3D Multiclass Digital Core Models via microCT, SEM-EDS and Deep Learning

Igor Varfolomeev[1*], *Vladimir* Svinin[1,2], and *Ivan* Yakimchuk[1]

[1]Schlumberger Moscow Research, Reservoir Potential Department, 125171 Moscow, Russia
[2]NRNU MEPHI, ICIS, 115409 Moscow, Russia

Abstract. We describe an integrated methodology for constructing a 3D multiclass model of a rock sample, based on X-ray microtomography (microCT) and quantitative evaluation of minerals (QEMSCAN) by automated SEM-EDS (Scanning Electron Microscopy, Energy Dispersive Spectroscopy). We focus on building an automated operator-independent workflow, allowing to distinguish between voxels featuring substantially different physical properties, such as void, quartz, denser and less dense clay aggregates. The workflow is demonstrated using a set of five ⌀8 mm Berea sandstone miniplugs. For each miniplug, a ~4000^3 voxel microCT image is acquired. Next, each miniplug is cut into smaller pieces, and the 45 resulting polished surfaces are subjected to the QEMSCAN analysis, producing ~4000^2 pixel mineral maps. Each mineral map is automatically spatially registered with the corresponding microCT image using an in-house surface-based algorithm. Further, the ground truth images for the supervised multiclass segmentation are constructed from the mineral maps. We compare 3D and 2D convolutional neural network (CNN) architectures with the baseline Naïve Bayes classifier, which is roughly equivalent to the approaches commonly used in practice today. We find that supervised CNN-based segmentation is fairly stable, despite microCT image quality non-uniformness and achieves higher quality scores compared to feature based and baseline approaches.

1 Introduction

Nowadays, the Digital Rock approach is a well-known technique for core analysis. Generally, it consists of numerical simulation of physical phenomena on a digital representation of a core sample.

Centimetre-scale Digital Rock models are usually created via X-ray microtomography (microCT) imaging and subsequent image segmentation. Binary solid/void segmentation using methods like Indicator Kriging [1] and Active Contours [2] is still most commonly used in practice. Such models proved to be suitable for estimating single-phase and multi-phase permeabilities of some rock samples, like Berea sandstone [3], where spatial resolution of the modern microCT scanners easily allows to resolve the pore throats. One major drawback, associated with these segmentation methods is the significant operator involvement; the operator's choices might greatly influence the ultimate result, and it might be not clear which segmentation parameters are more appropriate. Due to the nature of the hydrodynamics, the required parameter selection precision increases dramatically as the pore throat sizes get closer to the microCT resolution limit. Another drawback is that the substances featuring sub-resolution porosity could not be appropriately represented in such models. This might limit the ability to model, for example, clayey sandstones or carbonates. Both difficulties could be handled through the use of multiclass operator independent segmentation process and a suitable physical properties simulator.

Modern sophisticated simulation techniques allow to take into account a sizeable set of physicochemical properties (e.g., mineralogy, sub-resolution porosity, wettability) with respect to their volumetric spatial distributions [4,5]. In this paper, we focus on the segmentation part.

2 Related works

One of the first works, discussing the possibility of combining SEM and microCT images is [6,7]. The authors acquired microCT and QEMSCAN images of a sandstone sample, and spatially registered them. However, the segmentation technique, described by the authors is limited to a direct correlation between the microCT grayscale values and mineral types. Although such approach allows to distinguish minerals with significantly different X-ray effective linear attenuation coefficients, like quartz and pyrite, it is usually unreliable in most practical cases, like when distinguishing quartz and dense kaolinite. The contrast between different practically important minerals tends to be low on regular microCT systems, equipped with ~100 kV X-Ray tubes,

* Corresponding author: ivarfolomeev@slb.com

although it could be significantly higher on monochromatic lower-kV synchrotron-based systems.

2.1 Extracting features

Some works highlight the possibility to use additional features instead of just the grayscale value for CT image segmentation, for example, image gradients in [8]. The operator selects an area, attributed to a specific class on a 2D intensity-gradient histogram, which improves segmentation quality, compared to grayscale-only methods.

It is natural to assume that the segmentation quality could be further improved by increasing the number of the features (e.g. local average, median, variance, ...) taken into account. However, as the number of features grows, it becomes problematic to manually select ranges, attributed to specific classes. Thus, instead, operator directly specifies a set of voxels, attributed to each specific class, and the machine learning (ML) supervised segmentation task arises. The [9] was one of the first works to utilize the Fiji Trainable Weka Segmentation tool [10] for the microCT image segmentation task. The authors demonstrate that this approach allows better particle separation than the traditional, grayscale-based binary segmentation.

Nowadays, multiple software packages, similar to the Fiji Trainable Weka Segmentation are available, namely ilastik [11] and Zeiss Zen Intellesis™[12]. However, they share the same basic idea — using manual brush strokes as a training set for a ML classifier. Each tool provides a different set of feature extraction filters. Their common drawback is increased memory requirements due to in-memory architecture. Two more tools to mention are Thermo Scientific™ Amira-Avizo™ (including the XImage PAQ plugin) and ORS™ Dragonfly™, which also provide the ability to train pre-configured CNN models.

2.2 Training data

Some authors use semi-automatic segmentation results to train their ML models. In [13] authors train 2D SegNet CNN [14] using a set of 20 images, 256×256 each. The classes they consider are mostly easily distinguishable by their grayscale values, except a single phase, that got its own distinct microstructure. In [15] authors also use a "mainly grayscale-based" method to create a training set for their CNN, even though they do have two QEMSCAN images of the same sample. The motivation for such a decision is that those two slices do not provide a large enough training set. Some image registration difficulties are also mentioned. The CNN is able to almost perfectly reproduce that ground truth labeling, achieving ~99% accuracy. However, it is not entirely correct to directly compare this score with the score achieved with an independent ground-truth data (such as QEMSCAN-based).

Considerable efforts were spent on inventing a better ground truth segmentation, suitable for training ML models. Some works, discussing microCT image segmentation use an "image degradation" approach to create a ground truth segmentation and the "image to be segmented" from a single source. In [12] an already segmented volume is forward-projected into the projection domain, then shadow projections are blurred, and Gaussian noise is added to them; and finally a new volumetric image is reconstructed from these shadow projections. According to the authors, this emulates the real noise generating processes, taking place in a microCT device. However, it should be mentioned that this simple model does not take many other possible imperfectnesses of the real microCT device into account. In [16] the ground truth segmentation is produced from real full-quality X-ray projections, while the "image to be segmented" is produced from the decimated or downscaled subset of the same projections — this emulates faster microCT acquisition.

Although such approaches are well-suited for development purposes, we believe that the usage of a naturally higher-resolution SEM data represents a more promising approach.

In [17] authors use SEM-EDS to create a mineral map of the edge of a rock sample cylinder. This might not be the best option, because microCT images typically feature significant artefacts next to the outer edges of the sample being scanned, especially flat edges. In [18] both sample cylinder edge and a flat cut surface are scanned before the microCT, which might produce similar issues. The authors of the [19,20] also perform SEM-EDS on the top part of the cylindrical sample after the microCT measurement. They mention the imperfectness of the result, even after the histogram matching [19]. The authors utilize 2D SURF-based image registration [20] and also mention the imperfectness of the result, even though their sample contains large distinctive features. They conclude [19] that the feature based ML segmentation was unable to improve the quality, as compared to the grayscale-based segmentation. This result is also partially associated with the classes in question — all 3 classes being considered are easily distinguishable by their grayscale level.

In [21] authors use spatially registered QEMSCAN images of internal miniplug cross-sections to construct the ground truth labelling. They mainly focus on a feature-based segmentation of a coarse-grained sandstone via per-grain classification. Indeed, for larger, easily separable (using a watershed transform) and fairly uniform grains, this provides an ability to accumulate a massive amount of statistics that characterizes each grain. The non-zero threshold rotation-invariant Local Binary Patterns (LBP) [22] feature extraction technique was recognized as one of the most robust, allowing to distinguish even such a similar minerals, like quartz and albite. The demonstrated proof-of-concept per-pixel feature-based segmentation result considers 5 classes, and the whole experiment uses a single QEMSCAN image.

3 Imaging

3.1. X-ray microCT imaging

We use Bruker™ SkyScan™ 1172 cone-beam microCT scanner to obtain images of the five similar ⌀8 mm Berea sandstone miniplugs with ~2.2 µm voxel size, using 100 kV tube voltage and the maximum available 10 W tube power. Each sample is also ~8 mm high, which results in ~4000^3 voxel microCT image. It must be noted that the CCD detector being used got 4000×2096 pixels, and the conventional circular trajectory is used. Thus, the sample is scanned as two independent vertical segments, that are later stitched together by the scanner's image reconstruction software. We would later discuss how this affects the segmentation quality.

3.2 Sample slicing

After the acquisition of the microCT images, we use a precision diamond saw to cut each sample into smaller blocks and embed them into epoxy resin. Each block undergoes standard SEM-EDS sample preparation procedures, namely polishing and carbon deposition. The resulting geometry is illustrated in Fig. 1. Each cutting plain is roughly orthogonal to the plug cylinder axis, but up to 5° errors are possible, due to overall mechanical tolerances. We avoid using original outer miniplug edges for the SEM-EDS mapping, not only because these areas are associated with significant artifacts, but also because we would like to use 3D CNN receptive field later.

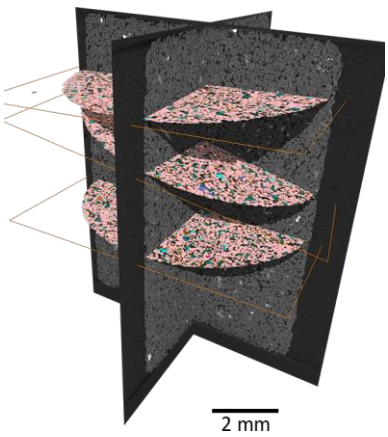


Fig. 1. Positions of the surfaces scanned with the SEM and the miniplug, imaged with microCT (purely schematic).

Furthermore, once the first set of the SEM images is acquired, we re-polish all blocks, to remove few extra microns, and repeat the imaging process. This allows us to obtain numerous slices from a limited physical volume. In this paper, about 9 slices were imaged from each of 5 miniplugs. A total of 45 slices allows to investigate the segmentation quality as a function of the size of the training set, and the effects, associated with non-uniformity of the microCT image quality.

In a daily practice, the "densely packed cross-sections" geometry allows to efficiently utilize the volume of the scanned sample, to save the most of the miniplug intact, e.g., for the subsequent laboratory experiments.

3.3 SEM imaging

We use Thermo Scientific™ QEMSCAN™ 650F SEM to acquire ~4000×4000 mineral maps of each cross-section. The bundled iMeasure® software acquires EDS spectra with about 1000 counts for each pixel of the mineral map and independently classifies each pixel in accordance with the pre-built mineral library [23]. We would refer to this image as QMS-mineralogy. Simultaneously, the BackScattered Electrons (BSE) image with the same resolution is acquired, we would refer to it as QMS-BSE. The software automatically moves the SEM table to sequentially cover the whole ⌀8 mm area of the sample with ~0.3×0.3 mm frames. The example mineral map is shown in Fig. 4a.

We also utilize Thermo Scientific™ Maps™ software, to obtain a higher-resolution BSE (Backscattered Electrons) image of the same surface, we would refer to it as Maps-BSE.

4 Image registration

To utilize classic supervised ML segmentation approaches, we must first spatially register the microCT and the SEM images. This task is complicated by the large size of the data (a single 4000^3 8-bit image takes 64 GB), and the lack of reliable "special" points — the whole image could be viewed as a semi-stochastic texture. Thus, most modern image registration techniques either incur a large computation cost [24], associated with the direct area-based registration, or require a rather fine manually selected initial starting point.

In this work, each mineral map is automatically spatially registered with the corresponding microCT image using the following in-house registration approach, which consists of the three steps.

4.1 Surface-based registration

The aim of the initial step is to roughly locate the global minima, with the precision, comparable with the size of the grain (or pore), as required for the subsequent area-based optimization process to start in the vicinity of the global minima.

The general rigid-body image registration problem could be formulated as an optimization of six translation-rotation parameters, and a scale parameter:

$$T_m = [x_m, y_m, z_m, \varphi_m, \theta_m, \psi_m, s_m].$$

Here, x_m, y_m, z_m are related to image shift along corresponding axes; $\varphi_m, \theta_m, \psi_m$ are x-y-z Euler angles; and s_m is image scaling.

In this notation, we can significantly reduce the volume of the parameter space to look through. For the binarized versions of the images, namely I_{3D} and I_{2D}, we extract the "center of mass" points. That is (x_{2Da}, y_{2Da}) for the 2D image. For 3D image, we use a number of z-slices and the least-squares approach to estimate the cylinder axis position $\left(x_{3Da}(z_{3D}), y_{3Da}(z_{3D})\right)$ for each z_{3D} within the 3D image.

The key point is the extraction of the outer contours of the images, as illustrated in Fig. 2. For 3D image, they could be formalized as

$$R_{2D}(\varphi_{2D}) = max(r){:}(I_{2D}(x_{2D}, y_{2D}) = 1)$$
$$R_{3D}(\varphi_{3D}, z_{3D}) = max(r){:}(I_{3D}(x_{3D}, y_{3D}, z_{3D}) = 1)$$
$$x_{2D} = x_{2Da} + rsin(\varphi_{2D})$$
$$y_{2D} = y_{2Da} + rcos(\varphi_{2D})$$
$$x_{3D} = x_{3Da}(z_{3D}) + rsin(\varphi_{3D})$$
$$y_{3D} = y_{3Da}(z_{3D}) + rcos(\varphi_{3D})$$

Having the surface profiles, we can estimate s_m using the average radius values. The x_m and y_m could be calculated based on the axis position, once the rest of the parameters would be known.

Thus, we would only need to optimize 4 parameters: $[z_m, \varphi_m, \theta_m, \psi_m]$, out of which for φ_m and θ_m we may consider only a quite limited discrete set of their values, because, in accordance with the assumption about the cutting geometry, they should not exceed 5°.

By the simple geometrical means, the $[z_m, \varphi_m, \theta_m, \psi_m]$ parameterize the sine-like R_{2D} curve position on the R_{3D} surface (Fig. 3), and the aim is to find the matching set of parameters, that would define the correct position of the R_{2D} curve (some additional details could be found in [21]).

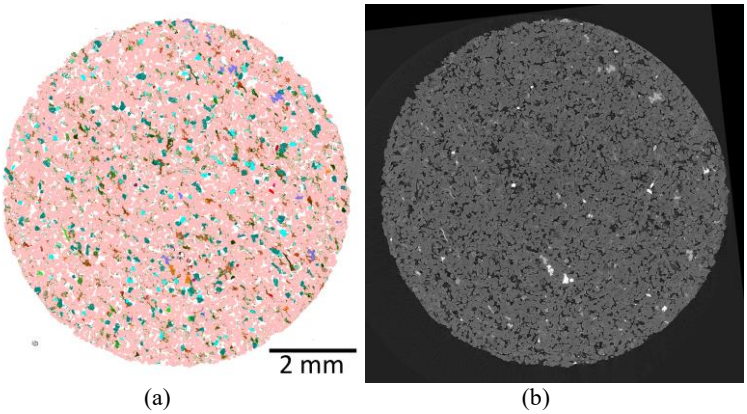
Due to relatively small size of the R_{2D} and R_{3D} arrays, which is unlikely to exceed 10 Mbytes, and a limited number of options to consider, even the exhaustive grid search only takes a few seconds on a modern CPU.

According to our practice, the surface profiles are good unique descriptors, and the rock sample surface is never smooth, which allows to automate the whole image registration process.

Moreover, the method works even if the outer surface of the sample is partially damaged. One could notice a perfectly smooth and round part of the surface in the top-right area of the sample shown in Fig. 4b. This artefact resulted from a sample being slightly outside the cylinder-shaped microCT reconstruction volume. Nevertheless, the surface-based registration works in this case as well.

4.2 Area-based registration

The aim of the second image registration step is to improve the result of the first step, still sticking to the rigid-body assumption. This is implemented as a direct iterative "black-box" optimization of all T_m parameters,



(a) (b)

Fig. 4. Example 3D-to-2D image registration procedure result: (a) de-warped QMS-mineralogy image; (b) spatially corresponding microCT cross-section

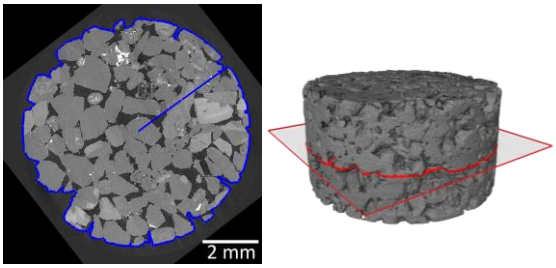


Fig. 2. Surface profiles of a miniplug sample (coarse-grained sandstone is used for readability).

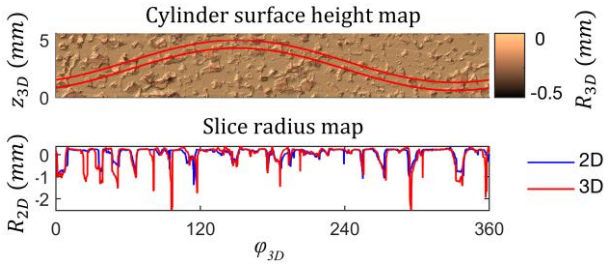


Fig. 3. Matching 2D and 3D surface profiles of a miniplug

using area-based registration norm. To compare discrete-index mineral map with continuous grayscale microCT image, we use mosaic-image Put'ev matching score [25] (or similarly, the explained variance score, calculated after assigning each mineral the grayscale value equal to the average of all the microCT image pixels, matching to that mineral, according to current T_m).

4.3 Smooth non-rigid distortions compensation

Although the miniplug usually represents a fairly good example of a rigid body, the rigid body approximation is insufficient for pixel-perfect SEM-microCT image registration. The issue arises mainly due to acquisition hardware imperfectnesses — even tiny sub-pixel SEM lens distortions tend to accumulate to several pixels–large smooth distortions, when multiple SEM images are stitched together.

One possible workaround is to calibrate the SEM distortions using a special regular sample, and later apply the corresponding de-warping to all SEM images [7]. One of the drawbacks of this method is that it assumes that the distortions would not change after the calibration. Another drawback is that this method is unable to compensate the distortions, that arise from a limited SEM stage movement precision and the SEM image stitching algorithm glitches.

Instead, we directly apply an Optical Flow estimation method [26] to the pair of images in question. We use smooth large-scale components of the resulting vector field to de-warp the SEM image. This ensures that small artefacts, like mechanically damaged grains would not result in major image distortions. The transformation is so smooth, that for each small sub-area, containing a few grains, it could be viewed as a plain translation.

It must be noted that this de-warping step only considers the 2D distortions in the plane of the SEM image. Thus, it is unable to fully replace

the area-based registration step, which is fully 3D, even if the surface-based registration step provides a very good initial approximation.

The ultimate result of the image registration procedure is shown in Fig. 4. Finally, to allow calculating features on the original microCT data, we scale the SEM image to match the original microCT voxel size. Also, for ML needs, we not only export a single matching microCT cross-section (Fig. 4b) but a 401-slices thick 3D block (e.g. ± 200 slices around the matching cross-section).

5. Relabelling

In [27] it was demonstrated that the QEMSCAN classification result could be misleading sometimes and not agree with the optical petrography results — the issues generally arise for minerals that are very close in their elemental concentration proportions, but considered very different from a petrography point of view. One simple example is the quartzite and quartz pair. These two minerals are usually easily visually distinguishable on a BSE image, but might be indistinguishable from the QEMSCAN's "independent pixels" point of view. This particular pair could be separated via a joint QMS-mineralogy and QMS-BSE images post-processing, honouring some minimal neighbourhood of the pixel in question. In this work, we do not question the accuracy of the QEMSCAN image. However, the original QEMSCAN image is still poorly suitable for direct use as a ground truth labelling, mainly due to the noise-like patterns, shown in Fig. 5.

Firstly, we note a number or single-pixel grains, that are definitely too small to be recognizable on a microCT image (Fig. 5a) — even though QEMSCAN pixel size (2 μm) is very close to microCT voxel size (~2.3 μm), the physical resolution of the SEM is significantly higher. Moreover, such objects would hardly influence any physical properties of the Digital Rock model anyway. However, the underlying grain still might be real, and might influence the microCT image somehow. Thus, the general principle we use is not to "smooth-out" such objects with a median filter or equivalent, but to re-assign them to the "Unknowns" class, which is later ignored by all the ML classifiers we use. The mask, highlighting such objects, is shown in the rightmost column in Fig. 5a.

Areas featuring high concentrations of such single-pixel grains are trickier. We found that removing relatively large areas around them stabilizes many classifiers, most probably because such objects are usually outliers in the ML sense. One might note that the

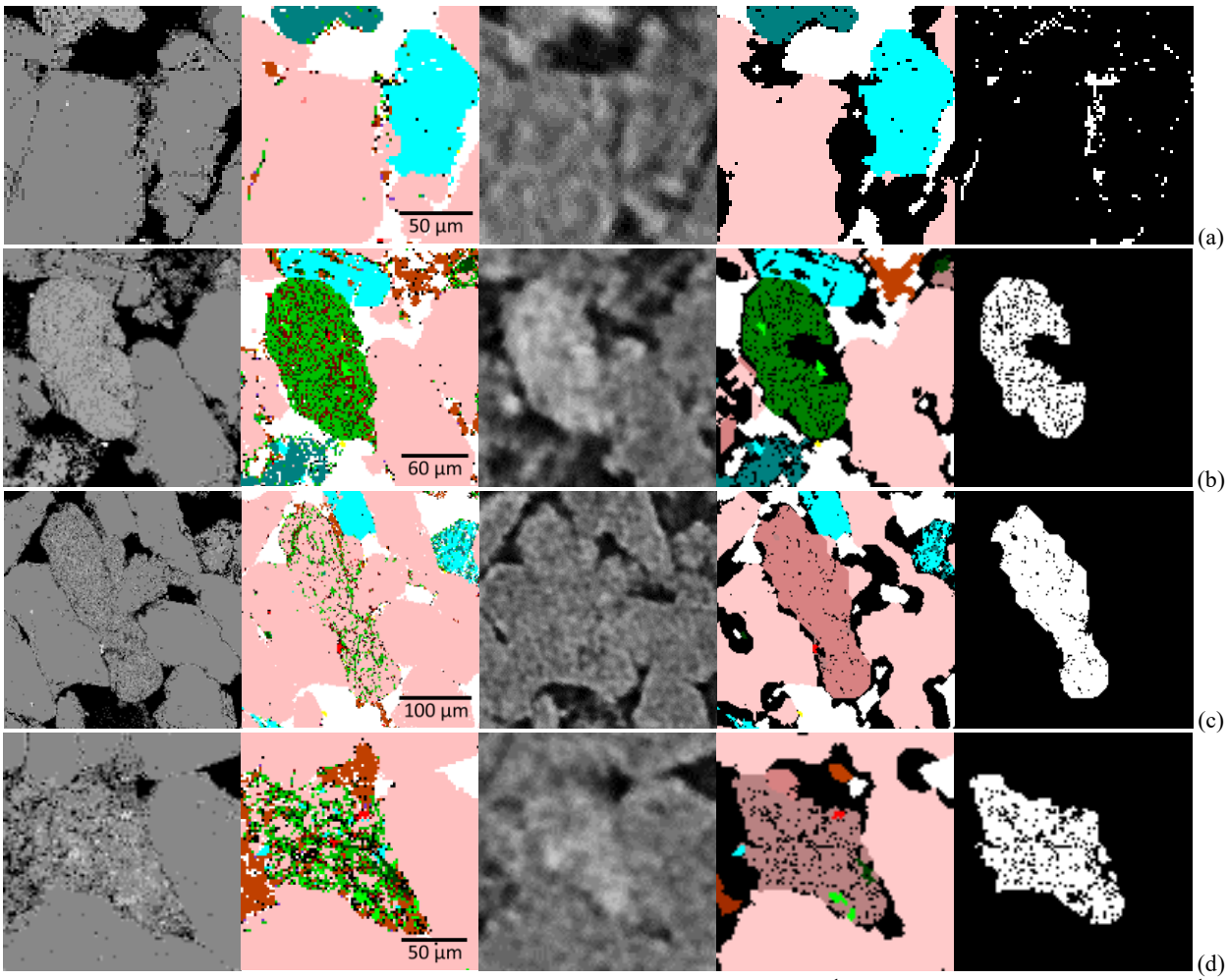


Fig. 5. QEMSCAN mineral map relabeling examples: 1st column - QMS-BSE image; 2nd column - QMS-mineralogy image; 3rd column – spatially registered microCT image; 4th column – relabeling result; 5th column – the relabeled pixels in question (see text); (a) small objects; (b) Clay_mixture class; (c) Quartz_with_clay class; (d) Quartz_clay_cement class.

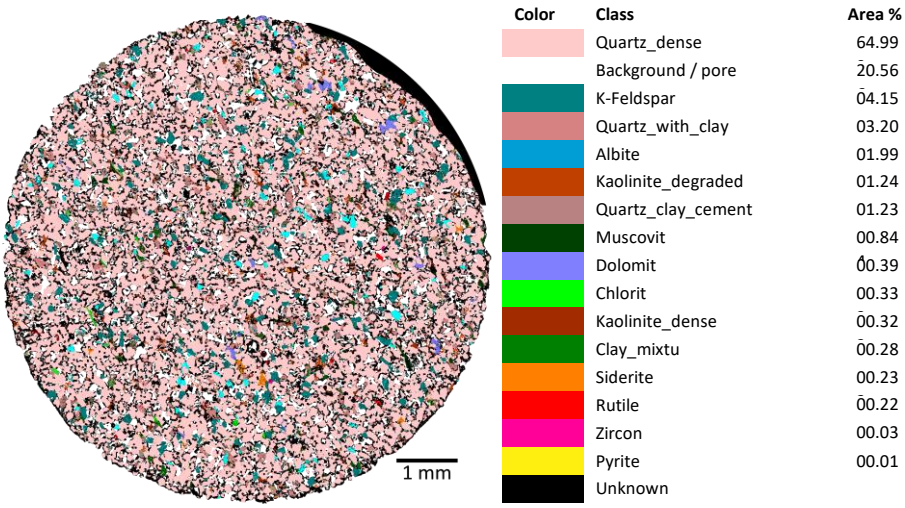


Fig. 6. The ground truth image of the validation slice — the result of the relabeling procedure for the image, shown on Fig. 4a. The legend result and the area percentage for all the classes on the relabeled image.

relabeled image in Fig. 5a contains more Unknowns than just the "small objects" mask, discussed in the previous paragraph.

More importantly, there is a multitude of clay-related classes. The Fig. 5b-d demonstrate that some grains look fairly uniform on the scale of tens of pixels (especially on the BSE image) but produce single-pixel noise-like patterns on both mineral map and BSE.

We relabel such objects, aiming to also distinguish the most important types of conglomerates. Most of them could be easily categorized into broad classes, each containing at least tens of such objects per a single QEMSCAN image. Namely, the grain shown in Fig. 5b is assigned to the Clay_mixture class. The grain, shown in Fig. 5c is assigned to the Quartz_with_clay class. The grain, shown in Fig. 5d is assigned to the Quartz_clay_cement class. Such relabelling not only makes sense from the further Digital Rock physical properties modelling point of view (and could be viewed as an upscaling step), but also greatly simplifies the segmentation task — now pixelwise classification losses begin to make sense.

All the said relabelling was implemented using basic morphological image processing operations, and carefully hand-tuned for the specific dataset in question (but not for individual slices). Theoretically, such a "clustering" operation could be automated to avoid operator involvement, but that would require considerable additional research.

Finally, slices are inspected manually, and the broad Unknown class strokes (black) are placed where major artefacts are noticed. An example of such a manual intervention is clearly visible in Fig. 6 on the top-right part of the image. The reason for this is the same artefact (resulting from a sample being slightly outside the cylinder-shaped microCT reconstruction volume) that was already discussed above. We also automatically detect the sample circle radius and fill the out-of-sample-circle area with Background/pore class colour (white). We ignore this area in all further area-percentage calculations, including segmentation accuracy calculations.

6. Feature-based segmentation

For the feature-based segmentation results, demonstrated below, we build on the results, demonstrated in [21,28]. We only use 2 QEMSCAN slices for training and a single QEMSCAN slice (shown in Fig. 6) for validation — feature-based approaches require less data before the result quality stops to increase, compared to the CNN-

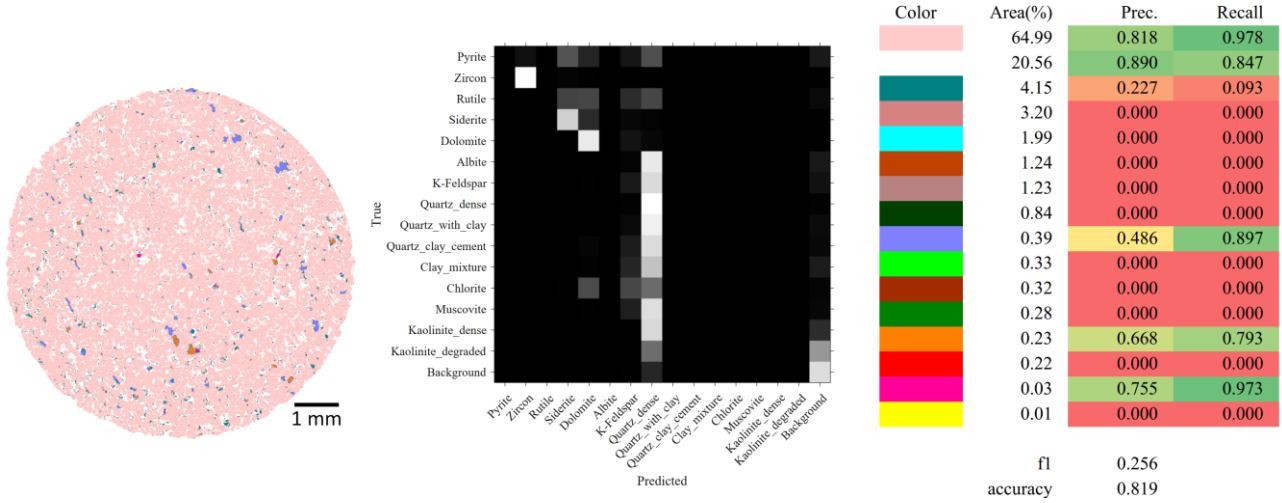


Fig. 7. The result of the "Naïve Bayes" segmentation approach: validation slice segmentation (see legend in Fig. 6), confusion matrix (row-normalized), Precision and Recall (green-yellow-red palette highlights good-average-poor values).

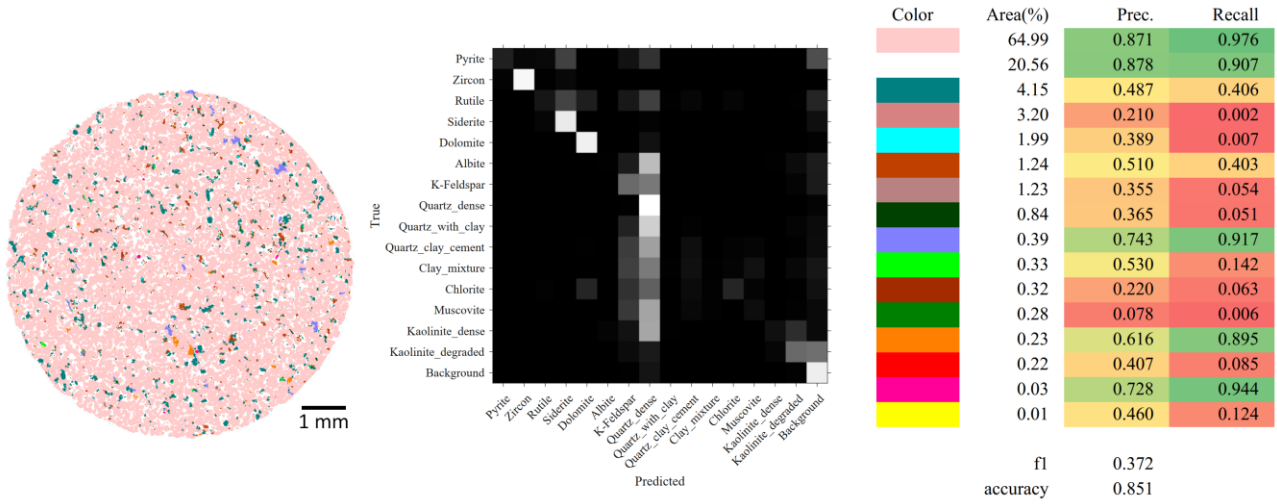


Fig. 8. The result of the "LightGBM" segmentation approach. Designations are equal to Fig. 7.

based. Depending on the circumstances, this could be either a strength or a weakness.

6.2 Naïve Bayes

Our baseline "Naïve Bayes" approach actually utilizes only a single feature channel – the original microCT image, filtered with the manually tuned bilateral filter to suppress noise. We use parametric Naïve Bayes classifier assuming normal distribution. This is roughly equivalent to manually choosing a set of a threshold values after image pre-filtering. The result, shown in Fig. 7 highlights a set of substances that are easily distinguishable by their grayscale values with green background in Precision and Recall columns (definitions of the "Precision", "Recall", "accuracy" and "F_1" terms could be found in [29]). A visual comparison with Fig. 4b confirms this assumption.

Still, Naïve Bayes looks significantly better than the binary Otsu segmentation approach (Fig. 15, Fig. 16) which only achieves 0.813 accuracy and 0.109 F_1, (we use micro-averaging) assuming that it distinguishes Quartz_dense and Background classes.

6.3 LightGBM

LightGBM [30] is a modern gradient boosting framework, providing both high performance and state-of-the-art classification accuracy in many benchmarks. We combine it with two 3D 8-bit cube-shaped rotation-invariant LBP patterns and a local window statistics feature vector (e.g. average, median, variance, skew, kurtosis, …). We fine-tune LBP meta-parameters using cross validation (namely histogram windows size, pattern scale and threshold value). Also, we post-process the soft-classification result, provided by the LGBM using the Graph Cut approach [31] to make it a bit smoother. The result demonstrated in Fig. 8 achieves higher scores compared to Fig. 7.

7. CNN-based segmentation

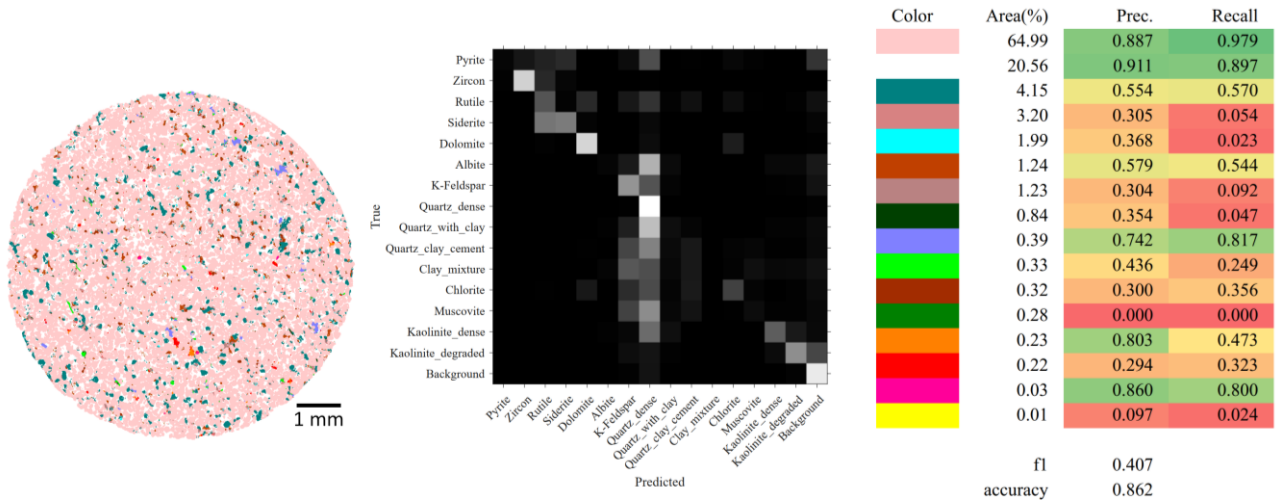


Fig. 9. The result of the "3-slice U-net" segmentation approach. Designations are equal to Fig. 7.

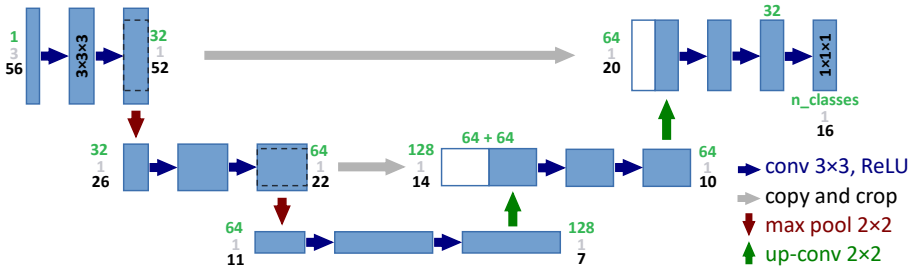


Fig. 10. The 3-slice U-net architecture. The number of feature channels is indicated in green; **z**-size is indicated in gray; **x** and **y** sizes are equal and are indicated in black.

The CNN-based segmentation results, demonstrated below, utilize the U-net architecture [32]. For the experiments in this section, we use the same single QEMSCAN slice for validation, and 44 slices for training. We use 3D receptive field, valid-mode convolutions and dense 2D output labelling (except that some pixels belong to the Unknown class). We argue that such a geometry is more efficient than sparse 3D labelling [33]. The training epoch corresponds to a number of randomly selected patches with a total area equal to the total area of the training set. We do not perform training data rebalancing or hard negative mining, despite severe imbalance. We use cross-entropy loss and Adam optimizer with learning rate 0.0002, β1=0.5, β2=0.99. Our experience suggests that further fine-tuning only provides specific minor improvements. Although they might be valuable in specific cases, they are out of the scope of the present article, aiming at a more general comparison.

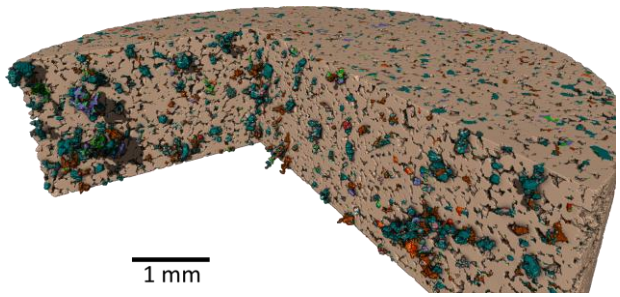


Fig. 11. The 3D render of the 3-slice U-net segmentation result

7.1 3-slice U-net

Our first, most lightweight U-net model uses 3-slice [34] architecture (Fig. 10), that is — only the first single convolution is 3×3×3, while the rest are 3×3, thus being almost 2D. Some additional details are available in [28]. The training performance is about 100 epochs per day, on a single GK210 chip of the Nvidia® K80 GPU. The Fig. 9 demonstrates the 1000-epoch training result, but scores only marginally improve after 200 epochs. On the large scale, this result is visually similar to the LightGBM result (Fig. 8), but the scores are significantly higher. Some differences could be seen in the Fig. 15 – the CNN generally produces less small infeasible details. The 3D rendering (Fig. 11) demonstrates adequate 3D geometry of the grains despite the "almost 2D" nature.

7.2 3D U-net

Our second U-net model (Fig. 13) is fully 3D and all 3 spatial dimensions are (almost) equal. It uses 204×204×92 input patches and 112×112 output patches. Specific patch sizes are used to ensure full symmetry in max pooling and up-convolution layers. Additional 3×3×1 convolution layer is added for the same purpose. Each convolution, except the last one, is followed by PReLU activation with 0.25 activation parameter. The last convolution uses softmax activation. Up-convolution layers use transposed convolutions and halve the number of the feature channels. The training performance is, again, about 100 epochs per day, but now a GPU sever with 7×RTX2080Ti is used. The training batch size is 7 (a single patch per a GPU) due to GPU memory limitations. After the first

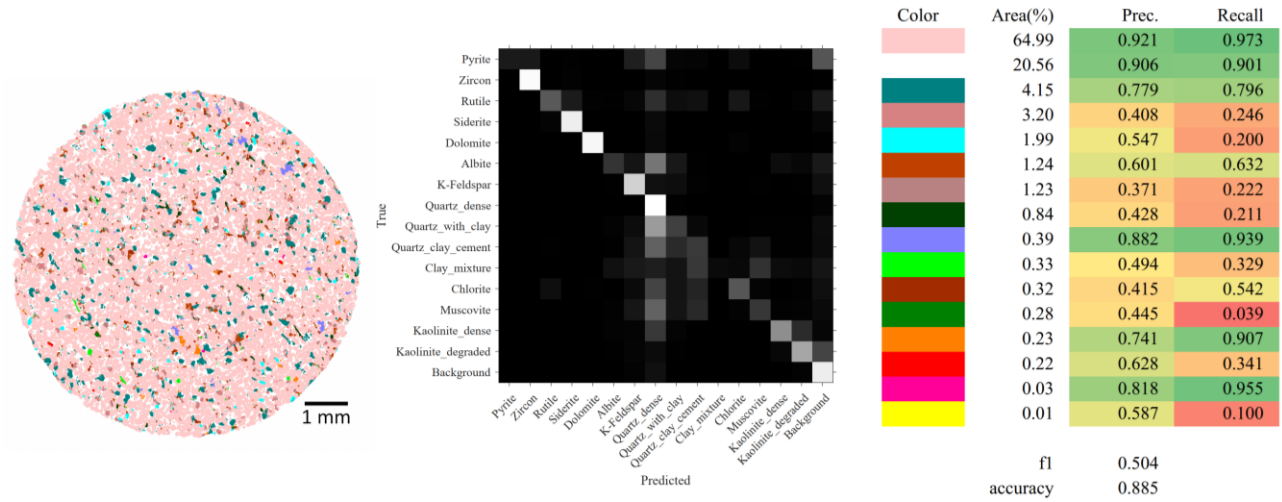


Fig. 12. The result of the "3D U-net" segmentation approach. Designations are equal to Fig. 7.

100 epochs, accuracy score, calculated on our only validation slice, oscillates in [0.088, 0.0885] range.

For the Fig. 12, we select a result with the maximum accuracy observed. This result demonstrates a significant step forward, as compared to our previous results — the confusion matrix now looks much more like a clean diagonal. One exception is the most-common Quartz_dense class that is thus considered "the safest choice" when the U-net is not certain about a specific pixel. The Precision score is above 0.37 for all classes, but the Recall is still below 0.25 for the most difficult-to-distinguish classes, including clays. However, the confusion matrix suggests that in many cases clays are mostly mixed with other clays, which might be acceptable for many practical scenarios. Classes like Quartz_with_clay and Quartz_clay_cement are similar in their nature — there is no clear boundary between them.

The fact that Pyrite is commonly mixed with Quartz and Background could be attributed to the small size of the Pyrite grains and mechanical damage of the sample surface during polishing.

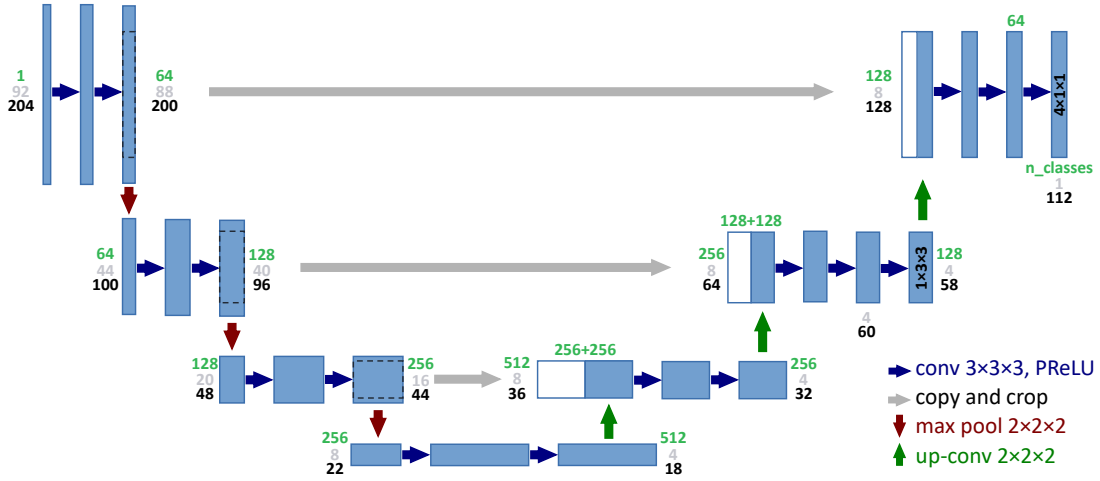


Fig. 13. The 3D U-net architecture. Designations are equal to Fig. 10.

7.2 3D U-net+GAN

The Generative Adversarial Networks (GAN) are algorithmic architectures that use two neural networks, competing against each other, e.g., while the first one attempts to generate realistic synthetic images, the other one attempts to distinguish between the synthetic and real images. This concept, originally proposed in [35], proved to be very successful for continuous data generation tasks, like grayscale or RGB image generation, including image-to-image translation [36]. But only a limited success should be expected, when directly applying GANs to discrete data generation tasks, like the image segmentation task in question. The issue arises from a limited amount of information, available to the generator network through the argmax layer.

Last but not least, almost all ground truth patches contain at least one Unknown (black) pixel (see Fig. 16b), while our segmentation methods are designed to produce none (see Fig. 16e-k). If we blindly pass such "real" and "fake" patches to the discriminator, it would be able to distinct between them by just that. Thus, it would not provide any meaningful information about the "fake" patch quality to the generator. To fix this, we make generated patches look like real in terms of the Unknown class distribution, by copying the Unknown pixels from the ground truth image.

Our third U-net model utilizes effectively the same architecture as the previous one, but features an additional 14-layer discriminator network, attached to its output.

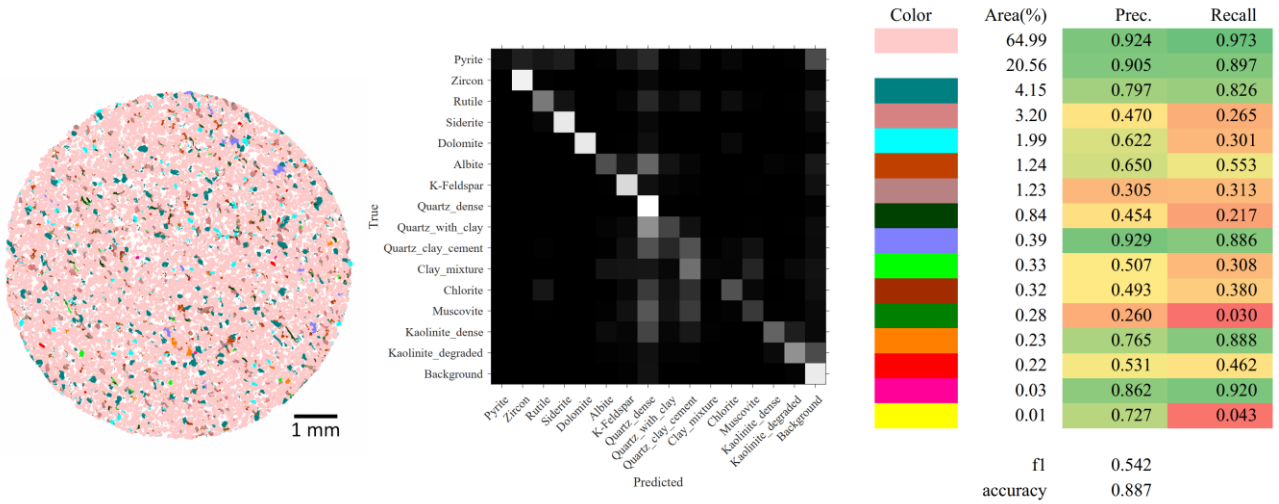


Fig. 14. The result of the "3D U-net+GAN" segmentation approach. Designations are equal to Fig. 7.

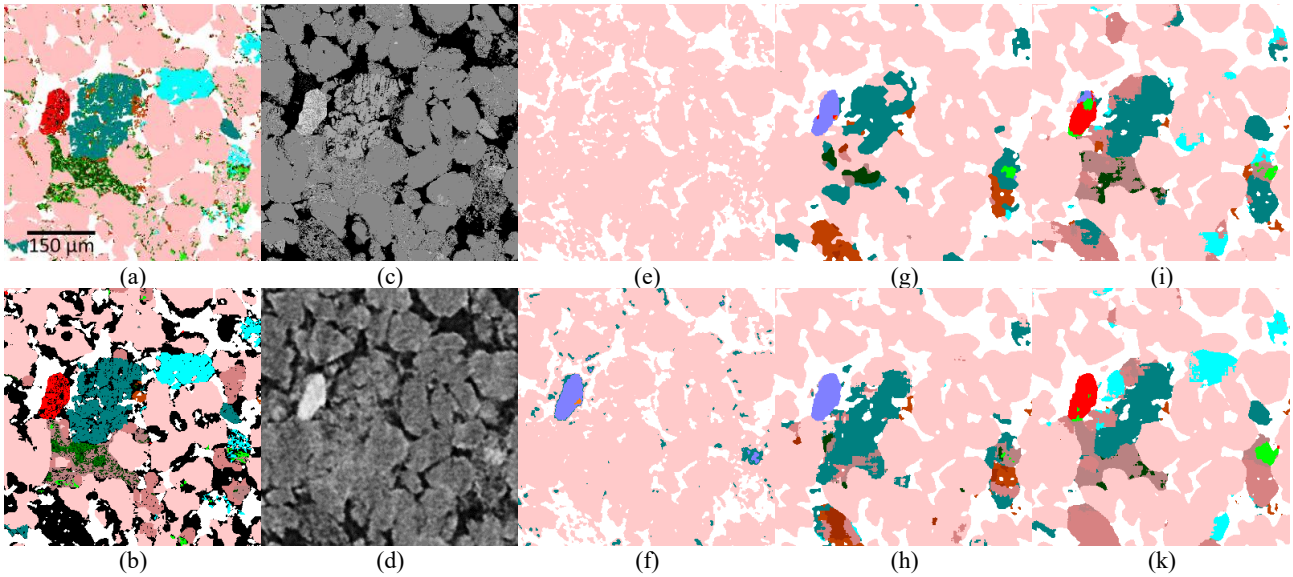


Fig. 15. Comparison of the different segmentation results for the fragment one: (a) original QMS-mineralogy; (b) Relabeling result (segmentation ground truth); (c) original QMS-BSE; (d) microCT; (e) Otsu; (f) Naïve Bayes; (g) LightGBM; (h) 3-slice U-net; (i) 3D Unet; (k) 3D Unet+GAN.

Only 2D 3×3 convolutions are used, because we lack 3D ground truth data. For the 4th, 7th and 10th convolutions, stride equals 2, for the rest stride equals 1. The number of feature channels first doubles after each filter, starting from 32 and ending up at 512, and then halves back down to 32. Each convolution is followed by a LeakyReLU activation with 0.2 parameter and instance normalization. The last convolution is rather followed with a sigmoid activation. During the training phase, generator and discriminator are updated with 1:1 ratio, using minimax GAN loss. Dropout and hard-negative mining are not used. The training performance drops to about 60 epochs per day.

Once again, in the Fig. 14 we just present the best result achieved. However, we must admit that in this case, the training performance is less stable, and results vary way more across different epochs, which is a typical for the architecture used. One promising point, however, is that this result is significantly better than those on Fig. 12 in terms of the F_1 score. Another promising point is that careful inspection of the smaller crops (Fig. 15) indicate some progress towards smoother, more realistic grains, with less unfeasible mineral mixtures. Even though this "smoothing" hardly ever provides a truly realistic result and sometimes results in an even more incorrect classification, like for a K-Feldspar grain in Fig. 16.

8. Image quality uniformness

To assess the microCT image quality uniformness, and the possible associated effects, we train 3-slice U-net on various sets of slices, and simultaneously evaluate it on all available slices (Fig. 17). One simple effect to mention is that training sets below 10 slices seem to be insufficient in terms of the F_1 score.

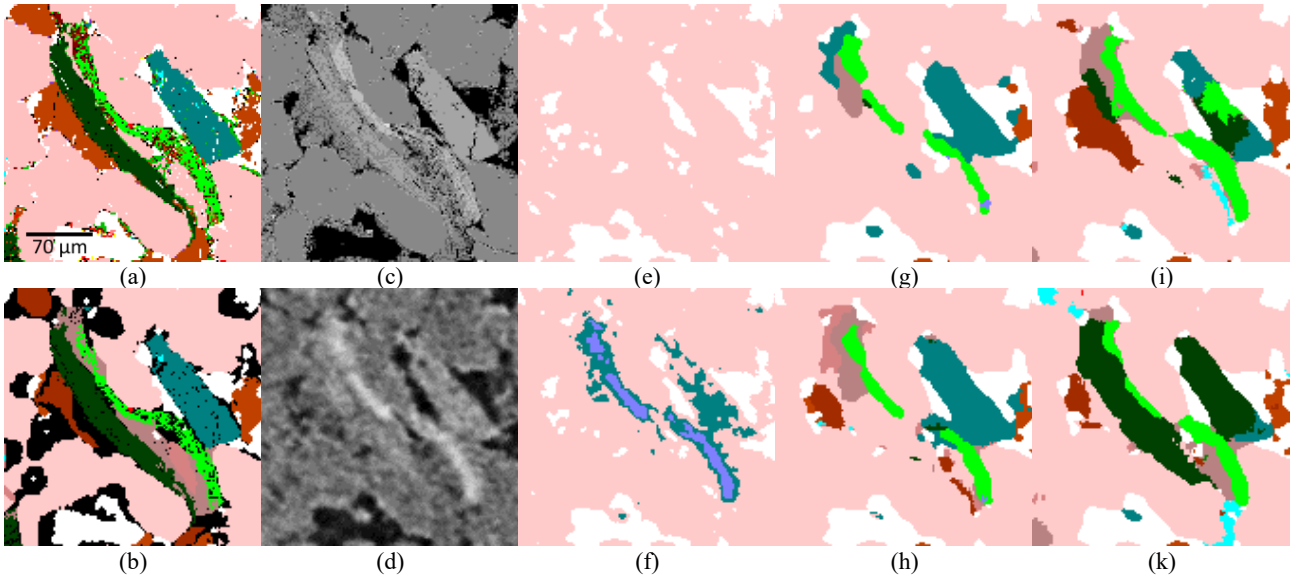


Fig. 16. Comparison of the different segmentation results for the fragment two: (a) original QMS-mineralogy; (b) Relabeling result (segmentation ground truth); (c) original QMS-BSE; (d) microCT; (e) Otsu; (f) Naïve Bayes; (g) LightGBM; (h) 3-slice U-net; (i) 3D Unet; (k) 3D Unet+GAN.

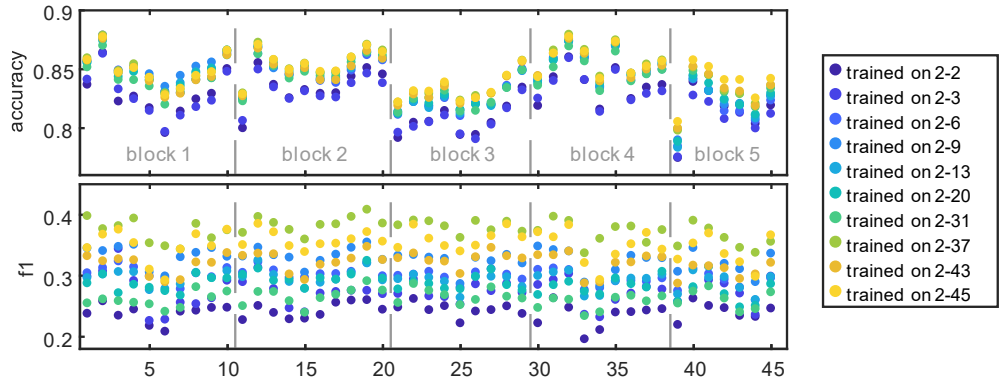


Fig. 17. The effect of the training set selection. Each block corresponds to a physical miniplug.

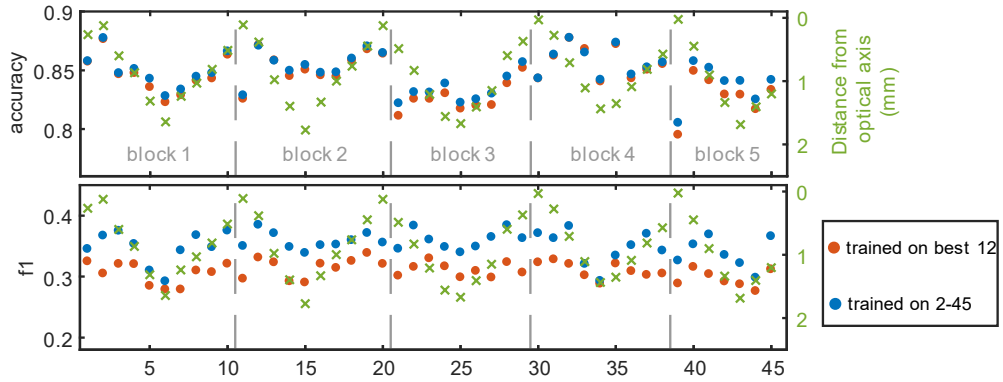


Fig. 18. The "distance from the optical axis" effect.

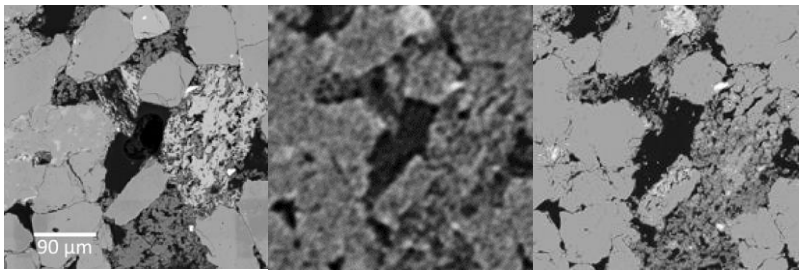
The second, more interesting effect we notice in Fig. 17 is that the segmentation quality is not growing monotonously when adding training slices. In particular, it significantly drops (in terms of the F_1 score) for most slices, when we add slices 38-43 to the training set. This might suggest that there are some "bad" slices, negatively affecting the segmentation quality. To assess this hypothesis, we train the same classifier using 12 best-accuracy slices, however, the segmentation quality decreases on all slices (Fig. 18). Thus, one might conclude that there are no "bad" slices, just an intrinsic non-uniformness, which is partially associated with microCT image quality non-uniformness.

We found a strong correlation between the distance from a slice to the microCT optical axis (central slice) and the segmentation quality for that slice (the fact that the first slice from each block demonstrates lower segmentation quality than the second one should be probably attributed to the sample preparation issues). Gradual minor quality drop in the areas further from the optical axis is the expected effect for conventional cone-beam circular-trajectory microCT scanners, utilizing Filtered Back Projection (FBP) reconstruction algorithm. The multiclass segmentation relies on barely visible tiny features and thus it is highly sensitive to such effects. However, this is certainly not the only source of imperfectnesses. We assume that the only feasible mitigation measure is to train models on a larger datasets, covering more-or-less all possible imperfectnesses. Later, such model could be fine-tuned for a specific smaller practical job.

9. Learned super-resolution

The ability of the CNN to resolve meaningful details in the microCT image could be also studied by using effectively the same CNN architecture and input data for a learned super-resolution task. For this purpose, we use a single full-slice Maps-BSE image for training, and another Maps-BSE image, from a different slice, for the visual comparison (Fig. 19). The result demonstrates that although our model is unable to reconstruct the Maps-BSE image exactly, and locations of specific tiny features may not match real in a pixel-perfect manner, on a larger-scale, the overall morphology and substance types are mostly reproduced correctly (which is actually more important than pixel-perfect positions for the subsequent physical properties modeling). Moreover, infeasible patches are uncommon. This is partially associated with the fact that continuous data generation task is more suitable for the modern GANs. This looks promising in a sense that there is still potential for segmentation quality improvement.

It should be noted that such a super-resolution tool is also valuable by itself. In [37] authors discuss microCT-to-microCT super-resolution and provide the visual comparison of the result with an SEM image. This super-resolution task is also close to the conditional image generation [38], which could be viewed as effectively the same task with a weaker preconditioning.



(a) (b) (c)

Fig. 19. Super-resolution experiment image fragments comparison:
(a) original Maps-BSE image; (b) microCT image; (c) super-resolution result

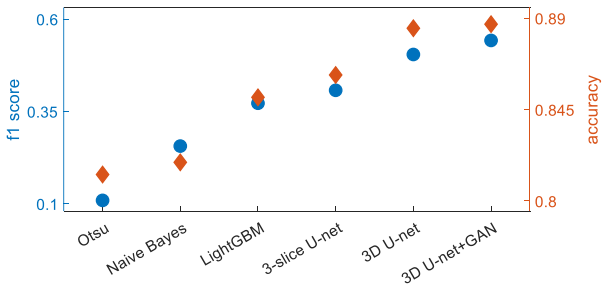


Fig. 20. The comparison of the segmentation quality scores.

10. Conclusions

Convolutional neural networks allow multiclass microCT image segmentation with a quality far beyond more traditional methods (Fig. 20). Larger 3D receptive fields are beneficial for the segmentation quality, but the associated computation cost is higher. Basic supervised training could be considered fairly stable and thus production ready. However, a tightly spatially matching sizable operator-independent ground truth labelling is required. The 3D GAN-based approaches have not demonstrated their full potential yet and are currently less stable in general. However, they already provide slight improvements and represent a promising subject for further research.

References

1. W. Oh et al., IEEE Transactions on Pattern Analysis and Machine Intelligence **21**, 590–602 (1999)
2. A. P. Sheppard et al., Physica A: Statistical Mechanics and Its Applications **339**, 145–151 (2004)
3. D. A. Koroteev et al., in *Proc. of International Symposium of the Society of Core Analysts* (California, USA, 2013), 12
4. O. Dinariev et al., in *Proc. of International Symposium of the Society of Core Analysts* (Pau, France, 2019)
5. J. Goral et al., Sci. Rep. **10**, (2020)
6. M. Knackstedt et al., in *Soc. Pet. Eng. - SPE Asia Pac. Oil Gas Conf. Exhib., APOGCE* (2010), 1589–1597
7. T. K. Varslot et al., US20110181701 A1 (2011)
8. A. C. Jones et al., Biomaterials **28**, 2491–2504 (2007)
9. Y. Wang et al., Minerals Engineering **83**, 185–191 (2015)
10. I. Arganda-Carreras et al., Bioinformatics **33**, 2424–2426 (2017)
11. C. Sommer et al., in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (2011), 230–233
12. M. Andrew, Computational Geosciences **22**, 1503–1512 (2018)
13. S. Karimpouli et al., Computers & Geosciences **126**, 142–150 (2019)
14. V. Badrinarayanan et al., IEEE Transactions on Pattern Analysis and Machine Intelligence **39**, 2481–2495 (2017)
15. Y. Da Wang et al., ArXiv:2002.05322 (2020)
16. I. A. Varfolomeev et al., Computers **8**, 21 (2019)
17. M. Andrew et al., Microscopy and Microanalysis **23**, 156–157 (2017)
18. M. R. Ball et al., Microscopy and Microanalysis **25**, 410–411 (2019)
19. G. Tiu, Classification of Drill Core Textures for Process Simulation in Geometallurgy: Aitik Mine, Sweden, Master's thesis in Natural Resources Engineering, Luleå University of Technology, (2017)
20. P. I. Guntoro et al., Minerals Eng **142**, 19 (2019)
21. I. A. Varfolomeev et al., in *Proceedings of the 3rd IAPR Asian Conference on Pattern Recognition – ACPR* (Kuala Lumpur, Malaysia, 2015), 346–350
22. M. Pietikäinen et al., in *Computer Vision Using Local Binary Patterns* (Springer, London, 2011), 13–47
23. A. R. Butcher et al., in *Seventh Mill Operators' Conference* (The Australasian Inst. of Mining & Metallurgy, 2000), 267–271
24. S. Latham et al., in *Proc. of International Symposium of the Society of Core Analysts* (Society of Core Analysts, Abu Dhabi, UAE, 2008)
25. Yu. P. Pyt'ev et al., *Methods of Morphological Analysis of Images [In Russian]* (Fizmatlit, Moscow, 2010)
26. C. Liu, Beyond Pixels: Exploring New Representations and Applications for Motion Analysis, PhD thesis in Electrical Engineering and Computer Science, Massachusetts Institute of Technology, (2009)
27. I. A. Varfolomeev et al., in *Proceedings of the SPE Russian Petroleum Technology Conference and Exhibition* (Society of Petroleum Engineers, 2016)
28. I. A. Varfolomeev et al., VKIT [In Russian] (**7**), 3–9 (2019)
29. D. M. W. Powers, ArXiv:2010.16061 (2010)
30. G. Ke et al., in *Advances in Neural Information Processing Systems* (2017), 3149–3157
31. Y. Boykov et al., Pattern Analysis and Machine Intelligence, IEEE Transactions On **23**, 1222–1239 (2001)
32. O. Ronneberger et al., in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Springer, Cham, 2015), 234–241
33. Ö. Çiçek et al., ArXiv:1606.06650 (2016)
34. M. Lai, ArXiv:1505.02000 (2015)
35. I. J. Goodfellow et al., ArXiv:1406.2661 (2014)
36. P. Isola et al., in *IEEE Conference on Computer Vision and Pattern Recognition – CVPR* (2017), 5967–5976
37. Y. D. Wang et al., Water Resources Research **56**, e2019WR026052 (2020)
38. L. Mosser et al., in *Proc. of 80th EAGE Conference and Exhibition* (EAGE, 2018), 1–5