



Comparative Study of Traditional and Deep-Learning Denoising Approaches for Image-Based Petrophysical Characterization of Porous Media

Miral S. Tawfik^{1*}, Amogh Subbkrishna Adishesha², Yuhan Hsi², Prakash Purswani³, Russell T. Johns^{3,4}, Parisa Shokouhi⁵, Xiaolei Huang² and Zuleima T. Karpyn^{3,4}

¹ Reservoir Engineering and Simulation, Chevron Technical Center, Houston, TX, United States, ² Information Science and Technology, The Pennsylvania State University, University Park, PA, United States, ³ The Energy Institute, The Pennsylvania State University, University Park, PA, United States, ⁴ John and Willie Leone Family Department of Energy and Mineral Engineering, The Pennsylvania State University, University Park, PA, United States, ⁵ Department of Engineering Science and Mechanics, The Pennsylvania State University, University Park, PA, United States

OPEN ACCESS

Edited by:

James E. McClure,
Virginia Tech, United States

Reviewed by:

Tim Scheibe,
Pacific Northwest National Laboratory
(DOE), United States
Xiaofan Yang,
Beijing Normal University, China

*Correspondence:

Miral S. Tawfik
mtawfik@alumni.psu.edu

Specialty section:

This article was submitted to
Water and Critical Zone,
a section of the journal
Frontiers in Water

Received: 23 October 2021

Accepted: 10 December 2021

Published: 25 January 2022

Citation:

Tawfik MS, Adishesha AS, Hsi Y, Purswani P, Johns RT, Shokouhi P, Huang X and Karpyn ZT (2022) Comparative Study of Traditional and Deep-Learning Denoising Approaches for Image-Based Petrophysical Characterization of Porous Media. *Front. Water* 3:800369. doi: 10.3389/frwa.2021.800369

Digital rock physics has seen significant advances owing to improvements in micro-computed tomography (MCT) imaging techniques and computing power. These advances allow for the visualization and accurate characterization of multiphase transport in porous media. Despite such advancements, image processing and particularly the task of denoising MCT images remains less explored. As such, selection of proper denoising method is a challenging optimization exercise of balancing the tradeoffs between minimizing noise and preserving original features. Despite its importance, there are no comparative studies in the geoscience domain that assess the performance of different denoising approaches, and their effect on image-based rock and fluid property estimates. Further, the application of machine learning and deep learning-based (DL) denoising models remains under-explored. In this research, we evaluate the performance of six commonly used denoising filters and compare them to five DL-based denoising protocols, namely, noise-to-clean (N2C), residual dense network (RDN), and cycle consistent generative adversarial network (CCGAN)—which require a clean reference (ground truth), as well as noise-to-noise (N2N) and noise-to-void (N2V)—which do not require a clean reference. We also propose hybrid or semi-supervised DL denoising models which only require a fraction of clean reference images. Using these models, we investigate the optimal number of high-exposure reference images that balances data acquisition cost and accurate petrophysical characterization. The performance of each denoising approach is evaluated using two sets of metrics: (1) standard denoising evaluation metrics, including peak signal-to-noise ratio (PSNR) and contrast-to-noise ratio (CNR), and (2) the resulting image-based petrophysical properties such as porosity, saturation, pore size distribution, phase connectivity, and specific surface area (SSA). Petrophysical estimates show that most traditional filters perform well when estimating bulk properties but show large errors for pore-scale properties like phase connectivity. Meanwhile, DL-based models give mixed outcomes, where supervised

methods like N2C show the best performance, and an unsupervised model like N2V shows the worst performance. N2N75, which is a newly proposed semi-supervised variation of the N2N model, where 75% of the clean reference data is used for training, shows very promising outcomes for both traditional denoising performance metrics and petrophysical properties including both bulk and pore-scale measures. Lastly, N2C is found to be the most computationally efficient, while CCGAN is found to be the least, among the DL-based models considered in this study. Overall, this investigation shows that application of sophisticated supervised and semi-supervised DL-based denoising models can significantly reduce petrophysical characterization errors introduced during the denoising step. Furthermore, with the advancement of semi-supervised DL-based models, requirement of clean reference or ground truth images for training can be reduced and deployment of fast X-ray scanning can be made possible.

Keywords: image processing, micro-computed tomography, deep learning, denoising, image enhancement, digital rock physics, carbon capture utilization and storage (CCUS), enhanced oil recovery

BACKGROUND AND INTRODUCTION

Micro-computed tomography (MCT) is a non-destructive technique used to visualize the internal structure of objects in a variety of disciplines including medicine, dentistry, tissue engineering, aerospace engineering, geology, and material and civil engineering (Orhan, 2020). The physical principle behind this technique is that X-rays attenuate differently as they penetrate through different materials depending on their density and atomic mass (Knoll, 2000; Attix, 2004; Ritman, 2004; Hsieh, 2015). This makes MCT an ideal tool for characterizing multiphase materials such as rocks with fluid phases of different densities.

The use of MCT imaging is becoming increasingly indispensable in several disciplines, including geotechnical and petrophysical characterization and understanding multiphase flow in porous media. Wang et al. (1984) first adopted medical CT imaging to monitor the injectant front to understand the effect of pore structure heterogeneity on oil recovery and interactions between oil and injected fluid in a Berea sandstone. However, medical CT imaging soon proved inadequate in describing the underlying pore-scale phenomena responsible for multiphase flow behavior in porous media (Cromwell et al., 1984). Therefore, high-resolution MCT imaging was used to obtain quantitative pore-scale information about structure–function relationships (Jasti et al., 1993). They characterized the 3D pore structure in a glass bead pack and three Berea sandstone samples to determine whether topological properties such as pore connectivity and phase features of individual fluid phases such as saturation can be resolved. Micro-computed tomography has since been successfully used for quantifying a wide range of petrophysical properties such as volume fraction for porosity and saturation quantification, specific surface area (SSA), pore- and blob-size distributions, *in-situ* contact angles, interface curvatures for local capillary pressures, grain sphericity, angularity, roughness as well as phase connectivity (Sharma and Yortsos, 1987; Prodanović and Bryant, 2006; Karpyn et al., 2010; Herring et al., 2013; Landry et al., 2014; Larpudomlert et al.,

2014; Berg et al., 2016; Klise et al., 2016; Scanziani et al., 2017; Chen et al., 2018; Tawfik et al., 2019; McClure et al., 2020). The accuracy with which we can estimate these pore-scale properties affect our ability to explain and predict multiphase fluid flow in porous media. For example, pore-network modeling, which is used in a variety of digital rock studies to explain and predict macroscopic transport properties such as absolute permeability, relative permeability, and capillary pressure, uses simplified pore structures composed of a network of pores and throats, which can be directly extracted from the MCT images (Valvatne and Blunt, 2004; Jia et al., 2007; Dong et al., 2009; Mostaghimi et al., 2013; Berg et al., 2016; Zahaf et al., 2017; Raeini et al., 2019). Similarly, Lattice Boltzmann simulation of multiphase flow is used to estimate the same transport properties but on real pore structures by directly using the binarized MCT images as input (Grader et al., 2010; Andrä et al., 2013; Landry et al., 2014; Liu et al., 2014; Armstrong et al., 2016; Kakouei et al., 2017; Li et al., 2018). The accuracy of the model predictions thus depends on the accuracy of processing the images that feed into these models.

Despite the fast-paced advances in MCT imaging, there are several factors that introduce noise to MCT images limiting the image quality and feature detectability. Those limitations can hinder our ability to accurately identify structure-function relationships and predict fluid flow behavior, which is essential in several applications including CO₂ storage, aquifer remediation, and hydrocarbon recovery. Some of the inherent limitations of MCT imaging include photon statistics, balance between representative elementary volume and resolution or feature detection, and partial volume consideration (Van Stappen, 2018). Apart from image resolution, there are numerous artifacts that arise during image acquisition. The prominent ones include photon starvation, detector saturation also known as ghosting, central rotation artifact, cone-beam effect, metal artifact, cupping artifact, streaks and dark bands, under-sampling, poor contrast, beam hardening, scatter, and ring artifacts (Boas and Fleischmann, 2012). These imaging artifacts result in different forms of noise in a CT image. Though noise in MCT has not been fully profiled, many studies including Diwakar and Kumar

(2018) and Lee et al. (2018) have attempted to profile the noise distributions through known statistical distributions. The most prevalent noise forms follow Poisson and Gaussian distributions. There are several other forms of noise including fractal noises like Perlin noise, periodic noise from helical sampling and finally random noise that may occur due to the change in X-ray quanta for multiple subject densities. Gaussian and Poisson noise are the most common types of noise observed in MCT images. Gaussian noise is additive and signal independent. Poisson noise is a consequence of low photon count, and results in random thin bright and dark streaks that appear preferentially along the direction of greatest attenuation (Boas and Fleischmann, 2012). The low photon count coupled with the error in quantization of the received photons leads to a mixture noise-model, which is a combination of Poisson and Gaussian distributions. Perlin noise is a pseudo-random texture gradient noise which is generated using octaves of noise flow vectors along with frequency and amplitude parameters to simulate spatially interpolated textures (Perlin, 1985). The granularity and texture of the noise profile in low quality (LQ) CT images can be captured by Perlin noise. Bae et al. (2018) showed that using Perlin noise profile to augment a lung CT dataset resulted in higher classification accuracy indicating that the features extracted from Perlin noise augmented images matched those of certain low-dose CT scans with lung conditions.

Mathematically, denoising is an inverse ill-posed problem where noisy data is a sparse representation of the real data and a unique solution for its restoration does not necessarily exist (Fan et al., 2019). For an exact solution to the problem, a priori knowledge of the types of noise, variance of noise within an image and along the image stack would be required. This information is usually not readily available. The basic principle behind most of the traditional spatial domain denoising filters is noise removal based on spatial correlation between an input pixel gray value and its neighboring pixels. Those filters can be classified into linear and non-linear filters. Linear filters are spatially invariant filters that output pixel values as a linear combination of the input pixel value and its neighborhood. Examples of linear filters include Gaussian and mean filters, which are optimal for Gaussian noise. However, linear filters tend to blur edges and blur small features. Conversely, non-linear filters can be spatially invariant like the median filter, or spatially variant like the anisotropic diffusion (AD) filter (Huang et al., 1979; Perona and Malik, 1990). Another approach to denoise MCT images involves machine-learning-based models. As such, the objective is to find the parameters that act on the sparse noisy data to obtain the closest representation of the clean signal. Several works in literature have approached this parameter estimation through the use of classic machine learning and modern neural network-based models (Tian et al., 2020). Improvements in the DL denoising field result from a clearer understanding of the specific noise forms and improvements in the architecture of neural networks. The literature points to three specific architectures, namely U-Nets, Residual Networks, and Generative Adversarial Networks (GANs), that have been established for addressing common noise forms including Poisson, Gaussian, and their mixtures which are predominant in CT images (Kulathilake et al., 2021).

Deep learning (DL) powered computer vision has seen phenomenal improvement since the introduction of AlexNet in 2012 (Krizhevsky et al., 2017). However, adaptation of imaging techniques for porous media is still work in progress. Kamrava et al. (2019) trained a 20-layer feedforward neural network using synthetic data generated by the stochastic cross-correlation-based simulation (CCSIM) algorithm and found that neural networks perform better compared to bi-cubic interpolation for image super-resolution while the synthetic data further improved the model's generalizability. Sidorenko et al. (2021) implemented an encoder-decoder network for denoising MCT images obtained from tight sandstone samples. The authors compared the effect of different loss functions and found that the least absolute deviation (L1) loss gave a better result in terms of peak signal to noise ratio (PSNR) and structural similarity index measure (SSIM) when compared to the least square errors (L2) loss, SSIM loss, and Visual Geometry Group (VGG) perceptual loss. Recently, Alsamadony et al. (2021) acquired low-exposure and high-exposure CT scans on the same carbonate sample and compared denoising performance using a pre-trained very deep super resolution (VDSR) network against a shallow U-Net. They demonstrated that DL-based image processing can improve image quality; and that pre-trained VDSR network with fine-tuning tends to out-perform VDSR trained from scratch. Wang et al. (2021) reviewed various DL applications in pore-scale imaging and modeling, including image segmentation, image super-resolution, petrophysical property prediction, flow simulation, as well as common convolutional neural network (CNN) architectures and various types of GANs for image generation. However, the discussion on image denoising was brief. A summary of findings from select literature on image denoising with applications in digital rock physics is provided in **Table 1**.

The literature survey reveals a need for a comprehensive comparison of denoising approaches for petrophysical applications. Existing studies mostly utilize non-learnable filters and rely on visual inspection to evaluate the effect of image denoising. The lack of quantitative metrics, both standard and physics-based, cause errors during this critical image processing step to remain largely unquantified. Finally, none of the reported studies consider the impact of denoising techniques on downstream tasks such as image segmentation and petrophysical property estimation, which is often the ultimate goal of quantitative imaging.

In this paper, we address these research gaps and compare the performance of traditional (user-based) denoising methods against more sophisticated DL-based denoising methods. Four types of models are considered in this study: (a) commonly used filters such as the Gaussian, non-local means (NLM), Median etc., (b) fully supervised models including the noise-to-clean (N2C), residual dense network (RDN), and cycle consistent generative adversarial network (CCGAN), (c) semi-supervised models that use a portion of clean reference images for learning, and (d) completely unsupervised techniques like the noise-to-noise (N2N) and noise-to-void (N2V). We evaluate the performance of the different denoising models using both qualitative and quantitative analyses against an independent ground truth.

TABLE 1 | A summary of selected publications that report on MCT imaging in the digital rock physics domain, highlighting the commonly used denoising methods.

| References | Porous medium | Resolution, microns | Denoising protocol(s) | Denoising evaluation | Image-based petrophysical analyses |
|-------------------------------|--|--------------------------------------|--|---------------------------------------|--|
| Culligan et al., 2004 | Glass beadpack | 18 | Median | – | Saturation, interfacial area |
| Porter and Wildenschild, 2010 | Beadpack | 11.8, 5.9 | Anisotropic diffusion, Median | Qualitative | Interfacial area |
| Landry et al., 2011 | Acrylic Beadpack | 25.9 | Median | – | Surface and interfacial area |
| Iglauer et al., 2012 | Clashach Sandstone | 9 | Anisotropic regularization | – | Porosity, residual oil saturation, residual oil cluster distribution |
| Brown et al., 2014 | SiO ₂ Beadpack | 10.6, 5.3 | Median, total variation, majority | Qualitative | Saturation, interfacial area, mean interface curvature |
| Andrew et al., 2014 | Bentheimer SS Doddington SS Ketton LS Mount Gambier LS Estailades LS | 6.16 5.39 4.60 4.44 3.93 | Non-local means | – | Pore size distribution, trapped residual scCO ₂ saturation, trapped scCO ₂ ganglia size distribution |
| Alyafei et al., 2015 | Berea SS Doddington SS Ketton LS Estailades LS | 2.7, 5.4, 10.8, 21.6 | Wavelet-Fourier filter, anisotropic regularization | – | Porosity, pore size distribution, permeability |
| Freire-Gormaly et al., 2015 | Indiana Limestone Pink Dolomite | 7.5, 8.3, 11.07 | Median | – | Porosity, pore size distribution, pore and throat radius, coordination number, pore-to-pore distance, capillary pressure |
| Menke et al., 2016 | Estailades LS Portland Base bed LS | 4.76 | Non-local means | – | Porosity, permeability, specific surface area |
| Al-Menhali et al., 2016 | Estailades LS | 4.9 | Non-local means | – | CO ₂ and N ₂ saturation, ganglia morphology, and distribution, <i>in-situ</i> contact angles |
| Berg et al., 2016 | Gildehauser SS | 2.2 | Non-local means | – | Porosity, saturation, permeability, relative permeability |
| AlRatrouf et al., 2017 | Ketton Limestone | 3.28 | Non-local means | – | <i>In-situ</i> contact angles |
| Verri et al., 2017 | Tight rock 1 Tight rock 2 Sandstone 1 Sandstone 2 | 2.175 2.73 2.13 2.05 | Bilateral, non-local means | Qualitative | Porosity, specific surface area, equivalent pore diameter, permeability, tortuosity |
| Singh et al., 2017 | Ketton LS | 3.28 | Non-local means | SNR | Local capillary pressure, oil saturation |
| Gao et al., 2017 | Bentheimer SS | 6 | Non-local means | – | Saturation, relative permeability |
| Lin et al., 2018 | Bentheimer SS | 3.58 | Non-local means | SNR, Average phase boundary sharpness | Interfacial curvature, local capillary pressure |
| Rücker et al., 2019 | Ketton LS | 3.28 | Non-local means | – | <i>In-situ</i> contact angles, fluid distribution as a function of pore size |
| Tawfik, 2020 | Indiana LS | 2.6 | Anisotropic diffusion | Qualitative | <i>In-situ</i> contact angles |
| Purswani et al., 2020 | Sintered glass frit | 6, 18 | Non-local means | – | Porosity, saturation, interfacial area, connectivity |
| Sidorenko et al., 2021 | Tight SS | 1.2 | Residual encoder-decoder network (RedNet) | SSIM, L1, L2, VGG-perceptual | – |
| Niu, 2021 | Sintered glass frit | 15 | Non-local means | Qualitative | Porosity, saturation, <i>in-situ</i> contact angles |
| Alsamadony et al., 2021 | Carbonate rock | 14 | Pre-trained very deep SR (VDSR), U-Net | SSIM, L2 | – |

LS, limestone; SS, sandstone.

TABLE 2 | Properties of the porous core sample (Niu, 2021).

| | |
|--|---------|
| Diameter, in | 0.5 |
| Length, in | 2.0 |
| Porosity, % | ~14.7 |
| Absolute permeability, mD | 630.4 |
| Pore size distribution, μm | 250–500 |
| Inner surface (BET), m^2/g | ~0.015 |

We also compare computational resource requirements and ease of implementation of the different methods and provide recommendations. From quantitative comparison, we show the utility of physics-informed evaluation based on petrophysical property estimation. From this exercise, we address the following research questions: can sophisticated denoising protocols yield a reasonable characterization of a saturated porous medium using fast X-ray scanning with low-exposure time? Can we use a reduced number of, or no clean reference images and still yield accurate petrophysical characterization estimates using low-exposure MCT images?

METHODOLOGY

Image Acquisition

In this section, we briefly describe the experiment conducted by Niu (2021) to generate the datasets used in this study. The MCT images used in this study are of a fritted Robu-type glass (borosilicate) filter procured from Adam & Chittenden Scientific Glass Cooperation. This porous medium was selected as a proxy for real rocks to minimize changes that typically take place in real rock samples when fluids are injected, such as dissolution, which can alter the results of the study. The core sample properties are given in **Table 2**.

The experimental procedure consisted of wrapping the core sample in a heat shrink jacket and placing it in an X-ray transparent Core Lab biaxial FCH series core holder (CoreLab, 2021; <https://www.corelab.com/cli/core-holders/x-ray-core-holder-fch-series>). This core holder is made of an aluminum body wrapped in carbon fiber composite, which reduces the holder's X-ray absorption capacity during scanning. The porous medium is saturated with high salinity brine that has a composition representative of a saline aquifer (Tawfik et al., 2019). Brine was pre-equilibrated with CO_2 at 1,500 psi and 45°C before being used to saturate the core. The core sample was then flooded with supercritical CO_2 (scCO_2) at a pore pressure of 1,180 psi and temperature of $41.7 \pm 0.2^\circ\text{C}$ ($107.06 \pm 0.36^\circ\text{F}$), as well as confining pressure of $1,430 \pm 50$ psi. The confining pressure was applied by injecting deionized water into the annular space of the core holder. The high pore pressure and temperature enabled the maintenance of scCO_2 conditions.

The MCT images were acquired using a GE v|tome|x L300 multi-scale nano/microCT system, using the 300 kV X-ray tube. Two datasets were acquired at the same voxel resolution of $15 \mu\text{m}$: a high-quality (HQ) scan, where more exposure time was allowed to reduce noise. The second dataset was a low-quality

TABLE 3 | MCT scanning parameters.

| | HQ | LQ |
|---------------------------|-------|----|
| Resolution, μm | 15 | |
| Voltage, kV | 200 | |
| Current, μA | 70 | |
| Magnification | 13.5X | |
| Scanning time, min | 90 | 12 |

(LQ) scan which was under the same experimental conditions, using the same scanning parameters except for exposure time, where less scanning time (~ 7.5 times less) was needed to produce a noisier version of the HQ dataset. The HQ dataset is used in this study as a ground truth to evaluate the different denoising methods and compare the petrophysical properties computed using the different denoised datasets. The HQ dataset is also used as a clean reference dataset to train the model for some of the DL denoising algorithms, which are discussed later in this section. The scanning details for both datasets are presented in **Table 3**. The sinograms generated during image acquisition were reconstructed using GE's proprietary GPU-based reconstruction software *datos|x 2 reconstruction*.

Selection and Implementation of Denoising Methods

The datasets obtained from the scans contain 2,024 32-bit images. Pre-processing of the datasets involved extracting a cylindrical sub-volume to eliminate the core holder and heat shrink jacket. Additionally, the top 103 and bottom 490 slices were removed to eliminate the cone-beam effect, which is typically observed at the top and bottom of the scan volume.

Non-learnable, Traditional Denoising Filters

Numerous filtering algorithms have been developed for reducing image noise to make the gray-scale CT images ready for the image segmentation step, after which feature extraction and quantitative analysis is carried out. In this study, we focus on the errors incurred particularly during the denoising step. Denoising may take place before image reconstruction by taking advantage of existing statistical models of noise, derived from an understanding of its physical origin. However, due to missing visual perception, operations performed on a sinogram might have unpredictable effects on real features post reconstruction (Matrecano et al., 2010). On the other hand, post-reconstruction filters are pixel-to-pixel transformations of an image. This transformation is based on the gray-scale intensity of each pixel and its neighboring pixels (Machado et al., 2013). They can be classified as low-pass or high-pass filters; a low-pass filter is used for image smoothing, whereas a high-pass filter is typically used for image sharpening. Some of the most commonly used filters include Gaussian filter, median filter, NLM, and AD filter.

The Gaussian filter is a low-pass filter that uses a Gaussian function. Applying this filter involves convolving an image with a Gaussian function, which may result in a blurrier image. A

median filter is also a low-pass smoothening filter. It replaces the gray-scale value of each pixel with the median of its neighborhood. The neighborhood is pre-specified by the user as 6 faces, 18 edges, or 26 vertices. The process may be repeated to further smooth the image. The median filter usually works well when images contain non-Gaussian noise and/or very small artifacts. The NLM filter first proposed by Buades et al. (2005) assumes the noise to be white noise. The value specified for the target voxel is a weighted-average of neighboring voxels within a user-specified search window. The weights are determined by applying a Gaussian kernel based on the similarity between the neighborhoods. The AD filter (Weickert, 1996) also known as the Perona-Malik diffusion, aims to reduce image noise within each phase while preserving edges and boundaries between phases. Typically, a diffusion stop criterion that exceeds the variation within a given phase is set to identify when diffusion should no longer take place to terminate diffusion and preserve edges. If this stop criterion is set to the maximum difference between voxel intensity values, diffusion never stops and the AD filter converges to a simple isotropic Gaussian filter. Some of the other used filters applied include Bilateral, SNN, Total variance, Nagao, Kuwahara, BM3D, Minimum, and Maximum filters.

The image processing software Avizo (v2020) was used to implement the simple filters, including the Gaussian, median, NLM, AD, bilateral, and SNN filters. For the Gaussian filter, the standard deviation was set to two voxels, and the kernel size was set to nine voxels in each direction. The median filter neighborhood was set to 26 (vertex neighborhood), and a total of three iterations are performed. The NLM filter was implemented using the adaptive-manifolds-based approach (Gastal and Oliveira, 2012). The search window was specified to be 500 voxels, and the local neighborhood was set to four. A large enough search window is selected to increase the chances of finding similar structures and phases. The local neighborhood value was determined such that it is similar in size to the fine structures that can be found in the image. Spatial standard deviation (which determines the relationship between how fast the similarity value decreases as distance between the target and neighboring voxel increases), was set to five. Similarly, the intensity standard deviation (which determines the relationship between how fast the similarity value decreases based on voxel intensities) was set to 0.2. For the AD filter, a total of five smoothing iterations per image is performed, with the diffusion stop threshold set to 0.048. That is, if the difference between the target voxel value and the value of its neighboring six voxels exceeds this stop threshold value, diffusion will not occur, hence preserving and enhancing phase interfaces. The bilateral filter parameters were kernel size = 3 and similarity = 20. Finally, for the SNN filter a kernel size of 3 was used. It is evident that these simple filters rely heavily on the user and their experience which lend themselves to likely bias. As such, there is a need for non-user-based denoising protocols.

Deep Learning Denoising Methods

A few notable DL-based denoising algorithms have been selected based on their architecture and input requirements. The most common architectures for encoder-decoder based denoising are

U-Nets, GANs and RDNs. U-Nets are a sequence of contracting and expanding convolution blocks (forming a “U” shape) originally proposed by Ronneberger et al. (2015) for the purpose of biomedical image segmentation. U-Nets have the ability to accumulate hierarchical features at multiple resolutions and maintain sharpness throughout the decoding process (Diwakar and Kumar, 2018). Residual dense networks, on the other hand, are inspired by the neural connections present in a human brain. Here, output of a convolution block is combined with its input to help train deeper neural networks and maintain a version of the original input features. Residual dense networks have lately shown that accumulating features through local and global residual learning can significantly improve image restoration (Zhang et al., 2021). Finally, GANs, are a recent addition to the neural network family which are designed to generate data in a semi-supervised setting. The primary reason for the success of GANs, however, is its robust loss function which inherently accommodates several noise removal objectives like L1 distance, distribution matching etc. (Yang et al., 2018). This advantage proves essential in denoising tasks where noisy inputs have to be mapped to clean outputs.

In each of the above architectures, a state-of-the-art model has been identified and used for comparison against the traditional filter-based denoising techniques. Noise-to-clean is a U-Net proposed by Lehtinen et al. (2018) which learns to minimize a mean absolute error (MAE) distance between the denoised image and its corresponding clean reference. Similarly, residual dense network by Zhang et al. (2021) is a residual network aimed at improving fine-grained features in the restoration of noisy images. Cycle Consistent Generative Adversarial Denoising Network (CCGAN) by Kang et al. (2019) is multi-stage GAN network originally designed to denoise coronary CT images. The three models mentioned above use a supervised approach where they extract the 1:1 mapping between noisy and clean pixel intensities. However, in practice, clean reference MCT images are costly and hard to obtain, since this equates to higher X-ray exposure, longer scan durations and other noise prevention techniques which may not be cost- and/or time-efficient. For this reason, an evaluation involving unsupervised machine learning approaches is essential for practical purposes. In this direction, noise-to-noise by Lehtinen et al. (2018) is a reference-less approach based on a U-Net architecture and used for removing common noise forms present in images. Similarly, N2V is another unsupervised technique introduced by Krull et al. (2019), which uses pixel manipulations to infer the presence and removal of noise. Noise-to-noise ratio is a unique model which tries to estimate the unvarying signal behind varying noise realizations. This is intuitively similar to removing moving tourists in an unchanging scenic photo. The training for N2N requires the addition of a synthetic noise profile which matches low dose signal. For our purpose, we use a mixture of Poisson-Gaussian along with Perlin noise textures to simulate random noise. For the Poisson-Gaussian mixture, we first simulate the Poisson process using the pixel intensities as the mean (λ). We then scale this value with a Gaussian factor using mean (μ) of 0 and standard deviation (σ) that is randomly picked between 0 and 0.005. For the Perlin noise, we use a simplex model with

scale set to 5, octaves to 6, persistence to 0.5, and lacunarity to 2.0 which showed the best match for the texture. Such profiles have been shown to emulate low-exposure noise in CT images (Lee et al., 2018). The model then learns to estimate the mean of the signal through several iterations. For test purposes, we do not add any noise and since the model is already accustomed to removing similar noise profiles, the denoising is successful.

Fully supervised and fully unsupervised DL models account for the best and worst-case scenarios of data availability, respectively. In this study, we propose new semi-supervised denoising models, where we use limited number of HQ high-exposure images to denoise a larger LQ low-exposure dataset. This can be applicable in several scenarios, including: (1) when the full length of a core sample needs to be scanned at a high resolution, yet resources in terms of time, availability, and money, are limited, and (2) when the processes or state of the core sample we are interested in visualizing changes quickly, and a fast, low exposure scan is deemed necessary (Gao et al., 2017), and (3) to avoid or minimize drift errors resulting from increased temperature of the experimental setup during longer scans, which leads to deviation of a feature from its original location due to mechanical and thermal stresses (Probst et al., 2016). Therefore, we extend the use of the N2N model to accommodate three HQ to LQ ratios where N2N25 is the scenario where 25% of the LQ data has a corresponding HQ reference image while 75% has no clean reference. Similarly, N2N50 and N2N75 indicate scenarios with a higher percentage of clean reference images. In the literature, fully supervised models use MAE between pixel intensities (L1 loss) to learn the mapping, whereas models like N2N and N2V, utilize a mean squared distance error (L2 loss). We propose a combination of the losses to reflect our semi-supervised task in training. For N2N25, we scale the L1 loss with a 0.25 weight and the L2 loss with a 0.75 weight. Similarly, for N2N50 and N2N75, we modify the weights to reflect the percentage of clean references. Using either L1 or L2 only yielded lower PSNR and SSIM values. **Table 4** details the models and their specifications.

For the DL models, we split the training and testing data using a 80–20% non-overlapping split where 286 slices are randomly sampled from each of the LQ and HQ datasets and reserved for testing and validation, while the remaining 1,140 slices are used for training. The slices are reshaped to 512×512 -pixel resolution using bicubic interpolation (Keys, 1981) in order to accommodate GPU memory limitations. After the model has been trained and evaluated, the original 800×800 resolution is obtained through similar bicubic interpolation. All the models are trained in accordance with their original implementation. Some hyperparameters are optimized to obtain the best possible performance from the model. Further details about the architecture, input requirements and choice of hyperparameters like batch sizes, learning rate and number of epochs are provided in **Table 4**. We perform training over 200 epochs and use the ADAM optimization algorithm for all the models. We pick the epoch with the lowest validation PSNR in order to perform testing. Upon the completion of training, the evaluation is performed on the held out 286 slices from the test set. For all the models, ADAM optimization was used (Kingma and Ba, 2015).

Standard Quantitative Denoising Evaluation Metrics

Traditionally, image degradations that take place during acquisition, reconstruction, processing, and storage have been evaluated qualitatively by visual inspection and subjective assessment based on the assumption that human visual perception is highly adapted to extract structural information (Wang et al., 2004). However, as more data is being generated and more complex structures are being studied, it becomes increasingly difficult to rely on just human inspection. Thus, efforts to automate image quality assessment have been undertaken since 1972, when Budrikis developed one of the first quantitative measures to predict perceived image quality (Budrikis, 1972). These metrics have since evolved to match modern imaging and perceptual standards.

Peak Signal to Noise Ratio (PSNR)

Peak signal to noise ratio is a comparative metric widely used to determine noise degradation in an image with respect to a clean reference. Peak signal to noise ratio is an expression for the ratio between the maximum possible value (power) of a signal and the power of distorting noise that affects the quality of its representation. It is measured in decibels (dB) and the higher the PSNR, the better the image. Theoretically, PSNR for the HQ image is not defined and so it is manually set to 100 dB. In our setting, we calculate the PSNR on the test set which has not been used for training the models. The PSNR value reported for a model is averaged over the previously 286 held-out slices from the test set. The PSNR of the LQ slices act as a baseline for comparison.

Structural Similarity Index Measure (SSIM)

Unlike PSNR, SSIM is a perceptual metric which is not based on pixel intensities but rather on structural similarity between the denoised and the clean reference images (Wang et al., 2004). This is especially important when trying to evaluate fine-grained details and edges. It ranges from 0 to 1, where higher SSIM indicates a cleaner image. Similar to PSNR each of the SSIM values are calculated and averaged. The equation for the SSIM calculation is detailed in the **Appendix** in Supplementary Material.

Blurring Index

Many denoising algorithms focus on removing high-frequency components which correspond to noise. However, this leads to blurry images with fuzzy edges and high smoothness. Blurring index penalizes such images with a lower blurring index score. Occasionally, some DL models, smoothen regions based on expected intensity values leading to the removal of sharp high frequency noise elements. Though this is beneficial, the models obtain lower blurring index (BI) scores. The implementation of BI is detailed in the **Appendix** in Supplementary Material.

Contrast to Noise Ratio (CNR)

In grayscale images, a vital requirement of restoration models is the ability to separate foreground region intensities from the background noise. Contrast-to-noise ratio (CNR) is a reference-less metric which yields a higher value for higher

TABLE 4 | Parameters selected for the DL-based denoising algorithms used in this study.

| Model | Architecture | Input Requirement | Selected Epoch* | Batch Size | Initial Learning Rate | Learning Rate Algorithm | Loss Function |
|---------|------------------------|-------------------|-----------------|------------------------|-----------------------|-------------------------|-------------------------------------|
| CCGAN | GAN | LQ, HQ | 198 | 4 slices | 0.001 | Plateau | $L_{Adv} + L_{Oyc} + L_{Iden}$ |
| RDN | Dense residual network | | 136 | | 0.0001 | Step | L1 (MAE) |
| N2C | U-Net | | 167 | 8 slices | 0.001 | Exponential | |
| N2N-75% | | LQ, 75% HQ | 177 | | | | $(0.75 \cdot L1) + (0.25 \cdot L2)$ |
| N2N-50% | | LQ, 50% HQ | 172 | | | | $(0.50 \cdot L1) + (0.50 \cdot L2)$ |
| N2N-25% | | LQ, 25% HQ | 138 | | | | $(0.25 \cdot L1) + (0.75 \cdot L2)$ |
| N2N | | LQ | 195 | | | | L2 (MSE) |
| N2V | | | 171 | 32 (128 × 128) patches | | Linear | |

*Selected epoch is based on the lowest validation PSNR.

contrasts and lower values for indistinguishable foreground and background regions. The higher the CNR, the cleaner the image, which makes it easier to segment the different phases. Our implementation follows Shahmoradi et al. (2016), where we identify six key regions of interest (ROI) including a background region from 10 random slices in the denoised test set. A detailed explanation of the CNR formula is provided in the **Appendix** in Supplementary Material.

Blind/Reference-Less Image Spatial Quality Evaluator (BRISQUE)

Blind/reference-less image spatial quality evaluator (BRISQUE) helps understand the presence of noise in an image without the use of a reference (Mittal et al., 2011). It relies on spatial statistics of the natural scene (NSS) model of locally normalized luminance coefficients in the spatial domain, as well as the model for pairwise products of these coefficients in order to extract spatially dense variances which occur in noisy images. We use the publicly available PyBRISQUE-1.0 software package (2020) for the implementation of this metric (<https://pypi.org/project/pybrisque/>).

Naturalness Image Quality Evaluator (NIQE)

Naturalness image quality evaluator (NIQE) unlike BRISQUE computes distortion specific features such as ringing, blur, and blocking to quantify possible losses of “naturalness” in the image due to the presence of distortions, thereby leading to a holistic measure of quality (Mittal et al., 2013). For this, we use the skvideo.measure.niqe package from the Scikit-Video library version 1.1.11 (Skvideo, 2013; <http://www.scikit-video.org/stable/modules/generated/skvideo.measure.niqe.html>). Each of the 286 slices are individually passed through the function and the scores are averaged for each model.

Physics-Informed Quantitative Denoising Evaluation Metrics: Petrophysical Characterization

Physics-informed denoising evaluation is performed through the estimation of petrophysical properties extracted from image data. To estimate these petrophysical properties, the denoised images

TABLE 5 | Features selected for segmentation for the different datasets using ilastik.

| | $\sigma 0$ | $\sigma 1$ | $\sigma 2$ | $\sigma 3$ | $\sigma 4$ |
|---------------------------------|------------|------------|------------|------------|------------|
| Feature | 0.3 | 0.7 | 1 | 1.6 | 3.5 |
| Color/Intensity | | | | | |
| Gaussian smoothing | ✓ | ✓ | ✓ | ✓ | ✓ |
| Edge | | | | | |
| Laplacian of Gaussian | ✓ | ✓ | ✓ | ✓ | ✓ |
| Gaussian gradient magnitude | ✓ | ✓ | ✓ | ✓ | ✓ |
| Difference of Gaussians | ✓ | ✓ | ✓ | ✓ | ✓ |
| Texture | | | | | |
| Structure tensor eigenvalues | ✓ | ✓ | ✓ | ✓ | ✓ |
| Hessian of Gaussian eigenvalues | ✓ | ✓ | ✓ | ✓ | ✓ |

need to be segmented to extract phase labels. In order to highlight the effect of different denoising approaches, and their effect on image-based rock and fluid property estimates, we used the same segmentation method on all denoised data. The segmentation method used was Ilastik¹, a supervised machine-learning based segmentation tool that is built using a fast random forest algorithm (Sommer et al., 2011). The user manually provides a limited number of voxel labels for training. We maintain the same features among the different denoised datasets (**Table 5**). An almost equal number of marker voxels were selected for each of the three phases (solid/glass, brine, and scCO₂). Also, the number of markers was kept to a minimum (~0.005% of sample voxels) to avoid overfitting. The same number of marker voxels per slice is maintained across all the datasets to minimize bias.

The use of quantitative denoising evaluation metrics is a good practice over qualitative and subjective assessment of image quality. However, a HQ image as determined by standard denoising evaluation metrics may not necessarily render the most accurate estimate of petrophysical properties,

¹Ilastik (2021). Available online at: <https://www.ilastik.org/> (accessed January 4, 2021).

either due to insufficient noise removal, or excessive filtering which results in information loss. Therefore, we evaluate the implemented denoising algorithms based on how various calculated petrophysical properties compare to those calculated using the HQ benchmark dataset. To achieve this, a toolbox was developed in Matlab R2019a to calculate different petrophysical properties, including (1) porosity, (2) saturation, (3) phase connectivity, which is represented by phase specific Euler number (phase Euler number divided by the phase volume), (4) phase SSA, which is calculated as the area of a given phase divided by its volume, (5) fluid-fluid interfacial area, (6) mean pore size, and (7) mean scCO_2 blob volume. We also calculate phase fractions per slice along the length of the core.

RESULTS AND DISCUSSION

In this section, we first describe the denoising results qualitatively by visual comparison of denoised datasets to the HQ ground truth images. Next, we present quantitative comparison results. The first set of quantitative comparison include standard metrics. Finally, we compare the quantitative petrophysical results of the denoised images post segmentation against the ground truth HQ segmented images.

Qualitative Evaluation of Denoising Methods

Figure 1 shows top view cross-sections from the HQ and the LQ data sets. The LQ image is evidently noisier than the HQ image. We have also marked regions to highlight spatial feature differences between the HQ and LQ images and associated implications for petrophysical property estimation. The red region shows that some small solid features can be inaccurately characterized as fluid, resulting in porosity estimation errors, while the green region demonstrates a blurry fluid-fluid interface in the LQ image, which can lead to errors in estimating fluid saturation and petrophysical properties pertaining to fluid-fluid interfaces. Similarly, the blue region shows an example where the solid phase can be falsely characterized as connected in the LQ image, while the HQ image shows some disconnection of the solid phase. This could result in errors in topological characterization, such as Euler number calculations for the solid phase.

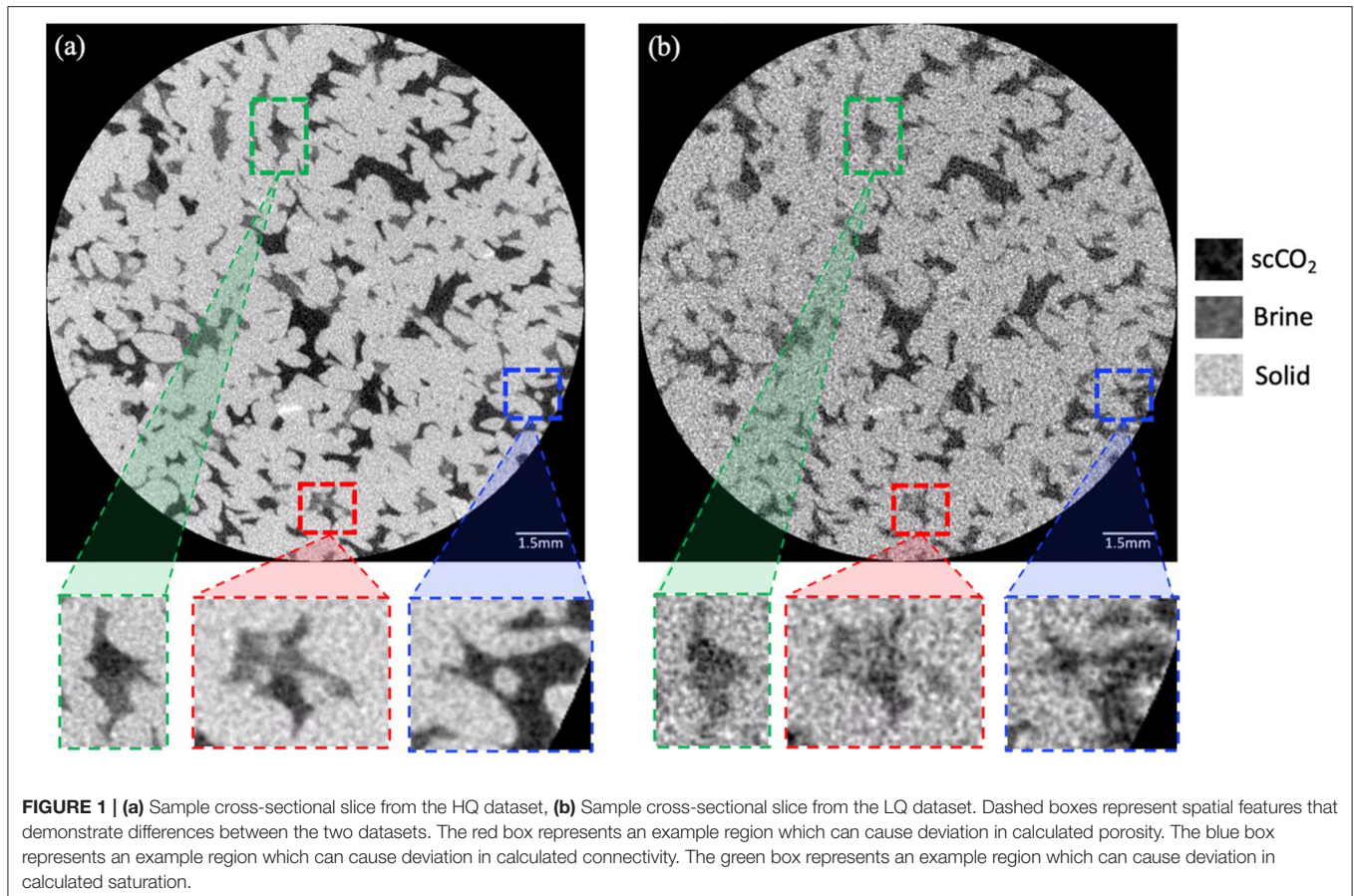
A cross-sectional image for each of the HQ dataset, LQ dataset and the denoised datasets are shown in **Figure 2**. Visually, the NLM filter results in an overall smoother image with significantly less variation within each phase (solid, brine, and scCO_2), as it replaces the value of each voxel with the mean of similar voxels within the search window. The median and SNN filters also visually exhibit smoother phases, compared to the AD and bilateral filters. Conversely, the Gaussian filter expectedly results in image blurring, which can highly affect the accuracy of calculated interfacial petrophysical properties, such as interfacial area and mean interface curvature. The reference-based DL methods qualitatively show good performance compared to traditional filters. Noise-to-clean and CCGAN show a superior performance over RDN, with smoother phases and sharper

edges. However, comparing the two models (N2C and CCGAN) with the HQ reference image, CCGAN resembles the reference image more closely in some features (highlighted in the blue circles), while N2C resembles the reference image more closely in other features (highlighted in the green circles). This exemplifies that qualitative or visual evaluation of denoising models is not a sufficient means of evaluation when deciding on the best denoising model to use. Reference-less DL methods (N2V and N2N) show little improvement compared to the LQ dataset. Finally, the semi-supervised DL methods show visually promising results, where the N2N-75% and N2N-50% show significant improvement compared to LQ.

The histogram representing the distribution of gray-scale intensities within a dataset can be used to qualitatively assess image quality. In our dataset, where we expect to see three phases—solid glass, brine, and scCO_2 , an ideal histogram would have three well-defined peaks with varying maximum voxel count, depending on the abundance of the phase represented by that peak in the imaged sample. The more distinguishable those peaks are, the sharper the phase boundaries and interfaces, and the easier the segmentation. Also, the smaller the spread of each peak, the smoother the bulk phase is (i.e., the phase is represented by less gray-scale values indicating a more homogeneous or smooth bulk phase). In **Figure 3**, we show the histograms of the HQ, LQ, and denoised datasets. The solid glass peak has the highest voxel count in all datasets, as expected, being the most abundant phase in our sample. We also observe a clear difference between the HQ and LQ datasets, where we can see three peaks in 3a, but only one visible peak in 3b, which makes the differentiation between scCO_2 and brine more difficult. From the denoised dataset histograms, the supervised and semi-supervised DL denoising models perform better than traditional and unsupervised DL denoising methods. Also, we observe that the issue of fluid-fluid differentiation (where we can't see all three peaks) persists when using the bilateral, N2V, N2N25, and N2N denoising methods. We also observe that some denoising models possibly outperformed the HQ dataset where the three peaks are even more distinguishable, including N2C, RDN, CCGAN, and N2N75. Additionally, upon comparing 3i and 3k with 3j, we can see results consistent with our observations from **Figure 2**, where N2C and CCGAN have more well-defined and narrower peaks compared to RDN. Despite the additional insight obtained from inspecting the gray-scale intensity histograms, histograms alone cannot be used to infer image quality. The histogram indicates the distribution of pixel intensities but not their location. This calls for a more localized inspection of the images to assess how voxel spatial distribution impacts the accuracy of petrophysical property estimates.

Quantitative Evaluation of Denoising Methods

Figure 4 demonstrates a more quantitative method for evaluating the performance of the different denoising models using an intensity profile along a line A-A'. Here, we use N2C as an example. From this profile, we can obtain insights on the (1) smoothing effect within each phase by comparing the plateau



of the noisy LQ dataset to that of the HQ and denoised dataset, and (2) contrast enhancement, which can be assessed by comparing the difference in intensities between the different phases. A larger difference implies an improvement in contrast, and finally (3) we can also assess phase boundary sharpness by calculating the steepness of the slope at each phase boundary. A steeper slope indicates a sharper boundary, which makes it more easily identifiable during image segmentation and petrophysical characterization. In **Figure 4**, we observe that the HQ profile exhibits oscillations within each phase, which shows that noise still exists in the HQ data, but the amplitude of those oscillations is smaller, which implies less noise and more homogeneous phases. We also observe slightly higher average intensity for the solid and brine phases in the HQ profile compared to the LQ profile. Additionally, we observe that N2C has improved the image quality on all three aspects: smoothing bulk phases, improving contrast between phases, and improving boundary sharpness. The scCO_2 -solid phase boundary has the most improvement in boundary sharpness with $\sim 28\%$ increase in slope.

Example regions are highlighted in **Figure 5a**, following the workflow proposed by Shahmoradi et al. (2016) to assess local image quality using CNR. ROI-0 corresponds to the background which is used to calculate the CNR of other regions. ROI-1 corresponds to brine with no interfaces, while ROI-2

corresponds to the scCO_2 phase, and ROI-3 corresponds to the solid phase. In addition, two interface regions, ROI-4, an interface between solid and brine and ROI-5, an interface between solid and scCO_2 are also considered. We pick a random slice and identify ROIs 1–5 in the next 22 consecutive slices. **Figures 5b,c** show localized average CNR values within bulk phases and at solid-fluid interfaces, respectively. We see that all denoising models, except for N2V, result in an improved CNR both in bulk phases and at interfaces. Supervised DL denoising models result in the largest CNR improvement, with N2C being a top performer. The N2N75 model also yields similar results as fully supervised denoising models, which is promising for cases where not all the HQ data is available. We also observe that the N2N25 model results in lower average CNR compared to N2N, which can be counter intuitive. However, this could be due to the inherent training objective of the model. The base N2N model estimates the mean of the signal and uses multiple noisy versions of the same signal to do so. However, when a small number of clean samples are provided, the network's estimate of the mean diverges from the expected value, which negatively affects learning. When provided with enough clean samples (50% or greater), the model converges to the mean more accurately compared to the base N2N or N2N25 versions. Comparing the improvement within bulk phases and at phase boundaries,

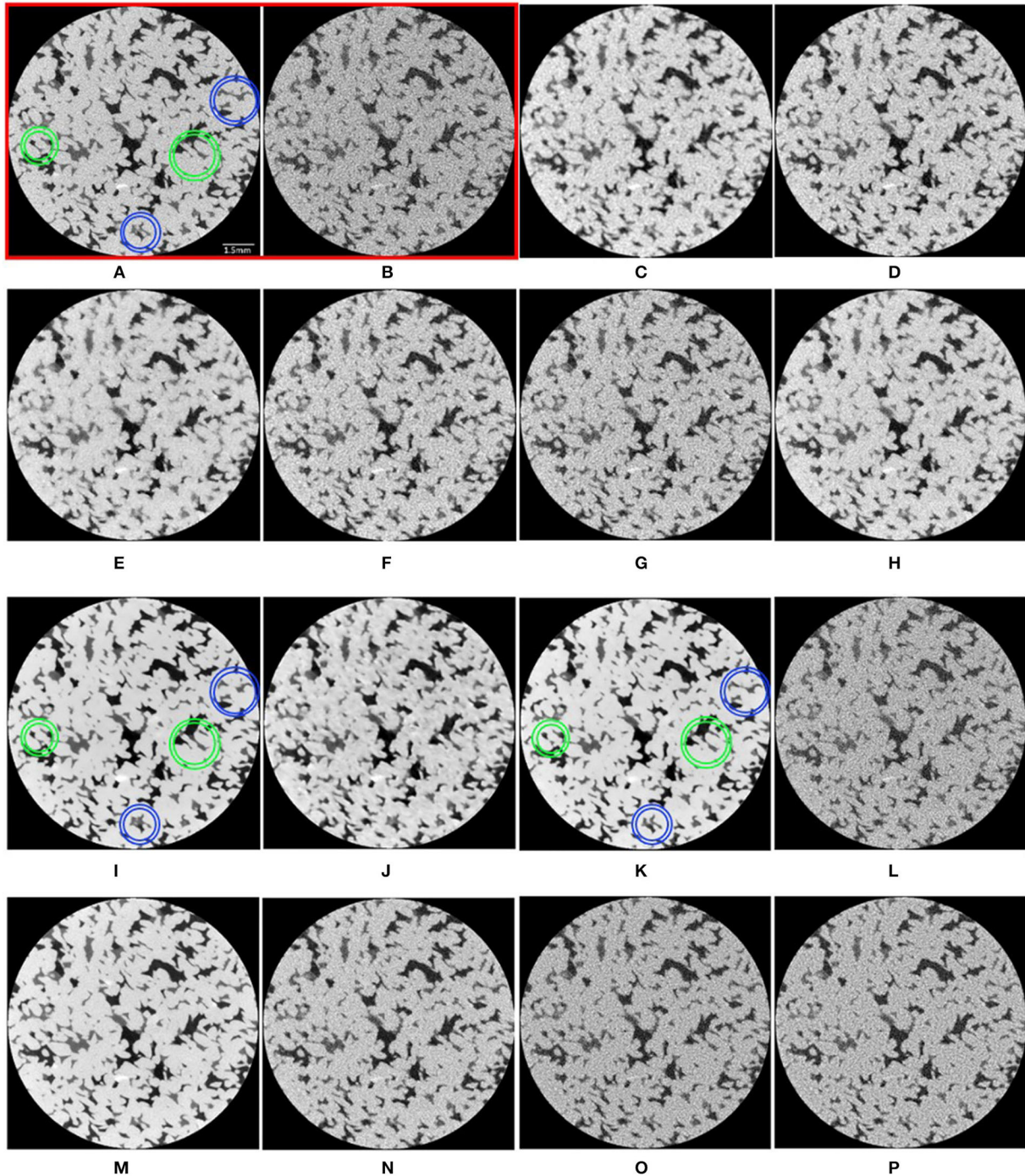
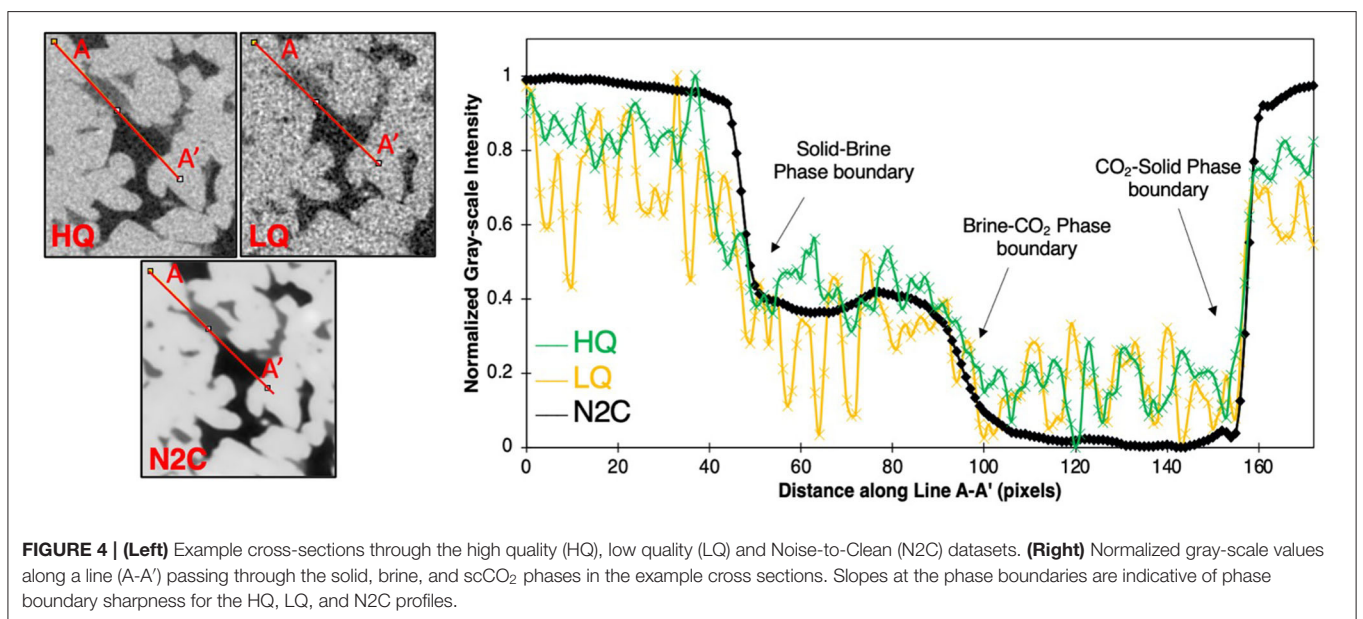
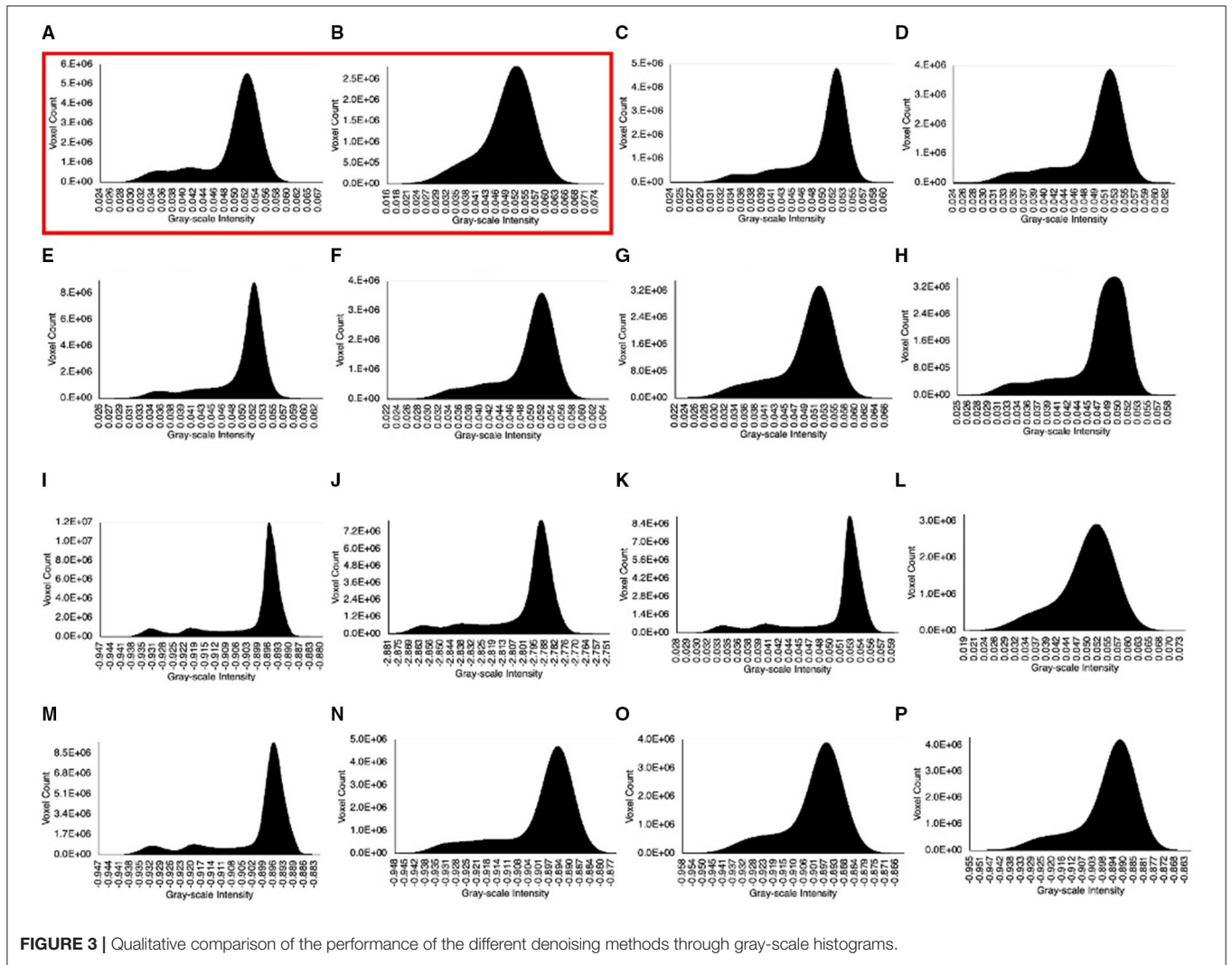


FIGURE 2 | Qualitative comparison of the performance of the different denoising methods through example cross-sectional slices. **(A)**, HQ reference; **(B)**, LQ; **(C)**, Gaussian; **(D)**, Median; **(E)**, Non-local means; **(F)**, Anisotropic diffusion; **(G)**, Bilateral; **(H)**, Symmetric nearest neighbor; **(I)**, Noise-to-clean; **(J)**, Residual dense network; **(K)**, Cycle consistent generative adversarial network; **(L)**, Noise-to-void; **(M)**, Noise-to-noise (75%); **(N)**, Noise-to-noise (50%); **(O)**, Noise-to-noise (25%); **(P)**, Noise-to-noise.

we do not see a significant difference, indicating that all denoising methods improve contrast homogeneously throughout the image. Although quantitative, one limitation of the type

of analysis performed in **Figures 4, 5** is that they evaluate the performance of the denoising models locally, which requires manual selection of specific sites. A comprehensive denoising



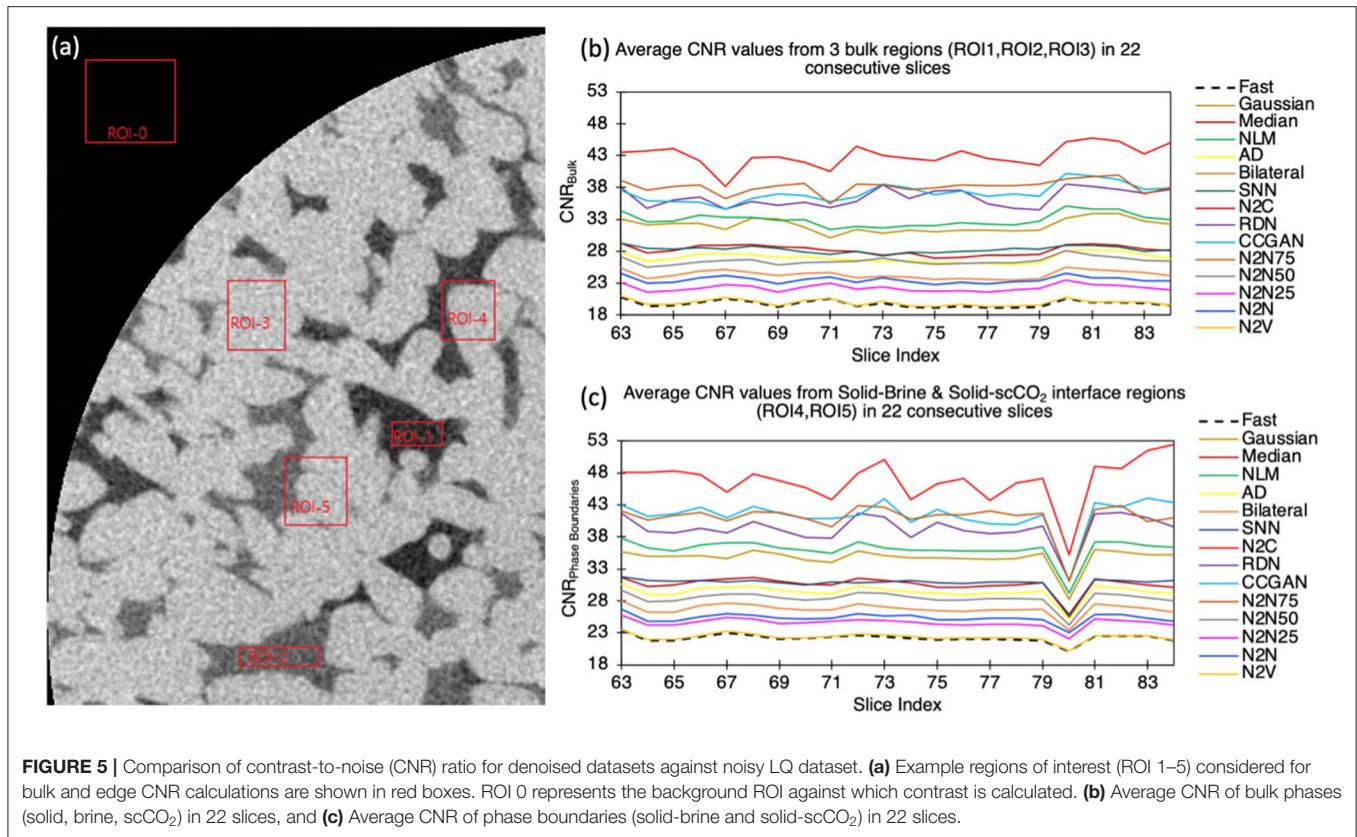


TABLE 6 | Quantitative comparison of denoising algorithms using standard denoising evaluation metrics.

| | Models | PSNR | SSIM | Blur index | CNR | BRISQUE | NIQE | |
|----------------------------|-----------------------|---------------|--------------|--------------|---------------|----------------|---------------|---|
| | HQ Images (Reference) | 100* | 1* | 0.724 | 19.233 | 118.183 | 26.645 | |
| | LQ Images | 18.224 | 0.507 | 0.634 | 21.472 | 106.619 | 25.517 | |
| Traditional filters | Gaussian | 23.378 | 0.769 | 0.783 | 34.584 | 116.833 | 26.875 | <div style="display: flex; align-items: center; justify-content: center;"> <div style="width: 10px; height: 100px; background: linear-gradient(to top, green, yellow, orange, red);"></div> <div style="margin-left: 5px; text-align: center;"> <p>Good Performance</p> <p>Poor Performance</p> </div> </div> |
| | Median | 24.593 | 0.716 | 0.753 | 30.504 | 118.237 | 27.173 | |
| | NLM | 25.396 | 0.760 | 0.715 | 36.018 | 116.948 | 26.900 | |
| | AD | 25.019 | 0.700 | 0.735 | 29.326 | 118.336 | 27.287 | |
| | Bilateral | 24.271 | 0.652 | 0.706 | 26.384 | 117.239 | 27.164 | |
| | SNN | 22.354 | 0.677 | 0.789 | 30.139 | 118.065 | 26.713 | |
| Fully supervised DL models | N2C | 24.702 | 0.798 | 0.752 | 46.517 | 116.387 | 26.874 | |
| | RDN | 25.397 | 0.785 | 0.730 | 40.313 | 116.443 | 26.902 | |
| | CCGAN | 24.716 | 0.795 | 0.728 | 39.702 | 116.536 | 26.875 | |
| Semi-supervised DL models | N2N-75% | 25.332 | 0.800 | 0.726 | 40.413 | 116.774 | 26.884 | |
| | N2N-50% | 24.518 | 0.716 | 0.711 | 28.356 | 118.799 | 27.325 | |
| | N2N-25% | 23.084 | 0.636 | 0.687 | 24.019 | 117.427 | 26.860 | |
| Un-supervised DL models | N2N | 23.339 | 0.650 | 0.718 | 25.099 | 117.707 | 26.981 | |
| | N2V | 20.607 | 0.546 | 0.662 | 21.569 | 112.048 | 26.356 | |

Colors represent how well a model performs compared to the rest of the models for a given denoising evaluation metric. *PSNR and SSIM are calculated with respect to HQ reference and thus are set to 100 and 1 for HQ inputs, respectively.

model performance evaluation is quantified using standard evaluation metrics detailed in **Table 6**.

The CNR value reported here represents the average of the five individual CNR regions detailed earlier (**Figure 5**).

Of the six metrics measured, the DL models performed the best in five of them. With Blur Index however, the metric measures the presence of sharpness which can be spiked by high frequency noise. The traditional filters failing to remove such

noise achieve a misleading high blur-index score. Though the unsupervised models performed poorly, they were comparable to the traditional filter-based methods. Noise-to-noise ratio had the lowest values across all the metrics. It is also observed that PSNR values are generally low across all denoising methods. However, PSNR uses pixel distance and may incorrectly favor models that have higher resemblance to the HQ images rather than actual signal accuracy. The DL models tend to produce images that appear cleaner than the HQ image but score lower PSNR due to this reason. Residual dense network had the highest PSNR gain with an average increase of 7.173 dB. Though it might be possible for an image to obtain a high PSNR, SSIM penalizes structural anomalies like blurring, noise contamination, wavelet ringing, and blocking (caused by image compressions like JPEG) leading to lower SSIM scores.

Each evaluation metric points to unique capabilities of the models. For example, a high CNR obtained for N2C makes it ideal for segmentation tasks while a high SSIM obtained through N2N-75% model makes it favorable for CT image artifact removal task. The differences observed, however, between supervised DL-based models and N2N-75% in terms of SSIM are not statistically significant, meaning that they all perform equally well for this metric. N2N-50% produced the most natural looking images (NIQE) and can be used to generate additional synthetic clean samples when only 50% of clean data is available.

We also evaluated the denoising methods using physics-based metrics, or more specifically—petrophysical properties. To perform the petrophysical characterization, we first segmented the datasets. **Figure 6** shows example segmented cross sections from the HQ and LQ datasets. It also shows a difference image, where the segmented LQ image was subtracted from the segmented HQ image. The percent mislabeling for each phase pair: brine-solid, scCO₂-brine, and scCO₂-solid is 2.46, 0.37, 1.57%, respectively. Based on the percent mislabeling of the denoised images, we see minor improvement compared to the LQ image. Similarly, image subtraction was performed between each of the denoised images and the HQ image. Noise-to-clean and N2N75 showed the most improvement especially in CO₂-solid and brine-solid phase mislabeling. The differences are presumed to be mainly due to the difference in the amount of noise in each image since the user interference was kept to a minimum in the segmentation step. Further, the differences within this single image might seem subtle, but it has a significant effect on petrophysical characterization performed using the entire image stack. The full petrophysical characterization and error quantification against the ground truth is presented in **Figure 7** and **Table 7**.

For petrophysical property estimations, we first consider bulk measures of phase fractions and saturations along the length of the core sample. **Figure 7** shows the percent errors in porosity, scCO₂ saturation, and brine fraction (measured as brine volume divided by the bulk volume for the sample space considered) profiles along the length of the sample for the different denoising methods post segmentation. The performance of denoising methods, namely, traditional methods, supervised, unsupervised, and semi-supervised DL-based methods are compared against

the HQ dataset. Generally, most of the traditional and semi-supervised DL-based methods show closer results to the ground truth (**Figure 7A**). Out of the traditional filters, the median filter seems to have the closest porosity profile to the ground truth HQ dataset, followed by the AD and bilateral filters. Conversely, the Gaussian filter performed poorly, consistently underestimating porosity. Similarly, NLM, which is one of the most commonly used filters in the digital rock physics literature, underestimates porosity throughout the length of the core and has the second worst performance. Unlike their performance in terms of standard evaluation metrics (**Table 6**), supervised methods show poor performance, consistently underestimating porosity along the entire length of the sample. Additionally, the error is increasing toward the bottom of the sample in most denoising methods, especially Gaussian, RDN, and N2V methods. Upon inspecting the root cause of this, the original LQ dataset was noisier toward the bottom of the sample as evidenced by lower PSNR (**Supplementary Figure 1A**). This is mainly due to the aluminum container, covering the bottom half of the sample, used to fix the coreholder in place during the scan and avoid wobbling which can result in image blurring (**Supplementary Figure 1B**). The Gaussian filter, which is a linear spatially invariant filter, cannot resolve the noise variation along the sample. Residual dense network, a very deep network, could be overfitting the data and therefore less able to adjust to the varying noise along the sample, compared to other supervised models (N2C and CCGAN).

When comparing the CO₂ saturation error profiles (**Figure 7B**), it is observed that traditional methods show close results, while most of the DL-based methods show poor performance when compared to the ground truth. Out of the traditional filters, SNN followed by median filters showed the most deviation from the ground truth. The Gaussian filter, on the other hand, had the closest results out of all the denoising models. Residual dense network and N2N significantly overestimate CO₂ saturation. The supervised DL-based methods, except for N2C, show weak performance while the semi-supervised DL-based methods show slightly better performance. Noise-to-noise ratio, as an exception, shows the second-best performance of all methods, while RDN significantly underestimates scCO₂ saturation. Larger error is seen at the top of the sample for all denoising methods except N2C and N2N75.

As the intermediate phase, brine fraction error profile is presented as opposed to brine saturation to account for errors that may result from misclassification of brine as solid or CO₂ post-denoising (**Figure 7C**). Noise-to-clean, followed by NLM show the least deviation compared to the ground truth, whereas RDN, CCGAN, and N2N show a significant deviation. Noise-to-noise ratio also shows a surprisingly close match to the ground truth brine fraction profile. Overall, these results demonstrate that even within bulk property estimates, since differences occur at fluid-fluid and fluid-solid interfaces, consideration of both porosity and saturation errors is critical. Again, here we see that N2C and N2N75 are robust enough to maintain consistent performance throughout the sample even though the dataset has a spatially varying noise profile.

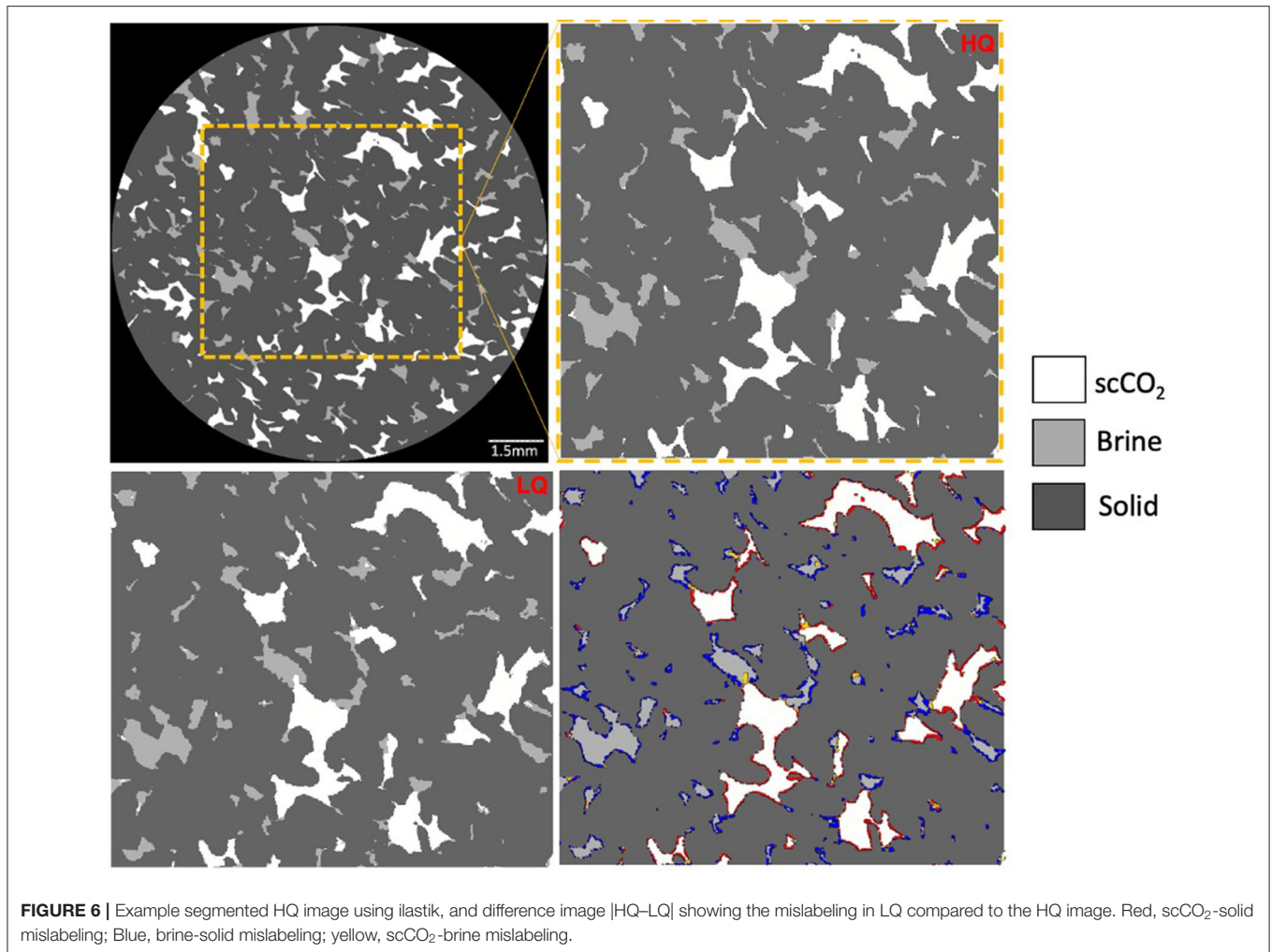


Table 7 summarizes the % errors associated with the estimates of bulk petrophysical properties, namely porosity and scCO₂ saturation, as well as pore-scale properties including, specific Euler number, SSAs, interfacial area, average pore size and average scCO₂ blob volume. These errors are quantified relative to the ground truth HQ images. The actual property estimates for the ground truth HQ images are presented in the first row of **Table 7**. Traditional filters like the median, AD and bilateral filters show an improved porosity estimate compared to the original LQ images, but the Gaussian filter results in the highest error for porosity (13.9%). This can be explained by the blurring caused by the Gaussian filter at phase interfaces, making it more difficult to locate boundaries, especially where the gray-scale intensities are less distinguishable like at the solid-brine interfaces. However, the Gaussian filter shows the best performance in terms of scCO₂ saturation, which may be due to a larger surface area of the scCO₂ phase being in contact with the solid phase rather than the brine phase, hence, enabling us to distinguish between the two phases more easily in terms of gray-scale intensity and labeling. The median filter on the other hand shows an opposite result where it exhibits better performance

when estimating porosity and poorer performance in estimating scCO₂ saturation. This can be explained by how a median filter works, where it involves replacing the gray-scale values of voxels within a user-specified window by their median value. This means that the small scCO₂ blobs observed in the dataset in the bulk brine phase will be replaced with gray-scale values that resemble brine, leading to a lower scCO₂ saturation (compared to HQ), but not affecting porosity as much. In general, most traditional filters perform better for bulk estimates of porosity and phase saturation, compared to DL-based methods.

For the pore-scale properties, we observe that supervised and semi-supervised methods show a better performance overall, except for scCO₂ SSA and average pore size, where bilateral and SNN filters, respectively, show better performance. This can be attributed to the edge preserving nature of both the bilateral and SNN filters. For the scCO₂ SSA, this allows fine-scale details like small scCO₂ blobs observed in our dataset to be preserved, leading to a more accurate representation of phase boundaries and surfaces. Noise-to-clean and N2N75 perform well for most of the properties with $\sim 10\%$ error or less. Other methods like RDN and N2N25 perform poorly with errors as large as 348.28%.

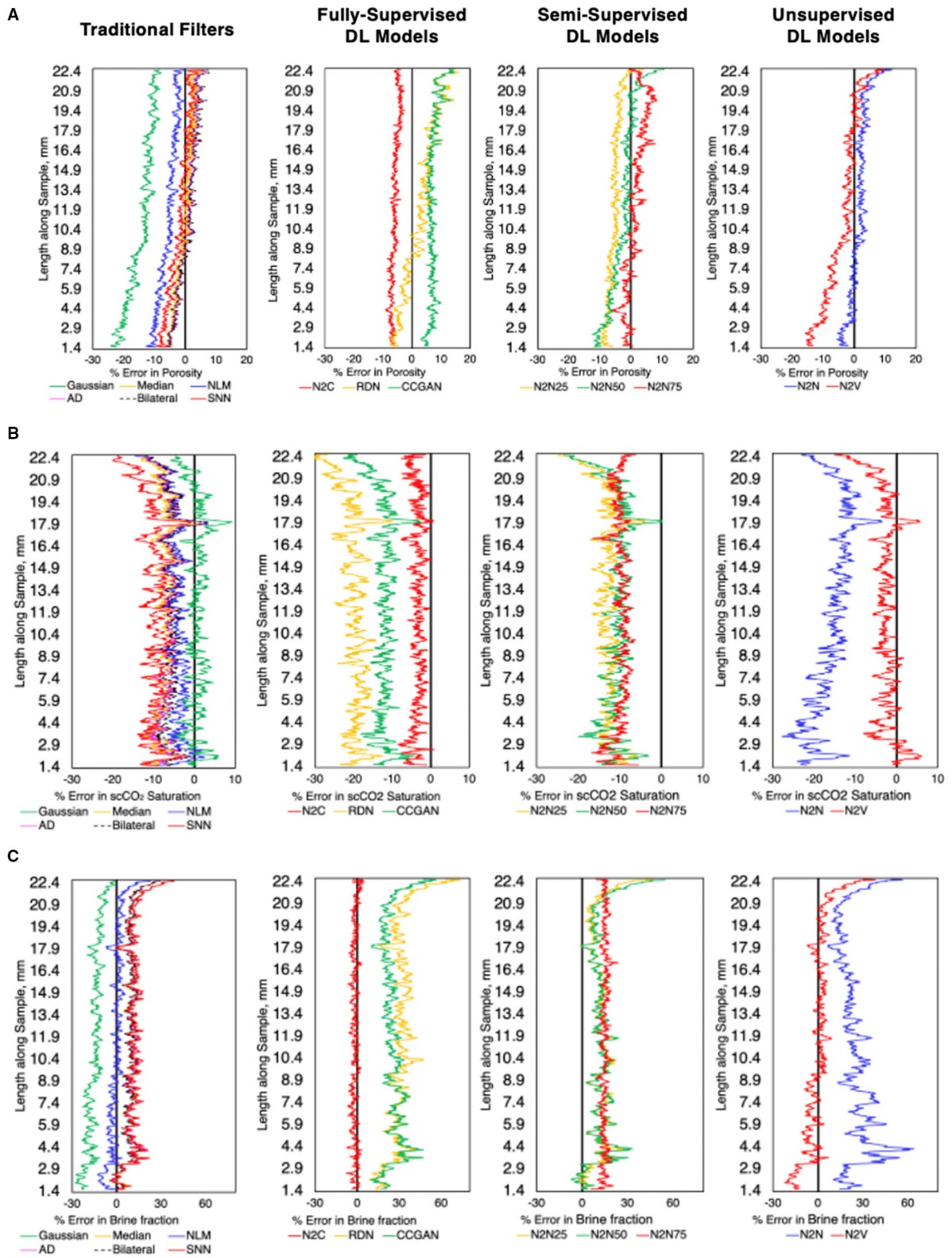


FIGURE 7 | (A) Porosity, (B) scCO₂ saturation, and (C) brine phase fraction | percent error (relative to HQ) along the length of the core for the different denoising methods.

TABLE 7 | Physics-based comparison of denoising algorithms using bulk and pore-scale petrophysical properties.

| | | Porosity, % | scCO ₂ saturation, % | SEN (solid), mm ⁻³ | SEN (pore), mm ⁻³ | SEN (Brine), mm ⁻³ | SEN (scCO ₂), mm ⁻³ | SSA (Solid), mm ⁻¹ | SSA (Brine), mm ⁻¹ | SSA (scCO ₂), mm ⁻¹ | Interfacial area, mm ⁻¹ | Avg. pore size, microns | Avg. scCO ₂ blob volume, mm ³ |
|---------------------|-----------|----------------|---------------------------------------|-------------------------------------|------------------------------------|-------------------------------------|--|-------------------------------------|-------------------------------------|--|--|-------------------------------|---|
| | HQ | 14.940 | 56.750 | -8.709 | -67.164 | 909.459 | 809.306 | 4.469 | 32.115 | 21.593 | 24.619 | 35.778 | 0.002 |
| | LQ | 1.14% | 7.75% | 17.53% | 169.22% | 70.12% | 71.34% | 0.51% | 12.81% | 7.36% | 11.63% | 13.52% | 118.32% |
| Traditional filters | Gaussian | 13.92% | 0.44% | 95.60% | 180.03% | 34.86% | 90.46% | 20.64% | 2.45% | 4.94% | 1.89% | 27.04% | 438.94% |
| | Median | 0.07% | 7.31% | 56.79% | 52.58% | 22.59% | 88.73% | 3.17% | 8.30% | 1.55% | 5.02% | 6.73% | 520.31% |
| | NLM | 5.42% | 4.21% | 48.45% | 97.74% | 11.18% | 96.78% | 9.85% | 3.69% | 3.31% | 2.97% | 21.17% | 795.99% |
| | AD | 0.80% | 6.63% | 13.92% | 143.82% | 4.32% | 86.44% | 2.22% | 10.31% | 2.68% | 8.10% | 9.86% | 314.17% |
| | Bilateral | 1.00% | 5.92% | 5.59% | 118.58% | 11.30% | 81.05% | 2.05% | 8.97% | 0.46% | 6.24% | 5.79% | 297.10% |
| | SNN | 1.67% | 10.20% | 62.59% | 123.40% | 20.61% | 105.95% | 5.63% | 1.09% | 4.53% | 0.77% | 0.28% | 1256.85% |
| Fully supervised | N2C | 5.89% | 3.47% | 6.39% | 7.92% | 11.93% | 57.78% | 7.39% | 1.73% | 2.77% | 0.76% | 24.55% | 56.40% |
| DL models | RDN | 3.01% | 20.28% | 12.40% | 194.80% | 53.78% | 54.32% | 2.63% | 0.87% | 4.69% | 1.73% | 57.89% | 163.94% |
| | CCGAN | 7.16% | 12.25% | 2.49% | 41.32% | 49.34% | 64.49% | 2.62% | 11.35% | 12.16% | 12.50% | 59.07% | 55.36% |
| Semi-supervised | N2N75 | 1.47% | 9.96% | 1.37% | 65.06% | 2.78% | 47.54% | 0.32% | 4.25% | 10.40% | 1.76% | 8.77% | 7.15% |
| DL models | N2N50 | 2.54% | 10.96% | 91.08% | 108.44% | 10.42% | 90.26% | 15.29% | 8.10% | 3.06% | 8.01% | 116.57% | 569.64% |
| | N2N25 | 5.29% | 13.02% | 108.40% | 143.58% | 86.10% | 74.32% | 15.87% | 16.29% | 7.60% | 15.40% | 68.18% | 348.28% |
| | N2N | 1.61% | 15.75% | 96.86% | 128.31% | 3.31% | 93.10% | 10.92% | 14.85% | 2.06% | 11.58% | 71.72% | 549.56% |
| Un-supervised | N2N | 1.61% | 15.75% | 96.86% | 128.31% | 3.31% | 93.10% | 10.92% | 14.85% | 2.06% | 11.58% | 71.72% | 549.56% |
| DL models | N2V | 3.61% | 2.85% | 93.15% | 161.36% | 117.84% | 93.12% | 16.64% | 10.92% | 2.87% | 9.89% | 35.40% | 345.72% |



Values represent the absolute % error in the estimate of each petrophysical property compared to the ground truth (HQ images), except for the first row, which represents the actual property values for the segmented HQ images. The best denoising model for each property is highlighted with bold borders. SEN, specific Euler number; SSA, specific surface area.

N2N75 also shows a more superior performance over N2C in terms of pore-scale estimates, which might not be expected due to the availability of more clean reference data to train the model for N2C, and the better overall standard metrics performance of N2C as shown in **Table 6**. However, it should be noted that the higher performance of N2N75 observed in **Table 7** can be attributed to being a closer representation of the HQ data, which in our study is used as a ground truth. The HQ data is not an idealistic ground truth data, it has its own noise associated with it. Therefore, the comparison between N2C and N2N75 is limited by the amount of noise that is present in the HQ data which is the reference of comparison. We also observe one exception where RDN performs better than N2C and N2N75 in terms of brine SSA. This can be attributed to the ability of the RDN model to preserve high frequency details, which translates to a better ability to preserve high contrast phase boundaries and surfaces where SSA is calculated (Zhang et al., 2021).

Unsupervised DL-based methods (N2N and N2V) generally exhibit poor performance for most properties compared to other denoising methods including some of the traditional filters. This is in line with previous observations made during the analysis of the standard metrics in **Table 6**. Semi-supervised models, as expected, are performing better than unsupervised models. Counter-intuitively, adding additional information in low quantities (e.g., N2N25) can sometimes affect the model's capability to converge, leading to larger errors, as seen when we compare N2N and N2N25 errors for most properties. As we continue to add further information, the model manages to converge at the best possible parameters, which can also be seen for most properties when we compare N2N25, N2N50 and N2N75. We believe this is possible only after an optimum threshold percentage of clean reference images are provided.

To decide which denoising methods perform better in terms of bulk properties (porosity and scCO₂ saturation) and pore-scale properties (SEN, SSA, interfacial area, average pore size and average blob volume), we sum up the absolute errors for each of the denoising methods for those two groups of properties. We also sum up all errors in **Table 7** to identify the overall accuracy of each denoising method in terms of combined petrophysical characterization. Additionally, an aggregated standard evaluation metric is calculated by summing up the values in **Table 6** for each method. The four aggregate metrics are presented in **Supplementary Table 1**. It is clear from **Supplementary Table 1** that the denoising method selection should be highly dependent on the type of analysis of interest. For example, a denoising method that can offer accurate bulk properties might not be able to provide accurate pore-scale properties and vice versa. Aggregating the error percentages point out that N2N75 has the most consistent performance with the lowest overall error across both bulk and pore-scale properties while SNN has the highest error in estimating pore-scale properties and combined petrophysical properties. Noise-to-clean comes in a close second, however, it requires the complete HQ reference dataset. Among the traditional filters, the bilateral filter was the best performer across all aggregated metrics.

Computational Requirements of DL-Based Denoising Methods

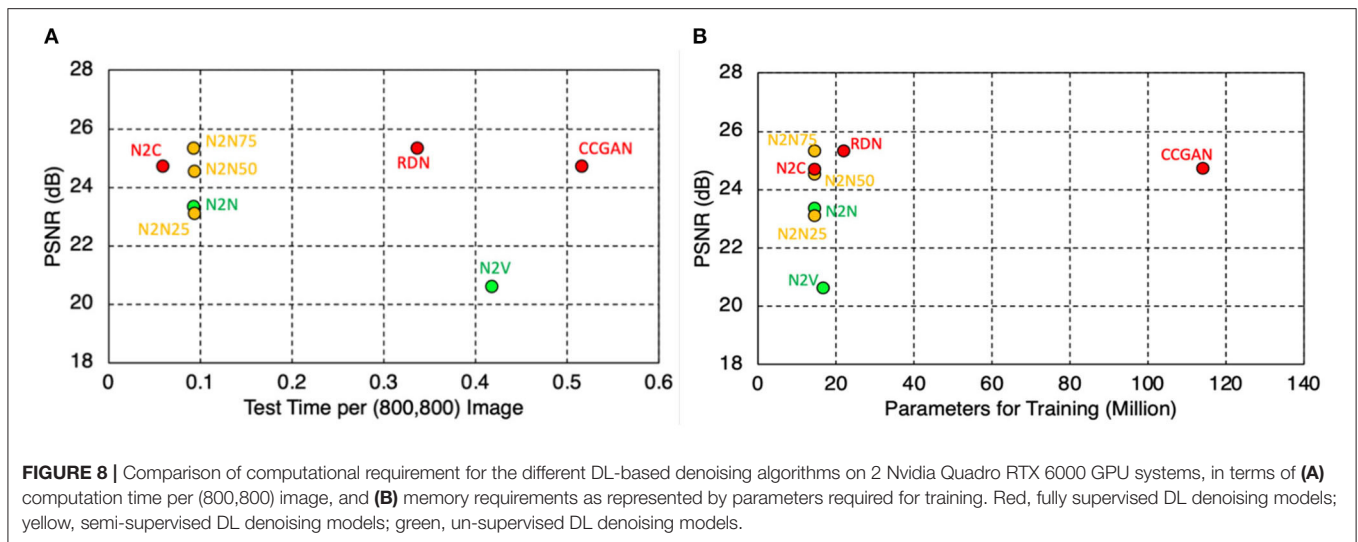
Another important factor to consider when selecting a DL-based denoising model is computational requirements. **Figure 8** shows the comparison of computational expenditure for the different DL-based models. Specifically, we compare the time required to process a single image slice of 800×800 pixels and the memory requirement which is computed from the parameters required for training. The system hardware available for implementing all models was the same (two Nvidia Quadro RTX 6000 GPU systems). Here, the comparison is presented against the traditional metric of PSNR to show how the models perform with the use of the required computational resources. We find that the supervised models, which show higher performance for PSNR, have substantially different computational needs, where RDN and CCGAN are approximately 5.6 and 8.6 times slower than N2C. Similarly, CCGAN is also found to be the most memory intensive primarily due to the training requirements of four individual networks while all other models share similar memory requirements ($\sim 1/6^{\text{th}}$ that of CCGAN). When comparing the semi-supervised models, all three models (N2N25, N2N50, and N2N75) require the exact same computational resources both in terms of processing time and memory requirements. Finally, the unsupervised models (N2V and N2N) require similar memory requirements, but N2V is approximately four times slower than N2N. This could be attributed to the original PIP² based implementation of the model. Overall, CCGAN is found to be the least computationally efficient, whereas N2C, N2N%, and N2N are the most efficient.

Denoising Method Recommendation

Denoising is a mainstay pre-processing step performed to improve our ability to detect features and perform more accurate quantitative analyses using CT images. Comparing denoising methods through standard metrics and petrophysical property estimates helps us understand the use-cases in which certain models perform better. For example, despite performing well in terms of standard evaluation metrics, supervised DL models like N2C and RDN rely upon the availability of clean HQ data and such models cannot be used when such reference data is scarce or not available. Similarly, downstream tasks also determine the choice of denoising model. N2N50 can be recommended for generative tasks where additional synthetic data is required while N2C can be used if image segmentation is the preceding step. Additionally, another important consideration is the availability of computational resources. Deep learning-based models require GPU-based systems to perform the training efficiently while traditional filters only require basic computational resources.

Based on the results we find that there is no one denoising method that ultimately performs better in all cases. Micro-computed tomography users are encouraged to adopt a similar evaluation workflow to derive an optimum denoising method that works best for their dataset(s) and use case. However, we present a high-level denoising model recommendation

²PIP (2021). Available online at: <https://pip.pyqa.io/en/stable/> (accessed January 4, 2021).



in **Supplementary Table 2**, based upon the different factors analyzed in this study. Though these factors influence the choice of a model individually, we list combinations of those factors to offer model recommendations for different scenarios. We derive the recommendations from **Supplementary Table 1** and **Figure 8** where we illustrate the superiority of some algorithms over others.

CONCLUSIONS AND FUTURE WORK

Micro-computed tomography image artifacts and alterations resulting from using a non-tailored denoising protocol can result in inaccurate image-based characterization of porous media. In this paper, we present a comprehensive comparison of the performance of various image denoising methods which are broadly categorized as traditional (user-based) non-learnable denoising filters and DL-based methods. The DL-based methods are further sub-categorized depending on their reliance on reference data for the training process. Common architectures are used for supervised and unsupervised methods. We also proposed new semi-supervised denoising models (N2N75, N2N50, and N2N25) to assess the value of information and explore whether faster, lower exposure MCT images can partially substitute high-exposure datasets, which can be costly and can also hinder our ability to capture phenomena that occur at smaller time scales. The datasets used are those of a porous sintered glass core that has been saturated with brine and scCO_2 . We use both qualitative and quantitative evaluations to compare the performance of different denoising methods. For the latter, we consider standard denoising evaluation metrics, as well as physics-based petrophysical property estimates. The following conclusions are derived based on the analysis of our results,

- Commonly used denoising filters in the digital rock physics literature, namely NLM and AD, show reasonable

performance in terms of traditional denoising metrics like PSNR and SSIM. These methods also show reasonable estimates for the bulk petrophysical properties. However, when estimating pore-scale properties like phase connectivity, these methods could result in significant errors. NLM and SNN filters were found to perform least favorably in terms of pore-scale petrophysical estimates.

- Qualitative evaluation is usually used in the digital rock physics literature. However, we show that visual image quality is not sufficient to select an appropriate denoising algorithm.
- The selection of an optimum denoising model cannot solely depend on the visual quality of the denoised image, or even on the standard denoising evaluation metrics. Several factors need to be considered, including the availability of computational resources, and the post-processing analysis of interest. Physics-based petrophysical evaluation metrics are key in selecting a fit-for-purpose denoising method since a superior performance based on standard evaluation metrics may not necessarily indicate a superior performance in terms of petrophysical characterization accuracy. Additionally, there is no single best denoising algorithm across all petrophysical properties.
- The performance of the newly proposed semi-supervised methods, especially N2N75, is very promising considering that less high-exposure data can be used to achieve accurate petrophysical characterization, while significantly reducing scanning time and cost. This can also be useful in cases where the phenomenon being investigated has a short time-scale like chemical processes and pore-scale flow events. It can also help minimize rotational drift errors during scans.
- The unsupervised DL models, in general, showed the weakest performance both on standard and petrophysical evaluation metrics, with N2V giving the least favorable outcomes across most properties of interest.
- N2C (fully supervised) and N2N75 (semi-supervised with 75% HQ data) overall showed the most favorable outcomes.

However, other supervised models like RDN and CCGAN and semi-supervised models like N2N50 and N2N25 showed weaknesses either on traditional metrics, petrophysical estimates, or computational requirements.

This study laid the groundwork for comparing and screening denoising methods for different use-cases within the digital rock physics domain. The following tasks are to be addressed in future work.

- Compare the performance of the different denoising methods using more complex petrophysical properties like surface roughness, tortuosity, interfacial curvature, *in-situ* contact angles, as well as fluid flow parameters such as absolute permeability, relative permeability, and capillary pressure. This comparison can enable optimum image processing workflow selection for creating accurate digital rocks that can be used for multiphase flow prediction and explanation.
- Optimize different denoising methods to accommodate different types of datasets. For that, we would need to first investigate the effect of porous medium structure complexity and image resolution on the performance of the different denoising methods.
- Use idealized simulated or synthetic ground truth reference images and test the effect of different types and levels of noise on the performance of each of the denoising methods, as well as combinations of the denoising methods.
- To identify the effect of noise in the HQ ground truth images on the conclusions drawn in this study, the same study can be conducted while using HQ datasets of varying exposure time.
- Optimize the newly proposed semi-supervised denoising models to determine the optimum threshold percentage of HQ high-exposure images that are needed while maintaining high accuracy of petrophysical analysis.
- Test the hypothesis of whether the sequential use of N2N25 can improve image quality to a point where accurate results are achievable.

REFERENCES

- Al-Menhali, A. S., Menke, H. P., Blunt, M. J., and Krevor, S. C. (2016). Pore Scale observations of trapped CO₂ in mixed-wet carbonate rock: applications to storage in oil fields. *Environ. Sci. Technol.* 50, 10282–10290. doi: 10.1021/acs.est.6b03111
- AlRatrou, A., Raeini, A. Q., Bijeljic, B., and Blunt, M. J. (2017). Automatic measurement of contact angle in pore-space images. *Adv. Water Resour.* 109, 158–169. doi: 10.1016/j.advwatres.2017.07.018
- Alsamadony, K. L., Yildirim, E. U., Glatz, G., Bin, W. U., and Hanafy, S. M. (2021). Deep learning driven noise reduction for reduced flux computed tomography. *Sensors* 21, 1–17. doi: 10.3390/s21051921
- Alyafei, N., Raeini, A. Q., Paluszny, A., and Blunt, M. J. (2015). A sensitivity study of the effect of image resolution on predicted petrophysical properties. *Transp. Porous Media.* 110, 157–169. doi: 10.1007/s11242-015-0563-0
- Andr , H., Combaret, N., Dvorkin, J., Glatt E., Han, J., Kabel, M., et al. (2013). Digital rock physics benchmarks-part II: computing effective properties. *Comput. Geosci.* 50, 33–43. doi: 10.1016/j.cageo.2012.09.008
- Andrew, M., Bijeljic, B., and Blunt, M. J. (2014). Pore-scale imaging of trapped supercritical carbon dioxide in sandstones and carbonates. *Int. J. Greenh. Gas Control.* 22, 1–14. doi: 10.1016/j.ijggc.2013.12.018
- Explore the merits and drawbacks of sequential vs. co-learning in DL-based image processing, including reconstruction, denoising, and segmentation steps.
- Explore more novel unsupervised denoising methods to remove dependency upon HQ, high-exposure images and make use of less time and cost intensive scans for accurate petrophysical characterization.
- Compare the performance of denoising methods pre-reconstruction vs. post-reconstruction.

DATA AVAILABILITY STATEMENT

The datasets generated and analyzed for this study can be found here: Tawfik, M., Karpyn, Z., and Huang, S. X. (2022). scCO₂-Brine-Glass Dataset for Comparing Image Denoising Algorithms [Data set]. Digital Rocks Portal. <https://doi.org/10.17612/A1QA-2A25>.

AUTHOR CONTRIBUTIONS

MT, AA, XH, and ZK: study conception and design. MT, ZK, and PP: data collection. MT, AA, YH, and PP: analysis and interpretation of results. MT, AA, YH, PP, RJ, PS, XH, and ZK: draft manuscript preparation. All authors reviewed the results and approved the final version of the manuscript.

ACKNOWLEDGMENTS

The authors thank the financial support of the National Energy Technology Laboratory's ongoing research under the RSS contract number 89243318CFE000003.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frwa.2021.800369/full#supplementary-material>

- Armstrong, R. T., McClure, J. E., Berrill, M. A., R cker, M., Schl ter, S., and Berg, S. (2016). Beyond Darcy's law: the role of phase topology and ganglion dynamics for two-fluid flow. *Phys Rev E.* 94, 1–10. doi: 10.1103/PhysRevE.94.043113
- Attix, F. H. (2004). *Introduction to Radiological Physics and Radiation Dosimetry*. Weinheim: WILEY-VCH Verlag GmbH & Co. KGaA.
- Bae, H.-J., Kim, C.-W., Kim, N., Park, B. H., Kim, N., Seo, J. B., et al. (2018). A perlin noise-based augmentation strategy for deep learning with small data samples of HRCT images. *Sci. Rep.* 8, 1–7. doi: 10.1038/s41598-018-36047-2
- Berg, S., R cker, M., Ott, H., Georgiadisa, A., van der Linde, H., Enzmann, F., et al. (2016). Connected pathway relative permeability from pore-scale imaging of imbibition. *Adv. Water Resour.* 90, 24–35. doi: 10.1016/j.advwatres.2016.01.010
- Boas, F. E., Fleischmann, D. (2012). CT artifacts: causes and reduction techniques. *Imaging Med.* 4, 229–240. doi: 10.2217/iim.12.13
- Brown, K., Schluter, S., Sheppard, A., and Wildenschild, D. (2014). On the challenges of measuring interfacial characteristics of three-phase fluid flow with x-ray microtomography. *J. Microsc.* 253, 171–182. doi: 10.1111/jmi.12106
- Buades, A., Coll, B., and Morel, J. M. (2005). "A non-local algorithm for image denoising," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (San Diego, CA), 60–65. doi: 10.1109/CVPR.2005.38

- Budrikis, Z. L. (1972). Visual fidelity criterion and modeling. *Proc. IEEE* 60, 771–779. doi: 10.1109/PROC.1972.8776
- Chen, X., Verma, R., Espinoza, D. N., and Prodanović, M. (2018). Pore-scale determination of gas relative permeability in hydrate-bearing sediments using x-ray computed micro-tomography and lattice boltzmann method. *Water Resour. Res.* 54, 600–608. doi: 10.1002/2017WR021851
- CoreLab (2021). *X-ray Transparent CoreLab Biaxial FCH Series Coreholder*. Available online at: <https://www.corelab.com/cli/core-holders/x-ray-core-holder-fch-series> (accessed January 9, 2021).
- Cromwell, V., Kortum, D. J., and Bradley, D. J. (1984). “The use of a medical computer tomography (CT) system to observe multiphase flow in porous media,” in *Proceedings of the SPE Annual Technical Conference and Exhibition* (Houston, TX), 1–4. doi: 10.2523/13098-ms
- Culligan, K. A., Wildenschild, D., Christensen, B. S. B., Gray, W. G., Rivers, M. L., and Tompson, A. F. B. (2004). Interfacial area measurements for unsaturated flow through a porous medium. *Water Resour. Res.* 40, 1–12. doi: 10.1029/2004WR003278
- Diwakar, M., and Kumar, M. (2018). A review on CT image noise and its denoising. *Biomed. Signal Process Control* 42, 73–88. doi: 10.1016/j.bspc.2018.01.010
- Dong, H., Blunt, M. J., (2009). Pore-network extraction from micro-computerized-tomography images. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* 80, 1–11. doi: 10.1103/PhysRevE.80.036307
- Fan, L., Zhang, F., Fan, H., and Zhang, C. (2019). Brief review of image denoising techniques. *Vis. Comput. Ind. Biomed. Art.* 2:7. doi: 10.1186/s42492-019-0016-7
- Freire-Gormaly, M., Ellis, J. S., MacLean, H. L., and Bazylak, A. (2015). Pore structure characterization of Indiana limestone and pink dolomite from pore network reconstructions. *Oil Gas Sci. Technol.* 71:33. doi: 10.2516/ogst/2015004
- Gao, Y., Lin, Q., Bijeljic, B., and Blunt, M. J. (2017). X-ray microtomography of intermittency in multiphase flow at steady state using a differential imaging method. *Water Resour. Res.* 53, 10274–10292. doi: 10.1002/2017WR021736
- Gastal, E. S. L., and Oliveira, M. M. (2012). Adaptive manifolds for real-time high-dimensional filtering. *ACM Trans. Graph.* 31, 33:1–33:13. doi: 10.1145/2185520.2185529
- Grader, A., Mu, Y., Toelke, J., Baldwin, C., Fang, Q., Carpio, G., et al. (2010). “SPE 138591. Estimation of relative permeability using the lattice boltzmann method for fluid flows in a cretaceous formation, Abu Dhabi,” in Paper presented at the Abu Dhabi International Petroleum Exhibition and Conference (Abu Dhabi).
- Herring, A. L., Harper, E. J., Andersson, L., Sheppard, A., Bay, B. K., and Wildenschild, D. (2013). Effect of fluid topology on residual nonwetting phase trapping: implications for geologic CO₂ sequestration. *Adv Water Resour.* 62, 47–58. doi: 10.1016/j.advwatres.2013.09.015
- Hsieh, J. (2015). *Computed Tomography: Principles, Design, Artifacts, and Recent Advances*, 3rd Edn. Bellingham WA: SPIE.
- Huang, T. S., Yang, G. J., and Tang, G. Y. (1979). A fast two-dimensional median filtering algorithm. *IEEE Trans. Acoust.* 27, 13–18. doi: 10.1109/TASSP.1979.1163188
- Iglauer, S., Fernø, M. A., Shearing, P., and Blunt, M. J. (2012). Comparison of residual oil cluster size distribution, morphology and saturation in oil-wet and water-wet sandstone. *J. Colloid Interface Sci.* 375, 187–192. doi: 10.1016/j.jcis.2012.02.025
- Jasti, J. K., Jesion, G., and Feldkamp, L. (1993). Microscopic imaging of porous media with X-ray computer tomography. *SPE Form. Eval.* 8, 189–193. doi: 10.2118/20495-PA
- Jia, L., Ross, C. M., and Kovalick, A. R. (2007). A pore-network-modeling approach to predict petrophysical properties of diatomaceous reservoir rock. *SPE Reserv. Eval. Eng.* 10, 597–608. doi: 10.2118/93806-pa
- Kakouei, A., Vatani, A., Rasaei, M. R., Sadaee Sola, B., and Moqtaderi, H. (2017). Cessation of Darcy regime in gas flow through porous media using LBM: comparison of pressure gradient approaches. *J. Nat. Gas Sci. Eng.* 45, 693–705. doi: 10.1016/j.jngse.2017.06.018
- Kamrava, S., Tahmasebi, P., and Sahimi, M. (2019). Enhancing images of shale formations by a hybrid stochastic and deep learning algorithm. *Neural Netw.* 118, 310–320. doi: 10.1016/j.neunet.2019.07.009
- Kang, E., Koo, H. J., Yang, D. H., Seo, J. B., and Ye, J. C. (2019). Cycle-consistent adversarial denoising network for multiphase coronary CT angiography. *Med. Phys.* 46, 550–562. doi: 10.1002/mp.13284
- Karpyn, Z. T., Piri, M., and Singh, G. (2010). Experimental investigation of trapped oil clusters in a water-wet bead pack using X-ray microtomography. *Water Resour. Res.* 46, 1–25. doi: 10.1029/2008WR007539
- Keys, R. G. (1981). Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoust. Speech Signal Process.* 29, 1153. doi: 10.1093/nq/s10-III.69.308-b
- Kingma, D. P., and Ba, J. L. (2015). “Adam: a method for stochastic optimization,” in *3rd International Conference on Learning Representations, ICLR 2015* (San Diego, CA), 1–15.
- Klise, K. A., Moriarty, D., Yoon, H., and Karpyn, Z. (2016). Automated contact angle estimation for three-dimensional X-ray microtomography data. *Adv. Water Resour.* 95, 152–160. doi: 10.1016/j.advwatres.2015.11.006
- Knoll, G. F. (2000). *Radiation Detection and Measurement*, 3rd Edn. New York, NY: Wiley. doi: 10.1002/hep.22108
- Krizhevsky, B. A., Sutskever, I., and Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *ACM Commun.* 60, 84–90. doi: 10.1145/3065386
- Krull, A., Buchholz, T. O., and Jug, F. (2019). “Noise2void-learning denoising from single noisy images,” in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (Long Beach, CA), 2124–2132. doi: 10.1109/CVPR.2019.00223
- Kulathilake, K. A. S. H., Abdullah, N. A., Sabri, A. Q. M., and Lai, K. W. (2021). *A Review on Deep Learning Approaches for Low-Dose Computed Tomography Restoration*. Kuala Lumpur: Springer International Publishing. doi: 10.1007/s40747-021-00405-x
- Landry, C. J., Karpyn, Z. T., and Piri, M. (2011). Pore-scale analysis of trapped immiscible fluid structures and fluid interfacial areas in oil-wet and water-wet bead packs. *Geofluids* 11, 209–227. doi: 10.1111/j.1468-8123.2011.00333.x
- Landry, J. C., Karpyn, Z. T., Ayala, O., (2014). Relative permeability of homogenous-wet and mixed-wet porous media as determined by pore-scale lattice Boltzmann modeling. *Water Resour. Res.* 50, 3672–3689. doi: 10.1002/2013WR014979.Reply
- Larpudomlert, R., Torrealba, V. A., Karpyn, Z. T., and Halleck, P. (2014). Experimental investigation of residual saturation in mixed-wet porous media using a pore-scale approach. *J. Pet. Explor. Prod. Technol.* 4, 175–187. doi: 10.1007/s13202-013-0076-8
- Lee, S., Lee, M. S., and Kang, M. G. (2018). Poisson-gaussian noise analysis and estimation for low-dose X-ray images in the NSCT domain. *Sensors (Switzerland)* 18, 1–22. doi: 10.3390/s18041019
- Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., et al. (2018). “Noise2Noise: learning image restoration without clean data,” in *Proceedings of the 35th International Conference on Machine Learning*, Vol. 7, (Stockholm), 4620–4631.
- Li, Z., Galindo-Torres, S., Yan, G., Scheuermann, A., and Li, L. (2018). A lattice Boltzmann investigation of steady-state fluid distribution, capillary pressure and relative permeability of a porous medium: effects of fluid and geometrical properties. *Adv. Water Resour.* 116, 153–166. doi: 10.1016/j.advwatres.2018.04.009
- Lin, Q., Bijeljic, B., Pini, R., Blunt, M. J., and Krevor, S. (2018). Imaging and measurement of pore-scale interfacial curvature to determine capillary pressure simultaneously with relative permeability. *Water Resour. Res.* 54, 7046–7060. doi: 10.1029/2018WR023214
- Liu, H., Valocchi, A. J., Werth, C., Kang, Q., and Oostrom, M., (2014). Pore-scale simulation of liquid CO₂ displacement of water using a two-phase lattice Boltzmann model. *Adv. Water Resour.* 73, 144–158. doi: 10.1016/j.advwatres.2014.07.010
- Machado, A. C. S., Machado, A. C. S., and Lopes, R. T. (2013). “Influence of mathematical filters in the volume measurement of a travertine sample using high resolution x-ray microtomography,” in *2013 International Nuclear Atlantic Conference - INAC 2013* (Recife, PE).
- Matrecano, M., Poderico, M., Poggi, G., Romano, M., and Cesarelli, M. (2010). “Application of denoising techniques to micro-tomographic images,” in *Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine* (Corfu), 1–4. doi: 10.1109/ITAB.2010.5687713
- McClure, J. E., Ramstad, T., Li, Z., Armstrong, R. T., and Berg, S. (2020). Modeling geometric state for fluids in porous media: evolution of the euler characteristic. *Transp. Porous Media* 133, 229–250. doi: 10.1007/s11242-020-01420-1
- Menke, H. P., Andrew, M. G., Blunt, M. J., and Bijeljic, B. (2016). Reservoir condition imaging of reactive transport in heterogeneous carbonates using fast synchrotron tomography - effect of initial pore structure and flow conditions. *Chem. Geol.* 428, 15–26. doi: 10.1016/j.chemgeo.2016.02.030

- Mittal, A., Moorthy, A. K., and Bovik, A. C. (2011). "Blind/referenceless image spatial quality evaluator," in *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)* (Pacific Grove, CA).
- Mittal, A., Soundararajan, R., and Bovik, A. C. (2013). Making a "completely blind" image quality analyzer. *IEEE Signal Process. Lett.* 20, 209–212. doi: 10.1109/LSP.2012.2227726
- Mostaghimi, P., Blunt, M. J., and Bijeljic, B. (2013). Computations of absolute permeability on micro-CT images. *Math. Geosci.* 45, 103–125. doi: 10.1007/s11004-012-9431-4
- Niu, F. (2021). *Experimental Investigation of Pressure Dependence of Contact Angle in CO₂-Brine Systems*. Master Thesis, Pennsylvania State University.
- Orhan, K. (2020). *Micro-Computed Tomography (Micro-CT) in Medicine and Engineering*. Cham: Springer. doi: 10.1007/978-3-030-16641-0
- Perlin, K. (1985). An image synthesizer. *Comput Graph.* 19, 287–296. doi: 10.1145/325165.325247
- Perona, P., and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 12, 629–639.
- Porter, M. L., and Wildenschild, D. (2010). Image analysis algorithms for estimating porous media multiphase flow variables from computed microtomography data: a validation study. *Comput. Geosci.* 14, 15–30. doi: 10.1007/s10596-009-9130-5
- Probst, G., Boeckmans, B., Kruth, J.-P., and Dewulf, W. (2016). "Compensation of drift in an industrial computed tomography system," *6th Conference on Industrial Computed Tomography* (Wels), 1–6.
- Prodanović, M., and Bryant, S. L. (2006). A level set method for determining critical curvatures for drainage and imbibition. *J Colloid Interface Sci.* 304, 442–458. doi: 10.1016/j.jcis.2006.08.048
- Purswani, P., Karpyn, Z. T., Enab, K., Xue, Y., and Huang, X. (2020). Evaluation of image segmentation techniques for image-based rock property estimation. *J. Pet. Sci. Eng.* 195, 107890. doi: 10.1016/j.petrol.2020.107890
- PyBRISQUE-1.0 software package (2020). Available online at: <https://pypi.org/project/pybrisque/> (accessed January 4, 2021).
- Raeini, A. Q., Yang, J., Bondino, I., Bultreys, T., Blunt, M. J., and Bijeljic, B. (2019). Validating the generalized pore network model using micro-CT Images of two-phase flow. *Transp. Porous Media.* 130, 405–424. doi: 10.1007/s11242-019-01317-8
- Ritman, E. L. (2004). Micro-computed tomography—current status and developments. *Annu. Rev. Biomed. Eng.* 6, 185–208. doi: 10.1146/annurev.bioeng.6.040803.140130
- Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: convolutional networks for biomedical image segmentation. *IEEE Access.* 9, 16591–16603. doi: 10.1109/ACCESS.2021.3053408
- Rücker, M., Bartels, W. B., Singh, K., Brussee, N., Coorn, A., van der Linde, H. A., et al. (2019). The effect of mixed wettability on pore-scale flow regimes based on a flooding experiment in ketton limestone. *Geophys. Res. Lett.* 46, 3225–3234. doi: 10.1029/2018GL081784
- Scanziani, A., Singh, K., Blunt, M. J., and Guadagnini, A. (2017). Automatic method for estimation of *in situ* effective contact angle from X-ray micro tomography images of two-phase flow in porous media. *J. Colloid. Interface Sci.* 496, 51–59. doi: 10.1016/j.jcis.2017.02.005
- Shahmoradi, M., Lashgari, M., Rabbani, H., Qin, J., and Swain, M. (2016). A comparative study of new and current methods for dental micro-CT image denoising. *Dentomaxillofacial Radiol.* 45, 20150302. doi: 10.1259/dmfr.20150302
- Sharma, M. M., and Yortsos, Y. C. (1987). Fines migration in porous media. *AIChE J.* 33, 1654–1662. doi: 10.1002/aic.690331009
- Sidorenko, M., Orlov, D., Ebadi, M., and Koroteev, D. (2021). Deep learning in denoising of micro-computed tomography images of rock samples. *Comput. Geosci.* 151:104716. doi: 10.1016/j.cageo.2021.104716
- Singh, K., Menke, H., Andrew, M., Lin, Q., Rau, C., Blunt, M. J., et al. (2017). Dynamics of snap-off and pore-filling events during two-phase fluid flow in permeable media. *Sci. Rep.* 7, 5192. doi: 10.1038/s41598-017-05204-4
- Skvideo (2013). *Skvideo.measure.niqe Package*. Available online at: <http://www.scikit-video.org/stable/modules/generated/skvideo.measure.niqe.html> (accessed January 4, 2021).
- Sommer, C., Straehle, C., Kothe, U., and Hamprecht, F. A. (2011). Ilastik: interactive learning and segmentation toolkit. *Proc. Int. Symp. Biomed. Imaging.* 2011, 230–233. doi: 10.1109/ISBI.2011.5872394
- Tawfik, M. S. (2020). *Multi-Scale Investigation of Wettability Alteration via Chemically-tuned Waterflooding in Carbonate Rocks: Understanding the Effect of Oil Chemistry*. Dissertation, Pennsylvania State University.
- Tawfik, M. S., Karpyn, Z., and Johns, R. (2019). "Multiscale study of chemically-tuned waterflooding in carbonate rocks using micro-computed tomography," in *IOR 2019 - 20th Eur Symp Improv Oil Recover* (Pau). doi: 10.3997/2214-4609.201900074
- Tian, C., Fei, L., Zheng, W., Xu, Y., Zuo, W., and Lin, C. W. (2020). Deep learning on image denoising: an overview. *Neural Networks.* 131, 251–275. doi: 10.1016/j.neunet.2020.07.025
- Valvatne, P. H., and Blunt, M. J. (2004). Predictive pore-scale modeling of two-phase flow in mixed wet media. *Water Resour. Res.* 40, 1–21. doi: 10.1029/2003WR002627
- Van Stappen, J. (2018). *A Micro-CT investigation into pore-Scale CO₂ Sequestration Processes in Fractured Reservoir Rocks*. Thesis, PhD. in Geology, Ghent University.
- Verri, I., Della Torre, A., Montenegro, G., Onorati, A., Ducab S., Morab, C. A et al., (2017). Development of a digital rock physics workflow for the analysis of sandstones and tight rocks. *J. Pet. Sci. Eng.* 156, 790–800. doi: 10.1016/j.petrol.2017.06.053
- Wang, S. Y., Ayril, S., and Gryte, C. C. (1984). Computer-assisted tomography for the observation of oil displacement in porous media. *Soc Pet Eng. J.* 24, 53–55. doi: 10.2118/11758-PA
- Wang, Y. D., Blunt, M. J., Armstrong, R. T., and Mostaghimi, P. (2021). Deep learning in pore scale imaging and modeling. *Earth Sci. Rev.* 215:103555. doi: 10.1016/j.earscirev.2021.103555
- Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13, 600–612. doi: 10.1109/TIP.2003.819861
- Weickert, J. (1996). *Anisotropic Diffusion in Image Processing*. Stuttgart: B.G. Teubner. Available online at: <https://www.mia.uni-saarland.de/weickert/Papers/book.pdf>
- Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., et al. (2018). Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Trans. Med. Imaging.* 37, 1348–1357. doi: 10.1109/TMI.2018.2827462
- Zahaf, K., Lecoq, T. F., Badi, B. S. A. L., Roth, S., Dong, H., Blunt, M. J. (2017). "Prediction of relative permeability and capillary pressure using digital rock physics : case study on two giant middle eastern carbonate reservoirs," in *International Symposium of the Society of Core Analysts* (Vienna), 1–9.
- Zhang, Y., Tian, Y., Kong, Y., Zhong, B., and Fu, Y. (2021). Residual dense network for image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 2480–2495. doi: 10.1109/TPAMI.2020.2968521

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Tawfik, Adishesha, Hsi, Purswani, Johns, Shokouhi, Huang and Karpyn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

NOMENCLATURE

| | |
|-------------------|---|
| AD | Anisotropic diffusion |
| BET | Brunauer, Emmett and Teller |
| BI | Blurring index |
| BM3D | Block-matching and 3D filtering |
| BRISQUE | Blind/reference-less image spatial quality evaluator |
| CCGAN | Cycle consistent generative adversarial network |
| CCSIM | Cross-correlation-based simulation |
| CCUS | Carbon capture, utilization, and storage |
| CNN | Convolutional neural network |
| CNR | Contrast-to-noise ratio |
| dB | Decibels |
| DL | Deep learning |
| GANs | Generative adversarial networks |
| GPU | Graphics processing unit |
| HQ | High quality |
| L1 | loss Least absolute deviations loss function |
| L2 | loss Least square errors loss function |
| LQ | low quality |
| LS | Limestone |
| MAE | Mean absolute error |
| MCT | Micro-computed tomography |
| N2C | Noise-to-clean |
| N2N | Noise-to-noise with no reference ground truth images |
| N2N25 | Noise-to-noise with 25% of the dataset having reference ground truth images |
| N2N50 | Noise-to-noise with 50% of the dataset having reference ground truth images |
| N2N75 | Noise-to-noise with 75% of the dataset having reference ground truth images |
| N2V | Noise-to-void |
| NIQE | Naturalness image quality evaluator |
| NLM | Non-local means |
| PSNR | Peak signal-to-noise ratio |
| RDN | Residual dense network |
| ROI | Region of interest |
| scCO ₂ | Supercritical carbon dioxide |
| SEN | Specific Euler number |
| SNN | Symmetric nearest neighbor |
| SS | Sandstone |
| SSA | Specific surface area |
| SSIM | Structural similarity index |
| VDSR | Very deep super resolution |
| VGG | Visual Geometry Group |
| λ | Poisson process mean |
| μ | Gaussian distribution mean |
| σ | Gaussian distribution standard deviation |