Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Learning Fast-to-Long Acquisition Mapping for Denoising Micro-CT

Luan Vieira

luan@cos.ufrj.br

Universidade Federal do Rio de Janeiro

Aurea Pereira Martins Neta

Universidade Federal do Rio de Janeiro

Felipe Bevilaqua Foldes Guimarães

Universidade Federal do Rio de Janeiro

Júlio de Castro Vargas Fernandes

Laboratório Nacional de Computação Científica

Carlos Eduardo Menezes dos Anjos

Universidade Federal do Rio de Janeiro

Alyne Duarte Vidal

Universidade Federal do Rio de Janeiro

Ricardo Alencar

Universidade Federal do Rio de Janeiro

Lizianne Carvalho Medeiros

Universidade Federal do Rio de Janeiro

Rodrigo Luna

Universidade Federal do Rio de Janeiro

Rodrigo Surmas

Petrobras (Brazil)

Alexandre Evsukoff

Universidade Federal do Rio de Janeiro

Research Article

Keywords:

Posted Date: September 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10856502/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Learning Fast-to-Long Acquisition Mapping for Denoising Micro-CT

Luan Vieira[1,*], Aurea Pereira Martins Neta[1], Felipe Bevilaqua Foldes Guimarães[1], Júlio de Castro Vargas Fernandes[2], Carlos Eduardo Menezes dos Anjos[1], Rodrigo Luna[1], Alyne Duarte Vidal[1], Lizianne Carvalho Medeiros[1], Ricardo Alencar[1], Rodrigo Surmas[3], and Alexandre Evsukoff[1]

[1]COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil
[2]Laboratório Nacional de Computação Científica (LNCC), Petrópolis, RJ, Brazil
[3]CENPES - Petrobras, Rio de Janeiro, RJ, Brazil
[*]Corresponding author: luan@cos.ufrj.br

ABSTRACT

High-quality X-ray microtomography (micro-CT) is essential for digital rock analysis, but long acquisition times increase laboratory cost and reduce experimental throughput. Fast acquisitions provide a practical alternative, although the resulting images are degraded by stronger noise and reduced structural definition, which can affect downstream pore-scale characterization. In this work, we investigate supervised deep learning as a fast-to-long acquisition mapping for denoising micro-CT images of heterogeneous Brazilian pre-salt carbonate rocks. A dataset of 12 plug samples was acquired at three scan times: 2 min 35 s, 8 min, and 60 min, and the 60 min scan was used as the high-quality target for model training. Since the samples were not repositioned between scans, the volumes were naturally registered, enabling direct voxel-wise pairing between the 2 min 35 s fast inputs and the 60 min reference targets. Restormer and RED-CNN models were trained using a leave-one-out strategy at the plug level and evaluated over full 3D cropped volumes using PSNR and SSIM metrics. The trained Restormer models substantially improved image quality relative to the 2 min 35 s acquisition and outperformed the 8 min intermediate acquisition on average. A contrast-adjusted evaluation showed that acquisition-dependent gray-value shifts strongly affect metrics, but the trained Restormer models remained comparable to the 8 min scan in most samples, despite using less than one-third of its acquisition time. Qualitative and pore-structure analyses showed that the models reduce the noise pattern of fast acquisitions and improve pore and solid texture visibility, but small details absent from the fast input cannot be reliably recovered. These results indicate that learning a fast-to-long acquisition mapping can reduce the trade-off between scan time and image quality in carbonate micro-CT imaging. However, they also highlight limitations near the image resolution limit and under acquisition-dependent contrast variability.

1 Introduction

Reservoir characterization is essential for understanding the structural and transport properties of porous rocks, supporting the prediction of fluid flow behavior [1]. In this context, digital rock analysis has emerged as a powerful framework for quantifying rock properties directly from X-ray microtomography (micro-CT) [2]. This imaging technique enables non-destructive visualization of the internal structure of rock samples [3, 4] and provides the basis for both numerical simulations [5] and data-driven methods for estimating petrophysical and flow properties [6, 7]. However, achieving high-fidelity micro-CT reconstructions requires long exposure and scanning times. These extended acquisitions lead to higher laboratory costs and lower experimental throughput [8]. To mitigate these costs, scan protocols can be accelerated by adjusting acquisition parameters [9], including the X-ray dose, exposure time per projection, number of projection views, frame averaging and resolution. Nevertheless, images with shorter acquisition routines are typically noisier, which can compromise downstream tasks such as segmentation, rock characterization, and numerical flow modeling [10]. This introduces a challenging trade-off between imaging efficiency and image quality.

Brazilian pre-salt carbonates provide a particularly challenging case for image-based reservoir characterization. These rocks are marked by strong mineralogical, textural, and pore-scale heterogeneity due to complex depositional and diagenetic processes such as dissolution, dolomitization, silicification, and cementation [11]. Their highly complex pore systems make it difficult to define petrophysical properties such as permeability and to interpret flow behavior [12]. As a result, even subtle losses in image quality can obscure fine details, increasing uncertainty in digital rock characterization.

Data-driven image denoising methods have become an increasingly common strategy for recovering high-quality images from degraded observations [13, 14]. Such degradations may arise from noise, blur, compression artifacts, limited resolution,

and are also commonly encountered in tomographic imaging. Learning-based models are often trained to suppress artifacts while preserving meaningful structural content. In medical computed tomography (CT), this strategy has been extensively explored for denoising low-dose CT images toward high-dose image quality, reducing noise while preserving diagnostically relevant structures [15, 16]. Similar image-to-image learning strategies have been extended to micro-CT and digital rock workflows, although paired training data are often constructed using synthetically degraded images [17] or filter-enhanced reference targets [18] rather than independently acquired high-quality scans.

To evaluate a denoising strategy for overcoming the trade-off between acquisition efficiency and image quality, we use a dataset of 12 heterogeneous Brazilian pre-salt carbonate plug samples scanned under multiple acquisition times. Each plug was imaged using a fast 2-minute-and-35-second scan, an intermediate 8-minute scan, and a long 60-minute scan used as the high-quality reference. The 60-minute acquisition was used as the reference because its longer exposure results in a higher number of detected photons, lower noise, and the highest image quality among the evaluated protocols. Samples were kept fixed inside the scanner chamber during the multi-time acquisitions, with no repositioning and fixed spatial resolution for each plug. This ensured naturally registered volumes and direct pairing between fast scans and long-acquisition references.

Using these paired acquisitions, we formulate the problem as supervised image denoising from fast to long micro-CT scans [19]. The learning task is to estimate the reference image from the corresponding fast input while suppressing acquisition-induced noise and preserving structural features relevant to digital rock analysis. We benchmark both attention-based [20] and convolutional deep learning [21] architectures for fast-to-long micro-CT denoising. All models are trained and evaluated using a leave-one-out strategy [22] across the 12 plug samples.

Using these paired acquisitions, we formulate the problem as supervised fast-to-long micro-CT denoising [19]. In this terminology, denoising refers to suppressing the acquisition-induced noise pattern in the fast scans while improving their agreement with a long-acquisition reference. However, the 60 min target is not assumed to be a noise-free ground truth of the rock microstructure. It is a high-quality reference acquisition with protocol-dependent contrast. The learning task is to estimate the reference image from the corresponding fast input while suppressing acquisition-induced noise and preserving structural features relevant to digital rock analysis. We benchmark both attention-based [20] and convolutional deep learning [21] architectures for fast-to-long micro-CT denoising. All models are trained and evaluated using a leave-one-out strategy [22] across the 12 plug samples.

The remainder of this paper is organized as follows. Section 2 describes the micro-CT dataset, acquisition protocols, preprocessing steps, segmentation procedure, model architectures, training strategy, and evaluation metrics. Section 3 presents quantitative and qualitative comparisons between the fast acquisition, intermediate acquisition, and denoised model outputs, including an additional contrast-adjusted evaluation and a pore-structure analysis. Finally, Section 4 summarizes the main findings, discusses the limitations, and outlines future directions for the proposed approach.

2 Materials and Methods

This section describes the experimental data, image processing procedures, model configurations, and evaluation strategy used to assess fast-to-long micro-CT image denoising. The workflow begins with the acquisition and cropping of paired carbonate rock volumes obtained under fast, intermediate, and reference scan protocols. A diagnostic contrast-adjusted evaluation is then introduced to separate volume gray-value mismatch from structural denoising performance. Binary pore–solid segmentations are used to further investigate possible limitations. Different neural network architectures are evaluated under a leave-one-sample-out validation strategy. Finally, full-volume metrics for quantifying denoising quality with respect to the 60 min reference scans are defined.

Data

This study was carried out using 12 cylindrical carbonate plug samples collected from Brazilian pre-salt oilfields (Table 1). The voxel size ranged from 43.48 µm to 53.93 µm. Micro-CT projection data were acquired using a `VTOMEX M System Scanner` manufactured by Baker Hughes and subsequently reconstructed to generate the 3D image volumes. The scans were performed with an X-ray tube voltage of 160 kV and a tube current of 250 µA.

For each sample, the three acquisitions were performed sequentially without repositioning the plug inside the scanner, preserving the same spatial resolution and sample position across protocols. As a result, the reconstructed volumes were naturally registered, allowing direct voxel-wise comparison between acquisition times. For all acquisition protocols, 801 projection images were acquired before reconstruction. The reconstruction was performed using the scanner software, with beam-hardening correction applied during the reconstruction process. The acquisition times and scanner acquisition parameters used for the fast, intermediate, and reference scans are summarized in Table 2.

The material standards visible in Fig. 1 were included as reference materials with distinct X-ray attenuation responses. In laboratory micro-CT, beam hardening arises from the polychromatic nature of the X-ray source, producing nonlinear attenuation and spatial variations in reconstructed gray values. The aluminium, quartz, and teflon rings therefore provide stable reference

Table 1. Voxel size and image dimensions for the original and cropped volumes.

Sample	Voxel Size (µm)	Original Dimensions	Cropped Dimensions
Sample 1	43.48	$2016 \times 1229 \times 1232$	$1492 \times 537 \times 611$
Sample 2	43.48	$2016 \times 1230 \times 1231$	$1477 \times 618 \times 557$
Sample 3	43.25	$2016 \times 1240 \times 1235$	$1380 \times 585 \times 600$
Sample 4	43.48	$2016 \times 1229 \times 1232$	$1380 \times 585 \times 600$
Sample 5	53.93	$2016 \times 993 \times 994$	$1120 \times 420 \times 550$
Sample 6	46.75	$2016 \times 1145 \times 1146$	$1320 \times 530 \times 560$
Sample 7	43.48	$2016 \times 1232 \times 1229$	$1160 \times 585 \times 600$
Sample 8	53.32	$2016 \times 992 \times 995$	$1180 \times 440 \times 530$
Sample 9	53.93	$2016 \times 992 \times 995$	$1320 \times 450 \times 530$
Sample 10	53.93	$2016 \times 995 \times 993$	$1295 \times 460 \times 495$
Sample 11	53.32	$2016 \times 994 \times 992$	$1280 \times 464 \times 492$
Sample 12	53.32	$2016 \times 993 \times 994$	$1060 \times 470 \times 490$

Table 2. Acquisition times and scanner parameters for the three micro-CT scan protocols. All protocols used 801 projections before reconstruction.

Protocol	Acquisition time	Projections	`TimingVal`	`Averaging`	`Skip`
Fast	2 min 35 s	801	50	1	0
Intermediate	8 min	801	100	3	1
Reference	60 min	801	100	40	1

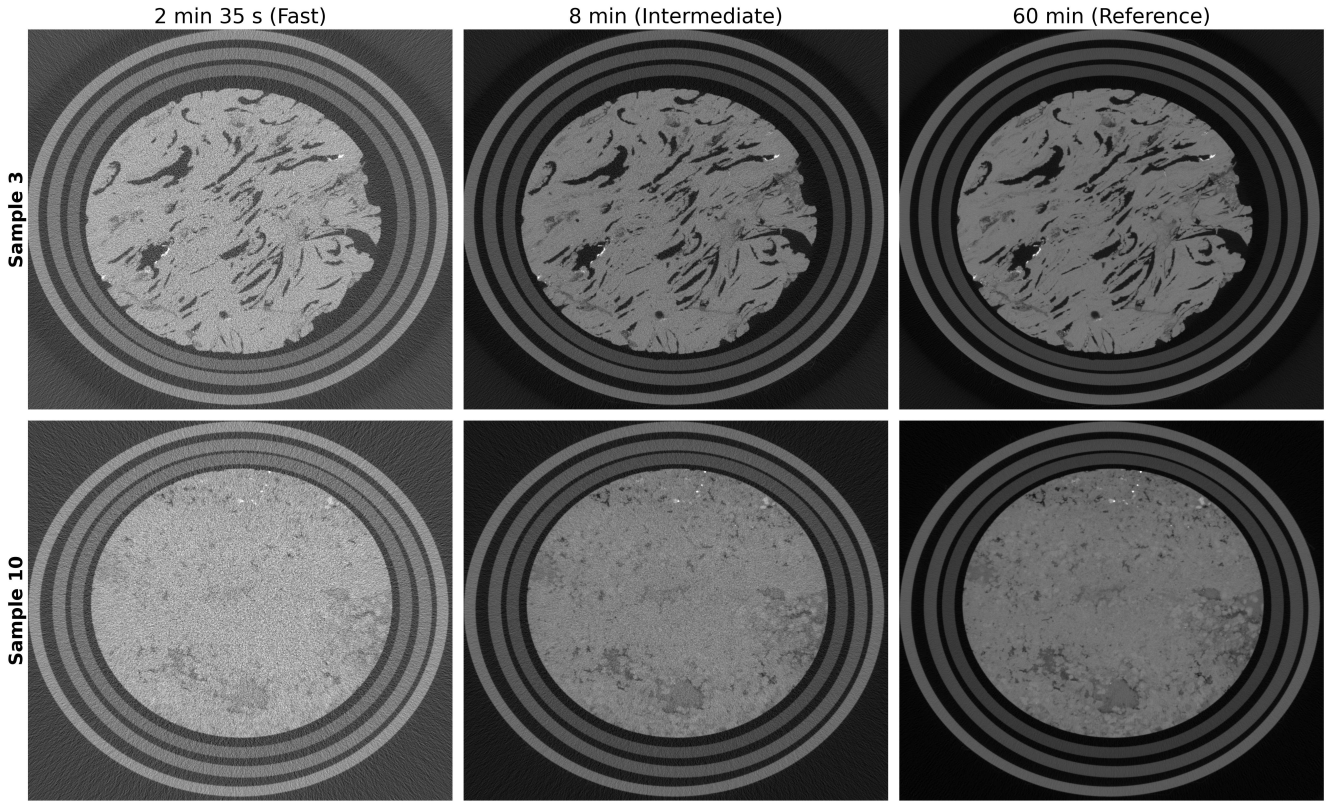


Figure 1. Center axial slices of Samples 3 and 10 acquired with fast, intermediate, and reference scan times: 2 min 35 s, 8 min, and 60 min, respectively. The outer rings correspond to material standards, ordered from outside to inside as aluminium, quartz, and teflon.

regions that can be used to assess gray-value consistency across scans and support beam-hardening evaluation or correction procedures [23, 24].

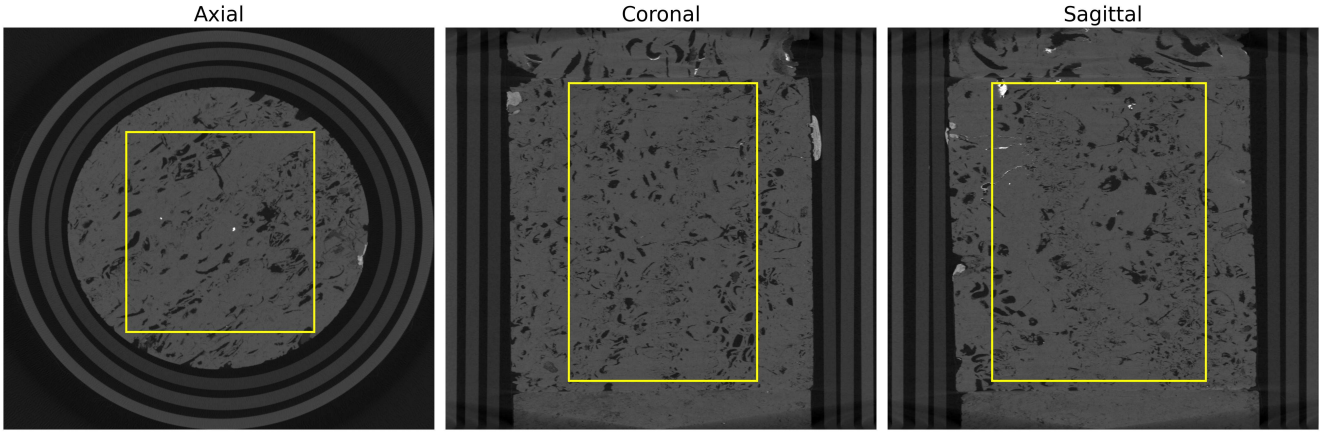


Figure 2. Orthogonal views of Sample 1 from the 60-minute reference acquisition, with the selected crop region highlighted in yellow.

Figure 2 shows the orthogonal views of Sample 1 from the 60-minute reference acquisition, together with the selected crop region indicated by the yellow boxes. The crop corresponds to the dimensions reported in Table 1 and contains only the internal rock volume used in the subsequent analyses and quantitative comparisons. The region was selected to avoid the surrounding calibration rings, external stabilization structures, and boundary voids caused by slight misalignment or plug crookedness. The same procedure was applied to all 12 samples, with each crop kept fixed across all acquisition times.

Figure 3 shows the cropped center axial slices of the 12 carbonate samples for the three acquisition protocols: 2 min 35 s, 8 min, and 60 min. The figure highlights the variability in pore structure, texture, and gray-value distribution across the sample set, while also showing the effect of acquisition time on image quality. The fast scans generally present stronger noise and reduced structural definition, whereas the intermediate and reference acquisitions progressively provide better structural definition. Since the same crop region was used for each acquisition time of a given sample, the figure provides a direct qualitative comparison between scan protocols.

Contrast adjustment

A large contrast variability can be observed across the images shown in Figure 3, even when comparing different acquisition times for the same sample. Although the tube energy was kept fixed during the scans, the different acquisition times required changes in acquisition parameters (Table 2), which may affect the reconstructed gray-value scale. Moreover, even under nominally fixed acquisition conditions, micro-CT gray values are not absolute material constants; images of the same sample may present different intensity ranges due to instrumental variability [25].

This supports the use of intensity-adjusted evaluations as diagnostic analysis steps. To investigate this effect, an additional contrast-adjusted evaluation was performed. For each evaluated 3D cropped volume (X), which may correspond to a baseline acquisition or to the output of a denoising model, a volume-wise linear transformation was estimated with respect to the corresponding 60 min reference volume (Y_{ref}). The intercept β_0 and slope β_1 were obtained by least-squares linear regression over all voxels in the cropped volume, minimizing the sum of squared residuals between the adjusted volume and the reference over all voxels:

$$X_{\text{adj}} = \beta_0 + \beta_1 X \tag{1}$$

where β_0 is the fitted volume offset and β_1 is the fitted volume intensity scaling coefficient. The parameters were estimated using all corresponding voxels of the full 3D crop. The same β_0 and β_1 values were then applied to the entire volume, rather than being estimated slice by slice or locally.

Comparing Figures 3 and 4, the contrast adjustment reduces part of the volume gray-value mismatch between acquisition protocols. However, when the adjusted images are displayed under the same gray-level range, fine structural details are difficult to distinguish. This is particularly evident for Samples 10–12 under the fast acquisition protocol, where the adjusted 2 min 35 s images show compressed contrast and reveal little pore or texture information compared with the corresponding 8 min and 60 min scans.

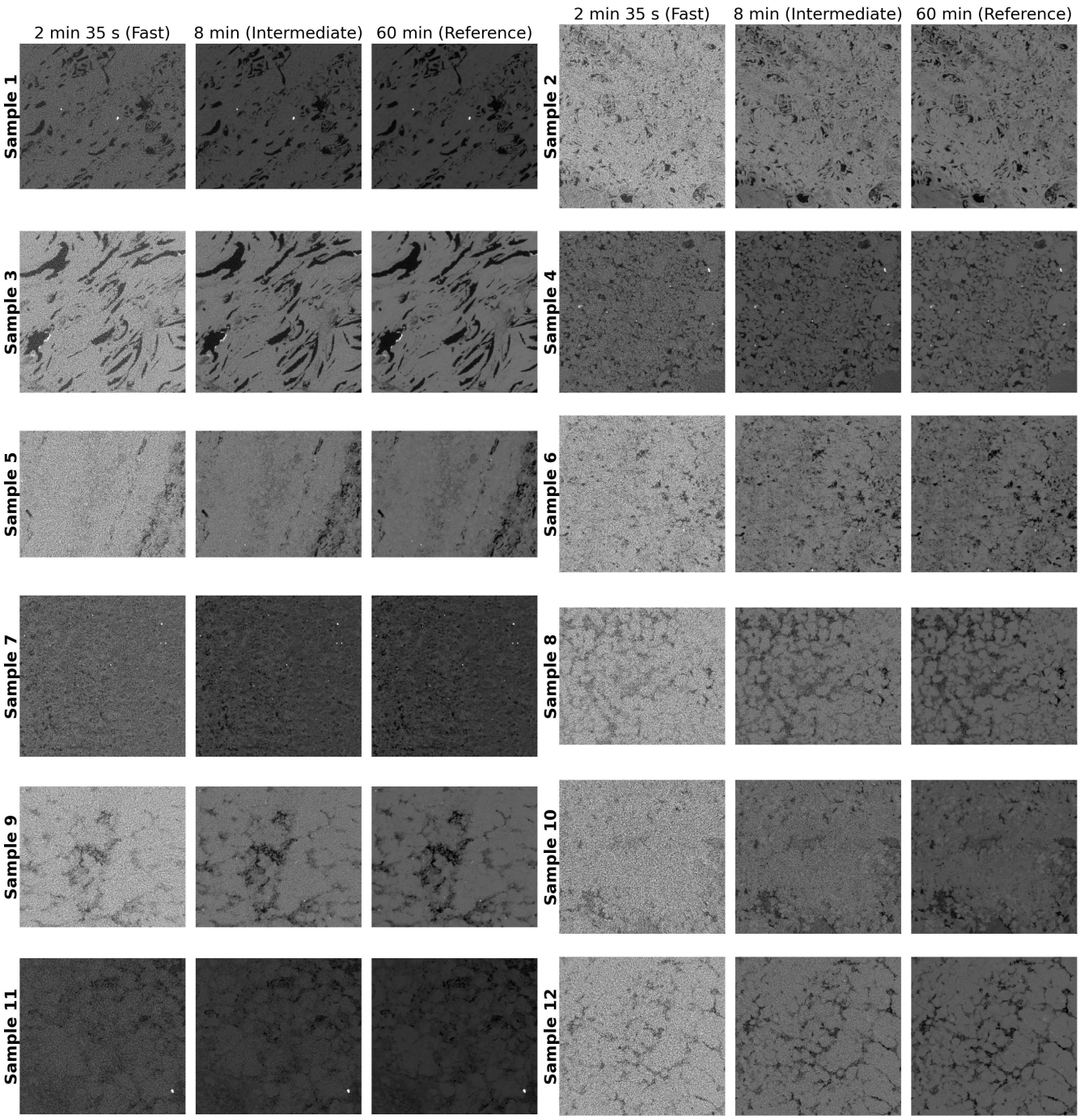


Figure 3. Cropped center axial slices of all 12 samples acquired with fast, intermediate, and reference scan times.

This suggests that the fast acquisitions differ from the reference not only by a volume intensity offset or scale, but also by protocol-dependent gray-value relationships. In medical imaging, such intensity standardization often requires non-linear, piecewise transformations [26, 27]. We chose a milder alternative to histogram matching, reducing volume intensity mismatch without fully reshaping the gray-value distribution, since stronger histogram-based corrections could make PSNR and SSIM metrics less informative by improving intensity agreement independently of actual structural recovery.

Binary phase segmentation

Binary pore–solid masks were generated from the cropped micro-CT volumes using Thermo Scientific PerGeos Software [28]. The segmentation used Otsu's thresholding method [29], followed by a watershed step [30]. Each voxel was assigned to either

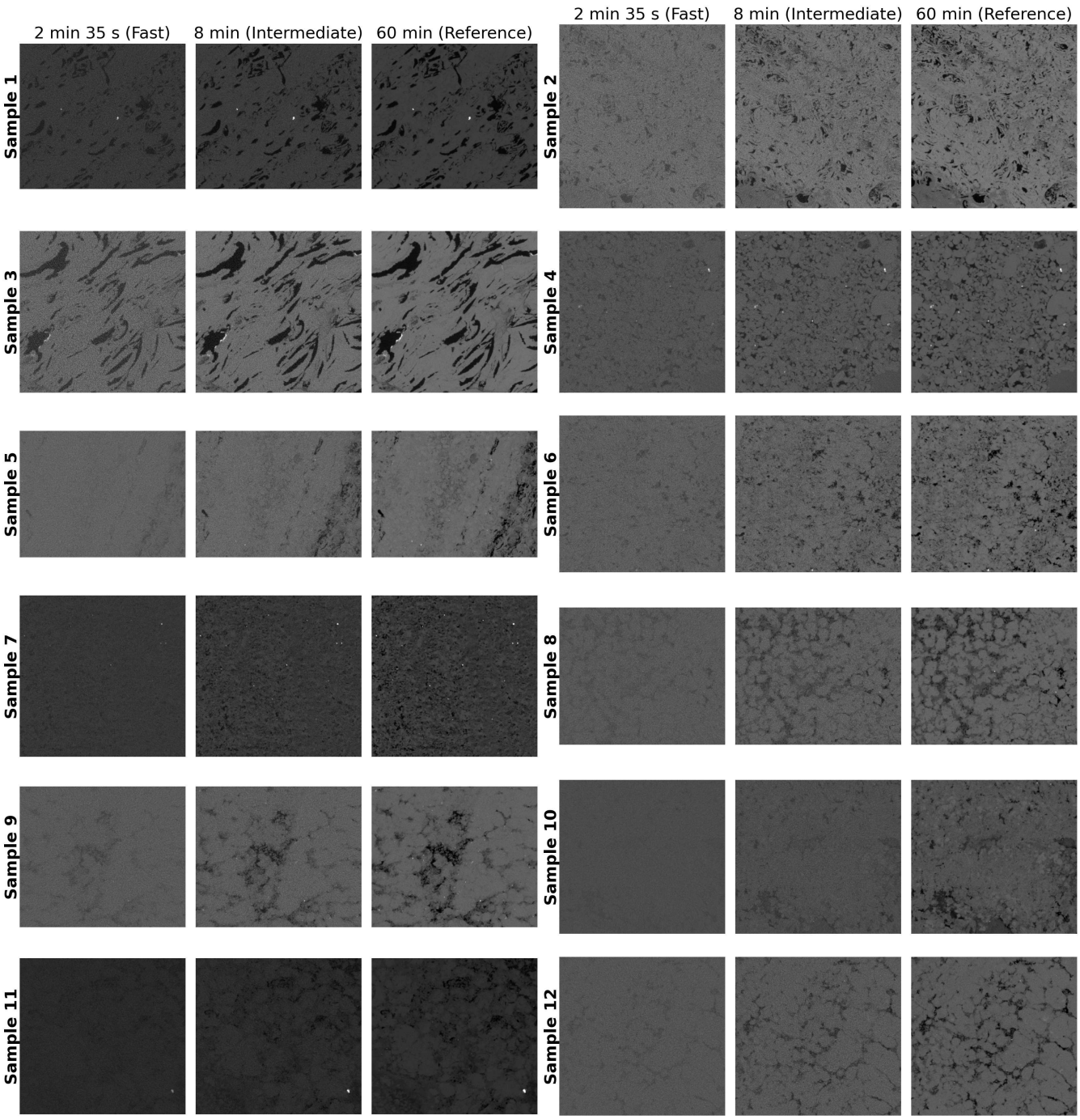


Figure 4. Contrast-adjusted cropped center axial slices of the 12 samples acquired with the fast, intermediate, and reference scan times.

the pore phase or the solid rock phase, as shown in Figure 5. These phase segmentations were used to extract pore-scale attributes for a more detailed evaluation of the denoising results.

Model architectures

Two deep learning architectures were evaluated for fast-to-long micro-CT denoising. The first architecture, Restormer [13], is a transformer-based image architecture designed for high-resolution denoising tasks. Unlike conventional vision transformers that compute self-attention over spatial tokens, which can become computationally expensive for large images, Restormer applies self-attention across channels. This design reduces the computational cost while still allowing the network to model

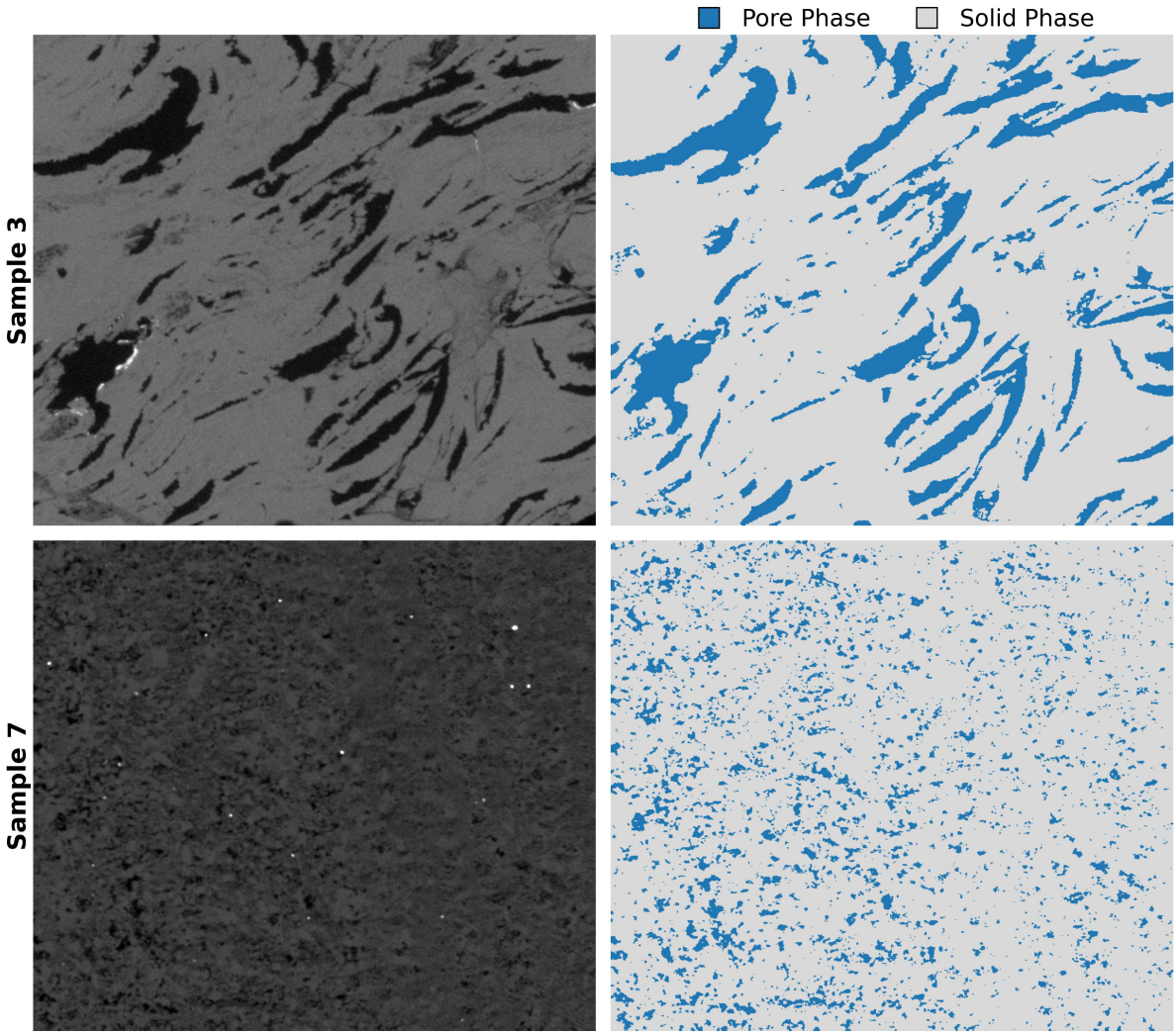


Figure 5. Center axial grayscale micro-CT slices and corresponding binary phase segmentations for Samples 3 and 7. The left column shows the cropped grayscale images, while the right column shows the watershed segmentation, with the pore phase shown in blue and the solid phase shown in light gray.

long-range dependencies and structural information in high-resolution images.

Three Restormer configurations were evaluated to assess the impact of domain transfer. In the first configuration, the model was initialized with weights pre-trained on the Smartphone Image Denoising Dataset (SIDD) [31] and used directly for inference (zero-shot) on the micro-CT data without using target-domain samples for additional training [32]. In the second, the pre-trained model was fine-tuned on the paired micro-CT rock data to adapt its learned representations to the new domain [33]. In the third configuration, Restormer was trained from scratch, with random initialization, using only the micro-CT training data.

The second architecture, RED-CNN, was included as a convolutional baseline model. RED-CNN was originally proposed for low-dose CT denoising and combines an encoder–decoder convolutional structure with residual shortcut connections. In the original study [34], RED-CNN was evaluated on both simulated and clinical low-dose CT data. The method achieved results that were among the best when compared with several established low-dose CT denoising methods.

Training and Validation Strategy

The goal of the denoising task is to learn a mapping function, $f(\cdot)$, that transforms a noisy, fast-acquisition 3D image volume into an estimate of the corresponding long-acquisition reference volume. In this work, the term denoised denotes the output of this supervised fast-to-long mapping. The 60 min scan provides the optimization target because it has higher acquisition quality. In practice, to optimize computational efficiency and memory usage, the volumes are first partitioned into 2D axial slices, from

which smaller sub-regions are extracted for model training. Specifically, given an input 2D crop x fixed at 400×400 pixels with randomly sampled position, extracted from an axial slice of the fast 2 min 35 s acquisition, the model estimates a denoised slice, $\hat{y} = f(x)$. During training, the difference between $\hat{y}$ and the reference target y is minimized by backpropagation [35], updating the model parameters to reduce the reconstruction error between the fast and long acquisitions.

A batch size of 2 and a learning rate of 10^{-4} were used for training. In all cases, the denoising objective was based on the L1 loss. Before training, each fast-acquisition volume was normalized using its own mean and standard deviation, and the corresponding 60 min reference volume was normalized using the same statistics from the paired fast volume. Therefore, no statistics from the long-acquisition target were used during normalization. After inference, the model outputs were transformed back to the original intensity scale using the same fast-volume statistics.

To avoid data leakage and evaluate generalization across rock samples, all splits were performed at the plug level. No slices from the same physical sample were shared between training, validation, and testing. For epoch selection, a nested leave-one-out validation strategy was used. In each fold, one complete plug volume was held out as the independent test sample, one additional complete plug volume was used for validation, and the remaining 10 plug volumes were used for training, as shown in Figure 6.

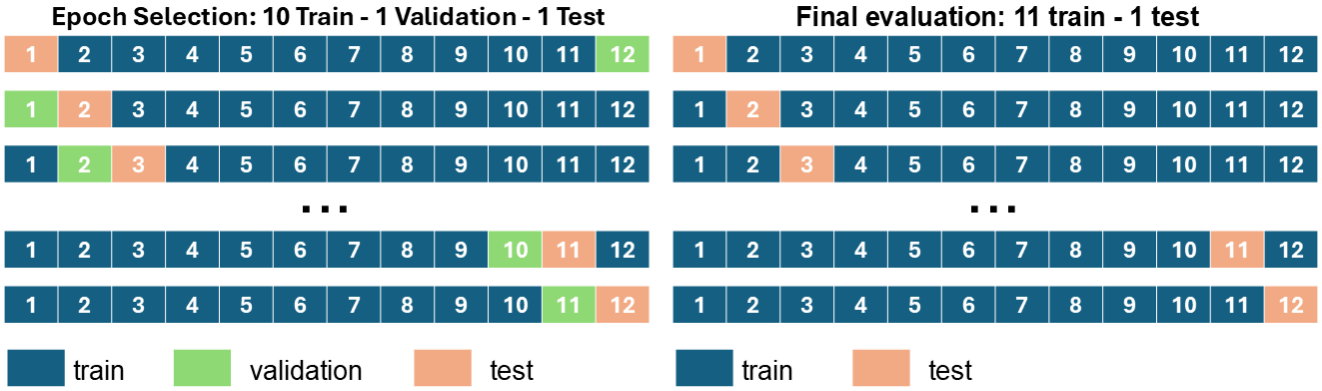


Figure 6. Training and evaluation protocol used for leave-one-out testing. The left panel illustrates the epoch-selection stage, where each fold uses 10 samples for training, one sample for validation, and one sample as the held-out test sample. The validation results across folds are used to select the training epoch. The right panel illustrates the final evaluation stage, where the model is retrained using 11 samples and evaluated on the remaining held-out sample.

After the epoch was selected from the validation analysis, the final leave-one-out experiments were repeated using all 11 non-test samples for training. For each of the 12 folds, one complete plug volume was held out for testing, the model was trained on the remaining 11 samples for the selected number of epochs, and inference was then performed on the full axial slices of the held-out test volume. The resulting restored slices were assembled into the final denoised 3D prediction used for test evaluation.

Evaluation Metrics

The denoised volumes were quantitatively compared with the corresponding 60 min reference volumes using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). Both metrics were computed over the full 3D cropped volumes using the implementations available in the scikit-image library [36]. Since the reconstructed micro-CT volumes were stored as 16-bit images, with gray values ranging from 0 to 65535, the data range was set to 65535 for both PSNR and SSIM computations.

PSNR quantifies how large the reconstruction error is relative to the available image intensity range. Higher PSNR values indicate smaller voxel-wise differences between the evaluated volume and the reference volume. Therefore, this metric is useful for assessing how closely the denoised gray values match the reference acquisition in an absolute intensity sense. Instead of focusing only on voxel-wise intensity differences, SSIM compares local image properties such as luminance, contrast, and structural patterns between the denoised and reference volumes. This is achieved by evaluating local statistics over small image regions, including mean intensity, intensity variance, and covariance between the two images. Values closer to 1 indicate higher structural similarity. Because SSIM emphasizes local structural consistency, it can remain relatively high even when volume intensity differences are present, whereas PSNR is more directly affected by such absolute gray-value shifts [37].

3 Results

This section evaluates the models described in Section 2 as learned mappings from fast to long micro-CT acquisitions. In this formulation, each model learns a function that takes the 2 min 35 s fast acquisition as input and estimates the corresponding 60 min reference volume. The denoised outputs are compared with both the original fast acquisitions and the 8 min intermediate acquisitions to assess whether the learned mapping can recover image quality. Quantitative performance is assessed using PSNR and SSIM computed over the full 3D cropped volumes, while qualitative comparisons are used to examine the structural behavior of the denoised outputs.

Validation results per epoch

Table 3 reports the mean and standard deviation of PSNR and SSIM across validation samples for each model and training epoch. Although individual results show considerable variability across epochs, the overall mean PSNR and SSIM do not improve consistently with additional training. Comparable average values are already obtained at epoch 1, and epochs 1–4 show only modest changes in the aggregate metrics for the Restormer models and RED-CNN. Therefore, epoch 1 was selected.

Table 3. Mean and standard deviation of PSNR and SSIM across validation samples for each model and epoch, computed against the corresponding 60 min reference acquisition. Models were trained using 10 samples in the training split.

			Epoch			
Model	Metric	zero-shot	1	2	3	4
Restormer	PSNR	29.109 ± 3.176	41.037 ± 3.911	40.875 ± 4.463	40.339 ± 4.982	40.801 ± 4.945
pretrained weights	SSIM	0.8925 ± 0.0232	0.9712 ± 0.0104	0.9700 ± 0.0146	0.9693 ± 0.0151	0.9678 ± 0.0184
Restormer	PSNR	–	40.357 ± 3.834	39.830 ± 4.895	39.423 ± 4.078	40.819 ± 3.937
random initialization	SSIM	–	0.9691 ± 0.0114	0.9680 ± 0.0142	0.9690 ± 0.0119	0.9710 ± 0.0103
RED-CNN	PSNR	–	39.882 ± 3.779	39.578 ± 3.553	40.307 ± 4.086	40.088 ± 3.960
random initialization	SSIM	–	0.9631 ± 0.0130	0.9640 ± 0.0113	0.9649 ± 0.0141	0.9640 ± 0.0146

Results without contrast adjustment

All subsequent trained-model results are reported using the one-epoch leave-one-sample-out setting, in which each held-out validation sample was evaluated with a model trained on the remaining 11 samples. As shown in Table 4, the trained Restormer models substantially improved reconstruction quality compared with the 2 min 35 s acquisition and the zero-shot Restormer. In terms of average performance, the trained Restormer variants achieved the best overall results, with mean PSNR values of 41.082 dB for the fine-tuned model and 41.618 dB for the randomly initialized model, and mean SSIM values close to 0.97. Both trained Restormer variants outperformed the 8 min baseline and RED-CNN on average. This trend is also reflected in the sample-wise best values: for PSNR, Restormer random init. achieved the best result in 5 out of 12 samples, followed by Restormer fine-tuned in 4 samples, RED-CNN in 2 samples, and the 8 min baseline in 1 sample.

Table 4. PSNR and SSIM values for each method and sample, computed against the corresponding 60-minute reference acquisition.

		Sample												
Method	Metric	1	2	3	4	5	6	7	8	9	10	11	12	$\mu \pm \sigma$
2 min	PSNR	26.712	28.200	27.284	34.453	28.474	26.940	28.540	27.175	24.555	28.354	29.984	26.172	28.070 ± 2.438
	SSIM	0.691	0.730	0.717	0.777	0.743	0.719	0.762	0.713	0.645	0.756	0.772	0.698	0.727 ± 0.038
8 min	PSNR	38.590	39.562	40.098	**38.621**	39.329	37.450	40.829	37.386	35.481	38.091	38.252	36.700	38.366 ± 1.493
	SSIM	0.927	0.937	0.939	0.945	0.943	0.933	0.951	0.934	0.913	0.945	0.947	0.927	0.937 ± 0.011
Restormer	PSNR	27.597	29.286	28.337	37.987	29.261	27.658	29.266	27.937	25.113	29.043	31.104	26.714	29.109 ± 3.176
zero-shot	SSIM	0.866	0.875	0.872	0.919	0.916	0.881	0.901	0.896	0.860	0.913	0.930	0.882	0.893 ± 0.023
Restormer	PSNR	**42.959**	38.712	**42.579**	32.950	43.766	39.341	41.153	41.049	**40.102**	44.551	**42.689**	43.135	41.082 ± 3.127
fine-tuned	SSIM	**0.972**	0.949	**0.977**	0.959	**0.982**	0.964	0.959	0.977	**0.974**	**0.983**	**0.980**	**0.979**	0.971 ± 0.011
Restormer	PSNR	42.523	**41.511**	41.586	33.246	42.608	41.330	**43.660**	**44.920**	36.566	**45.737**	41.296	**44.439**	41.618 ± 3.537
random init.	SSIM	0.970	**0.956**	0.975	**0.961**	0.980	**0.965**	**0.966**	**0.978**	0.968	**0.983**	0.979	**0.979**	0.972 ± 0.009
RED-CNN	PSNR	40.717	40.957	36.472	31.576	**44.073**	**41.963**	42.371	41.807	36.410	44.374	41.395	38.408	40.044 ± 3.701
	SSIM	0.961	0.950	0.961	0.953	0.977	0.959	0.960	0.973	0.965	0.980	0.976	0.969	0.965 ± 0.010

All values were rounded to three decimal places before identifying the best sample-wise results. Under this rounding criterion, ties occurred only for SSIM, which explains why more than one method can be highlighted for the same sample. Considering the best SSIM values after rounding to three decimal places, including ties, Restormer fine-tuned was among the best results in 7 samples, Restormer random initialization was among the best results in 7 samples, and RED-CNN did not achieve the best SSIM in any sample. When both trained Restormer variants are considered together, Restormer was among the best SSIM results in all 12 samples.

Figure 7 provides a qualitative comparison for Sample 5 before contrast adjustment. A visible contrast difference can be observed between the fast input and the 60 min reference, in addition to the stronger static-like noise pattern present in the fast acquisition. The zero-shot Restormer output does not substantially improve the image toward the reference; instead, it behaves mainly as a noise-suppression filter, reducing part of the high-frequency noise but without adequately recovering the target-domain texture and pore contrast. In contrast, the trained models reduce noise and produce images that are closer to the 60 min reference, especially in the broader pore and rock texture. However, since the input is a noisy 2 min 35 s acquisition, some fine details are not available to the model and therefore cannot be reliably recovered. This qualitative behavior is consistent with Table 4, where the trained models improve PSNR and SSIM relative to the fast input and the zero-shot Restormer for this sample.

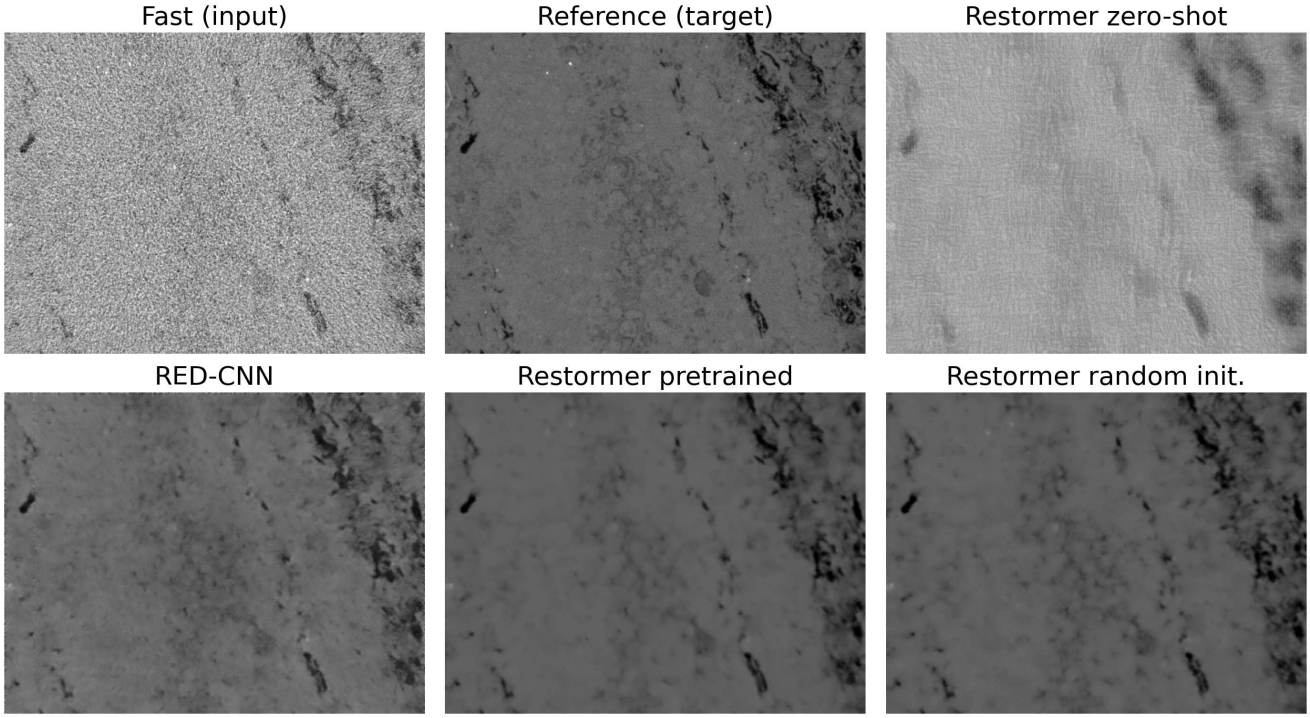


Figure 7. Qualitative comparison of the center axial slice of Sample 5. The panels show the fast input, the 60 min reference target, and the outputs of Restormer zero-shot, RED-CNN, Restormer fine-tuned from pretrained weights, and Restormer trained from random initialization.

Figure 8 illustrates the sample-wise behavior of the metrics across the selected methods. The PSNR curves show a clear improvement from the 2 min acquisition to the 8 min baseline for all samples, confirming the expected benefit of longer acquisition time, while the Restormer zero-shot result produced only a small PSNR increase relative to the 2 min scan and remained below the 8 min baseline in all cases. In contrast, the trained models generally produced the highest PSNR values, with large positive ΔPSNR values relative to the 2 min acquisition and positive gains over the 8 min baseline for most samples. However, the ΔPSNR vs 8 min plot also shows that these gains were not uniform, with the models underperforming the 8 min baseline in a few cases.

The SSIM curves follow a similar overall trend, but the model results show a stronger improvement in SSIM than in PSNR. Even without dataset-specific training, the zero-shot Restormer achieves SSIM values close to those of the 8 min intermediate acquisition. The trained models further increased SSIM, with most values clustered close to 0.97–0.98. The ΔSSIM results show that the trained models outperformed both the 2 min and 8 min baselines for all samples.

However, the improvement was not uniform across all cases. Specifically, Sample 4 represents a failure case in terms of

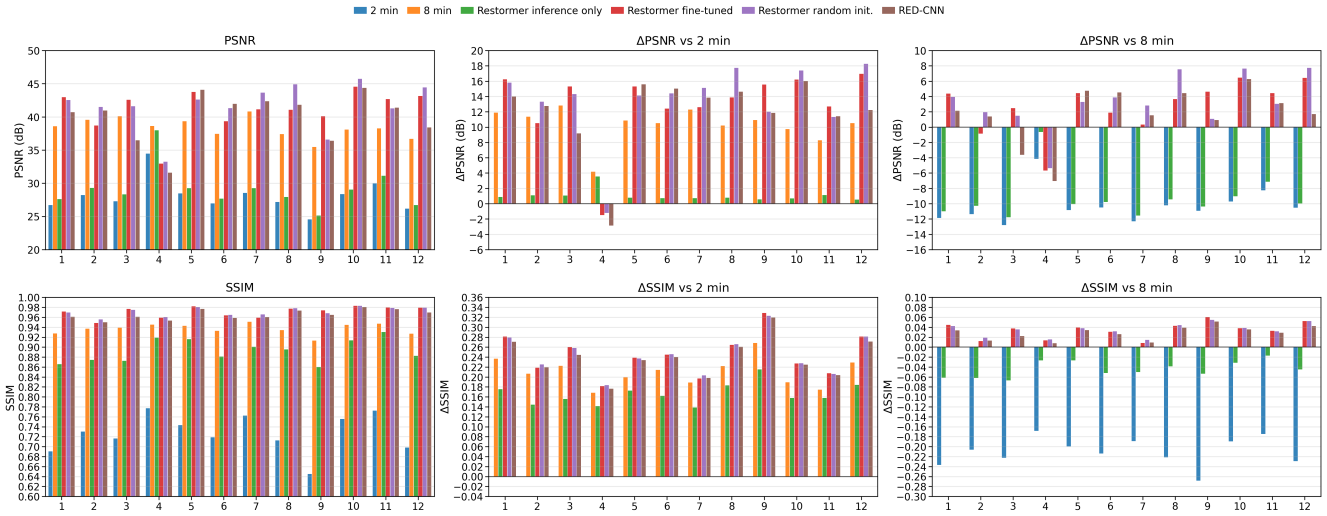


Figure 8. PSNR and SSIM results for each sample across the selected acquisition and denoising methods. The top row shows PSNR and the corresponding PSNR differences relative to the 2 min 35 s and 8 min baselines, while the bottom row shows the same comparison for SSIM. Positive Δ values indicate improvement over the corresponding baseline.

PSNR, since all trained models produced lower PSNR values than the 8 min baseline. This behavior is even more evident in Figure 8, where Sample 4 corresponds to the red curve: for the trained models, the PSNR drops below the 2 min 35 s baseline, despite the corresponding SSIM values remaining relatively high. Figure 9 helps explain this discrepancy visually, as the model output still preserves much of the structural content, but shows a noticeable contrast mismatch relative to the reference. This mismatch is reflected in the Model–Reference residuals, which are predominantly blue, indicating a systematic difference in gray-value scale rather than purely structural errors. Since PSNR is based on absolute voxel-wise differences, it is more sensitive to these shifts, whereas SSIM can remain high when the main structural features are preserved.

The model output achieves high SSIM against the 60 min reference even when PSNR decreases, suggesting that part of the error may be associated with contrast shifts rather than a loss of structural information. This interpretation is also supported by the qualitative comparison in Figure 9, where the restored output preserves much of the structural content but exhibits a noticeable contrast difference relative to the 60 min reference.

Results with contrast adjustment

The resulting PSNR and SSIM values, adjusted with respect to the corresponding 60 min reference volume for all evaluated methods (Equation 1), are reported in Table 5. After contrast adjustment, PSNR values increased substantially for all methods. However, the largest gains were observed for the 2 min 35 s acquisition and the zero-shot Restormer, whose mean PSNR values increased from 28.070 dB to 39.779 dB and from 29.109 dB to 41.060 dB, respectively. Their SSIM values also increased, from 0.727 to 0.933 and from 0.893 to 0.944. This indicates that the initial evaluation was strongly affected by contrast differences, especially for the baseline and untrained cases. Nevertheless, even after this correction reduced the gap between methods, the trained Restormer models remained the best-performing approaches overall, outperforming the 2 min 35 s acquisition across all average metrics.

This clear improvement against the 2 min 35 s baseline differs from the initial evaluation (Table 4), where Sample 4 appeared as an underperforming case, and the trained Restormer models produced lower PSNR values than both the 2 min 35 s and 8 min acquisitions. After contrast adjustment, however, the trained Restormer models achieved the best PSNR values in 10 out of 12 samples and were among the best SSIM results in 11 out of 12 samples. The remaining cases were surpassed by the 8 min acquisition, which is consistent with its longer acquisition time, since it was acquired for more than three times longer than the 2 min 35 s scan used as input to the models.

Figure 10 shows the sample-wise metric curves after contrast adjustment. Compared with the previous curves in Figure 8, the adjusted evaluation shifts all methods to higher PSNR and SSIM values. The gains relative to the 2 min scan become less pronounced. The zero-shot Restormer provides limited improvement and remains below the 8 min acquisition in both metrics for all samples. In contrast, the trained models produce results comparable to, and often better than, the 8 min acquisition. This is especially evident for the Restormer architecture. Relative to the 8 min acquisition, the fine-tuned Restormer achieved higher PSNR in 9 out of 12 samples and higher SSIM in 10 out of 12 samples. The Restormer trained from scratch achieved higher PSNR in 10 out of 12 samples and higher SSIM in 11 out of 12 samples, as evidenced by the ΔPSNR and ΔSSIM curves

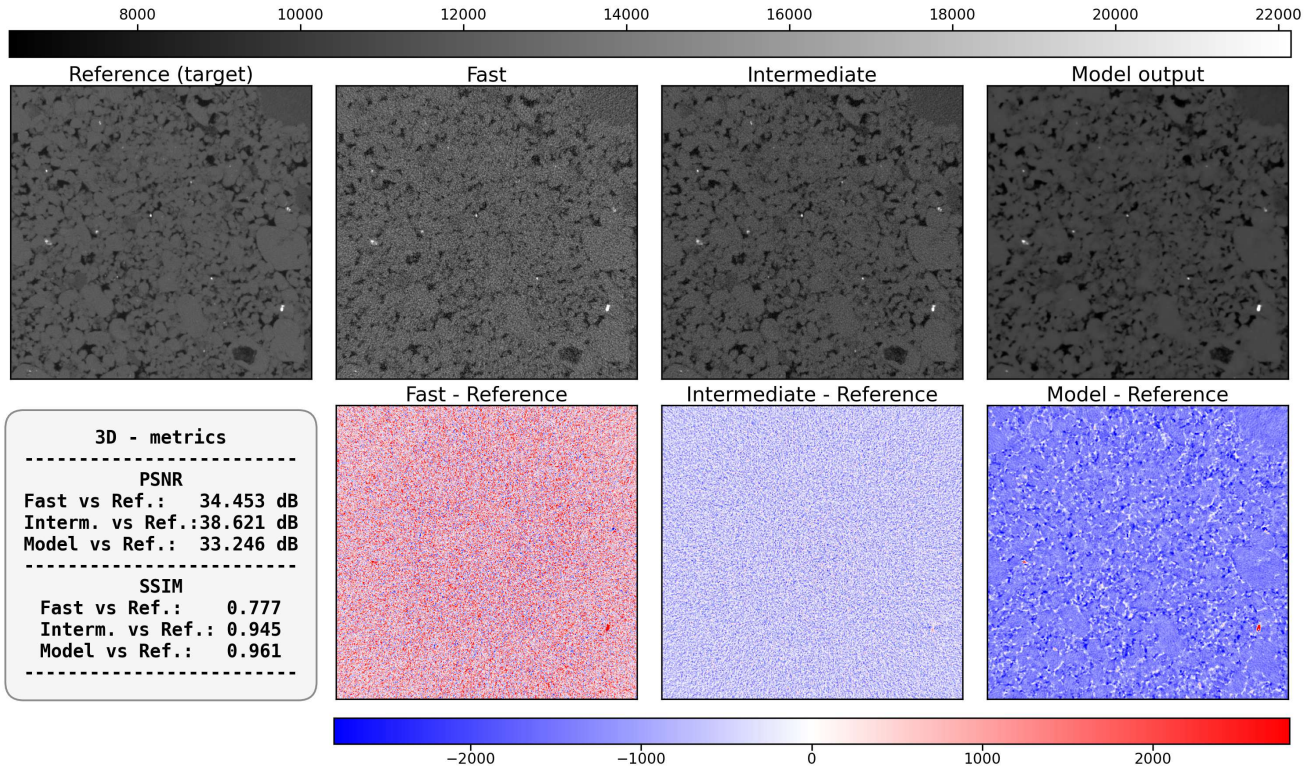


Figure 9. Qualitative comparison of the center axial slice of Sample 4. The top row shows the 60 min reference, the 2 min 35 s fast acquisition, the 8 min intermediate acquisition, and the denoised output obtained with the Restormer model with random initialization. The bottom row shows the corresponding difference maps relative to the reference. The PSNR and SSIM values reported were computed over the full 3D cropped volume.

Table 5. Contrast-adjusted PSNR and SSIM values for each method and sample.

Method	Metric	Sample												
		1	2	3	4	5	6	7	8	9	10	11	12	$\mu \pm \sigma$
2 min	PSNR	37.022	37.609	36.465	38.824	41.479	39.121	40.977	40.518	39.331	43.014	41.658	41.327	39.779 ± 2.047
	SSIM	0.891	0.896	0.866	0.920	0.965	0.925	0.943	0.955	0.942	0.972	0.963	0.959	0.933 ± 0.034
8 min	PSNR	41.679	**42.091**	42.098	43.288	44.364	42.722	**44.304**	43.586	42.387	45.292	44.672	43.770	43.354 ± 1.166
	SSIM	0.945	0.951	0.948	0.961	0.971	0.956	**0.968**	0.965	0.955	0.977	0.972	0.967	0.961 ± 0.010
Restormer zero-shot	PSNR	38.188	38.646	38.452	39.427	43.185	40.287	40.831	42.173	41.457	44.390	43.158	42.531	41.061 ± 2.084
	SSIM	0.914	0.907	0.924	0.929	0.967	0.931	0.945	0.961	0.949	0.973	0.966	0.960	0.944 ± 0.022
Restormer fine-tuned	PSNR	**44.049**	39.268	**44.758**	43.679	**46.857**	**43.969**	42.520	42.605	**45.258**	46.738	**46.657**	**46.310**	44.389 ± 2.237
	SSIM	**0.972**	0.947	**0.978**	0.968	**0.983**	**0.966**	0.961	**0.978**	**0.976**	**0.984**	**0.981**	**0.980**	0.973 ± 0.011
Restormer random init.	PSNR	42.744	41.794	43.481	**43.824**	46.845	43.111	43.829	**45.150**	44.140	**46.813**	46.342	46.244	44.526 ± 1.709
	SSIM	0.970	**0.956**	0.976	**0.969**	0.982	0.965	0.966	**0.978**	0.974	**0.984**	**0.981**	**0.980**	0.973 ± 0.008
RED-CNN	PSNR	42.175	41.596	42.261	43.092	44.734	42.349	42.994	44.434	43.724	46.002	45.291	44.800	43.621 ± 1.417
	SSIM	0.962	0.950	0.968	0.966	0.978	0.959	0.961	0.976	0.972	0.982	0.979	0.976	0.969 ± 0.010

relative to the 8 min acquisition.

The qualitative comparison before intensity adjustment (Figure 9) and the corresponding contrast-adjusted comparison (Figure 11) explain the improvement in the metrics for Sample 4, especially in PSNR. After intensity adjustment, the gray-value scale becomes better aligned with the reference, and the error map becomes predominantly white, with residual values concentrated around zero.

Results for Samples 4 (Figure 11) and 9 (Figure 12) show that, in the fast acquisition, the image contains a strong noise pattern with a static-like appearance, which makes the pore structure and matrix texture difficult to distinguish clearly. Although the contrast-adjusted evaluation improves the quantitative agreement, the residual maps suggest that contrast mismatch is

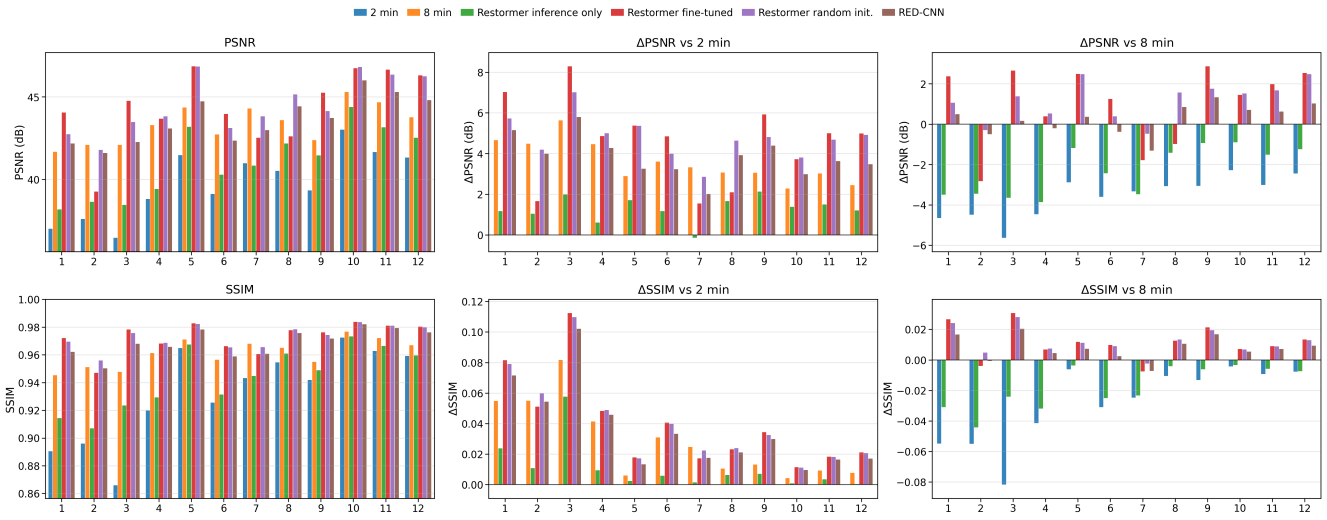


Figure 10. PSNR and SSIM results after intensity adjustment for each sample across the selected acquisition and denoising methods. The top row shows PSNR and the corresponding PSNR differences relative to the 2 min 35 s and 8 min baselines, while the bottom row shows the same comparison for SSIM. Positive Δ values indicate improvement over the corresponding baseline.

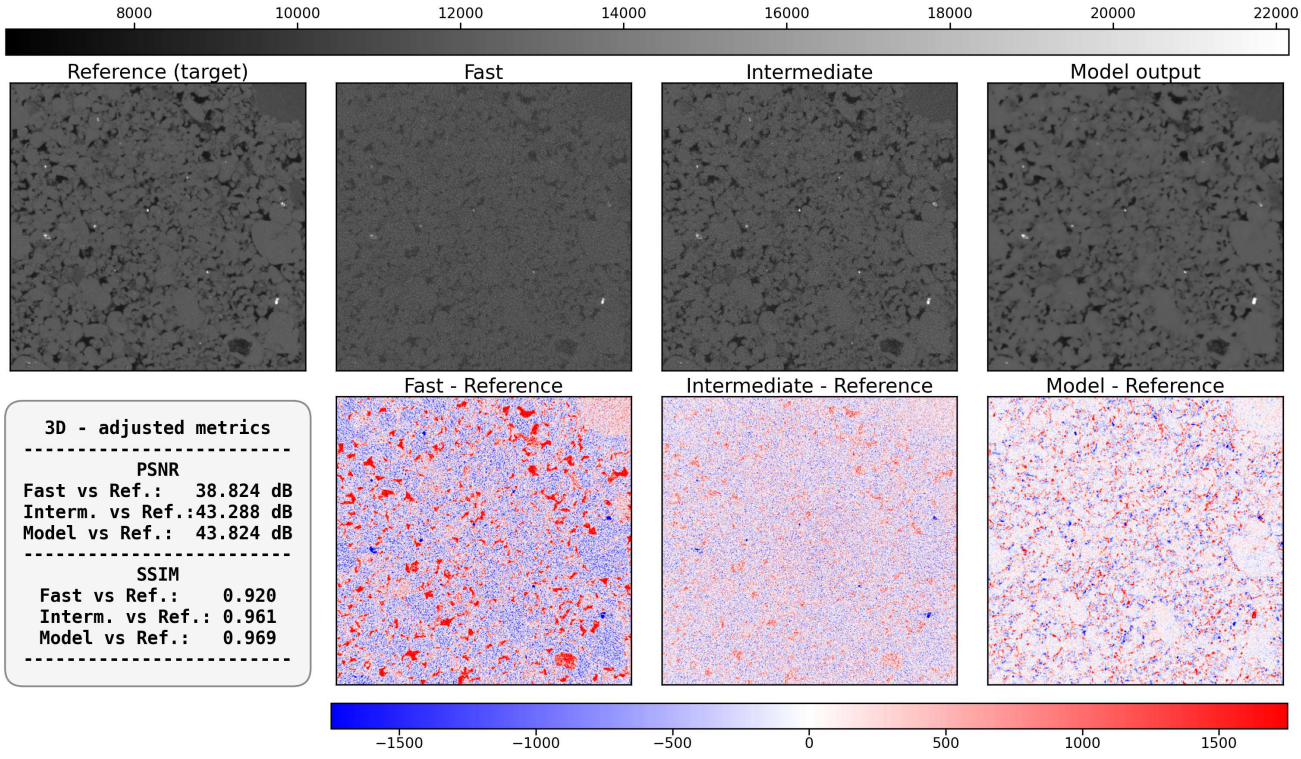


Figure 11. Qualitative comparison of the center axial slice of Sample 4 after intensity adjustment. The top row shows the 60 min reference, the 2 min 35 s fast acquisition, the 8 min intermediate acquisition, and the denoised output obtained with the Restormer model with random initialization. The bottom row shows the corresponding difference maps relative to the reference. The PSNR and SSIM values reported were computed over the full 3D cropped volume.

still not fully removed. In particular, the pore regions tend to present higher gray values while parts of the matrix become relatively lower, so adjusting the volume intensity scale to improve one phase can worsen the agreement in the other. The

8 min acquisition appears more consistent with the reference, while the model output shows visual improvement, with a more white-dominant residual map.

The model outputs show a clear texture improvement compared with the fast acquisition, which is also reflected in the higher PSNR and SSIM values. However, especially for Sample 9 (Figure 12), some small high-density materials, visible as bright points, are not fully recovered. This is expected, since they are barely visible or absent in the 2 min input image. However, it also represents an important limitation: the model can enhance and denoise structures that are present in the input signal, but it cannot reliably recover details that were not captured by the fast acquisition.

For the model output, the residuals are mainly concentrated along pore boundaries. Since pores have lower gray values than the rock matrix, negative residuals, shown in blue, indicate regions where the model is darker than the reference, which may suggest pore enlargement or the creation of pore-like regions. Positive residuals, shown in red, indicate regions where the model is brighter than the reference, which may suggest pore loss or partial filling. These effects are especially relevant for threshold-based segmentation, where small gray-level shifts near pore boundaries can change the resulting binary pore structure and affect derived physical measurements [38].

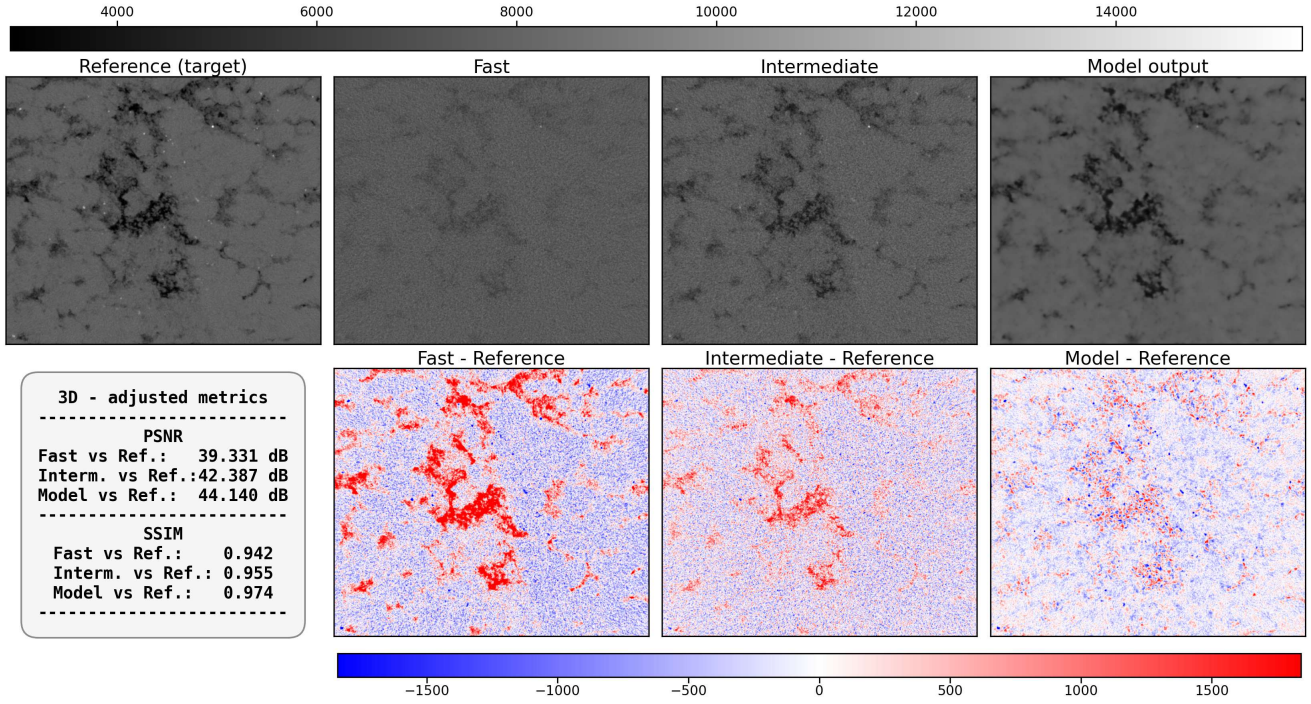


Figure 12. Qualitative comparison of the center axial slice of Sample 9 after intensity adjustment. The top row shows the 60 min reference, the 2 min 35 s fast acquisition, the 8 min intermediate acquisition, and the denoised output obtained with the Restormer model with random initialization. The bottom row shows the corresponding difference maps relative to the reference. The PSNR and SSIM values reported were computed over the full 3D cropped volume.

Figure 12 shows the residual differences after intensity adjustment. A pore–solid intensity imbalance remains visible, indicating that a single intensity shift cannot correct both phases simultaneously. This is especially clear in the Fast–Reference residual map: the residuals tend to have opposite signs in the pore phase and in the rock matrix. Since the solid phase occupies most of the volume, the adjustment is driven mainly by the rock phase. It therefore reduces rock-phase intensity differences more effectively, while the pore phase retains larger absolute residuals, visible as darker regions in the residual map. Thus, the contrast mismatch appears to be phase-dependent. Correcting this effect would require changing the intensity distribution, not only applying an intensity shift, which could distort the data and affect the quantitative results. However, even before the contrast adjustment, the model performed comparably to the 8 min acquisition, as shown in Table 4.

Pore structure and performance gains

Table 6 summarizes the pore-structure variability among the samples. The small-pore fraction was computed as the fraction of pore voxels with local diameter below 3 voxels. Given the voxel sizes in Table 1, this threshold corresponds to different physical lengths across samples, approximately below 130–162 µm. The threshold was intentionally kept in voxel units

because the analysis aims to probe structures represented by only a few voxels in the reconstructed grid, rather than to define a common micrometer-scale pore-size class across all samples. Local diameter was estimated as twice the Euclidean distance to the nearest solid voxel in the binary pore mask. Therefore, this value should be interpreted as a voxel-based measure of locally narrow or near-boundary pore space, rather than as a count of separate small pores or as a uniform physical pore-size threshold. This definition is simpler because it does not require individual pores to be segmented or labeled; it only uses the local distance-transform value at each pore voxel. Sample 3 has the highest porosity and the lowest small-pore fraction, indicating that its pore volume is less concentrated in locally narrow regions. In contrast, several lower-porosity samples have larger small-pore fractions, suggesting that a greater share of their pore space is associated with narrow or near-boundary regions.

Table 6. Porosity and small-pore fraction for each sample. The small-pore fraction is the fraction of pore-phase voxels with local diameter below 3 voxels, calculated from the Euclidean distance transform of the binary pore mask.

Pore attribute	Sample												
	1	2	3	4	5	6	7	8	9	10	11	12	$\mu \pm \sigma$
Porosity	0.136	0.120	0.159	0.083	0.054	0.087	0.104	0.036	0.038	0.061	0.078	0.083	0.087 ± 0.038
Small-pore fraction	0.518	0.683	0.360	0.760	0.725	0.760	0.901	0.849	0.727	0.775	0.772	0.756	0.716 ± 0.145

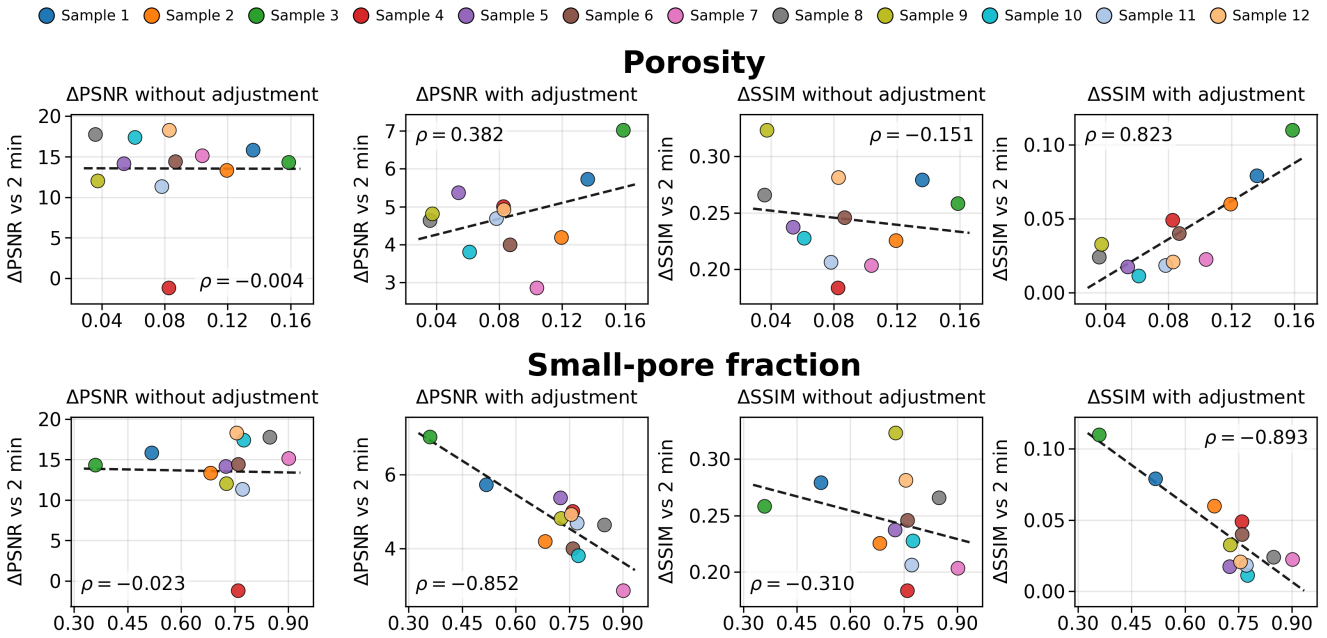


Figure 13. Relationship between pore attributes and performance gains of the Restormer random-initialization model relative to the 2 min baseline. Each point represents one sample. Rows correspond to porosity and the small-pore fraction, defined as the fraction of pore-phase voxels with local diameter below 3 voxels. Columns show ΔPSNR and ΔSSIM before and after intensity adjustment. Dashed lines indicate fitted linear regressions, and ρ denotes the Pearson correlation coefficient.

Figure 13 relates the pore attributes of each sample to the performance gains obtained by the Restormer random-initialization model relative to the 2 min baseline. Before intensity adjustment, the gains show only weak associations with pore structure. For porosity, the correlations are close to zero or weakly negative, with $\rho = -0.004$ for ΔPSNR and $\rho = -0.151$ for ΔSSIM. The small-pore fraction shows a similar pattern, with $\rho = -0.023$ for ΔPSNR and $\rho = -0.310$ for ΔSSIM. These weak correlations suggest that, before adjustment, protocol-dependent intensity and contrast differences dominate the PSNR and SSIM response, making the relationship between denoising performance and pore structure difficult to interpret.

After intensity adjustment, the performance gains relative to the 2 min baseline become smaller overall, as reflected by the reduced y-axis scale. This indicates that part of the apparent gain before adjustment was associated with intensity mismatch. However, the adjusted comparison is more informative for assessing pore-scale effects because contrast differences between acquisitions have been reduced. After intensity adjustment, clearer pore-structure trends emerge. Porosity is strongly positively

correlated with adjusted ΔSSIM ($\rho = 0.823$), whereas the small-pore fraction is strongly negatively correlated with adjusted ΔPSNR and ΔSSIM ($\rho = -0.852$ and $\rho = -0.893$, respectively).

This suggests that samples with more pore volume represented by locally narrow or boundary-adjacent regions benefit less from denoising, likely because these regions are more easily obscured by static-like noise and are closer to the image resolution limit in voxel units. Because voxel size varies across samples, this trend should not be interpreted as a single physical pore-size cutoff shared by all plugs. Rather, it indicates that model gains decrease when a larger fraction of the pore space is represented by only a few voxels on the reconstructed grid.

4 Conclusion

The present work investigated whether supervised deep learning can reduce the trade-off between acquisition time and image quality in micro-CT imaging of carbonate rock samples. The objective was to learn a fast-to-long acquisition mapping from 2 min 35 s scans to the corresponding 60 min reference acquisitions. This problem is particularly relevant for heterogeneous pre-salt carbonates, where acquisition noise and reduced structural definition can obscure complex pore structures and affect downstream digital rock analysis.

The proposed methodology was based on paired acquisitions of 12 carbonate plug samples scanned at 2 min 35 s, 8 min, and 60 min, with no sample repositioning between scans. Restormer and RED-CNN models were evaluated using a leave-one-out strategy at the sample level, ensuring that each test volume came from a plug not used during training. The models were compared against the 2 min fast acquisition and the 8 min intermediate acquisition using PSNR and SSIM computed over the full 3D cropped volumes. An additional contrast-adjusted evaluation was also performed to focus on structural differences and minimize gray-value mismatches.

The results show that the trained models substantially improved image quality relative to the fast acquisition. Without contrast adjustment, both trained Restormer variants outperformed the 8 min baseline on average, achieving the highest mean PSNR and SSIM values among the evaluated methods. The contrast-adjusted evaluation showed that acquisition-dependent gray-value differences strongly affect PSNR and SSIM, especially for the fast acquisition. After applying a volume-wise affine intensity adjustment, the fast acquisition metrics improved substantially. Nevertheless, the trained Restormer models remained the best-performing approaches overall. In the sample-wise analysis, the trained Restormer models achieved the best PSNR values in 10 out of 12 samples and were among the best SSIM results in 11 out of 12 samples. These results indicate that the learned fast-to-long mapping can often produce image quality comparable to the 8 min acquisition, despite using only the 2 min 35 s scan as input.

Qualitative analyses support the quantitative findings. The model outputs reduced the static-like noise pattern observed in the fast acquisitions and improved the visibility of pore and solid textures. However, the residual maps also showed errors concentrated near pore boundaries, where small gray-value differences can change the apparent pore–solid interface. In addition, small high-density structures that were absent or barely visible in the fast input could not be reliably recovered by the models. After contrast adjustment, samples with higher small-pore fractions tended to show smaller improvements against the fast scan, with strong negative correlations between small-pore fraction and both adjusted ΔPSNR and adjusted ΔSSIM. This suggests that performance is partly limited by image resolution.

Overall, the results demonstrate that learning a fast-to-long acquisition mapping using paired acquisitions is a promising strategy for improving the quality of fast micro-CT scans of carbonate rocks. The trained Restormer models were able to recover much of the image quality associated with longer acquisitions and obtain results similar to the intermediate scan. However, the study is limited by the relatively small number of samples, the use of a single scanner, and a fixed set of acquisition conditions. Results indicate that gray-value variability can occur across acquisitions even under nominally fixed scanning conditions. In this study, some acquisition parameters were intentionally changed to enable the faster 2 min 35 s scan, which may further contribute to acquisition-dependent intensity shifts when comparing micro-CT volumes. Future work should therefore evaluate the approach using different micro-CT equipment, broader acquisition parameters, and even faster scan protocols. This would support the development of faster acquisition workflows and help enable time-resolved 4D micro-CT experiments under dynamic laboratory conditions.

Code availability

The deep learning framework and inference workflows were written in `Python` using the `PyTorch` library. The full implementation code is available via GitHub at: https://github.com/luancvieira/image-denoising.

Data availability

This dataset is available upon request with permission from the Brazilian National Agency of Petroleum, Natural Gas and Biofuels (https://www.gov.br/anp/pt-br).

Funding

Open access funding provided by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) within the CAPES–Springer Nature agreement. This work was also supported by Petrobras, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

Author contributions

Conceptualization: Luan Vieira; Aurea Pereira Martins Neta; Rodrigo Surmas; Alexandre Evsukoff. Methodology: Rodrigo Surmas; Luan Vieira; Aurea Pereira Martins Neta; Alexandre Evsukoff. Data acquisition: Aurea Pereira Martins Neta. Data processing: Luan Vieira; Aurea Pereira Martins Neta; Alyne Duarte Vidal. Writing – Original Draft: Luan Vieira; Aurea Pereira Martins Neta; Alexandre Evsukoff. Computational analysis: Luan Vieira. Investigation: Luan Vieira; Aurea Pereira Martins Neta; Felipe Bevilaqua Foldes Guimarães; Júlio de Castro Vargas Fernandes; Carlos Eduardo Menezes dos Anjos; Rodrigo Luna; Alyne Duarte Vidal; Lizianne Carvalho Medeiros; Ricardo Alencar; Rodrigo Surmas; Alexandre Evsukoff. Writing – Review & Editing: Luan Vieira; Alexandre Evsukoff. Supervision: Rodrigo Surmas; Alexandre Evsukoff. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

References

1. Ahmed, T. *Reservoir engineering handbook* (Gulf professional publishing, 2018).
2. Bera, A., Shukla, B. & Jogani, D. A perspective review of applications of the computed tomography (ct) scan imaging technique for microscopic reservoir rock characterization. *Deep. Undergr. Sci. Eng.* **5**, 4–27 (2026).
3. Withers, P. J. *et al.* X-ray computed tomography. *Nat. Rev. Methods Primers* **1**, 18 (2021).
4. Jyoti, A. & Haese, R. R. Comparison of petrophysical properties of porous rocks using nmr, micro-ct, and fluid flow simulations. *Geosciences* **11**, 500 (2021).
5. Blunt, M. J. *et al.* Pore-scale imaging and modelling. *Adv. Water resources* **51**, 197–216 (2013).
6. da Silva, L. C. V. *et al.* Entropy-based measure of rock sample heterogeneity derived from micro-ct images: Lcv da silva et al. *Transp. Porous Media* **152**, 44 (2025).
7. Li, H., Aslam, B. & Yan, B. An integrated deep learning and physics-constrained upscaling workflow for robust permeability prediction in digital rock physics. *SPE J.* **30**, 3220–3239 (2025).
8. Roslin, A. *et al.* Processing of micro-ct images of granodiorite rock samples using convolutional neural networks (cnn), part i: Super-resolution enhancement using a 3d cnn. *Miner. Eng.* **188**, 107748 (2022).
9. Irie, M. S., Spin-Neto, R., Borges, J. S., Wenzel, A. & Soares, P. B. F. Effect of data binning and frame averaging for micro-ct image acquisition on the morphometric outcome of bone repair assessment. *Sci. Reports* **12**, 1424 (2022).
10. Saxena, N. *et al.* Rock properties from micro-ct images: Digital rock transforms for resolution, pore volume, and field of view. *Adv. Water Resour.* **134**, 103419, 10.1016/j.advwatres.2019.103419 (2019).
11. Cataldo, R. A., Leite, E. P., Rebelo, T. B. & Mattos, N. H. Depth-variant pore type modeling in a pre-salt carbonate field offshore brazil. *Frontiers Earth Sci.* **10**, 1014573 (2022).
12. Belila, A. M. P., Basso, M., Chinelatto, G. F., Kuroda, M. C. & Vidal, A. C. Pore typing using nuclear magnetic resonance, an example with samples from cretaceous pre-salt lacustrine carbonates in the santos basin, brazil. *J. Petroleum Sci. Eng.* **190**, 107079 (2020).
13. Zamir, S. W. *et al.* Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5728–5739 (2022).
14. Su, J., Xu, B. & Yin, H. A survey of deep learning approaches to image restoration. *Neurocomputing* **487**, 46–65, 10.1016/j.neucom.2022.02.046 (2022).

15. Zhang, J. *et al.* A review of deep learning methods for denoising of medical low-dose ct images. *Comput. Biology Medicine* **171**, 108112, 10.1016/j.compbiomed.2024.108112 (2024).

16. Li, L. *et al.* Ct-mamba: A hybrid convolutional state space model for low-dose ct denoising. *Comput. Medical Imaging Graph.* **124**, 102595, 10.1016/j.compmedimag.2025.102595 (2025).

17. Bizhani, M., Ardakani, O. H. & Little, E. Reconstructing high fidelity digital rock images using deep convolutional neural networks. *Sci. reports* **12**, 4264 (2022).

18. Sidorenko, M., Orlov, D., Ebadi, M. & Koroteev, D. Deep learning in denoising of micro-computed tomography images of rock samples. *Comput. & Geosci.* **151**, 104716 (2021).

19. Alsamadony, K. L., Yildirim, E. U., Glatz, G., Waheed, U. B. & Hanafy, S. M. Deep learning driven noise reduction for reduced flux computed tomography. *Sensors* **21**, 1921, 10.3390/s21051921 (2021).

20. Dosovitskiy, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations* (2021).

21. LeCun, Y. & Bengio, Y. Convolutional networks for images, speech, and time series. *The handbook brain theory neural networks* (1998).

22. Wong, T.-T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern recognition* **48**, 2839–2846 (2015).

23. Davis, G. R. Development of two-dimensional beam hardening correction for x-ray micro-ct. *J. X-Ray Sci. Technol.* **30**, 863–874, 10.3233/XST-221178 (2022).

24. Evershed, A. N. Z., Mills, D. & Davis, G. R. Multi-species beam hardening calibration device for x-ray microtomography. In *Developments in X-Ray Tomography VIII*, vol. 8506 of *Proceedings of SPIE*, 85061N, 10.1117/12.928933 (2012).

25. Zhang, Y., Mostaghimi, P. & Armstrong, R. T. On the challenges of greyscale-based quantifications using x-ray computed microtomography. *J. Microsc.* **275**, 82–96 (2019).

26. Bagci, U., Udupa, J. K. & Bai, L. The influence of intensity standardization on medical image registration. In *Medical Imaging 2010: Visualization, Image-Guided Procedures, and Modeling*, vol. 7625, 602–613 (SPIE, 2010).

27. Nyúl, L. G. & Udupa, J. K. On standardizing the mr image intensity scale. *Magn. Reson. Medicine: An Off. J. Int. Soc. for Magn. Reson. Medicine* **42**, 1072–1081 (1999).

28. Thermo Fisher Scientific. Thermo Scientific PerGeos Software (2024). Digital rock analysis software. Available at: https://www.thermofisher.com/us/en/home/electron-microscopy/products/software-em-3d-vis/pergeos-software.html. Accessed: 30 June 2026.

29. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. on Syst. Man, Cybern.* **9**, 62–66, 10.1109/TSMC.1979.4310076 (1979).

30. Vincent, L. & Soille, P. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Trans. on Pattern Anal. Mach. Intell.* **13**, 583–598, 10.1109/34.87344 (1991).

31. Abdelhamed, A., Lin, S. & Brown, M. S. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1692–1700 (2018).

32. Xian, Y., Lampert, C. H., Schiele, B. & Akata, Z. Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly. *IEEE Trans. on Pattern Anal. Mach. Intell.* **41**, 2251–2265 (2018).

33. Long, M., Cao, Y., Wang, J. & Jordan, M. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, 97–105 (PMLR, 2015).

34. Chen, H. *et al.* Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE transactions on medical imaging* **36**, 2524–2535 (2017).

35. Wythoff, B. J. Backpropagation neural networks: a tutorial. *Chemom. Intell. Lab. Syst.* **18**, 115–155 (1993).

36. Van der Walt, S. *et al.* scikit-image: image processing in python. *PeerJ* **2**, e453 (2014).

37. Mudeng, V., Kim, M. & Choe, S.-w. Prospects of structural similarity index for medical image analysis. *Appl. Sci.* **12**, 3754 (2022).

38. Moreira, A. C. *et al.* Effects of manual threshold setting on image analysis results of a sandstone sample structural characterization by x-ray microtomography. *Appl. Radiat. Isot.* **70**, 937–941, 10.1016/j.apradiso.2012.03.001 (2012).