Canadian Geotechnical Journal

Research Article

Enhancement and assessment of large vision models for 3D particle reconstruction from X-ray tomography

Ruidong Li[a], **Zhen-Yu Yin**[a], **Shaoheng He**[a], **and Brian Sheil**[b]

[a]Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong, China; [b]Department of Engineering, University of Cambridge, Cambridge, United Kingdom

Corresponding author: **Zhen-Yu Yin** (email: zhenyu.yin@polyu.edu.hk)

Abstract

Three-dimensional (3D) particle reconstruction from X-ray micro-computed tomography (μCT) images is essential for digital twins and understanding the micromechanical behaviors of granular media. Despite large vision models (LVMs) having shown remarkable effectiveness across various domains, their application to accurate 3D particle reconstruction remains underexplored. This study proposes a systematic framework that enhances and leverages LVMs for the reconstruction of arbitrary 3D particles. The proposed framework includes three key steps: (1) enhancing LVMs with higher computational efficiency for two-dimensional (2D) label extraction, (2) mapping stacked 2D labels to 3D, and (3) extracting particle surfaces to generate 3D models. The enhanced approach is applied to reconstruct four distinct samples to validate feasibility. Six conventional and four lightweight LVMs are selected to explore the influence of model size and the number of prompts on reconstruction accuracy. The H-extreme watershed method is chosen as a benchmark for comparison. The results demonstrate that the enhanced framework can accurately reconstruct irregular and complex samples, such as carbonate sands, with a greater than 50% improvement in accuracy compared to the benchmark prediction. Additionally, the framework effectively reduces over- and under-segmentation errors, resulting in accurate reconstruction of microstructural characteristics. This versatile framework presents a promising alternative for investigating complex micromechanical mechanisms of granular media.

Key words: X-ray tomography, machine learning, large vision models, particle reconstruction, particle shape

Résumé

La reconstruction tridimensionnelle (3D) des particules à partir d'images de tomographie micro-informatique à rayons X (μCT) est essentielle pour les jumeaux numériques et la compréhension des comportements micromécaniques des milieux granulaires. Bien que les modèles de vision à grande échelle (LVM) aient démontré une efficacité remarquable dans divers domaines, leur application à la reconstruction 3D précise de particules demeure peu explorée. Cette étude propose un cadre systématique qui améliore et exploite les LVM pour la reconstruction de particules 3D arbitraires. Le cadre proposé comprend trois étapes clés : (1) l'amélioration des LVM avec une plus grande efficacité de calcul pour l'extraction d'étiquettes bidimensionnelles (2D), (2) le mappage des étiquettes 2D empilées en 3D, et (3) l'extraction des surfaces de particules pour générer des modèles 3D. L'approche améliorée est appliquée à la reconstruction de quatre échantillons distincts afin d'en valider la faisabilité. Six LVM conventionnels et quatre LVM légers sont sélectionnés pour étudier l'influence de la taille du modèle et du nombre d'invites sur la précision de la reconstruction. La méthode des bassins versants H-extreme est choisie comme référence pour la comparaison. Les résultats démontrent que le cadre amélioré permet de reconstruire avec précision des échantillons irréguliers et complexes, tels que les sables carbonatés, avec une amélioration de plus de 50 % de la précision par rapport à la prédiction de référence. En outre, le cadre réduit efficacement les erreurs de sur-segmentation et de sous-segmentation, ce qui permet de reconstruire avec précision les caractéristiques microstructurelles. Ce cadre polyvalent constitue une alternative prometteuse pour l'étude des mécanismes micromécaniques complexes des milieux granulaires.

Mots-clés : tomographie à rayons X, apprentissage automatique, modèles de vision à grande échelle, reconstruction de particules, forme des particules

1. Introduction

Granular media such as sands are ubiquitous in nature and play extremely important roles in various engineering and industrial processes, including pharmaceuticals, mining, chemical, and civil engineering (Jaeger et al. 1996; Huang et al. 2022; Lin et al. 2023, 2024; Chang et al. 2024; Xie et al. 2024, 2025). Composed of particles packed in an inhomogeneous nature, granular media exhibit extremely complex

Fig. 1. Overview of the proposed framework. μCT, micro-computed tomography; LVM, large vision model.

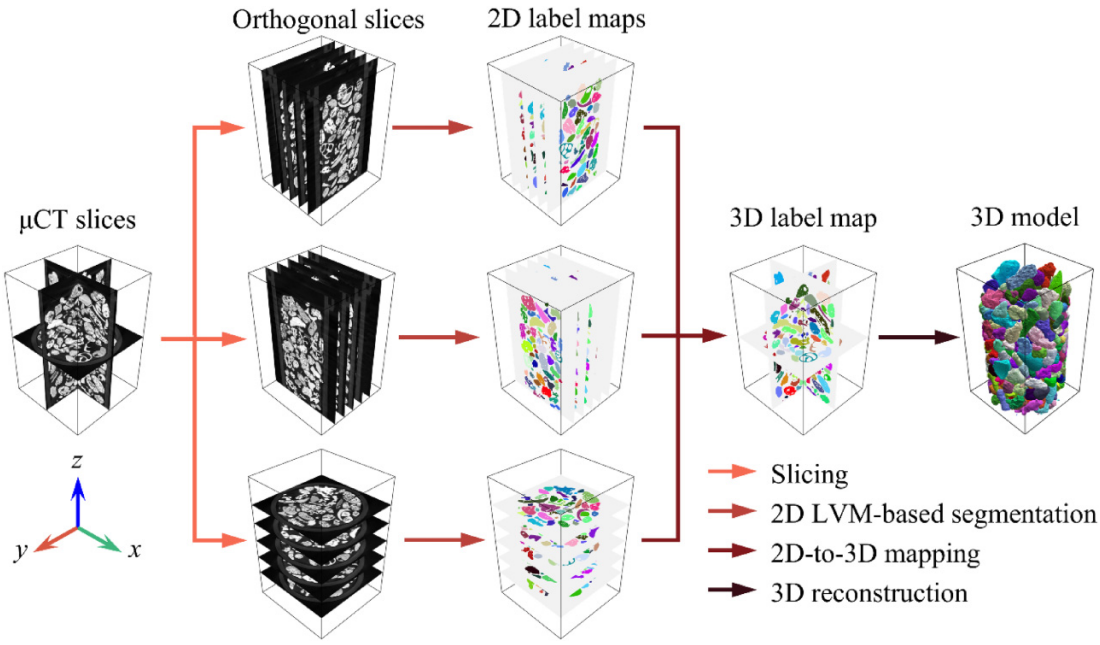


Table 1. Overview of the considered large vision models and their corresponding properties; note: M denotes million.

Name	Model	Branch	Model size (M)	Release date
SAM-ViT-H	SAM	ViT-H	641.1	04/2023
SAM-ViT-L	SAM	ViT-L	308	04/2023
SAM-ViT-B	SAM	ViT-B	91	04/2023
HQ-SAM-ViT-H	HQ-SAM	ViT-H	636.044	06/2023
HQ-SAM-ViT-L	HQ-SAM	ViT-L	308.044	06/2023
HQ-SAM-ViT-B	HQ-SAM	ViT-B	91.044	06/2023
HQ-SAM-ViT-T	HQ-SAM	ViT-Tiny	21.044	06/2023
MobileSAM	MobileSAM	\	9.8	06/2023
EfficientViT-SAM-l1	EfficientViT-SAM	ViT-l1	47.7	02/2024
EfficientViT-SAM-l0	EfficientViT-SAM	ViT-l0	34.8	02/2024

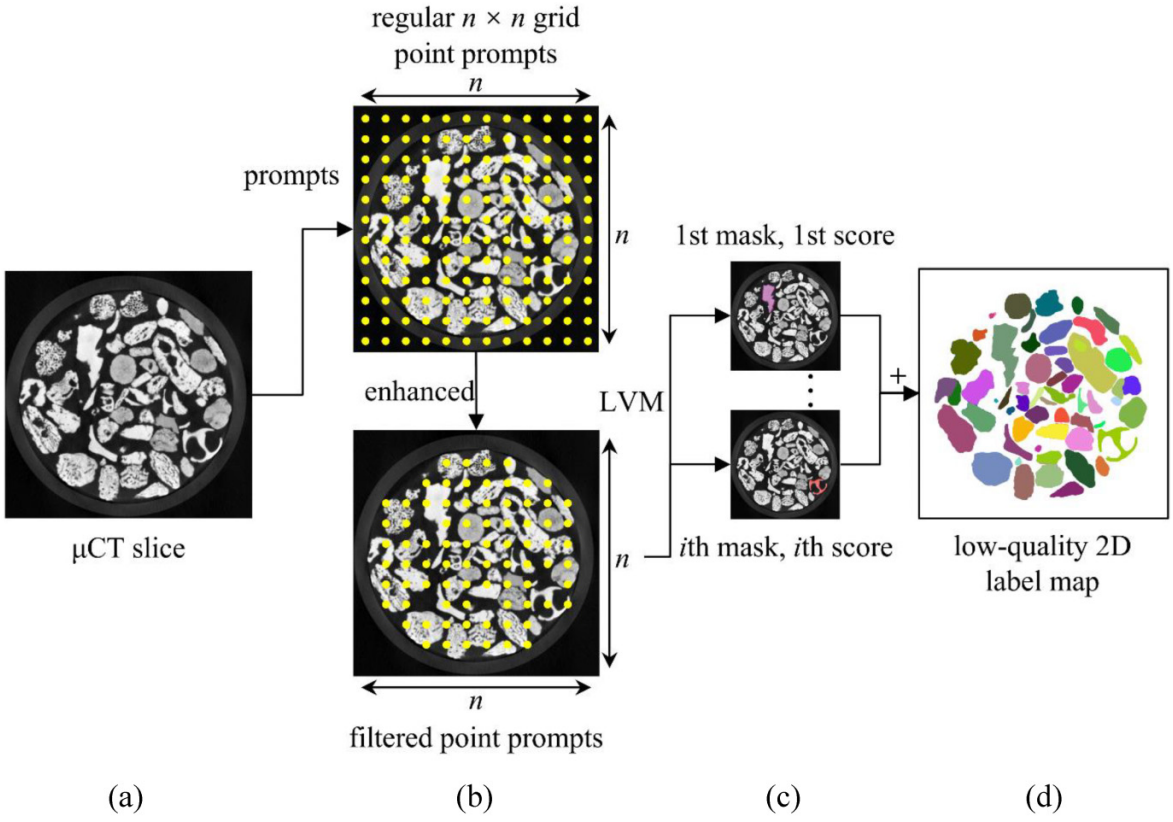
Note: SAM, segment anything model; ViT, vision transformer.

macroscopic mechanical behaviors, such as stress path dependency (Cai et al. 2013; He et al. 2025), stress dilatancy (Xiao et al. 2019), anisotropy (Guo and Zhao 2013; He et al. 2024*b*), critical state flow (Zhao and Guo 2013), particle breakage (He et al. 2023), and jamming transition (Corwin et al. 2005). However, the origin and mechanism of complicated macroscopic responses in granular media remain poorly understood.

Microstructure characteristics, such as morphology and topology, of granular media have been gradually recongized to influence the macroscopic behaviors significantly, as evidenced by many microscale tests and numerical simulations (Darve et al. 2007; Tang et al. 2024*a*, 2024*b*). The advent of X-ray micro-computed tomographic (μCT) imaging has paved the way for more accurate quantification of the microstructure of real-world granular media (Beister et al. 2012). μCT imaging enables measurement and visualization of the internal structure of granular media non-destructively through three-dimensional (3D) particle reconstruction. This technique allows the identification of all individual particles in a sample of interest directly from μCT images, facilitating a deeper understanding of the intrinsic micromechanical behaviors of granular media.

There are three traditional image-processing approaches for 3D particle reconstruction: thresholding segmentation, topology-aware segmentation, and watershed segmentation. Thresholding segmentation uses single- or multiple thresholds to distinguish objects with different grayscale values (of images) (Kohler 1981). This approach works well in segmenting background and particles but fails to identify and separate particles in contact. Topology-aware segmentation, such as the Morse theory-based framework (Pandey et al. 2022), is a relatively new method that enables the segmentation of connected particles. However, a primary weakness of topology-aware segmentation is the complicated method

Fig. 2. Image segmentation by large vision model (LVM): (*a*) micro-computed tomographic (μCT) slice; (*b*) μCT slice with prompts; (*c*) scored prompt and corresponding label; and (*d*) low-quality 2D label map.

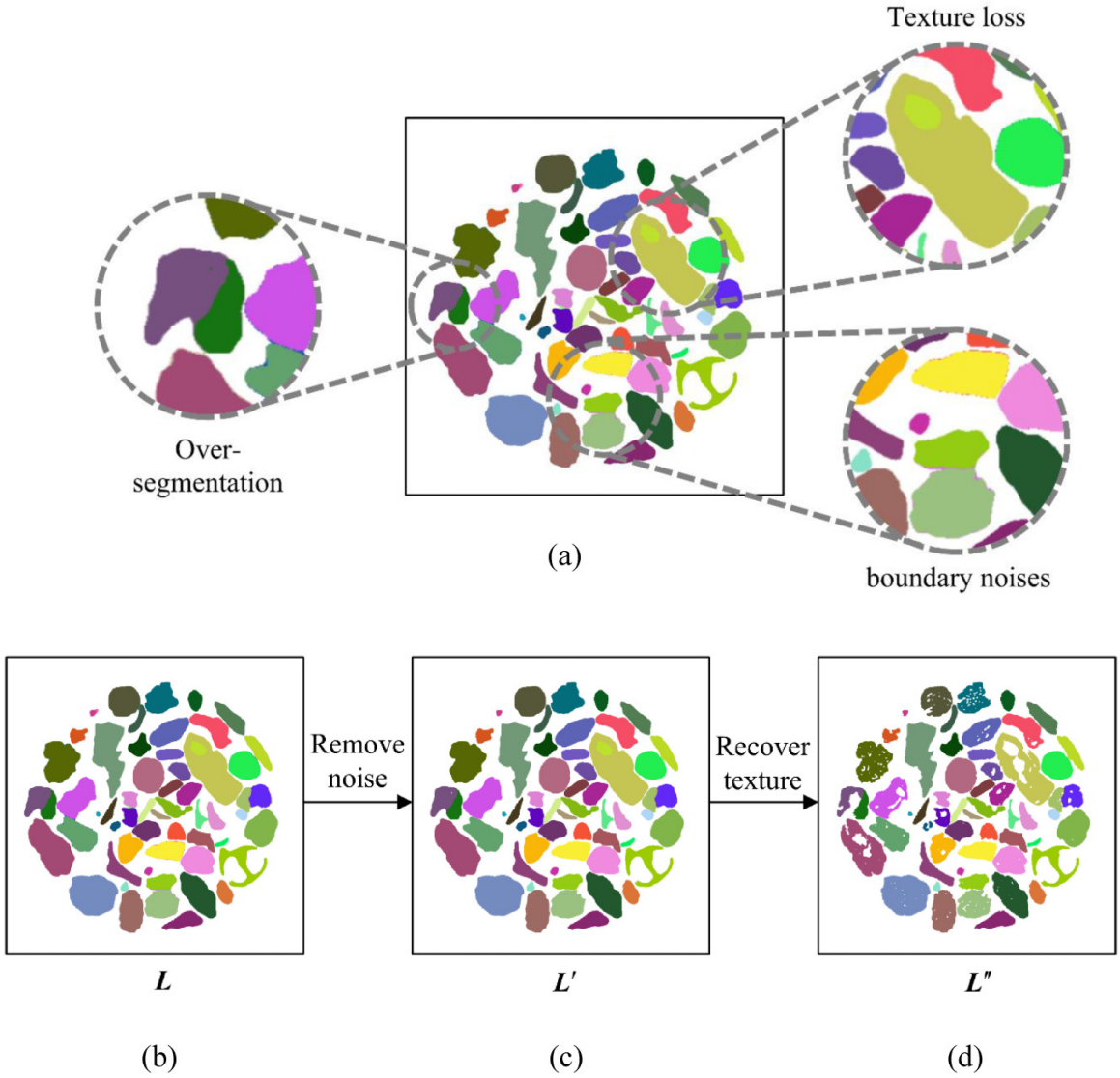


pipeline that necessitates considerable specialist expertise and lacks validation on complex granular media (e.g., carbonate sand). Watershed segmentation is the most popular method for 3D particle reconstruction (Shi and Yan 2015; Sun et al. 2019) and separates particles in contact by identifying the watershed lines between local minima in the inverted Euclidean distance transform map. Determining the local minima correctly is highly challenging, as many local minima arise due to noise artifacts such as irregular particle morphology and interior voids. While some modifications have improved segmentation accuracy (e.g., the H-extreme transform (Burgmann et al. 2022) and filling holes before segmentation), the watershed algorithm remains highly sensitive to noise and performs poorly when applied to irregular particles.

Recently, machine learning (ML), such as random forest (Zhao et al. 2024), has become a popular approach for more robust particle reconstruction. For example, the Trainable Weka Segmentation (TWS) algorithm (Lai and Chen 2019) uses pixel features in blurred μCT training images to segment particles from the background and then apply level set evolution for 3D reconstruction manually. However, preparing training datasets for each case and manually assigning level sets are expensive and time-consuming. Zhang et al. (2022) improved the assignment of the initial level set using a convolution kernel, though the approach was inefficient due to redundant assignments. To address this problem, Li et al. (2024) proposed a two-dimensional (2D) segmentation scheme to quickly locate the initial level sets. The improved TWS method performs well in particle reconstruction from noisy images but only applies to single-layer particles, limiting broader application. Other ML-based approaches, such as convolutional neural networks (CNNs) focusing on 2D images, face similar challenges (Cevallos et al. 2023; Zhao et al. 2023*a*, 2023*b*).

Large vision models (LVMs), such as the segment anything model (SAM) (Kirillov et al. 2023) and "SAMHQ" (Ke et al. 2023), have received considerable recent interest for image segmentation. These models are trained on large-scale datasets and are adaptable to a variety of subsequent downstream tasks, meaning they are often considered superior alternatives to conventional ML methods (Brown et al. 2020; Devlin et al. 2019; Radford et al. 2021). Their growing use in two-dimensional (2D) medical image segmentation has demonstrated their feasibility and robustness for real-world applications (Wu et al. 2023; Zhang et al. 2023). While LVMs currently enable only 2D image segmentation, they also hold significant potential for improving 3D particle reconstruction accuracy. However, no LVMs have yet been applied to reconstruct 3D particles from μCT images. The main challenges rest in the lack of a rigorous strategy for restoring internal features of grains (e.g., holes) in 2D images and the conversion from stacked 2D images to 3D reconstructions. In addition, no systematic evaluation of LVMs for 2D segmentation of granular media has been conducted, nor has there been a comparative study to highlight the differences

Fig. 3. 2D label maps: (*a*) problems; (*b*) (***L***); (*c*) ***L***′; and (*d*) ***L***″.



between LVMs for this task. Therefore, it is of interest to the geotechnical community to explore, verify, and benchmark the application of LVMs to 3D particle reconstruction problems.

To this end, this paper introduces an enhanced framework for automatically reconstructing 3D particle reconstruction from μCT images using LVMs. The main academic contributions of this paper include: (1) A novel framework that integrates and enhances LVMs into 3D arbitrary particle reconstruction is proposed; (2) a systematic quantitative assessment of the performance of a range of LVMs for 3D particle reconstruction tasks, including the influence of model types, model sizes, and number of prompts, is conducted; (3) a robust algorithm for converting stacked 2D images into 3D models is proposed based on the optimal transport (OT) technique is developed; (4) a rigorous validation of the effectiveness and accuracy of reconstruction through analysis of microstructure characteristics is provided.

2. Methodology

2.1. Overview

The proposed reconstruction framework comprises three steps: 2D particle segmentation, 2D-to-3D mapping, and 3D particle reconstruction, as illustrated in Fig. 1. The 2D segmentation step is achieved in an automated way using LVMs, turning raw 2D image "slices" of the soil sample along the *x*-, *y*-, and *z*-axes (in turn) into corresponding high-quality 2D label maps. The 2D-to-3D mapping involves stitching 2D label maps along the three principal axes and applying voxel mapping to produce a single 3D label map. Finally, the 3D particle reconstruction step generates mesh-based 3D particles using the marching cube (MC) algorithm.

2.2. 2D particle segmentation

The first LVM designed for image segmentation tasks is the SAM released by Meta AI (Kirillov et al. 2023). SAM consists of

Fig. 4. Schematic illustration of the present 2D-to-3D mapping process.

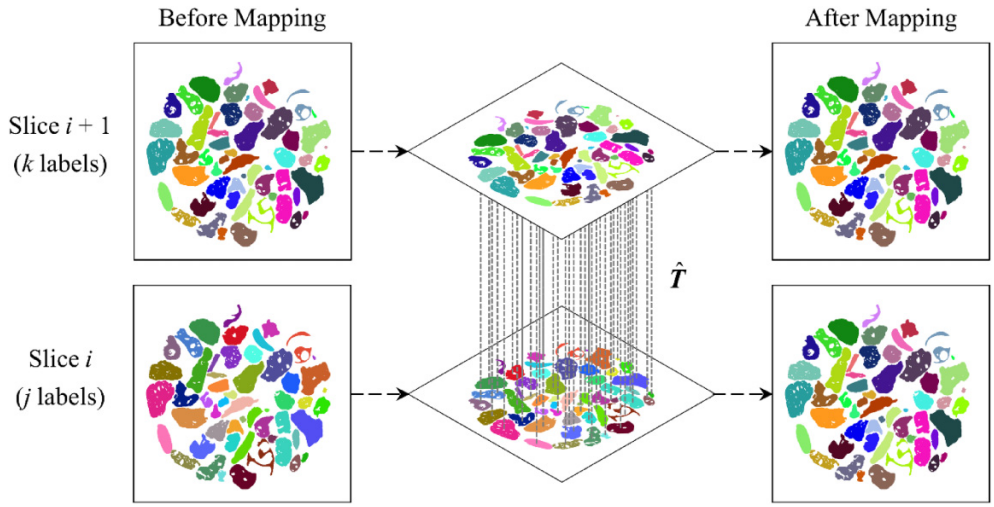


Table 2. Summary of present particle datasets.

Dataset	Material	10% passing diameter, D_{10} (mm)	90% passing diameter, D_{90} (mm)	Particle numbers	Image size ($X \times Y \times Z$ pixels)
CS1	Carbonate sand	1.03	1.62	2123	$700 \times 700 \times 800$
CS2	Carbonate sand	1.39	2.95	501	$700 \times 700 \times 1056$
FJS	Fujian sand	1.00	1.41	4424	$750 \times 750 \times 1032$
CGB	Concave glass bead	1.11	1.74	3138	$800 \times 800 \times 1066$

three components: an image encoder, a prompt encoder, and a mask decoder. The image encoder downsamples the input using a Vision Transformer (ViT), which splits the image into patches, embeds them, and applies self-attention to model global relationships for image recognition. This approach enables it to capture both local and global features without relying on traditional convolutions used in CNNs. Prompts in SAM refer to user inputs that guide the model in identifying the regions of interest within an image. The prompt encoder enables (a) dense masks or (b) sparse (points, boxes, text) prompts (Radford et al. 2021; Ren et al. 2023). Finally, the mask decoder generates a segmentation mask based on the input prompt and assigns a confidence score to indicate how certain it is about the accuracy of the mask. SAM can be classified into three different types based on the model's size and capacity, primarily determined by the type of ViT backbone: ViT-B (91 M parameters), ViT-L (308 M), and ViT-H (636 M), where "B", "L", and "H" stand for "base", "large", and "huge", respectively.

Recently, several SAM variants have been introduced to improve computational efficiency and accuracy, including HQ-SAM, MobileSAM, and EfficientViT-SAM (Ke et al. 2023; Zhang et al. 2023, 2024). For instance, HQ-SAM incorporates a learnable "high-quality output token" into the mask decoder, which is fused with ViT features twice to enhance mask detail. HQ-SAM was trained on a dataset of 44k masks over just four hours using eight Nvidia GeForce RTX 3090 GPUs, yet outperformed the original SAM on the COCO dataset which is the most widely used dataset for general image segmentation (Lin et al. 2014). MobileSAM replaced the heavyweight image encoder with a lighter alternative to reduce computational time. The resulting inference speed of MobileSAM is more than 50 times faster than that of SAM. Similarly, EfficientViT-SAM used EfficientViT as its lightweight image encoder, achieving a major boost in efficiency without sacrificing performance. In total, ten different types of LVMs are considered to explore the influence of model type, model size, and number of prompts on segmentation performance, as summarized in Table 1. The first six models are classified as conventional LVMs due to their larger model sizes, while the remaining models are classified as lightweight LVMs. It is evident that lightweight SAMs are becoming a growing trend. However, it remains unclear whether these lightweight SAMs are sufficient or even superior to the original SAM for image segmentation tasks.

The LVMs are used for automatic segmentation of all particles in a given μCT soil slice image, such as the one depicted in Fig. 2*a*. The original prompts recommended by LVMs are a regular grid of $n \times n$ points, but increase computational costs significantly due to unnecessary assignment of prompts located at the background, as shown in Fig. 2*b*. To enhance the performance and efficiency of LVMs, only prompts corresponding to gray-scale values exceeding the threshold determined by Otsu segmentation (Otsu 1979) are retained, as illustrated in Fig. 2*b*. This enhancement will reduce computational time by at least 50% while guarantee robust segmentation. To investigate the influence of prompt numbers, values of the "prompt size" variable n of 8, 16, 24, 32, and 40

Fig. 5. Example micro-computed tomographic (μCT) slices corresponding to the four particle datasets: (*a*) CS1; (*b*) CS2; (*c*) FJS; and (*d*) CGB.

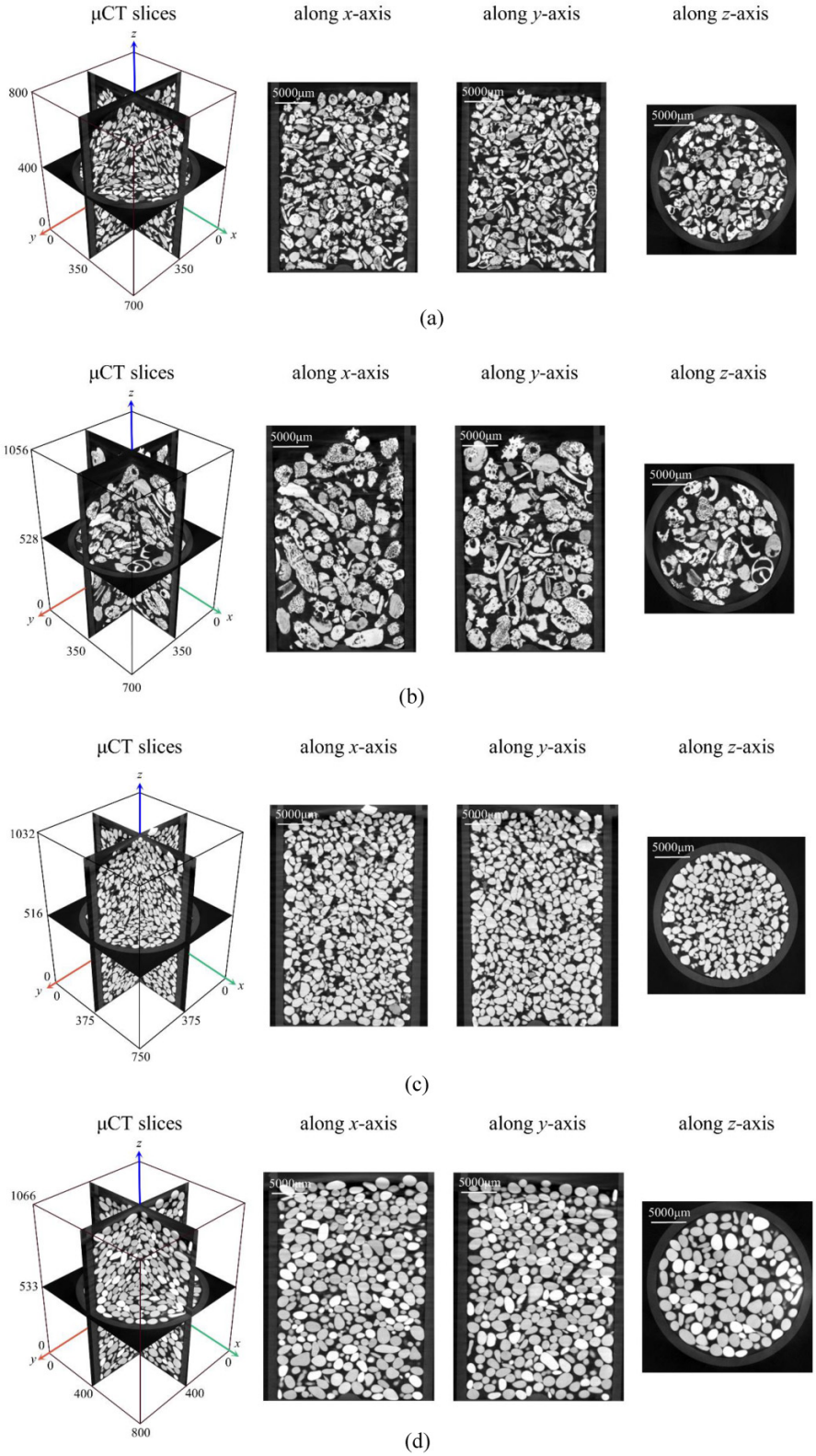


Fig. 6. Example ground truth labelling for the four particle datasets: (*a*) CS1; (*b*) CS2; (*c*) FJS; and (*d*) CGB. μCT, micro-computed tomography.

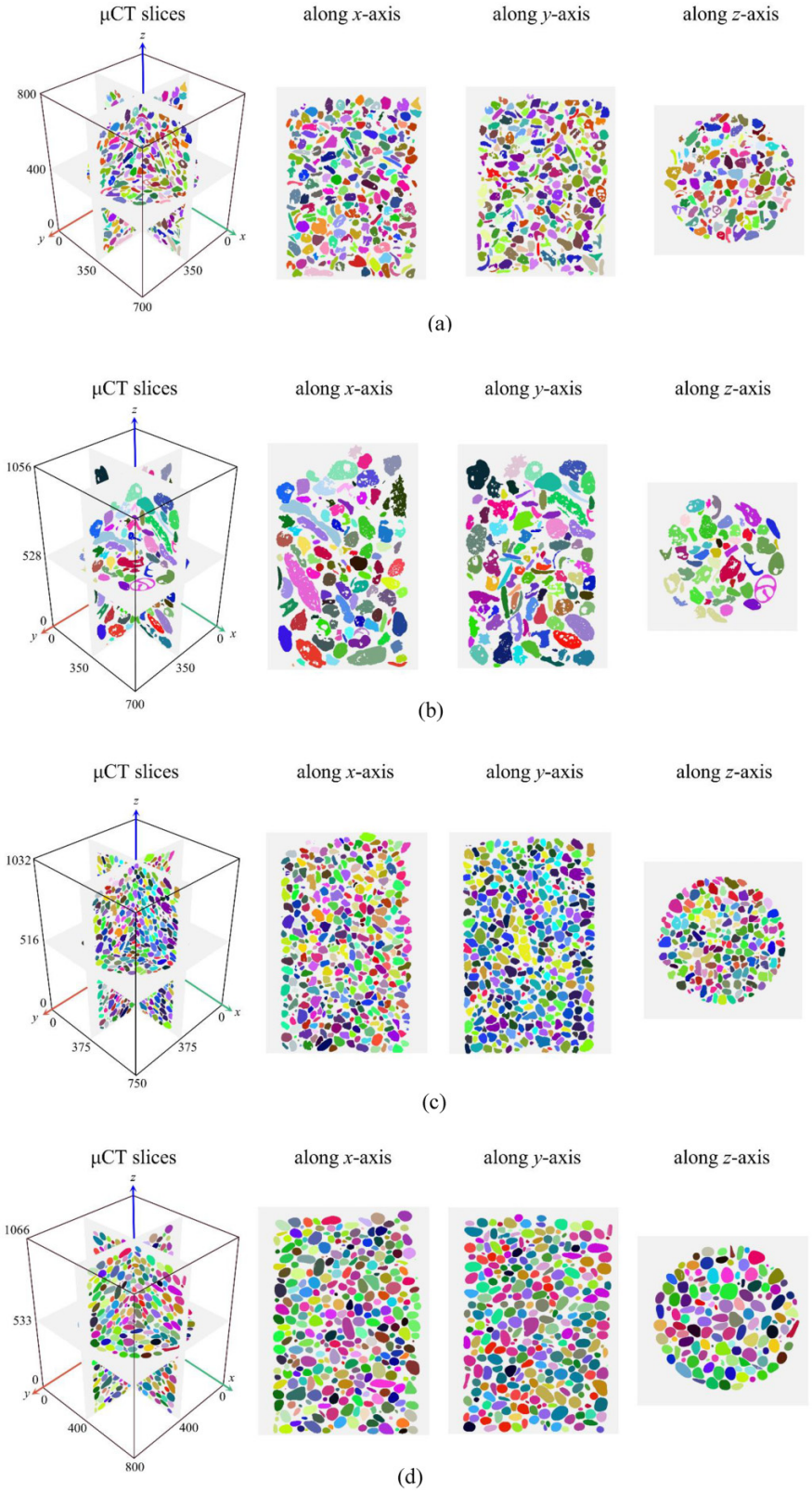


Fig. 7. Sensitivity analysis of ξ. AP, average precision.

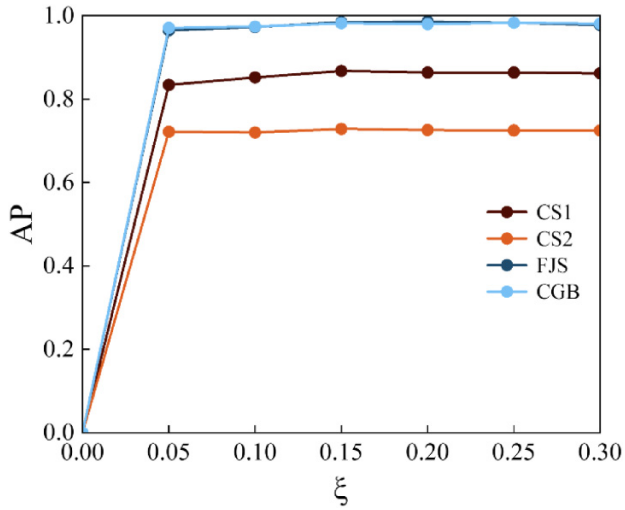
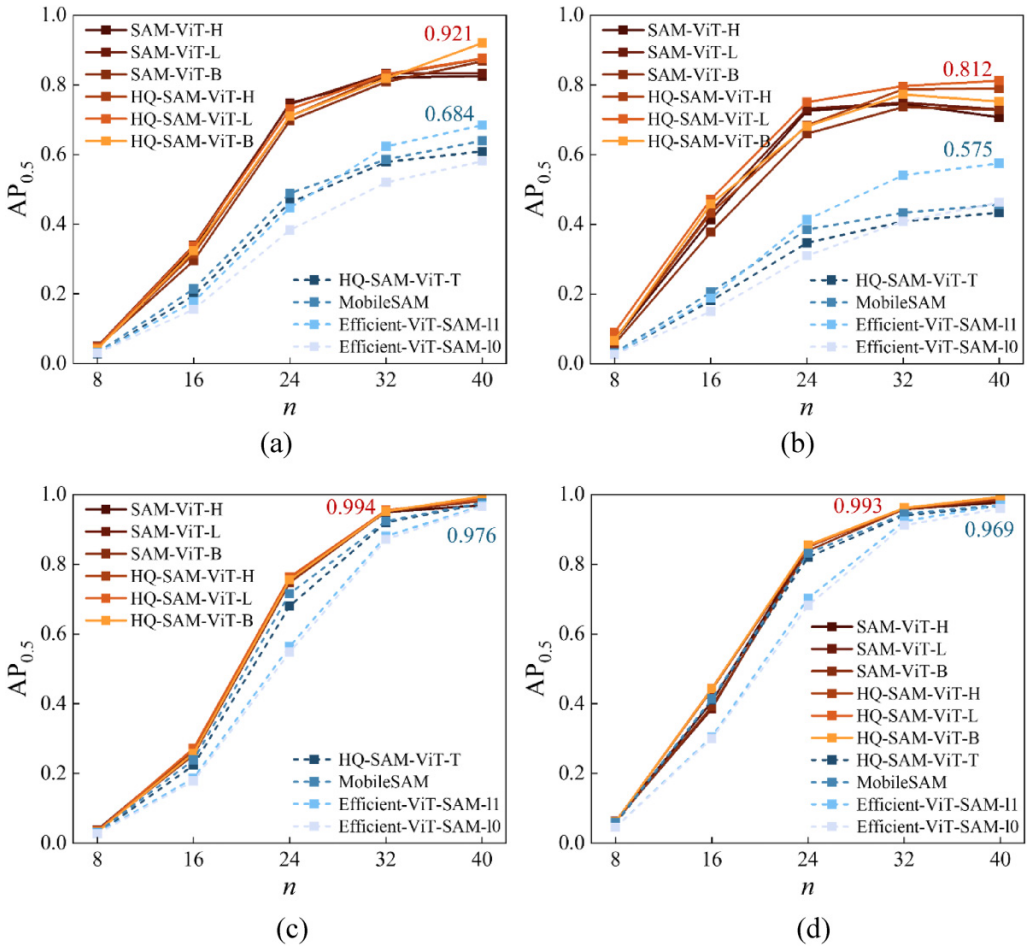
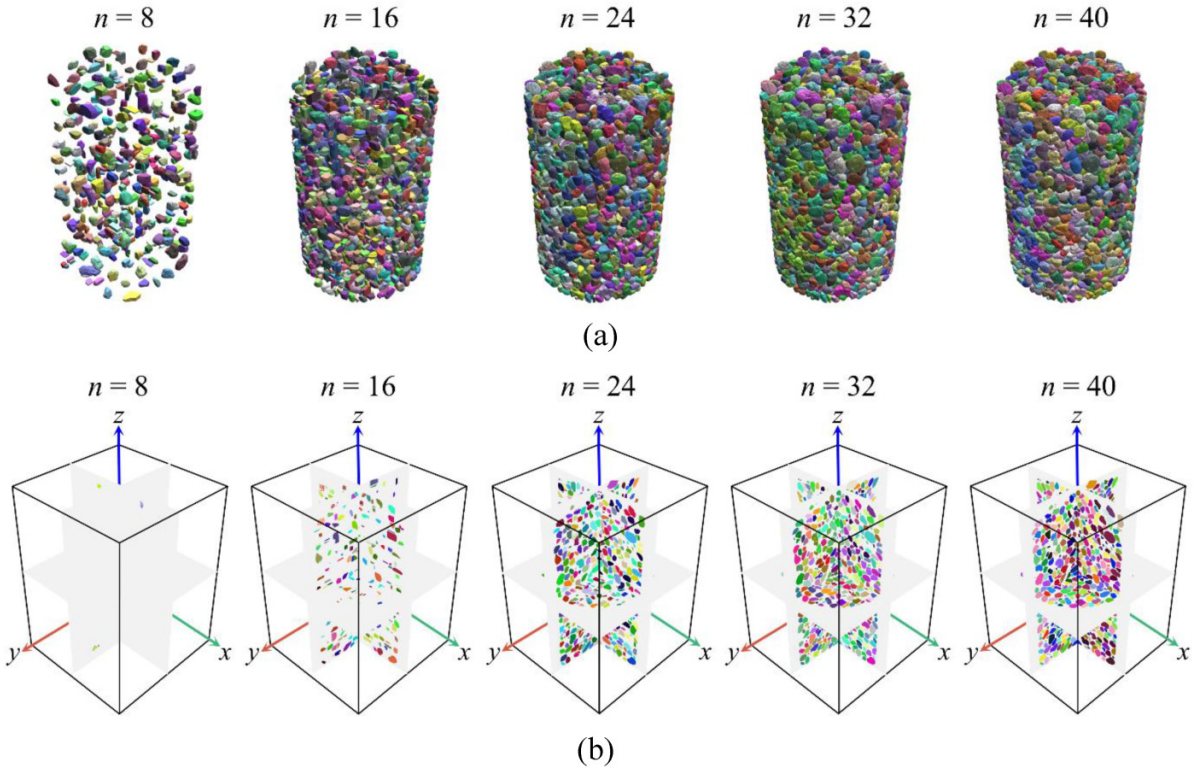


Fig. 8. Influence of large vision model type and prompt size parameter (n) on the segmentation performance ($AP_{0.5}$) using the proposed framework: (*a*) CS1; (*b*) CS2; (*c*) FJS; and (*d*) CGB. The $AP_{0.5}$ values of the best-performing conventional and lightweight models are superimposed in red and blue text, respectively. AP, average precision.



are considered. For each input prompt, the LVM generates a mask with an associated confidence score. Masks with unsatisfactory scores are filtered out, and the remaining masks are sorted by size in descending order, as presented in Fig. 2*c*. These masks are then mapped to create a low-quality 2D label map, as shown in Fig. 2*d*.

Fig. 9. Exemplar reconstruction results showing the influence of prompt size (n) using the HQ-SAM-ViT-H model for CS1: (*a*) 3D models; and (*b*) orthogonal slices.



While particle boundaries are preserved, they are coarse and inadequate for direct use. For example, numerous small connected pixels, or "noise", are visible around the particle boundaries and need to be removed, as shown in Fig. 3*a*. Additionally, internal textures such as voids are ignored. To address these issues, a two-step approach is employed, as illustrated in Fig. 3*b*. The first step is to eliminate the connected pixels around particle boundaries, based on their size, in an updated label map $\boldsymbol{L}'$ (see Fig. 3*c*):

$$\boldsymbol{L}' = \{\boldsymbol{l} \in \boldsymbol{L} | A(\boldsymbol{l}) \geq A_0\} \tag{1}$$

where $\boldsymbol{L}$ denotes the original low-quality label map, $\boldsymbol{l}$ is the label map for a given single label in $\boldsymbol{L}$, $A(\boldsymbol{l})$ is the label area in $\boldsymbol{l}$, and A_0 is the predetermined threshold area. The next procedure is to recover interior textures by pixel intensity threshold resulting in a further update to the label map ($\boldsymbol{L}''$):

$$\boldsymbol{L}'' = \left\{\boldsymbol{l} \in \boldsymbol{L}' | Q(\boldsymbol{l})\right\} \tag{2}$$

$$Q(\boldsymbol{l}) = \begin{cases} \text{OTSU}\left(\boldsymbol{I} \circ \boldsymbol{l}\right), & \text{if } \bar{I} \geq I_0 \\ \boldsymbol{O}, & \text{if } \bar{I} < I_0 \end{cases} \tag{3}$$

where the symbol $\circ$ represents the Hadamard product, OTSU $(\cdot)$ is the Otsu segmentation operator (Otsu 1979), $\boldsymbol{I}$ denotes the raw image, $\boldsymbol{O}$ is a zero matrix, $\bar{I}$ denotes the average pixel value at the position of the label in $\boldsymbol{I}$, and I_0 is the threshold grey-scale value. These steps transform low-quality 2D label maps into high-quality versions, as demonstrated in Fig. 3*d*. All low-quality 2D label maps obtained from the LVM along the x, y, and z-axes are subsequently processed using this method.

2.3. 2D-to-3D mapping

The 2D-to-3D mapping process involves stitching the 2D label maps and voxel mapping to generate a complete 3D label map, as shown in Fig. 4. The stitching of 2D label maps is achieved using an automated OT method, which sequentially updates a stack of 2D label maps, effectively combining them into a single 3D label map (Liu et al. 2023). Specifically, any adjacent 2D label maps $\boldsymbol{L}_i \in \mathbb{R}$ and $\boldsymbol{L}_{i+1} \in \mathbb{R}$, can be treated as uniform discrete probability distributions containing j and k items (or labels), respectively, in terms of OT:

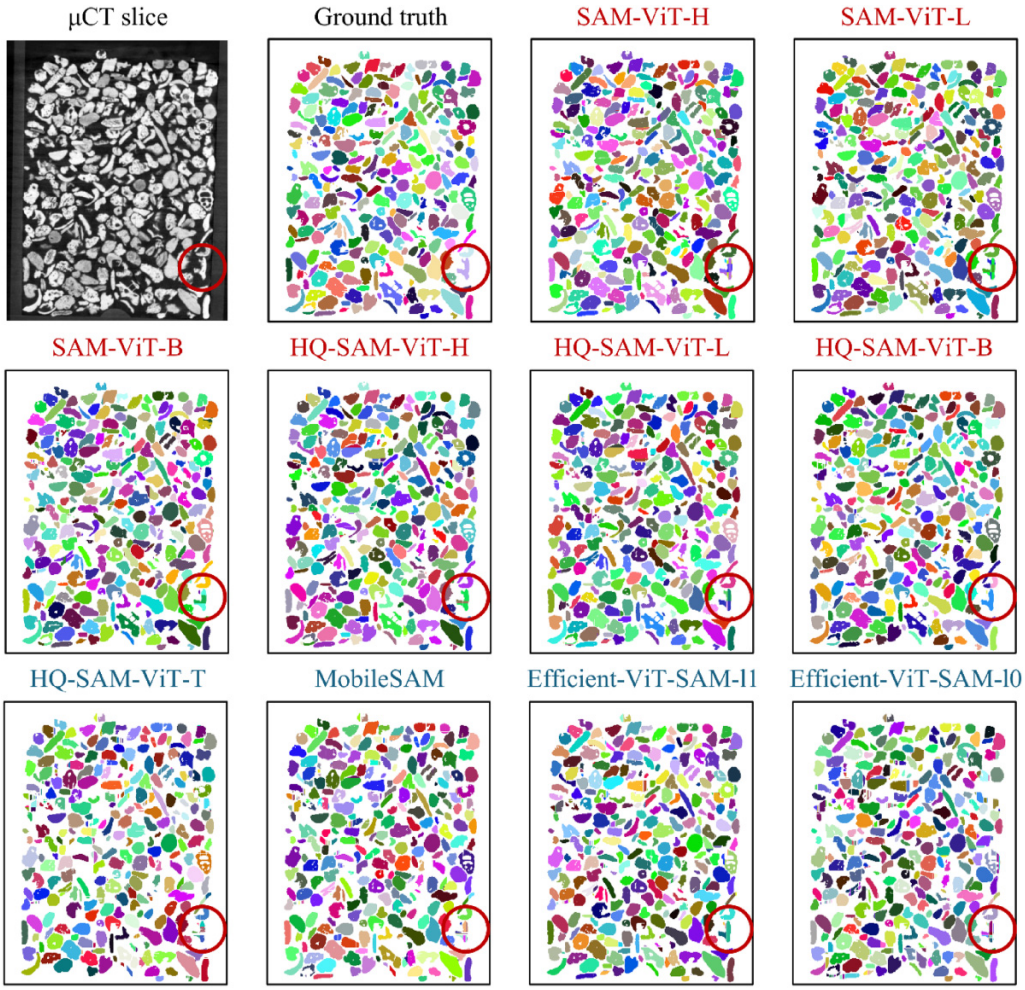
$$\boldsymbol{L}_i(p) = \frac{1}{j} \tag{4}$$

$$\boldsymbol{L}_{i+1}(q) = \frac{1}{k} \tag{5}$$

where p and q denote the label in $\boldsymbol{L}_i$ and $\boldsymbol{L}_{i+1}$, respectively. OT outputs a plan $\widehat{\boldsymbol{T}}_{i,i+1}$ that describes how to update labels in $\boldsymbol{L}_{i+1}$ so that the same particles in $\boldsymbol{L}_i$ and $\boldsymbol{L}_{i+1}$ are denoted by the same label. The calculation of OT is summarized as follows:

$$\widehat{\boldsymbol{T}}_{i,i+1} = \arg\min_{T \in \Gamma(L_i, L_{i+1})} \sum_{p,q} C_{p,q} T_{p,q} \tag{6}$$

Fig. 10. Comparisons between the raw micro-computed tomographic (μCT) slice, ground truth labels, and representative slices of a reconstructed CS1 sample generated by the 10 different enhanced large vision models ($n = 40$); note the red circular inset highlights specific regions used for comparison.



$$\Gamma(\boldsymbol{L}_i, \boldsymbol{L}_{i+1}) = \left\{ \boldsymbol{T} \in \mathbb{R} : \sum_q T_{p,q} = \boldsymbol{L}_i, \sum_p T_{p,q} = \boldsymbol{L}_{i+1} \right\} \tag{7}$$

where $T_{p,q}$ represents the matching between labels p and q, and $C_{p,q}$ is the cost function. The definition of the cost function $C_{p,q}$ influences key properties of the OT problem, including the uniqueness of solutions and the computational cost of solving the optimization problem. A popular choice for computing the cost matrix is the l_p-norm distance, where $l_p(x, y) = \|x - y\|^p$. In this study, a new cost function is proposed:

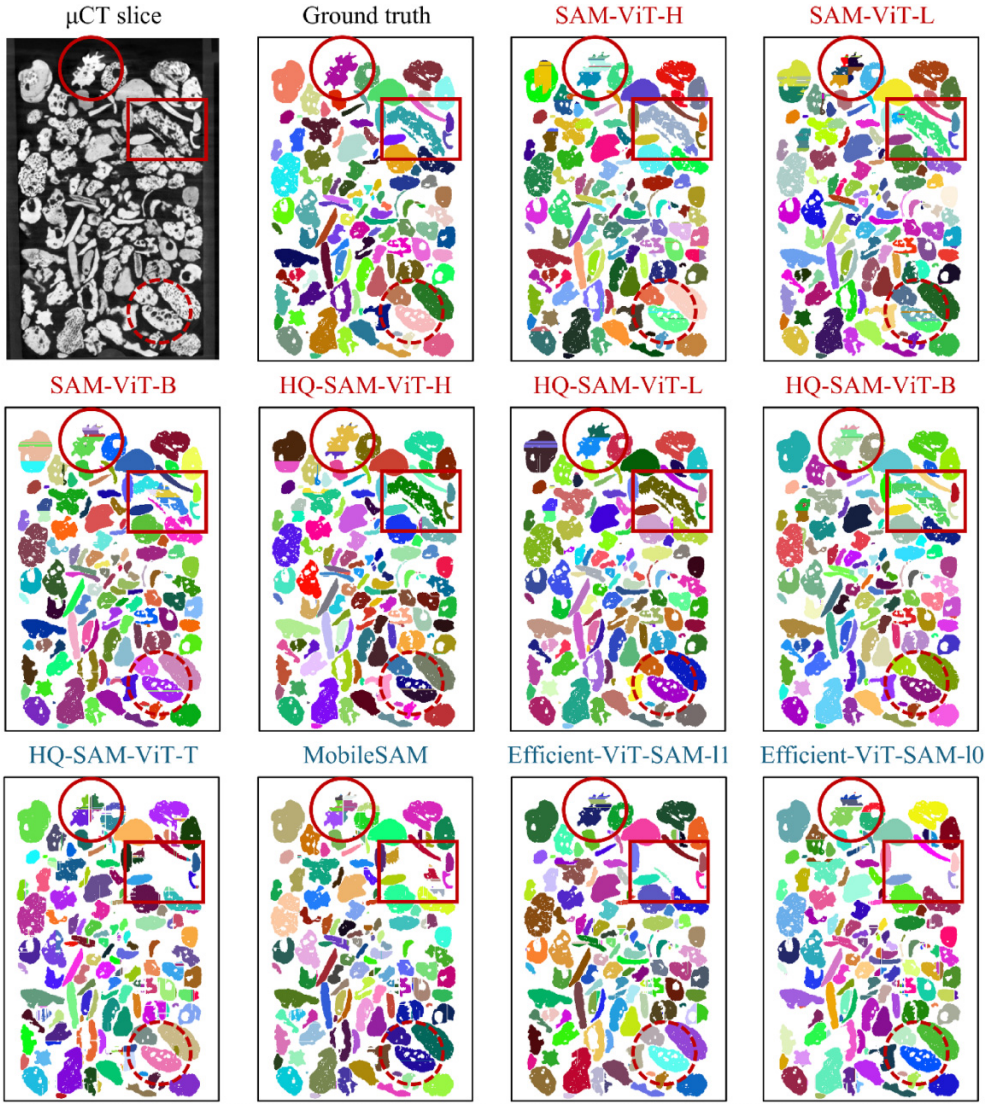
$$C_{p,q} = 1 - \max\left\{ \frac{I(p,q)}{U(p,q)}, \frac{I(p,q)}{N(p)}, \frac{I(p,q)}{N(q)} \right\} \tag{8}$$

where max(·) denotes the maximum value, $I(p, q)$ is the number of pixels in the intersection of label p in $\boldsymbol{L}_i$ and label q in $\boldsymbol{L}_{i+1}$ and $U(p, q)$ is the number of pixels in the union of label p in $\boldsymbol{L}_i$ and label q in $\boldsymbol{L}_{i+1}$. The ratio between $I(p, q)$ and $U(p, q)$ is the so-called Intersection over Union (IoU). IoU is defined as the ratio of the intersecting volume between the bounding boxes of the ground truth and candidate label to the total volumes encompassed by both (Rezatofighi et al. 2019). IoU provides a quantitative measure of the similarity between two arbitrary shapes. $N(p)$ and $N(q)$ represent the number of pixels associated with label p in $\boldsymbol{L}_i$ and label q in $\boldsymbol{L}_{i+1}$, respectively. This novel cost function incorporates IoU and the ratio between the intersection area and two label areas. The Sinkhorn algorithm is a smoothed version of the original OT with an entropic regularization term ξ (Cuturi 2013). The incorporation of the entropic regularization term ξ turns the original OT into a regularized OT which can be solved by iteration faster. The updated calculation of OT is

$$\widehat{\boldsymbol{T}}_{i,i+1} = \arg\min_{T \in \Gamma(L_i, L_{i+1})} \sum_{p,q} C_{p,q} T_{p,q} + \xi \sum_{p,q} T_{p,q} \log T_{p,q} \tag{9}$$

Notably, $\widehat{\boldsymbol{T}}_{i,i+1}$ describes how to update labels in $\boldsymbol{L}_{i+1}$, but ignores the labeling in $\boldsymbol{L}_i$. For instance, suppose there are two or more labels in $\boldsymbol{L}_i$ belonging to the same label in $\boldsymbol{L}_{i+1}$. $\widehat{\boldsymbol{T}}_{i,i+1}$ will only choose one label with the lowest cost in $\boldsymbol{L}_i$ for matching, which is incorrect. To solve this problem, an inverse OT solution $\widehat{\boldsymbol{T}}_{i+1,i} \in \mathbb{R}$ is computed to reassign labels in $\boldsymbol{L}_i$. The final OT solution $\widehat{\boldsymbol{T}}$ for updating labels between $\boldsymbol{L}_i$ and $\boldsymbol{L}_{i+1}$ is

Fig. 11. Comparisons between the raw micro-computed tomographic (μCT) slice, ground truth labels, and representative slices of a reconstructed CS2 sample generated by the 10 different enhanced large vision models ($n = 40$); note the red circular and rectangular insets highlight specific regions used for comparison.



the union of $\widehat{\boldsymbol{T}}_{i,i+1}$ and $\widehat{\boldsymbol{T}}_{i+1,i}$, as shown in Fig. 4:

$$\widehat{\boldsymbol{T}} = \widehat{\boldsymbol{T}}_{i,i+1} \cup \widehat{\boldsymbol{T}}_{i+1,i} \quad (10)$$

For each label q in $\boldsymbol{L}_{i+1}$, the optimal matching label $\widehat{p}$ in $\boldsymbol{L}_i$ is the one with the lowest cost in $\widehat{\boldsymbol{T}}$. If $\widehat{p} = 0$, q is considered a new label and is assigned a new label value $j + 1$. Additionally, if the label p is determined to correspond to a label $\widehat{q}$ in $\boldsymbol{L}_{i+1}$, then all masks in the processed 2D label maps (e.g., $\boldsymbol{L}_{i\text{-}1}$, $\boldsymbol{L}_{i\text{-}2}$, …, $\boldsymbol{L}_1$) with the same value as p are also relabeled to $\widehat{q}$. The stitching is implemented between all adjacent 2D label maps along the same axis sequentially until all 2D label maps are relabeled and converted to one 3D label map.

Suppose the 3D label map created by stitched slices along the x-, y-, and z-axes are denoted $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$, respectively. These 3D label maps cannot be used directly due to the presence of connected particles along the stitch directions. Therefore, voxel mapping is required to combine the three 3D label maps into a single complete map. For each voxel at position (x, y, z), let its label values in $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ be a, b, and c, respectively. Voxels with identical label values across $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$ are considered to belong to the same part and are assigned a unique label in the mapped 3D label map $\boldsymbol{S}$. After all voxels are mapped into $\boldsymbol{S}$, the Hadamard product is applied to $\boldsymbol{S}$, along with the binarized versions of $\boldsymbol{X}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$, as illustrated below:

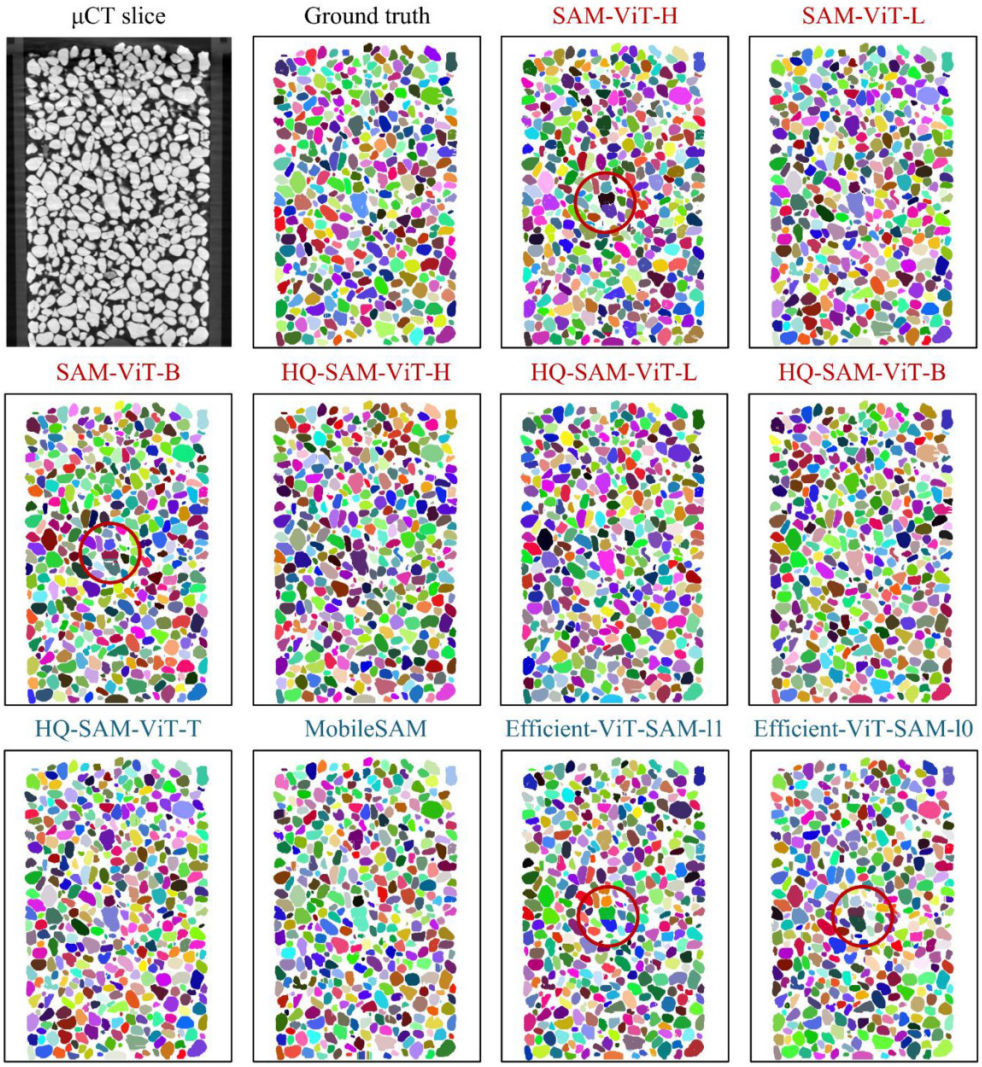
$$\boldsymbol{M} = \boldsymbol{S} \circ B(\boldsymbol{X}) \circ B(\boldsymbol{Y}) \circ B(\boldsymbol{Z}) \quad (11)$$

where $\boldsymbol{M}$ is the final 3D label map, and $B(\cdot)$ denotes the binarization operation.

2.4. 3D particle reconstruction

The MC algorithm is employed to reconstruct individual 3D particles from the combined 3D label map $\boldsymbol{M}$ (Lorensen and Cline 1987). For each label in $\boldsymbol{M}$, a binary matrix is created where voxels corresponding to the label are set to 1 and all others are set to 0. The MC algorithm evaluates each cube of

Fig. 12. Comparisons between the raw micro-computed tomographic (μCT) slice, ground truth labels, and representative slices of a reconstructed FJS sample generated by the 10 different enhanced large vision models ($n = 40$); note the red circular inset highlights a specific region used for comparison.



eight neighboring voxels and determines the polygonal representation of the isosurface through interpolation. This process generates a set of vertices and triangular faces that approximate the surface of the particle. Finally, these triangular faces are smoothed using Gaussian filtering with a kernel size of 3×3.

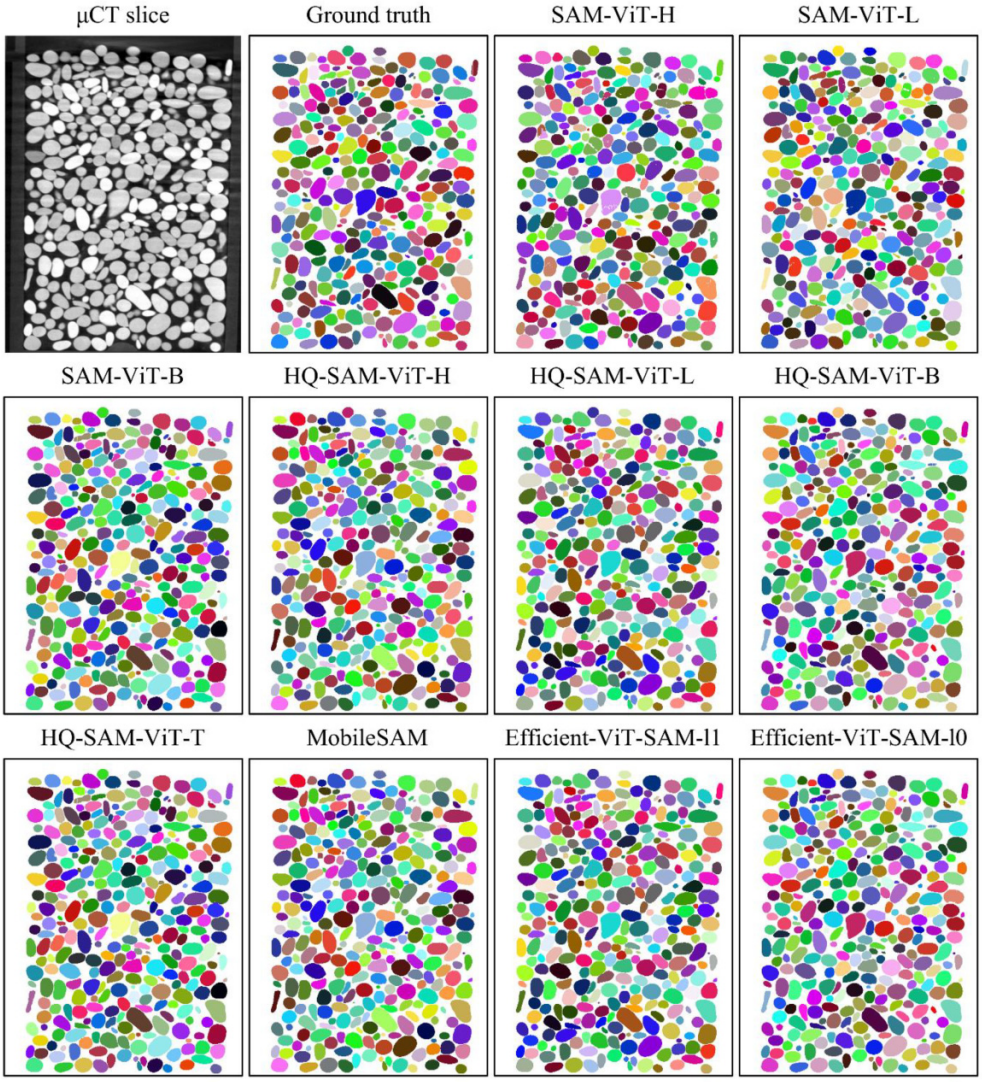
3. Experimental data

3.1. Test apparatus and datasets

A Nikon XT H225ST 2X Computed Tomography system was used for the X-ray scanning in this research. The main components of this system include an X-ray beam tube (or source), a 225 kV high-voltage generator, a sample manipulator, and a flat panel detector. During the scanning process, the sample was placed on the manipulator, which rotated 360° while X-rays passed through the sample and projected a shadow image onto the flat panel detector. The X-rays are attenuated by the sample material, with higher density material causing greater attenuation. These differences are reflected in the image intensity, with high-density materials appearing much darker than low-density regions. Each X-ray pass generates a shadow image, representing a single-angle projection of the sample. The μCT slices along the *x*-, *y*-, and *z*-axes can be obtained using an iterative reconstruction algorithm based on all projections (Beister et al. 2012). To ensure high-quality μCT images, each scan in this study involved capturing 3000 projections as the sample underwent a full 360° rotation.

Four samples (CS1, CS2, FJS, and CGB) were scanned to validate the proposed 3D particle reconstruction framework and to evaluate the performance of different LVMs, as summarized in Table 2. The spatial resolution of these 3D images is 34.96 μm and representative μCT slices of CS1, CS2, FJS, and CGB are presented in Figs. 5*a*–5*d*, respectively. It can be observed that CS1 and CS2 exhibit significant irregularity and porous characteristics while FJS and CGB are more regular in

Fig. 13. Comparisons on the CGB dataset between the raw micro-computed tomographic (μCT) slice, ground truth labels, and representative slices of a reconstructed CGB sample generated by the 10 different enhanced large vision models ($n = 40$).



shape. These complex morphological characteristics are beneficial for the rigorous evaluation of the proposed method. The corresponding ground truth of CS1, CS2, FJS, and CGB, represented in the form of 3D label maps, are illustrated in Figs. 6*a*–6*d*, respectively. These ground truths are determined through manual labeling and are made available as open-source data.

3.2. Benchmark and comparative tests

The watershed segmentation technique is widely used for segmenting granular materials from 3D CT images (Shi and Yan 2015). Although the watershed algorithm has been shown to deliver impressive performance, its segmentation can often be unsatisfactory due to issues with over- and under-segmentation. The H-extreme watershed segmentation technique, which is integrated into the commercial software Avizo, is a common remedy and serves as the benchmark for comparison in this study (Kornilov et al. 2022).

To evaluate the accuracy of the proposed reconstruction method, a series of comparative tests are conducted on all four datasets. For the H-extreme watershed method, strategies proposed by Kong and Fonseca (2018) to mitigate over-segmentation are applied, including a hole-filling step and a hole-repairing step before and after reconstruction. The resulting best H-extreme watershed is used for comparison. All computations are performed on a PC equipped with a 13th Gen Inter(R) Core i5-13600KF CPU, 64GB RAM, and an NVIDIA GeForce RTX 4090 GPU.

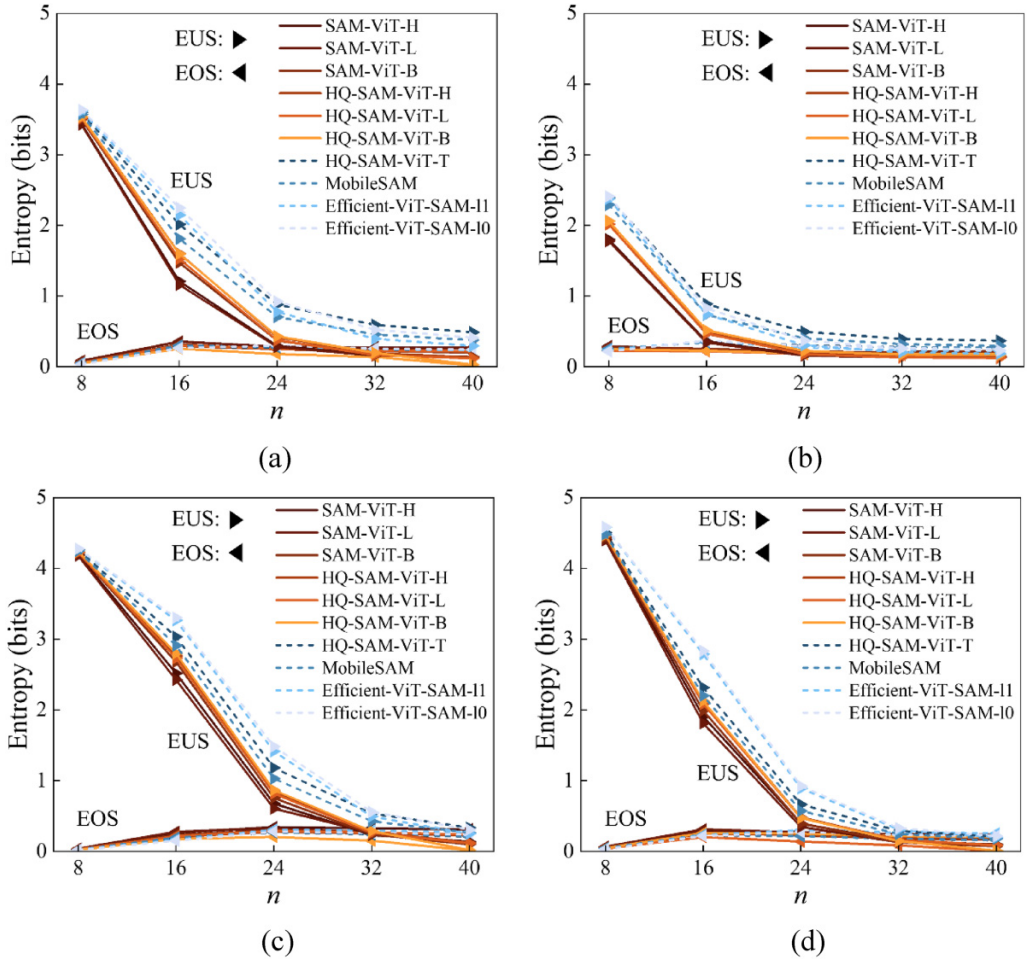
3.3. Evaluation metrics

The standard metric, average precision (AP), is adopted to evaluate the reconstruction accuracy, defined as follows (Davis and Goadrich 2006):

$$\text{AP} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} \tag{12}$$

where TP is the number of true positive labels, FP is the number of false positive labels, and FN is the number of false neg-

Fig. 14. Influence of large vision model type and prompt size on entropy of under-segmentation (EUS) and entropy of over-segmentation (EOS) for: (*a*) CS1; (*b*) CS2; (*c*) FJS; and (*d*) CGB.



ative labels. In this study, a ground truth label and a corresponding candidate prediction label are considered a match if the IoU between the two is greater than a specified threshold which is conventionally taken as 0.5 and denoted $AP_{0.5}$.

The variation of information (VI) is used to quantify the amount of over- and under-segmentations and is arguably the only appropriate criterion from an axiomatic point of view, as demonstrated by Meilă (2007). Suppose there are two segmentations: the ground truth $\boldsymbol{G} = \{G_1, G_2, \ldots, G_i, \ldots, G_m\}$ which contains m labels, and the candidate segmentation $\boldsymbol{S} = \{S_1, S_2, \ldots, S_i, \ldots, S_k\}$ composed of k labels. The VI of the segmentations $\boldsymbol{G}$ and $\boldsymbol{S}$, or the distance between the candidate segmentation and ground truth, is defined as the sum of the conditional entropies:

$$\text{VI}(\boldsymbol{G}, \boldsymbol{S}) = \text{EUS} + \text{EOS} \tag{13}$$

where EUS denotes the entropy of under-segmentation and EOS represents the entropy of over-segmentation and are defined as follows:

$$\text{EUS} = \sum_{i,j} \frac{|G_i \cap S_j|}{N} \log\left(\frac{|S_j|}{|G_i \cap S_j|}\right) \tag{14}$$

$$\text{EOS} = \sum_{i,j} \frac{|G_i \cap S_j|}{N} \log\left(\frac{|G_i|}{|G_i \cap S_j|}\right) \tag{15}$$

where G_i and S_j are the ith and jth labels in $\boldsymbol{G}$ and $\boldsymbol{S}$, respectively, and N is the total number of image pixels, defined as follows:

$$N = \sum_i |G_i| = \sum_j |S_j| \tag{16}$$

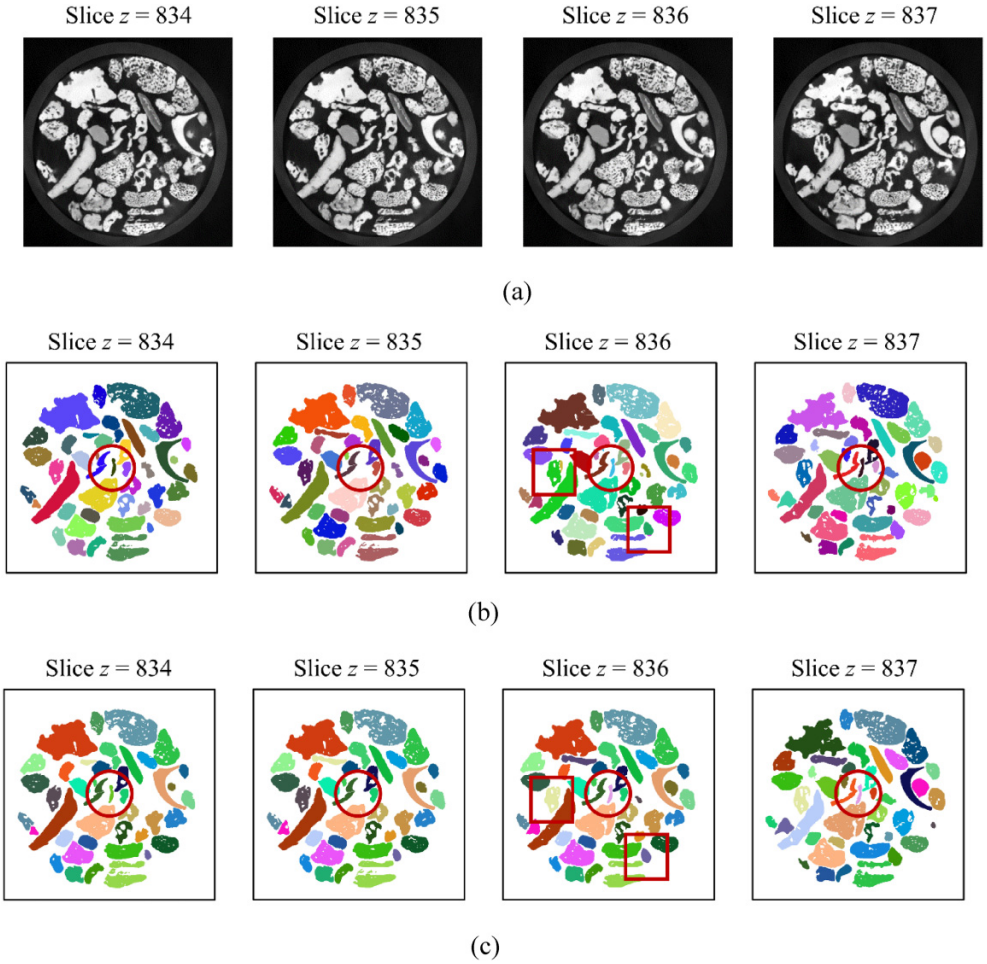
EUS and EOS are measured in bits and are always non-negative. A value of 0 for EUS and EOS indicates that there is no under-segmentation and no over-segmentation respectively, and thus $\text{VI}(\boldsymbol{G}, \boldsymbol{S}) = 0$ indicates perfect segmentation.

4. Results and comparison

4.1. Sensitivity analysis of entropic regularization term

The entropic regularization term ξ significantly influences reconstruction accuracy in OT. Empirical studies have suggested a value of ξ approximately 0.01 to guarantee solution uniqueness and robustness in their applications (Karlsson

Fig. 15. Development of over-segmentation in four exemplar consecutive soil slices of CS2: (*a*) original micro-computed tomographic slices; (*b*) 2D label map; (*c*) 3D label map.



and Ringh 2017). To identify an appropriate value for this parameter in the present study, a sensitivity analysis was performed across four datasets using the SAM model with ($n = 40$), and the results are illustrated in Fig. 7. It is clear that even a small ξ significantly improves the AP, underscoring the effectiveness of smoothness in solving OT. Notably, AP appears to be insensitive to the specific value of the ξ once it exceeds zero. Among the four datasets, the case with $\xi = 0.15$ achieves optimal performance and is adopted in this study.

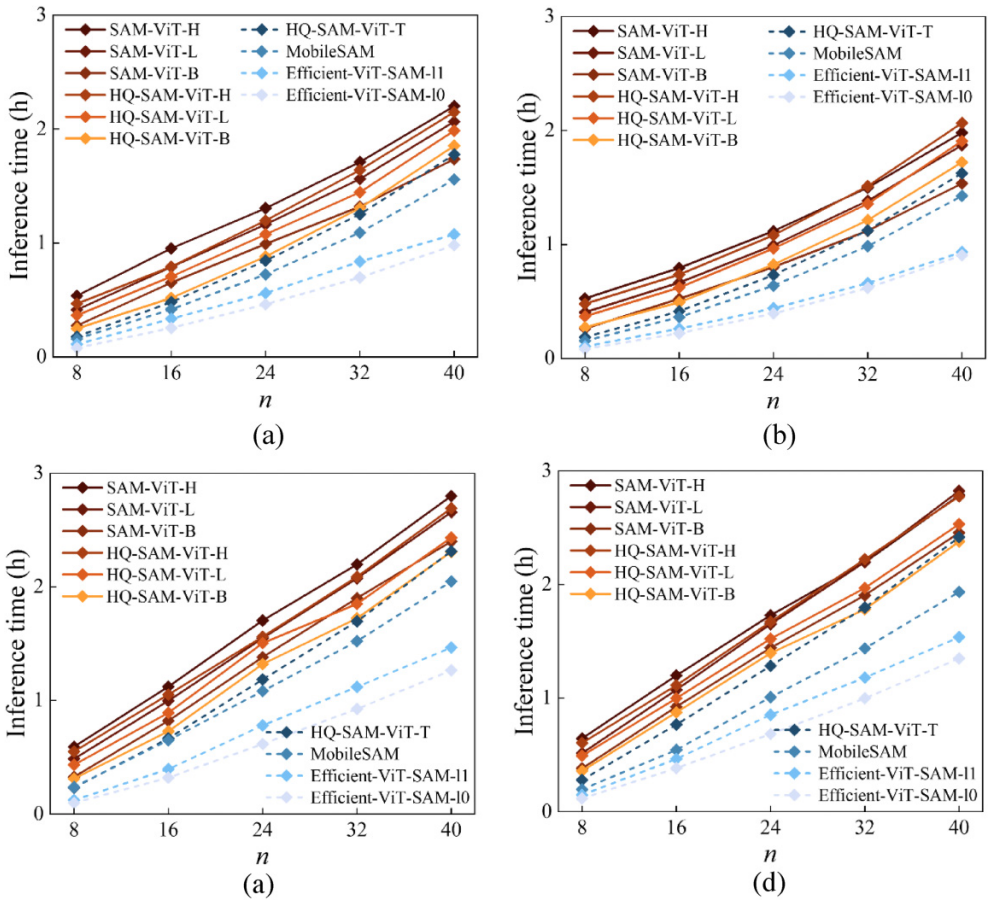
4.2. Reconstruction accuracy analysis

Figure 8 shows the influence of LVM type and the prompt size parameter on $AP_{0.5}$ for the four particle datasets (CS1, CS2, FJS, and CGB). The conventional and lightweight models with the best performance ($AP_{0.5}$) are highlighted in red and blue, respectively. For all cases, $AP_{0.5}$ increases with n up to $n = 32$, after which the growth becomes either more gradual or plateaus; discrepancies between results are likely due to differences in the number of particles in each of the four samples. Note excessive prompts may cause over-segmentation and reduced accuracy, such is the case for CS2 with $n = 40$. Except for CS2, where the HQ-SAM-ViT-L model provides the best performance, model HQ-SAM-ViT-B with $n = 40$ provides the best performance across all particle types. Specifically, for FJS and CGB, it achieves a near-perfect $AP_{0.5}$ of 0.99; for CS1, the same values exceed 0.9. These results demonstrate the effectiveness of the enhanced LVM-based reconstruction method in handling various granular materials with complex morphologies.

For FJS and CGB, the performances of the lightweight LVMs are comparable—and in some cases superior—to that of conventional LVMs indicating that LVM model effectiveness is not necessarily proportional to model size. While larger models may have a greater capacity for label extraction, they are also more prone to over-segmentation. In contrast, lightweight LVMs are sufficient for segmenting simple-shaped granular particles. However, for CS1 and CS2, conventional LVMs significantly outperform lightweight LVMs, with performance improvements ranging from 20% to 30%. This observation is logical given that matching irregular shapes is inherently more challenging—lightweight LVMs are less effective at reconstructing particles with complex morphologies, such as carbonate sands.

Figure 9 shows the influence of the prompt size on the reconstruction of FJS using the HQ-SAM-ViT-H model, in the

Fig. 16. Influence of n on the inference time for all 10 enhanced LVM models: (*a*) CS1; (*b*) CS2; (*c*) FJS; and (*d*) CGB. LVM, large vision model.



form of both 3D models and orthogonal slices. When $n = 8$, only a small number of particles are reconstructed in the incomplete morphology state, evidenced by the near-blank orthogonal slices. This issue is attributed to insufficient input prompts for positioning all particles accurately. While an increase in n improves performance, a value of $n \geq 32$ is required to accurately capture particle morphology and numbers, and to preserve features such as holes and boundaries well. These results imply that the required number of input prompts does not linearly correlate with model performance and a sufficient number of prompts, such as a 32×32 regular grid, is crucial to achieve satisfactory LVM-based reconstruction.

To further compare the performance of the different enhanced LVMs, the middle slices along the *y*-axis are considered for each sample and compared to the ground truth label map and the raw μCT image; the results are presented in Figs. 10, 11, 12, and 13 for samples CS1, CS2, FJS, and CGB, respectively. For CS1, visual inspection reveals no under-segmentation but minor over-segmentation of particles, particularly when using the lightweight LVMs, as shown in the red insets of Fig. 10. These results show that inaccuracies generated by the various LVMs tend to be similar and thus may have similar underpinning mechanisms, indicating potential to improve model robustness with fine-tuning techniques.

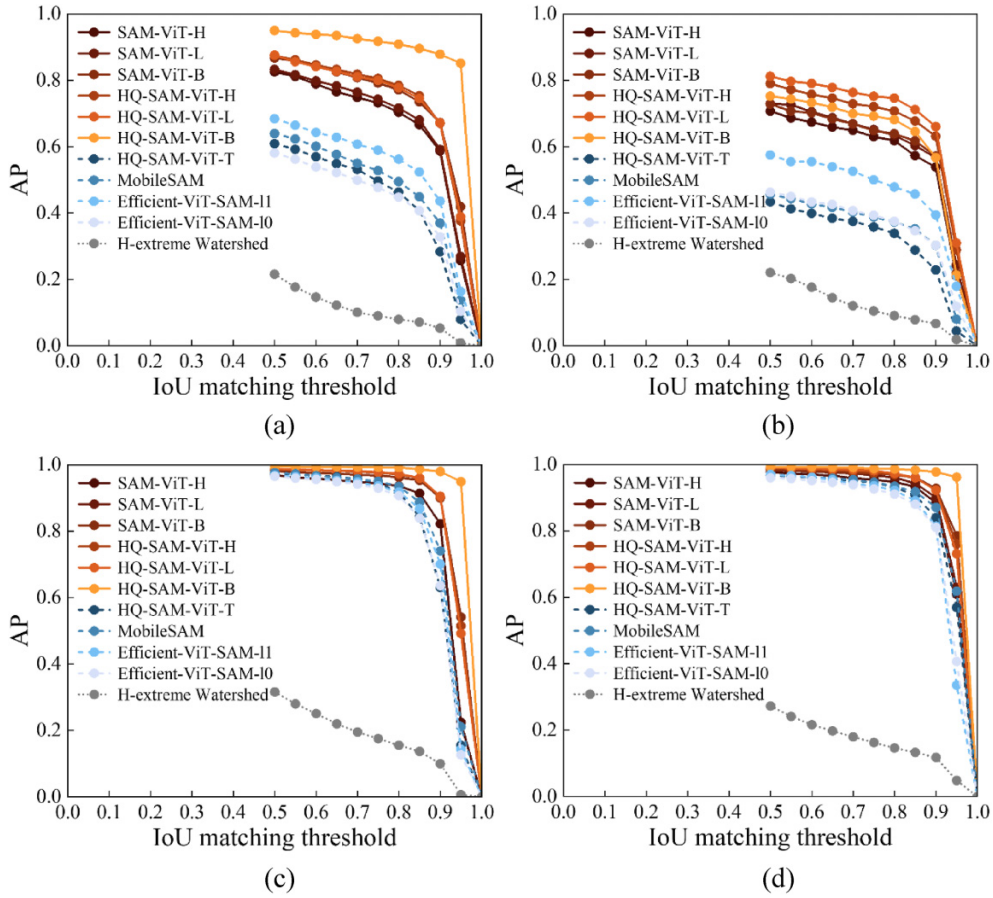
From Fig. 11, over-segmentation is much more pronounced in CS2 compared to CS1 due to the more complex particle shape where Model H1-SAM-ViT-B again performs best. Notably, all LVMs fail to reconstruct the particle in the red insert, showing different levels of over-segmentation. For instance, SAM-ViT-H exhibits over-segmentation along the *z*-axis, whereas over-segmentation along both *y*- and *z*-axes is obtained when using MobileSAM. Furthermore, the lightweight LVMs fail to reconstruct the particles contained within the red rectangular due to their inability to handle complex-shaped particles whereas all enhanced LVMs show good performance for simple particle shapes.

For FJS, almost all particles are reconstructed accurately by both conventional and lightweight LVMs (Fig. 12). Only one over-segmentation is observed, as highlighted by the red insets. For CGB (Fig. 13), all models reach a near-perfect state where all particle boundaries and interior textures are well-preserved with no segmentation errors. These results demonstrate that the most appropriate LVM for segmentation is highly dependent on the complexity of the particle shape.

4.3. Reconstruction error analysis

Figure 14 describes the influence of LVM type and prompt size on EUS and EOS, again for all four particle datasets. For all cases, EUS experiences a sharp drop as n increases to

Fig. 17. Influence of Intersection over Union (IoU) matching threshold on average precision for all 10 enhanced large vision models ($n = 40$) compared to the H-extreme watershed benchmark for: (*a*) CS1; (*b*) CS2; (*c*) FJS; and (*d*) CGB.



$n = 24$, followed by convergence to a small value. For $n < 24$, inaccurate reconstruction is predominantly attributable to under-segmentation due to insufficient prompts. It is noteworthy that the EUS values corresponding to the lightweight LVMs are significantly larger than those obtained using the conventional LVMs. In contrast, EOS values are similar for both conventional and lightweight LVMs and remain at low values. It is also worth remarking that CS2 exhibits notably lower EUS values due to the smaller number of particles. When $n = 40$, all cases have similarly low values of EUS and EOS, demonstrating the robustness of the proposed method; at this point, over-segmentation becomes the main source of error.

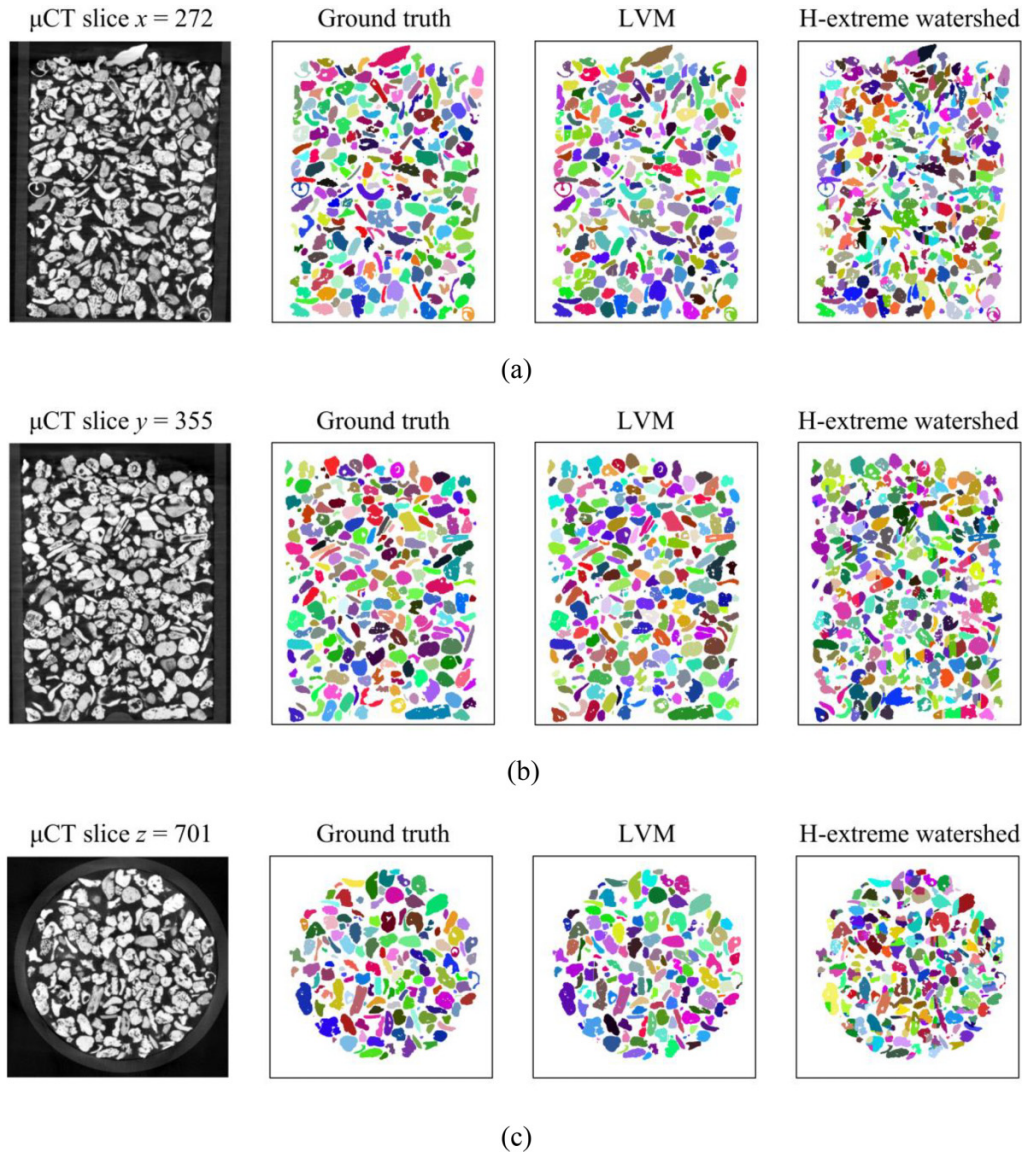
The primary cause of over-segmentation is label loss in certain slices. By way of example, Fig. 15*a* shows μCT slices of CS2 at four consecutive slice positions: $z = 834$, 835, 836, and 837. From Fig. 15*b*, the missing label in the red circle of slice $z = 835$ leads to over-segmentation in the subsequent slice along the z-axis (i.e., at $z = 836$; see Fig. 15*c*). This over-segmentation is also evident in slices along the x and y axes. Similarly, missing labels along the x and y axes cause over-segmentations in the z-axis slices. These over-segmentations accumulate, resulting in fragmented particles, as highlighted in the red solid circles of Fig. 11. A potential solution is to employ more advanced input prompts. For example, using the centre points determined by the object detection algorithm can significantly reduce the number of prompts while guaranteeing reconstruction accuracy. However, training effective object detection algorithms can be challenging and time-consuming, with the risk of overfitting.

Additionally, the LVM-supported reconstruction framework helps reduce over-segmentation errors. For example, the 2D over-segmentation in the red rectangle of Fig. 15*b* is corrected via the proposed 2D-to-3D mapping, as shown in the final 3D label map (Fig. 15*c*).

4.4. Inference time analysis

The influence of n on the inference time for all 10 enhanced LVM models is illustrated in Fig. 16 for all four samples. Nearly all reconstructions are completed within 3 h, which is acceptable for a real-world project. The inference time increases linearly with n across all cases. Conventional LVMs are notably slower than lightweight LVMs, with the fastest conventional LVM still slower than the slowest lightweight LVM. Among the conventional LVMs, the ViT-H branch is the slowest, followed by ViT-L and ViT-B, highlighting the impact of model size. Notably, the fastest model is EfficientViT-SAM-l0, which has 34.8 M parameters, while the smallest LVM is MobileSAM with only 9.8 M parameters. This disparity in performance may be due to differences in model

Fig. 18. Comparisons between exemplar ground truth micro-computed tomographic (μCT) slides and 2D visualizations with ground truth labelling, and label maps predicted by the present LVM model and the H-extreme benchmark for the CS1 dataset: (*a*) slice $x = 272$; (*b*) slice $y = 355$; and (*c*) slice $z = 701$. LVM, large vision model.

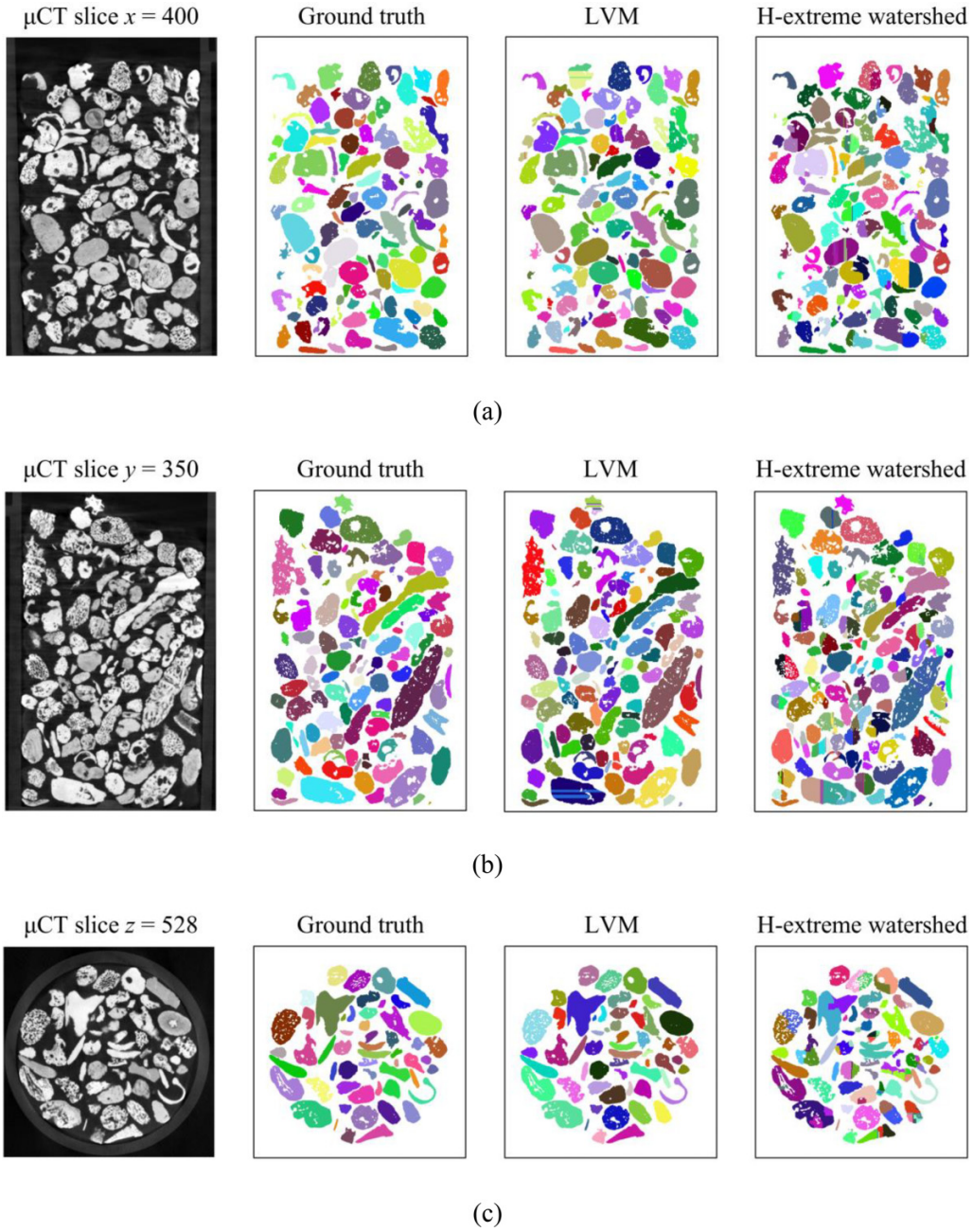


architecture. Additionally, the average inference time for FJS and CGB is longer compared to CS1 and CS2, corresponding to the higher number of particles in these datasets. In practice, excessively large or complex datasets can significantly diminish both the efficiency and accuracy of our method. Therefore, it is advisable to optimize the dataset by including only essential information and excluding any unnecessary details, such as background. More lightweight LVMs can be used to improve computational efficiency. For instance, replacing the heavy image encoder with a lighter one can significantly reduce computational time (Li et al. 2025). Along this way, the challenge lies in balancing accuracy and efficiency which needs further research.

4.5. Comparison with benchmark

The relationship between AP and IoU matching threshold (ranging between 0.5 and 1 in intervals of 0.05) for all 10 LVMs ($n = 40$) is compared to the best-performance H-extreme watershed benchmark in Fig. 17. A higher IoU matching threshold indicates stricter matching requirements. The results show a gradual decrease in AP up to an IoU matching threshold of 0.95, after which the AP values drop sharply, reaching 0 at a threshold of 1. It is evident that the proposed method significantly outperforms the benchmark H-extreme watershed at all IoU matching thresholds and across all samples. Even the worst-performing exceeds the benchmark by at least 50%. Conventional LVMs outperform the lightweight models for CS1 and CS2. Among the lightweight LVMs, Efficient-ViT-SAM-l1 achieves the best performance, providing notably superior performance when reconstructing CS2, although there are no such distinct differences for the other particle types. For FJS and CGB, conventional and lightweight LVMs are comparable yet still substantially outperform the H-extreme watershed benchmark.

Fig. 19. Comparisons between exemplar ground truth micro-computed tomographic (μCT) slides and 2D visualizations with ground truth labelling, and label maps predicted by the present LVM model and the H-extreme benchmark for the CS2 dataset: (*a*) slice $x = 400$; (*b*) slice $y = 350$; and (*c*) slice $z = 528$. LVM, large vision model.

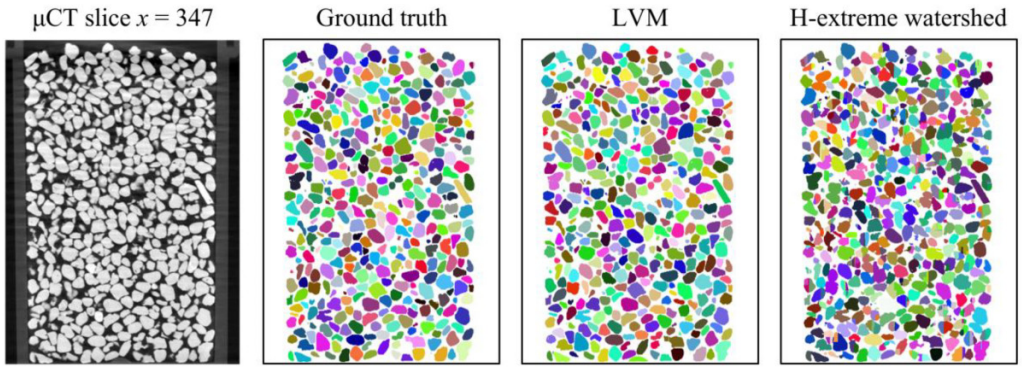


2D visualizations of slices parallel to the *x*, *y*, and *z* axes, processed by the best-performing LVM (model HQSAM-ViT-B with $n = 40$) and the H-extreme watershed algorithm are compared for CS1, CS2, FJS, and CGB in Figs. 18–21, respectively. These figures illustrate the superior performance of the LVM models over the watershed algorithm, with the former exhibiting significantly fewer instances of over- and under-segmentation. In comparison to the benchmark, our approach demonstrates near-perfect accuracy in processing 3D particle reconstructions for simpler morphologies, such as FJS and CGB. Even for complicated cases, such as carbonate sands, the LVM-based reconstructions perform well, with minimal over- or under-segmentation observed.

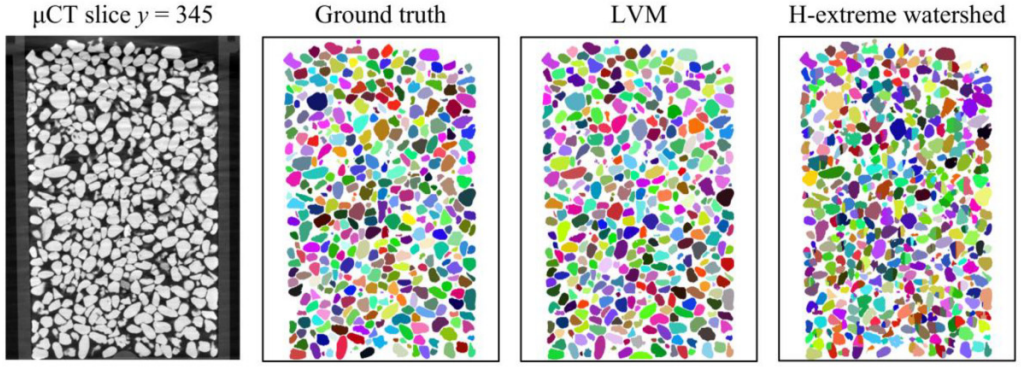
5. Microstructure characteristics

Complex micromechanical phenomena occur inside granular materials such as interlocking, particle breakage, and anisotropy (Harkness 2009; He et al. 2023; Pinzón et al. 2023), all intrinsically linked to microstructural characteristics (Li et al. 2009). Therefore, assessing reconstruction quality should consider microstructural properties, rather than solely relying on accuracy comparisons. This section evaluates the representative reconstructed model (enhanced SAMHQ-ViT-B at $n = 40$) against the ground truth based on three microstructural characteristics: morphological features, coordination number, and interparticle contact orientation.

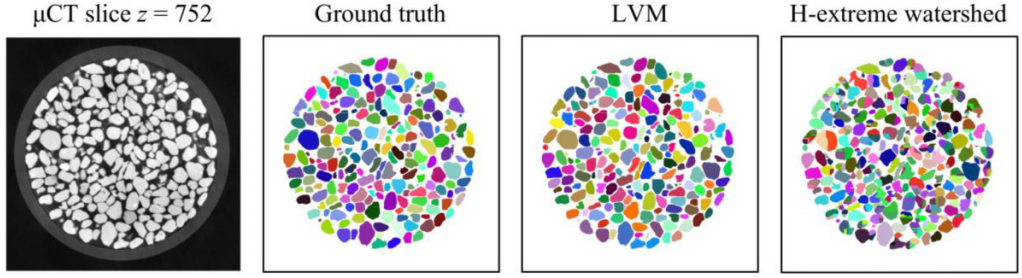
Fig. 20. Comparisons between exemplar ground truth micro-computed tomographic (μCT) slides and 2D visualizations with ground truth labelling, and label maps predicted by the present LVM model and the H-extreme benchmark for the FJS dataset: (*a*) slice $x = 347$; (*b*) slice $y = 345$; and (*c*) slice $z = 752$. LVM, large vision model.



(a)



(b)



(c)

5.1. Morphological features

Particle morphology significantly influences the macroscopic mechanical response of granular media (Zhao et al. 2016). The morphological characteristic of a particle can be quantified by shape indices (Wu and Wang 2023), including elongation index (EI), flatness index (FI), sphericity index (SI), and convexity index (CI), defined as follows:

$$\text{EI} = \frac{b}{a} \tag{17}$$

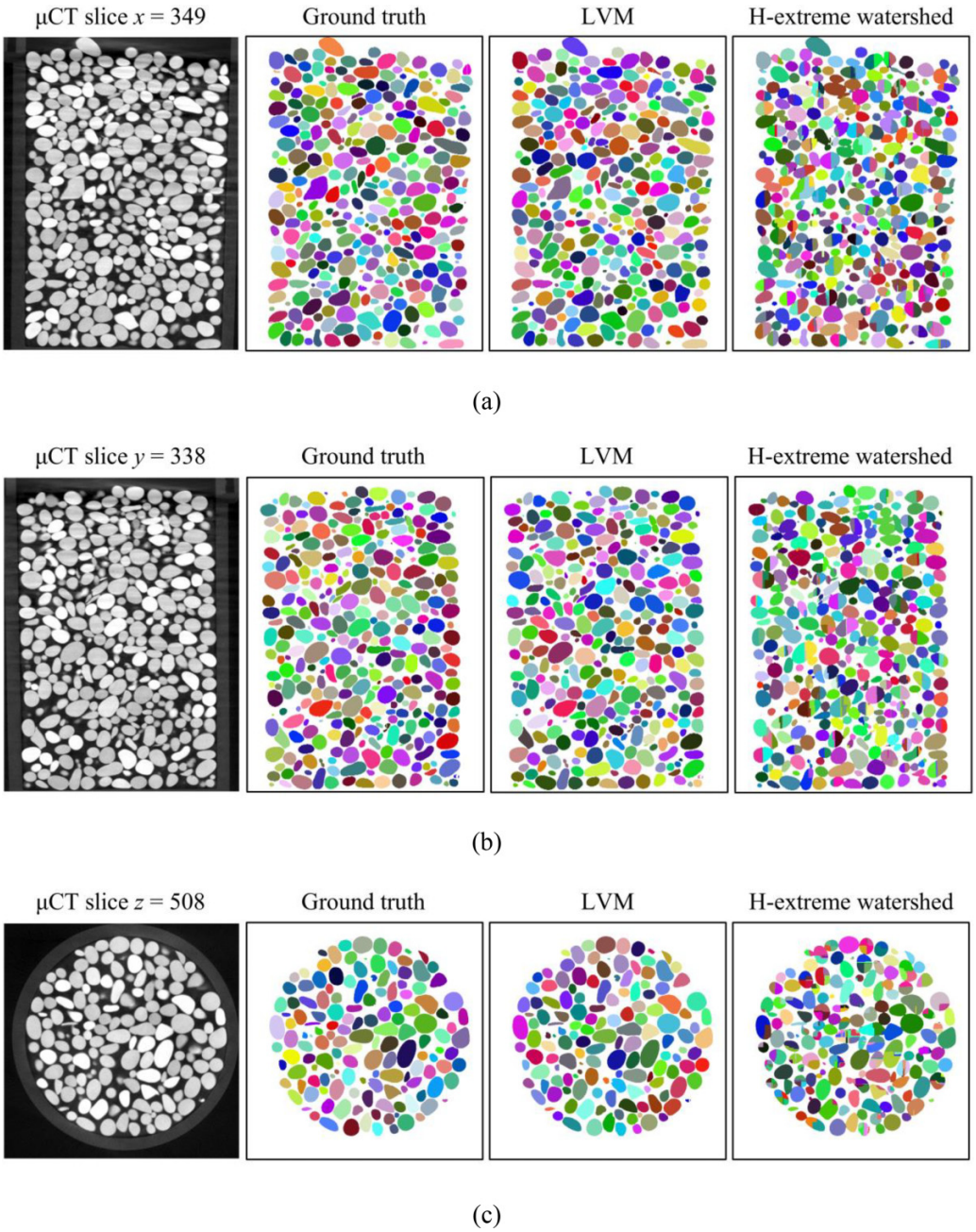
$$\text{FI} = \frac{c}{b} \tag{18}$$

$$\text{SI} = \frac{\sqrt[3]{36\pi V^2}}{S} \tag{19}$$

$$\text{CI} = \frac{V}{V_c} \tag{20}$$

where a, b, and c are the major, intermediate, and minor dimensions of a particle, respectively; V and S are the volume and surface area of a particle, respectively; V_c denotes the volume of the convex hull of the particle. Figure 22 presents the distribution of particle shape indices across four datasets, comparing our reconstruction with the ground truth. The reconstructed particle morphologies closely resemble the ground truth, as demonstrated by nearly perfect overlap, particularly for CS1, FJS, and CGB. These results indicate that the reconstructed particles accurately preserve the original morphological features, allowing them to be used in subsequent

Fig. 21. Comparisons between exemplar ground truth micro-computed tomographic (μCT) slides and 2D visualizations with ground truth labelling, and label maps predicted by the present LVM model and the H-extreme benchmark for the CGB dataset: (*a*) slice $x = 349$; (*b*) slice $y = 338$; and (*c*) slice $z = 508$. LVM, large vision model.



numerical simulations without modification. The high precision of the particle morphology ensures that the macroscopic behavior of the granular media aligns closely with the ground truth.

To quantify the relationship between the complexity of particle shapes and AP, the overall regularity (OR) is adopted as follows (He et al. 2024*a*; Yang and Luo 2015):

$$\text{OR} = (\text{EI} + \text{SI} + \text{CI})/3 \tag{21}$$

The OR provides a collective description of the complexity of the particle shape in granular materials. Figure 23 illustrates the OR values of four datasets and their corresponding AP values, where OR_P and OR_G denote the OR value of the predicted and ground truth values, respectively. It is observed that the AP values exhibit a positive correlation with the OR values, with AP values reaching high levels ($\geq$0.9) when OR values exceed 0.76. Furthermore, the discrepancy between the predicted OR values and the ground truth decreases as the OR values increase. These findings suggest that the model achieves more satisfactory performance on datasets of rounded materials with higher OR values.

Fig. 22. Comparison of the distribution of particle shape indices from the present 3D sample reconstruction compared to the ground truth for all four datasets: (*a*) CS1; (*b*) CS2; (*c*) FJS; and (*d*) CGB; note dashed lines denote results from the present reconstruction and solid lines represent the ground truth.

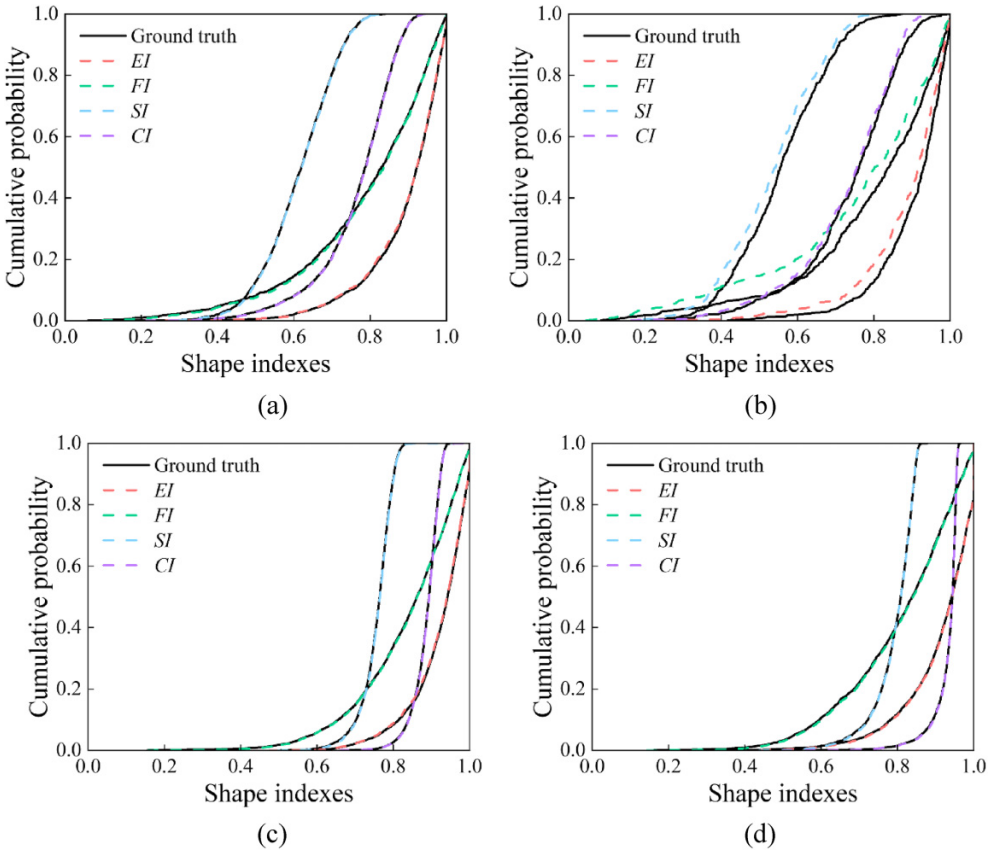
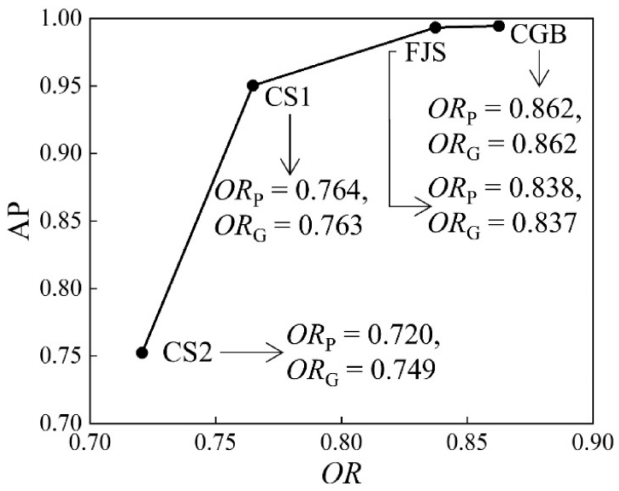


Fig. 23. Relationship between the complexity of particle shapes and average precision (AP). OR, overall regularity.



5.2. Coordination number

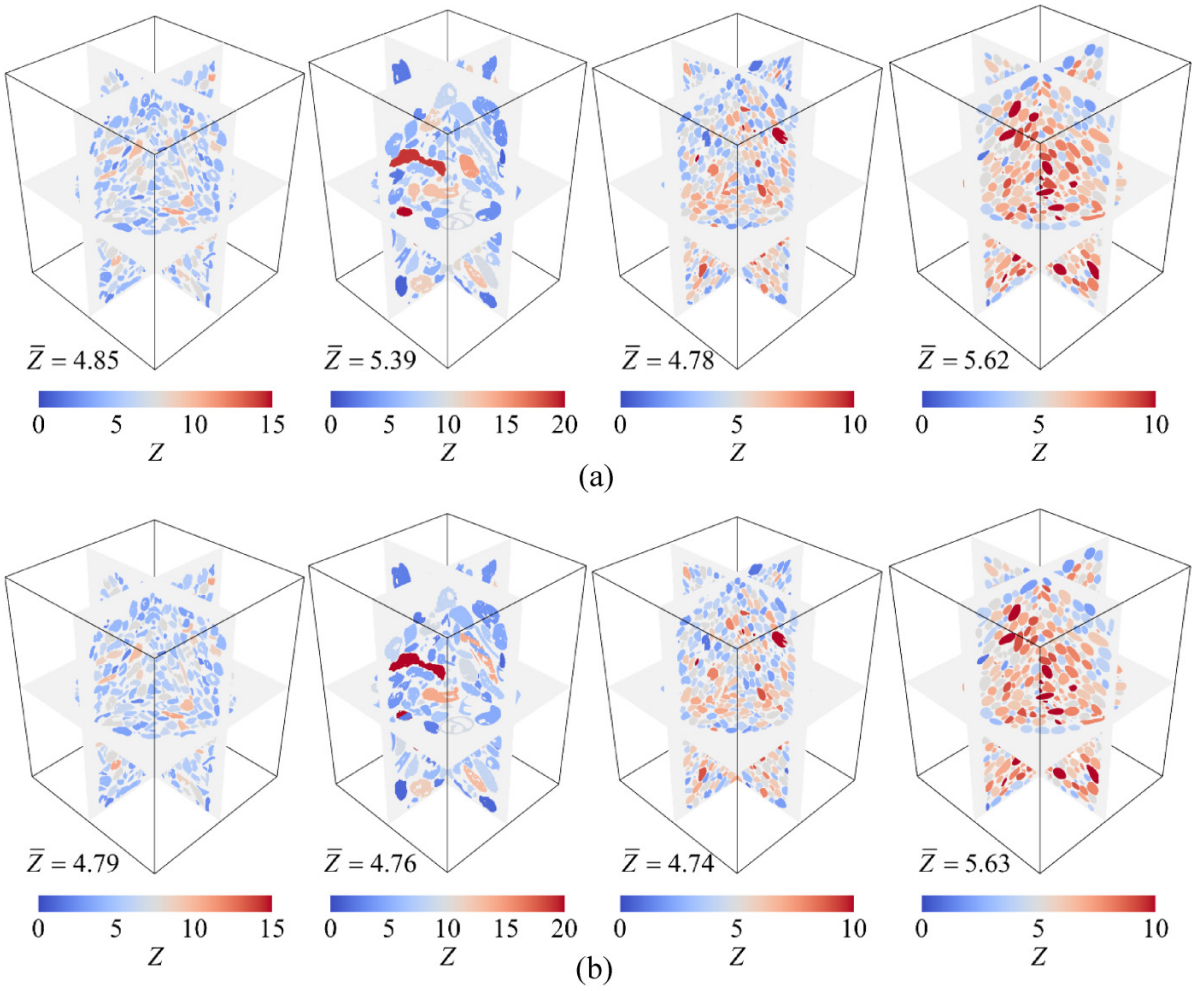
The coordination number *Z*, representing the number of contacts per particle, is a crucial metric for elucidating the geometric arrangement of densely packed particles. Contact detection was facilitated by dilating all 3D labels and subsequently identifying which labels were in contact with the dilated counterparts through pixel connectivity (Kong and Fonseca 2019). The mean coordination number $\bar{Z}$ within an assembly is defined as follows:

$$\bar{Z} = \frac{2N_c}{N_p} \quad (22)$$

where N_c and N_p denote the number of contacts and particles within the assembly, respectively.

Figure 24 shows the distribution of coordination number per particle of four datasets. The correct mean coordination numbers for CS1, CS2, FJS, and CGB are 4.85, 5.39, 4.78, and 5.62, respectively. These results indicate that CGB exhibits a denser packing compared to the other datasets. The high *Z* value observed in CS2 may be attributed to the larger particle size, which typically correlates with a higher *Z* value, or the tendency to be surrounded by more particles, consistent with the findings of Zhang et al. (2023). Notably, the mean coordination numbers for CS1, FJS, and CGB, as derived from reconstruction, align well with the ground truth. These findings suggest that our reconstruction of these three datasets is satisfactory, as evidenced by the indistinguishable orthogonal slices of coordination number distribution between the reconstruction and the ground truth. Conversely, the relatively poorer performance for sample CS2 highlights the critical importance of accurate 3D

Fig. 24. Distribution of coordination number per particle of four datasets: (*a*) ground truth; and (*b*) reconstructed result. Note that in (*a*) and (*b*), from left to right are datasets CS1, CS2, FJS, and CGB.



particle reconstruction for the quality of microstructural quantification.

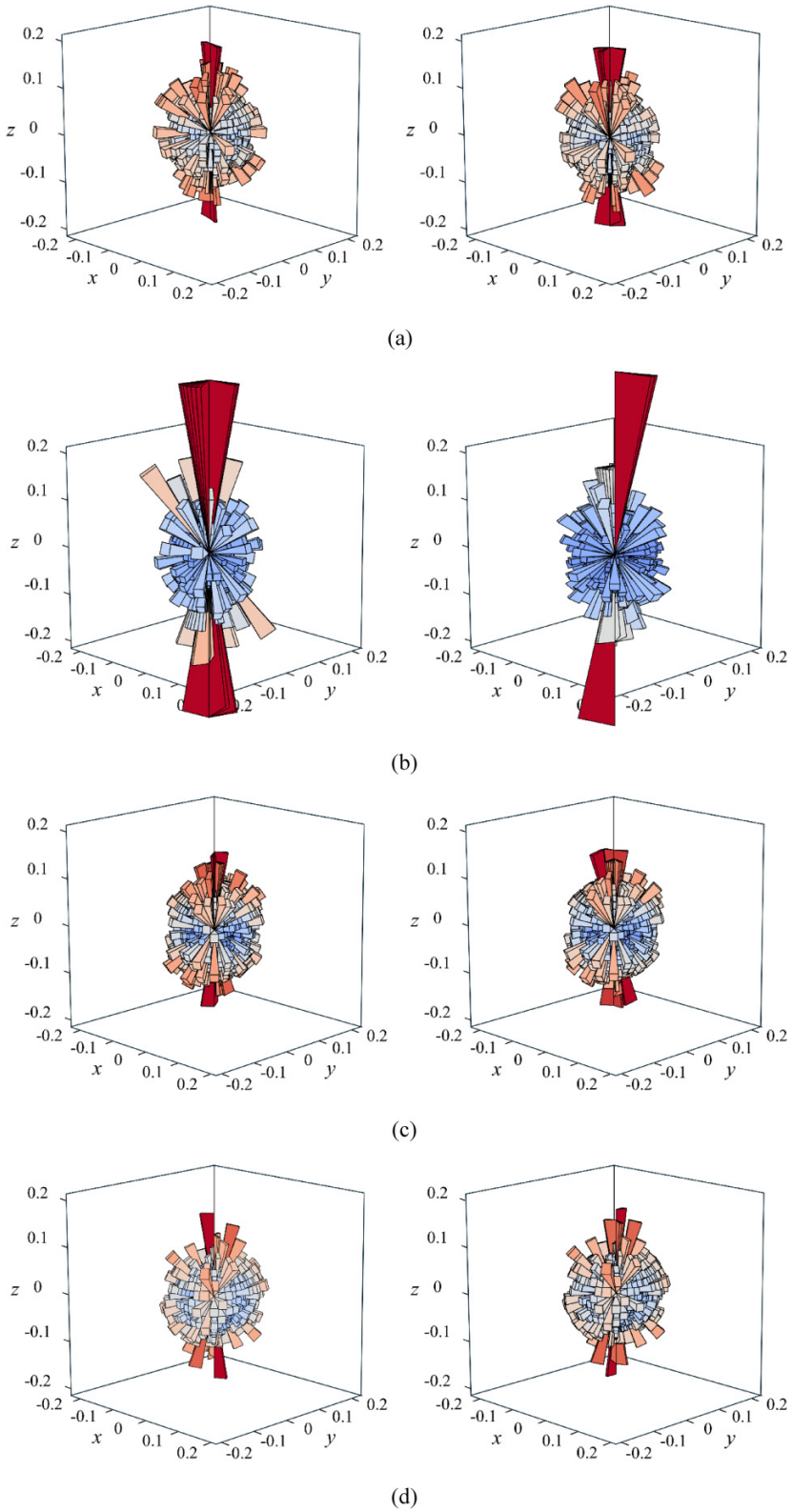
5.3. Interparticle contact orientation

Interparticle contact orientation plays an important role in granular micromechanics and is defined by the normal vector of the contact area (Wiebicke et al. 2017). Figure 25 presents a comparison of the 3D spherical histograms of contact normal orientation vectors across four datasets. In the case of CS1, both the ground truth and the reconstructed results reveal a distinctly anisotropic distribution, characterized by a prominent peak along the *z*-axis, which indicates a robust force chain oriented in this direction. Despite the reconstructed distribution successfully capturing the majority of structural features, significant deviations are observed, particularly concerning non-vertical contacts. This indicates a potential underestimation of the actual values by the reconstructed model, underscoring the need for further refinement to enhance model accuracy. The concentration of orientations along the *z*-axis is particularly pronounced in dataset CS2, while it is mitigated in FJS and CGB. Overall, all reconstructed cases exhibit a similar distribution to ground truth, capturing microstructural characteristics accurately.

6. Discussions

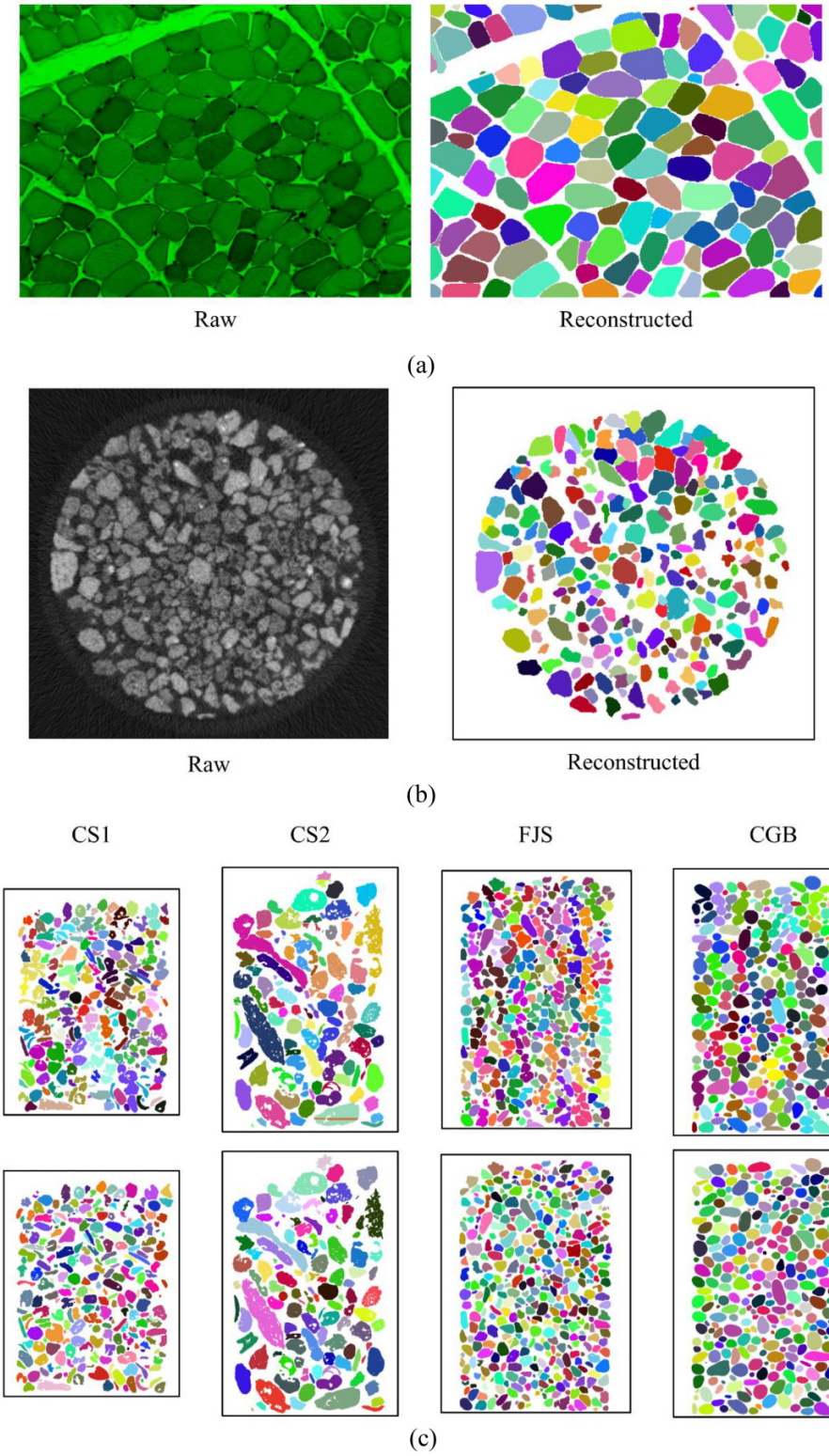
In this section, the potential of our proposed method in other fields was validated and discussed, leveraging the HQ-SAM-ViT-B model with $n = 32$. Figure 26*a* presents a cell image from the Cellpose dataset (Pachitariu and Stringer 2022) and corresponding 2D segmentation by our method. The cells are accurately separated with distinct boundaries, suggesting the potential applicability of our approach in biomedical imaging. Furthermore, Fig. 26*b* shows the raw image and reconstruction result on a highly noisy image provided by Li et al. (2024). Notably, our method exhibits satisfactory performance in reconstructing particles even from noisy images, underscoring its robustness and potential for application in datasets where image quality is compromised. In addition, Fig. 26*c* illustrates the reconstructed results obtained using segmentations along a single axis (e.g., *z*-axis) and corresponding ground truths. The $AP_{0.5}$ for the reconstructed results across four datasets (CS1, CS2, FJS, and CGB) is re-

Fig. 25. 3D spherical histograms of contact normal orientation vectors for all four datasets: (*a*) CS1; (*b*) CS2; (*c*) FJS; and (*d*) CGB; note that the left column represents the ground truth, while the right column refers to the reconstructed result.



ported as 30.36%, 57.19%, 30.68%, and 30.87%, respectively. A notable observation is the presence of numerous connected particles along the *z*-axis, which can be attributed to the partial volume effect (Rousset et al. 1998). This phenomenon highlights the limitations of relying solely on segmentations along one single axis and underscores the importance of utilizing segmentations along three axes to ensure accurate reconstruction.

Fig. 26. Potential applications: (a) cell image; (b) highly noisy image; (c) reconstructions only using 2D segmentation along the *z*-axis.



7. Conclusions

This paper has presented a novel framework enhancing LVMs and robust algorithms to reconstruct 3D particles from μCT images. The framework converted μCT images into low-quality 2D label maps, refined them through a two-step process, and stitched them into a unified 3D label map using the OT method. The final 3D map was reconstructed with the MC algorithm. Validated on four sample datasets, the framework operated fully automatically without additional training.

The study evaluated ten different foundation models to assess the impact of LVM type, model size, and prompt size on 2D label extraction. The result indicated that LVMs with varying model sizes but the same architecture provided comparable performance. A minimum grid of 32 × 32 point prompts was necessary to achieve satisfactory results. The enhanced framework demonstrated over 50% improvement compared to the benchmark H-extreme watershed method and proved robust in minimizing over- and under-segmentation errors. The reconstructed microstructure characteristics accurately revealed the micromechanical properties of densely packed particles.

The proposed framework is generic and applicable to the reconstruction of arbitrary 3D particles from μCT images. It offers a straightforward and automated method for users to reconstruct 3D particles. Future research could explore its application to a wider range of 3D particles or materials, utilizing lightweight foundation models and more efficient input prompts.

Acknowledgements

This research was financially supported by the Research Grants Council (RGC) of Hong Kong Special Administrative Region Government (HKSARG) of China (Grant Nos.: 15229223, 15227923, 15226822, 15220221).

Article information

History dates

Received: 28 November 2024
Accepted: 8 March 2025
Accepted manuscript online: 24 March 2025
Version of record online: 23 May 2025

Copyright

Data availability

All algorithms and data are available upon reasonable request.

Author information

Author ORCIDs

Zhen-Yu Yin https://orcid.org/0000-0003-4154-7304
Shaoheng He https://orcid.org/0000-0001-7883-971X

Author notes

Zhen-Yu Yin served as Editorial Board Member at the time of manuscript review and acceptance; peer review and editorial decisions regarding this manuscript were handled by another editorial board member.

Author contributions

Conceptualization: RL, ZY
Data curation: RL, BS
Funding acquisition: ZY
Methodology: RL
Software: RL
Writing – original draft: RL
Writing – review & editing: ZY, SH, BS

References

Beister, M., Kolditz, D., and Kalender, W.A. 2012. Iterative reconstruction methods in X-ray CT. Physica Medica, **28**: 94–108. doi:10.1016/j.ejmp.2012.01.003. PMID: 22316498.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., et al. 2020. Language models are few-shot learners. Advances in Neural Information Processing Systems, **33**: 1877–1901.

Burgmann, S., Godehardt, M., Schladitz, K., and Breit, W. 2022. Separation of sand and gravel particles in 3D images using the adaptive h-extrema transform. Powder Technology, **404**: 117468. doi:10.1016/j.powtec.2022.117468.

Cai, Y., Yu, H.-S., Wanatowski, D., and Li, X. 2013. Noncoaxial behavior of sand under various stress paths. Journal of Geotechnical and Geoenvironmental Engineering, **139**: 1381–1395. doi:10.1061/(ASCE)GT.1943-5606.0000854.

Cevallos, S.B., Jerves, A.X., Mital, U., Medina, D.A., Quinteros, V.S., Mulas, M., and Torgersrud, Ø. 2023. Towards a more accurate characterization of granular media 2.0: involving AI in the process. Computers and Geotechnics, **160**: 105510. doi:10.1016/j.compgeo.2023.105510.

Chang, J., Huang, H., Thewes, M., Zhang, D., and Wu, H. 2024. Data-based postural prediction of shield tunneling via machine learning with physical information. Computers and Geotechnics, **174**: 106584. doi:10.1016/j.compgeo.2024.106584.

Corwin, E.I., Jaeger, H.M., and Nagel, S.R. 2005. Structural signature of jamming in granular media. Nature, **435**: 1075–1078. doi:10.1038/nature03698. PMID: 15973404.

Cuturi, M. 2013. Sinkhorn distances: lightspeed computation of optimal transport. Advances in Neural Information Processing Systems, **26**.

Darve, F., Sibille, L., Daouadji, A., and Nicot, F. 2007. Bifurcations in granular media: macro-and micro-mechanics approaches. Comptes Rendus Mécanique, **335**: 496–515.

Davis, J., and Goadrich, M. 2006. The relationship between Precision-recall and ROC curves. *In* Proceedings of the 23rd International Conference on Machine Learning—ICML '06. Presented at the the 23rd international conference. ACM Press, Pittsburgh, Pennsylvania. pp. 233–240. doi:10.1145/1143844.1143874.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. 2019. BERT: pre-training of deep bidirectional transformers for language understanding.

Guo, N., and Zhao, J. 2013. The signature of shear-induced anisotropy in granular media. Computers and Geotechnics, **47**: 1–15. doi:10.1016/j.compgeo.2012.07.002.

Harkness, J. 2009. Potential particles for the modelling of interlocking media in three dimensions. International Journal for Numerical Methods in Engineering, **80**: 1573–1594. doi:10.1002/nme.2669.

He, S.-H., Goudarzy, M., Ding, Z., and Sun, Y. 2023. Strength, deformation, and particle breakage behavior of calcareous sand: role of anisotropic consolidation. Journal of Geotechnical and Geoenvironmental Engineering, **149**: 04023002. doi:10.1061/JGGEFK.GTENG-10501.

He, S.-H., Yin, Z.-Y., and Ding, Z. 2024a. Cyclic liquefaction of granular materials with varied forms under multidirectional loads. Géotechnique, **75**: 243–257. doi:10.1680/jgeot.22.00386.

He, S.-H., Yin, Z.-Y., Ding, Z., and Li, R.-D. 2024b. Particle morphology and principal stress direction dependent strength anisotropy through torsional shear testing. Canadian Geotechnical Journal. doi:10.1139/cgj-2023-0717.

He, S.-H., Yin, Z.-Y., Ibraim, E., and Ding, Z. 2025. Face mask chips-reinforced sands under monotonic and cyclic torsional shearing. Géotechnique, 1–16. doi:10.1680/jgeot.24.01180.

Huang, H., Chang, J., Zhang, D., Zhang, J., Wu, H., and Li, G. 2022. Machine learning-based automatic control of tunneling posture of shield machine. Journal of Rock Mechanics and Geotechnical Engineering, **14**: 1153–1164. doi:10.1016/j.jrmge.2022.06.001.

Jaeger, H.M., Nagel, S.R., and Behringer, R.P. 1996. The physics of granular materials. Physics Today, **49**: 32–38.

Karlsson, J., and Ringh, A. 2017. Generalized Sinkhorn Iterations for Regularizing Inverse Problems Using Optimal Mass Transport. SIAM Journal on Imaging Sciences, **10**: 1935–1962. doi:10.1137/17M111208X.

Ke, L., Ye, M., Danelljan, M., Liu, Y., Tai, Y.-W., Tang, C.-K., and Yu, F. 2023. Segment anything in high quality.Advances in Neural Information Processing Systems.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., et al. 2023. Segment anything. Proceedings of the IEEE/CVF international conference on computer vision.

Kohler, R. 1981. A segmentation system based on thresholding. Computer Graphics and Image Processing, **15**: 319–338. doi:10.1016/S0146-664X(81)80015-9.

Kong, D., and Fonseca, J. 2018. Quantification of the morphology of shelly carbonate sands using 3D images. Géotechnique, **68**: 249–261. doi:10.1680/jgeot.16.P.278.

Kong, D., and Fonseca, J. 2019. On the kinematics of shelly carbonate sand using X-ray micro tomography. Engineering Geology, **261**: 105268. doi:10.1016/j.enggeo.2019.105268.

Kornilov, A., Safonov, I., and Yakimchuk, I. 2022. A review of watershed implementations for segmentation of volumetric images. Journal of Imaging, **8**: 127. doi:10.3390/jimaging8050127. PMID: 35621890.

Lai, Z., and Chen, Q. 2019. Reconstructing granular particles from X-ray computed tomography using the TWS machine learning tool and the level set method. Acta Geotechnica, **14**: 1–18. doi:10.1007/s11440-018-0759-x.

Li, C., Qi, L., and Geng, X. 2025. A SAM-guided two-stream lightweight model for anomaly detection. ACM Transactions on Multimedia Computing, Communications, and Applications, **21**: 1–23. doi:10.1145/3706574.

Li, R., Zhang, P., Yin, Z., and Sheil, B. 2024. Enhanced hybrid algorithms for segmentation and reconstruction of granular grains from X-ray micro computed-tomography images. International Journal for Numerical and Analytical Methods in Geomechanics, **48**, 4206. doi:10.1002/nag.3832.

Li, X., Yu, H.S., and Li, X.S. 2009. Macro–micro relations in granular mechanics. International Journal of Solids and Structures, **46**: 4331–4341. doi:10.1016/j.ijsolstr.2009.08.018.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. 2014. Microsoft COCO: Common objects in context. *In* Computer Vision—ECCV 2014, Lecture Notes in Computer Science. *Edited by* D. Fleet, T. Pajdla, B. Schiele and T. Tuytelaars. Springer International Publishing, Cham. pp. 740–755. doi:10.1007/978-3-319-10602-1_48.

Lin, W., Li, P., Xie, X., Cao, Y., and Zhang, Y. 2023. A novel back-analysis approach for the external loads on shield tunnel lining in service based on monitored deformation. Structural Control and Health Monitoring, **2023**: 1–13. doi:10.1155/2023/8128701.

Lin, W., Sheil, B., Zhang, P., Zhou, B., Wang, C., and Xie, X. 2024. Seg2Tunnel: A hierarchical point cloud dataset and benchmarks for segmentation of segmental tunnel linings. Tunnelling and Underground Space Technology, **147**: 105735. doi:10.1016/j.tust.2024.105735.

Liu, Y., Jin, Y., Azizi, E., and Blumberg, A.J. 2023. CellStitch: 3D cellular anisotropic image segmentation via optimal transport. BMC Bioinformatics 2023–06. doi:10.1186/s12859-023-05608-2.

Lorensen, W.E., and Cline, H.E. 1987. Marching cubes: a high resolution 3D surface construction algorithm. ACM SIGGRAPH Computer Graphics, **21**: 163–169. doi:10.1145/37402.37422.

Meilă, M. 2007. Comparing clusterings—an information based distance. Journal of Multivariate Analysis, **98**: 873–895. doi:10.1016/j.jmva.2006.11.013.

Otsu, N. 1979. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, **9**: 62–66. doi:10.1109/TSMC.1979.4310076.

Pachitariu, M., and Stringer, C. 2022. Cellpose 2.0: how to train your own model. Nature Methods, **19**: 1634–1641. doi:10.1038/s41592-022-01663-4.

Pandey, K., Bin Masood, T., Singh, S., Hotz, I., Natarajan, V., and Murthy, T.G. 2022. Morse theory-based segmentation and fabric quantification of granular materials. Granular Matter, **24**: 1–20. doi:10.1007/s10035-021-01182-7.

Pinzón, G., Andò, E., Desrues, J., and Viggiani, G. 2023. Fabric evolution and strain localisation in inherently anisotropic specimens of anisometric particles (lentils) under triaxial compression. Granular Matter, **25**: 15. doi:10.1007/s10035-022-01305-8.

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., et al. 2021. Learning transferable visual models from natural language supervision. *In* International Conference on Machine Learning. PMLR. pp. 8748–8763.

Ren, S., Luzi, F., Lahrichi, S., Kassaw, K., Collins, L.M., Bradbury, K., and Malof, J.M. 2023. Segment anything, from space?

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., and Savarese, S. 2019. Generalized intersection over union: a metric and a loss for bounding box regression. *In* Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 658–666.

Rousset, O.G., Ma, Y., and Evans, A.C. 1998. Correction for partial volume effects in PET: principle and validation. Journal of Nuclear Medicine, **39**: 904–911.

Shi, Y., and Yan, W.M. 2015. Segmentation of irregular porous particles of various sizes from X-ray microfocus computer tomography images using a novel adaptive watershed approach. Géotechnique Letters, **5**: 299–305. doi:10.1680/jgele.15.00100.

Sun, Q., Zheng, J., and Li, C. 2019. Improved watershed analysis for segmenting contacting particles of coarse granular soils in volumetric images. Powder Technology, **356**: 295–303. doi:10.1016/j.powtec.2019.08.028.

Tang, Z.-Q., Yin, Z.-Y., Jin, Y.-F., and Zhou, X.-W. 2024a. A novel mesoscale modelling method for steel fibre-reinforced concrete with the combined finite-discrete element method. Cement and Concrete Composites, **149**: 105479. doi:10.1016/j.cemconcomp.2024.105479.

Tang, Z.-Q., Zhou, X.-W., Jin, Y.-F., Yin, Z.-Y., and Zhang, Q. 2024b. A novel coupled bES-FEM formulation with SUPG stabilization for thermo-hydro-mechanical analysis in saturated porous media. Computers and Geotechnics, **173**: 106454. doi:10.1016/j.compgeo.2024.106454.

Wiebicke, M., Andò, E., Herle, I., and Viggiani, G. 2017. On the metrology of interparticle contacts in sand from x-ray tomography images. Measurement Science and Technology, **28**: 124007. doi:10.1088/1361-6501/aa8dbf.

Wu, J., Zhang, Y., Fu, R., Fang, H., Liu, Y., Wang, Z., et al. 2023. Medical SAM adapter: adapting segment anything model for Medical image segmentation.

Wu, M., and Wang, J. 2023. Exploring particle breakage in sand under triaxial shearing using combined X-ray tomography and particle tracking method. Géotechnique, **74**: 1684–1699. doi:10.1680/jgeot.22.00351.

Xiao, Y., Long, L., Matthew Evans, T., Zhou, H., Liu, H., and Stuedlein, A.W. 2019. Effect of particle shape on stress-dilatancy responses of medium-dense sands. Journal of Geotechnical and Geoenvironmental Engineering, **145**: 04018105. doi:10.1061/(ASCE)GT.1943-5606.0001994.

Xie, J., Huang, X., and Jin, G. 2024. Analytical model for the sealant performance of tunnel gasketed joints based on multi-scale contact and percolation theories. Underground Space, **14**: 319–337. doi:10.1016/j.undsp.2023.08.004.

Xie, J., Wang, N., Huang, X., Lei, Q., and Molins, C. 2025. A novel A-star algorithm-based approach to predicting the sealant performance of shield tunnel's gasketed joints. Tunnelling and Underground Space Technology, **158**: 106379. doi:10.1016/j.tust.2025.106379.

Yang, J., and Luo, X.D. 2015. Exploring the relationship between critical state and particle shape for granular materials. Journal of the Mechanics and Physics of Solids, **84**: 196–213. doi:10.1016/j.jmps.2015.08.001.

Zhang, C., Han, D., Qiao, Y., Kim, J.U., Bae, S.-H., Lee, S., and Hong, C.S. 2023. Faster segment anything: towards lightweight SAM for mobile applications.

Zhang, P., Yin, Z.-Y., and Chen, Q. 2022. Image-based 3D reconstruction of granular grains via hybrid algorithm and level set with convolution kernel. Journal of Geotechnical and Geoenvironmental Engineering, **148**: 04022021. doi:10.1061/(ASCE)GT.1943-5606.0002790.

Zhang, Z., Cai, H., and Han, S. 2024. EfficientViT-SAM: accelerated segment anything model without performance loss.Proceedings

of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.

Zhao, J., and Guo, N. 2013. Unique critical state characteristics in granular media considering fabric anisotropy. Géotechnique, **63**: 695–704. doi:10.1680/geot.12.P.040.

Zhao, J., Jiang, M., Soga, K., and Luding, S. 2016. Micro origins for macro behavior in granular media. Granular Matter, **18**: 59. doi:10.1007/s10035-016-0662-9.

Zhao, S., Tan, D., Lin, S., Yin, Z., and Yin, J. 2023a. A deep learning-based approach with anti-noise ability for identification of rock microcracks using distributed fibre optic sensing data. International Journal of Rock Mechanics and Mining Sciences, **170**: 105525. doi:10.1016/j.ijrmms.2023.105525.

Zhao, S., Wang, F.-Y., Tan, D.-Y., and Yang, A.-W. 2024. A deep learning informed-mesoscale cohesive numerical model for investigating the mechanical behavior of shield tunnels with crack damage. Structures, **66**: 106902.

Zhao, S., Zhang, G., Zhang, D., Tan, D., and Huang, H. 2023b. A hybrid attention deep learning network for refined segmentation of cracks from shield tunnel lining images. Journal of Rock Mechanics and Geotechnical Engineering, **15**: 3105–3117. doi:10.1016/j.jrmge.2023.02.025.