

Cross-Scale Micro-CT Super-Resolution for Digital Rock Reconstruction in Heterogeneous Conglomerate Reservoirs

Yumin Li

China University of Petroleum, Beijing

Yang Zhang

2023216752@st.cupk.edu.cn

China University of Petroleum Beijing at Karamay

Yuwei Zu

Shengli Oilfield Lusheng Petroleum Development Company

Hao Liu

China University of Petroleum Beijing at Karamay

Ji Li

PetroChina Xinjiang Oilfield Company

Haoyu Wu

China University of Petroleum Beijing at Karamay

Jiexin Xu

China University of Petroleum Beijing at Karamay

Xiaoan Wang

China University of Petroleum Beijing at Karamay

Research Article

Keywords: Digital rock physics, Micro-CT, Super-resolution, Heterogeneous conglomerates, Pore-scale flow, Reservoir characterization

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11106599/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Cross-Scale Micro-CT Super-Resolution for Digital Rock Reconstruction in Heterogeneous Conglomerate Reservoirs

Yumin Li^a, Yang Zhang^{b,*}, Yuwei Zu^d, Hao Liu^b, Ji Li^c, Haoyu Wu^b, Jiexin Xu^b, Xiaoran Wang^b

^a School of Geosciences, China University of Petroleum (Beijing), Beijing 102249, China

^b School of Petroleum, China University of Petroleum (Beijing) at Karamay, Karamay,
Xinjiang 834000, China

^c Research Institute of Exploration and Development, PetroChina Xinjiang Oilfield Company,
Karamay, Xinjiang 834000, China

^d Shengli Oilfield Lusheng Petroleum Development Company, Limited, Dongying, Shandong
257000, China

* Corresponding author: Yang Zhang

Email: 2023216752@st.cupk.edu.cn

Orcid: <https://orcid.org/0000-0002-6876-9997>

Abstract

Accurate cross-scale characterization of heterogeneous reservoirs is needed for digital-rock analysis, hydrocarbon production, and subsurface energy storage assessment. Micro-computed tomography (micro-CT) faces a field-of-view-resolution trade-off: large-volume scans capture gravel-scale heterogeneity but miss pore-scale features. This study presents Hy-MambaIR, a hybrid Mamba-Transformer network for super-resolution reconstruction of micro-CT images from heterogeneous conglomerate reservoirs. A local-to-global serial architecture combines windowed self-attention and dynamic Tanh normalization for fine-pore recovery with Mamba-based global modeling and spatial-channel attention for large-scale gravel structures. The model was evaluated with paired multi-resolution micro-CT images and against representative convolutional and Transformer baselines. Hy-MambaIR achieved a peak signal-to-noise ratio of 33.36 dB, exceeding HAT by 1.31 dB, and reached 30 dB with 6.6-fold fewer training iterations than HAT. Physical validation showed that the reconstructed images preserved segmentation accuracy, pore topology,

pore-throat distributions, and single-phase flow fields; the mean relative error of three-directional absolute permeability on the 855^3 -voxel representative elementary volume was 18.56%. Supervised fine-tuning on pebbly sandstone reduced this MRE to 3.59%. These results establish a computational route for physically informed digital-rock reconstruction and reservoir characterization in heterogeneous geoenergy systems.

Keywords Digital rock physics; Micro-CT; Super-resolution; Heterogeneous conglomerates; Pore-scale flow; Reservoir characterization

1. Introduction

With the acceleration of the global energy transition, precise microstructural characterization of subsurface porous media has emerged as a fundamental basis for addressing critical scientific challenges in the energy sector (Heinemann et al., 2021; Bhadariya et al., 2024; Wang et al., 2024). Whether for efficient conventional hydrocarbon recovery or emerging technologies such as Carbon Capture, Utilization, and Storage (CCUS) (S. Li et al., 2025; Malik et al., 2024) and Underground Hydrogen Storage (UHS) (Pantaleo et al., 2024), the optimization of engineering schemes hinges on an accurate understanding of reservoir microscopic pore structures and fluid transport mechanisms. Digital Rock Physics (DRP) utilizes high-resolution imaging and numerical simulation (Khan, 2024) to provide a non-destructive and efficient means for predicting permeability and multiphase flow characteristics, serving as a pivotal bridge between microscopic pore mechanisms and macroscopic energy development (Behbahani et al., 2023; Malik et al., 2024).

Strongly heterogeneous reservoirs, typified by conglomerates, have become the focal point of deep energy exploration due to their vast resource potential. However, current imaging techniques face significant technical bottlenecks when characterizing such complex reservoirs. These rocks exhibit pronounced multiscale structural features, comprising centimeter-scale macroscopic gravel frameworks of varying morphologies and microscopic flow networks formed by micrometer-scale matrix and pores (Zhou et al., 2025). This structural complexity imposes a fundamental trade-off between field-of-view (FOV) and resolution. While large-FOV scanning captures the macroscopic gravel distribution, it fails to identify the critical microscopic pores that govern flow capacity; conversely, high-resolution scanning delineates microscopic details but is limited by a FOV far smaller than the Representative Elementary Volume (REV), making it difficult to characterize the

overall geological heterogeneity (Saxena et al., 2019; Ma et al., 2025). Consequently, Super-Resolution (SR) technology, which computationally reconstructs high-resolution details from low-resolution large-FOV images, has emerged as an effective approach to resolve this trade-off and achieve high-fidelity characterization of complex reservoirs (Li et al., 2024; Nabipour et al., 2024). Recent work has demonstrated the feasibility of deep neural networks for 3D digital-rock super-resolution (You et al., 2024), yet physically reliable reconstruction of strongly heterogeneous conglomerates remains challenging.

Over the past decade, deep learning-driven super-resolution (SR) techniques for digital rock physics have evolved rapidly, yet they continue to face severe challenges when processing large-scale industrial data. Early Convolutional Neural Networks (CNNs), such as SRCNN and RCAN, effectively improved image signal-to-noise ratios. However, constrained by their inherent local receptive fields, they struggle to capture the macroscopic contours and long-range geometric features of centimeter-scale clasts in conglomerates, resulting in structural distortions (Khan and Khanal, 2024; Niu et al., 2021; Wang et al., 2019). To enhance textural fidelity, Generative Adversarial Network (GAN)-based methods (e.g., SRGAN) have been widely explored. Nevertheless, these methods often yield artifacts that are geologically inconsistent, which compromises the reliability of subsequent fluid flow simulations (Mosser et al., 2017; Shan et al., 2022; Zhang et al., 2023). To capture long-range dependencies, Vision Transformers (ViTs) (e.g., SwinIR, HAT) were introduced, achieving significant performance gains. However, the quadratic computational complexity ($O(N^2)$) of the self-attention mechanism in ViTs results in prohibitive memory consumption and inference latency. This not only limits their scalability for processing large-scale 3D core data but also fails to align with the energy sector's demand for efficient, low-carbon computing (Kim et al., 2024; Xing et al., 2024). Consequently, existing methods struggle to simultaneously reconcile computational efficiency, macroscopic structural consistency, and microscopic physical fidelity. Therefore, developing a novel SR architecture capable of efficiently modeling global structures with linear complexity while precisely recovering microscopic pore details is of significant scientific and practical value (Meng et al., 2022; Yuan et al., 2023; Qiao et al., 2024).

To address these challenges, we propose Hy-MambaIR, a hybrid Mamba-Transformer super-resolution network tailored for strongly heterogeneous media. Built upon a "Local-Global Synergy"

design philosophy, the network incorporates the Hierarchical Attentive State Space Block (H-ASSB) as its core unit. The architecture adheres to a "Local-to-Global" principle: it first utilizes a Window-based Multi-Head Self-Attention (Window-MHSA) mechanism (Zraqou et al., 2025), coupled with Dynamic Tanh Normalization (DyT), to precisely capture high-frequency microscopic pore details within local receptive fields. Subsequently, leveraging the linear complexity of the State Space Model (Mamba) (Zhang et al., 2025), it efficiently establishes long-range dependencies between pixels, accurately reconstructing the spatial distribution of the macroscopic gravel framework. Furthermore, a Spatial-Channel Attention Block (SCAB) is integrated within the Mamba branch to perform a secondary refinement of global features. This design aims to overcome the computational bottlenecks of traditional ViTs, achieving efficient, high-fidelity representation of cross-scale features in strongly heterogeneous media.

The main contributions of this work are summarized as follows:

1. We propose Hy-MambaIR, a task-oriented hybrid Mamba-Transformer framework for cross-scale digital rock super-resolution in strongly heterogeneous conglomerate reservoirs, designed to mitigate the field-of-view–resolution trade-off in REV-scale reservoir characterization.

2. We introduce a local-to-global serial architecture that integrates a window-based local Transformer unit, an attentive state-space global modeling branch, and SCAB-based feature refinement, enabling coordinated recovery of microscopic pore details and macroscopic gravel structures.

3. We show that Hy-MambaIR achieves strong improvements in both reconstruction quality and computational efficiency, outperforming representative CNN- and Transformer-based baselines by more than 1.2 dB in PSNR and converging $6.6\times$ faster than comparable Transformer models.

4. We establish a physically grounded evaluation framework to validate reconstruction fidelity beyond image quality, including segmentation consistency, pore-network topology, and flow-related properties, and further demonstrate transferability across lithofacies through fine-tuning.

The remainder of this paper is organized as follows: Section 2 details the network architecture and core modules of Hy-MambaIR. Section 3 describes the experimental settings, datasets, and evaluation metrics. Section 4 presents the quantitative and qualitative results, including comparisons with representative baselines, ablation studies, and downstream application validation. Finally, Section 5 concludes the paper and outlines future work.

2. Methodology

This section details the proposed Hy-MambaIR, a hybrid Mamba–Transformer network designed for the super-resolution reconstruction of strongly heterogeneous digital rocks. To balance reconstruction fidelity, computational efficiency, and scalability for large digital-rock volumes, the framework adopts a slice-wise reconstruction scheme in which 2D cross-sectional images are super-resolved and then assembled into a 3D high-resolution (HR) volume. Although the reconstruction is performed in a slice-wise manner, the Mamba-based global modeling module helps preserve long-range structural consistency within each section, while the Transformer-based local modeling module enhances fine-scale pore textures and boundaries. We first present the overall network architecture (Section 2.1), whose core consists of multiple cascaded Attentive State Space Groups (ASSGs). Subsequently, we describe the main components in a bottom-up manner: the Hierarchical Attentive State Space Block (H-ASSB) (Section 2.2), and its key modules, H-SSM (Section 2.3), SCAB (Section 2.4), and DyT (Section 2.5).

2.1. Network Architecture

Hy-MambaIR is designed as a layered serial fusion architecture integrating convolution, window-based self-attention, and selective state space models, specifically tailored for the high-fidelity reconstruction of strongly heterogeneous media such as conglomerates. The overall architecture, illustrated in Fig. 1, comprises three core stages: shallow feature extraction, deep feature refinement, and high-resolution image reconstruction.

Shallow Feature Extraction: Given a low-resolution (LR) 2D digital rock slice input $X_{LR} \in R^{C_{in} \times H \times W}$, where C_{in} denotes the input channel number, and H and W represent height and width, respectively. A standard 3×3 convolutional layer first processes it, mapping it into a high-dimensional embedding space of dimension C , yielding the initial feature $F_0 \in R^{C \times H \times W}$.

Deep Feature Refinement: The network backbone consists of L cascaded Attentive State Space Groups (ASSGs). Each ASSG serves as an independent residual feature extraction unit, designed to capture complex cross-scale features through deep stacking. Within each ASSG, the core computation is performed by D stacked Hierarchical Attentive State Space Blocks (H-ASSBs). To facilitate effective transition between the image and sequence domains, features undergo a PatchUnembed operation before entering the stacked H-ASSBs and are restored via PatchEmbed

after processing. Additionally, a local 3×3 convolutional layer is placed at the end of the ASSG, forming a stable residual structure via a Long Skip Connection (LSC) with the input features. This deep network body focuses on learning high-frequency residual information, denoted as F_{res} :

$$F_{res} = H_{net}(F_0) \quad (1)$$

where $H_{net}(\cdot)$ represents the deep feature extractor composed of L ASSGs. The final deep feature F_{deep} is obtained by summing the shallow feature and the residual feature via the LSC:

$$F_{deep} = F_0 + F_{res} \quad (2)$$

High-Resolution Image Reconstruction: The deep feature F_{deep} is processed by an upsampling module (e.g., sub-pixel convolution) and a final 3×3 convolutional layer to generate the high-resolution (HR) image $Y_{SR} \in R^{C_{out} \times sH \times sW}$, where (C_{out}) is the output channel number and (s) is the scale factor.

Loss Function: The training objective is to minimize the discrepancy between the reconstructed image and the ground truth HR image. To prioritize network architecture performance and ensure physical fidelity, we adopt the Charbonnier loss. This loss function emphasizes pixel-level reconstruction accuracy, ensuring fair comparison with existing state-of-the-art (SOTA) methods while avoiding artifacts often introduced by complex loss functions. The calculation is defined as follows (where (ϵ) is a small constant set to 10^{-3} for numerical stability):

$$L = \sqrt{\|Y_{SR} - Y_{HR}\|_2^2 + \epsilon^2} \quad (3)$$

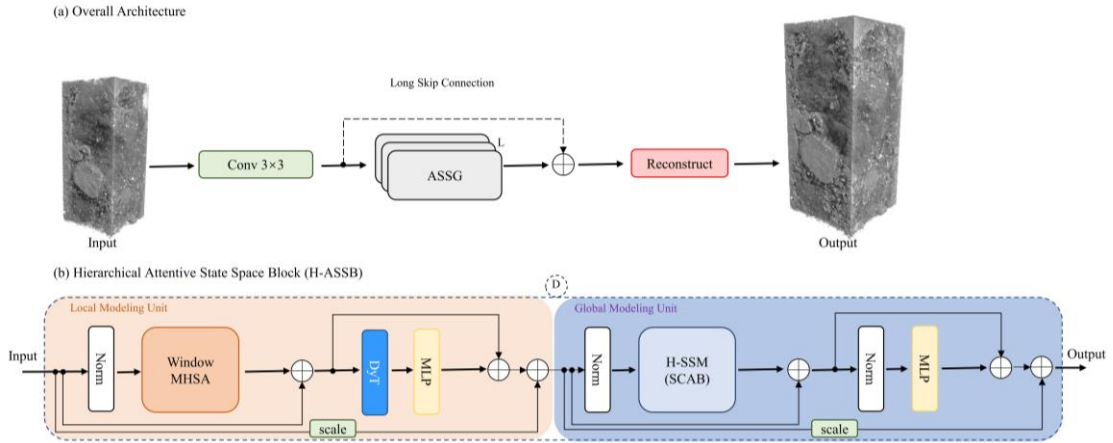


Fig. 1. Architecture of the Hy-MambaIR model.

2.2. Hierarchical Attentive State Space Block

The hierarchical attentive state space block (H-ASSB) is the core building unit of Hy-MambaIR and is designed to perform hierarchical, serial feature extraction that first emphasizes local patterns and then progressively incorporates global information. As illustrated in Fig. 1(b), H-

ASSB adopts a serial internal architecture in which each attentive layer is composed of two functionally complementary units.

2.2.1. Local Modeling Unit

The local modeling unit is built upon a Transformer architecture and is responsible for capturing the high-frequency local details in the images. This unit contains a shifted-window multi-head self-attention (SW-MHSA) module. To enhance the representational capacity of local features, a dynamic Tanh normalization (DyT) layer (Zhu et al., 2025) is inserted between the window-based MHSA and the multilayer perceptron (MLP), as illustrated in Fig. 1(b). As a nonlinear pre-processing mechanism, the DyT layer helps the MLP extract and amplify key features that are crucial for reconstructing high-frequency details. Its formulation is expressed as:

$$DyT(x) = \gamma \odot \tanh(\alpha \cdot x) + \beta \quad (4)$$

where x denotes the input features; α is a learnable scalar that dynamically adjusts the curvature of the activation function; and γ and β are learnable scaling and bias vectors that act as an affine transformation to restore the expressive capacity of the features.

2.2.2. H-SSM (Hierarchical Attentive State Space Module)

Global modeling unit. After the local details are extracted, the feature maps are fed into the global modeling unit. This unit is implemented as a hierarchical attentive state space module (H-SSM), which aims to exploit the local representations and, with linear computational complexity, build an efficient global perception of the macroscopic geological structure.

The H-SSM integrates a two-stage refinement process to hierarchically handle global context and local details. As shown in Fig. 2, its workflow mainly consists of three components.

Input pre-processing and semantic reordering. The input features are first processed by a combination of a 1×1 convolution and a 3×3 depthwise separable convolution, forming a convolutional positional encoding (CPE) and increasing the channel dimension. Then, a semantic-guided neighborhood (SGN) algorithm (Guo et al., 2025) is used to reorder the features such that semantically similar pixels are clustered in sequence space.

Attentive state space equation (ASE). The reordered sequence is passed to the selective state-space core based on the ASE. By dynamically injecting prompts, the ASE performs non-causal global modeling over the sequence and outputs features F_{global} that embed global contextual information.

Secondary local enhancement. To further strengthen the feature representations, a spatial-channel attention block (SCAB) is applied to F_{global} for secondary refinement, and the refined features are adaptively fused using a learnable scaling factor σ to obtain the final output of H-SSM:

$$F_{H-SSM} = F_{global} + \sigma \cdot SCAB(F_{global}) \quad (5)$$

where F_{global} denotes the global features produced by the ASE, and σ is a learnable scaling factor that balances the contributions from the global features and the refined local features.

To increase the nonlinear transformation capacity, convolutional feed-forward networks (ConvFFNs) are attached to the end of both the local and global modeling units. In addition, a layer scale mechanism is introduced at each residual connection, where a learnable diagonal matrix is used to reweight the residual branch. This mechanism effectively regulates the signal contributions of the local and global units and stabilizes the training of deep networks, without incurring noticeable computational overhead.

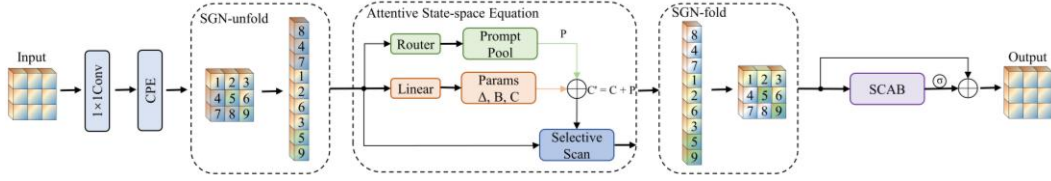


Fig. 2. Schematic diagram of the H-SSM architecture.

2.3. Spatial-channel attention block (SCAB)

The SCAB adopts a cascaded structure and consists of a channel attention block (CAB), a parameter-free spatial attention module (SimAM), and a spatially adaptive feature modulation (SAFM) module (Sun et al., 2023), as depicted in Fig. 3. Given an input feature map $F_{in} \in R^{C \times H \times W}$, the output of SCAB, F_{out} , can be formulated as the following cascaded nonlinear transformations:

$$F_{out} = \mathcal{F}_{SAFM}(\mathcal{F}_{SimAM}(\mathcal{F}_{CAB}(F_{in}))) \quad (6)$$

where $\mathcal{F}_{CAB}$, $\mathcal{F}_{SimAM}$, and $\mathcal{F}_{SAFM}$ denote the nonlinear mappings of the CAB, SimAM, and SAFM submodules, respectively.

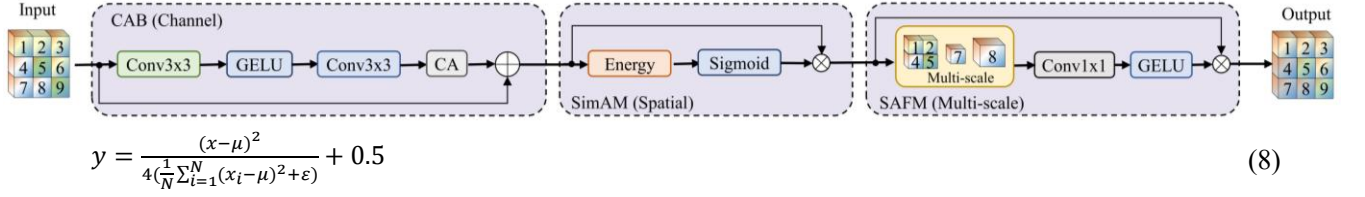
Local context encoding and channel reweighting (CAB). CAB comprises two 3×3 2D convolution layers ($\text{Conv}_{3 \times 3}$) followed by a channel attention module $\mathcal{A}_{ch}$, which selects feature channels that are critical for reconstructing local details:

$$\mathcal{F}_{CAB} = \mathcal{A}_{ch}(\text{Conv}_{3 \times 3}(\text{GELU}(\text{Conv}_{3 \times 3}(F_{in})))), \quad (7)$$

where $\text{GELU}(\cdot)$ denotes the Gaussian error linear unit, and $\mathcal{A}_{ch}(\cdot)$ represents the channel attention.

Parameter-free spatial saliency enhancement (SimAM). The features F_{CAB} are then fed into the

SimAM module (Yang et al., 2021). SimAM evaluates the importance of each neuron in the feature map using an energy function, and the simplified energy map y is computed as:



where μ is the mean of all pixels in the channel, N is the total number of pixels, and ϵ is a regularization term. The output of SimAM is finally obtained by applying a sigmoid activation to the energy map and multiplying it with the original features:

$$F_{SimAM} = F_{CAB} \odot \sigma(y) \quad (9)$$

where $\sigma(\cdot)$ is the sigmoid function and $\odot$ denotes element-wise multiplication.

Multi-scale spatial feature modulation (SAFM). Finally, F_{SimAM} is fed into the spatially adaptive feature modulation (SAFM) module to generate the final output:

$$F_{out} = GELU(M) \odot F_{SimAM} \quad (10)$$

$$M = \mathcal{F}_{ms}(F_{SimAM}) \quad (11)$$

where M is a dynamic modulation mask generated by the multi-scale processing branches $\mathcal{F}_{ms}$.

Fig. 3. Schematic diagram of the SCAB architecture.

3. Experimental settings and evaluation indicators

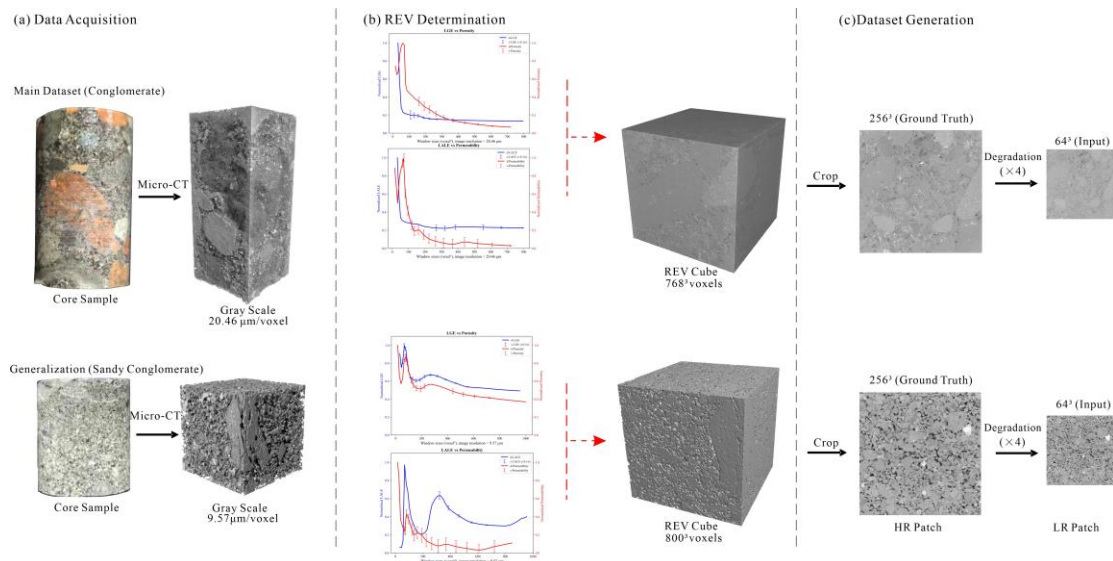
3.1. Dataset preprocessing

To overcome the limitations of existing open-source datasets in terms of lithological representativeness and physical dimensions, and thereby remove their constraints on holistic investigations of pore structures, we construct a new dataset specifically dedicated to super-resolution studies of strongly heterogeneous conglomerate reservoirs. The core samples used in this study were collected from the Mahu sag of the Junggar Basin, China. They belong to the Baikouquan Formation, and the lithology is dominated by fan-delta conglomerates (X. Li et al., 2025; Wu et al., 2021). The workflow for constructing the dataset is illustrated in Fig. 4.

First, in the data acquisition stage (Fig. 4(a)), a NanoVoxel-5000 micro-CT imaging system was used to scan 2.5 cm-diameter plug samples, producing three-dimensional grayscale volumes with a voxel resolution of 20.46 μm . To determine a statistically stable representative elementary volume (REV) for the present samples, we analyzed the data using the gray-level size zone matrix

(GLSZM) method (Bruns et al., 2017) (Fig. 4(b)). By computing two key statistics—the low gray-level emphasis (LGE), which describes the overall distribution of low gray levels, and the large-area low gray-level emphasis (LALE), which characterizes the distribution of large connected regions—we quantitatively characterize the size distribution and connectivity of the internal pore-throat structures. The analysis indicates that, as the analysis window size increases, the LGE and LALE statistics first exhibit fluctuations and then converge (see Supplementary Fig. S1 for the detailed convergence curves). The results indicate that when the window size reaches 768^3 voxels, the statistics become stable. Based on this convergence behavior, the volume is further cropped to 855^3 voxels as the high-resolution (HR) reference volume for subsequent reconstruction and physical evaluation.

In the dataset generation stage (Fig. 4(c)), sub-volumes of size 256^3 are randomly cropped from the REV volume and used as HR samples. The corresponding low-resolution (LR) images are real low-resolution micro-CT measurements that are spatially matched to the HR samples to construct



paired HR/LR data. The dataset is split into training, validation, and test subsets with a ratio of 8:1:1. After training, the reconstructed models are further evaluated in terms of rock-physics accuracy to ensure the reliability of the reconstruction.

Model generalization capability is crucial. We therefore adopt weight fine-tuning to perform high-fidelity reconstructions for different lithofacies. We further select pebbly sandstones from the Mahu sag and perform the same acquisition procedure (Fig. 4(a)). A 2.5 cm-diameter plug is scanned by CT, and an REV is then determined, yielding a dataset of size 800^3 voxels. The same registration and paired-data construction procedure was used for this dataset to form the generalization test set.

The paired real LR/HR measurements were then used for supervised fine-tuning and evaluation on this previously unseen lithofacies.

Fig. 4. Schematic diagram of the data processing workflow.

3.2. Evaluation metrics

The evaluation framework in this study comprises two categories: image reconstruction quality and rock-physics accuracy. Image reconstruction quality is primarily evaluated using peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) (Wang et al., 2004). In addition, we adopt the learned perceptual image patch similarity (LPIPS) metric to assess the perceptual quality of the reconstructions and to better preserve the morphological features of different minerals, which aligns with recent evaluation trends. Meanwhile, for the reconstructed images, an iterative adaptive thresholding method is employed for binarization, and the intersection over union (IoU) and Dice coefficient are adopted as segmentation accuracy metrics to quantify the reconstruction accuracy (Andrä et al., 2013a; Suo et al., 2024). PSNR is defined as:

$$\text{PSNR} = 20\log_{10}\left(\frac{\text{MAX}}{\sqrt{\text{MSE}}}\right) \quad (12)$$

$$\text{MSE} = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W (x_{ij} - y_{ij})^2 \quad (13)$$

where MAX is 255; H and W denote the height and width of the image, respectively; and x and y are the images being compared. A higher PSNR generally indicates better reconstruction quality; however, because it only accounts for pixel-wise errors, it tends to favor overly smooth reconstructions.

SSIM compares the statistics within local windows in terms of luminance, contrast, and structure, thereby evaluating whether structural information is preserved. It is defined as:

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (14)$$

where the luminance μ and contrast σ are computed from the mean and standard deviation of the image intensities, and the constants are set to $C_1 = (0.01 \cdot 255)^2$, $C_2 = (0.03 \cdot 255)^2$. Values closer to 1 indicate more similar structures; compared with PSNR, SSIM places more emphasis on perceptual factors such as texture and contrast.

The LPIPS metric exhibits a strong correlation with human subjective evaluations and is particularly effective at capturing texture and semantic details (Zhang et al., 2018). It is defined as:

$$\text{LPIPS}(x, y) = \sum_l \frac{1}{H_l W_l} \sum_{h=1}^{H_l} \sum_{w=1}^{W_l} |w_l \odot (\hat{\phi}^l(x)_{hw} - \hat{\phi}^l(y)_{hw})|^2 \quad (15)$$

where (ϕ_l) denotes the features extracted at the l -th layer of the backbone network (e.g.,

AlexNet, VGG); $\hat{\phi}_l(\cdot) = \phi_l(\cdot)/|\phi_l(\cdot)|_2$ denotes the channel-wise normalized features; w_l are learned channel weights; and (H_l, W_l) are the spatial dimensions of the feature maps at layer l .

The segmentation accuracy metrics directly quantify the geometric fidelity of pore structures in the segmented reconstructions and constitute the basis for evaluating physical accuracy. The intersection over union (IoU), also known as the Jaccard index, computes the ratio between the intersection and union of the predicted pore region and the ground-truth pore region, and intuitively reflects the overall overlap of the segmentation results:

$$IoU = \frac{|B_{SR} \cap B_{HR}|}{|B_{SR} \cup B_{HR}|} = \frac{TP}{TP + FP + FN} \quad (16)$$

The Dice coefficient is another commonly used metric for segmentation accuracy, and computationally it places greater emphasis on the intersection region:

$$Dice = \frac{2|B_{SR} \cap B_{HR}|}{|B_{SR}| + |B_{HR}|} = \frac{2TP}{2TP + FP + FN} \quad (17)$$

where TP (true positives), FP (false positives), and FN (false negatives) denote the numbers of pore voxels correctly identified as pores, solid voxels incorrectly classified as pores, and pore voxels incorrectly classified as solid, respectively.

Rock-physics accuracy should be given the same priority as pixel-level accuracy and is a key indicator for evaluating reconstruction performance (Behbahani et al., 2023). In this study, physical accuracy is assessed through two complementary approaches: topological analysis and flow simulation. The topological analysis employs parameters such as porosity, two-point correlation function, tortuosity, pore-size distribution, and Euler characteristic (Andr  et al., 2013b; Qin et al., 2022). Flow behavior is further characterized by single-phase flow simulations and three-direction absolute permeability computed using the Dragonfly software (Andr  et al., 2013b; Pramana and Thaib, 2022).

Porosity, defined as the ratio of pore volume to the total core volume, is the most fundamental parameter for evaluating reservoir storage capacity:

$$\phi = \frac{V_p}{V_{total}} = \frac{\sum B(i,j,k)}{D \times H \times W} \quad (18)$$

where $B(i, j, k)$ equals 1 at pore voxels and 0 at solid voxels.

The two-point correlation function $S_2(r)$ quantifies the probability that two points separated by a distance r simultaneously lie in the pore phase, thus capturing the spatial correlation of the pore structure across different scales. Here, x and $x + r$ denote two random points in space separated by

distance r (Gonzalez et al., 2020):

$$S_2(r) = P(x \in V_p, x + r \in V_p) \quad (19)$$

Tortuosity describes the ratio between the actual path length L_e of a fluid flowing through the porous medium and the macroscopic sample length L , and reflects the winding degree of the flow paths (Fu et al., 2021):

$$\tau = \left(\frac{L_e}{L}\right) \quad (20)$$

The pore-size distribution serves as a bridge between image reconstruction quality and macroscopic physical accuracy and also reflects the ability to recover structures across scales. The Euler characteristic is a key indicator for quantifying the topological connectivity of the pore-throat network, and the specific Euler characteristic is obtained by normalizing the Euler characteristic by the total volume (Ávila et al., 2022):

$$\chi_v = \frac{\beta_0 - \beta_1}{V_{total}} \quad (21)$$

where β_0 denotes the number of isolated pore connected components, β_1 denotes the number of independent internal loops, and V is the total volume. In addition, single-phase flow simulations are performed on the digital rocks, and Darcy's law is used to compute the absolute permeability in three orthogonal directions, thereby characterizing the macroscopic flow capacity of the rock.

3.3. Experimental settings

The experimental environment and training strategy in this study are carefully designed, and all experiments are conducted on a high-performance computing platform. The hardware configuration consists of an NVIDIA RTX 5090 GPU (32 GB) and 25 vCPUs on an Intel(R) Xeon(R) Platinum 8470Q processor. The software environment includes Python 3.10, PyTorch 2.8.0, and CUDA 12.8. The Hy-MambaIR model takes low-resolution (LR) patches of size 64 as input and outputs high-resolution (HR) patches of size 256. It is trained using the Adam optimizer for 15,000 iterations, with an initial learning rate of 2×10^{-5} and $\beta = (0.9, 0.999)$. The learning rate is scheduled using MultiStepLR with a decay factor of 0.5, and the effective number of training patches per update is fixed at 72. The luminance (Y) channel of the YCbCr color space is used for training. To improve training efficiency and resource utilization, we employ mixed-precision training, gradient accumulation, and multi-process data loading.

4. Results and Discussion

4.1. Comparative analysis with state-of-the-art (SOTA) models

To comprehensively evaluate the overall performance of Hy-MambaIR in processing strongly heterogeneous conglomerate reservoirs, we systematically compare it with representative super-resolution baselines. The comparison baselines cover a traditional interpolation method (Bicubic), classical CNN-based architectures (SRCNN, EDSR, RCAN, RRDB), and advanced Transformer-based architectures (SwinIR, HAT). We conduct an in-depth analysis along two dimensions: image reconstruction quality (quantitative and qualitative) and model computational efficiency. All models are trained and tested under exactly the same data splits and evaluation metrics to ensure fair and comparable results. The selected baselines are intended to cover representative interpolation, CNN-based, and Transformer-based SR paradigms widely used in image reconstruction studies.

4.1.1. Image reconstruction quality analysis

Quantitative evaluation. As shown in Table 1, Hy-MambaIR achieves the best accuracy and stability across all evaluation metrics. For the core PSNR metric, Hy-MambaIR reaches 33.36 dB. Compared with the state-of-the-art Transformer model HAT (32.05 dB) and the CNN model RCAN (32.10 dB), our model achieves a significant gain of more than 1.2 dB. The lowest MAE (0.01677) and RMSE (0.02184) further demonstrate its high regression accuracy, while the low standard deviation (± 0.36) indicates that Hy-MambaIR is highly robust, maintaining consistently strong performance across different test samples, which is crucial for accurately recovering the physical properties of the rock matrix. For the SSIM metric, Hy-MambaIR achieves the highest score of 0.7941, indicating that its reconstructions are closest to the real cores in terms of structural information, such as pore structures. For the LPIPS metric, which reflects human visual perception, Hy-MambaIR attains the lowest value of 0.1974, corresponding to the best perceptual quality among all methods. Combined with the Edge Preservation Index (EPI, 0.3164) and the contrast quality index (CQI, 6.6463), these results indicate that the model preserves edge sharpness while effectively suppressing high-frequency artifacts introduced by over-sharpening, thus achieving an optimal balance between structural fidelity and natural texture appearance.

Qualitative evaluation. To intuitively demonstrate the capability of the model in handling cross-scale geological features, visual reconstruction comparisons at different scales are presented in Figs. 5–7.

Table 1. Comparison of model performance.

Model	PSNR (dB)	SSIM	LPIPS	MAE	RMSE	CQI	EPI
Bicubic	31.45±0.37	0.7415±0.0109	0.3209±0.0219	0.02057	0.02677	7.5997	0.2887
Hy-MambaIR	33.36±0.36	0.7941±0.0081	0.1974±0.0136	0.01677	0.02184	6.6463	0.3164
EDSR	32.10±0.47	0.7595±0.0083	0.2076±0.0299	0.01907	0.02488	6.7288	0.3229
HAT	32.05±0.48	0.7587±0.0083	0.2215±0.0314	0.01917	0.02501	6.9001	0.3230
RCAN	32.10±0.47	0.7596±0.0083	0.2075±0.0299	0.01906	0.02487	6.7284	0.3225
RRDB	32.04±0.47	0.7581±0.0084	0.2249±0.0259	0.01918	0.02503	6.8328	0.3187
SRCNN	31.89±0.44	0.7466±0.0096	0.2740±0.0207	0.01959	0.02547	6.9350	0.2942
SwinIR	32.07±0.47	0.7587±0.0082	0.2072±0.0289	0.01911	0.02494	6.7132	0.3233

Microscopic pore and grain recovery (small scale, Fig. 5). In the 256×256 -pixel local view (highlighted by red arrows and boxes), Bicubic and SRCNN produce severe blurring, causing fine pore structures to almost disappear. Although RCAN recovers partial edges, its limited local receptive field leads to distorted pore morphology, while grain boundaries reconstructed by other models remain relatively blurred. In contrast, Hy-MambaIR accurately reconstructs the geometric shapes of micron-scale pores and grains, with sharp edges that closely match the HR reference. Detailed comparisons at additional locations and viewing angles are provided in Supplementary Fig. S2.

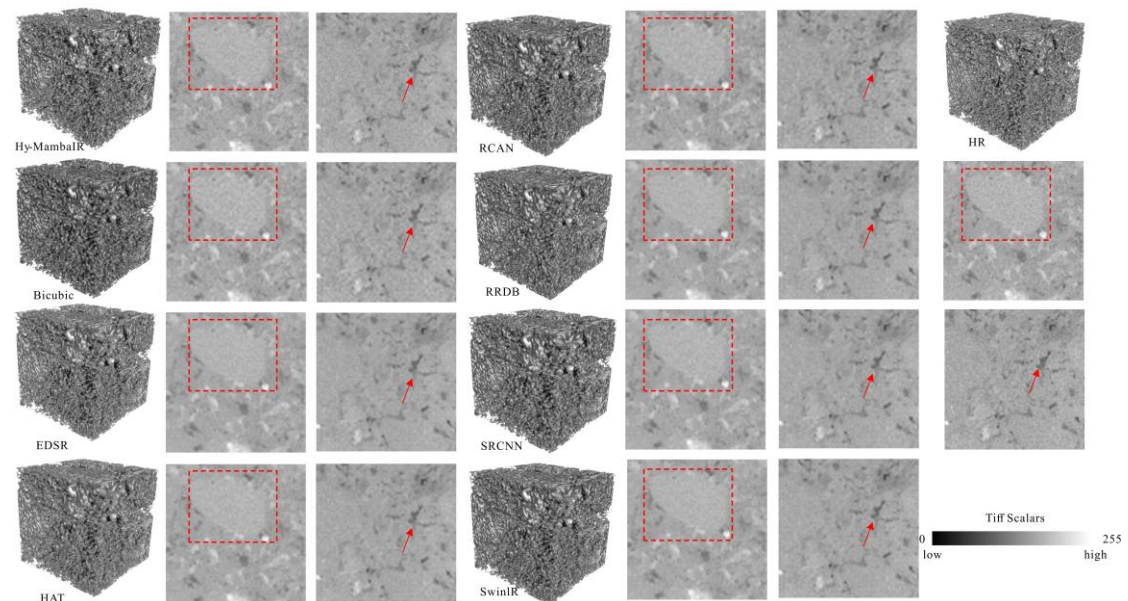


Fig. 5. Comparison of small-scale (256 pixels) reconstruction results.

Complex boundaries and fractures (medium scale, Fig. 6). In the 342-pixel view, we focus on the contact boundaries between gravels and matrix, as well as micro-fractures (highlighted by yellow dashed circles). HAT and SwinIR exhibit ringing artifacts when processing these high-frequency transition regions. In contrast, Hy-MambaIR, benefiting from the nonlinear optimization of DyT, successfully reconstructs continuous, smooth, and sharp mineral boundaries

and fracture paths without introducing artificial textures.

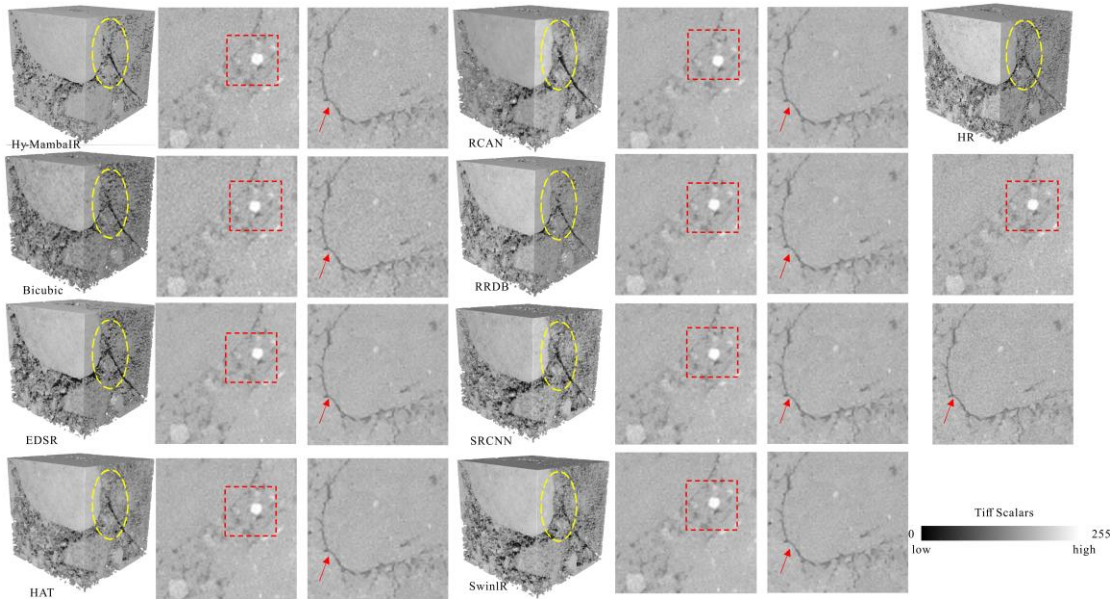


Fig. 6. Comparison of medium-scale (342 pixels) reconstruction results.

Macroscopic structural consistency (large scale, Fig. 7). Under a large field of view of 855 pixels, the advantages of Hy-MambaIR become particularly evident. Benefiting from the global perception capability of H-SSM, our model not only excels in local texture reconstruction but also preserves the integrity of the gravel framework and the overall uniformity of the rock structure at the macroscopic scale, avoiding the blocky artifacts commonly observed in CNN-based methods during large-scale reconstruction.

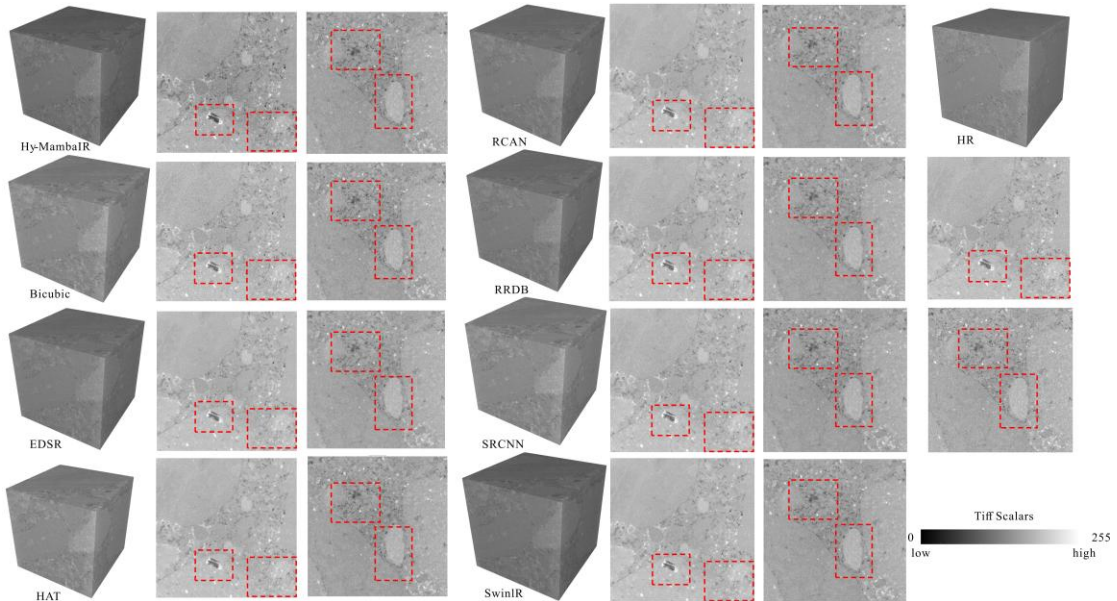
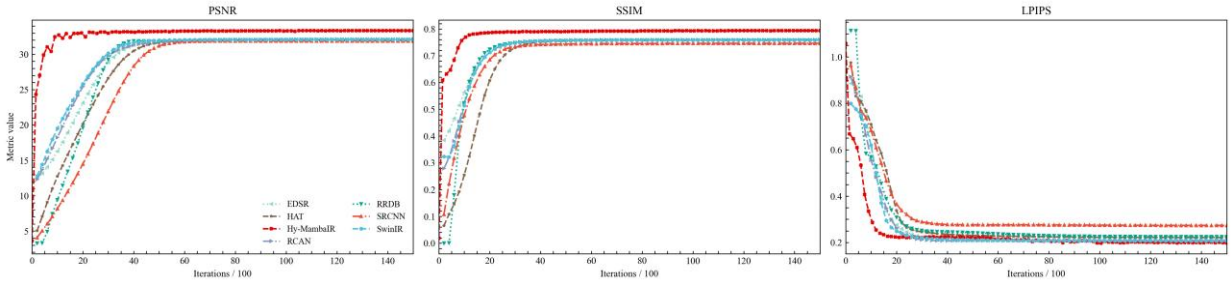


Fig. 7. Comparison of large-scale (855 pixels) reconstruction results.

4.1.2. Efficiency and convergence analysis

In addition to its exceptional image reconstruction quality, Hy-MambaIR demonstrates substantial potential for engineering applications, owing to its high training efficiency and low computational cost, which together reduce computational burden and facilitate industrial deployment. Table 2 and Fig. 8 highlight these significant advantages over Transformer-based methods.

Fast convergence characteristics. As illustrated in Fig. 8, the training curve of Hy-MambaIR (solid red line) exhibits a steep ascent in the initial phase and is the first to reach a steady state. To quantify this, we calculate the average growth rate during early training (metric increment per 100 iterations, Avg Δ /100 iters). Table 2 shows that Hy-MambaIR achieves an Avg Δ PSNR of 0.2067 dB, exceeding HAT (0.1825) and the CNN-based RCAN (0.1337). This indicates higher early-training efficiency under identical computational resources. Crucially, the time-to-target performance is improved. To reach a PSNR benchmark of 30 dB, Hy-MambaIR requires only 600 iterations, whereas HAT and EDSR require 4,000 and 3,400 iterations, respectively. This corresponds to a $6.6\times$ acceleration in convergence speed, substantially reducing the total training



time to just 1.62 hours (compared to 2.48 hours for HAT).

Fig. 8. Comparison of efficiency and performance curves among different models.

Balance between parameters and inference efficiency. Despite incorporating a hybrid Transformer architecture, Hy-MambaIR leverages the linear-time complexity of its Mamba-based modules to achieve superior efficiency over pure Transformer models. As shown in Table 2, Hy-MambaIR contains 14.93 M parameters, reflecting its lightweight design. Compared with its closest performance competitor, HAT (20.77 M), the parameter count of our model is reduced by approximately 28%, and it is also lighter than RCAN (15.59 M). In terms of inference speed, under the same hardware environment, Hy-MambaIR requires 0.0305 s to process a single image. Although lightweight models such as SRCNN and EDSR exhibit lower inference latency, their relatively shallow network depth prevents them from capturing the complex cross-scale features of

conglomerate reservoirs, leading to significantly inferior reconstruction accuracy (with PSNR differences exceeding 1.2 dB) and making them unable to satisfy the high-fidelity requirements of downstream rock-physics tasks. In contrast, while maintaining competitive accuracy, Hy-MambaIR achieves an inference speed comparable to that of HAT and higher than that of RCAN (0.036 s), thereby validating the high efficiency of its architectural design.

Table. 2. Comparison of efficiency metrics and parameter counts among different models.

Model	Number of parameters (M)	Training time (hours)	Reconstruction time (seconds)	Avg Δ PSNR /100 iters	Avg Δ SSIM /100 iters	Avg Δ LPIPS /100 iters	Iters-PSNR 30	Iters-SSIM 0.7
Bicubic	-	-	0.0007	-	-	-	-	-
Hy-MambaIR	14.93	1.62	0.0305	0.2067	0.0054	0.0086	600	750
EDSR	1.52	0.62	0.0028	0.1329	0.0028	0.0148	3400	2000
HAT	20.77	2.48	0.0308	0.1825	0.0047	0.0055	4000	2600
RCAN	15.59	1.91	0.036	0.1337	0.0038	0.0186	3200	2000
RRDB	16.70	2.16	0.0189	0.1942	0.0051	0.0064	3200	1800
SRCNN	0.057	0.98	0.0013	0.1879	0.0046	0.0057	4600	2200
SwinIR	11.90	1.68	0.0168	0.1303	0.0036	0.004	3000	2000

4.2. Ablation study

To systematically validate the effectiveness of each innovative component in Hy-MambaIR for handling strongly heterogeneous conglomerate reservoirs and addressing the underlying physics of CT imaging, we conducted a series of comprehensive ablation experiments. Using a strong MambaIR architecture as the baseline, we progressively introduced architectural ordering refinements, DyT normalization, and the SCAB module to evaluate their individual contributions. All experiments were conducted on the same dataset under identical training configurations. The quantitative results are summarized in Table 3, and visual reconstruction comparisons are presented in Fig. 9.

4.2.1. Architectural ordering for cross-scale feature modeling

The core challenge of conglomerate reservoirs lies in their pronounced cross-scale structures, spanning centimeter-scale macroscopic gravel frameworks and micron-scale matrix pores. To address this issue, we compared two architectural variants: placing the local enhancement module at the input stage (+CAB on Input) versus placing it after the global state space module (+CAB on SSM). When local enhancement is applied first, the model tends to overfit microscopic textures while neglecting the continuity of macroscopic lithology, leading to artifacts within the gravel framework. By contrast, placing local refinement after the SSM aligns with a reconstruction logic that first establishes the global pore structure and then refines local details.

As shown in Table 3, compared with the baseline MambaIR, adding CAB at the input (+CAB on Input) increases the PSNR by only 0.03 dB, with a slight decrease in SSIM. However, when CAB is placed after the SSM module (+CAB on SSM), although the PSNR remains at 31.93 dB, the LPIPS value significantly improves from 0.2395 to 0.2298, corresponding to a reduction of approximately 4%. These results demonstrate that performing local detail refinement after global context modeling yields reconstructions with higher perceptual quality and geologically more plausible structures. This verifies the superiority of our serial architecture; accordingly, all subsequent experiments are based on this optimized “global-first, then-local” configuration.

Table 3. Ablation study of the Hy-MambaIR model.

Model	PSNR (dB)	SSIM	LPIPS	MAE	RMSE	CQI	EPI
MambaIR	31.90±0.51	0.7558±0.0080	0.2222±0.0236	0.01947	0.02546	7.1231	0.3167
+ CAB on Input	31.93±0.48	0.7552±0.0083	0.2395±0.0205	0.01943	0.02537	7.0839	0.3156
+ CAB on SSM	31.93±0.48	0.7549±0.0084	0.2298±0.0282	0.01943	0.02535	7.2132	0.3180
+ CAB on SSM w/ DyT	29.73±0.77	0.6673±0.0115	0.1280±0.0294	0.02530	0.03277	7.4260	0.2182
Hy-MambaIR	33.36±0.36	0.7941±0.0081	0.1974±0.0136	0.01677	0.02184	6.6463	0.3164

4.2.2. Addressing CT imaging noise and nonlinear mapping

After establishing the architectural ordering, we further investigated the synergistic mechanism between SCAB and DyT in addressing the low signal-to-noise ratio and nonlinear mapping characteristics of CT images. When DyT normalization was applied in isolation, the PSNR suffered a catastrophic drop to 29.73 dB, and the SSIM fell to 0.6673. From a physical viewpoint, raw CT images, in which voxel intensities are approximately proportional to rock density, are often contaminated by Poisson noise and various imaging artifacts. The strong nonlinear tanh activation in DyT, designed to enhance edge contrast, indiscriminately amplifies background noise in the absence of effective feature selection. This distorts the approximately linear relationship between grayscale values and rock density and ultimately destroys the physical mapping between intensity and density, causing the model to collapse.

To resolve these issues, we introduced the SCAB module (Fig. 3). Within SCAB, SimAM employs an energy function to automatically focus on high-frequency pore edges, while SAFM smooths matrix noise at multiple scales. SCAB thus acts as a feature purifier, precisely separating effective signals corresponding to pore and framework boundaries from cluttered CT responses while suppressing imaging noise. When these purified, high-quality features are fed into DyT, the

synergistic effect between the two modules becomes apparent: the strong nonlinearity of DyT no longer amplifies noise but instead performs precise sharpening of the pore–framework transitions. This synergy boosts the PSNR to 33.36 dB, increases the SSIM to 0.7941, and further reduces the LPIPS to 0.1974. As shown in Fig. 9, Hy-MambaIR produces reconstructions that are significantly superior to those of all other variants in terms of sharpness, contrast, and texture details, and are closest to the ground truth. Notably, it avoids the blurring and smearing artifacts commonly observed in other variants when recovering tiny pores.

The ablation experiments confirm that the superior performance of Hy-MambaIR does not arise from naive module stacking. Instead, it is achieved by adopting a “global-first, then-local” architectural strategy and combining SCAB-based feature purification with DyT-based nonlinear sharpening to jointly tackle the dual challenges of cross-scale geological structures and CT imaging noise. Furthermore, single-phase flow simulations on different ablation variants (detailed in Supplementary Fig. S3) provide fluid-dynamic evidence that the full architecture offers an advantage in restoring pore connectivity.

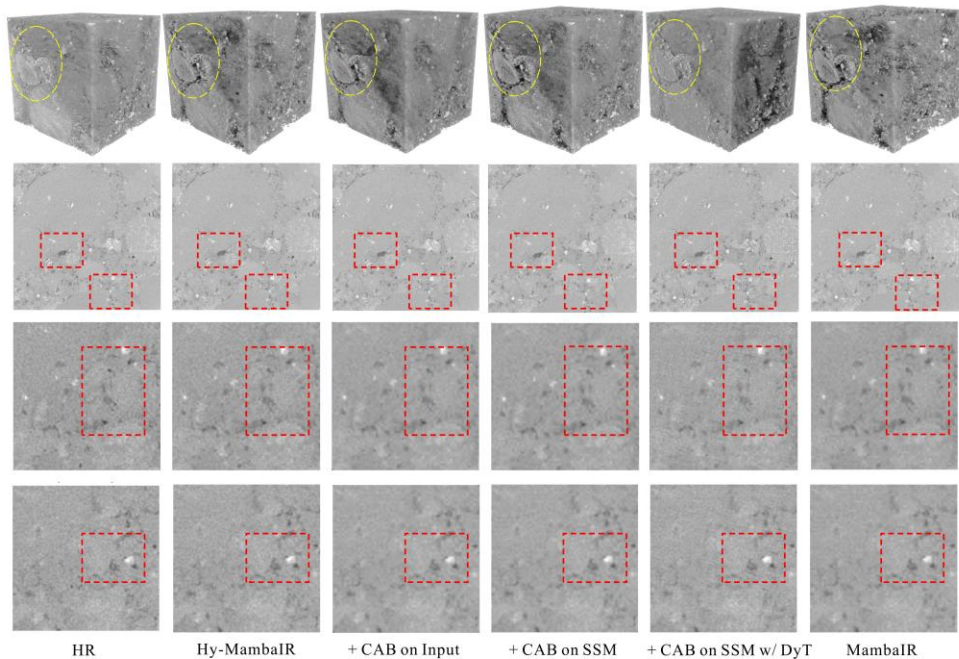


Fig. 9. Multi-dimensional visual comparison of ablation study variants.

4.3. Evaluation of physical accuracy

The ultimate value of image reconstruction lies in its capability to provide a reliable data foundation for downstream scientific analysis and engineering applications. It should be emphasized that the benefit of Hy-MambaIR is not limited to visual enhancement. In this study, the gains in

image-level metrics are accompanied by consistent improvements in segmentation fidelity, pore-network topology, connectivity-related descriptors, and permeability prediction, indicating that the reconstruction improvement is physically meaningful rather than merely cosmetic. To this end, we constructed a comprehensive physical accuracy evaluation framework that combines topological structure analysis with fluid dynamics simulation, aiming to rigorously verify the reliability of the reconstruction results under realistic geological conditions.

4.3.1. Validation of segmentation accuracy

When computing physical parameters such as porosity and permeability, the sharpness of the pore-framework boundaries in the reconstructed images is a critical prerequisite for accurate binary segmentation. If the super-resolved (SR) images cannot be reliably segmented, the credibility of any subsequent physical parameters derived from them will be significantly compromised. Therefore, we propose using segmentation accuracy—quantified by the Intersection-over-Union (IoU) and Dice coefficient—as a direct indicator for assessing the physical fidelity of SR models.

To eliminate systematic errors introduced by threshold selection bias, we adopted a unified thresholding strategy. Based on the high-resolution reference images (ground truth, GT), an optimal segmentation threshold (Threshold = 167) was determined using an iterative adaptive algorithm. This threshold was then uniformly applied to binarize all reconstructed images from different models. This unified standard ensures objective comparison and effectively reveals discrepancies in gray-level distributions and rock-physics characteristics among models. Quantitative and qualitative evaluation results are presented in Table 4 and Fig. 10, respectively. Additionally, bar charts comparing segmentation metrics across different methods are provided in Supplementary Fig. S4.

Quantitative analysis. As shown in Table 4, Hy-MambaIR exhibits clear advantages in the segmentation task, achieving an IoU of 0.901 and a Dice coefficient of 0.948. These values are substantially higher than those of classical CNN-based models (e.g., EDSR with an IoU of 0.820) and Transformer-based models (e.g., HAT with an IoU of 0.723). Notably, although HAT performs reasonably well in terms of PSNR, its segmentation accuracy (IoU = 0.723) is unexpectedly low. This suggests that its reconstructions may contain numerous artifacts or blurred boundaries that are visually imperceptible but severely interfere with threshold-based segmentation, leading to misinterpretation of the underlying physical structures. Compared with Bicubic interpolation (IoU = 0.707), Hy-MambaIR improves the IoU by approximately 0.19, indicating that the model

successfully recovers highly usable pore structures from heavily blurred inputs.

Table 4. Segmentation accuracy metrics of different models.

Model	IoU	Dice	Model	IoU	Dice
Bicubic	0.707	0.829	RCAN	<u>0.820</u>	<u>0.901</u>
Hy-MambaIR	0.901	0.948	RRDB	0.711	0.831
EDSR	<u>0.820</u>	<u>0.901</u>	SRCNN	0.714	0.833
HAT	0.723	0.839	SwinIR	0.718	0.836

Qualitative analysis. Fig. 10 presents a visual comparison of pore-network segmentation results (white regions represent pores). In the segmentation maps produced by Bicubic, SRCNN, and RRDB, significant noise persists, causing confusion between matrix impurities and pore signals, thereby altering the topological characteristics of the pore network. By contrast, the segmentation results of Hy-MambaIR are closest to the ground truth. This method effectively removes noise while preserving the connectivity of micro-fractures and throats (see yellow boxes in Fig. 10). Although HAT and SwinIR provide some background noise suppression, they fail to accurately reconstruct fine throats, resulting in disconnected pore pathways, which in turn introduces bias into subsequent permeability calculations.

The significant advantages of Hy-MambaIR in terms of IoU and Dice demonstrate that its reconstructed images possess high structural fidelity and sharp boundary features. This not only confirms the model's superior image reconstruction capability but also ensures the geometric fidelity of the digital rock, thereby providing a reliable data foundation for subsequent physical parameter computation and geological analysis.

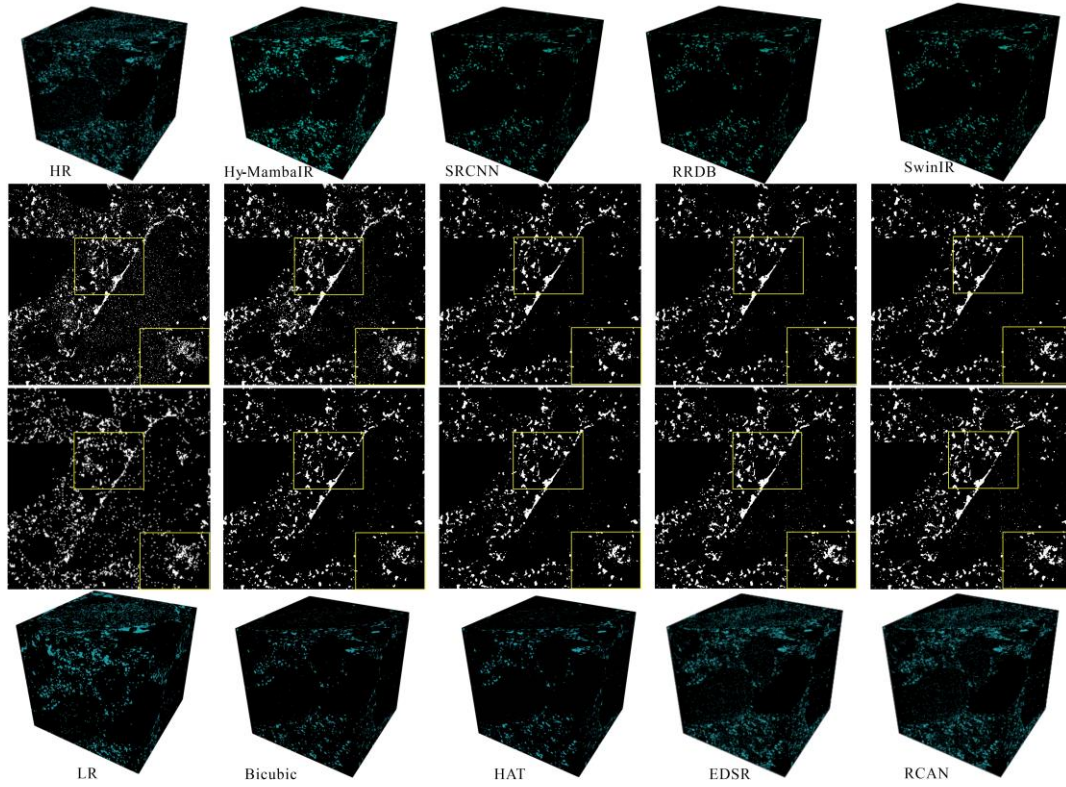


Fig. 10. Comparison of segmentation results on 855-pixel slices.

4.3.2. Validation of rock-physics properties and fluid dynamics

To evaluate the physical fidelity of the reconstructed digital rocks and their suitability for studying flow mechanisms, we performed quantitative analyses of the pore-network topology and validated single-phase fluid-dynamics behavior.

Pore structure and connectivity analysis. The geometrical and topological characteristics of the pore network determine the reservoir's storage capacity and flow performance (transport properties). As summarized in Table 5, Hy-MambaIR achieves the best agreement with the reference images (HR) in terms of key structural parameters. First, in terms of porosity (ϕ), Hy-MambaIR yields a predicted value of 8.05%, which has the smallest relative error compared with the true value of 8.45%. In contrast, methods such as Bicubic (5.07%), HAT (5.78%), and SRCNN (5.89%) exhibit significant underestimation, indicating that these models lose a large number of small pores during reconstruction. Second, for topological connectivity, the Euler characteristic density (χ_v) serves as a key indicator. For the HR images, χ_v is $-13.73 \times 10^{-9} \mu\text{m}^{-3}$, reflecting a highly connected and complex pore network. Among all models, Hy-MambaIR produces the value closest to the ground truth ($\chi_v = -8.444$). In contrast, Bicubic ($\chi_v = -1.869$) and EDSR ($\chi_v = -3.970$) yield much smaller absolute values, indicating poorer pore-network connectivity with more isolated pores and broken

throats. Finally, regarding multiscale spatial correlation, Fig. 11 shows the two-point correlation function $S_2(r)$ and the coordination-number distribution. The $S_2(r)$ curve of Hy-MambaIR (light-cyan dashed line) closely matches the HR curve (dark-blue solid line) over the entire range of length scales, confirming that the model effectively preserves the spatial distribution characteristics of the pores. In terms of coordination-number distribution, Hy-MambaIR accurately reproduces the multi-peak behavior observed in the real rock sample, further demonstrating its superiority in reconstructing pore-connectivity relationships. These consistent multiscale structural indicators provide a solid basis for reliable single-phase flow simulations and subsequent analysis of flow mechanisms.

Table 5. Comparison of physical accuracy metrics and flow properties for different models.

Model	Porosity (%)	χ_v ($\times 10^{-9} \mu\text{m}^{-3}$)	Permeability (mD)			MRE(k) (%)	Tortuosity (Mean)	Path Count
			k_x	k_y	k_z			
HR	8.45	-13.73	18.8	6.64	9.93	0	1.32	488
Bicubic	5.07	-1.869	6.52	3.94	1.15	64.80	1.77	107
Hy-MambaIR	8.05	-8.444	18.1	6.22	5.4	18.56	1.29	246
EDSR	6.60	-3.970	17.5	4.6	1.74	40.04	1.41	140
HAT	5.78	-2.171	69.4	6.57	2.05	116.52	1.67	91
RCAN	6.60	-3.990	16	36.8	1.47	184.77	1.40	141
RRDB	5.66	-2.038	53.8	22.8	1.41	171.78	1.89	79
SRCNN	5.89	-2.445	134	8.71	1.67	242.37	1.59	95
SwinIR	5.72	-2.110	64.4	5.86	2.29	110.41	1.71	81

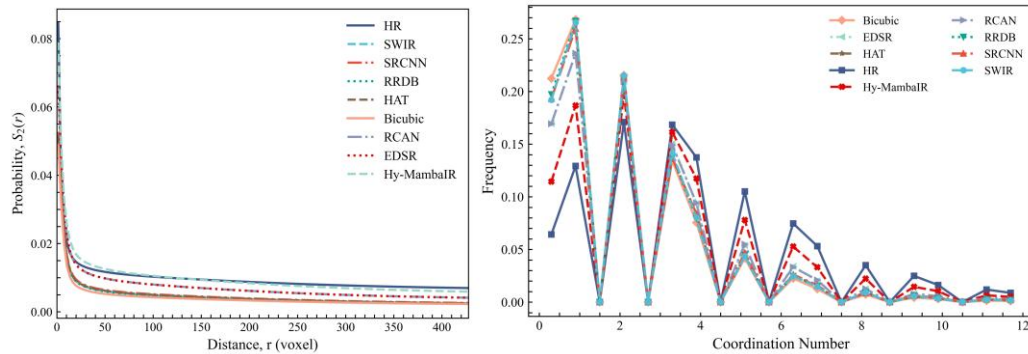


Fig. 11. Two-point correlation function $S_2(r)$ and coordination-number distribution for real and reconstructed samples.

Pore and throat size distributions. The pore-throat size distribution is a critical parameter linking microscopic pore structure to macroscopic physical properties. Fig. 12 presents the reconstructed pore-radius and throat-radius distributions obtained by different models. The real conglomerate sample exhibits a characteristic broad-peak distribution. For the Bicubic interpolation method (orange line), the distribution shifts toward smaller radii. This shift is attributed to the

smoothing effect of interpolation, which blurs the boundaries of large pores, causing them to be misclassified as smaller ones and leading to the loss of critical large-pore features. In contrast, Hy-MambaIR (red line) accurately captures the dominant large-pore peak in the 100–200 μm range, and its distribution in the sub-50 μm throat-radius region also agrees closely with the HR reference. Such accurate reconstruction of pore–throat characteristics across the entire scale range provides a solid physical basis for high-precision permeability prediction. In addition to pore size, we analyzed the distributions of pore volume, specific surface area, and throat length. Supplementary Figs. S6–S8 demonstrate that Hy-MambaIR yields results that are highly consistent with the ground truth across all of these microscopic geometric parameters, and that this consistency is maintained even when different pore-network extraction algorithms (e.g., OpenPNM) are employed.

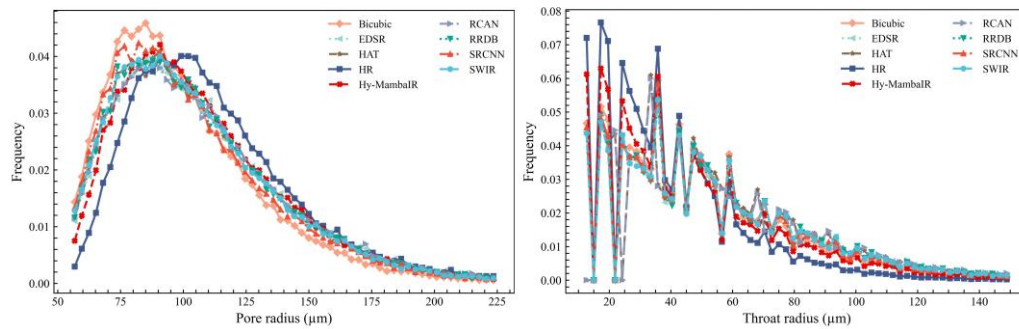


Fig. 12. Distributions of pore radius and throat radius for the real and reconstructed samples.

To definitively assess the flow capacity of the reconstructed rock cores, we analyzed the geometrical characteristics of flow paths and performed single-phase flow simulations. The simulations were conducted on digital rock models of 342^3 voxels, using a unified segmentation threshold of 167. The fluid viscosity was set to 0.01 Pa·s, with inlet and outlet pressures of 33.05 MPa and 0.1 MPa, respectively, and the fluid was constrained to flow along the Z-axis. Based on these settings, we examined both the geometric properties of the flow paths and the spatial distribution of the simulated flow field.

Tortuosity (τ) and the number of effective flow paths (path count) are pivotal parameters governing fluid-transport efficiency. As reported in Table 5, the mean tortuosity of the ground-truth (HR) sample is 1.32. Hy-MambaIR yields a value of 1.29, very close to the HR reference, and its tortuosity distribution in Fig. 13 also closely matches that of HR. In contrast, RRDB (1.89) and Bicubic (1.77) exhibit significantly higher tortuosity, indicating that their reconstructed pore networks contain non-physical tortuous channels that hinder efficient transport. The path count

effectively reflects the recovery of pore-network connectivity. While the HR reference contains 488 effective flow paths, most competing models (e.g., HAT, SwinIR, RRDB) recover fewer than 100, implying that many critical throats remain closed. Hy-MambaIR, however, recovers 246 effective paths—the highest among all compared methods (approximately 2.5 times that of the other methods). This indicates that Hy-MambaIR not only maintains low tortuosity but also increases the number of effective pathways, thereby more faithfully restoring the rock’s intrinsic flow capacity.

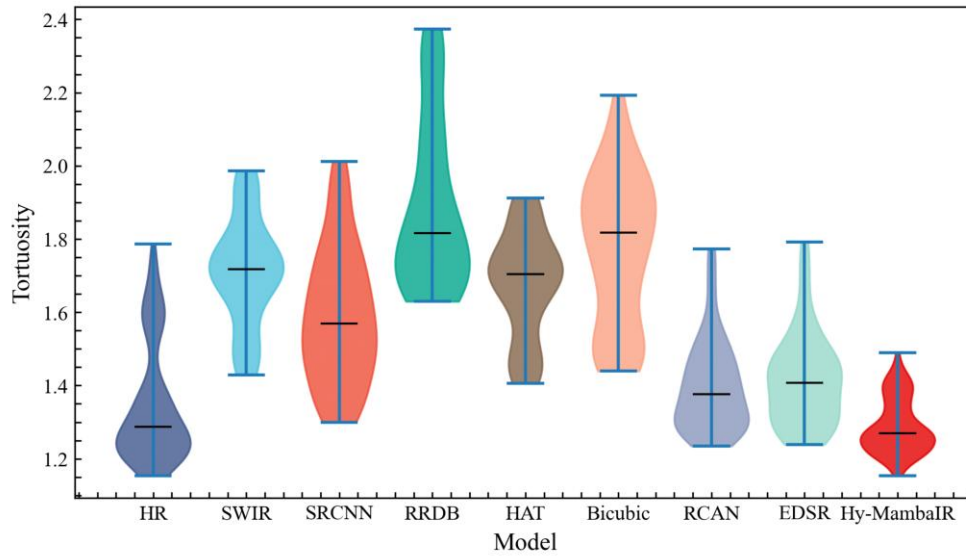


Fig. 13. Tortuosity distributions for the real and reconstructed samples.

Flow visualization. Based on these geometrical characteristics, Fig. 14 illustrates the three-dimensional streamline patterns reconstructed by different methods. The streamline density and spatial topology produced by Hy-MambaIR are highly consistent with the HR reference, accurately reconstructing the dominant preferential flow channels. The 342^3 volume was used for the three-dimensional flow visualization in Fig. 14. Quantitatively, the permeability predicted by Hy-MambaIR ($k=3.93 \times 10^{-5}$ D) is close to the HR value ($k=4.04 \times 10^{-5}$ D), indicating close agreement under the 342^3 flow-visualization setup. In contrast, the streamlines reconstructed by HAT and EDSR are sparse and exhibit poor continuity. Regarding the Bicubic method, although its simulated permeability value ($k = 4.56 \times 10^{-5}$ D) is higher than the HR value, the visualized streamline features suggest that this high permeability is mainly dominated by a few excessively enlarged pores rather than a realistically connected pore network, indicating pronounced computational artifacts and a failure to reflect the true connectivity mechanisms of the pore space.

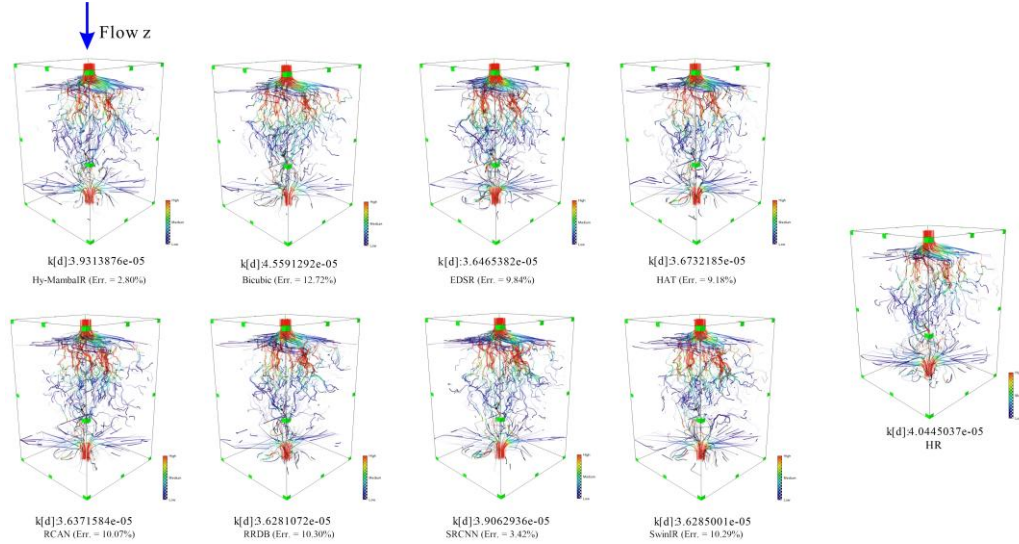


Fig. 14. Single-phase flow visualizations for the 342^3 volume; permeability values are reported in Darcy (D).

To investigate flow behavior at the pore scale, we analyzed velocity fields in a sub-volume of 125^3 voxels (Fig. 15). Detailed quantitative data are provided in Supplementary Table S1. The HR images clearly capture fluid acceleration and stagnation within complex pore–throat structures (deep red regions indicate high velocities). Hy-MambaIR effectively reconstructs both high-velocity preferential channels and low-velocity stagnant regions, with a velocity distribution that closely matches the HR reference. For this 125^3 local sub-volume, the mean relative error of three-directional absolute permeability is 11.19%. In contrast, velocity fields from CNN-based methods (e.g., RCAN, SRCNN) appear overly smoothed, with larger errors (15%–22%), and fail to capture rapid acceleration dynamics in narrow throats. The Bicubic interpolation method, unable to reconstruct connectivity, yields predominantly low-velocity regions with almost no flow.

Macroscopic permeability. To obtain statistically representative macroscopic properties, we performed three-directional absolute-permeability simulations on the full-size REV (855^3 voxels), as reported in Table 5. Hy-MambaIR exhibits superior accuracy, with a mean relative error (MRE) of 18.56%. For a visual comparison of permeability across models on the full-size REV, see Supplementary Fig. S5. All other compared methods show substantial deviations: Bicubic (MRE = 64.80%) and EDSR (MRE = 40.04%) significantly underestimate permeability, which, combined with their low path counts (Path Count < 150), confirms that they fail to reconstruct throat structures and result in blocked channels. In contrast, SRCNN (MRE = 242.37%), RCAN (MRE = 184.77%), and HAT (MRE = 116.52%) drastically overestimate permeability. This suggests that, while these

models improve visual sharpness, they introduce non-physical artifacts during reconstruction, erroneously connecting pores that should remain isolated and creating spurious preferential flow channels.

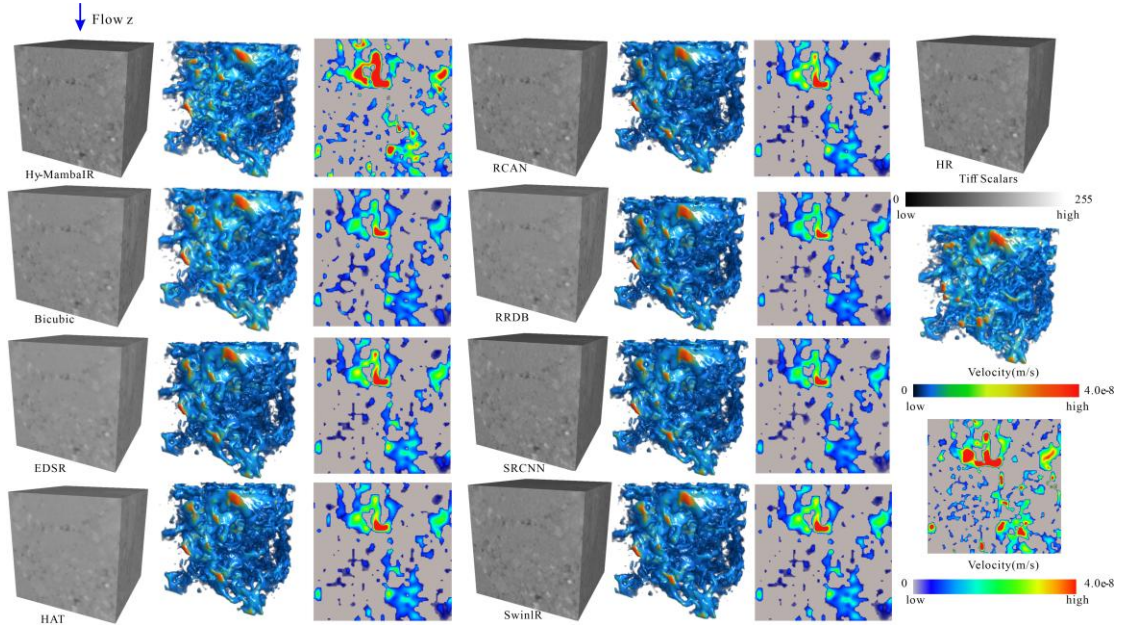


Fig. 15. Comparison of three-dimensional velocity fields and cross-sectional views for the real and reconstructed samples.

4.4. Evaluation of generalization capability

To evaluate the generalization and transferability of Hy-MambaIR under complex geological conditions, we conducted transfer-learning experiments. The model pretrained on the conglomerate dataset was transferred to pebbly sandstone samples with markedly different lithological characteristics. We first assessed its direct inference capability on the unseen lithofacies in a zero-shot setting and then performed supervised fine-tuning with paired LR/HR images from the target domain. The corresponding quantitative and qualitative results are reported in Table 6, Table 7, Fig. 16, and Fig. 17.

4.4.1. Image reconstruction

Table 6 quantitatively summarizes the changes in image-quality metrics before and after fine-tuning. In the zero-shot condition, even without any prior exposure to pebbly sandstone samples, Hy-MambaIR achieves PSNR and SSIM values of 29.09 dB and 0.7392, respectively, significantly outperforming conventional Bicubic interpolation. This demonstrates that the hybrid Mamba-Transformer architecture can effectively learn generalizable priors of porous-media structures, endowing the model with strong initial generalization capability. After fine-tuning on the target

domain, the performance is further improved: PSNR increases by 1.15 dB to 30.24 dB, while the perceptual error measured by LPIPS is substantially reduced.

Table 6. Generalization performance of Hy-MambaIR on pebbly sandstone (unseen lithofacies).

model	PSNR (dB)	SSIM	LPIPS
Bicubic	28.3051±0.4427	0.7156 ±0.0145	0.4100 ±0.0271
Hy-MambaIR (Zero-Shot)	29.0850±0.8082	0.7392±0.0064	0.2961 ±0.0202
Hy-MambaIR (Fine-tuned)	30.2375±0.4157	0.7592 ±0.0111	0.2903±0.0188

The visual comparisons in Fig. 16 further corroborate these findings. In the zero-shot reconstructions, the overall structures remain intact, but mineral boundaries exhibit noticeable blurring due to lithofacies differences. By contrast, the fine-tuned results present much sharper textural details and are highly consistent with the ground truth (GT), indicating that the model effectively mitigates domain shift between different geological samples.

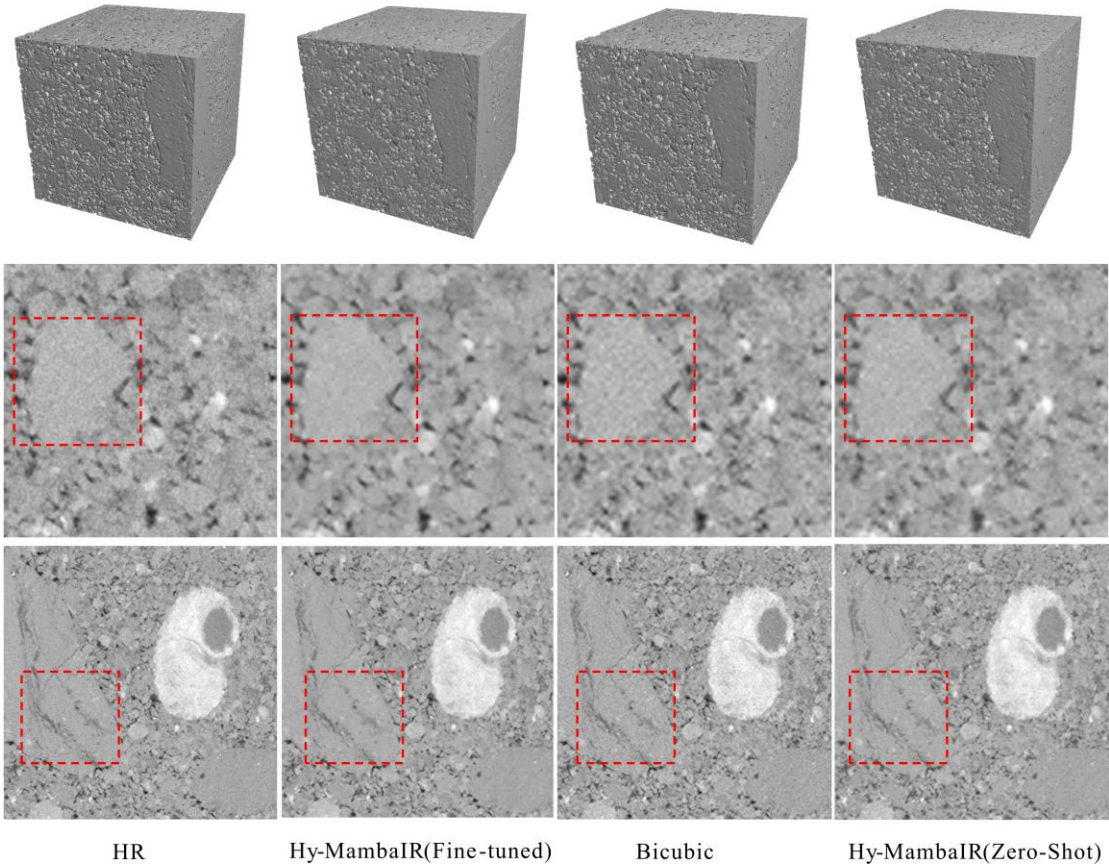


Fig. 16. Visual comparison between real and reconstructed samples on pebbly sandstone.

4.4.2. Validation of physical accuracy

To verify the model's ability to generalize pore-structure representation to previously unseen rocks, we conducted a quantitative evaluation of the physical properties derived from the fine-tuned reconstructions (Table 7). In terms of topological connectivity, Hy-MambaIR (Fine-tuned) produces key structural parameters—such as porosity, specific Euler characteristic density (χ_v), and tortuosity—that are closest to the high-resolution (HR) reference. In particular, for the core connectivity indicator, namely the number of effective flow paths (Path Count), the fine-tuned model recovers 126 paths, significantly outperforming Bicubic (98 paths) and approaching the ground-truth value of 135 paths. These results confirm that the model not only provides accurate morphological restoration but also faithfully reconstructs the complex topological network structures of new samples.

Table 7. Comparison of physical accuracy and performance between real samples and reconstructed samples on pebbly sandstone.

Model	Porosity (%)	χ_v ($\times 10^{-6} \mu\text{m}^{-3}$)	Permeability (D)			MRE(k) (%)	Tortuosity (Mean)	Path Count
			k_x	k_y	k_z			
Bicubic	15.68	$-1.27 \times 10^{-6} \mu\text{m}^{-3}$	3.895	4.143	3.117	7.73	1.42	98
Hy-MambaIR (Zero-Shot)	16.31	$-1.48 \times 10^{-6} \mu\text{m}^{-3}$	3.726	4.077	2.914	4.86	1.41	119
Hy-MambaIR (Fine-tuned)	15.92	$-1.55 \times 10^{-6} \mu\text{m}^{-3}$	3.741	3.954	2.942	3.59	1.38	126
HR	16.06	$-1.73 \times 10^{-6} \mu\text{m}^{-3}$	3.583	3.766	2.983	0	1.35	135

Regarding the fluid-dynamics response, single-phase flow simulations based on the fine-tuned digital rock indicate that the three-directional permeabilities predicted by Hy-MambaIR yield a mean relative error (MRE) of 3.59%. The three-dimensional velocity fields and two-dimensional cross-sectional views shown in Fig. 17 show that the fine-tuned model reproduces the main velocity-distribution patterns inside the pebbly sandstone. Detailed distributions of physical parameters involved in the generalization test, such as throat radius and specific surface area, are provided in Supplementary Figs. S9-S10. Automatic iterative thresholding for reconstruction (Supplementary Fig. S9 and Table S2) leads to substantial deviations of key physical parameters from the HR reference, highlighting the value of fine-tuning for this lithofacies. Together, the structural and flow results support the applicability of the fine-tuning strategy to the evaluated pebbly sandstone and motivate further validation across reservoir conditions and lithofacies.

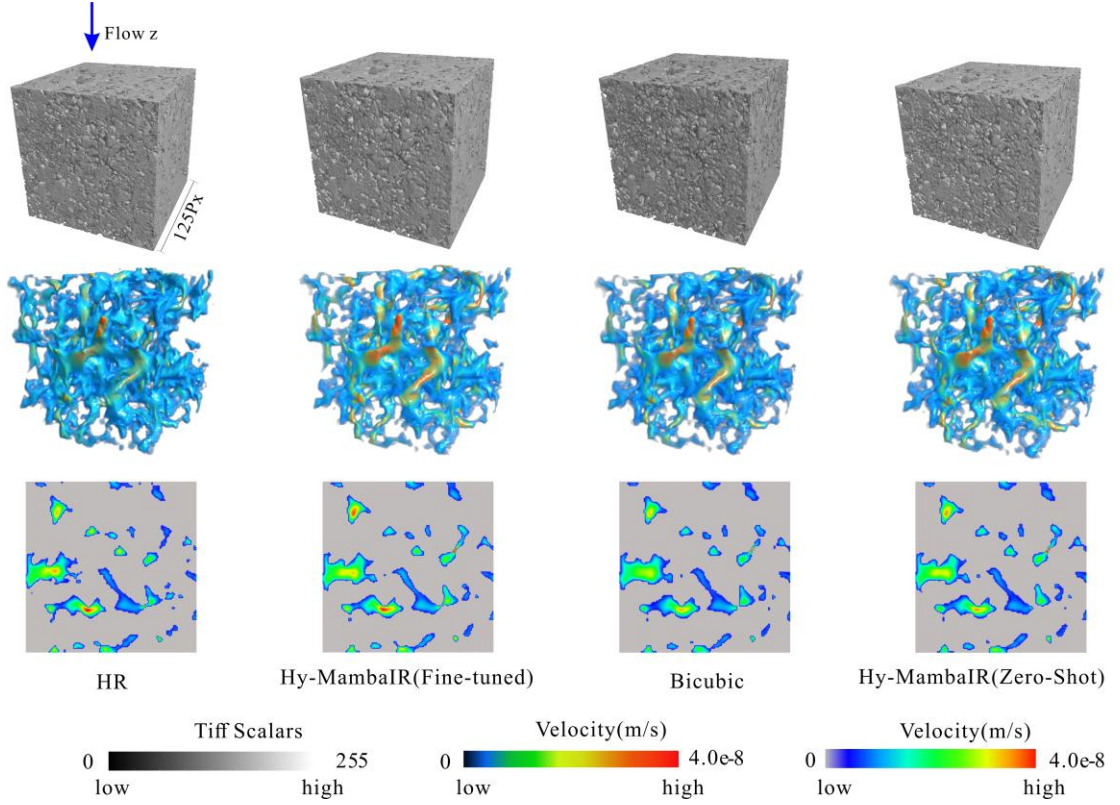


Fig. 17. Comparison of three-dimensional velocity fields and cross-sectional views for real and reconstructed samples on pebbly sandstone.

5. Conclusion

To address the inherent trade-off between field-of-view (FOV) and resolution in digital rock imaging, particularly the challenge of reconstructing cross-scale structures in strongly heterogeneous conglomerate reservoirs, this study proposes a hybrid Mamba–Transformer network, termed Hy-MambaIR. The method adopts a serial “local-to-global” progressive architecture that fuses the fine-scale detail-capturing capability of windowed self-attention with the macroscopic structural modeling capacity of Mamba. The main conclusions are as follows:

(a) Hy-MambaIR establishes a new benchmark for image reconstruction quality on strongly heterogeneous conglomerate datasets. It achieves a PSNR of 33.36 dB, exceeding advanced Transformer (HAT) and CNN (RCAN) baselines by more than 1.2 dB. In addition, the significant reduction in LPIPS (0.1974) confirms that the method effectively recovers geologically realistic textural details while suppressing artifacts that are commonly observed in conventional approaches.

(b) Physical validation extends the evaluation beyond image-quality metrics. Hy-MambaIR attains an IoU of 0.901 in segmentation, recovers pore-network connectivity and geometric

characteristics, and yields a mean relative error of 18.56% for three-directional absolute permeability. These results support its use for digital-rock flow analysis under the evaluated single-phase conditions.

(c) Ablation experiments show that the SCAB and DyT modules contribute jointly to the reconstruction gains. Hy-MambaIR reaches 30 dB in 600 iterations, compared with 4,000 iterations for HAT, reducing the training burden for the evaluated digital-rock reconstruction workflow.

(d) Supervised fine-tuning with paired low- and high-resolution images demonstrates transferability to pebbly sandstone. The fine-tuned model reduces the mean relative error of three-directional absolute permeability to 3.59% while preserving pore-scale topology and flow descriptors. This result supports the use of the framework for reservoir characterization across lithofacies.

Despite its strong performance, the current implementation of Hy-MambaIR still relies on a “2D-to-3D” slice-reconstruction strategy. Although explicit volumetric continuity is not directly enforced, the proposed framework preserves continuity-related structural characteristics to a considerable extent, as evidenced by the improvements in IoU, Path Count, velocity-field reconstruction, and permeability prediction. Future work will focus on exploring voxel-wise generation architectures based on 3D Mamba, and on incorporating physics-informed neural networks (PINNs) by embedding fluid-dynamics governing equations as regularization terms in the loss function, so as to guide the model toward pore structures with stronger physical consistency and more realistic connectivity. In addition, we will investigate end-to-end optimization that couples the super-resolution model with pore-network extraction algorithms, further enhancing both the automation and the accuracy of physical-parameter prediction.

CRedit authorship contribution statement

Yumin Li: Methodology, Writing - original draft. Yang Zhang: Supervision, Project administration, Funding acquisition. Hao Liu: Investigation. Ji Li: Data curation. Yuwei Zu: Visualization. Haoyu Wu: Writing - review & editing. Jiexin Xu: Validation. Xiaonan Wang: Software.

Declaration of interests

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

Data availability

The raw micro-CT images and other supporting data cannot be deposited in a public repository because they are subject to a confidentiality agreement. Access may be considered upon reasonable request to the corresponding author, subject to the terms of that agreement and any required approvals.

Code availability

The source code used for model training, reconstruction, and analysis is publicly available at <https://github.com/76483393/Hy-MambaIR>.

Funding

This research was co-funded by the National Natural Science Foundation of China (42562021), the Tianshan Talent Training Program (2023TSYCCX0005), and the Key R&D Program of Xinjiang Uygur Autonomous Region (2024B01017-2).

References

- Andrä, H., Combaret, N., Dvorkin, J., Glatt, E., Han, J., Kabel, M., Keehm, Y., Krzikalla, F., Lee, M., Madonna, C., Marsh, M., Mukerji, T., Saenger, E.H., Sain, R., Saxena, N., Ricker, S., Wiegmann, A., Zhan, X., 2013a. Digital rock physics benchmarks—part I: imaging and segmentation. *Comput. Geosci.*, Benchmark problems, datasets and methodologies for the computational geosciences 50, 25–32. <https://doi.org/10.1016/j.cageo.2012.09.005>
- Andrä, H., Combaret, N., Dvorkin, J., Glatt, E., Han, J., Kabel, M., Keehm, Y., Krzikalla, F., Lee, M., Madonna, C., Marsh, M., Mukerji, T., Saenger, E.H., Sain, R., Saxena, N., Ricker, S., Wiegmann, A., Zhan, X., 2013b. Digital rock physics benchmarks—part II: computing effective properties. *Comput. Geosci.*, Benchmark problems, datasets and methodologies for the computational geosciences 50, 33–43. <https://doi.org/10.1016/j.cageo.2012.09.008>
- Ávila, J., Pagalo, J., Espinoza-Andaluz, M., 2022. Evaluation of geometric tortuosity for 3D digitally generated porous media considering the pore size distribution and the a-star algorithm. *Sci. Rep.* 12, 19463. <https://doi.org/10.1038/s41598-022-23643-6>
- Behbahani, H., Azin, R., Osfouri, S., 2023. A comprehensive review of digital rock physics: from tomographic images to pore network modeling. *Chem. Process Des.* 2, 6–31. <https://doi.org/10.22111/cpd.2023.44680.1018>
- Bhadariya, V., Kaur, J., Sapale, P., Rasane, P., Singh, J., 2024. Hydrogen storage in porous media: understanding and mitigating microbial risks for a sustainable future. *Int. J. Hydrogen Energy* 67, 681–693. <https://doi.org/10.1016/j.ijhydene.2024.04.178>

783 Bruns, S., Stipp, S.L.S., Sørensen, H.O., 2017. Statistical representative elementary volumes of
784 porous media determined using greyscale analysis of 3D tomograms. *Adv. Water Resour.* 107, 32–
785 42. <https://doi.org/10.1016/j.advwatres.2017.06.002>

786 Fu, J., Thomas, H.R., Li, C., 2021. Tortuosity of porous media: image analysis and physical
787 simulation. *Earth Sci. Rev.* 212, 103439. <https://doi.org/10.1016/j.earscirev.2020.103439>

788 Gonzalez, J.L., de Faria, E.L., Albuquerque, Marcelo P., Albuquerque, Marcio P., Bom, C.R., Freitas,
789 J.C.C., Correia, M.D., 2020. Simulations of NMR relaxation in a real porous structure: pre-
790 asymptotic behavior to the localization regime. *Appl. Magn. Reson.* 51, 581–595.
791 <https://doi.org/10.1007/s00723-020-01200-6>

792 Guo, H., Guo, Y., Zha, Y., Zhang, Y., Li, W., Dai, T., Xia, S.-T., Li, Y., 2025. MambaIRv2: attentive
793 state space restoration. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and*
794 *Pattern Recognition*, pp. 28124–28133. <https://doi.org/10.1109/CVPR52734.2025.02619>

795 Heinemann, N., Alcalde, J., Miocic, J.M., Hangx, S.J.T., Kallmeyer, J., Ostertag-Henning, C.,
796 Hassanpouryouzband, A., Thaysen, E.M., Strobel, G.J., Schmidt-Hattenberger, C., Edlmann, K.,
797 Wilkinson, M., Bentham, M., Haszeldine, R.S., Carbonell, R., Rudloff, A., 2021. Enabling large-
798 scale hydrogen storage in porous media – the scientific challenges. *Energy Environ. Sci.* 14, 853–
799 864. <https://doi.org/10.1039/D0EE03536J>

800 Khan, M.I., 2024. Application of digital rock physics and machine learning to improve the
801 understanding of fluid flow behavior through porous media. Master's thesis. University of Texas at
802 Tyler, Tyler, TX, USA. Mechanical Engineering Theses, Paper 39. <http://hdl.handle.net/10950/4794>.

803 Khan, M.I., Khanal, A., 2024. Machine learning assisted prediction of porosity and related
804 properties using digital rock images. *ACS Omega* 9, 30205–30223.
805 <https://doi.org/10.1021/acsomega.3c10131>

806 Kim, J., Nang, J., Choe, J., 2024. LMLT: low-to-high multi-level vision transformer for image super-
807 resolution [preprint]. *arXiv*. <https://doi.org/10.48550/arXiv.2409.03516>.

808 Li, C., Medina, M.R., Warren, M., Tasrianto, R., Macy, B., Bowling, J., Roy, B., Sena, A., 2024.
809 Seismic super-resolution method: enhancing reservoir delineation and characterization through
810 high-resolution seismic data. *Leading Edge* 43, 843–851. <https://doi.org/10.1190/tle43120843.1>

811 Li, S., Wang, X., Jiang, L., Song, Y., 2025. Pore-scale observations of CO₂ thermodynamic
812 properties and dissolution trapping in porous media. *J. Hydrol.* 658, 133118.
813 <https://doi.org/10.1016/j.jhydrol.2025.133118>

814 Li, X., Li, F., Huang, Z., Ji, H., Wang, H., Chen, L., Zhang, Z., Shi, X., 2025. Characterization and
815 simulation of conglomerate reservoirs using core data of triassic baikouquan formation, mahu
816 depression. *Sci. Rep.* 15, 29293. <https://doi.org/10.1038/s41598-025-09954-4>

817 Ma, Z., Bijeljic, B., Wen, G., Tang, K., Wang, Y., Blunt, M.J., 2025. Super-resolution imaging of
818 multiphase fluid distributions in porous media using deep learning. *Transp. Porous Media* 152, 85.

819 <https://doi.org/10.1007/s11242-025-02210-3>

820 Malik, S., Makauskas, P., Sharma, R., Pal, M., 2024. Evaluating petrophysical properties using
821 digital rock physics analysis: a CO₂ storage feasibility study of lithuanian reservoirs. *Appl. Sci.* 14,
822 10826. <https://doi.org/10.3390/app142310826>

823 Meng, Z., Zhang, J., Li, X., Zhang, L., 2022. Lightweight image super-resolution based on local
824 interaction of multi-scale features and global fusion. *Mathematics* 10, 1096.
825 <https://doi.org/10.3390/math10071096>

826 Mosser, L., Dubrule, O., Blunt, M.J., 2017. Reconstruction of three-dimensional porous media using
827 generative adversarial neural networks. *Phys. Rev. E* 96, 043309.
828 <https://doi.org/10.1103/PhysRevE.96.043309>

829 Nabipour, I., Raoof, A., Cnudde, V., Aghaei, H., Qajar, J., 2024. A computationally efficient
830 modeling of flow in complex porous media by coupling multiscale digital rock physics and deep
831 learning: improving the tradeoff between resolution and field-of-view. *Adv. Water Resour.* 188,
832 104695. <https://doi.org/10.1016/j.advwatres.2024.104695>

833 Niu, Y., Jackson, S.J., Alqahtani, N., Mostaghimi, P., Armstrong, R.T., 2021. A comparative study
834 of paired versus unpaired deep learning methods for physically enhancing digital rock image
835 resolution [preprint]. *arXiv*. <https://doi.org/10.48550/arXiv.2112.08644>.

836 Pantaleo, G., Roncoroni, G., Pipan, M., 2024. Monitoring elastic parameters changes during
837 underground hydrogen storage using rock physics parametrized FWI. Presented at the Fifth EAGE
838 Global Energy Transition Conference & Exhibition (GET 2024), European Association of
839 Geoscientists & Engineers, pp. 1–5. <https://doi.org/10.3997/2214-4609.202421200>

840 Pramana, A.A., Thaib, U.M., 2022. Digital core analysis for absolute permeability of shale. *AIP*
841 *Conf. Proc.* 2708, 050001. <https://doi.org/10.1063/5.0122815>

842 Qiao, N., Di, L., Sun, C., 2024. Single image super-resolution based on global-local information
843 synergy [preprint]. *arXiv*. <https://doi.org/10.48550/arXiv.2405.01085>.

844 Qin, X., Xia, Y., Wu, J., Sun, C., Zeng, J., Xu, K., Cai, J., 2022. Influence of pore morphology on
845 permeability through digital rock modeling: new insights from the Euler number and shape factor.
846 *Energy Fuels* 36(14), 7519–7530. <https://doi.org/10.1021/acs.energyfuels.2c01359>.

847 Saxena, N., Hows, A., Hofmann, R., Alpak, F.O., Dietderich, J., Appel, M., Freeman, J., De Jong,
848 H., 2019. Rock properties from micro-CT images: digital rock transforms for resolution, pore
849 volume, and field of view. *Adv. Water Resour.* 134, 103419.
850 <https://doi.org/10.1016/j.advwatres.2019.103419>

851 Shan, L., Liu, C., Liu, Y., Kong, W., Hei, X., 2022. Rock CT image super-resolution using residual
852 dual-channel attention generative adversarial network. *Energies* 15, 5115.
853 <https://doi.org/10.3390/en15145115>

854 Sun, L., Dong, J., Tang, J., Pan, J., 2023. Spatially-adaptive feature modulation for efficient image
855 super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision,
856 pp. 13190–13199. <https://doi.org/10.1109/ICCV51070.2023.01213>.

857 Suo, L., Wang, Z., Liu, H., Cui, L., Sun, X., Qin, X., 2024. Innovative deep learning approaches for
858 high-precision segmentation and characterization of sandstone pore structures in reservoirs. *Appl.*
859 *Sci.* 14, 7178. <https://doi.org/10.3390/app14167178>

860 Wang, H., Zhang, M., Xia, X., Tian, Z., Qin, X., Cai, J., 2024. Lattice boltzmann prediction of CO₂
861 and CH₄ competitive adsorption in shale porous media accelerated by machine learning for CO₂
862 sequestration and enhanced CH₄ recovery. *Appl. Energy* 370, 123638.
863 <https://doi.org/10.1016/j.apenergy.2024.123638>

864 Wang, Y.D., Armstrong, R.T., Mostaghimi, P., 2019. Enhancing resolution of digital rock images
865 with super resolution convolutional neural networks. *J. Pet. Sci. Eng.* 182, 106261.
866 <https://doi.org/10.1016/j.petrol.2019.106261>

867 Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error
868 visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612.
869 <https://doi.org/10.1109/TIP.2003.819861>

870 Wu, J., Yang, S., Gan, B., Cao, Y., Zhou, W., Kou, G., Wang, Z., Li, Q., Dong, W., Zhao, B., 2021.
871 Pore structure and movable fluid characteristics of typical sedimentary lithofacies in a tight
872 conglomerate reservoir, mahu depression, northwest China. *ACS Omega* 6, 23243–23261.
873 <https://doi.org/10.1021/acsomega.1c02952>

874 Xing, Z., Yao, J., Liu, L., Sun, H., 2024. Efficiently reconstructing high-quality details of 3D digital
875 rocks with super-resolution transformer. *Energy* 300, 131499.
876 <https://doi.org/10.1016/j.energy.2024.131499>

877 Yang, L., Zhang, R.-Y., Li, L., Xie, X., 2021. SimAM: a simple, parameter-free attention module
878 for convolutional neural networks, in: Proceedings of the 38th International Conference on Machine
879 Learning. Presented at the International Conference on Machine Learning, PMLR, pp. 11863–11874.

880 You, S., Liao, Q., Yan, Z., Li, G., Tian, S., Song, X., Wang, H., Xue, L., Lei, G., Liu, X., Patil, S.,
881 2024. Super-resolution reconstruction of 3D digital rocks by deep neural networks. *Geoenergy Sci.*
882 *Eng.* 237, 212781. <https://doi.org/10.1016/j.geoen.2024.212781>

883 Yuan, B., Li, H., Du, Q., 2023. Enhancing identification of digital rock images using super-
884 resolution deep neural network. *Geoenergy Sci. Eng.* 229, 212130.
885 <https://doi.org/10.1016/j.geoen.2023.212130>

886 Zhang, J., Gao, M., Li, W., Fang, D., Li, C., 2025. Visual state space model for image super-
887 resolution. *IEEE Trans. Instrum. Meas.* 74, 1–8. <https://doi.org/10.1109/TIM.2024.3522426>

888 Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O., 2018. The unreasonable effectiveness of
889 deep features as a perceptual metric. In: Proceedings of the IEEE/CVF Conference on Computer

890 Vision and Pattern Recognition, pp. 586–595. <https://doi.org/10.1109/CVPR.2018.00068>.

891 Zhang, T., Ni, M., Guan, Q., Li, D., Zhou, S., Du, Y., 2023. Reconstruction of three-dimensional
892 porous media using multi-scale generative adversarial networks. *J. Appl. Geophys.* 213, 105042.
893 <https://doi.org/10.1016/j.jappgeo.2023.105042>

894 Zhou, B., Du, S., Wei, Y., Zong, Z., Duan, X., Wang, Y., 2025. Heterogeneity and anisotropy of tight
895 conglomerates: mechanisms and implications. *Adv. Geo-Energy Res.* 18, 7–20.
896 <https://doi.org/10.46690/ager.2025.10.02>

897 Zhu, J., Chen, X., He, K., LeCun, Y., Liu, Z., 2025. Transformers without normalization. In:
898 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14901–
899 14911. <https://doi.org/10.1109/CVPR52734.2025.01388>.

900 Zraqou, J., Alrousan, R., Sowan, B., Alkhatib, J., 2025. Utilizing image super-resolution to
901 overcome information bottlenecks in vision transformers. *Int. J. Innov. Res. Sci. Stud.* 8, 3734–3749.
902 <https://doi.org/10.53894/ijirss.v8i3.7383>

903

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementalFile.docx](#)