Advances in Geo-Energy Research

Vol. 12, No. 2, p. 89-101, 2024

Original article

Drilling rock image segmentation and analysis using segment anything model

Liqun Shan[1], Yanchang Liu[2]*, Ke Du[2], Shovon Paul[1], Xingli Zhang[1], Xiali Hei[1]

[1]*School of Computing and Informatics, University of Louisiana at Lafayette, Lafayette 70503, USA*

[2]*School of Physical and Electrical Engineering, Northeast Petroleum University, Daqing 163318, P. R. China*

Keywords:
Segmentation
segment anything model
underwater rock image
granular analysis

Cited as:
Shan, L., Liu, Y., Du, K., Paul, S., Zhang, X., Hei, X. Drilling rock image segmentation and analysis using segment anything model. Advances in Geo-Energy Research, 2024, 12(2): 89-101.
https://doi.org/10.46690/ager.2024.05.02

Abstract:
Image processing and analysis techniques are commonly utilized in various fields such as geology, underwater engineering, environmental conservation, marine resource exploration, and soil and geological assessments, particularly for examining drilling rock samples. However, processing images of rocks drilled underwater is challenging due to the intricate nature of aquatic settings, where factors such as light reflection and refraction, irregular sizes of rocks, and overlapping particles introduce noise, obscure textures, and distort colors in the images. Although improved versions of the mask region-based convolutional neural network have shown promise for quick and accurate analysis of large sets of underwater rock images, these methods can still be affected by inconsistencies in rock appearance, texture, and lighting. To address these issues, a comprehensive approach is introduced using the segment anything model. Our methodology begins with the application of Gaussian filters to reduce noise and smooth images, followed by the deployment of underwater image enhancement. Further, histogram equalization is applied to better the contrast and employ the segment anything model approach for the detailed understanding of rock features by extracting information on rock size and shape. EeEquivalent area circle diameter and axial ratio are used to generate particle size alignment maps and to ascertain shape details. Our approach has achieved an average precision rate of 80.6%, outperforming other strategies and yielding more precise rock information analysis.

1. Introduction

Rock image segmentation, a crucial technique in image processing, plays a vital role in areas such as geoscience and the development of resources. Assigning pixels in rock images to regions with specific semantic or instance relationships enables obtaining important information related to rock structure, features, and genesis, supporting geological exploration and resource development (Cai et al., 2022; Shan et al., 2022; Akkaynak et al., 2023). Drilling rocks are widely distributed and complex, and studying underwater rock images is particularly difficult. Due to the limitations of the unique aquatic imaging environment, underwater photos often have many problems, such as noise interference, blurred texture features, low contrast, and color distortion. Therefore, an urgent problem to be solved is how to accurately, quickly, and stably detect, identify, and track underwater target objects under poor image visibility.

Traditional image segmentation methods mainly include semantic segmentation and instance segmentation. Semantic segmentation labels each pixel of an image into a specific semantic category, such as rock particles and background. Commonly used algorithms, such as fully convolutional networks (Long et al., 2015), point-wise spatial attention network (Zhao et al., 2018), and DeepLab (Minaee et al., 2021), can perform rock image segmentation, but they need more segmentation accuracy to accurately calculate the size of rock particles. Instance segmentation, on the other hand, requires the separation of different object instances from the image to achieve pixel-level classification. The most representative instance segmentation algorithm is mask region-based convolutional neural network (Mask R-CNN), which is extended

Yandy Scientific Press

*Corresponding author.
E-mail address: liqun.shan1@louisiana.edu (L. Shan); ychang_liu@nepu.edu.cn (Y. Liu); d15153065697@163.com (K. Du); shovon.paul1@louisiana.edu (S. Paul); xingli.zhang@louisiana.edu (X. Zhang); xiali.hei@louisiana.edu (X. Hei).

Received January 28, 2024; revised February 19, 2024; accepted March 17, 2024; available online March 24, 2024.

based on a target detection network and can simultaneously predict each object's bounding box, category, and pixel-level mask, thus achieving high-quality instance segmentation (He et al., 2017).

Introducing the Mask R-CNN method brings new possibilities and technological breakthroughs for the rock image segmentation task. Firstly, He et al. (2017) proposed Mask R-CNN and successfully applied it to target detection and segmentation. Subsequently, Qadir et al. (2019) adapted the Mask R-CNN and used different modern CNNs as its feature extractors to evaluate its performance for polyp detection and segmentation, and their experiments showed that good segmentation results were obtained for all of them. Roland et al. (2019) added a new prediction header to Mask R-CNN. The edge protocol header encourages predicted masks to have image gradients similar to those of ground truth masks using edge detection filters, which never improves the training speed of the algorithm. Bello et al. (2021) presents an enhanced Mask R-CNN model for multi-objective instance segmentation to support fuzzy boundaries and irregular shapes of animals for precision animal husbandry. The method combines an optimal filter size smaller than the residual network for extracting smaller composite features and region suggestions utilizing multi-scale semantic features to improve the performance of the algorithm. Researchers have continued optimizing and extending Mask R-CNN's application in rock image processing in recent years. Ullo et al. (2021) proposed a combination of Mask R-CNN and migration learning for pre-training on large-scale datasets, further improving the model's performance on the rock image recognition task. Fan et al. (2022) proposed a Mask R-CNN-based segmentation and analysis method for stacked stone materials by adding the squeeze-and-excitation block to enhance the feature extraction capability of the model. These studies have provided powerful tools for rock image segmentation and have demonstrated excellent performance on various tasks and datasets.

In addition, several researchers have focused on combining particle recognition and morphological analysis. Frei and Kruis (2020) proposed a convolutional neural network for size analysis of agglomerated and partially sintered particles and improved and optimized the Mask R-CNN by employing advanced training strategies, and hyperparameter tuning to achieve pixel-level detection and size measurements of agglomerated, occluded, raw particles. Chen et al. (2018) proposed a new holistic nested convolutional neural network for bone cement segmentation and shape analysis by supervised learning of the network using labeled images of rock particles, enabling it to learn morphological features of the particles. Liang et al. (2019) proposed a method based on a lightweight U-net model combined with image processing techniques such as threshold segmentation for the extraction and sizing of rock particles by modeling the 3-Dimensional structure of the particles and analyzing the parameters such as size, contour, and curvature using morphological feature descriptors.

However, complex texture modeling may limit conventional convolutional neural networks when dealing with underwater rock images. Factors such as illumination variations and noise make rock image segmentation more challenging, while rock morphology diversity and texture complexity also increase the difficulty of the segmentation task.

Recently, Kirillov et al. (2023) proposed segment anything model (SAM) which has made significant progress in breaking the segmentation boundaries and has contributed significantly to developing fundamental computer vision models. SAM is a cue-based model trained on 11 million images with over 1 billion masks, enabling powerful zero-sample generalization for image processing with simple cues and text instructions. In terms of image editing, SAM can generate accurate masks with simple hints that can help in image editing scenarios. Yamagiwa et al. (2024) proposed a zero-shot edge detection technique that improves the SAM by reducing edge over-detection. They employ a three-step process: eliminating small masks, merging masks via spectral clustering based on position and overlap, and removing artifacts to significantly improve SAM's utility and offer a new avenue in zero-shot edge detection techniques. SAM segments different objects and regions in the image, assigning specific labels to each area. Then, erosion and expansion operations (Suvorov et al., 2022) perform functions such as filling, resulting in more accurate and high-quality image restoration.

Regarding style conversion, Ganugula et al. (2023) utilized SAM to conduct arbitrary style transfer to multiple objects within an image. By integrating style embedding with object segmentation strategies, this approach can produce high-quality stylized images. In real-life scenarios, SAM shows a wide range of potential applications. Julka and Granitzer (2023) applied SAM model for quicker annotation of planetary images, specifically for mapping skylights. This approach significantly reduces manual labeling effort and improves segmentation efficiency, offering a promising tool for accelerating the exploration and analysis of extraterrestrial landforms. Giannakis et al. (2024) used SAM for crater detection. SAM shows good performance and potential in all aspects. In addition, SAM has demonstrated the ability to handle image segmentation problems in complex scenes and adapt them to various domains and data types by combining post-processing steps or using specific training methods. Osco et al. (2023) applied SAM to remote sensing image analysis, exploiting its superior generalization and zero-shot learning for aerial and orbital image processing. Testing across diverse datasets with various prompts, they enhanced SAM's accuracy for remote sensing through an innovative combination of text prompts and one-shot training, despite challenges with lower resolution images. Mazurowski et al. (2023) evaluated SAM's ability to segment medical images across 19 datasets, finding varied performance dependent on the dataset and task specifics. Their key insights include SAM's better performance with box prompts and for distinct objects, outperforming similar tools in single-point prompt scenarios. While SAM shows limited improvement with multiple prompts, it displays considerable potential for automated image segmentation, albeit with performance inconsistencies across different scenarios. Overall, SAM is a powerful image processing model with many potential applications in image segmentation, target detection, image generation, and video segmentation and tracking. It can provide accurate segmentation results and offer fundamental

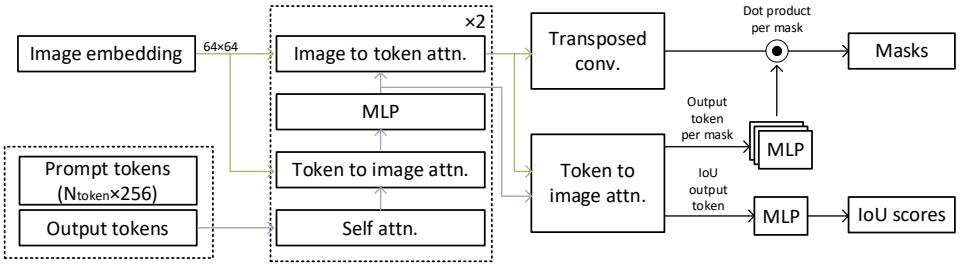


Fig. 1. Mask decoder of SAM.

and innovative solutions for more complex visual tasks.

In addressing the intricate challenges associated with processing rock images within complex environments, particularly underwater scenarios, this study introduces the SAM algorithm for drilling rock image segmentation. SAM excels in extracting densely arranged particle contours from images featuring intricate rock formations. Through comparative experiments, our work demonstrates the superior performance of the SAM algorithm over the Mask R-CNN algorithm, leveraging various backbones, including ResNet-101 (Anantharaman et al., 2018), ResNet-50 (Tahir et al., 2021), Wide ResNet-101 (Ullo et al., 2021), and ResNeXt-101 (Yamada and Di Santo, 2024) in the context of rock image segmentation. Furthermore, our approach extends beyond segmentation, encompassing post-processing and in-depth shape analysis of the segmented images. This comprehensive analysis aims to unravel the underlying structure and properties of rocks. By calculating essential shape features such as equivalent area, circular diameter, and axial ratio, we gain valuable insights into the size and shape characteristics of the particles. In essence, rock image segmentation poses a pivotal yet intricate challenge.

This paper seamlessly integrates the SAM model with advanced techniques for processing rock images, yielding enhanced segmentation and shape analysis results. Our notable contributions encompass:

1) A novel method was developed for processing underwater drilling rock images, effectively addressing challenges such as reflection, refraction, uneven rock size, and particle stacking.
2) The proposed method achieves an impressive average precision of 80.6%, surpassing the capabilities of other masked R-CNN methods.
3) Experiments show that our approach outperforms existing techniques and provides more accurate insights into the complex details of rock structures, significantly advancing the understanding of their properties.

2. Methodology

2.1 Segment anything model

The SAM is a robust data engine for image segmentation tasks. Trained on the extensive SA-1B dataset, comprising approximately 11 million images and 1 billion segmentation masks, SAM exhibits notable generalization capabilities. Leveraging the Transformer visual model (Carion et al., 2020), SAM includes distinct components: an image encoder, a cue encoder, and a mask decoder (Rahmon et al., 2021). SAM's image encoder utilizes a pre-trained Vision Transformer (ViT) (Dosovitskiy et al., 2021), employing a self-attentive mechanism to convert images into a high-dimensional feature representation. This process captures global and local contextual information, generating an effective image-embedding representation. The subsequent step involves instance detection and feature extraction, where SAM accurately locates and identifies each object instance in the image using an object detection module. A CNN extracts feature embeddings from the bounding box of each sample. These embeddings encapsulate visual and semantic information about the object instances, offering crucial input for subsequent instance segmentation tasks. SAM's mask decoder employs an advanced Transformer decoder block, incorporating self-attention and cross-attention mechanisms. This decoder fuses image embeddings and object embeddings (Schult et al., 2023), mapping them to a dynamic linear classifier via a multilayer perceptron (MLP) (Goodfellow et al., 2016). The result is the generation of per-pixel instance masks. SAM computes the loss by comparing these masks with annotated instance segmentation masks, utilizing backpropagation to optimize model parameters. Through a comprehensive process of pre-training and fine-tuning, SAM adeptly learns the semantic information associated with each object instance, translating this knowledge into pixel-level instance segmentation results.

While Mask R-CNN relies on a two-stage process involving region proposal and then classification and mask prediction, SAM uses a Transformer-based architecture, potentially enabling more direct and parallel processing of image segments. SAM's use of the ViT allows it to consider the global context of the image more effectively, which can be a limitation in the region-based approach of Mask R-CNN. The detailed semantic feature extraction in SAM, followed by the advanced Transformer decoder, provides a more nuanced understanding of each object instance, leading to more accurate pixel-level segmentation. The extensive training on the SA-1B dataset gives SAM a significant advantage in terms of learning from a large variety of segmentation scenarios, contributing to its generalization capability.

The mask decoder, depicted in Fig. 1, incorporates a learnable output token within the prompt embeddings for decoder output. This token comprises two essential components: The intersection over union (IoU) token and the mask token. The IoU token, illustrated on the right side of the structure

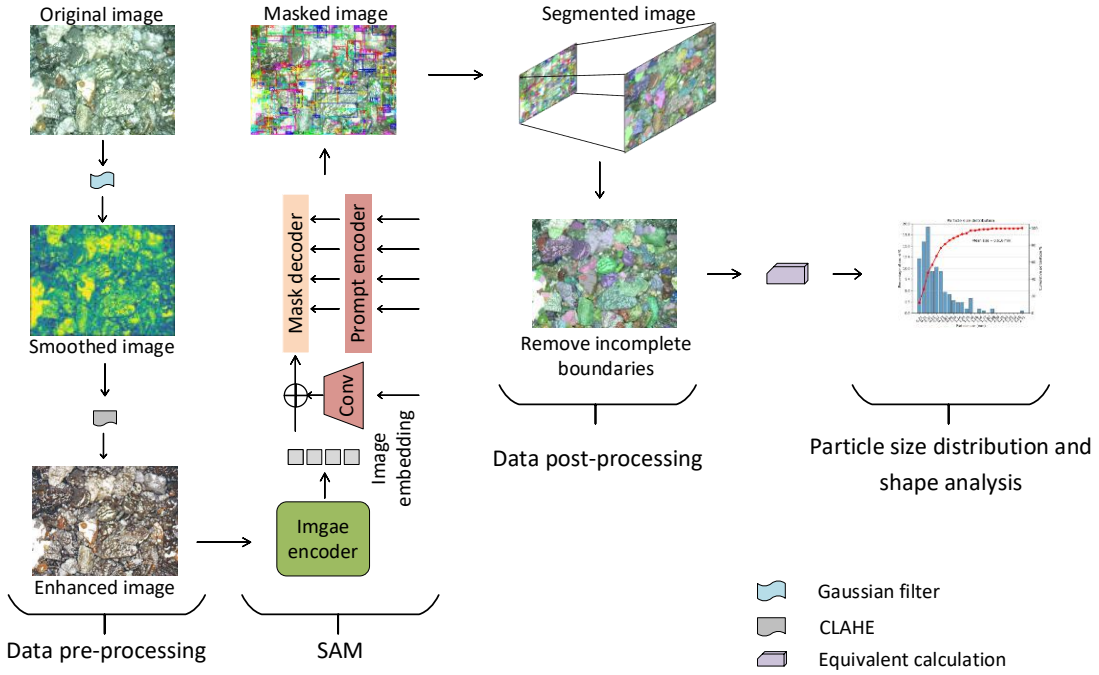


Fig. 2. Proposed workflow for segmenting rock images.

diagram, is later isolated to predict the reliability of IoU. This prediction is supervised by the mean squared error (MSE) loss between the model-computed IoUs, model-computed masks, and the actual IoUs from the ground truth (Murphy, 2012). Simultaneously, the mask token, corresponding to the output token per mask on the right side of the structure picture, is also separated for participation in the final mask prediction. The operation involves several key steps:

1) Combining prompt and output Tokenstokens: The model combines prompt and output tokens for self-attention operations.
2) Cross attention with image embedding: The obtained token is used as query in cross-attention with the image embedding.
3) Point-wise MLP token update: Tokens are updated using point-wise multi layer perceptron.
4) Cross attention with updated token and image embedding: Cross attention is again applied, using the updated token as query and image embedding.

The decoder repeats the above steps twice, concatenating the attention results through residuals to produce a final output of masks and IoU scores.

This study leverages the described methodology in the context of rock images, applying the SAM to enhance segmentation accuracy. SAM adopts the ViT architecture, known for its potent representation learning ability, enabling the accurate capture of complex texture and shape information in rock images. SAM generates three masks and assigns scores to each, facilitating the selection of the most appropriate segmentation result based on the expected segmentation area. This fine-grained segmentation capability allows SAM to accommodate rocks of varying sizes and provide highly accurate segmentation results.

2.2 Workflow

This study aims to present an automated approach for the identification, segmentation, and shape analysis of rock particles. The essential procedures encompass (1) image pre-processing, (2) image recognition and segmentation, (3) data post-processing, and (4) shape analysis. In the rock image segmentation process (refer to Fig. 2), the initially obtained dataset undergoes pre-processing to enhance underwater image clarity and contrast. This process uses fuzzy kernel filtering, color correction, and enhancement techniques to facilitate more distinct segmentation of rock features. Subsequently, the contrast limited adaptive histogram equalization (CLAHE) (Sahu et al., 2019) technique is applied for local contrast enhancement, followed by image denoising and smoothing to reduce noise and enhance image details. The pre-processed dataset is then fed into the SAM model, generating an initial segmentation map.

Post-processing steps are implemented following image processing to eliminate incomplete or inaccurate particle segmentation contours. In addition, measures were taken to repair or fill segmented particles, improving the accuracy of the results. The segmented particles are then subjected to size estimation, where characteristics such as diameter, area, or volume are measured, providing crucial information about the rock particles. Finally, shape estimation and ranking are carried out based on shape features like roundness and aspect ratio. This comprehensive analysis allows for a thorough understanding of the spatial structure and composition of the rock by comparing the shape characteristics of different particles.

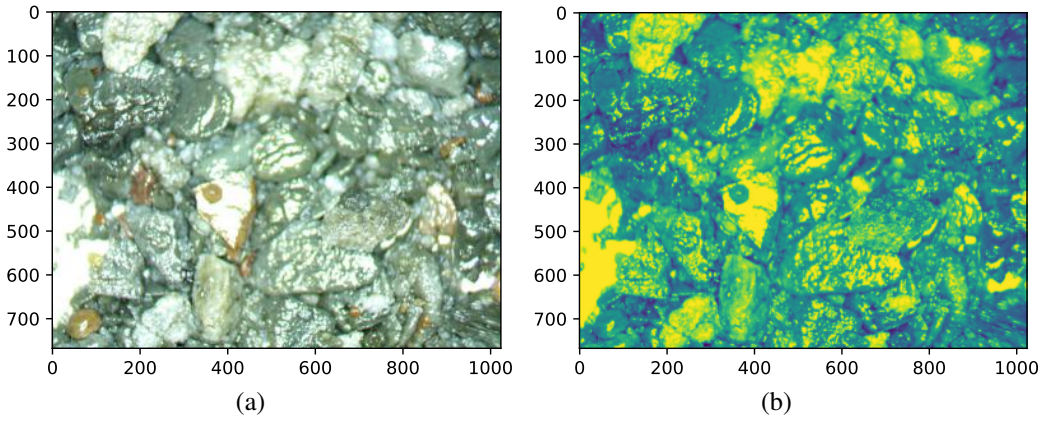


Fig. 3. Comparison of (a) original and (b) smoothed images.

3. Experiments

3.1 Dataset

The underwater rock image acquisition, processing, and analysis field encounter many intricate challenges. Factors such as light attenuation, scattering, and absorption significantly impact image quality, leading to contrast reduction and increased noise. The complex nature of underwater rock images, marked by uneven sizes and particle stacking, further complicates segmentation and analysis. The coexistence of rocks with varying sizes in the images poses a segmentation challenge, while particle stacking hinders the accurate location and segmentation of rock boundaries. Light attenuation induces blurriness in rock areas distant from the light source, making recognition challenging. Additionally, water currents introduce blurring, distortion, and motion blur effects. Various noise sources within the underwater environment exacerbate image quality issues.

To confront these challenges, drilling rock image data of several formation zones from a drilling company and labeled these data were collected under the guidance of drilling experts. The dataset comprises 310 sets of image data with different resolutions. The original images are cropped into small patches with a resolution of 512×512 pixels, providing a comprehensive representation of underwater rock image diversity. This dataset covers a variety of underwater rock images, including many types of data, such as different size distributions, particle stacking scenarios, light attenuation variations, and noise instances. To ensure data quality and usability, the images are partitioned into a training and test set, with a 70%-30% ratio for training and testing, respectively.

3.2 Data pre-processing

3.2.1 Image denoising and smoothing

Underwater images, especially those of rock samples, are prone to various types of noise due to factors such as floating particles and lighting conditions. Gaussian filters are effective in reducing this noise because they apply a smoothing effect that blurs the image slightly. This blur helps in diminishing the impact of grainy noise, making the underlying rock textures and features more distinguishable. The intrinsic smooth nature of the Gaussian filter makes it ideal for softening the sharp edges and irregularities in the rock images. This smoothing effect is particularly beneficial for images with irregular sizes of rocks and overlapping particles as shown in Fig. 3. By smoothing the image, the Gaussian filter helps in reducing the distortions caused by these irregularities, making the rocks' shapes and sizes clearer for analysis.

The formula for a 2-Dimensional Gaussian function is expressed as:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

where $G(x,y)$ is the value of the Gaussian function at point (x,y), σ represents the standard deviation of the Gaussian distribution, which controls the Gaussian kernel's spread, with higher values causing wider, blurrier effects due to slower weight decreases from the center, and lower values leading to sharper images with fast-decreasing weights. Kernel size, typically an odd number, dictates the filter's dimension and the extent of neighboring pixel consideration, affecting blurriness: Larger kernels increase smoothing by involving more pixels, while smaller ones localize the effect.

In our experiment, Gaussian filter is applied as the filter function to an input image using conv2d function for convolution, with $\sigma = 10$ and kernel size = 5. The σ controls the extent of the smoothing effect; a larger σ results in a greater blurring effect. The selection of σ is critical and can be adjusted based on the specific requirements of the underwater rock images to balance between noise reduction and preservation of essential details. The initial application of Gaussian filters sets a strong foundation for the following steps in the methodology. With reduced noise and smoothed images, techniques like underwater image enhancement and histogram equalization become more effective. These subsequent steps can better improve contrast and detail visibility, leading to more accurate analysis by the SAM.

3.2.2 Underwater image enhancement

Underwater rock images pose distinct challenges in processing and analysis due to their unique optical properties

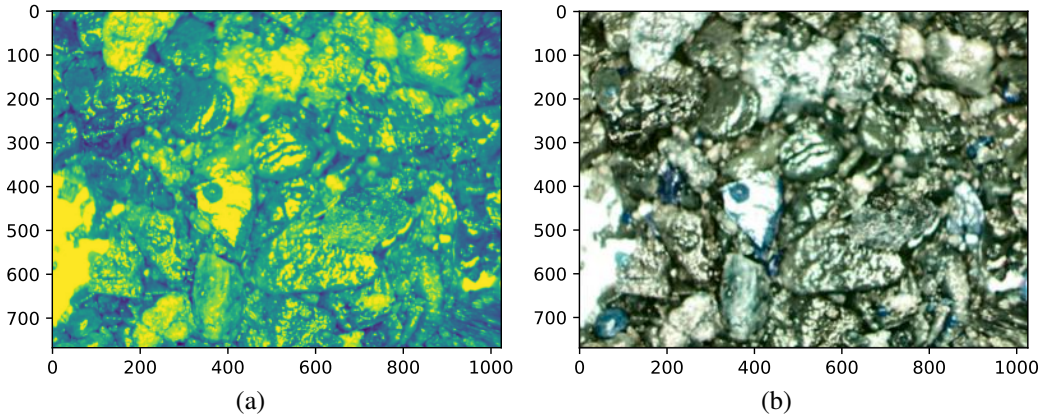


Fig. 4. Comparison of (a) smoothed and (b) CLAHE images.

and environmental constraints. The absorption and scattering of light in water introduce color distortion, with light waves of different colors experiencing varying speeds and levels of attenuation, causing deviations from the actual colors of rocks. Reflection and refraction between water and rock surfaces create variations in contours, surface features, and shadows, resulting in light spots, reflections, and shadows that interfere with shape and edge information. Consequently, underwater rock images commonly exhibit low contrast, color distortion, blurred details, and uneven lighting.

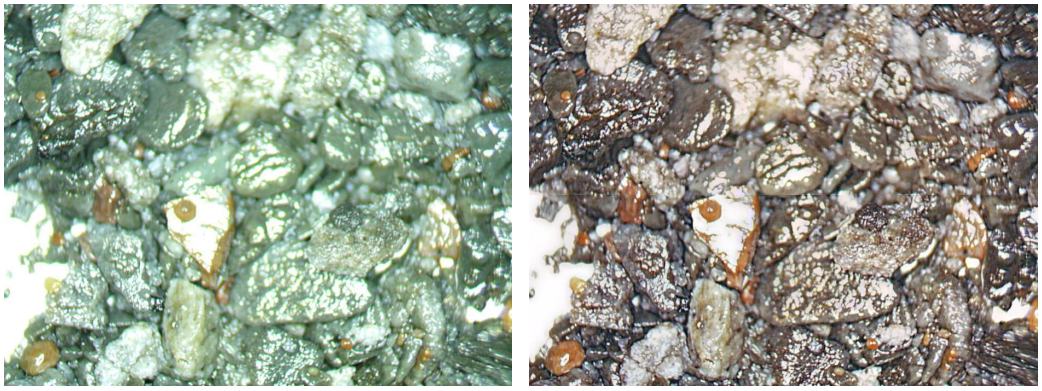
CLAHE is employed to enhance the contrast of underwater rock images. CLAHE, an improved histogram equalization method, adaptively improves contrast without being sensitive to noise or over-enhancing local features. By segmenting the image into small chunks and performing histogram equalization within each sub-region, CLAHE enhances image details and visual appearance. A side-by-side comparison of two images shows the effects of different image processing techniques on underwater rock imagery in Fig. 4. On the left is a smoothed image, which appears to have undergone a filtering process to reduce noise, resulting in a more uniform appearance of the rock surfaces with softened edges and textures. On the right is a CLAHE Image, where the CLAHE technique has been applied. The CLAHE process has visibly enhanced the contrast of the rock textures and colors, making individual rocks and mineral features stand out more clearly. Unlike the smoothed image, the CLAHE image exhibits sharper details, improved visibility of rock features, and a more varied color palette that likely represents the rocks' true colors and details more accurately.

Next, SyreaNet (Wen et al., 2023) is incorporated for enhancing underwater image data to address these challenges. SyreaNet, a generative adversarial network (Goodfellow et al., 2014) algorithm, employs a modified underwater image formation model (Akkaynak and Trebibitz, 2018) and a domain adaptation strategy (Toldo et al., 2020) for training and prediction. By using synthetic and real data, SyreaNet increases the number of data samples while maintaining the original data distribution, enabling better adaptation to the complex underwater environment concerning light, noise, and perturbation. The components of SyreaNet include the physically guided synthesis module, which generates synthetic underwater images based on principles of underwater imaging and optical transmission models, and the physically guided detangling network, designed for predicting detangling components and clear images, as illustrated in Fig. 5. Enhanced image shows the result of applying SyreaNet for image enhancement. This processed image has a markedly increased contrast, with the colors of the rocks appearing much more vivid and the textures more defined. The distinction between individual rocks and their features is noticeably clearer in the enhanced image. The whites appear brighter, the darks deeper, and the overall visibility of the underwater scene is significantly improved.

3.3 Drilling rock image segmentation

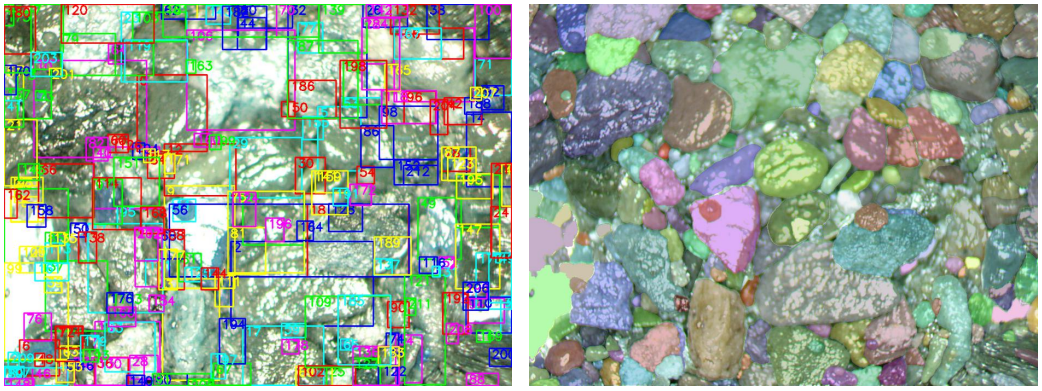
For rock image segmentation, the SAM network initially conducts target detection on the original image to identify the target region shown in the Fig. 6(a). Subsequently, a segmentation operation is applied based on this detection, extracting effective aggregated pixels within each bounding box. These aggregated pixels represent the particle shapes and boundaries of the target object. By extracting valid aggregated pixels, the network generates a corresponding accurate mask shown in Fig. 6(b), accurately outlining the boundaries of the target object and distinguishing it from the background.

SAM has demonstrated exceptional efficacy in rock image segmentation as shown in Fig. 6. The segmentation outcomes reveal precise delineation and extraction of small-sized rock particles, showcasing well-defined boundaries. Notably, SAM adeptly distinguishes the intricate boundary between fine rock particles and the surrounding background, yielding clear and sharp features. This proficiency signifies SAM's capability to capture fine textures and morphological characteristics, enhancing the accuracy and reliability of identifying fine rock particles. Furthermore, SAM effectively preserves shape and size information across diverse rock particles in the segmentation results. Each particle exhibits distinct morphological features, such as round, oval, or irregular shapes, and maintains a consistent size relationship with the original image. Signif-



(a) (b)

Fig. 5. Impact of underwater image enhancement, (a) original and (b) enhanced images.



(a) (b)

Fig. 6. Image segmentation using SAM, (a) masked and (b) segmented images.

icantly, SAM mitigates challenges related to overlapping and blurring between rock particles. The method ensures accurate separation, eliminating overlaps and interference between individual particles. Hence, it highlights SAM's commendable performance in addressing rock particle contact and overlap issues.

The segmentation performance on rock images from different scenarios is further described in Fig. 7. It is a collection of paired images. On the left of each pair is the original rock image, and on the right is the segmented image that highlights individual rocks in various colors to differentiate them. The segmentation appears to be quite detailed, with the segmented images showing a clear distinction between different rocks. Some of the original images have a variety of textures and color tones, indicating different types of rocks and minerals, while others are more uniform.

The robustness of the proposed method is evident from the segmented images. The method is able to distinguish between rocks with similar colors and textures, which is often challenging in geological image analysis. The segmentation is particularly impressive in complex geological conditions, where rock formations can be highly irregular and have overlapping features. The method's effectiveness in different scenarios suggests that it has a strong ability to generalize across different geological formations, which is a desirable attribute in segmentation tasks for geological assessments.

3.4 Data post-processing

Irregularly shaped particles in the underwater rock image challenge the conventional segmentation method, leading to incomplete particle masks (Zhou et al., 2021). To enhance the SAM prediction accuracy, a watershed-based post-processing approach is introduced for refining the final segmentation masks of rock particles. This technique involves applying the watershed transformation to the region within the predicted segmentation mask, utilizing the predicted mask minus its boundary as a marker. The algorithm feeds the resulting mask back to itself for further refinement. The integration of watershed post-processing notably enhances the accuracy of subsequent shape analyses. The segmentation mask map in Fig. 8(a) reveals over-segmented and under-segmented particles. The post-processing map is obtained by removing incomplete particles and addressing inaccuracies inherent in the segmentation algorithm, as shown in Fig. 8(b). This post-processing step ensures that the shape of segmented particles is more precise and reliable, overcoming limitations in the initial segmentation and contributing to improved overall performance.

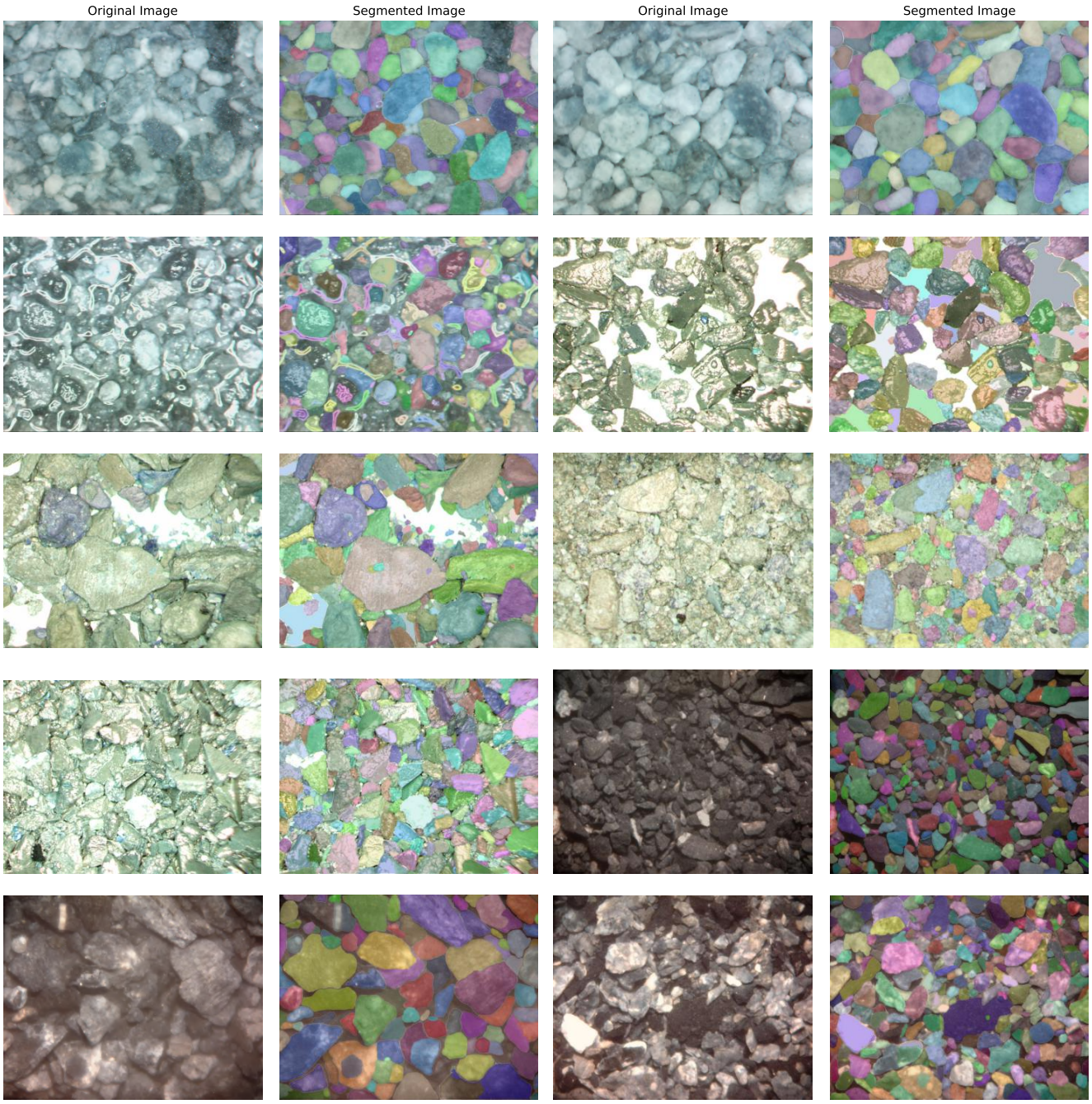


Fig. 7. Drilling rock image segmentation in different scenarios.

4. Evaluation

4.1 Segmentation results and analysis

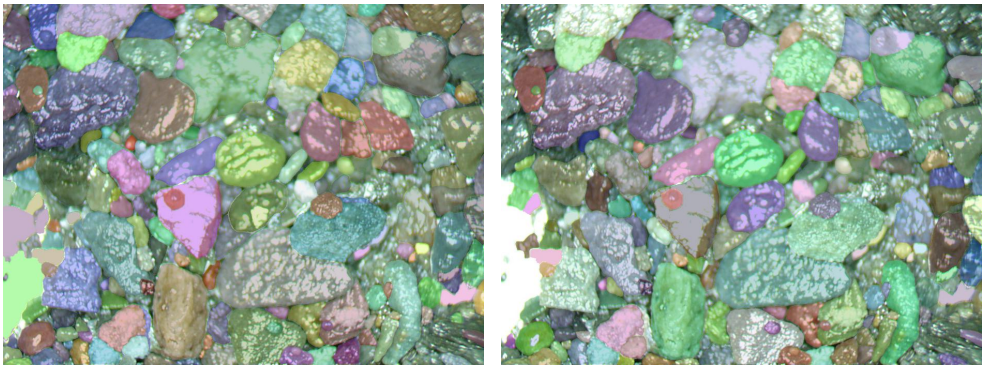
Commonly used evaluation metrics in instance segmentation tasks include average precision (AP), average recall (AR), and F_1. AP indicates how many instances predicted as positive samples are truly positive. It measures the model's accuracy under different IoU thresholds and target sizes. AR denotes the detection ability of the model used to measure real targets in the instance segmentation task. In the instance segmentation task, it denotes the average recall of detected targets at different IoU thresholds. A higher AR value indicates that the model can detect the target better. The mathematical expressions of AP and AR are expressed as follows:

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{2}$$

$$\mathrm{AP} = \frac{1}{N}\sum_{k=1}^{N} P(k) \tag{3}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{4}$$

$$\mathrm{AR} = \frac{1}{N}\sum_{k=1}^{N} R(k) \tag{5}$$



(a) (b)

Fig. 8. (a) Segmented image without and (b) incomplete particles at the edge.

where TP denotes a true positive sample (correctly detected target), FP denotes a false positive sample (incorrectly detected target), and FN denotes a false negative sample (to detect target correctly). k refers to a threshold. N is the number of thresholds considered. The F_1 value is a metric that combines the precision and recall of the model. It is the reconciled average of precision and recall and is used to assess the model's overall performance. A higher F_1 value indicates that the model performs better regarding accuracy and recall. The mathematical expression for F_1 is as follows:

$$F_1 = 2\frac{\mathrm{AP}\cdot\mathrm{AR}}{\mathrm{AP}+\mathrm{AR}} \tag{6}$$

These metrics can help evaluate the model's performance on different target sizes and provide the robustness and adaptability of the model. Higher values of the metrics indicate better segmentation of the model for different particle sizes. To summarize, AP, AR, and F_1 values are essential metrics for measuring the performance of an instance segmentation model and can provide information on model accuracy, recall, and overall comprehensive performance. At the same time, they can also help to evaluate the model's ability to adapt to different target sizes.

The proposed method is compared with alternative approaches across three key performance metrics: Precision, recall, and F_1-score shown in Table 1. The data unequivocally highlights the substantial superiority of SAM over the Mask R-CNN model in the context of rock image segmentation. The SAM model exhibits markedly higher precision, recall, and overall performance when contrasted with the Mask R-CNN model, with improvements exceeding 10% across all evaluation indices. Notably, SAM attains a superior AP score of 0.806 and an AR score of 0.813 among all backbone models, achieving an impressive F_1-score of 0.808. These compelling outcomes underscore SAM's capacity to detect and segment rock particles with heightened accuracy. Importantly, this enhanced performance positions SAM as a more reliable tool, offering robust data support for analyzing and exploring rock formations.

Table 1. Comparison between SAM and Mask R-CNN.

Model	Backbone	*AP*	*AR*	F_1
Mask R-CNN	ResNet-50	0.669	0.712	0.693
	ResNet-101	0.663	0.716	0.681
	Wide ResNet-101	0.694	0.735	0.709
	ResNeXt-101	0.668	0.715	0.692
SAM	Vit-h	0.806	0.813	0.808

4.2 Particle size distribution and shape analysis

4.2.1 Size distribution

With the above, the proposed approachwe can automatically distinguish the particles in a rock image from the background and obtain boundary information for each particle. Based on this boundary information, automatic size and shape estimation of individual particle contours are performed. By segmenting the rock image and estimating the particle size and shape, one can obtain the size information of each particle to construct a rock particle size arrangement map. As shown in Fig. 9, the pixel size of each particle is determined by the diameter of an equivalent area circle, D, in mm. This diameter D can be derived from the basic formula for the area of a circle.

Calculating the equivalent circle diameter of each particle provides a size arrangement map, and the diameter distribution of the equivalent area circle helps determine the overall distribution of particle sizes. This approach effectively solves the tedious process of traditional manual feature extraction while providing more accurate and efficient shape analysis results.

A comprehensive histogram is featured displaying the size distribution of particles in Fig. 10. It demonstrates a cumulative distribution curve of equivalent particle sizes corresponding to the original image. The horizontal axis denotes particle size, and the vertical axis represents the percentage of particles. Notably, the distribution plot reveals that particles with a size of 0.41 mm dominate the raw image dataset, with an average size of 0.61 mm. The arrangement of equivalent

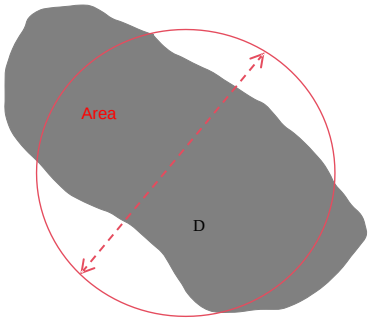


Fig. 9. Diameter of the equivalent area circle.

areas for the particles depicted in Fig. 8(b) is shown in Fig. 10. These size distribution maps serve beyond providing statistical insights into particle size; they extend to facilitating detailed analysis of shape characteristics. The study quantitatively emphasizes the prominence of 0.41 mm particles in the dataset, establishing a crucial foundation for further research. It offers substantial support for applications ranging from classifying and identifying rock particle shapes to assessing engineering performance.

The distribution of rock particles' size and shape characteristics is crucial for gaining insights into geological formations. Smaller particles, although numerous, occupy a relatively small portion of the total area, while a few larger particles dominate the overall area.

To enhance the intuitive representation of this information, the number distribution and area distribution of particles are visually presented. The number of particles in various area ranges as a proportion of the total count is illustrated in Fig. 11(a). At the same time, the distribution of different sizes as a proportion of the particle area is shown. In Fig. 11(b). These particles contribute significantly to areas greater than 0.5, constituting 52.85 percent of the overall space. This observation highlights that particles in the range of areas larger than 0.5 occupy a larger area and contribute more to the overall area. Through this visual representation, the proposed method effectively convey the size distribution of rock particles and provide insights into their inherent shape characteristics.

4.2.2 Shape analysis

In the context of segmentation results, the geometric features of rock particle contours provide essential information for estimating their 2-Dimensional shape. Commonly employed methods for characterizing these shapes include the minimum boundary circle, minimum boundary rectangle, and equivalent ellipse. In this paper, the equivalent ellipse is strategically opted for as the chosen characterization method due to its compatibility with the distinctive appearance characteristics of rock particles. In our approach to establishing a shape-based size distribution, each rock chip is treated as an ellipsoid, with 'a' and 'b' representing the major and minor axes of the equivalent ellipse, as illustrated in Fig. 12. The axis ratios derived from these axes serve as key parameters for analyzing the shape variations among particles in the rock images. The axis ratio is defined as the ratio of the minor axis length to the major axis length.

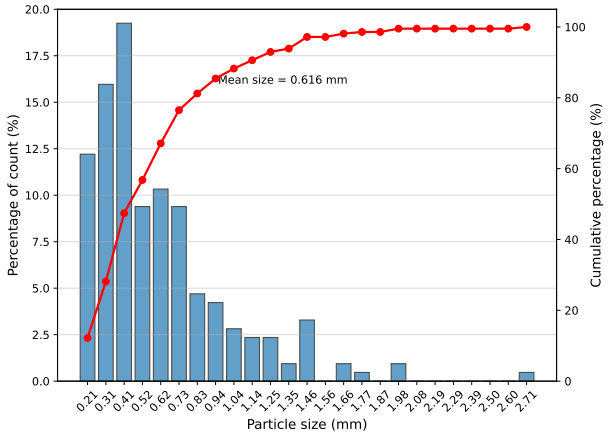


Fig. 10. Size distribution of underwater rock images.

The axial ratio, a crucial metric in particle characterization, delineates the proportional relationship between a particle's length and width. Computed by dividing the shortest axis by the longest, an axial ratio nearing 1 signifies a particle's closeness to a round or spherical shape.

In contrast, a significantly lower ratio suggests elongation or flattening. Analysis of axial ratios unveils valuable insights into particle shapes. The distribution of rock particles based on their axial ratios in our dataset is illustrated in Fig. 13. Notably, a prevalence of ratios exceeding 0.5 indicates a bias toward roundness in the rock particles. This observation is reinforced by a visual inspection of selected dataset images, revealing a predominance of circular or sub-circular shapes and a smaller proportion of elongated particles. The prevalence of axial ratios greater than 0.5 suggests a tendency toward circular or sub-circular shapes in the rock particles of this sample. Leveraging axial ratios enables a quantitative assessment of particle morphology, providing a foundation for rock particles' classification, identification, and engineering performance evaluation.

5. Discussion

Given the complexity of our segmentation approach, particularly its reliance on transformers, it requires significant computational power and memory. This could limit its applicability in resource-constrained environments. Meanwhile, the sophisticated architecture of this approach might result in slower inference times, making it less suitable for applications requiring real-time drilling rock segmentation. In the future, the focus will be on research into more efficient transformer models or pruning techniques helping reduce the computational demands of SAM without sacrificing performance, enabling its deployment on a wider range of platforms. Methods, such as quantization, model distillation, or optimized hardware acceleration, will be developed to speed up inference and make SAM more practical for real-time applications. Also, the implementation of advanced transfer learning techniques is planned to improve SAM's ability to adapt to new domains with limited data, enhancing its performance on specialized tasks.

In rock segmentation, especially in the context of drilling

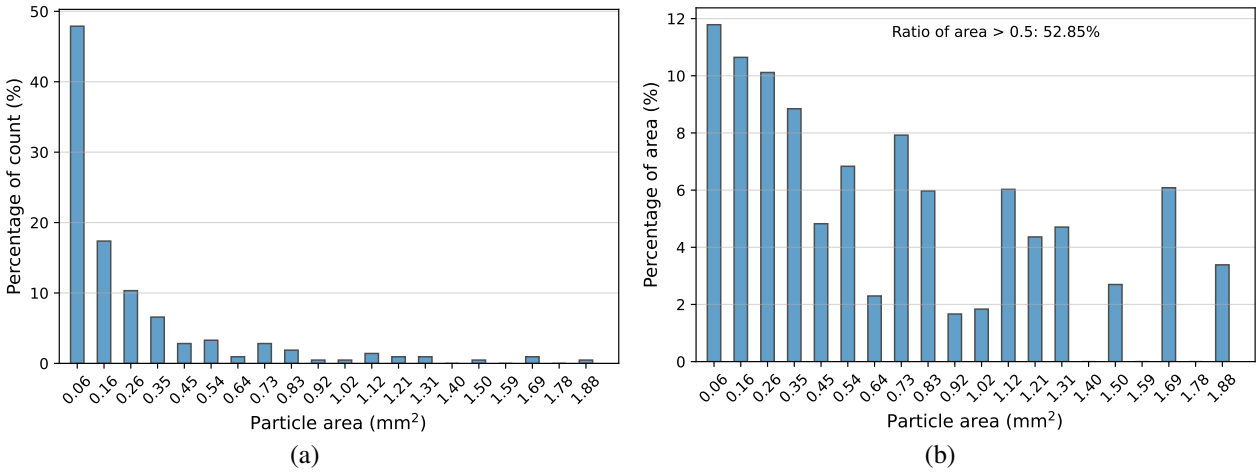


Fig. 11. (a) Number distribution and (b) area distribution of particles.

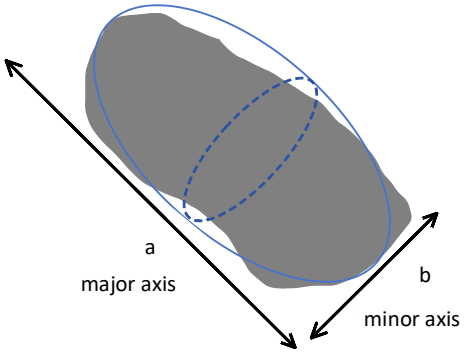


Fig. 12. Equivalent ellipsoid.

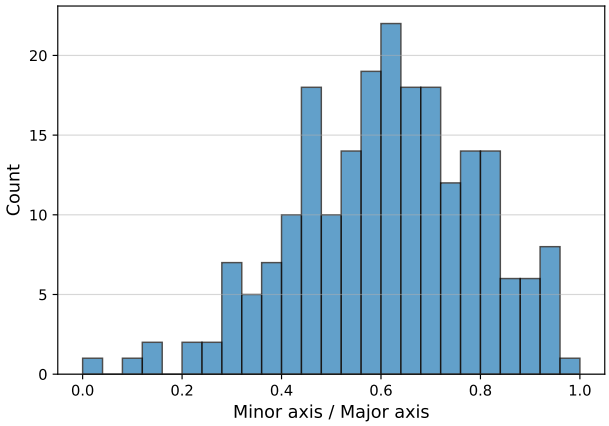


Fig. 13. Axis ratio distribution of rock particles.

where rock particles may be partially covered by other particles, the morphology of these particles can indeed be affected in the images used for analysis. This overlapping or partial coverage can obscure the edges and true shapes of the particles, complicating the segmentation process and potentially leading to inaccuracies in size distribution, shape characterization, and texture analysis.

Dealing with overlapping or partially covered particles is challenging. To mitigate these issues, the application of morphological operations such as erosion, dilation, opening, and closing can be beneficial for separating touching objects and adjusting feature prominence. Additionally, the watershed algorithm serves as a critical tool for segmenting closely situated or overlapping objects by treating the grayscale image as a topographic landscape to delineate object boundaries.

Further advancements in this domain may revolve around the development and utilization of custom-trained models based on deep learning techniques, specifically designed to recognize and accurately segment partially obscured objects within a dataset enriched with examples of overlapping particles. The meticulous annotation of such images, highlighting the contours of obscured particles, is paramount to enhancing model performance. Moreover, the employment of post-segmentation analysis and refinement techniques, focusing on the distributions of particle size, shape, and orientation, presents a promising avenue for correcting segmentation anomalies attributed to partial coverage. This comprehensive approach, combining enhanced segmentation algorithms with deep learning and post-processing adjustments, aims to significantly improve the accuracy and reliability of rock particle segmentation in complex scenarios.

In our study, the issue of particles overlapping at the image's edge was addressed. The challenge of managing overlapping or partially covered particles within the image is designated for exploration in our forthcoming research.

6. Conclusions

In this study, an innovative and comprehensive method for segmenting drilling rock images is introduced, particularly tailored for underwater settings. The multifaceted approach combines the SAM with a suite of sophisticated pre-processing and post-processing techniques to enhance the quality and accuracy of segmentation. These techniques include advanced denoising algorithms, contrast enhancement using CLAHE, and the application of Gaussian filters for smoothing, which together significantly mitigate issues such as overlap, blurring, and textural ambiguity often encountered in underwater imagery. The method demonstrates a marked improvement in

performance metrics, achieving a 10% increase in precision, recall, and F_1 scores over traditional methods like the Mask R-CNN model. The robustness of our approach is further exemplified by its exceptional ability to delineate intricate rock textures and shapes across a diverse array of real-world underwater images. Moreover, the method provides detailed information on rock particle size and shape, which is imperative for geological assessments and rock morphology analysis.

Acknowledgements

This research was financially supported by the Natural Science Foundation of Hebei Province (No. E2021107005). We gratefully acknowledge the helpful comments of the editor and anonymous reviewers.

Conflict of interest

The authors declare no competing interest.

References

Akkaynak, D., Treibitz, T. A revised underwater image formation model. Paper Presented at Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 18-23 June, 2018.

Almutairi, A., Saira, S., Wang, Y., et al. Effect of fines migration on oil recovery from carbonate rocks. Advances in Geo-Energy Research, 2023, 8(1): 61-70.

Anantharaman, R., Velazquez, M., Lee, Y. Utilizing mask R-CNN for detection and segmentation of oral diseases. Paper Presented at Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, 3-6 December, 2018.

Bello, R. W., Mohamed, A. S. A., Talib, A. Z. Enhanced mask R-CNN for herd segmentation. International Journal of Agricultural and Biological Engineering, 2021, 14(4): 238-244.

Cai, J., Zhao, L., Zhang, F., et al. Advances in multiscale rock physics for unconventional reservoirs. Advances in Geo-Energy Research, 2022, 6(4): 271-275.

Carion, N., Massa, F., Synnaeve, G., et al. End-to-end object detection with transformers. Paper Presented at Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23-28 August, 2020.

Chen, H., Jin, Y., Li, G., et al. Automated cement fragment image segmentation and distribution estimation via a holistically-nested convolutional network and morphological analysis. Powder Technology, 2018, 339: 306-313.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. An image is worth 16×16 words: Transformers for image recognition at scale. Paper Presented at International Conference on Learning Representations, Vienna, Austria, 3-7 May, 2021.

Fan, H., Tian, Z., Xu, X., et al. Rockfill material segmentation and gradation calculation based on deep learning. Case Studies in Construction Materials, 2022, 17: e01216.

Frei, M., Kruis, F. Image-based size analysis of agglomerated and partially sintered particles via convolutional neural networks. Powder Technology, 2020, 360: 324-336.

Ganugula, P., Kumar, Y. S. S. S., Reddy, N. K., et al. MOSAIC: Multi-object segmented arbitrary stylization using CLIP. Paper Presented at Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2-3 October, 2023.

Giannakis, I., Bhardwaj, A., Sam, L., et al. Deep learning universal crater detection using Segment Anything Model (SAM). Icarus, 2024, 408: 115797.

Goodfellow, I., Bengio, Y., Courville, A. Deep Learning. Cambridge, USA, MIT Press, 2016.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. Generative adversarial nets. Paper Presented at Advances in Neural Information Processing Systems, Montreal, Canada, 8-13 December, 2014.

He, K., Gkioxari, G., Dollár, P., et al. Mask R-CNN. Paper Presented at Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22-29 October, 2017.

Julka, S., Granitzer, M. Knowledge distillation with segment anything (SAM) model for planetary geological mapping. Paper Presented at International Conference on Machine Learning, Optimization, and Data Science, Grasmere, UK, 22-26 September, 2023.

Kirillov, A., Mintun, E., Ravi, N., et al. Segment anything. Paper Presented at Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1-6 October, 2023.

Liang, Z., Nie, Z., An, A., et al. A particle shape extraction and evaluation method using a deep convolutional neural network and digital image processing. Powder Technology, 2019, 353: 156-170.

Long, J., Shelhamer, E., Darrell, T. Fully convolutional networks for semantic segmentation. Paper Presented at Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 7-12 June, 2015.

Mazurowski, M. A., Dong, H., Gu, H., et al. Segment anything model for medical image analysis: An experimental study. Medical Image Analysis, 2023, 89: 102918.

Minaee, S., Boykov, Y., Porikli, F., et al. Image segmentation using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(7): 3523-3542.

Murphy, K. Machine Learning: A Probabilistic Perspective. Cambridge, USA, MIT Press, 2012.

Osco, L. P., Wu, Q., de Lemos, E. L., et al. The segment anything model (SAM) for remote sensing applications: From zero to one shot. International Journal of Applied Earth Observation and Geoinformation, 2023, 124: 103540.

Qadir, H. A., Shin, Y., Solhusvik, J., et al. Polyp detection and segmentation using Mask R-CNN: Does a deeper feature extractor CNN always perform better? Paper Presented at Proceedings of the International Symposium

on Medical Information and Communication Technology, Oslo, Norway, 8-10 May, 2019.

Rahmon, G., Bunyak, F., Seetharaman, G., et al. Motion U-Net: Multi-cue encoder-decoder network for motion segmentation. Paper Presented at Proceedings of the International Conference on Pattern Recognition, Milan, Italy, 10-15 January, 2021.

Roland, S., Zimmermann, Julien, N., et al. Faster training of Mask R-CNN by focusing on instance boundaries. Computer Vision and Image Understanding, 2019, 188: 102795.

Sahu, S., Singh, A., Ghrera, S., et al. An approach for denoising and contrast enhancement of retinal fundus image using CLAHE. Optics & Laser Technology, 2019, 110: 87-98.

Schult, J., Engelmann, F., Hermans, A., et al. Mask3d: Mask transformer for 3d semantic instance segmentation. Paper Presented at 2023 IEEE International Conference on Robotics and Automation, London, UK, 29 May-2 June, 2023.

Shan, L., Bai, X., Liu, C., et al. Super-resolution reconstruction of digital rock CT images based on residual attention mechanism. Advances in Geo-Energy Research, 2022, 6(2): 157-168.

Suvorov, R., Logacheva, E., Mashikhin, A., et al. Resolution-robust large mask inpainting with fourier convolutions. Paper Presented at Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, USA, 3-8 January, 2022.

Tahir, H., Khan, M. S., Tariq, M. O. Performance analysis and comparison of faster R-CNN, mask R-CNN and ResNet50 for the detection and counting of vehicles. Paper Presented at Proceedings of the International Conference on Computing, Communication, and Intelligent Systems, Greater Noida, India, 19-20 February, 2021.

Toldo, M., Maracani, A., Michieli, U., et al. Unsupervised domain adaptation in semantic segmentation: A review. Technologies, 2020, 8(2): 35.

Ullo, S., Mohan, A., Sebastianelli, A., et al. A new mask R-CNN-based method for improved landslide detection. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 3799-3810.

Wen, J., Cui, J., Zhao, Z., et al. SyreaNet: A physically guided underwater image enhancement framework integrating synthetic and real images. Paper Presented at Proceedings of the IEEE International Conference on Robotics and Automation, London, United Kingdom, 29 May-2 June, 2023.

Yamada, T., Di Santo, S. Instance segmentation of piled rock particles based on mask R-CNN. Paper Presented at Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17-22 July, 2022.

Yamagiwa, H., Takase, Y., Kambe, H., et al. Zero-shot edge detection with SCESAME: Spectral clustering-based ensemble for segment anything model estimation. Paper Presented at Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Hawaii, USA, 4-8 January, 2024.

Zhao, H., Zhang, Y., Liu, S., et al. PSANet: Point-wise spatial attention network for scene parsing. Paper Presented at Proceedings of the European Conference on Computer Vision, Munich, Germany, 8-14 September, 2018.

Zhou, X., Gong, Q., Liu, Y., et al. Automatic segmentation of TBM muck images via a deep-learning approach to estimate the size and shape of rock chips. Automation in Construction, 2021, 126: 103685.