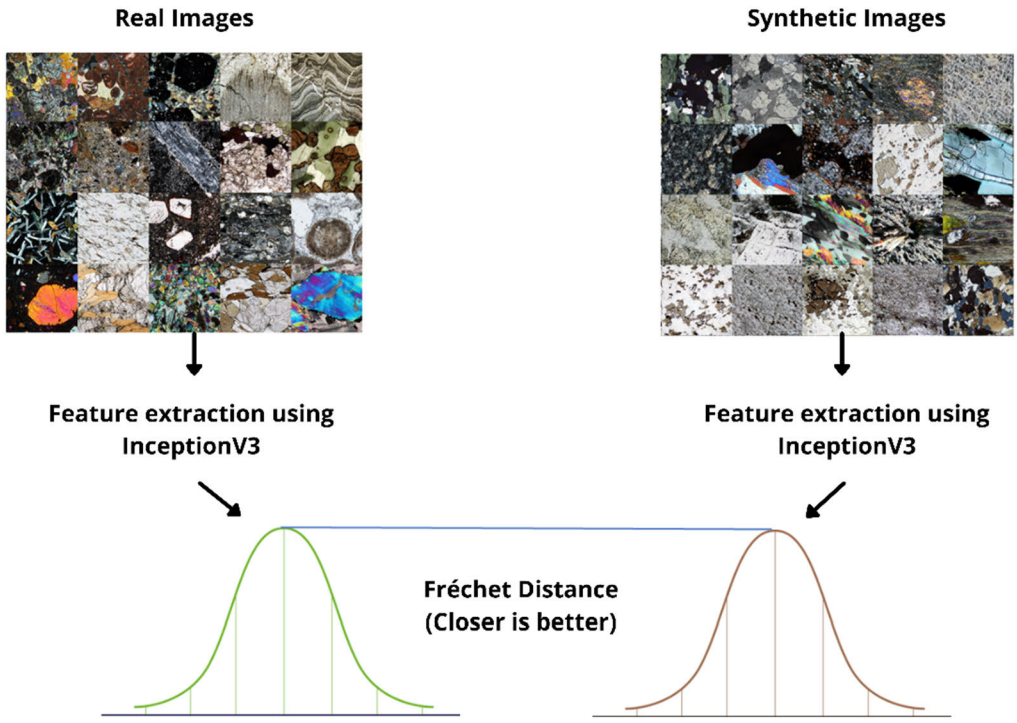

FIGURE 6. Flowchart to calculate the FID score, taken from [15].

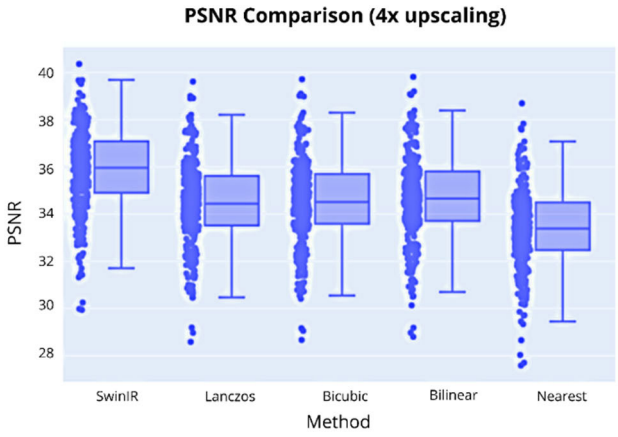

FIGURE 7. PSNR results for a 4x upscaling (200px2 to 400px2). Comparison between SwinIR and the interpolations using the Lanczos kernel, bicubic bilinear and Nearest Neighbors interpolation.

This assumption was based on the Frechet Inception Distance (FID) Score achieved by this architecture and its variations, with the StyleganXL being the current state-of-the-art implementation for image generation tasks [33], [36]. Among other unconditional and style-based GAN models, the StyleGAN2 architecture has a more robust and flexible implementation that allows experimenting with the model and its associated latent space, allowing not only to generate but also to model synthetic images. For such a case, the algorithm extracts styles of interest in the dataset of images from which it is trained. These styles can be interpreted in faces as hair color, face orientation, or skin tone [30], while for petrographic datasets, it translates into grain size and color of minerals in the thin section [15].

D. METRICS

In evaluating the models, several metrics were used in this study. We use Peak Signal-To-Noise ratio, PSNR (eq. 1), to evaluate the super-resolution model as a standard metric for assessing super-resolution models and scaling algorithms. This metric has been calculated in the literature for both standard computer vision datasets [21], [22] and in datasets of geological interest [37], [38].

In addition, this metric calculates the logarithm of the ratio between the maximum value of a signal, the 255-pixel value for a grayscale image, and the mean squared error of the image

(MSE), eq. 1. In this metric, the higher value indicates a better image reconstruction, with values greater than 40 expected from standard image compression algorithms and undefined when the images are the same, as the MSE goes to zero (refer to eq.1).

$$PSNR = 10 \times \log_{10}\left(\frac{255^2}{MSE}\right) \tag{1}$$

Values obtained in the literature for geological datasets ranging from 25 to 45 PSNR for different image datasets, from petrographic to micro-computed tomography images [34], [38], [39].

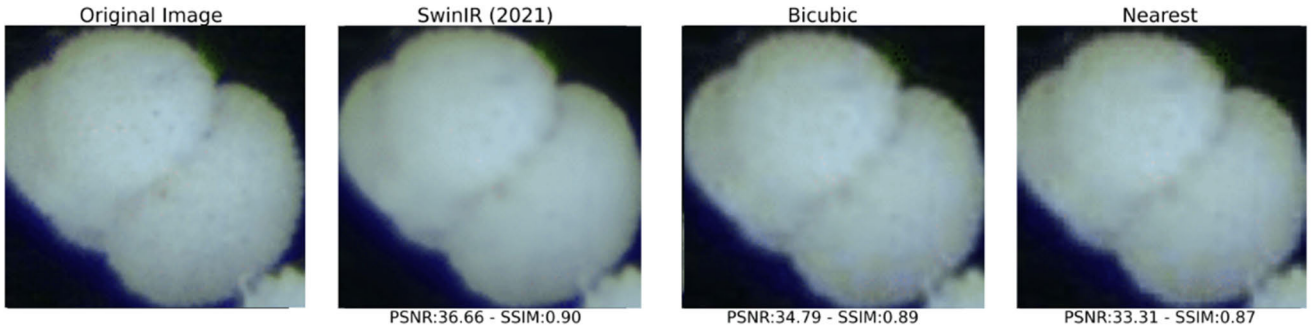

FIGURE 8. 4x magnification of foraminifera specimens, PSNR: Peak Signal-to-Noise Ratio SSIM: Structural Similarity Index.

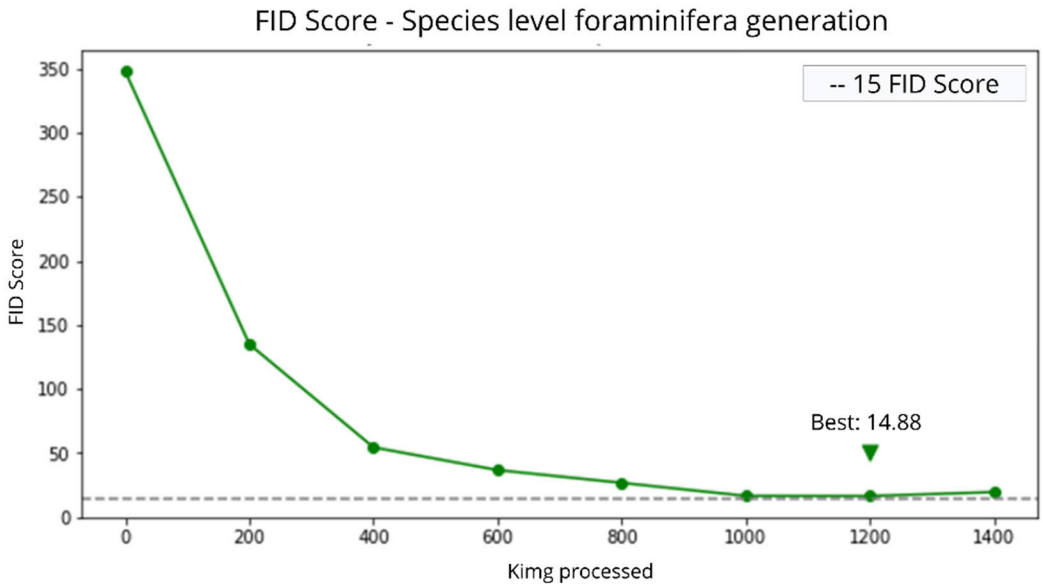

FIGURE 9. FID score for foraminiferal species, indicating the best score obtained, 14.88.

The generative algorithm is evaluated using the FID Score using InceptionV3 [40], [41]. For this metric, the images are sampled, real and fake, and sent as input to the InceptionV3 architecture [41]. The activation results of the last layers of this classification architecture are calculated for both populations, generating a distribution for the synthetic and real images.

The Fréchet distance between these two distributions is calculated where the smaller the value, the closer the two distributions are, i.e., the more the synthetic images resemble the actual population (Fig. 6). This is a standard metric used for the evaluation of generative models, FID score values below 5 expected in face generation using 10^5 image datasets and values around 10 obtained in the generation of synthetic petrographic images [15], [32], [40].

III. RESULTS

A. SUPER-RESOLUTION ALGORITHM

The SwinIR architecture was used to train it and compare its performance to traditional interpolation methods, as detailed by [42] and Liang et al. [21]. The best PSNR value within 50000 iterations of each of the models (2×,4×,8×) is chosen and compared, for the case of the 4× scale, it is also compared with other conventional algorithms (Fig. 7). The Average PSNR values obtained for the test set in each of the training shown in Table 1, together with the iteration of the model in which it was obtained.

TABLE 1. PSNR values obtained for each scaling task and the iteration in which it was reached.

Scaling	*Low resolution (px^2)*	*High Resolution (px^2)*	*PSNR value*	*Iteration*
2×	400	800	39.12	45000
4×	200	800	35.81	50000
8×	100	800	33.29	35000

In addition, a visual comparison is conducted for foraminifera and its magnifications, both done by SwinIR and by bicubic interpolation and nearest neighbors interpolation

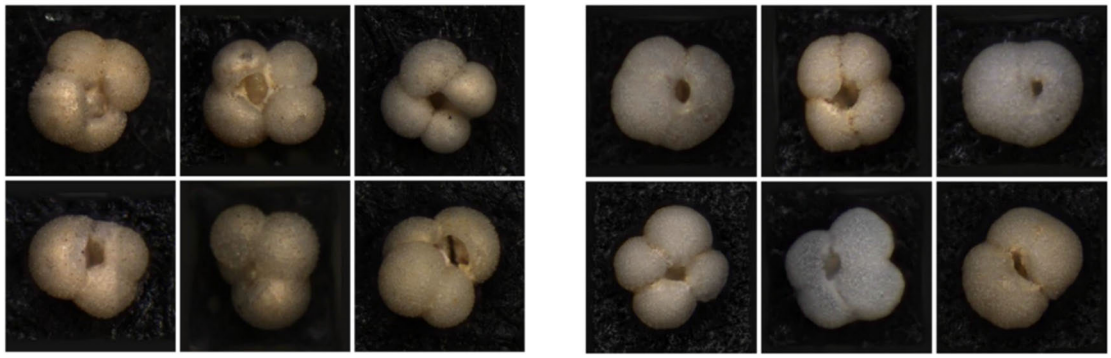

FIGURE 10. Generation of species-specific foraminifera (a) Real foraminifera of the species Globigerina bulloides (b) Generated specimens of G. bulloides.

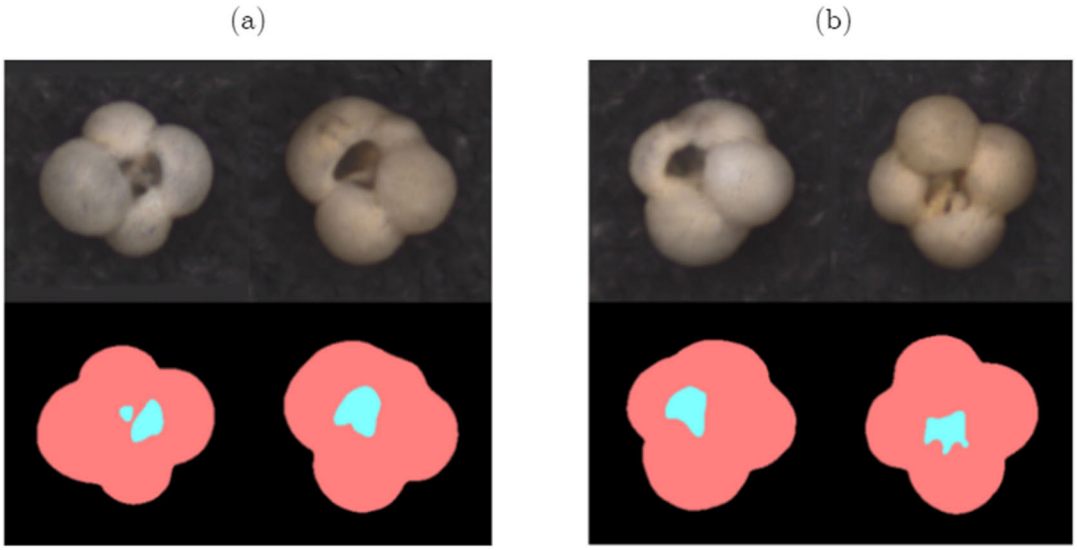

FIGURE 11. Few-Shot segmentation of Globigerina bulloides using the pretrained GAN and five specimes labelled (a) sample of specimens labelled (b) model inferences.

(Fig. 8). The results showed that SwinIR outperformed these traditional approaches, demonstrating its potential to surpass conventional methods for image restoration.

B. GENERATIVE ALGORITHM

The StyleGAN2 architecture was also trained to generate synthetic foraminifera specimens and explore their associated latent space [34], [35]. The performance of the generative model was evaluated both quantitatively, with the FID score (Fig. 9), and qualitatively with a visual inspection of the generated foraminifera (Fig. 10). This was done to ensure that the model converged (Fig. 9) as well as that the model did not collapse generating a single foraminifera specimen (Fig. 10). The FID score achieved for the generator at the species level was 14.88. The score at the unconditioned level is 32.32, suggesting that the model performs better when conditioned to generate specific foraminifera species than unconditioned.

IV. DISCUSSION AND CONCLUSIONS

In addition to the fossil classification and segmentation tasks, the current development of deep learning architectures in using generative adversarial networks (GAN) and Vision Transformers (ViT) to generate and restore geological image datasets automatically has fueled the interest to explore and unravel the potential of these models in geosciences. One of the major tasks is image super-resolution (i.e., upscaling image resolution with deep learning), which is instrumental in accurately classifying various geological analyses. Hence, it is essential to devise a workflow that could bring low-resolution images or images with a considerable amount of noise to better quality, directly translating into saving resources when obtaining datasets. In addition, this allows legacy datasets to be re-used and accessible for newer algorithms. On the other hand, generating synthetic data allows for replicating both aesthetic and statistical characteristics of a set of geoscientific datasets, allowing data

to be modeled, visualized, and augmented for subsequent geological workflows [15], [37], [38], [39], [43], [44].

This work seeks to find alternatives to classical microfossil image scaling algorithms, managing to train for the first time a specific architecture, SwinIR. For this purpose, its performance is better than classical interpolations such as Lanczos, bicubic, and bilinear when compared using standard metrics such as PSNR. This demonstrates the potential of this our proposed model to surpass traditional approaches for image restoration, even when this algorithm is trained with a dataset of moderate size, 3263 images, by computer vision standards [45], while also being of specific scientific importance.

In addition, the StyleGAN2 architecture is trained, opening the possibility of generating foraminifera specimens and experimenting with the associated latent space, obtaining a model that reaches FID scores less than 20, being close to the values reported for other geological datasets, which are capable of being indistinguishable from natural images to the trained eye [15]. These foraminifera were visually compared between some selected species. However, more detailed analysis with experts in the field is required to know how well it replicates morphotypes between different species. The trained model can generate images of foraminifera at the species level, and also, another model is trained for the unconditional generation of foraminifera. A possible use of this tool is increasing the diversity of micropaleontological datasets and releasing confidential information as a synthetic copy of the original dataset.

The model's ability to generate synthetic samples can also be applied to few-shot segmentation tasks and learning how to generate weights that can be used for other downstream tasks, such as semantic segmentation [46], [47]. Our study demonstrated the potential of using the proposed workflow for automatically segmenting foraminifera chambers (Fig. 11).

For this task, the segmentation of foraminifera chambers was performed by only labeling five samples which allowed a Few Shots CNN to learn and segment both the aperture and the contour of a specimen, in this case, a synthetic *Globigerina bulloides.*

APPENDIX A
DATA AVAILABILITY

The data used are available from the cited publication Marchant et al. [3], the preprocessed dataset and the synthetic self-labeled datasets are available upon reasonable request from the authors.

The models, sample synthetic images, and images used to evaluate the super-resolution model are available upon request.

ACKNOWLEDGMENT

The authors would like to thank the Editor and the reviewers of the early versions of this paper for their valuable contributions to the manuscript. They appreciate a correction from Prof. Bridget Wade on the foraminifera species naming.

REFERENCES

[1] B. Sen, *Modern Foraminifera*, 1st ed. Dordrecht, The Netherlands: Kluwer, 2003.

[2] R. Jones, *Foraminifera and Their Applications*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[3] R. Marchant, M. Tetard, A. Pratiwi, M. Adebayo, and T. de Garidel-Thoron, "Automated analysis of foraminifera fossil records by image classification using a convolutional neural network," *J. Micropalaeontol.*, vol. 39, no. 2, pp. 183–202, Oct. 2020, doi: 10.5194/jm-39-183-2020.

[4] L. Beaufort and D. Dollfus, "Automatic recognition of coccoliths by dynamical neural networks," *Mar. Micropaleontol.*, vol. 51, nos. 1–2, pp. 57–73, Apr. 2004, doi: 10.1016/j.marmicro.2003.09.003.

[5] P. Culverhouse, R. Simpson, R. Ellis, J. Lindley, R. Williams, T. Parisini, B. Reguera, I. Bravo, R. Zoppoli, G. Earnshaw, H. McCall, and G. Smith, "Automatic classification of field-collected dinoflagellates by artificial neural network," *Mar. Ecol. Prog. Ser.*, vol. 139, nos. 1–3, pp. 281–287, 1996, doi: 10.3354/meps139281.

[6] S. Liu, M. Thonnat, and M. Berthod, "Automatic classification of planktonic foraminifera by a knowledge-based system," in *Proc. 10th Conf. Artif. Intell. Appl.*, 1994, pp. 358–364.

[7] A. Pedraza, G. Bueno, O. Deniz, G. Cristóbal, S. Blanco, and M. Borrego-Ramos, "Automated diatom classification (Part B): A deep learning approach," *Appl. Sci.*, vol. 7, no. 5, p. 460, May 2017, doi: 10.3390/app7050460.

[8] R. Mitra, T. M. Marchitto, Q. Ge, B. Zhong, B. Kanakiya, M. S. Cook, J. S. Fehrenbacher, J. D. Ortiz, A. Tripati, and E. Lobaton, "Automated species-level identification of planktic foraminifera using convolutional neural networks, with comparison to human performance," *Mar. Micropaleontol.*, vol. 147, pp. 16–24, Mar. 2019, doi: 10.1016/j.marmicro.2019.01.005.

[9] L. E. Carvalho, G. Fauth, F. S. Baecker, G. Krahl, A. C. Moreira, C. P. Fernandes, and A. von Wangenheim, "Automated microfossil identification and segmentation using a deep learning approach," *Mar. Micropaleontol.*, vol. 158, Jun. 2020, Art. no. 101890, doi: 10.1016/j.marmicro.2020.101890.

[10] A. Koeshidayatullah, M. Morsilli, D. J. Lehrmann, K. Al-Ramadan, and J. L. Payne, "Fully automated carbonate petrography using deep convolutional neural networks," *Mar. Petroleum Geol.*, vol. 122, Dec. 2020, Art. no. 104687, doi: 10.1016/j.marpetgeo.2020.104687.

[11] T. H. Johansen, S. A. Sørensen, K. Møllersen, and F. Godtliebsen, "Instance segmentation of microscopic foraminifera," *Appl. Sci.*, vol. 11, no. 14, p. 6543, Jul. 2021, doi: 10.3390/app11146543.

[12] M. Ho, S. Idgunji, J. L. Payne, and A. Koeshidayatullah, "Hierarchical multi-label taxonomic classification of carbonate skeletal grains with deep learning," *Sedimentary Geol.*, vol. 443, Jan. 2023, Art. no. 106298, doi: 10.1016/j.sedgeo.2022.106298.

[13] D. Shoji, R. Noguchi, S. Otsuki, and H. Hino, "Classification of volcanic ash particles using a convolutional neural network and probability," *Sci. Rep.*, vol. 8, no. 1, pp. 1–12, May 2018, doi: 10.1038/s41598-018-26200-2.

[14] R. P. de Lima and D. Duarte, "Pretraining convolutional neural networks for mudstone petrographic thin-section image classification," *Geosciences*, vol. 11, no. 8, p. 336, Aug. 2021, doi: 10.3390/GEOSCIENCES11080336.

[15] I. Ferreira, L. Ochoa, and A. Koeshidayatullah, "On the generation of realistic synthetic petrographic datasets using a style-based GAN," *Sci. Rep.*, vol. 12, no. 1, Jul. 2022, Art. no. 12845, doi: 10.1038/s41598-022-16034-4.

[16] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," 2017, *arXiv:1712.04621*.

[17] A. Abdellatif, A. H. Elsheikh, G. Graham, D. Busby, and P. Berthet, "Generating unrepresented proportions of geological facies using generative adversarial networks," *Comput. Geosci.*, vol. 162, May 2022, Art. no. 105085, doi: 10.1016/j.cageo.2022.105085.

[18] A. Koeshidayatullah, "Optimizing image-based deep learning for energy geoscience via an effortless end-to-end approach," *J. Petroleum Sci. Eng.*, vol. 215, Aug. 2022, Art. no. 110681, doi: 10.1016/j.petrol.2022.110681.

[19] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," 2017, *arXiv:1706.03762*, doi: 10.48550/ARXIV.1706.03762.

[20] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.
[21] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using swin transformer," 2021, *arXiv:2108.10257*.
[22] Z. Lin, P. Garg, A. Banerjee, S. A. Magid, D. Sun, Y. Zhang, L. Van Gool, D. Wei, and H. Pfister, "Revisiting RCAN: Improved training for image super-resolution," 2022, *arXiv:2201.11279*.
[23] A. Parnami and M. Lee, "Learning from few examples: A summary of approaches to few-shot learning," Mar. 2022, *arXiv:2203.04291*.
[24] O. Vinyals, "Model vs optimization meta learning," in *Proc. Meta-Learn. Symp. NIPS*, 2017.
[25] A. Y. Hsiang et al., "Endless Forams: >34,000 modern planktonic foraminiferal images for taxonomic training and automated species recognition using convolutional neural networks," *Paleoceanography Paleoclimatology*, vol. 34, no. 7, pp. 1157–1177, 2019, doi: 10.1029/2019PA003612.
[26] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," 2021, *arXiv:2103.14030*.
[27] A. Koeshidayatullah, S. Al-Azani, E. E. Baraboshkin, and M. Alfarraj, "FaciesViT: Vision transformer for an improved core lithofacies prediction," *Frontiers Earth Sci.*, vol. 10, Oct. 2022, Art. no. 992442, doi: 10.3389/feart.2022.992442.
[28] I. Ferreira, A. Koeshidayatullah, F. Baharudin, S. Yusmananto, and S. Allam, "PoreFormer: A novel microporosity characterization in mudstone using a deep learning vision transformer approach," in *Proc. 84th EAGE Annu. Conf. Exhib., Eur. Assoc. Geoscientists Eng.*, Jun. 2023, pp. 1–5, doi: 10.3997/2214-4609.202310211.
[29] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, vol. 2, 2014, pp. 2672–2680.
[30] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," 2018, *arXiv:1812.04948*.
[31] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," 2020, *arXiv:2006.06676*.
[32] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," 2019, *arXiv:1912.04958*.
[33] T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, and T. Aila, "Alias-free generative adversarial networks," 2021, *arXiv:2106.12423*.
[34] F. Lago, C. Pasquini, R. Böhme, H. Dumont, V. Goffaux, and G. Boato, "More real than real: A study on human visual perception of synthetic faces," 2021, *arXiv:2106.07226v2*, doi: 10.1109/MSP.2021.3120982.
[35] S. J. Nightingale and H. Farid, "AI-synthesized faces are indistinguishable from real faces and more trustworthy," *Proc. Nat. Acad. Sci. USA*, vol. 119, no. 8, Feb. 2022, Art. no. e2120481119, doi: 10.1073/pnas.2120481119.
[36] A. Sauer, K. Schwarz, and A. Geiger, "StyleGAN-XL: Scaling StyleGAN to large diverse datasets," 2022, *arXiv:2202.00273*.
[37] M. Bizhani, O. H. Ardakani, and E. Little, "Reconstructing high fidelity digital rock images using deep convolutional neural networks," *Sci. Rep.*, vol. 12, no. 1, p. 4264, Mar. 2022, doi: 10.1038/s41598-022-08170-8.
[38] Y. Niu, Y. D. Wang, P. Mostaghimi, P. Swietojanski, and R. T. Armstrong, "An innovative application of generative adversarial networks for physically accurate rock images with an unprecedented field of view," *Geophys. Res. Lett.*, vol. 47, no. 23, Dec. 2020, Art. no. e2020GL089029, doi: 10.1029/2020GL089029.
[39] Y. Liu, C. Guo, J. Cao, Z. Cheng, X. Ding, L. Lv, F. Li, and M. Gong, "A new resolution enhancement method for sandstone thin-section images using perceptual GAN," *J. Petroleum Sci. Eng.*, vol. 195, Dec. 2020, Art. no. 107921, doi: 10.1016/j.petrol.2020.107921.
[40] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," 2017, *arXiv:1706.08500*.
[41] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015, *arXiv:1512.00567*.
[42] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.
[43] L. Mosser, O. Dubrule, and M. J. Blunt, "Reconstruction of three-dimensional porous media using generative adversarial neural networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 96, no. 4, Oct. 2017, Art. no. 043309, doi: 10.1103/PhysRevE.96.043309.
[44] T. Nanjo and S. Tanaka, "Carbonate lithology identification with generative adversarial networks," in *Proc. Int. Petroleum Technol. Conf.*, 2020.
[45] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. Comput. Vis. Pattern Recognit.*, 2009.
[46] N. Tritrong, P. Rewatbowornwong, and S. Suwajanakorn, "Repurposing GANs for one-shot semantic part segmentation," 2021, *arXiv:2103.04379*.
[47] I. Ferreira and A. Koeshidayatullah, "Generating self-labeled geological datasets for semantic segmentation using pretrained GANs," in *Proc. EGU Gen. Assem.*, Vienna, Austria, Apr. 2023, doi: 10.5194/egusphere-egu23-888.2023.

IVAN FERREIRA-CHACUA received the Bachelor of Science degree from the National University of Colombia, Bogotá. He is currently pursuing the Master of Science degree in geology with the King Fahd University of Petroleum and Minerals with a focus on computer vision applications in the geosciences for automatic petrographic analysis and data science applied to the earth sciences. With experience in data science applications to geological problems.

ARDIANSYAH IBNU KOESHIDAYATULLAH is currently an Assistant Professor with the Geosciences Department, College of Petroleum Engineering and Geosciences (CPG), KFUPM. His research interests include integrated carbonate reservoir characterization, including stratigraphic modeling and novel geochemistry and artificial intelligence in geosciences. He was a Presidential Doctoral Scholar with The University of Manchester, U.K., and a Visiting Research Student with the Rosenstiel School of Marine and Atmospheric Sciences, University of Miami, USA. Before joining KFUPM, he was a Joint Postdoctoral Research Fellow with the Paleobiology and Basin Petroleum System Modeling (BPSM) Group, Stanford University, USA.