Intelligent Geoengineering 3 (2026) 1–10

Contents lists available at ScienceDirect

Intelligent Geoengineering

journal homepage: www.keaipublishing.com/en/journals/ige/

Full length article

Generalizable digital rock image segmentation under limited data with the segment anything model

Ziqiang Wang[a], Zhiyu Hou[a,b], Shuai Hou[a], Danping Cao[a,*]

[a] State Key Laboratory of Deep Oil and Gas, China University of Petroleum (East China), Qingdao 266580, China
[b] Institute of Earth Sciences, University of Lausanne, Lausanne CH-1015, Switzerland

ARTICLE INFO

Keywords:
Digital rock physics
Image segmentation
Segment Anything Model

ABSTRACT

Accurate segmentation of digital rock images is essential for characterizing pore–matrix systems and predicting petrophysical properties. However, the diversity of rock textures across different lithologies poses a significant challenge for conventional segmentation networks, especially under limited training data. To address this, we introduce DRI-SAM (Digital Rock Image – Segment Anything Model), a hybrid segmentation framework that leverages the powerful visual prior of the Segment Anything Model (SAM) and adapts it to the digital rock domain. Specifically, we apply LoRA-based fine-tuning to SAM's image encoder to better capture rock-specific microstructures, while U-Net is employed to generate prompt points, guiding SAM toward accurate pore–matrix delineation. This approach retains the encoder's representational power while allowing domain-specific adaptation via LoRA, enabling effective cross-domain generalization under limited training data. The model is trained exclusively on 200 annotated images of Bentheimer sandstone, covering two distinct voxel resolutions, and is evaluated on digital rock images of varying lithologies, resolutions and imaging modalities. The results confirm that DRI-SAM achieves accurate segmentation on both sandstone and more challenging carbonate samples, including synthetic and SEM images, without additional retraining or parameter adjustments. Compared to DeepLabV3 + and the only LoRA-tuned SAM, DRI-SAM demonstrates superior performance under limited supervision, highlighting its strong generalization and practical value in digital rock image analysis. Moreover, the findings suggest that foundation models like SAM, when properly adapted, also hold great promise for broader geoscientific imaging tasks.

1. Introduction

Accurate characterization of subsurface materials underpins a wide range of geoscience and engineering applications, spanning hydrocarbon recovery, groundwater management, carbon storage, and geothermal energy (Blunt et al., 2013; Li et al., 2023; Lubis and Harith, 2014; Shan et al., 2024). The ability of researchers to reliably predict subsurface behavior is fundamentally contingent on a precise understanding of its microstructure (Dvorkin et al., 2009). However, traditional experimental methods are often hindered by their high cost, significant time demands, and limited capacity to resolve critical pore-scale details (Chi et al., 2024; Hou et al., 2023b; Karimpouli and Tahmasebi, 2019a; Soltanmohammadi et al., 2024). Over the past decade, digital rock physics has emerged as a transformative paradigm, integrating high-resolution imaging techniques (e.g., micro-CT, FIB-SEM) with numerical simulation and data-driven modeling to bridge microscopic structures with macroscopic properties (Al-Marzouqi, 2018; Alkhimenkov, 2025; Andrä et al., 2013a, 2013b; Madonna et al., 2012). Within this framework, a central challenge lies in processing, analyzing, and interpreting digital images, transforming their intricate visual information into quantitative models that can underpin robust physical predictions.

A pivotal step in this transformation is image segmentation — the process of assigning each voxel in a 3D volume or each pixel in a 2D image to specific material classes such as pores and matrix. Only through precise segmentation can researchers reliably: (1) Quantify essential pore-scale structural properties, such as porosity, pore size distribution, pore connectivity (percolation), pore throat geometry, and

Peer review under responsibility of Chinese Society for Rock Mechanics and Engineering
* Corresponding author at: China University of Petroleum (East China), Qingdao 266580, China.
E-mail address: caodp@upc.edu.cn (D. Cao).

https://doi.org/10.1016/j.ige.2025.12.001
Received 5 October 2025; Received in revised form 27 November 2025; Accepted 9 December 2025
Available online 10 December 2025

specific surface area – parameters crucial for understanding fluid storage and flow capacity (Arns et al., 2002; Purswani et al., 2020; Reinhardt et al., 2022); (2) Simulate fundamental transport phenomena, including single and multi-phase fluid flow (permeability), electrical conductivity, and molecular diffusion, enabling predictions of reservoir performance and recovery efficiency (Iraji et al., 2023; Soltanmohammadi et al., 2024; Wang and Zai, 2023; Yang et al., 2023); and (3) Predict key elastic and mechanical rock properties, such as bulk and shear moduli, seismic velocities, compressibility, and strength, bridging the gap between microstructure and geomechanical response (Andhumoudine et al., 2021; Cao et al., 2022; Cui et al., 2021; Hayatdavoudi et al., 2025; Hou and Cao, 2022; Karimpouli et al., 2018; Saxena and Mavko, 2016). Thus, segmentation quality is not merely a preliminary step; it is the foundational bottleneck upon which the validity and predictive power of all downstream digital rock physics analyses critically depend (Hou et al., 2023a; Wang et al., 2026; Ye et al., 2025).

However, the segmentation of digital rock images is often hindered by limitations in image quality and the significant diversity of rock textures across different lithologies and imaging resolutions (Hou et al., 2021; Ibrahim et al., 2020; Karimpouli and Tahmasebi, 2019b). High-performance segmentation networks, predominantly based on deep learning architectures like U-Net (Ronneberger et al., 2015), typically require large amounts of annotated data for training, which is both expensive and labor-intensive in the digital rock domain due to the need for expert geological knowledge. Moreover, a core challenge lies in achieving robust model generalization. As a result, achieving accurate and generalizable segmentation across different lithologies and imaging resolutions remains exceptionally challenging under limited training data availability, often necessitating resource-intensive retraining or parameter tuning for new datasets.

Recently, the emergence of general-purpose vision foundation models such as the Segment Anything Model (SAM) has opened new possibilities for segmentation across diverse image domains (Kirillov et al., 2023). SAM demonstrates impressive zero-shot performance on natural images by leveraging strong structural priors and prompt-based interaction. However, directly applying SAM to digital rock images presents substantial domain-specific challenges. Ma et al. (2023) first applied SAM to digital rock segmentation and showed that weak contrast often leads to segmentation failure, while direct fine-tuning can partially alleviate this issue. Wang et al. (2025b) further proposed EG-SAM, which enhances SAM by fusing encoder-derived semantic embeddings with edge features to improve boundary sensitivity. However, both approaches rely solely on direct encoder fine-tuning and do not introduce additional strategies for domain adaptation. SAM relies heavily on clear visual boundaries to produce accurate masks, whereas boundaries in CT images of rocks are often ambiguous. This is because grayscale values in CT do not directly reflect the absolute density or specific material composition; instead, they result from complex interactions of X-rays with heterogeneous samples. Consequently, grayscale only indicates relative contrasts and cannot reliably distinguish between solid matrix and pore spaces. Moreover, when voxel sizes exceed the scale of microstructural features, a single voxel may represent a mixture of minerals and pores, which further blurs boundary definitions. In addition, prompt-based segmentation using sparse points is less effective in digital rock images. A single field of view may contain numerous pore regions, making one-point prompts insufficient. Automatic point generation via grid sampling often fails to provide meaningful guidance, while manually crafted prompts can yield better results but are time-consuming to prepare. As illustrated in Fig. 1, both direct use of SAM and point-based prompting exhibit limitations when applied to digital rock image segmentation.

To address the aforementioned limitations, we propose DRI-SAM (Digital Rock Image – Segment Anything Model), a segmentation framework that adapts SAM to digital rock analysis through encoder fine-tuning and domain-specific prompt generation. Since segmentation accuracy hinges on the encoder's ability to extract meaningful features, we fine-tune the SAM encoder using LoRA to enhance its sensitivity to microstructural textures often overlooked by models trained on natural images. Furthermore, we reformulate prompt generation as a sparse point prediction task: using U-Net, we predict representative foreground points derived from the centers of maximum inscribed spheres within binary-labeled pore regions. This design enables DRI-SAM to achieve accurate and transferable segmentation across varying lithologies and imaging settings.

2. Methodology

Building on the original Segment Anything Model architecture, DRI-SAM incorporates LoRA-based fine-tuning of the encoder and introduces a U-Net network for automated prompt point generation, with the overall framework illustrated in Fig. 2. Section 2.1 provides a brief overview of the SAM model and its key components. Section 2.2 describes the LoRA fine-tuning process applied to the encoder in detail. Section 2.3 presents the U-Net-based point generation network, including the design of point labels for automated prompt acquisition. Finally, Section 2.4 outlines the dataset and experimental setup, covering Bentheimer sandstone samples at two voxel resolutions as well as relevant hyperparameter configurations.

2.1. Overview of Segment Anything Model (SAM)

The Segment Anything Model (SAM) is a general-purpose image segmentation model that combines interactive and automated segmentation. Pretrained on the large-scale SA-1B dataset, which includes 11 million images and over 1 billion high-quality masks, SAM enables robust zero-shot generalization across diverse domains. Its architecture consists of three core components: an image encoder, a prompt encoder, and a mask decoder:

1) The image encoder is built upon a Vision Transformer (ViT) architecture that has been pretrained using a masked autoencoding (MAE) strategy. This pretraining approach enhances the encoder's ability to capture comprehensive contextual and structural information from input imagery. The encoder incorporates a hybrid attention mechanism that effectively combines both local and global attention. Local attention focuses on fine-grained details and texture characteristics, while global attention captures broader structural patterns and long-range dependencies. This dual attention strategy enables the model to accurately recognize small-scale features as well as large-scale compositional layouts.
2) The prompt encoder provides flexible user interaction, supporting sparse prompts (such as points, boxes, or clicks) as well as dense prompts (such as masks). This design allows SAM to adapt dynamically to different segmentation tasks and user input, making it versatile for interactive and automated workflows.
3) The mask decoder synthesizes feature embeddings from the image and prompt encoders through a lightweight transformer-based architecture. It generates high-fidelity, pixel-accurate segmentation masks by integrating multi-scale feature representations. This multi-scale processing helps preserve fine details and object boundaries while ensuring overall structural coherence and spatial consistency in the output segmentation.

SAM's modular architecture and strong visual priors, gained from large-scale pretraining, enable efficient transfer to specialized domains, supporting both zero-shot generalization and scenarios with limited annotated data. This makes it a powerful foundation for digital rock image analysis, where high-quality labeled samples are rare and manual annotation is labor-intensive. By leveraging SAM's pretrained feature representations, it becomes possible to achieve accurate segmentation of complex pore–matrix systems without requiring extensive

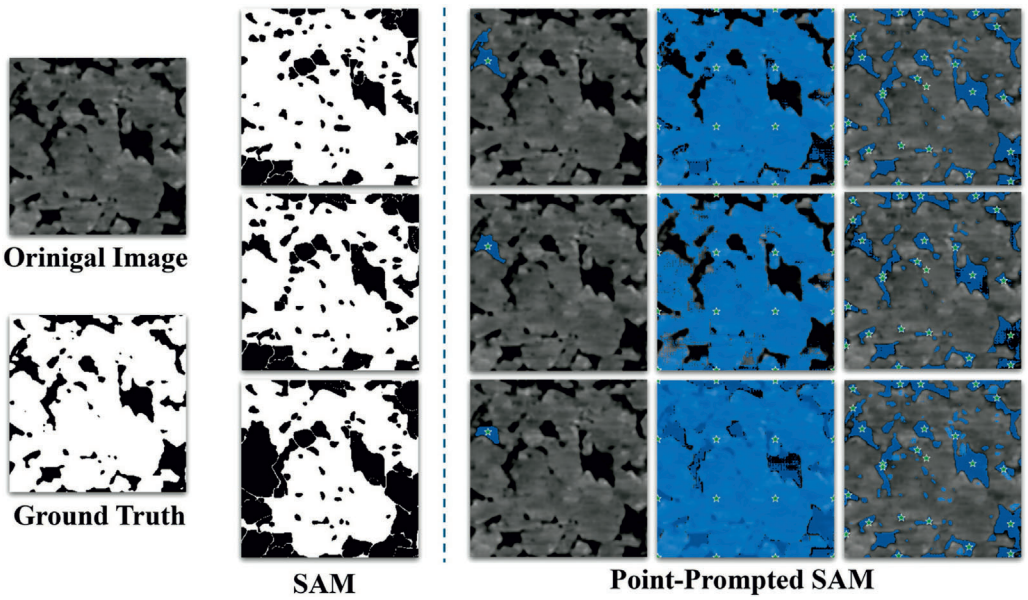

Fig. 1. Segmentation results using the Segment Anything Model (SAM) on digital rock images. Left: From top to bottom are the results of direct SAM inference using three variants: ViT-B, ViT-L, and ViT-H. Right: From left to right are the multimask outputs guided by different types of point prompts: a single point, automatically generated grid-based points, and manually annotated points.

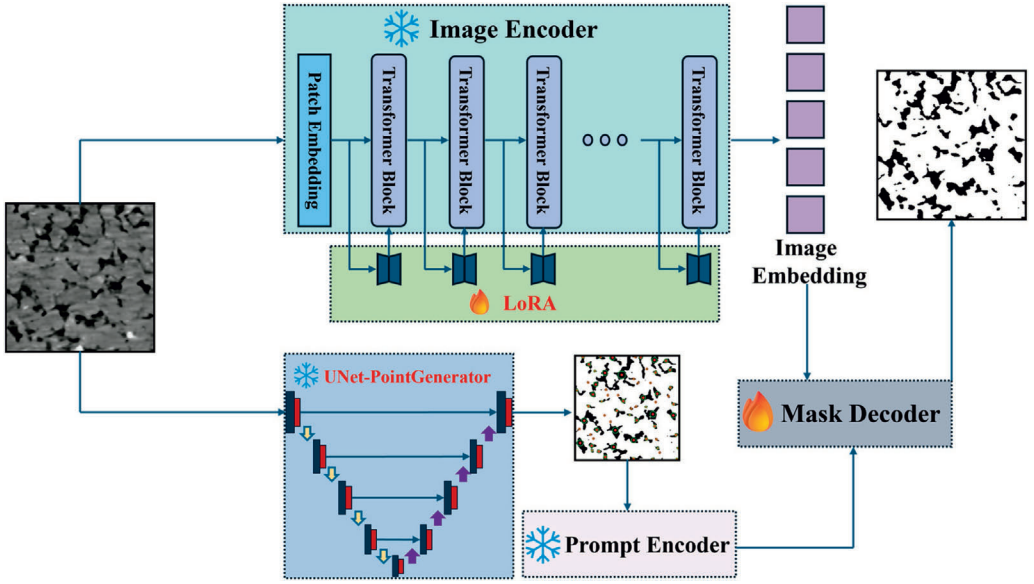

Fig. 2. Overall framework of DRI-SAM, where the newly added modules are highlighted in red.

domain-specific annotations. However, despite these advantages, direct application of SAM to digital rock images may still struggle to capture subtle microstructural details arising from specific imaging modalities, highlighting the need for domain-adaptive strategies, such as LoRA-based encoder fine-tuning.

2.2. Lora-based fine-tuning method

Fine-tuning large-scale pretrained models has become a common strategy for adapting general-purpose architectures to domain-specific tasks (Gao et al., 2026; Guo et al., 2025). Common approaches include full fine-tuning of both the encoder and decoder (Ma et al., 2024), and freezing the encoder while fine-tuning only the decoder (Hu et al., 2023). The former maximizes task-specific learning capacity but requires large amounts of annotated data and substantial computational resources. The latter allows the model to leverage the feature extraction capabilities acquired from large-scale pretraining while adapting to new tasks, but it may limit the refinement of segmentation accuracy. In the context of digital rock images, full fine-tuning would necessitate the collection of extensive sample–label pairs and high computational cost, which has not been widely pursued. On the other hand, decoder-only fine-tuning enables SAM to adapt to the digital rock domain, but it often lacks sufficient fine-grained representation.This limitation arises because the encoder, pretrained on general images, does not fully perceive the characteristics of specialized imaging modalities such as micro-CT, an issue that will be further analyzed and discussed in subsequent comparisons. It is also worth noting that the image encoder contains the vast majority of SAM's parameters, whereas the mask decoder is designed to be lightweight. As a result, most existing studies focus on how to efficiently fine-tune the encoder, while the decoder can be fully fine-tuned at minimal computational cost.

To balance these trade-offs, LoRA (Low-Rank Adaptation) has recently emerged as an effective technique for parameter-efficient adaptation of large models. By injecting low-rank learnable matrices into pretrained weight layers, LoRA allows the model to adapt to new domains without updating the entire set of parameters, preserving the general visual priors while focusing on domain-specific features. In the context of digital rock image segmentation, LoRA-based fine-tuning

provides an efficient and effective solution for improving segmentation performance while minimizing the computational burden, enabling the encoded features to capture the unique textural patterns of digital rocks.

Specifically, each transformer block's multi-head self-attention module contains a combined query-key-value (QKV) projection layer, originally parameterized by a weight matrix $W \in R^{C_{out} \times C_{in}}$. We replace this layer with a LoRA-augmented module, decomposing the fine-tuning update ΔW as the product of two low-rank matrices $A \in R^{r \times C_{in}}$ and $B \in R^{C_{out} \times r}$, where r $\ll \min(C_{in}, C_{out})$. The adapted weight becomes:

$$\hat{W} = W + \Delta W = W + BA \tag{1}$$

This update is applied simultaneously and independently to the query, key, and value projections by enabling LoRA on all three components of the fused QKV layer. Furthermore, the feed-forward network (MLP) within each transformer block contains two linear layers, both of which are similarly augmented with LoRA modules, enabling the network to better capture the microstructural nuances of digital rock images.

During fine-tuning, the original SAM weights W remain frozen, and only the low-rank matrices A and B are optimized, ensuring parameter efficiency. Formally, the attention mechanism with LoRA-augmented projections is given by:

$$Att(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \tag{2}$$

where the adaped queries, keys, and values are computed as $Q = \hat{W}_q F = W_q F + B_q A_q F$, $K = \hat{W}_k F = W_k F + B_k A_k F$, $V = \hat{W}_v F = W_v F + B_v A_v F$, with $W_{\{\cdot\}}$ denoting frozen original projection weights and $A_{\{\cdot\}}$ and $B_{\{\cdot\}}$ trainable LoRA parameters. This fine-grained LoRA injection enables the SAM image encoder to adapt to domain-specific textures and structures in digital rock images effectively, while maintaining the robustness of the pretrained model and minimizing training cost.

2.3. UNet-based automated prompt point generation

Prompt-based segmentation with SAM relies on sparse, user-provided input points to guide mask generation. However, for digital rock images, formulating meaningful prompts is inherently challenging. A single 2D field of view often contains numerous disconnected pore regions, making manual annotation impractical and random sampling strategies largely ineffective. Moreover, the fine-scale heterogeneity and irregular shapes of pores in micro-CT or SEM images further complicate the selection of representative points.

These challenges motivate us to recast prompt generation as a sparse point prediction problem, where each prompt corresponds to a representative point within an individual pore. In this formulation, the task can be viewed analogously to instance-level center point detection or minimal-point semantic labeling. By framing prompt generation in this manner, it becomes possible to automate the process, reduce human effort, and ensure that each pore is adequately represented during segmentation. In practice, a standard U-Net architecture is employed to accomplish this sparse point prediction task.

To obtain reliable training labels for this task, we leverage the binary ground-truth masks and extract geometrically meaningful center points for each pore region (as illustrated in Fig. 3). Specifically, we first perform connected-component labeling on the pore space (i.e., background in binary masks), and compute the Euclidean distance transform within each connected pore region, where pixels with higher values (shown in red) are farther from the boundary and highlight the geometric center of the pore. The point with the maximum distance from the boundary corresponds to the center of the largest inscribed sphere within that region, offering a natural and robust choice for sparse prompting. This center point is not only spatially stable but also inherently captures the geometric core of each pore body.

Formally, let $\Omega \subset R^2$ denote the binary pore mask, and let $R_i \subset \Omega$ be the i-th connected pore region. We compute the distance transform $D_i(x) = \min_{y \in \delta R_i} \|x - y\|_2$ for all $x \in R_i$. The maximum of $D_i(x)$ determines the location of the center point $p_i \in R_i$ such that:

$$p_i = arg\, max_{x \in R_i} D_i(x) \tag{3}$$

where p_i defines the center of the maximal inscribed disk in R_i. The collection of such points $\{p_i\}$ over all pore regions provides the target coordinates for training the point prediction module.

The network employed for sparse point prediction is a 2D UNet consisting of an encoder-decoder architecture with skip connections. The encoder progressively extracts multi-scale features through four convolutional blocks with an increasing number of channels: 8, 16, 32, and 64, each block consisting of two consecutive convolutional layers, each followed by batch normalization and ReLU activation, and followed by strided convolutions for downsampling. each followed by strided convolutions for downsampling. The decoder restores spatial resolution via transposed convolutions and concatenates features from the corresponding encoder layers, followed by convolutional blocks to refine the feature maps. The final output layer is a single-channel convolution that produces a pixel-wise probability map of pore center locations. The network is pre-trained prior to deployment, allowing efficient inference of sparse prompts from new images without human input. The automated nature of this labeling strategy eliminates the need for manual annotations or heuristic sampling and naturally aligns with SAM's point-based prompt interface.

2.4. Data and training details

Generally, the performance of neural networks improves with increasing training data, and conventional segmentation networks often require thousands of images to learn effective representations. In contrast, this study aims to demonstrate the advantages of fine-tuning large pretrained models under limited data conditions, and therefore, we use only 200 annotated images for training. This design highlights the capability of large models to adapt effectively in a weakly supervised setting. Furthermore, we select a single lithology, Bentheimer sandstone, as the training material. Characterized by high porosity, strong pore connectivity, and low clay content, this sandstone exhibits a relatively homogeneous internal structure, making it a widely used benchmark material in both experimental and digital rock studies. By choosing this lithology, we leverage its well-characterized and uniform microstructure for training, while employing two voxel resolutions

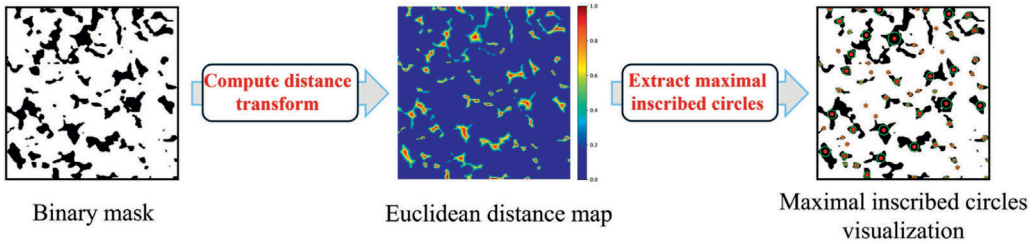

Fig. 3. Workflow for generating pore center point labels. A binary pore mask is processed to compute the Euclidean distance map for each connected pore region, where the intensity of each pixel indicates its distance to the nearest boundary and highlights the geometric center of the pore. The geometrically meaningful centers are then identified as the maximal inscribed circles, with the circle centers serving as reliable sparse point labels.

(1.657314 μm and 2.25 μm) ensures that the model is exposed to slightly varying spatial scales. Although the training dataset consists of a single lithology, the test dataset includes a wider variety of rock types, voxel resolutions, and imaging modalities to comprehensively evaluate the robust cross-domain generalization of DRI-SAM. A detailed analysis of these experimental results is presented in Section 3.

The model is trained for 100 epochs using the Adam optimizer with a learning rate of 5×10^{-4}, with each epoch taking approximately 25 s. The loss function is BCEWithLogitsLoss, which is suitable for binary segmentation tasks. For a predicted mask p and ground-truth mask g the loss is defined as:

$$loss = -\frac{1}{N}\sum_{i=1}^{n}\left[g_i \cdot \log\sigma(p_i) + (1-g_i)\cdot\log(1-\sigma(p_i))\right] \quad (4)$$

Considering that large pretrained models may require substantial adjustment when adapting to a new domain, an initially relatively high learning rate is used to enable rapid alignment of the encoder features with the digital rock images. As training progresses, the learning rate is gradually decreased to facilitate the learning of fine-grained microstructural details. Accordingly, a ReduceLROnPlateau scheduler is employed with patience $= 10$ and factor $= 0.5$, which reduces the learning rate when the validation loss plateaus. The U-Net network used for automated prompt point generation is trained on the same dataset as DRI-SAM and uses the BCEWithLogitsLoss loss function. A smaller learning rate of 1×10^{-4} and the Adam optimizer are used. A longer training schedule of 600 epochs is adopted because the target points occupy a small spatial region in each image and require more iterations to converge. The additional computational cost remains low due to the relatively small number of parameters in the U-Net.All experiments are conducted on two Nvidia A100 80 GB GPUs.

3. Results

After training, DRI-SAM is directly applied to various test datasets to evaluate its generalizable segmentation performance without any additional parameter tuning. Specifically, Section 3.1 presents segmentation results on eight different sandstone samples, Section 3.2 evaluates segmentation on carbonate rocks with varying voxel resolutions, and Section 3.3 examines performance across different imaging modalities, including SEM images and diffusion model–generated images. Section 3.4 further computes several elastic parameters based on the segmentation results, quantitatively assessing the impact of segmentation details on parameter estimation.

3.1. Generalizable segmentation on sandstones

Fig. 4 presents segmentation results on eight sandstone types (e.g., Bandera Brown, Bandera Gray, Berea, etc.), all with a voxel resolution of 2.25 μm, which is also included in the training set. For each sandstone type, segmentation performance was quantitatively evaluated on 100 images using the Dice coefficient. The results consistently high values across all samples, ranging from 0.9683 to 0.9881. Berea sandstone achieves the highest score, whereas Bandera Brown and Bandera Gray show relatively lower performance, likely due to their finer, more fragmented pores and complex pore–matrix interfaces. These characteristics are visually reflected in the segmentation maps, where the pores appear more intricate, disconnected, and irregularly shaped. Additionally, the presence of high-density minerals, which appear as bright regions in the original grayscale images, introduces further challenges for accurate mask delineation. In contrast, other sandstone types tend to feature larger, better-connected pores that are more easily separable, facilitating more precise segmentation.

The difference maps indicate that most deviations are localized at the pore–matrix interfaces, where the transition is often gradual and structurally complex. These discrepancies do not necessarily imply model failure, but rather reflect the inherent uncertainty and potential inaccuracy in the ground truth itself, particularly in regions where a definitive boundary is difficult to establish even under expert supervision.

Overall, the results preliminarily indicate that DRI-SAM generalizes well to various sandstone types at the same resolution, even under limited

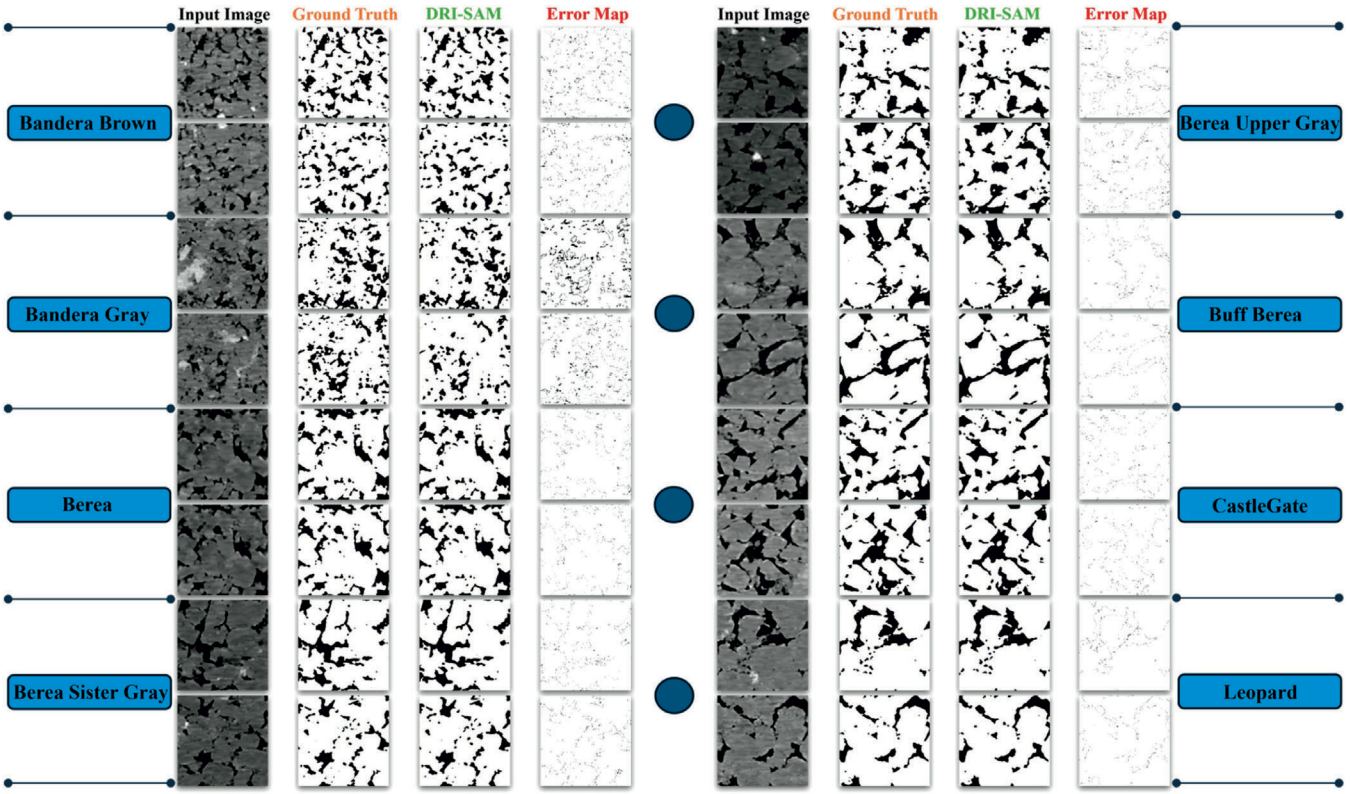

Fig. 4. Segmentation results of DRI-SAM on eight types of sandstone.

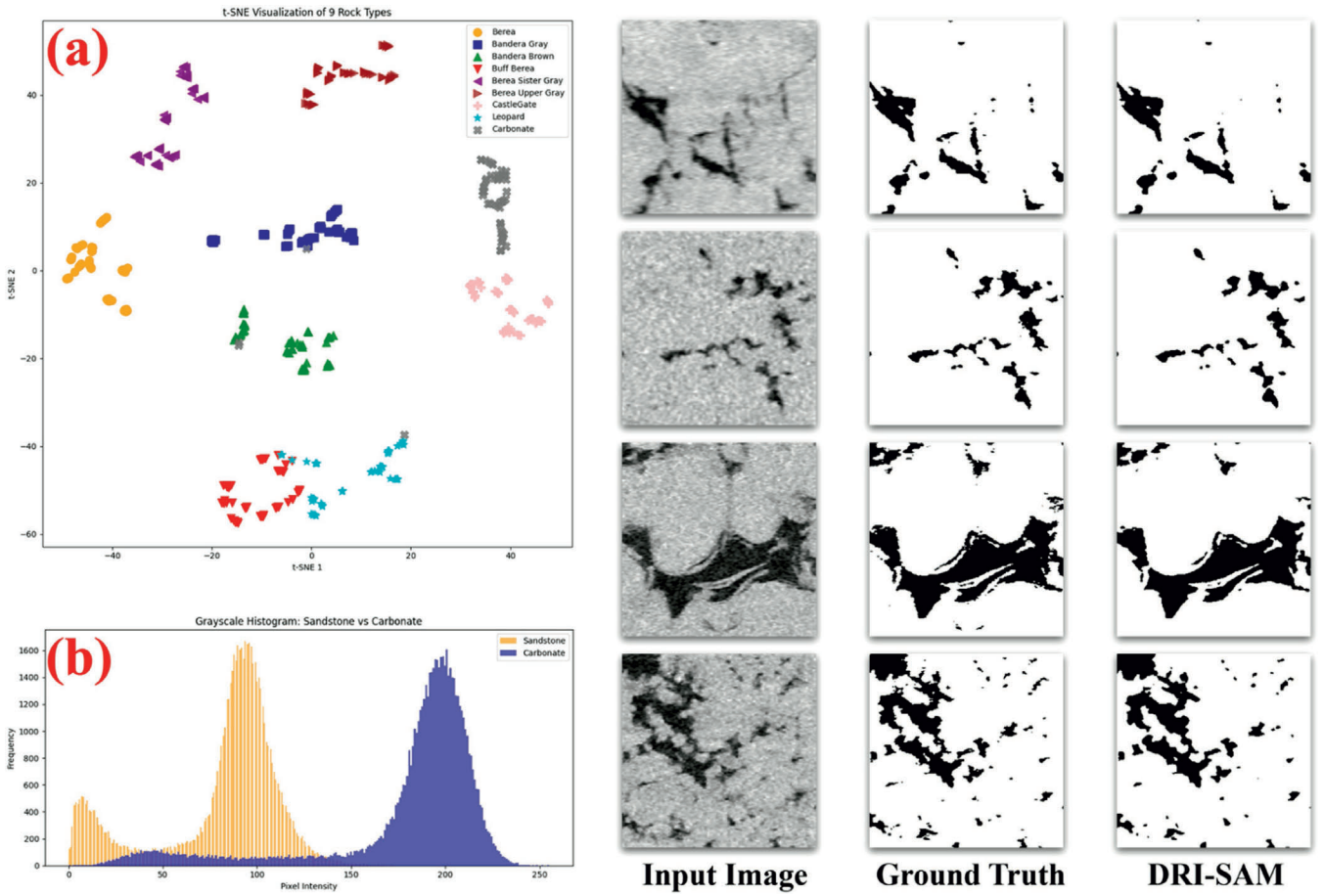

Fig. 5. Rock difference analysis and DRI-SAM segmentation results on carbonate samples: (a) *t*-SNE visualization showing distinct feature distributions between rock types; (b) Grayscale histogram highlighting intensity differences between sandstone and carbonate.

data. The combination of high Dice scores and visual consistency across multiple lithologies suggests that the model robustly captures micro-scale pore details, highlighting its potential utility for cross-lithology digital rock segmentation. Nevertheless, the observed variations underscore that certain complex pore morphologies and mineral heterogeneities remain challenging, pointing to avenues for further refinement or incorporation of additional domain-specific priors.

3.2. Generalizable segmentation on carbonates

Subsequently, the same DRI-SAM model was applied to carbonate rock samples with a voxel resolution of 8 μm. As illustrated in Fig. 5, panel (a) presents the t-SNE visualization of nine rock types, including eight sandstones and one carbonate. Different colors and marker shapes correspond to distinct lithologies. The results demonstrate that samples of the same lithology cluster closely together, indicating high intra-class feature similarity extracted by the model. Conversely, distinct lithologies exhibit well-separated clusters in the feature space, reflecting strong inter-class discriminability. This indicates that the selected image features successfully represent the intrinsic differences among rock types, highlighting the segmentation model's robust generalization performance when applied to diverse lithologies with distinct textural properties.

Panel (b) displays the grayscale histograms of sandstone and carbonate images. Sandstone exhibits two prominent peaks: a lower-intensity peak around 0–40 corresponding to pore spaces, indicating that pores generally appear as dark regions; and a distinct mid-intensity peak around 80–110 representing the rock matrix, suggesting a relatively uniform and homogeneous matrix grayscale distribution. In contrast, carbonate rock images are primarily distributed in a higher grayscale range approximately between 170 and 220. Their histogram is broader and smoother, lacking a pronounced low-intensity peak, which implies a smaller proportion of pore spaces or pores with grayscale values close to the matrix. This pattern reflects the carbonate's complex mineralogy and less distinct pore boundaries, potentially due to the presence of high-density mineral phases such as dolomite or calcite, which obscure clear structural delineation.

Despite these complexities, DRI-SAM accurately segments the carbonate rock images, as demonstrated in the right of Fig. 5. These results confirm that the model can generalize from training on homogeneous sandstone to segmenting heterogeneous carbonate lithologies, underscoring its strong cross-lithology generalization capability.

3.3. Generalizable segmentation across imaging modalities

To further evaluate the generalization capability of DRI-SAM, we tested it on four complex digital rock images without ground-truth labels: Savonnières limestone (voxel size: 3.8 μm), a mono-mineralic, calcitic, layered oolithic limestone from the Oolithe Vacuolaire formation in France; two diffusion model-generated sandstone images (Esmaeili, 2024); and a North Sea sandstone SEM image with 0.5 μm resolution. These samples contain intricate pore structures such as inter- and intra-oolithic microporosity and hollow ooliths with limited macro-pore connectivity. In addition, they differ significantly from the training data in imaging modality and texture. Specifically, the training set consists solely of micro-CT sandstone images, while the test set includes SEM images and synthetic data, posing a greater challenge for cross-domain generalization.

We compared DRI-SAM with two baselines: DeepLabV3+, a classical semantic segmentation network known for its strong performance across various segmentation tasks, and the original SAM enhanced with LoRA-based fine-tuning, applied without automated prompt point generation. As

Table 1
Presents the quantitative comparison across four segmentation metrics.

	Dice	IoU	Recall	Precision
DeepLabV3+	0.897	0.885	0.902	0.896
SAM	0.952	0.934	0.956	0.958
DRI-SAM	**0.971**	**0.952**	**0.974**	**0.988**

shown in Table 1, we calculated the average values of four evaluation metrics for the three models. Compared with SAM, DRI-SAM improves the IoU from 0.934 to 0.952. Although the numerical improvement is modest, the visual enhancement is more noticeable and reliable. This is because SAM already identifies most pore regions correctly, and the remaining gains mainly come from reducing a small number of local segmentation errors. As shown in Fig. 6 (highlighted in red), DRI-SAM consistently produces the most accurate segmentation, effectively capturing subtle pore structures and fine-scale connectivity. In contrast, DeepLabV3+ struggles to generalize when trained on a single annotated dataset and applied to previously unseen rock types, often failing to identify many pores. The LoRA-tuned SAM performs reasonably well, but still lacks sufficient structural detail in critical regions, particularly where pore boundaries are ambiguous or highly complex.

4. Discussion

In summary, DRI-SAM addresses two major challenges inherent to automated segmentation of digital rock images: limited annotated data and robust generalization. Remarkably, the model was trained on only 200 labeled images from a single lithology over 100 training epochs, yet it consistently generalized to other rock types and imaging modalities. This demonstrates that even with a relatively small number of training iterations, DRI-SAM is capable of capturing the essential microstructural features necessary for accurate segmentation across diverse datasets. Previous studies have attempted to enhance digital rock image segmentation through edge guidance or structural constraints (Wang et al., 2025a), but none have achieved the level of stable generalization demonstrated by DRI-SAM when trained on limited data from a single lithology. The underlying reason lies in the difference between large-scale pretrained models and conventional segmentation networks. SAM's pretrained encoder already possesses the ability to extract fundamental image features across diverse domains. While it may initially lack sensitivity to the fine-scale microstructures specific to digital rock images, LoRA-based fine-tuning allows the encoder to rapidly adapt to these patterns, enabling effective segmentation. In contrast, traditional small-parameter networks tend to overfit or underfit when trained on limited datasets, often failing to learn sufficiently representative features and, consequently, struggling to generalize to unseen images. This limitation has long been a critical challenge in intelligent segmentation of various types of images. For DRI-SAM, failure cases may occur in boundary regions or areas with extremely small, low-contrast pores, where pore locations can be slightly misidentified or adjacent micropores merged. These issues could be mitigated by incorporating multi-resolution images for multi-scale feature extraction and fusion.

To further investigate why SAM requires fine-tuning for digital rock image segmentation rather than directly relying on its pretrained encoder features, we conducted a feature visualization analysis. Specifically, we examined the multi-channel feature maps ($256 \times 64 \times 64$) produced by the encoder. On the one hand, we randomly selected several individual channels from the feature maps for inspection; on the other hand, we applied principal component analysis (PCA) across all channels and projected the first three principal components into the RGB space, thereby providing a global visualization of the encoded representations. As shown in Fig. 7, the first row presents feature maps extracted directly from SAM's encoder, while the second row shows the features obtained after LoRA-based fine-tuning.

When features are extracted directly from SAM's encoder without any fine-tuning, the resulting activation maps tend to be diffuse and spatially scattered, showing limited sensitivity to pore–matrix boundaries and fine

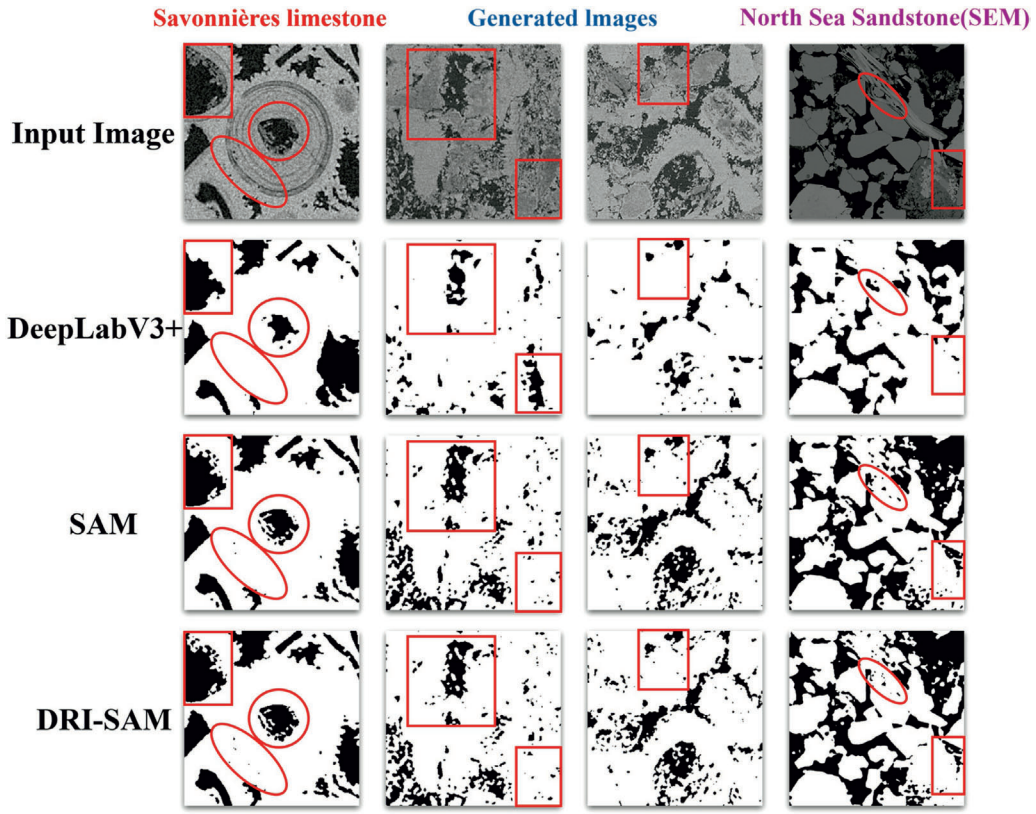

Fig. 6. Segmentation results of DRI-SAM and baseline models on more complex rock images.

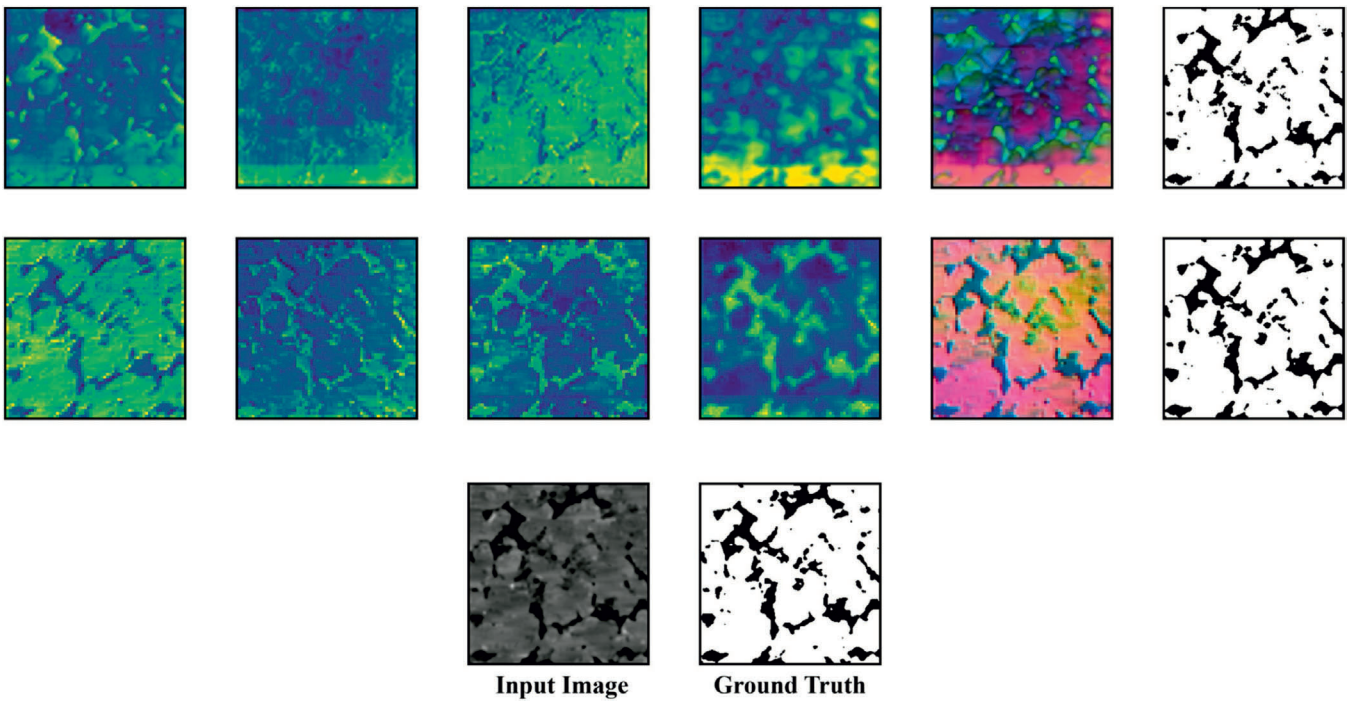

Fig. 7. Visualization of SAM encoder feature maps for digital rock images.

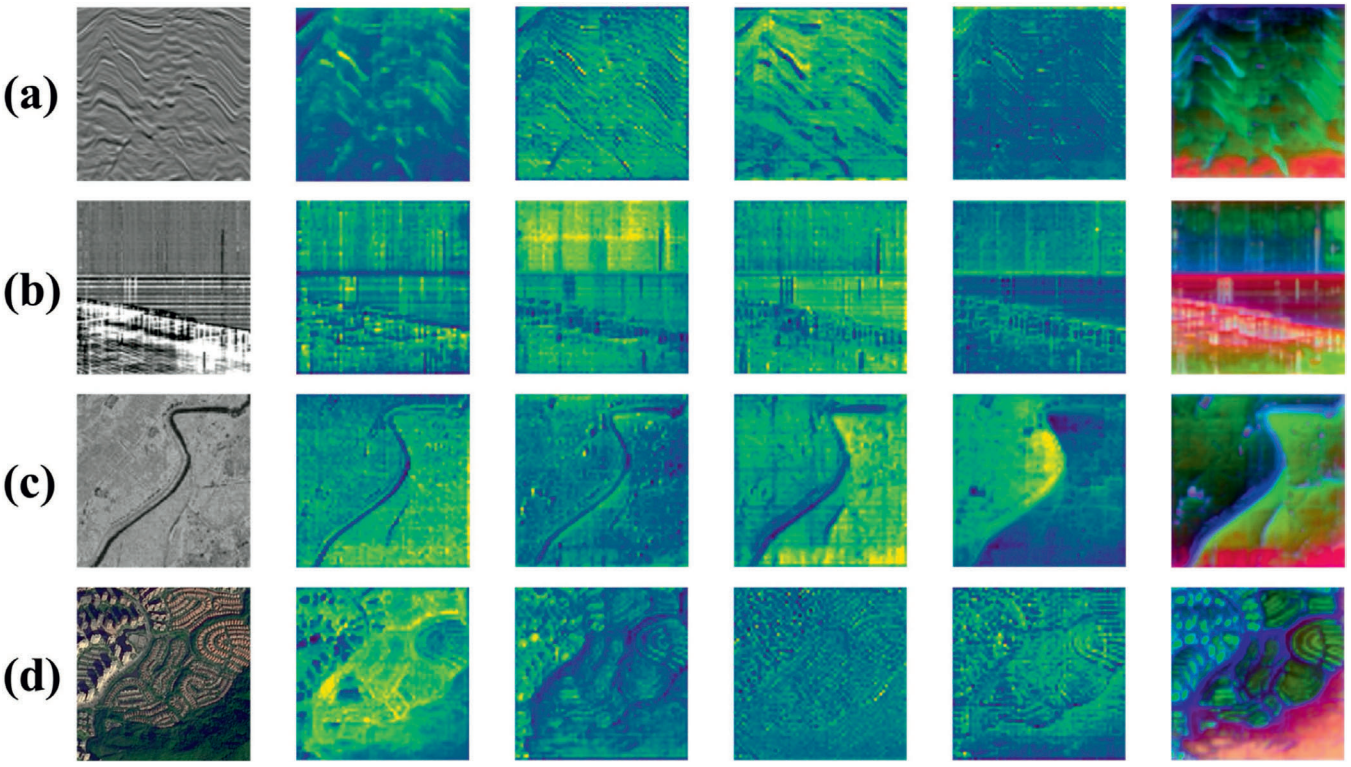

Fig. 8. Visualization of SAM encoder features across different geoscientific imaging modalities. (a) Seismic image, (b) distributed acoustic sensing (DAS) VSP data, (c) synthetic aperture radar (SAR) image, and (d) remote sensing image.

microstructural patterns. Feeding these raw encoder features into the decoder means that, regardless of how well the decoder is trained, it cannot recover the missing fine-grained information, leading to segmentation outputs that are less precise in regions with subtle or complex structures. Therefore, LoRA-based fine-tuning of the encoder is necessary to adapt SAM to digital rock images, enabling it to capture finer structural details and produce more accurate segmentation results.

As a foundation model for segmentation, SAM demonstrates strong generalization capability in the domain of digital rock images, and this adaptability is likely to extend to other geoscientific imaging tasks as well. To further examine this potential, we explored SAM's performance on seismic image, distributed acoustic sensing (DAS) VSP data, synthetic aperture radar (SAR) image, and remote sensing image, as illustrated in Fig. 8. The visualized feature maps reveal that, even without

domain-specific fine-tuning, SAM effectively captures fundamental representations such as textures, edges, and structural patterns. These features are essential not only for segmentation but also for downstream tasks including classification and denoising (Sheng et al., 2024). Compared with the feature maps obtained from digital rock images, it is evident that certain geoscientific image types align more readily with SAM's segmentation logic. This contrast underscores the particular difficulty of digital rock image segmentation, where subtle pore–matrix transitions and fine-scale heterogeneity present far greater challenges.

5. Conclusion

Intelligent image segmentation faces enduring challenges, particularly the scarcity of annotated data and the need for robust generalization across diverse domains. In this study, we proposed DRI-SAM, a domain-adaptive segmentation framework that combines LoRA-enhanced fine-tuning of SAM's encoder with automated prompt generation via a pretrained U-Net. This design enables accurate and stable segmentation of digital rock images, effectively tackling boundary ambiguity, textural heterogeneity, and limited supervision. Despite being trained on a small, homogeneous dataset, DRI-SAM demonstrates strong generalization across a wide range of rock types, voxel resolutions, and imaging sources. Its performance remains competitive even on complex carbonate and cross-modality images, significantly outperforming traditional segmentation networks and original SAM variants. These results highlight the effectiveness of prompt-guided adaptation for scientific image domains and suggest that foundation models like SAM, when appropriately tuned, can serve as powerful tools for digital rock physics. Beyond digital rock physics, the findings suggest that foundation models like SAM hold great promise for broader geoscientific imaging tasks. When properly adapted, they can achieve robust feature extraction and generalization, which are equally critical in these domains. Future research will further extend DRI-SAM to 3D scenarios, multiphase segmentation, and integration with downstream physical simulations to maximize its applicability in earth science.

CRediT authorship contribution statement

Ziqiang Wang: Writing – original draft, Validation, Data curation. **Zhiyu Hou:** Validation, Supervision, Software. **Shuai Hou:** Visualization, Supervision. **Danping Cao:** Supervision, Methodology, Conceptualization.

Data availability statement

The data used in this paper are obtained from the Digital Rocks Portal.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Danping Cao gratefully acknowledges financial support of this work through National Natural Science Foundation of China (42325403) and National Science and Technology Major Project of China (2024ZD1004201). Zhiyu Hou gratefully acknowledges financial support from China Scholarship Council (CSC202306450071).

References

Alkhimenkov, Y., 2025. Digital rock physics: calculation of effective elastic properties of heterogeneous materials using graphical processing units (GPUs). Comput. Geosci. 194, 105749. https://doi.org/10.1016/j.cageo.2024.105749

Al-Marzouqi, H., 2018. Digital rock physics: using CT scans to compute rock properties. IEEE Signal Process. Mag. 35 (2), 121–131. https://doi.org/10.1109/MSP.2017.2784459

Andhumoudine, A.B., Nie, X., Zhou, Q., Yu, J., Kane, O.Ibrahima, Jin, L., Djaroun, R.Rana, 2021. Investigation of coal elastic properties based on digital core technology and finite element method. Adv. GeoEnergy Res. 5 (1), 53–63. https://doi.org/10.46690/ager.2021.01.06

Andrä, H., et al., 2013a. Digital rock physics benchmarks—Part I: imaging and segmentation. Comput. Geosci. 50, 25–32. https://doi.org/10.1016/j.cageo.2012.09.005

Andrä, H., et al., 2013b. Digital rock physics benchmarks—part II: computing effective properties. Comput. Geosci. 50, 33–43. https://doi.org/10.1016/j.cageo.2012.09.008

Arns, C.H., Knackstedt, M.A., Pinczewski, W.V., Garboczi, E.J., 2002. Computation of linear elastic properties from microtomographic images: methodology and agreement between theory and experiment. Geophysics 67 (5), 1396–1405. https://doi.org/10.1190/1.1512785

Blunt, M.J., Bijeljic, B., Dong, H., Gharbi, O., Iglauer, S., Mostaghimi, P., Paluszny, A., Pentland, C., 2013. Pore-scale imaging and modelling. Adv. Water Resour. 51, 197–216. https://doi.org/10.1016/j.advwatres.2012.03.003

Cao, D., Ji, S., Cui, R., Liu, Q., 2022. Multi-task learning for digital rock segmentation and characteristic parameters computation. J. Pet. Sci. Eng. 208. https://doi.org/10.1016/j.petrol.2021.109202

Chi, P., Sun, J., Yan, W., Cui, L., 2024. From digital rock to digital wellbore: multiscale reconstruction and simulation. Adv. GeoEnergy Res. 13 (1), 1–6. https://doi.org/10.46690/ager.2024.07.01

Cui, R., Cao, D., Liu, Q., Zhu, Z., Jia, Y., 2021. VP and VS prediction from digital rock images using a combination of U-Net and convolutional neural networks. Geophysics 86 (1), MR27–MR37. https://doi.org/10.1190/geo2020-0162.1

Dvorkin, J., Derzhi, N., Fang, Q., Nur, A., Nur, B., Grader, A., Baldwin, C., Tono, H., Diaz, E., 2009. From micro to reservoir scale: Permeability from digital experiments. Lead. Edge 28 (12), 1446–1452. https://doi.org/10.1190/1.3272699

Esmaeili, M., 2024. Enhancing digital rock analysis through generative artificial intelligence: diffusion models. Neurocomputing 587, 127676. https://doi.org/10.1016/j.neucom.2024.127676

Gao, H., Wu, X., Liang, L., Sheng, H., Si, X., Gao, H., Li, Y., 2026. A foundation model enpowered by a multi-modal prompt engine for universal seismic geobody interpretation across surveys. Inf. Fusion 125, 103437. https://doi.org/10.1016/j.inffus.2025.103437

Guo, Z., Wu, X., Liang, L., Sheng, H., Chen, N., Bi, Z., 2025. Cross-domain foundation model adaptation: pioneering computer vision models for geophysical data analysis. J. Geophys. Res. Mach. Learn. Comput. 2 (1), e2025JH000601. https://doi.org/10.1029/2025JH000601

Hayatdavoudi, M., Niri, M.E., Kalhor, A., 2025. Comparative analysis of sandstone microtomographic image segmentation using advanced convolutional neural networks with pixelwise and physical accuracy evaluation. Sci. Rep. 15 (1), 22164. https://doi.org/10.1038/s41598-025-07211-2

Hou, Z., Cao, D., 2022. Estimating elastic parameters from digital rock images based on multi-task learning with multi-gate mixture-of-experts. J. Pet. Sci. Eng. 213, 110310. https://doi.org/10.1016/j.petrol.2022.110310

Hou, Z., Cao, D., Liu, Q., 2021. Segmentation of digital rock images guided by edge feature using deep learning. 2021 (1), 1–5. https://doi.org/10.3997/2214-4609.202113323

Hou, Z., Cao, D., Liu, Q., Su, Y., Ma, Y., Zhou, Z., 2023a. An intelligent method for reconstructing large-size digital rocks by joining multi-dimension information. Geoenergy Sci. Eng. 228, 212049. https://doi.org/10.1016/j.geoen.2023.212049

Hou, Z., Cao, D., Wang, X., 2023b. Intelligent digital rock physics assisting quantitative seismic interpretation, Third International Meeting for Applied Geoscience & Energy Expanded Abstracts, 743-747. https://doi.org/10.1190/image2023-3911212.1

Hu, X., Xu, X., Shi, Y., 2023. How to Efficiently Adapt Large Segmentation Model(SAM) to Medical Images, ArXiv, abs/2306.13731. https://doi.org/10.48550/arXiv.2306.13731

Ibrahim, A., Alqahtani, N., Wang, Y.D., Shabaninejad, M., Armstrong, R., Mostaghimi, P., 2020. Segmentation of X-Ray Images of Rocks Using Deep Learning. https://doi.org/10.2118/201282-MS

Iraji, S., Soltanmohammadi, R., Matheus, G.F., Basso, M., Vidal, A.C., 2023. Application of unsupervised learning and deep learning for rock type prediction and petrophysical characterization using multi-scale data. Geoenergy Sci. Eng. 230, 212241. https://doi.org/10.1016/j.geoen.2023.212241

Karimpouli, S., Tahmasebi, P., 2019a. Image-based velocity estimation of rock using Convolutional Neural Networks. Neural Netw. 111, 89–97. https://doi.org/10.1016/j.neunet.2018.12.006

Karimpouli, S., Tahmasebi, P., 2019b. Segmentation of digital rock images using deep convolutional autoencoder networks. Comput. Geosci. 126, 142–150. https://doi.org/10.1016/j.cageo.2019.02.003

Karimpouli, S., Tahmasebi, P., Saenger, E.H., 2018. Estimating 3D elastic moduli of rock from 2D thin-section images using differential effective medium theory. Geophysics 83 (4), MR211–MR219. https://doi.org/10.1190/geo2017-0504.1

Kirillov, A., et al., 2023. Segment anything. 2023 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 3992–4003.https://doi.org/10.1109/ICCV51070.2023.00371

Li, X., Li, B., Liu, F., Li, T., Nie, X., 2023. Advances in the application of deep learning methods to digital rock technology. Adv. GeoEnergy Res. 8 (1), 5–18. https://doi.org/10.46690/ager.2023.04.02

Lubis, L.A., Harith, Z.Z.T., 2014. Pore type classification on carbonate reservoir in offshore sarawak using rock physics model and rock digital images. IOP Conference

Series Earth Environmental Science 19, 012003. https://doi.org/10.1088/1755-1315/19/1/012003

Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B., 2024. Segment anything in medical images. Nat. Commun. 15 (1), 654. https://doi.org/10.1038/s41467-024-44824-z

Ma, Z., He, X., Sun, S., Yan, B., Kwak, H., Gao, J., 2023. Zero-shot digital rock image segmentation with a Fine-Tuned Segment Anything Model, ArXiv, abs/2311.10865. https://doi.org/10.48550/arXiv.2311.10865

Madonna, C., Almqvist, B.S.G., Saenger, E.H., 2012. Digital rock physics: numerical prediction of pressure-dependent ultrasonic velocities using micro-CT imaging. Geophys. J. Int. 189 (3), 1475–1482. https://doi.org/10.1111/j.1365-246X.2012.05437.x

Purswani, P., Karpyn, Z.T., Enab, K., Xue, Y., Huang, X., 2020. Evaluation of image segmentation techniques for image-based rock property estimation. J. Pet. Sci. Eng. 195, 107890. https://doi.org/10.1016/j.petrol.2020.107890

Reinhardt, M., Jacob, A., Sadeghnejad, S., Cappuccio, F., Arnold, P., Frank, S., Enzmann, F., Kersten, M., 2022. Benchmarking conventional and machine learning segmentation techniques for digital rock physics analysis of fractured rocks. Environ. Earth Sci. 81 (3), 71. https://doi.org/10.1007/s12665-021-10133-7

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. Paper Presented at Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing, Cham.

Saxena, N., Mavko, G., 2016. Estimating elastic moduli of rocks from thin sections: digital rock study of 3D properties from 2D images. Comput. Geosci. 88, 9–21. https://doi.org/10.1016/j.cageo.2015.12.008

Shan, L., Liu, Y., Du, K., Paul, S., Zhang, X., Hei, X., 2024. Drilling rock image segmentation and analysis using segment anything model. Adv. GeoEnergy Res. 12 (2), 89–101. https://doi.org/10.46690/ager.2024.05.02

Sheng, H., Wu, X., Si, X., Li, J., Zhang, S., Duan, X., 2024. Seismic foundation model: a next generation deep-learning model in geophysics. GEOPHYSICS 90 (2), IM59–IM79. https://doi.org/10.1190/geo2024-0262.1

Soltanmohammadi, R., Iraji, S., Rodrigues de Almeida, T., Basso, M., Ruidiaz Munoz, E., Campane Vidal, A., 2024. Investigation of pore geometry influence on fluid flow in heterogeneous porous media: A pore-scale study. Energy Geosci. 5 (1), 100222. https://doi.org/10.1016/j.engeos.2023.100222

Wang, F., Zai, Y., 2023. Image segmentation and flow prediction of digital rock with U-net network. Adv. Water Resour. 172, 104384. https://doi.org/10.1016/j.advwatres.2023.104384

Wang, Z., Hou, Z., Cao, D., 2025a. Edge-guided segmentation of digital rock images: integrating a pretrained edge aware path with the main segmentation path. Comput. Geosci. 197, 105884. https://doi.org/10.1016/j.cageo.2025.105884

Wang, Z., Hou, Z., Cao, D., 2025b. Enhancing SAM-based digital rock image segmentation via edge-semantics fusion. Appl. Comput. Geosci. 28, 100292. https://doi.org/10.1016/j.acags.2025.100292

Wang, Z., Hou, Z., Cao, D., 2026. Deep-learning-based digital rock physics analysis: from image segmentation and edge detection by few-shot learning to mechanical properties prediction. Geoenergy Sci. Eng. 256, 214133. https://doi.org/10.1016/j.geoen.2025.214133

Yang, Y., Horne, R.N., Cai, J., Yao, J., 2023. Recent advances on fluid flow in porous media using digital core analysis technology. Adv. GeoEnergy Res. 9 (2), 71–75. https://doi.org/10.46690/ager.2023.08.01

Ye, S., Song, X., Ma, Z., Gao, Y., Zhu, L., Zhou, M., Xiao, L., Wen, G., Bijeljic, B., Blunt, M.J., 2025. A noise-resistant and annotation-free supervoxel-based algorithm for rapid segmentation of multiphase X-ray images. Adv. GeoEnergy Res. 16 (1), 50–59. https://doi.org/10.46690/ager.2025.04.06