U-Net Segmentation Methods for Variable-Contrast XCT Images of Methane-Bearing Sand

F. J. Alvarez-Borges[1], O. N. F. King[1†], B. N. Madhusudhan[2], T. Connolley[1], M. Basham[1 3] and S. I. Ahmed[1]

[1]Diamond Light Source Ltd.

[2]University of Southampton.

[3] The Rosalind Franklin Institute.

† Corresponding author: Oliver N. F. King (olly.king@diamond.ac.uk)

Key Points:

- Minimally trained U-Net models can perform multiphase segmentation of variable-contrast X-ray CT images of methane-bearing sand.
- U-Net models trained on low-contrast images can accurately segment different, higher contrast data sets without additional training.
- U-Net segmentations deliver accurate 3D visualizations of the soil fabric and porosity/methane saturation profiles.

Abstract

Methane (CH_4) hydrate dissociation and CH_4 release are potential geohazards currently investigated using X-ray computed tomography (XCT) imaging in laboratory experiments. Image segmentation constitutes an important data processing step for this type of research, but it is often time consuming, computing resource-intensive and operator-dependent. Furthermore, segmentation procedures are frequently tailored for each XCT data set due to differences in image characteristics, such as greyscale contrast variations. To address these issues, an investigation has been carried out using U-Nets, a novel class of Convolutional Neural Network, to segment synchrotron radiation XCT (SRXCT) images of CH_4-bearing sand during hydrate formation. Three U-Net deployment methodologies previously untried for this task were assessed: (1) 3D hierarchical, (2) 2D multilabel and (3) RootPainter, a 2D application that implements interactive corrections. Results show high segmentation accuracy, with RootPainter slightly outperforming the alternative approaches. Greyscale contrast between material phases was found to affect segmentation performance, with the lowest metrics corresponding to data exhibiting the lowest contrast. Segmentation accuracy affected derived parameters such as CH_4-saturation and porosity, but errors were small compared with gravimetric methods. It was also found that U-Net models trained on low greyscale contrast images could be used to segment higher-contrast data sets and produce accurate 3D visualizations of CH_4 distribution, demonstrating model portability. Such portability is anticipated to be advantageous when the segmentation of large XCT data sets needs to be delivered over short timespans.

Plain Language Summary

Methane hydrates are ice-like solids present in deep ocean sediments and frozen ground and contain large volumes of methane gas. Recently, geoscientists have used X-ray computer tomography to produce 3D images of hydrate formation and melting in controlled experiments. They then classify the images into the soil grains, water and methane that composed the sample to measure changes in soil structure. This process is called segmentation, and often needs to be tailored for each image depending on the difference in tone between the features being classified, known as contrast. Therefore, segmentation can be time-consuming, and results might vary depending on the person who performs it. Looking to overcome this, we evaluated the use of a class of machine learning algorithm called U-Net to perform segmentations. U-Nets use a set of verified segmented images as training data to ‘learn’ how to segment similar images. We investigated three ways of implementing U-Nets and found that they all produced accurate segmentations, but accuracy diminished for low contrast images. U-Net segmentations were then used to accurately calculate parameters like porosity and methane saturation. Finally, we discovered that U-Net algorithms trained on low-contrast images could be used to segment higher-contrast images without additional training.

1 Introduction

Deep sea sediments and permafrost host large quantities of methane (CH_4), an energy source and potent greenhouse gas that may be a contributor to climate change (Dean et al., 2018; IPCC, 2013). Much of this CH_4 is present as hydrates (clathrates), that is, crystalline lattices of frozen water that enclose CH_4 molecules. 164 m^3 of CH_4 gas at normal temperature and pressure can be stored in one m^3 of hydrate (Kvenvolden, 1993). However, the extent of the world-wide CH_4 hydrate inventory is subject to considerable uncertainty (James et al., 2016; Ruppel & Kessler, 2017). This is in part due to discrepancies between measurements produced by geophysical and electrical resistivity methods (Sahoo, Marín-Moreno et al., 2018; Yokohama et al., 2011), which are potentially associated with hydrate heterogeneity in the host soils (Sahoo, Madhusudhan et al., 2018). Uncertainties on the global CH_4 hydrate inventory affect resource estimation and CH_4 emission prediction models (Moridis et al., 2011; Ruppel & Kessler, 2017; Saunois et al., 2020).

In addition to hydrocarbon resource and greenhouse gas emission prediction challenges, CH_4 hydrate formation and dissociation has also been associated with changes in the mechanical characteristics of the host sediment, which may result in geohazards. For instance, hydrates may strengthen and stiffen the sediment by creating inter-grain cementation bonds (Madhusudhan et al., 2019; Song et al., 2019). Hydrate dissociation may then reverse these gains and is thus speculated to lead to, for example, underwater slides that may trigger tsunami or damage seabed infrastructure such as cables and pipelines, which are vital for communications and energy transport (Maslin et al., 2010; Mienert, 2009; Vanneste et al., 2014).

Recently, researchers have shown that X-ray computed tomography (XCT) can be used to successfully detect hydrate heterogeneity and characterize changes in sediment microstructure associated with hydrate formation and dissociation (Holland & Schultheiss, 2014; Kerkar et al., 2014; Lei et al., 2018; Sahoo, Madhusudhan et al., 2018). This has been possible in great part due to advancements in image segmentation techniques. Segmentation is the process of classifying 2D pixels or 3D voxels into regions, for example, the solids (e.g. soil grains and cement bonds), liquids (e.g. water or brine) and gases (e.g. air or CH_4) present in an image. Microstructural parameters such as porosity (or void ratio) and grain and pore size, shape and orientation can then be derived from the segmented image, as well as volumetric quantities like CH_4 gas and hydrate saturation ratios.

Some of the most common segmentation techniques used in geomechanics and geoscience are greyscale thresholding and watershed algorithms (Fonseca et al., 2009; Iassonov et al., 2009). The former involves the selection of a greyscale range to classify pixels or voxels into regions of interest. Watershed algorithms redefine the image as a geographical map, were greyscale intensities form topographical elevations and catchment basins. Pixel/voxel markers within these basins are used to define the materials or 'labels' present in the image, and the algorithm then morphologically dilates these markers until they 'fill' their catchment basins (Rogowska, 2000; Zhang et al., 2014). Greyscale range determination in thresholding techniques and marker grey value and location in watershed techniques are operator and/or method dependent (Baveye et al., 2010; Fonseca et al., 2009; Koyuncu et al., 2012). The values assigned to these parameters also depend on the recorded greyscale contrast, which is highly reliant on the X-ray imaging instrument used and how it was optimized. For example, Brunke et al. (2008) showed that for the same sample, a significant difference in image contrast can be present between data taken using laboratory-based X-ray tube sources and synchrotron radiation sources.

Sample heterogeneity or density changes during an in-situ experiment will also introduce contrast variability in space and time (Fonseca et al., 2009; Kong & Fonseca, 2018). As a result, thresholding and watershed segmentation are typically optimized per XCT scan. Consequently, objective comparison is difficult given that the data treatment varies between data sets. These issues often result in segmentation procedures in geomechanics and geoscience that are highly demanding of computing resources and operator time.

Novel alternative approaches have employed machine learning solutions to segment the material phases present in XCT images of soil and rock samples (Chauhan, Rühaakm Anbergen et al., 2016; Chauhan, Rühaakm Khan et al., 2016; Chauhan et al., 2020). For these applications, machine learning algorithms produce segmentations via a mathematical model optimized or 'trained' using a series of 'ground truth' example segmentations of XCT images provided by the user. Within the realm of machine learning, convolutional neural networks (CNNs) are a class of deep neural networks that are commonly used in computer vision and image processing. Researchers are now beginning to explore their application to segment XCT images of soil and rock (Douarre et al., 2018; Karimpouli & Tahmasebi, 2019; Varfolomeev et al., 2019). CNNs employ multiple convolutional layers where the filters ('kernels') used to separate image features are learned (Krizhevsky et al., 2017). U-Nets are a class of CNN originally designed to segment biomedical images (Ronneberger et al., 2015). The U-Net architecture is composed of downsampling (encoding/contracting) and upsampling (decoding/expanding) paths. The former reduces the spatial dimensions of the data while increasing feature information while the latter recombines spatial and feature data to generate the label image. Additionally, the encoding and decoding sections of the network are linked by connections that can feed the output from each level of the contracting path directly into the corresponding level of the expanding path. These connections provide a path for spatial information that allows the preservation of fine-grained details in the upsampled output image. A limitation to the implementation of U-Nets (and CNNs in general) to segment XCT images of soil and rock is the preparation of training and validation data sets, which often require labor-intensive manual segmentation (hand annotation) of many images.

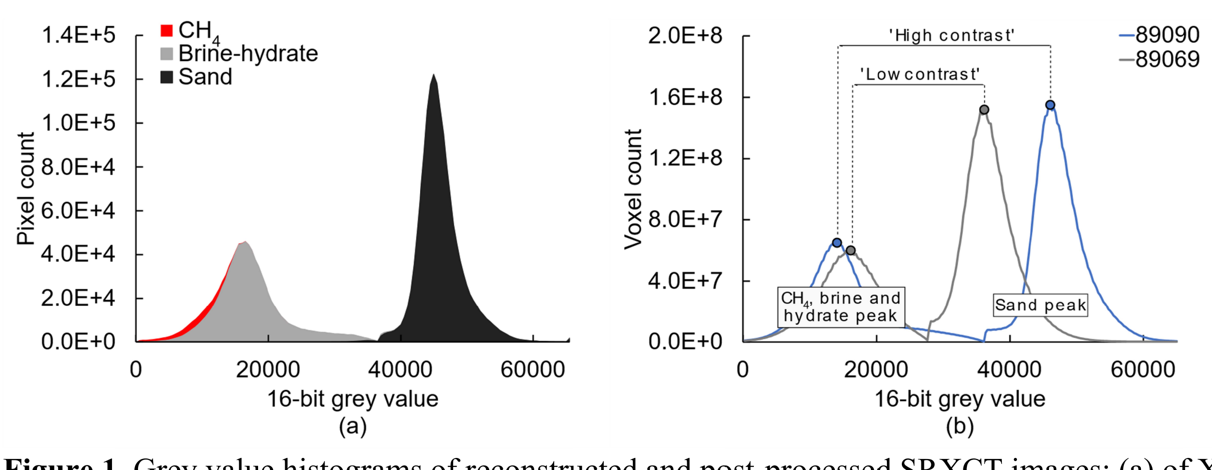

Figure 1. Grey value histograms of reconstructed and post-processed SRXCT images: (a) of XY slice 1050 of scan 89062, showing the frequency distribution of pixels for each material; (b) of two whole 3D images showing the grey value difference between histogram peaks as a measure of image contrast.

This paper examines the use of U-Nets to segment synchrotron radiation XCT (SRXCT) images of CH_4-bearing sand. The SRXCT data was obtained from in-situ imaging of hydrate formation and dissociation experiments. The reconstructed volumes exhibited different greyscale contrast amongst them. Furthermore, contrast between the three main material phases present in the images was low, as shown in the example image histogram for a reconstructed slice in **Figure 1(a)**. This rendered the use of 'standard' thresholding or watershed techniques unsuitable. Instead, three different U-Net implementation strategies, previously untried for this purpose, have been developed and applied. The U-Net segmentation procedures targeted the three main material phases present in the images: (1) sand, (2) CH_4 gas bubbles and (3) brine combined with hydrates, since the contrast between these two materials was minimal. Special focus has been given to the CH_4 gas phase, as it not only exhibited low contrast with regards to the brine-hydrate phase but was also uncommon in the data compared to the other materials, as evidenced in **Figure 1(a)**. The aim of the investigation was to determine if U-Nets can accurately segment XCT images of soil samples with varying greyscale contrast between material phases using only a small number of training and validation images, thus reducing operator/computing time and allowing objective comparison of data. The starting hypotheses were (1) that U-Net models trained on a small portion of the reconstructed SRXCT 3D image can be used to accurately segment the entire volume, (2) that segmentation accuracy is directly linked to greyscale contrast between materials, and (3) that accurate U-Net segmentation models produced from training on a given SRXCT data set can deliver accurate segmentations for similar data sets without additional training (model portability).

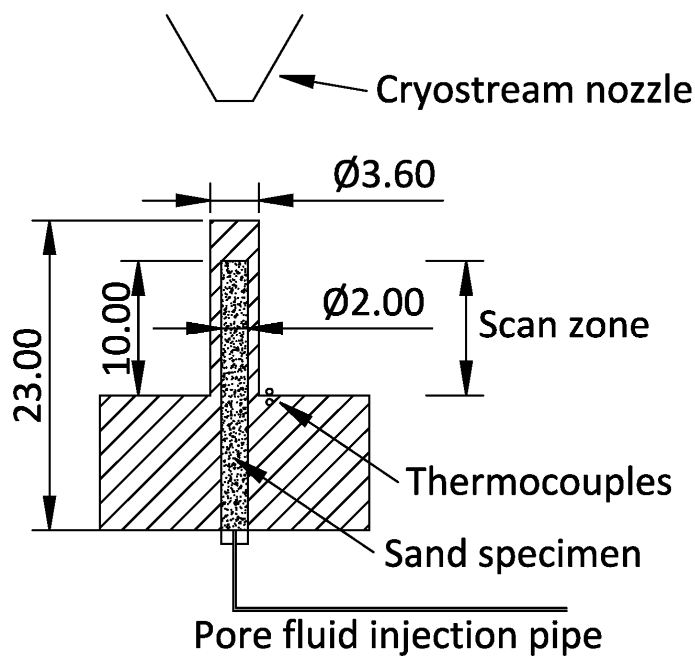

Figure 2. Cross-section sketch of hydrate test rig. Monolithic PEEK element denoted by hatched area. All units mm.

2 Materials and Methods

2.1 Methane gas hydrate formation and dissociation experiments

A custom rig designed and manufactured by Sahoo, Madhusudhan et al. (2018) for in-situ SRXCT imaging of gas hydrate formation and dissociation was used in the present study. The rig is made of polyether ether ketone (PEEK) and consists of a monolithic 2 mm internal diameter by 23 mm tall cylindrical vessel with 0.8 mm thick walls and an enlarged base, as shown in **Figure 2**. The SRXCT imaging zone in this study corresponds to a 1.755 mm section of the 10 mm-tall portion of the vessel that protrudes from the enlarged base. The soil sample is placed through the bottom of the rig. The pore fluid injection pipe is connected to this inlet, as depicted in **Figure 2**. The rig features thermocouples at the base of the SRXCT imaging zone to measure sample temperature.

Leighton Buzzard sand Fraction E (LBE) with mean grain diameter of 100 μm was used as surrogate marine sediment. LBE is an angular silica sand widely used as a standard laboratory material in geotechnical research. The sand was tamped into the PEEK vessel to a target porosity of 35%. A vacuum pressure of less than 1 Pa was applied through the injection pipe to reduce air presence in the pore space. A calculated volume of brine solution (3.5% NaCl by weight) was thereafter injected into the sample, such that approximately 90% of the pore volume became saturated. CH_4 gas was then injected at 10 MPa and the valve to the sample closed. The sample was gradually cooled to a target constant temperature of 2 °C using a N_2 cryostream. This thermobaric condition enabled hydrate formation in the pore space instead of ice. The target temperature was maintained for 30 hours to complete the hydrate formation process (Madhusudhan et al., 2019).

2.2 Synchrotron X-ray Computed Tomography

2.2.1 Set-up and Image Acquisition

Data was collected on beamline I13-2 at Diamond Light Source (DLS). Scans were performed using a polychromatic ‘pink beam’ at 30 keV peak energy. The detector system used was a scintillator-coupled pco.edge 5.5 camera fitted with a 4x optic magnification lens, resulting in an effective pixel size of 0.8125 μm. The X-ray projection size was 2560×2160 pixels (width × height).

Table 1. *SRXCT Scan Summary.*

Data set	Time at 2°C (h)	Number of projections	Exposure time per projection (ms)
89062	0.00	1501	200
89064	1.53	1501	200
89069	5.38	3001	30
89075	10.72	3001	30
89090	20.77	1501	30
89113	30.02	1501	30

Scans were carried out in-situ at various time intervals after reaching 2 °C. The number of projections and the exposure time per projection varied amongst scans to reduce acquisition times at specific moments of the CH_4 hydrate formation process. **Table 1** correlates each scan discussed in this paper with the time after the start of the 30-hour sustained 2 °C period, as well as the scan set-up used.

2.2.2. Tomographic Reconstruction and Post-Processing

Tomographic reconstruction was carried out using Savu 2.4 (Wadeson et al., 2019; see also Atwood et al., 2015, Wadeson & Basham, 2016). Two Savu reconstruction pipelines were used: one with and one without Paganin phase retrieval (Paganin et al., 2002). These pipelines were labelled ‘phase contrast’ (**Figure 3(b)**) and ‘absorption contrast’ (**Figure 3(a)**), respectively. The Savu plugins used and their descriptions are included in the supporting information (Text S1). To obtain a single reconstructed volume per SRXCT scan that retained both edge detail and phase contrast, the output from both reconstruction pipelines was averaged and median and unsharp masking image filters applied using Fiji (Schindelin et al., 2012; Schneider et al., 2012), as detailed in Text S2 of the supporting information. An example slice resulting from this procedure is shown in **Figure 3(c**). Finally, to mitigate the halo-like artefact caused by the preferential attenuation of lower-energy X-rays close to the specimen surface, known as ‘beam hardening’ (Hsieh, 2015), each slice was convolved with a bump-shaped mollifier function that flattened the horizontal (XY) grey value profile. This procedure is explained in Text S3 of the supporting information, and an example output slice is presented in **Figure 3(d)**.

Limited greyscale contrast between the CH_4 gas and the brine-hydrate phase persisted after reconstruction and post-processing. Distinction between these two phases became increasingly difficult as the distance between the 3D image histogram peaks for the sand and non-sand phases reduced, as exemplified in **Figure 1(b)**. This distance is therefore used in this paper as an overall measure for image contrast, with regards to the ease with which the material phases could be identified and segmented. Considering this, ‘intermediate contrast’ data set 89062 was selected initially to investigate the suitability of U-Nets to segment the three main material phases present in the images.

2.3 U-Net Segmentation

Three different methodologies were used to create trained U-Net models that could classify the SRXCT data into three labels: sand, brine and hydrates, and CH_4 gas. These methods were:

1. A 3D hierarchical approach where two separate 3D U-Net models were trained to perform binary segmentations on the sand phase vs the others and the CH_4 gas phase vs the others.
2. A 2D multilabel approach where a single 2D U-Net was trained to classify the three labels. The encoder section of this U-Net implementation was pre-trained on the ImageNet data set (Russakovsky et al., 2015), meaning that the network should only

require a small amount of 'transfer' training in order to achieve acceptable results on new data.

3. RootPainter software, which uses a graphical user interface (GUI) and human intervention by interactive corrections to train a lightweight binary 2D U-Net model.

The models produced by each method were used to segment a 1554×1554×2000 voxel region of the 2560×2560×2000 reconstructed and post-processed volumes. This region was inscribed within the cylindrical FOV of the post-processed volumes and omitted the black pseudo-background generated during reconstruction. **Figure 4(a)** shows the 1554×1554×2000 volume for data set 89062. All 1554×1554×2000 3D images discussed in this paper are available in Alvarez-Borges et al. (2021).

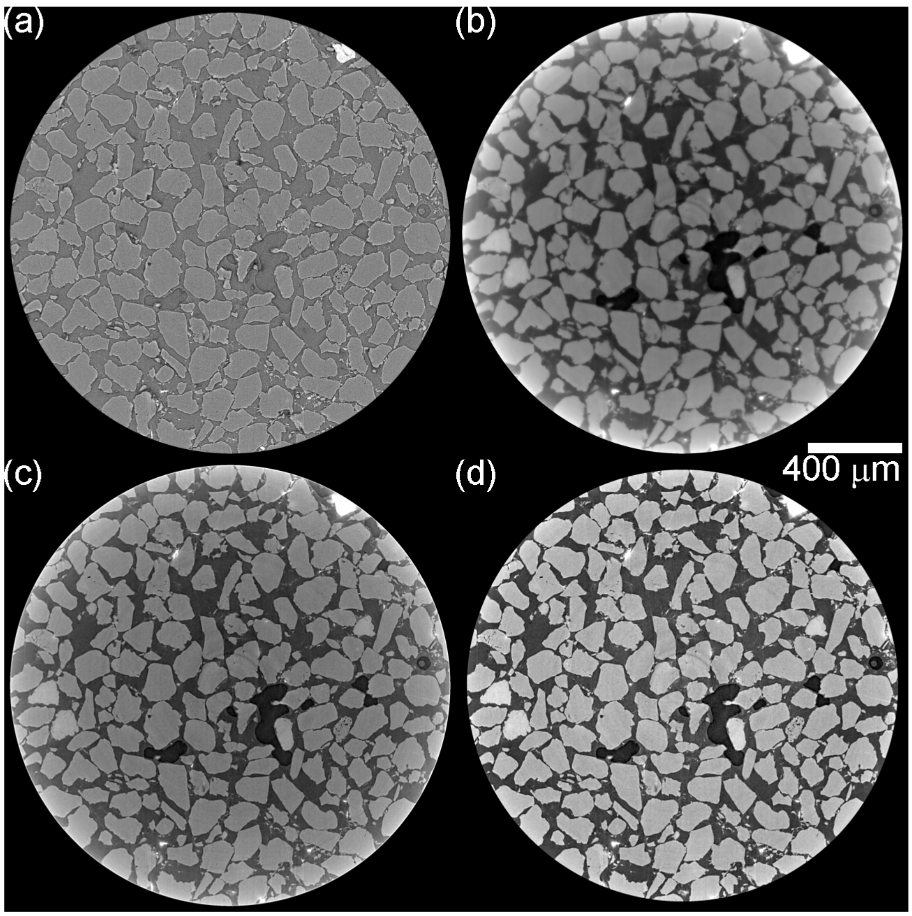

Figure 3. Slice 1050 of data set 89062 showing the output of the reconstruction and post-processing stages: **(a)** reconstruction through absorption contrast pipeline; **(b)** reconstruction through phase contrast pipeline; **(c)** First post-processing output (volume averaging); **(d)** Second post-processing output (beam hardening correction).

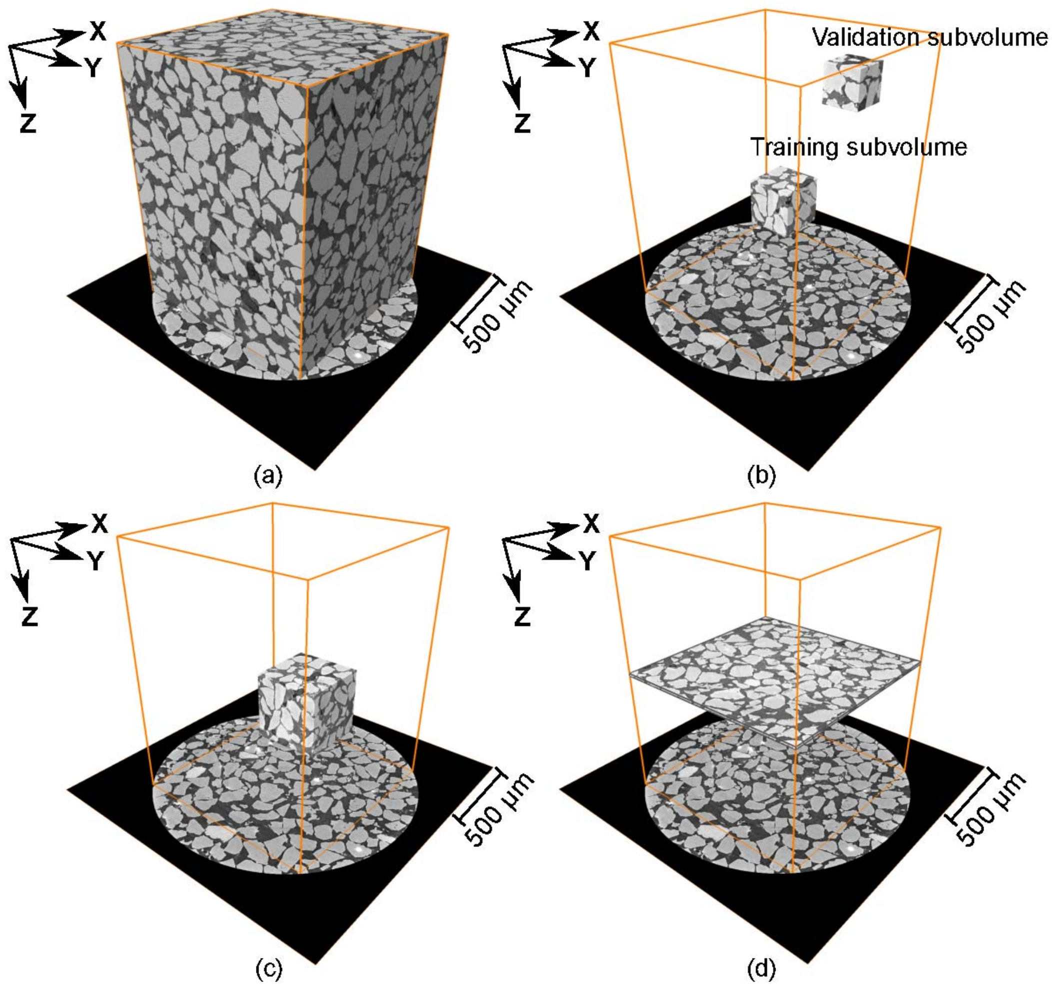

Figure 4. For data set 89062: **(a)** 1554×1554×2000 voxel central region used for U-Net segmentation; **(b)** Location of 384×384×384 and 256×256×256 voxel training and validation volumes, respectively; **(c)** 572×572×572 voxel training and validation subvolume; **(d)** Central 40 slices used for quantitative analysis.

2.3.1 Training and Validation Data

U-Net training and validation data sets were created from subregions of the 1554×1554×2000 volumes. The 3D hierarchical approach used a 384×384×384 voxel training sub-volume and a 256×256×256 voxel validation sub-volume selected from a different region of the 3D image. RootPainter software requires 2D images (slices) of at least 572×572 pixels in size for training and validation. Therefore, a 572×572×572 sub-volume was delimited for this purpose. The same sub-volume was used to train the 2D multilabel U-Net models. The training and validation sub-volume coordinate origins relative to the global origin of the reconstructed 2560×2560×2000 data set are listed in **Table 2**. The global coordinate system origin is indicated in **Figure 4**, which also presents the location of the training and validation volumes (**Figure 4b-c**).

The U-Net training procedures required both greyscale and label data sets. The latter was the 'ground truth' information used during training and validation. The label data sets were produced by manually annotating the sand, CH_4 gas and brine-hydrate regions of each slice using Avizo Lite® software. All training, validation and segmented data used in this investigation are available in Alvarez-Borges et al. (2021).

Table 2. *Training and validation sub-volume origin voxel coordinates relative to global origin of the 2560×2560×2000 volume (shown in **Figure 4**).*

Size (voxels)	X	Y	Z
256×256×256	1133	1753	50
384×384×384	1343	943	1158
572×572×572	1343	943	1158

2.3.2. 3D Hierarchical

The 3D hierarchical U-Net model used was implemented in the Python library PyTorch (Paszke et al., 2019) and based upon an existing implementation of a residual 3D U-Net from the literature (Lee et al., 2017; Wolny et al., 2020). The voxel datatype of the training and validation sub-volumes was rescaled from 16-bit to 8-bit depth. To mitigate the skewing effect of extreme outliers, voxel intensities were clipped to be within 2.575 standard deviations of the mean before rescaling. This cut-off incorporated 99% of all values in a normally distributed range of intensities. The ground truth label volumes (with three labels: sand, brine-hydrates and CH_4 gas) were used to create separate binary label volumes, one with sand vs background and the other with CH_4 gas vs background. These volumes were used as the label data for training the separate binary 3D U-Net models.

Unlike the multilabel 2D U-Net implementation described later, this model had not been pre-trained on ImageNet and was therefore likely to require a larger amount of training data to reach a high segmentation accuracy. To overcome this, the TorchIO library (Pérez-García et al., 2020) was used to sample 128×128×128 voxel sub-volumes from the $(384)^3$ voxel training data. For each training epoch (i.e., a full training cycle), 48 sub-volumes were generated with random noise, flips, blurs, affine and elastic transformations. In addition, the validation volume was randomly sampled, creating 12 sub-volumes for mode validation after each training epoch. During training, parameter optimization was carried out with a variant of adaptive moment estimation with decoupled weight decay, known as AdamW (Loshchilov & Hutter, 2019). The learning rate was cycled up and down every epoch (Smith, 2017). Binary cross entropy was used as the loss function and mean Intersection Over Union (IOU) was used as the evaluation metric.

Monitoring of the validation loss was used as the basis for an early stopping regime during training. If either no improvement in validation loss occurred after 40 training epochs or 100 epochs were completed, the model with the lowest validation loss was saved. This was aimed at preventing overfitting. Final training metrics are given in the supporting material (Table S1). Software source code for this method is available from King & Alvarez-Borges (2021).

When predicting segmentation for the 1554×1554×2000 SRXCT volumes, two binary predictions were produced for each data set, one for sand vs background and the other for CH_4

gas vs background. In both cases the image data was split into blocks of 192×192×192 voxels with an overlap of 32 voxels between blocks using the TorchIO library before being fed into the U-Net for label prediction. These two volumes were then combined using a label hierarchy: first, a new 1554×1554×2000 volume was created with all voxel labels set to brine-hydrates, then the labels corresponding to CH_4 gas were transferred from the CH_4 vs background prediction, and lastly the labels corresponding to sand were transferred from the sand vs background prediction.

2.3.3. 2D Multilabel

Training of the 2D U-Net with multiple labels was performed on the $(572)^3$ voxel sub-volume using two approaches. The first mimicked that of RootPainter, with the network being trained on horizontal 2D (XY) slices through the image volume. The second, multiplane approach, utilized slices taken in the XY, XZ and YZ planes (coordinate system shown in **Figure 4**). Prior to training, for both approaches, the voxel intensities in the selected volume were rescaled to 8-bit depth, as in the 3D hierarchical method. A 2D U-Net was used with a ResNet34 encoder (He et al., 2016). This encoder was loaded with pre-trained weights from ImageNet. The model was created with Fastai (Howard & Gugger, 2020), a Python library which has a high-level interface that utilizes PyTorch. During training, default Fastai image transformations and augmentations were used. The loss function used was cross entropy and the evaluation metric used was the number of correctly labelled voxels expressed as a percentage. Training was carried out for 15 epochs.

For the single-plane implementation, the XY training stack and corresponding label stack of 572 images, with dimensions 572×572, were split into training (80%) and validation (20%) sets. When predicting the segmentation for the 1554×1554×2000 SRXCT volumes, data was fed into the network in the form of 2000 XY slices of size 1554×1554 pixels.

For the multi-plane approach, the training data and corresponding label volume with dimensions 572×572×572 voxels were sliced into 2D images in the XY, XZ and YZ planes, resulting in 1716 training image and label pairs. These images were also split into a training (80%) and validation (20%) set. When predicting the segmentations for the 1554×1554×2000 SRXCT volumes, an averaging approach for data produced from each plane was used as described by Tun et al. (2020), but with a modification to take the multiple labels into account. In short, this averaging approach consisted in slicing, segmenting, and rotating the SRXCT volume across the XY 4-fold symmetry plane and then splitting and hierarchically recombining the 12 resulting segmentation volumes so that two label volumes were obtained, one containing labels for sand vs background and the other for CH_4 vs background. These two binary label volumes were then combined into a multilabel volume in a similar hierarchical manner as for the data output from the 3D hierarchical method described in Section 2.3.2. The averaging approach is further described in the supporting information (Text S4).

Final training metrics for both the single- and multi-plane approaches are also given in the supporting information (Table S2). Software source code for this method is available from King & Alvarez-Borges (2021).

2.3.4. RootPainter

RootPainter (Smith & Ørting, 2020) is a client-server application originally developed to segment plant root features from photographs of soil profiles (Smith, Han et al., 2020; Smith, Petersen et al., 2020). The client GUI is employed to annotate 2D images from a dataset, such as a tomography image stack of horizontal (XY) slices, as in the present case. The tomography slices and corresponding annotations are then read by the server and used to train the segmentation model using a U-Net variant implemented in PyTorch and described by Smith, Petersen et al. (2020) and Smith, Han et al. (2020). To execute the training routine, the software creates a validation dataset by randomly selecting one annotation image out of every five created. The accuracy of the model produced at the end of each training epoch is evaluated using the F-score parameter described by Smith, Petersen et al. (2020). At the end of each training epoch, F-score values for the current and previous model are compared and the one with the highest value is saved. Training is stopped if 60 epochs are completed without improvements in F-score.

A feature that distinguishes RootPainter is the use of human intervention via interactive corrections. These are carried out by the user by annotating slices overlaid with the segmentations produced by the best model available at that moment during the training process. The annotations are targeted to 'correct' erroneously labelled pixels. These corrective annotation slices are added to the training and validation datasets so that the five to one ratio is maintained.

At present, RootPainter can only predict binary segmentations with one material label, termed 'foreground'. The rest of the image is considered 'background'. Thus, RootPainter was initially used to segment the CH_4 gas phase only, which exhibited limited contrast with regards to the brine-hydrate phase (**Figure 1(a)**). The $(572)^3$ voxel label sub-volume created in Avizo Lite® was used for training and validation. A procedure described in the supporting information was applied to produce arbitrarily sparsely annotated images from the label data (Text S5), as Gonda et al. (2017) and Smith, Han et al. (2020) suggest that sparse annotations produce better results than dense/intensive annotations when interactively training a U-Net. This procedure essentially converted all CH_4 gas labels into foreground and enclosed them with RootPainter background labels that included brine-hydrates and sand pixels, as shown in **Figure 5(a-b)**. The annotated slices were then copied in batches of five into the RootPainter annotations folder. One slice from each batch was copied into the validation folder as well, to maintain the five to one ratio. A user could alternatively annotate the material of interest using the GUI, as explained by Smith, Han et al. (2020). Training was initiated via GUI command after copying the first image batch. Further batches were added if a training epoch finalized without further improvements in F-score and the model could not segment most CH_4 pixels, or if the erroneously segmented pixels were patently greater than the correctly segmented ones, as shown in **Figure 5(c)**. Corrective annotation was started after a training epoch produced a model that segmented most of the CH_4 regions with a roughly equivalent number of erroneously labelled pixels, as presented in **Figure 5(d-e)**.

Once a model was produced that could segment CH_4 without error pixels that could be visually identified, the software was left to carry on training until the 60-epoch limit was reached. The resulting model was then used to segment the 1554×1554×2000 SRXCT volume.

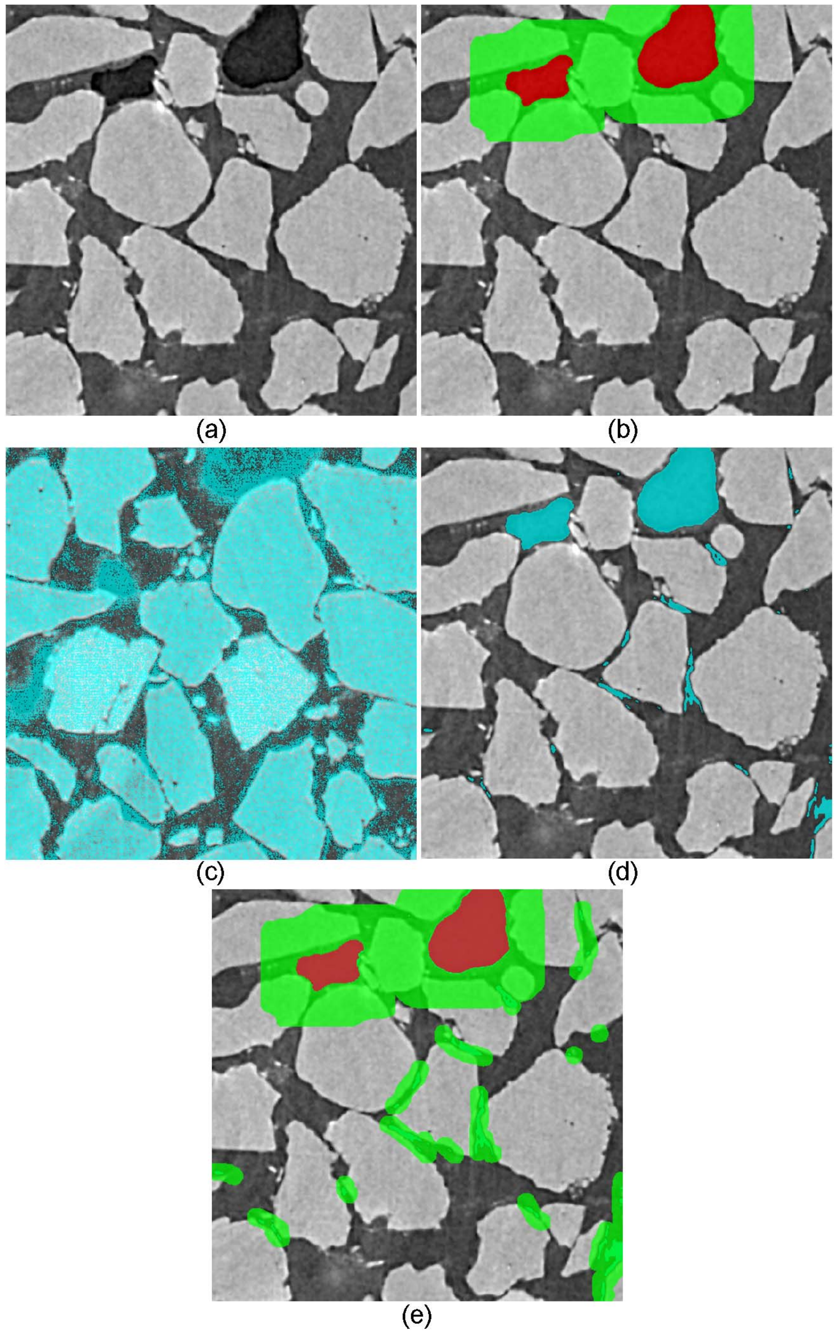

Figure 5. RootPainter usage example (on data from 89062): **(a)** XY slice from $(572)^3$ sub-volume; **(b)** Slice annotations used for training and validation with CH_4 (foreground) shown in red and background shown in green; **(c)** Initial segmentation output (blue) with a large number of erroneously labelled voxels; **(d)** Improved segmentation with a small number of erroneously labelled voxels; **(e)** Annotative corrections on mislabeled voxels.

2.3.5. Quantitative Analysis

The central 40 XY slices of the U-Net-segmented 1554×1554×2000 volumes were compared with manually annotated counterparts created in Avizo Lite® and considered to represent 'ground truth' labels. These ground truth volumes are available in Alvarez-Borges et al. (2021). The previously mentioned IOU metric, also known as the Jaccard Index (Jaccard, 1901), was used to evaluate segmentation performance. IOU is defined as:

$$\mathrm{IOU} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN} + \mathrm{FP}} \tag{1}$$

where TP refers to the number of voxels or pixels correctly predicted to correspond to the label of interest ('true positive'), and FP and FN are the number of voxels or pixels incorrectly predicted to be part of the label of interest ('false positive') and voxels/pixels incorrectly predicted to belong to any of the other material phases ('false negative'), in each case. A comparable analysis of U-Net accuracy has been done by, e.g., Karabağ et al. (2020).

IOU returns a value between 0 and 1, where the latter corresponds to the scenario were the segmentation matches the validation image pixel by pixel (or voxel by voxel).

3 Results and Discussion

3.1. U-Net Performance Comparison

Figure 6 compares the original and segmented central slice for dataset 89062, produced using each of the three methods described in Section 2.3. Training in both the 2D multilabel approach and RootPainter was carried out using XY slices only (i.e., single plane). **Figure 7a** presents segmentation accuracy metrics for the central 40 XY slices of dataset 89062. It may be noted that RootPainter delivered slightly higher metrics than the other two methods, but this difference in performance cannot be readily identified in **Figure 6**.

The slightly lower performance metrics observed in **Figure 7(a)** for the 3D hierarchical output may be preliminary attributed to the smaller training sub-volume used ($[384]^3$). To present a more balanced comparison, a further 3D hierarchical model was trained on a sub-volume of the same size as the one used for both 2D methods, i.e. $(572)^3$. This comparison is presented in **Figure 7(b)**, where it is evident that RootPainter still outperformed the 3D hierarchical approach, though the difference between methods reduced.

It may also be noted from **Figure 7(a)** and **Figure 7(b)** that pre-training on the ImageNet database for the 2D multilabel method did not result in a significant segmentation performance advantage over the 3D hierarchical method. A similar outcome on the effect of transfer learning has been reported by He et al. (2019). They remarked that, ultimately, pre-training primes the U-Net for feature identification, which leads to fewer training iterations rather than greater segmentation accuracy. Such appears to be the present case, as the 2D multilabel approach produced similar results to the 3D hierarchical method with up to six times fewer training epochs (Table S1 and Table S2).

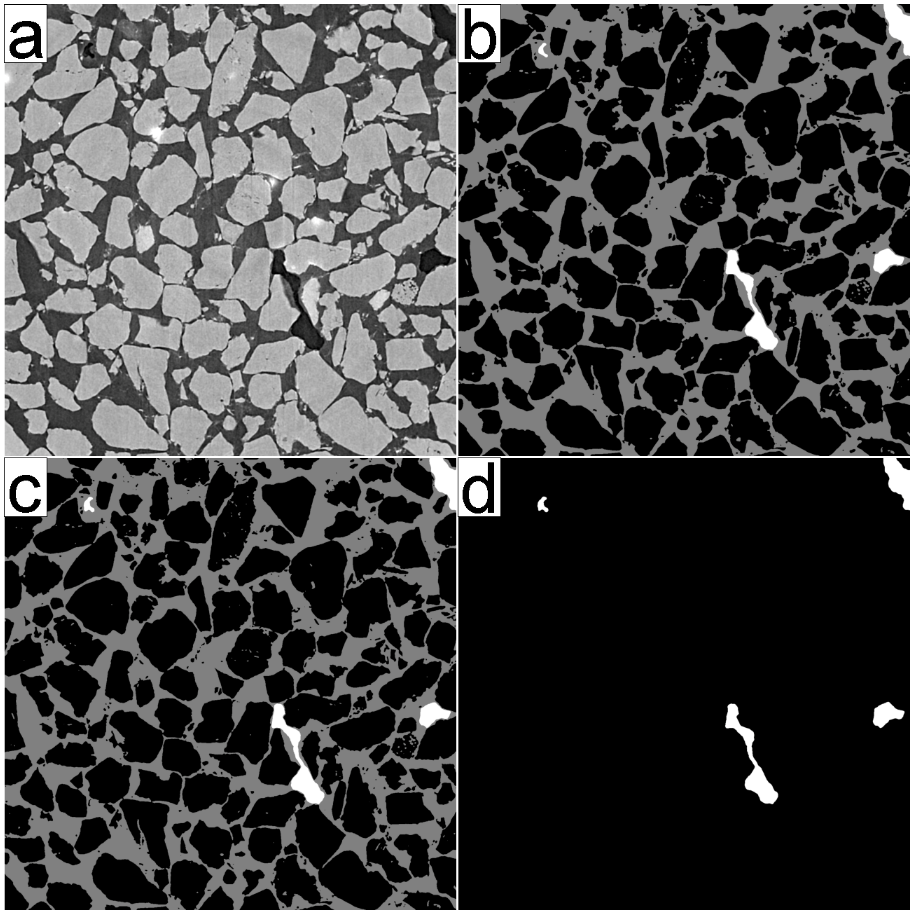

Figure 6. (a) Original XY central slice of data set 89062; **(b)** Segmented slice using the 3D hierarchical method with the $(384)^3$ training subvolume; **(c)** Segmented slice using the 2D multilabel single-plane approach; **(d)** RootPainter segmentation of the CH_4 gas phase. CH_4 gas shown in white.

A disadvantage of the use of 2D U-Net segmentation methods that operate solely with XY slices, as RootPainter, is that horizontal stripe artefacts may appear in the vertical (YZ or XY) slices of the segmented volume. This occurs because training and segmentation does not account for feature continuity between slices, that is, along the vertical (Z) axis. These artefacts, though minor for the present case, may be observed in Figure S2 of the supporting information. Such artefacts are absent in the output of the 3D hierarchical implementation, which can also be observed by considering the "smoothness" of the line showing the per-slice metrics for the 3D approach in **Figure 7**. These artefacts are naturally also present in the output of the 2D multilabel single-plane approach, but can be mitigated if the method is applied to SRXCT data slices produced from different angular directions and the output recombined into a single volume, as described in Section 2.3.3 for the 2D multi-plane method. This also improves the algorithm segmentation performance metrics, as shown in **Figure 7(b)**, but at the expense of greater computation times, as depicted in **Figure 8**.

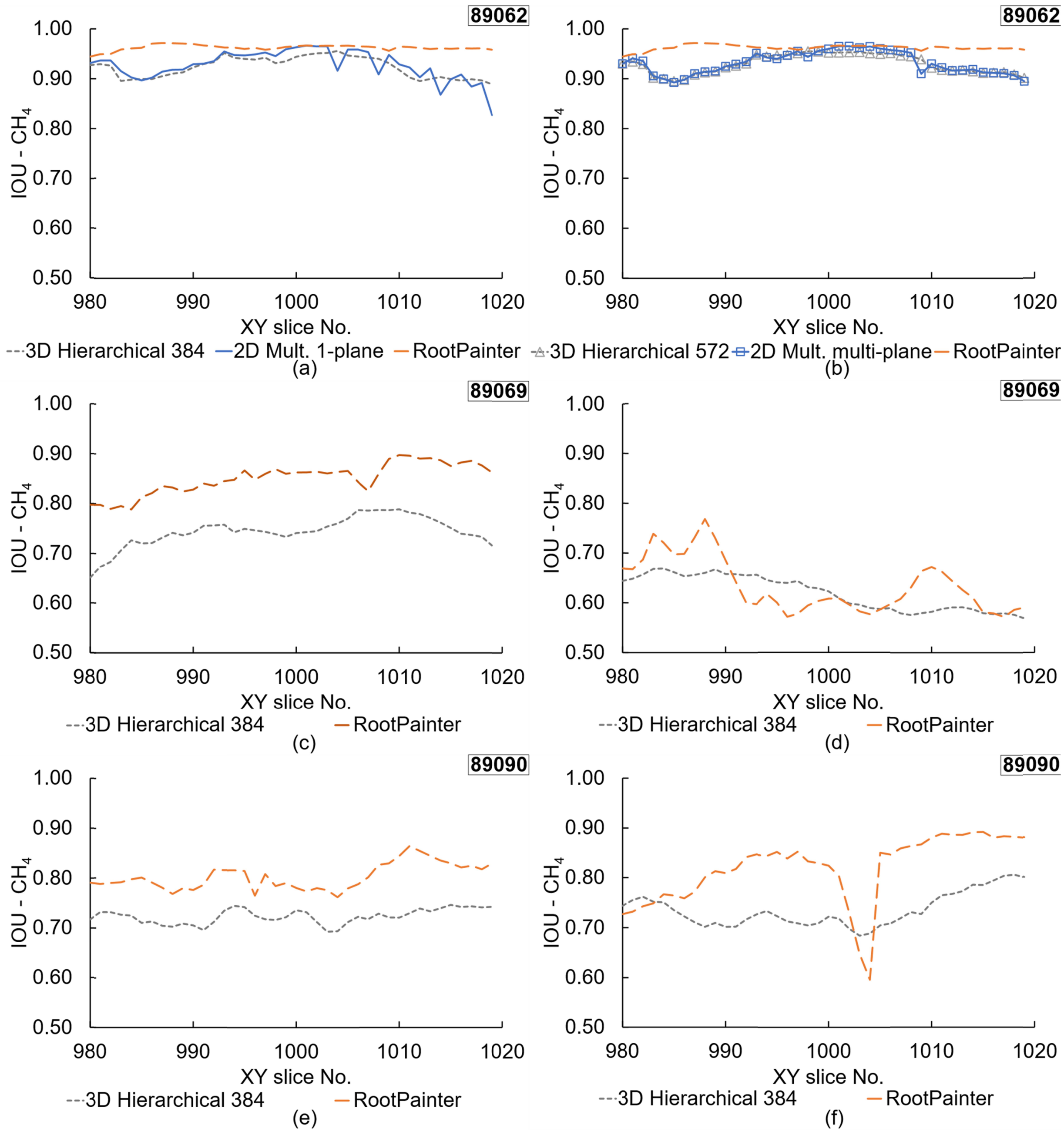

Figure 7. Performance metrics for the segmentation of CH_4 gas on the central 40 XY slices of: **(a)** 89062 using the 3D hierarchical method ($[384]^3$ voxel training sub-volume), the single-plane 2D multilabel method and RootPainter; **(b)** 89062 using the 3D hierarchical method ($[572]^3$ voxel training sub-volume), the multi-plane 2D multilabel method and RootPainter; **(c)** 89069 using the 3D hierarchical method ($[384]^3$ voxel training sub-volume) and RootPainter; **(d)** 89069 using the 3D hierarchical ($[384]^3$ voxel training sub-volume) and RootPainter U-Nets trained on data from 89062; **(e)** 89090 using the 3D hierarchical ($[384]^3$ voxel training sub-volume) and RootPainter U-Nets trained on data from 89062; **(f)** 89090 using the 3D hierarchical ($[384]^3$ voxel training sub-volume) and RootPainter U-Nets trained on data from 89069.

It is thus proposed that, in general terms, RootPainter benefits from human intervention via annotative corrections in such way that it can deliver single-label segmentations that are marginally superior to the alternative procedures. However, the alternative methods are able to (1) segment three material labels with limited user intervention, which results in less user time, and (2) deliver segmentations where horizontal stripe artefacts are largely absent, which can result in higher quality data visualization outputs.

All three methods require the manual annotation of training and validation sub-volumes, which is labor-intensive. RootPainter was able to produce the best segmentation model using only 109 slices for training and validation, including annotative correction slices (included in Alvarez-Borges et al., 2021), but the method can currently only segment one label at a time. **Figure 8** also evidences that segmenting the CH_4 gas phase using RootPainter requires similar computing resources than segmenting all three labels using the 2D single plane multilabel approach. On the other hand, while the 3D hierarchical procedure required significantly longer computing times, it produced competitive results and segmented three labels using a $(384)^3$ training and $(256)^3$ voxel validation sub-volumes, which are small compared to the size of the entire 3D image.

3.2. Performance on similar data with different greyscale contrast

The models resulting from the three U-Net implementations produce suitable segmentations when trained on subsections of the ‘intermediate’ contrast data set 89062. To examine if similar results may be obtained using data sets exhibiting lower greyscale contrast, 3D hierarchical and RootPainter U-Nets have been used to segment ‘low’ contrast data set 89069 (**Table 1**, **Figure 1(b)**), using $(384)^3$ and $(572)^3$ sub-volumes of the same data for training, respectively. **Figure 7(c)** presents the performance metrics resulting from this approach. It may be noted that both methods return lower metrics than those of the intermediate contrast data set 89062. The IOU computations show that, on average, 74% and 85% of the voxels predicted to be CH_4 were true positives in the 3D hierarchical and RootPainter results, respectively. In comparison, these average values were 92 and 94% for 89062.

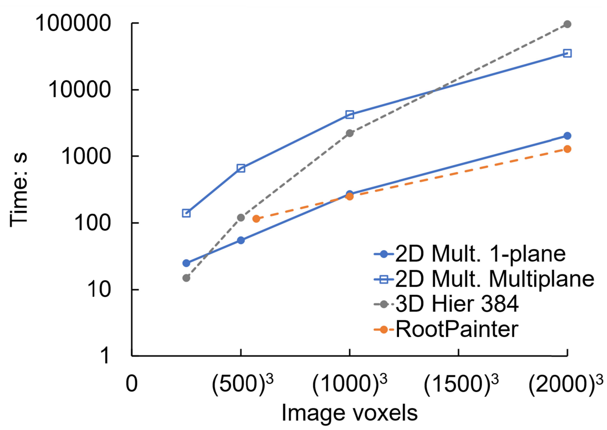

Figure 8. Segmentation time required for each method using a Nvidia Tesla V100® graphics computing unit. Benchmarking 3D images were extracted from the reconstructed and post-processed scan 89062 and are available from Alvarez-Borges et al. (2021).

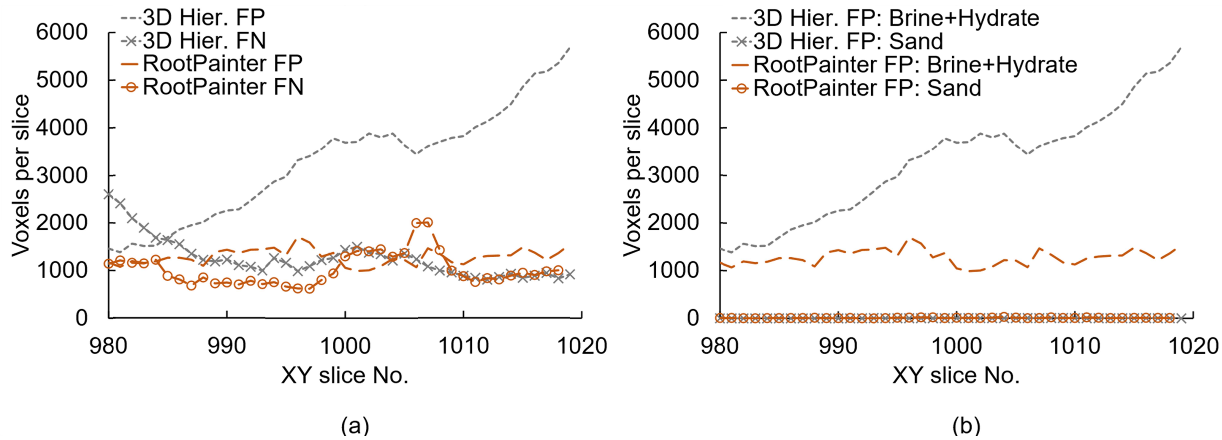

Figure 9. (a) False positive (FP) and false negative (FN) CH_4 gas voxels and **(b)** FP voxel labels per slice for the central 40 slices of data set 89069 segmented using RootPainter and the 3D hierarchical method (3D Hier).

Figure 9(a) shows that, for both methods, the lower performance metrics of the segmentation for 89069 are driven by false positives. However, false positives are over twice as numerous than false negatives in the results for the 3D hierarchical approach, whereas they only surpass false negatives by about 30% in the RootPainter segmentation. For both methods, most false positives correspond to ground truth brine-hydrate voxels incorrectly labelled as CH_4 gas, as depicted in **Figure 9(b)**. This indicates that the reduced grey value differentiation between CH_4 gas and brine-hydrate phases restricted U-Net segmentation accuracy, as anticipated.

3.3. Segmentation model suitability across data sets (model portability)

To examine if the U-Net models generated by training on a given data set produce suitable segmentations when applied to data with different greyscale contrast, 'low' and 'high' contrast data sets 89069 and 89090 (**Table 1**, **Figure 1(b)**) have been segmented using the models produced from training on 'intermediate' contrast data set 89062. **Figure 7(d, e)** presents the performance metrics of the resulting segmentations. It may be seen that segmentation accuracy is lowest for the case were the U-Net models from 89062 were applied to the lower contrast data set 89069. However, **Figure 7(a)** and **Figure 7(e)** show that the 89062 U-Net models produced segmentations for high-contrast 89090 that were of notably lower accuracy than those obtained for medium-contrast 89062, that is, for the dataset from which a sub-volume was used to train the model. That said, the performance metrics for the 89090 segmentations produced with 89062 models are comparable to those for the segmentation of low-contrast data 89069 produced with 89069-trained models (**Figure 7(c)**).

As segmentation performance appears to be higher when U-Net models trained on lower contrast data are used to segment higher contrast data, models trained on low-contrast 89069 images have been used to segment high-contrast data set 89090. Performance metrics are presented in **Figure 7(f)**. This Figure shows an overall improvement in performance metrics compared with segmentations produced with the 89062-trained U-Net models (**Figure 7(e)**).

However, a distinct poor performance for RootPainter can be observed in the profiles of **Figure 7(f)**, which resulted from a cluster of FP pixels (see Figure S3 in the supporting information). This denotes a broadly similar pattern of FP-driven model inaccuracy as for the results discussed previously in section 3.2 (**Figure 9**).

3.4. Applications and implications

The segmentation of XCT or SRXCT images of soil and rock samples is often carried out to determine parameters such as porosity or liquid/gas saturation, as discussed in Section 1. The varying performances of the U-Net methods used in the present investigation result in differences in the parameters calculated from the segmented images. This is exemplified in **Figure 10**, which compares porosity and CH_4 gas saturation ratios derived from the segmented volumes produced with the 3D hierarchical approach ($[384]^3$ training sub-volume) and RootPainter, which were the procedures that seemed to provide the best results with the least user time. Porosity was calculated as:

$$\text{Porosity (\%)} = \frac{volume\ of\ pores}{total\ volume} \times 100 \tag{2}$$

and CH_4 gas saturation was determined as:

$$\text{CH}_4\text{ saturation (\%)} = \frac{volume\ of\ \text{CH}_4}{volume\ of\ pores} \times 100 \tag{3}$$

where the volume of CH_4 gas amounts to the total number of CH_4 gas voxels, the volume of pores is the sum of CH_4 gas and brine-hydrate voxels, and the total volume is the total number of voxels in the image, in all cases multiplied by the voxel volume (0.8125×0.8125×0.8125 μm). These calculations were carried out on a slice by slice basis. For the RootPainter method, the sand phase has been segmented using the same approach used for CH_4 described in Section 2.3.4, but using sand labels and only one quadrant of each annotation slice to produce sparsely annotated training and validation images. Results presented in **Figure 10** correspond to two application cases, that is:

1. U-Nets trained on sub-volumes of the data set of interest and then used to segment the entire data set, shown in **Figure 10(a-d)**. As discussed in Section 3.5, differences in greyscale contrast affect the performance of the resulting segmentation.
2. U-Nets trained on sub-volumes of a low-greyscale contrast data set and then used to segment other data sets of higher greyscale contrast (model portability). This is presented in **Figure 10(e-f)**, corresponding to parameters derived for high-contrast data set 89090 using segmentations produced from U-Nets trained on sub-volumes of low-contrast data set 89069.

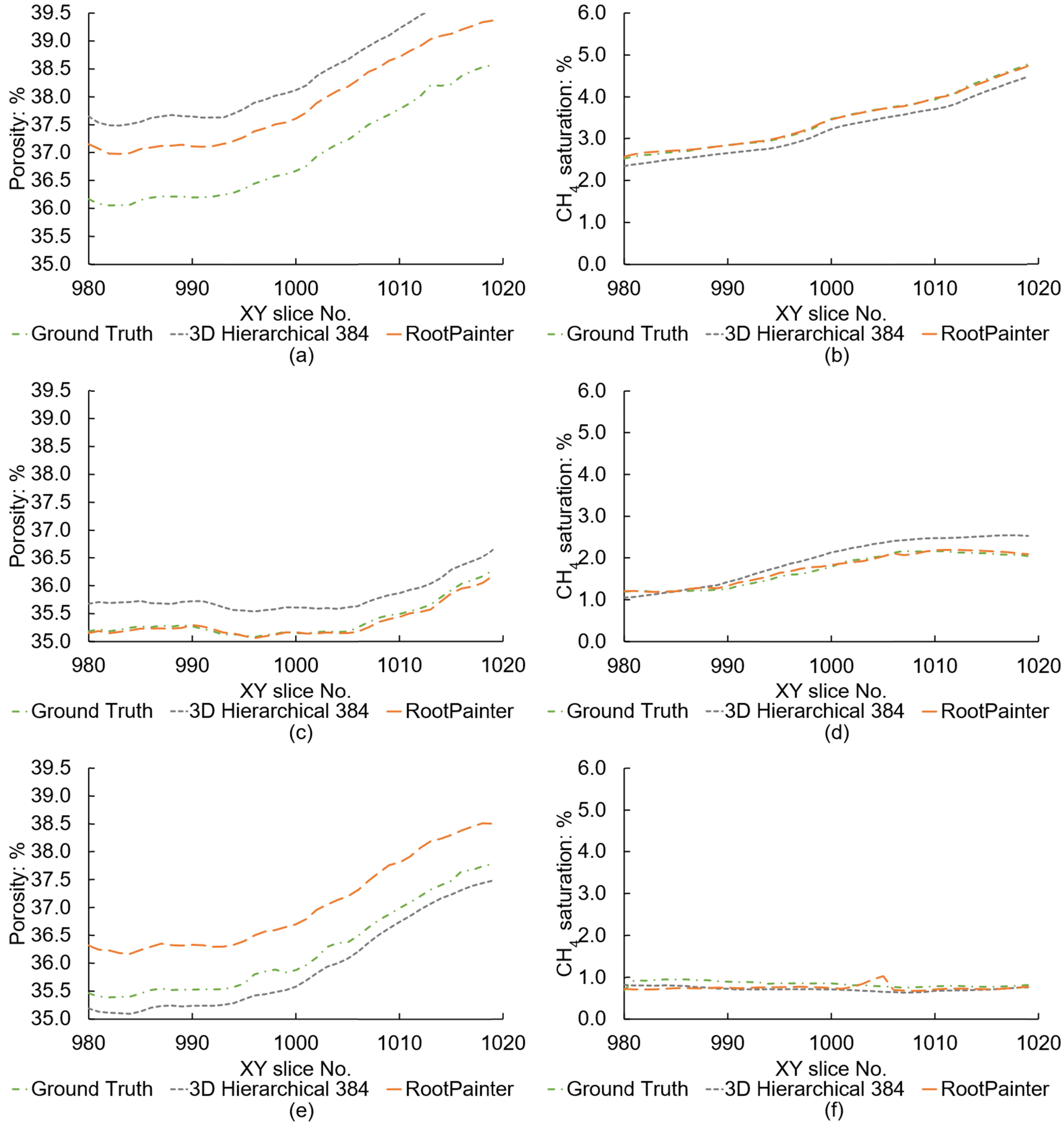

Figure 10. Porosity and CH_4 gas saturation profiles for the central 40 XY slices of data sets 89062 **(a, b)**, 89069 **(c, d)** and 89090 **(e, f)** derived using image segmentations obtained from 3D hierarchical and RootPainter U-Net models trained on sub-volumes of 89062 **(a, b)** and 89069 **(c-f)**.

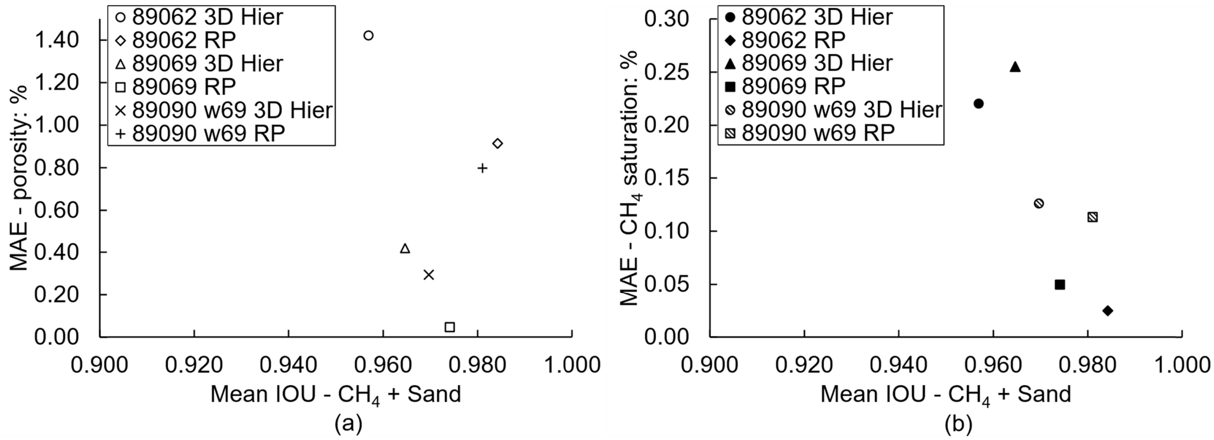

Figure 11. Comparison of mean absolute errors for **(a)** porosity and **(b)** CH_4 gas saturation estimations with mean IOU metrics for the segmentations used. w69 denotes the use of a U-Net model trained on a sub-volume of low-contrast dataset 89069; RP refers to RootPainter.

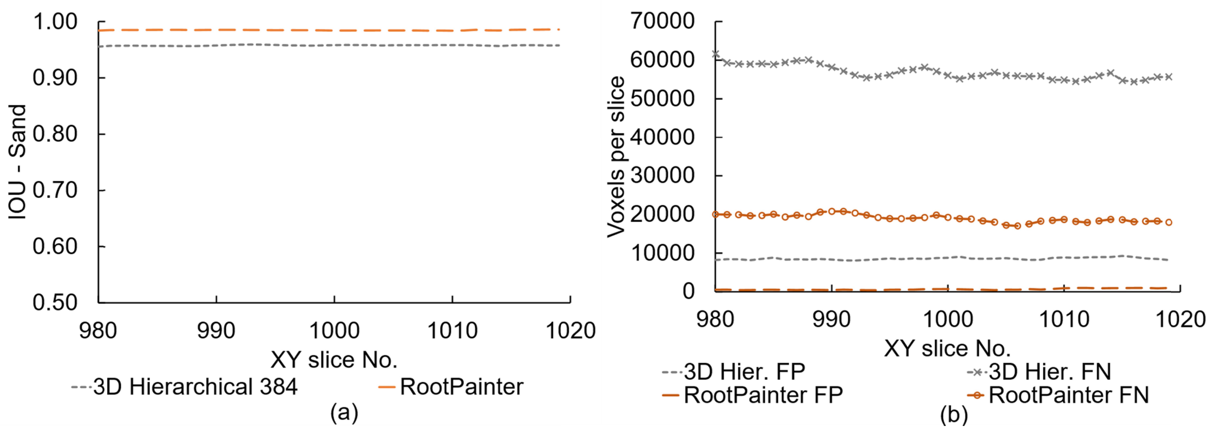

Figure 12. (a) IOU metrics for the segmentation of sand in data set 89062 using the 3D hierarchical method ($[384]^3$ training sub-volume) and RootPainter, and **(b)** associated false positive (FP) and false negative (FN) sand voxels per slice of the central 40 XY slices.

Figure 10 suggests that, while the segmentation of 89062 using U-Net models trained on a sub-volume of the same data produced high performance metrics for the CH_4 gas phase, the derived porosity and CH_4 saturation parameters deviated from ground truth values to some extent. A comparison between the mean absolute error (MAE) for porosity and CH_4 saturation calculations with mean IOU values for the combined CH_4 gas and sand labels of the three segmentation volumes used to generate **Figure 10** is shown in **Figure 11**. This Figure reveals that, while there is a general trend of lower MAE with higher segmentation accuracy, the correlation exhibits some scatter. Considering that both CH_4 gas saturation and porosity are in part derived using the number of sand voxels and that these are significantly more numerous than pore voxels (CH_4 gas and brine-hydrates), it may be proposed that errors in porosity/CH_4-saturation estimation originate from inaccuracies in the segmentation of the sand phase. This is evidenced in **Figure 12** for 89062, which presents **(a)** IOU metrics for the segmentation of the sand phase and **(b)** the number of FP and FN voxels. **Figure 12(a)** reveals that the inaccuracies in the segmentation of the sand phase are relatively small in terms of metrics, which are in fact higher than those of the CH_4 gas phase presented in **Figure 7(a)**. However, **Figure 12(b)** shows that the number of FP and FN voxels is large compared to the size of the CH_4 gas and brine-hydrate phase, which amounts to roughly 8.75×10^5 voxels per slice. This, in turn, affects parameters calculated from voxel counts. This denotes that the estimation of soil parameters based on ratios between material phases from segmented images is particularly sensitive to the relative size of said phases. Nevertheless, it should be noted that the maximum absolute errors presented in **Figure 11** (1.40% and 0.26% for porosity and CH_4 gas saturation, respectively) are smaller than those commonly reported for laboratory methods (Matula et al., 2016; Missimer & Lopez, 2018; Péron et al., 2007).

A further application for U-Net segmentations XCT/SRXCT images of soil and rock is 3D data visualization, which can then be used to investigate, for instance, CH_4 gas distribution within the pore matrix. Such application can greatly benefit from model portability. To exemplify this, **Figure 13** compares 3D views of the CH_4 gas phase produced by segmenting data sets obtained at different stages of hydrate formation using the RootPainter model trained on the low-contrast 89069 sub-volume. The model produces sensible 3D representations of the data, and changes in CH_4 gas distribution as it is consumed for hydrate formation can be clearly distinguished. In a further example, a 2D multilabel U-Net, trained using the single-plane approach on a $(572)^3$ volume from scan 89062, has been used to segment a higher-contrast SRXCT scan from a similar experiment carried out at the Swiss Light Source (SLS) originally reported by Sahoo, Madhusudhan et al. (2018). The post-processing steps described in Section 2.2.2, except beam hardening correction, were applied to the reconstructed data and a 1554×1554×2000 voxel region was extracted from the center of the 3D image (data is available from Alvarez-Borges et al. (2021)). Results are shown in **Figure 14**, where it is seen that the model delivers qualitatively accurate 3D views of the distribution of all three material phases, without any additional training or user input.

Both examples demonstrate the capability of U-Net models to segment multiple SRXCT images of CH_4-bearing soil, despite being obtained with different scan set-ups. The U-Net models used only a single (572)3 voxel sub-volume for training and did not require any additional training or user input to segment new images. A key implication is that training of a single U-Net model on a low greyscale contrast data set could be used to deliver insight on hydrate-induced variations in sediment morphology in other data sets. This has valuable applications. For example, segmentations are often required over a short period of time with limited operator input, like during data acquisition at a synchrotron or other X-ray facility. The availability of pre-trained U-Net models would allow to produce segmentations and sediment morphology/microstructure information within a short time after acquisition and reconstruction. Pre-trained models could also be used to segment numerous and/or large data sets with less user effort and bias and over shorter timespans.

4 Conclusions

The application of U-Nets, a class of Convolutional Neural Network, to segment SRXCT images of CH_4-bearing sand has been investigated. The general aim was to determine if U-Nets were capable of accurately segmenting SRXCT data of different greyscale contrast, with focus on the CH_4 gas phase, using a small number training images ($\leq [572]^3$ voxels). Training images were obtained from hand-annotated sub-volumes of the reconstructed SRXCT data. Three U-Net deployment methods were used: 3D hierarchical, 2D multilabel and RootPainter. Quantitative comparisons of image segmentation were carried out using the IOU metric. Major outcomes of this investigation are presented below.

1. For a given SRXCT data set, the three U-Net deployment methodologies produced models capable of delivering segmented images of the CH_4 gas phase with average IOU metrics above 0.740. This demonstrated that the U-Net methods used were reasonably capable of accurately identifying the CH_4 gas phase using a small number of training images. RootPainter delivered marginally higher IOU metrics than the other methods but suffered from minor horizontal stripping artefacts and required proportionally higher computing time.

2. Greyscale contrast between material phases in the different SRXCT data sets was a significant factor affecting segmentation accuracy. The lowest segmentation performance metrics corresponded to SRXCT data sets exhibiting the lowest greyscale contrast, while greater segmentation accuracy resulted from the use of higher contrast data.

3. Model portability, i.e. the segmentation of a given SRXCT data set using a U-Net model trained on a sub-volume of a different data set, was explored. It was found that models trained on lower-contrast images were able to produce accurate segmentations of higher-contrast data. In comparison, U-Net models trained on higher-contrast images were found to deliver poor results when used to segment lower-contrast data.

4. The effect of segmentation accuracy on image-derived material parameters was investigated by calculating porosity and CH_4 gas saturation profiles using U-Net segmentations. A general trend of lower mean absolute error of the derived parameter with greater segmentation accuracy was found, but the correlation exhibited some scatter. Considering that porosity, fluid saturation and other parameters are ratios between material phases, it was proposed that errors in U-Net derived parameters are not only linked to segmentation accuracy metrics but to the number of false positive and negative voxel labels of the largest phase relative to the other phases.

5. It was found that a single U-Net model could be used to segment multiple SRXCT data sets and produce qualitatively accurate 3D views of the sediment matrix and CH_4 gas distribution during hydrate formation without additional training, even when using independent data from other X-ray imaging facilities.

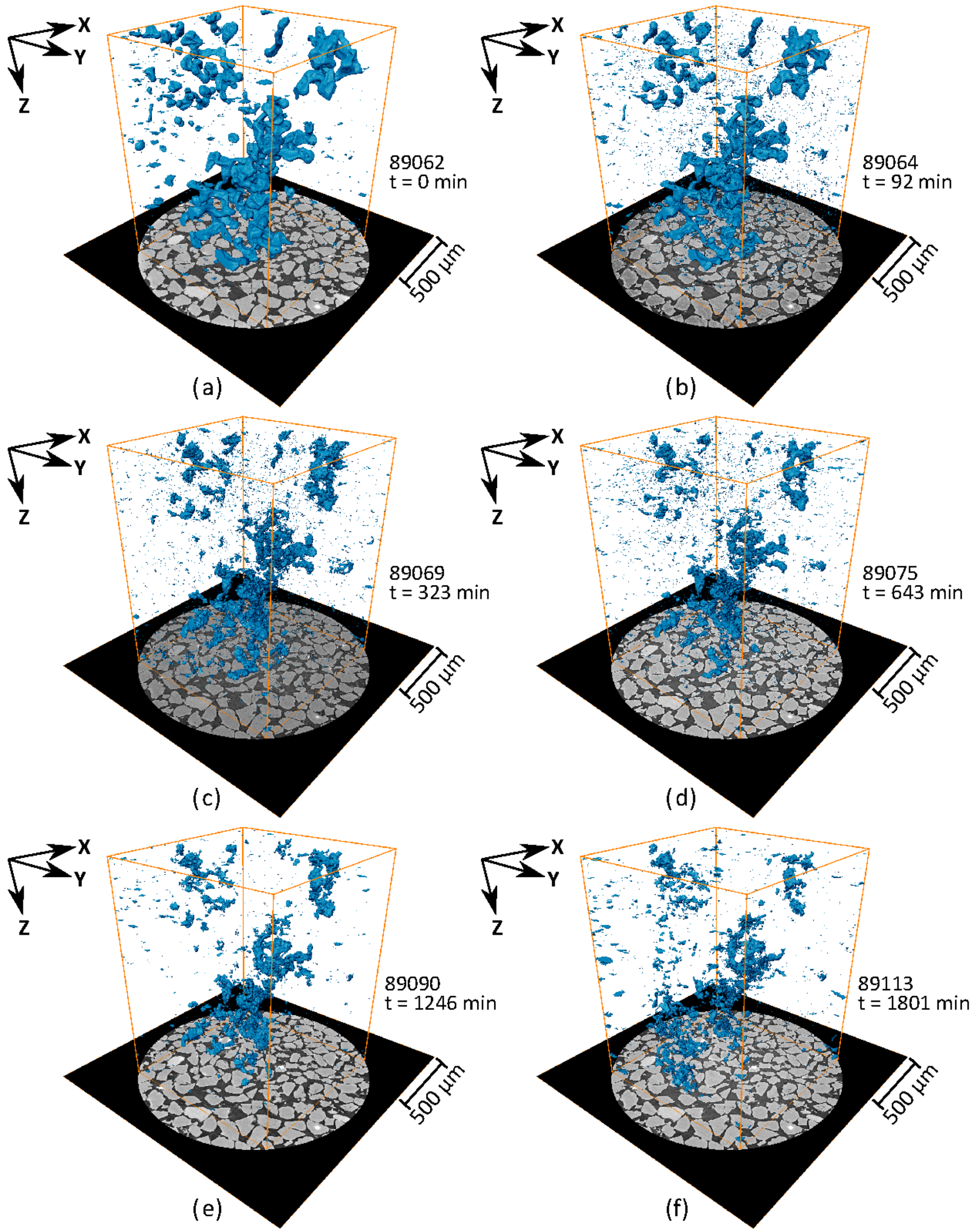

Figure 13. 3D views of the CH_4 gas phase segmented using a RootPainter U-Net model trained on the low-contrast 89069 data (t denotes cooling time in minutes after reaching 2°C).

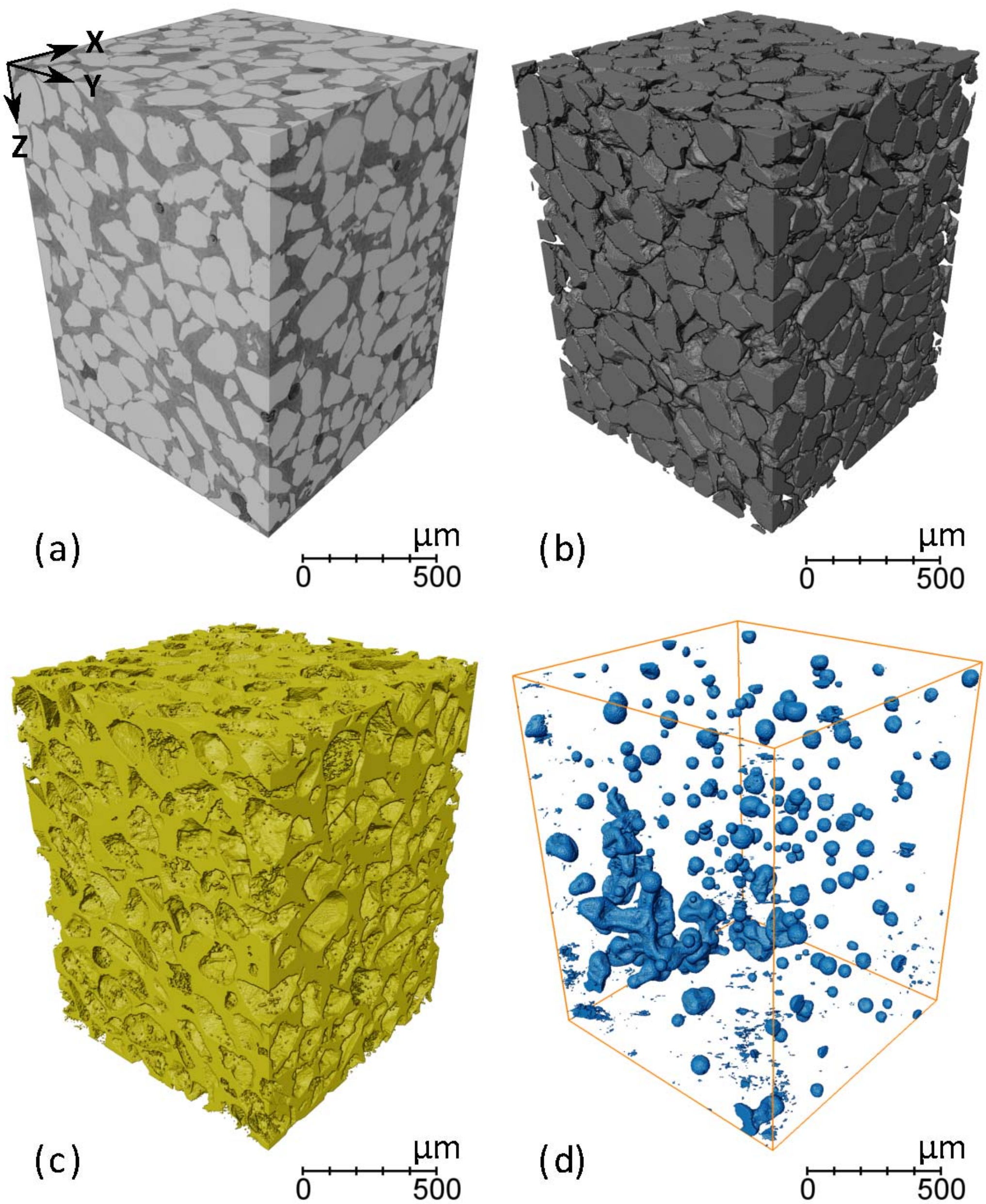

Figure 14. U-Net segmentation of independent data set from Sahoo, Madhusudhan et al. (2018) acquired at SLS, using a 2D multilabel single-plane U-Net model trained on a $(572)^3$ sub-volume of data set 89062: (a) reconstructed SLS volume; (b) sand; (c) brine-hydrate; (d) CH_4 gas.

Acknowledgments, Samples, and Data

This research was carried out with the support of Diamond Light Source, the University of Southampton and the Natural Environment Research Council (UK) grant No. NE/K00008X/1 awarded to B.N. Madhusudhan. The authors gratefully acknowledge Diamond Light Source for the provision of beamtime (proposal MT16205-1) and are much thankful for the support provided by beamline I13-2 staff. Datasets for this research are available in this in-text data

citation reference Alvarez-Borges et al. (2021) under an Apache V2 License. Software for this research is available in these in-text data citation references: Smith & Ørting (2020) [under a GNU General Public License v3.0] and King & Alvarez-Borges (2021) [under an Apache V2 License].

References

Alvarez-Borges F.J., King O.N.F., Madhusudhan B.N. & Ahmed S.I. (2021). Tomography data of methane-bearing sand used to investigate U-Net segmentation methods [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.4580278

Atwood R.C., Bodey A.J., Price S.W.T., Basham M. & Drakopoulos M. (2015) A high-throughput system for high-quality tomographic reconstruction of large datasets at Diamond Light Source. *Philosophical Transactions of the Royal Society A, 373 (2043)*, 2369–2393. https://doi.org/10.1098/rsta.2014.0398

Baveye P.C., Laba M., Otten W., Bouckaert L., Dello Sterpaio P., Goswami R.R., et al. (2010) Observer-dependent variability of the thresholding step in the quantitative analysis of soil images and X-ray microtomography data. *Geoderma, 157 (1)*, 51-63. https://doi.org/10.1016/j.geoderma.2010.03.015

Brunke O., Brockdorf K., Drews S., Müller B., Donath T., Herzen J. & Beckmann F. (2008) Comparison between x-ray tube-based and synchrotron radiation-based uCT. *Proceedings of SPIE 7078, Optical Engineering + Applications*. Bellingham, WA, USA: SPIE. https://do.org/10.1117/12.794789

Chauhan S., Rühaak W., Anbergen H., Kabdenov A., Freise M., Wille T. & Sass I. (2016) Phase segmentation of X-ray computer tomography rock images using machine learning techniques: an accuracy and performance study. *Solid Earth, 7 (4)*, 1125-1139. https://doi.org/10.5194/se-7-1125-2016

Chauhan S., Rühaak W., Khan F., Enzmann F., Mielke P., Kersten M. & Sass I. (2016) Processing of rock core microtomography images: Using seven different machine learning algorithms. *Computers & Geosciences, 86*, 120-128. https://doi.org/10.1016/j.cageo.2015.10.013

Chauhan S., Sell K., Rühaak W., Wille T. & Sass I. (2020) CobWeb 1.0: machine learning toolbox for tomographic imaging. *Geoscientific Model Development, 13 (1)*, 315-334. https://doi.org/10.5194/gmd-13-315-2020

Dean J.F., Middelburg J.J., Röckmann T., Aerts R., Blauw L.G., Egger M., et al. (2018) Methane Feedbacks to the Global Climate System in a Warmer World. *Reviews of Geophysics, 56 (1)*, 207-250. https://doi.org/10.1002/2017RG000559

Douarre C., Schielein R., Frindel C., Gerth S. & Rousseau D. (2018) Transfer Learning from Synthetic Data Applied to Soil–Root Segmentation in X-Ray Tomography Images. *Journal of Imaging, 4 (5)*. https://doi.org/10.3390/jimaging4050065

Fonseca J., O'Sullivan C. & Coop M.R. (2009) Image Segmentation Techniques for Granular Materials. In Nakagawa M. & Luding S. (Eds.), *Powders and Grains: Proceedings of the 6th International Conference on Micromechanics of Granular Media*, pp. 223-226. Melville NY, USA: American Institute of Physics. https://doi.org/10.1063/1.3179898

Gonda F., Kaynig V., Jones T.R., Haehn D., Lichtman J.W., Parag T. & Pfister H. (2017) ICON: An interactive approach to train deep neural networks for segmentation of neuronal structures. *IEEE 14th International Symposium on Biomedical Imaging*. Piscataway NJ, USA: IEEE. https://doi.org/10.1109/isbi.2017.7950530

He K., Zhang X., Ren S. & Sun J. (2016) Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778. Piscataway NJ, USA: IEEE. https://doi.org/10.1109/CVPR.2016.90

He K., Girshick R. & Dollar P. (2019) Rethinking ImageNet Pre-Training. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4918-4927. Seoul, KR: Computer Vision Foundation.

Holland M. & Schultheiss P. (2014) Comparison of methane mass balance and X-ray computed tomographic methods for calculation of gas hydrate content of pressure cores. *Marine and Petroleum Geology, 58*, 168-177. https://doi.org/10.1016/j.marpetgeo.2014.07.016

Howard J. & Gugger S. (2020) Fastai: A Layered API for Deep Learning. *Information, 11 (2)*, 108. https://doi.org/10.3390/info11020108

Hsieh J. (2015) *Computed Tomography: Principles, Design, Artifacts, and Recent Advances* (3rd Edition), Bellingham, WA, USA: SPIE.

Iassonov P., Gebrenegus T. & Tuller M. (2009) Segmentation of X-ray computed tomography images of porous materials: A crucial step for characterization and quantitative analysis of pore structures. *Water Resources Research, 45 (9)*, W09415. https://doi.org/10.1029/2009WR008087

IPCC (2013) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. In Stocker T.F., Qin D., Plattner G.-K., Tignor M., Allen S.K., Boschung J., Nauels A., Xia Y., Bex V. & Midgley P.M. (Eds.). Intergovernmental Panel on Climate Change. Cambridge, UK and New York City, NY, USA: Cambridge University Press.

Jaccard P. (1901) Distribution de la Flore Alpine dans le Bassin des Dranses et dans quelques régions voisines. *Bulletin de la Societe Vaudoise des Sciences Naturelles, 37 (142)*, 241-72. https://doi.org/10.5169/seals-266440

James R.H., Bousquet P., Bussmann I., Haeckel M., Kipfer R., Leifer I., et al. (2016) Effects of climate change on methane emissions from seafloor sediments in the Arctic Ocean: A review. *Limnology and Oceanography, 61 (S1)*, S283-S299. https://doi.org/10.1002/lno.10307

Karabağ C., Jones M.L., Peddie C.J., Weston A.E., Collinson L.M. & Reyes-Aldasoro C.C. (2020) Semantic segmentation of HeLa cells: An objective comparison between one traditional algorithm and four deep-learning architectures. *Plos One, 15 (10)*, e0230605. https://doi.org/10.1371/journal.pone.0230605

Karimpouli S. & Tahmasebi P. (2019) Segmentation of digital rock images using deep convolutional autoencoder networks. *Computers & Geosciences, 126*, 142-150. https://doi.org/10.1016/j.cageo.2019.02.003

Kerkar P.B., Horvat K., Jones K.W. & Mahajan D. (2014) Imaging methane hydrates growth dynamics in porous media using synchrotron X-ray computed microtomography. *Geochemistry, Geophysics, Geosystems, 15 (12)*, 4759-4768. https://doi.org/10.1002/2014GC005373

King O.N.F. & Alvarez-Borges F.J. (2021) Gas Hydrate Segmentation Using U-Nets. Code Repository. https://doi.org/10.5281/zenodo.4683002

Kong D. & Fonseca J. (2018) Quantification of the morphology of shelly carbonate sands using 3D images. *Géotechnique, 68 (3)*, 249-261. https://doi.org/10.1680/jgeot.16.P.278

Koyuncu C.F., Arslan S., Durmaz I., Cetin-Atalay R. & Gunduz-Demir C. (2012) Smart Markers for Watershed-Based Cell Segmentation. *Plos One, 7 (11)*, e48664. https://doi.org/10.1371/journal.pone.0048664

Krizhevsky A., Sutskever I. & Hinton G.E. (2017) ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60 (6)*, 84–90. https://doi.org/10.1145/3065386

Kvenvolden K.A. (1993) Gas hydrates—geological perspective and global change. *Reviews of Geophysics, 31 (2)*, 173-187. https://doi.org/10.1029/93RG00268

Lee K., Zung J., Li P., Jain V. & Seung H.S. (2017) Superhuman Accuracy on the SNEMI3D Connectomics Challenge. arXiv:1706.00120 [Preprint]

Lei L., Seol Y. & Jarvis K. (2018) Pore-Scale Visualization of Methane Hydrate-Bearing Sediments With Micro-CT. *Geophysical Research Letters, 45 (11)*, 5417-5426. https://doi.org/10.1029/2018GL078507

Loshchilov I. & Hutter F. (2019) Decoupled Weight Decay Regularization. arXiv:1711.05101 [Preprint]

Madhusudhan B.N., Clayton C.R.I. & Priest J.A. (2019) The Effects of Hydrate on the Strength and Stiffness of Some Sands. *Journal of Geophysical Research: Solid Earth, 124 (1)*, 65-75. https://doi.org/10.1029/2018JB015880

Maslin M., Owen M., Betts R., Day S., Jones T.D. & Ridgwell A. (2010) Gas hydrates: past and future geohazard? *Philosophical Transactions of the Royal Society A, 368 (1919)*, 2369-2393. https://doi.org/10.1098/rsta.2010.0065

Matula S., Báťková K. & Legese W.L. (2016) Laboratory Performance of Five Selected Soil Moisture Sensors Applying Factory and Own Calibration Equations for Two Soil Media of Different Bulk Density and Salinity Levels. *Sensors, 16 (11)*, 1912. https://doi.org/10.3390/s16111912

Mienert J. (2009) Methane Hydrate and Submarine Slides. In Steele J.H. (Ed.) *Encyclopedia of Ocean Sciences*, (2nd Edition), pp. 790-798. Oxford, UK: Academic Press. https://doi.org/10.1016/B978-012374473-9.00704-9

Missimer T.M. & Lopez O.M. (2018) Laboratory Measurement of Total Porosity in Unconsolidated Quartz Sand by Two Integrated Methods. *Journal of Geology and Geophysics, 7 (5)*, 1000448. https://doi.org/10.4172/2381-8719.1000448

Moridis G., Collett T.S., Pooladi-Darvish M., Hancock S.H., Santamarina C., Boswell R., et al. (2011) Challenges, Uncertainties, and Issues Facing Gas Production From Gas-Hydrate Deposits. *SPE Reservoir Evaluation & Engineering, 14 (1)*, 76-112. https://doi.org/10.2118/131792-PA

Paganin D., Mayo S.C., Gureyev T.E., Miller P.R. & Wilkins S.W. (2002) Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object. *Journal of Microscopy, 206 (1)*, 33-40. https://doi.org/10.1046/j.1365-2818.2002.01010.x

Paszke A., Gross S., Massa F., Lerer A., Bradbury J., Chanan G., et al. (2019) PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Wallach H., Larochelle H., Beygelzimer A., d'Alché-Buc F., Fox E. & Garnett R. (Eds.), *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Red Hook, NY, USA: Curran Associates, Inc.

Pérez-García F., Sparks R. & Ourselin S. (2020) TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. arXiv:2003.04696 [Preprint]

Péron H., Hueckel T. & Laloui L. (2007) An Improved Volume Measurement for Determining Soil Water Retention Curves. *Geotechnical Testing Journal, 30 (1)*, 1-8. https://doi.org/10.1520/GTJ100167

Rogowska J. (2000) 5 - Overview and Fundamentals of Medical Image Segmentation. In Bankman I.N. (Ed.) *Handbook of Medical Imaging*, pp. 69-85. San Diego: Academic Press. https://doi.org/10.1016/B978-012077790-7/50009-6

Ronneberger O., Fischer P. & Brox T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In Navab N., Hornegger J., Wells W.M. & Frangi A.F. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science, vol 9351.*, pp. 234-241. Cham, CH: Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

Ruppel C.D. & Kessler J.D. (2017) The interaction of climate change and methane hydrates. *Reviews of Geophysics, 55 (1)*, 126-168. https://doi.org/10.1002/2016RG000534

Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., et al. (2015) ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision, 115 (3)*, 211-252. https://doi.org/10.1007/s11263-015-0816-y

Sahoo S.K., Madhusudhan B.N., Marín-Moreno H., North L.J., Ahmed S., Falcon-Suarez I.H., et al. (2018) Laboratory Insights Into the Effect of Sediment-Hosted Methane Hydrate Morphology on Elastic Wave Velocity From Time-Lapse 4-D Synchrotron X-Ray Computed Tomography. *Geochemistry, Geophysics, Geosystems, 19 (11)*, 4502-4521. https://doi.org/10.1029/2018GC007710

Sahoo S.K., Marín-Moreno H., North L.J., Falcon-Suarez I., Madhusudhan B.N., Best A.I. & Minshull T.A. (2018) Presence and Consequences of Coexisting Methane Gas With Hydrate Under Two Phase Water-Hydrate Stability Conditions. *Journal of Geophysical Research: Solid Earth, 123 (5)*, 3377-3390. https://doi.org/10.1029/2018JB015598

Saunois M., Stavert A.R., Poulter B., Bousquet P., Canadell J.G., Jackson R.B., et al. (2020) The Global Methane Budget 2000–2017. *Earth System Science Data, 12 (3)*, 1561-1623. https://doi.org/10.5194/essd-12-1561-2020

Schindelin J., Arganda-Carreras I., Frise E., Kaynig V., Longair M., Pietzsch T., et al. (2012) Fiji: an open-source platform for biological-image analysis. *Nature Methods, 9 (7)*, 676-82. https://doi.org/10.1038/nmeth.2019

Schneider C.A., Rasband W.S. & Eliceiri K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nature Methods, 9 (7)*, 671-675. https://doi.org/10.1038/nmeth.2089

Smith A.G., Han E., Petersen J., Olsen N.A.F., Giese C., Athmann M., et al. (2020) RootPainter: Deep Learning Segmentation of Biological Images with Corrective Annotation. https://doi.org/10.1101/2020.04.16.044461 [Preprint]

Smith A.G. & Ørting S. (2020) RootPainter 0.2.5. https://github.com/Abe404/root_painter

Smith A.G., Petersen J., Selvan R. & Rasmussen C.R. (2020) Segmentation of roots in soil with U-Net. *Plant Methods, 16 (1)*. https://doi.org/10.1186/s13007-020-0563-0

Smith L.N. (2017) Cyclical Learning Rates for Training Neural Networks. *2017 IEEE Winter Conference on Applications of Computer Vision*, pp. 464-472. Piscataway, NJ, USA: IEEE. https://doi.org/10.1109/WACV.2017.58

Song Y., Luo T., Madhusudhan B.N., Sun X., Liu Y., Kong X. & Li Y. (2019) Strength behaviors of CH4 hydrate-bearing silty sediments during thermal decomposition. *Journal of Natural Gas Science and Engineering, 72*, 103031. https://doi.org/10.1016/j.jngse.2019.103031

Tun W.M., Poologasundarampillai G., Bischof H., Nye G., King O.N.F., Basham M., et al. (2020) A massively multi-scale approach to characterising tissue architecture by synchrotron micro-CT applied to the human placenta. https://doi.org/10.1101/2020.12.07.411462 [Preprint]

Vanneste M., Sultan N., Garziglia S., Forsberg C.F. & L'Heureux J.-S. (2014) Seafloor instabilities and sediment deformation processes: The need for integrated, multi-disciplinary investigations. *Marine Geology, 352*, 183-214. https://doi.org/10.1016/j.margeo.2014.01.005

Varfolomeev I., Yakimchuk I. & Safonov I. (2019) An Application of Deep Neural Networks for Segmentation of Microtomographic Images of Rock Samples. *Computers, 8 (4)*, 72. https://doi.org/10.3390/computers8040072

Wadeson N. & Basham M. (2016) Savu: A Python-based, MPI Framework for Simultaneous Processing of Multiple, N-dimensional, Large Tomography Datasets. arXiv:1610.08015 [Preprint]

Wadeson N., Basham M., Parsons A., Kazantsev D., Vo N.T., Schoonjans T., et al. (2019) DiamondLightSource/Savu: Version 2.4. https://doi.org/10.5281/zenodo.3541873

Wolny A., Cerrone L., Vijayan A., Tofanelli R., Barro A.V., Louveaux M., et al. (2020) Accurate and versatile 3D segmentation of plant tissues at cellular resolution. *eLife, 9*, e57613. https://doi.org/10.7554/eLife.57613

Yokohama T., Nakayama E., Kuwano S. & Saito H. (2011) Relationship between seismic wave velocities, eletric resistivities and saturation ratio of methane hydrate using core samples in laboratory experiments. *Proceedings of the 7th International Conference on Gas Hydrates*, pp. 1830-1833. Red Hook, NY, USA: Curran Associates Inc.

Zhang X., Jia F., Luo S., Liu G. & Hu Q. (2014) A marker-based watershed method for X-ray image segmentation. *Computer Methods and Programs in Biomedicine, 113 (3)*, 894-903. https://doi.org/10.1016/j.cmpb.2013.12.025

References from the Supporting Information

Titarenko V., Bradley R., Martin C., Withers P. & Titarenko S. (2010) Regularization methods for inverse problems in x-ray tomography. *SPIE Optical Engineering + Applications* (7804). Society of Photo-Optical Instrumentation Engineers (SPIE). https://doi.org/10.1117/12.860260

Vo N.T., Drakopoulos M., Atwood R.C. & Reinhard C. (2014) Reliable method for calculating the center of rotation in parallel-beam tomography. *Optics Express, 22 (16)*, 19078-19086. https://doi.org/10.1364/OE.22.019078