Enhancement of Scanco micro-CT images of granodiorite rocks using a 3D convolutional neural network super-resolution algorithm

Alexandra Roslin[1,1], Maxim Lebedev[2,2], Travis Ryan Mitchell[3,3], Italo Andres Onederra[1,1], and Christopher Ross Leonardi[3,3]

[1]University of Queensland
[2]Curtin University
[3]The University of Queensland

November 30, 2022

Abstract

X-Ray micro-computed tomography (micro-CT) is a standard method to perform three-dimensional analysis of the internal structure of a rock sample. 3D X-Ray microscopes, such as those from the XRadia Versa family, provide images of high resolution and contrast. Medical scanning machines can also be used for scanning rock samples to reduce operational cost and time, but they generally provide poorer spatial resolution and contrast compared to 3D X-Ray microscopes. Recent success in implementing deep learning algorithms to enhance image quality demonstrated that, in some cases, the application of convolutional neural network (CNN) models might significantly enhance the resolution of the micro-CT images. In this research, a super-resolution technique employing the U-Net 3D CNN architecture is applied to enhance the resolution of granodiorite rock sample images obtained by two different 3D scanning machines. The high-resolution dataset was obtained using the XRadia Versa XRM-500 microscope. It contained images with nominal resolutions of 10.3 and 5 microns. The low-resolution scanning was performed using a Scanco medical μCT 50 machine, and the images from this dataset had a nominal resolution of 10.3 microns. Several models were created to enhance the quality of the low-resolution images, and the results were analysed. It was observed that super-resolution processing could significantly improve the low-resolution micro-CT image quality and suppress noise that appeared on medical images. The results presented in this study are of particular interest and value to geoscientists that use medical scanners to study the structure of rock samples at large scale.

Enhancement of Scanco micro-CT images of granodiorite rocks using a 3D convolutional neural network super-resolution algorithm

A. Roslin[a,*], M. Lebedev[b], T.R. Mitchell[a], I.A. Onederra[a], C.R. Leonardi[a]

[a]*School of Mechanical and Mining Engineering, The University of Queensland, St Lucia, Australia*
[b]*WA School of Mines: Minerals, Energy and Chemical Engineering, Curtin University, Perth Campus, Australia*

Abstract

X-Ray micro-computed tomography (micro-CT) is a standard method to perform three-dimensional analysis of the internal structure of a rock sample. 3D X-Ray microscopes, such as those from the XRadia Versa family, provide images of high resolution and contrast. Medical scanning machines can also be used for scanning rock samples to reduce operational cost and time, but they generally provide poorer spatial resolution and contrast compared to 3D X-Ray microscopes. Recent success in implementing deep learning algorithms to enhance image quality demonstrated that, in some cases, the application of convolutional neural network (CNN) models might significantly enhance the resolution of the micro-CT images. In this research, a super-resolution technique employing the U-Net 3D CNN architecture is applied to enhance the resolution of granodiorite rock sample images obtained by two different 3D scanning machines. The high-resolution dataset was obtained using the XRadia Versa XRM-500 microscope. It contained images with nominal resolutions of 10.3 and 5 microns. The low-resolution scanning was performed using a Scanco medical μCT 50 machine, and the images from this dataset had a nominal resolution of 10.3 microns. Several models were created to enhance the quality of the low-resolution images, and the results were analysed. It was observed that super-resolution processing could significantly improve the low-resolution micro-CT image quality and suppress noise that appeared on medical images. The results presented in this study are of particular interest and value to geoscientists that use medical scanners to study the structure of rock samples at large scale.

Keywords: Convolutional Neural Network, Micro-CT, Super-Resolution, Igneous Rocks, Deep Learning, U-Net 3D, Scanco, XRadia

Plain language summary

In the field of geology, it is common to examine the structure of rocks through X-Ray imaging. It provides a non-destructive method to estimate the mineral content and other important properties of a rock when considering the extraction of natural resources. The issue of cost, time, and accessibility to quality X-Ray scanning equipment often leads geologist to perform this imaging on medical equipment, rather than that designed for the analysis of earth materials. As a result, the obtained scans can be poor quality leading to large uncertainties in the properties of interest. In this work, we used deep learning techniques to improve images obtained from the more accessible, medical X-Ray microscopes such that they were comparable to those obtained from a purpose built X-Ray machine. This was done by providing a neural network with medical X-Ray images, and training it to improve them to high-resolution images of the same material. Following this, unseen images were provided

*Corresponding author.
Email address: a.roslin@uq.edu.au (A. Roslin)

to the network and the improved output images were observed to accurately correspond to those obtained with a microscope designed for geomaterials.

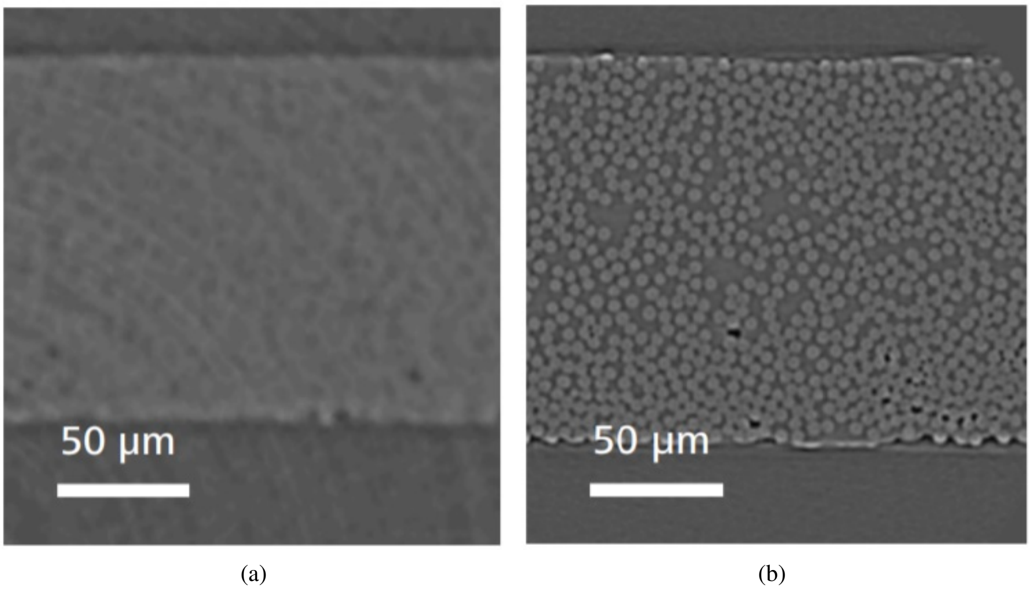
1. Introduction

X-Ray micro-computed tomography (micro-CT) is an imaging method that produces three-dimensional representations of the internal microstructure of materials without destroying the analysed samples [1, 2]. In order to obtain fine feature details and a representative image of the sample interior, high-resolution micro-CT images are required. The term "resolution" may seem intuitively clear, but various terminology is used in the literature to describe the resolution capacity of micro-CT instruments. Spatial resolution is the most comprehensive metric describing a micro-CT instrument. It measures the output of the system and accounts for multiple scanning characteristics, including X-ray source spot size, detector resolution, vibrational, electrical and thermal stability, magnification geometry, and imaging conditions [3]. Contrary to this, nominal resolution is a theoretical parameter that does not provide evidence of the true performance of a system. It is a resolution in ideal conditions. While the spatial resolution reflects the system performance, the nominal resolution is related to the system design. Practically, the term "nominal resolution" is often used to refer to the minimum achievable voxel size. The voxel size is a geometric calculation referring to a cross-sectional area in the sample that is imaged onto a single detector pixel. The voxel size accounts for only the detector pixel size and system geometry but does not consider the imaging conditions. Researchers that work with micro-CT images are often referencing the voxel size when talking about resolution. In this work, the authors are generally referring to the voxel size when mentioning the resolution of the micro-CT images and assume that the nominal resolution of the images is equal to their voxel size.

It has already been demonstrated that the same voxel size of the images produced by different CT instruments does not guarantee similar image quality. Figure 1 compares two images of a carbon fibre composite material. Both images were taken with the same voxel size (1 μm), but using different micro-CT systems (similar to the instruments used in the current research). However, only the image on the right can adequately resolve the carbon fibres. Another problem related to high-quality scanning is that a higher resolution requires physical reduction of the sample size, which reduces the possible volume of investigation [4]. Thus, the scanning system hardware, as well as the physical characteristics of the sample, often limit the ability to obtain high-resolution micro-CT images. In these circumstances, the potential to combine the output of different scanners and enhance the image resolution to maintain a large field of view (FOV) and sample size may have a significant practical value.

To enhance the quality of low-resolution images, super-resolution algorithms that employ neural networks can be applied to the image data [5–10]. As opposed to the convenient interpolation methods (nearest neighbour [11], random forest [12], linear [13], and bicubic [14] interpolation), the convolutional neural network (CNN) methods, which are built within a deep learning framework [15], may use the real high-resolution image data for training to enhance the quality of the low-resolution data [16, 17].

Various CNN architectures have been built for super-resolution processing, and their performance tested on the most common sedimentary rocks – sandstone, limestone, and coal. The effectiveness of super-resolution (SR) CNN models (SRCNN), such as the SR-Resnet, Enhanced Deep SR (EDSR), and Wide-Activation Deep SR (WDSR), for enhancing the quality of digital rock images has been demonstrated in the literature [10]. Most of these architectures are currently available for the users in 2D and 3D, along with the U-Net architecture which was used in this research. The U-Net architecture was first introduced by Ronneberger et al. [18] for medical



(a) (b)

Figure 1: Illustration of the difference between the voxel size and spatial resolution metrics for (a) a non-ZEISS commercially available micro-CT system at 1 μm and (b) a ZEISS Xradia Versa at 1 μm. Although the same carbon fibre composite sample was imaged at the same voxel size of 1 μm, the resulting image quality differs greatly (reproduced from [3]).

image segmentation. Since its introduction, the architecture and its modifications [19, 20] has become known as one of the most successful and reliable CNN architectures for segmentation purposes. It is presumably explained by the U-shape organisation of the structural blocks and the presence of concatenation blocks that combine the low-layer and high-layer features [21]. Although the U-Net is widely used for image segmentation, the architecture is not commonly used for super-resolution. However, the architectures that were built based on the U-Net, were previously used for SR [22, 23].

This research logically continues the work which was reported in Roslin et al. [24]. In the authors' earlier research, the U-Net 3D architecture was used to build a SRCNN model. The model was trained on two triplexes of micro-CT XRadia Versa image data, and it was demonstrated that the U-Net 3D architecture significantly enhanced the quality of the lower resolution images. The work also demonstrated how the enhanced image quality influences the segmentation results. This work investigates the validity and performance of the U-Net 3D model when trained on a combination of Scanco Medical and XRadia Versa datasets. It studies how the different training parameters are related to the quality of the processed images and presents the results for each. It is also discussed in this paper how SR processing can be optimised to reduce the time required for CNN model training.

To summarise, since the introduction of artificial intelligence methods for image processing, the evolution of these methods has gone in three main directions, namely: creating new techniques and architectures; improving the performance of the existing methods and complicating the architectures, and; applying the current methods for new materials and finding the practical application of the existing methods in the applied sciences.

The authors consider all the above-mentioned research areas as equally important and contributing to the field of science. The research described in this work can be classified as related to the third research area and is aimed at applying the previously written convolutional neural network architecture to explore how AI image processing may help improve the resolution of the rock images made by the low-resolution scanning machine. Many previous studies investigated image quality enhancement using high-resolution images and downsampled images to demonstrate the performance of the algorithms. Low-resolution images are generally downsampled using the predefined coefficient, as discussed in the literature review. Studies focused on combining real images of different origins, and thus, which resolution is not directly correlated, are not widely presented in the literature but, based on the authors' experience, present significant importance for the industry. Another key point highlighted in this work is the application of deep learning methods for solid ore rocks. Some attempts were made to apply deep learning segmentation for ore rocks, but ore rocks are still out of the scope of the majority of the researchers, although successful deep learning implementation examples may encourage the industry to adapt the described practices.

This paper is structured to, first, introduce the applied methodology and the rock samples used for the research (Section 2). This includes a description of the analysed samples, low- and high-resolution micro-CT imaging and an explanation of the U-Net architecture. Following this, the results of the application of super-resolution processing are demonstrated, different combinations of XRadia and Scanco micro-CT images are tested and the results are validated using several image quality assessment criteria (Section 3). The obtained results are discussed (Section 4), and conclusions (Section 5) are drawn on how the research can be used in practice.

2. Methodology

In this research, the 3D U-Net architecture was used for super-resolution processing. The structure of this CNN architecture is presented and described in this section. The input data for SR were obtained by micro-CT scanning of a granodiorite sample. The sample was scanned with different resolutions using two different micro-CT scans, namely XRadia for high-resolution scanning and Scanco for low-resolution imaging.

2.1. Samples

Figure 2 shows the sample of subvolcanic igneous rock that was used in this study. The mineral composition of the sample corresponds to granodiorite. The sample has a porphyritic texture and contains pores, partly-healed fractures, and grains of accessory minerals such as pyrite, rutile, sphalerite, molybdenite, and corundum. Some signs of hydrothermal alteration, such as quartzification and sericitisation, were observed. The sample was subjected to stress induced by a blast wave before the scanning. The blast wave produced several fractures which crossed the entire sample. The sample has a cuboidal shape with dimensions of approximately $18 \times 18 \times 34\,mm$.

2.2. Low-resolution micro-CT scanning

The rock sample was analysed with a micro-computed tomography medical tool (μCT50, SCANCO Medical AG, Brütisellen, Switzerland). The sample was positioned, and stabilised with foam padding, in $34\,mm$ tubes before being scanned with an isotropic voxel size of $10.3\,\mu m^3$. The sample was scanned in air at an energy of $90\,kVp$ and a current of $155\,\mu A$. The integration time was set at $1320\,ms$, once averaged, resulting in a $1.32\,s$ sample time. A $0.1\,mm$ copper filter was used. The images were exported as a DICOM stack to allow for further processing. Characteristics of the micro-CT Scanco dataset are listed in Table 1.

2.3. High-resolution micro-CT scanning

Three resolutions were used to scan the sample to provide sufficient data for training and validation of the deep learning algorithms applied in this study. The sample was scanned at nominal resolutions of 5, 10.3 and 36

Figure 2: Photographic images of the cuboidal rock sample used to examine SR processing of images obtained from medical scanners.

Table 1: Characteristics of the micro-CT data obtained with Scanco and XRadia scanners.

Image set	Width [px]	Height [px]	Depth [px]	Voxel size [μm]
Scanco 10.3	3388	3304	3000	10.3000
XRadia 5.0	988	1012	994	5.0000
XRadia 10.3	1000	1024	1009	10.3000
XRadia 36.0	1004	1024	1016	36.0002

microns (μm). The VersaXRM-500 can potentially achieve a true spatial resolution of 0.9 μm with a minimum achievable voxel size of 0.3 μm. Advanced absorption and phase contrast (for soft or transparent materials) provide greater versatility in overcoming the limitations of traditional computed tomography. Characteristics of the micro-CT XRadia dataset are additionally listed in Table 1. The parameters of the micro-CT scanning are summarised in Table 2. After scanning, the images were exported in a .txm file format for further processing.

Table 2: Scanning parameters used for the three resolutions obtained from the XRadia 3D microscope.

Image set	Lens	Voltage [kV]	Filter	Voxel size [μm]
XRadia 5.0	0.4×	80	Air	5.0
XRadia 10.3	0.4×	80	Air	10.3
XRadia 36.0	0.4×	80	Air	36.0

2.4. Convolutional neural network architecture

The current research employed a CNN architecture which is called U-Net 3D. The U-Net architecture was introduced by Ronneberger [18] to segment large medical images. This architecture can be described as a U-shape architecture consisting of two networks, namely the encoder and the decoder (Figure 3).

The U-Net architecture consists of a convolution operation, max pooling, rectified linear unit (ReLU) activation, concatenation, and upsampling layers. These blocks form a contracting path (left side) and expansive path (right side) [18]. The contracting path (encoder) is composed of the repeated convolutional layers in which the filters slide along the input data and produce specific feature maps, thus, extracting the key features from

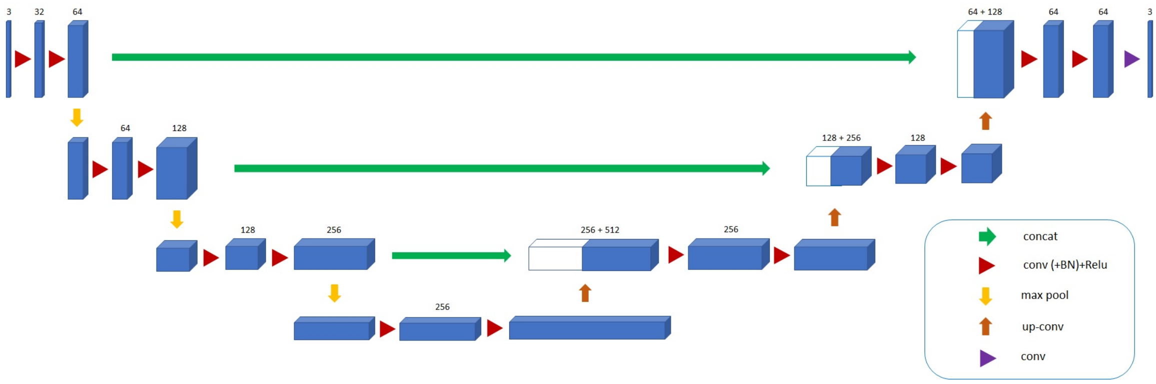


Figure 3: The 3D U-Net architecture in which blue boxes represent feature maps. The number of channels is denoted above each feature map (reproduced from [19]).

the input dataset. The size of the filters is $3 \times 3 \times 3$. The process of sliding the filter along the data is referred to as a stride. For example, a stride of one means that the filter moves one unit along the data matrix. Each convolution layer is followed by an activation function. The activation function defines how the weighted sum of the input from the previous layer is transformed into an output from a node or nodes in the next layer of the network. The U-Net uses the ReLU activation function, which is one of the most popular activation functions in deep learning due to its simplicity and effectiveness. The ReLU function is calculated as the maximum of zero and x, which means that if the input value, x, is negative, the function returns a value of zero, otherwise, the value x is returned,

$$f(x) = \begin{cases} 0, & \text{if } x \leq 0 \\ x, & \text{if } x > 0. \end{cases} \tag{1}$$

The activation function is followed by the pooling layer. This layer is used to downsample the feature maps by summarising the presence of the elements in the feature map's patches while preserving the essential structure of the data [25]. The U-Net uses a $2 \times 2 \times 2$ maximum pooling operation with a stride of two for downsampling.

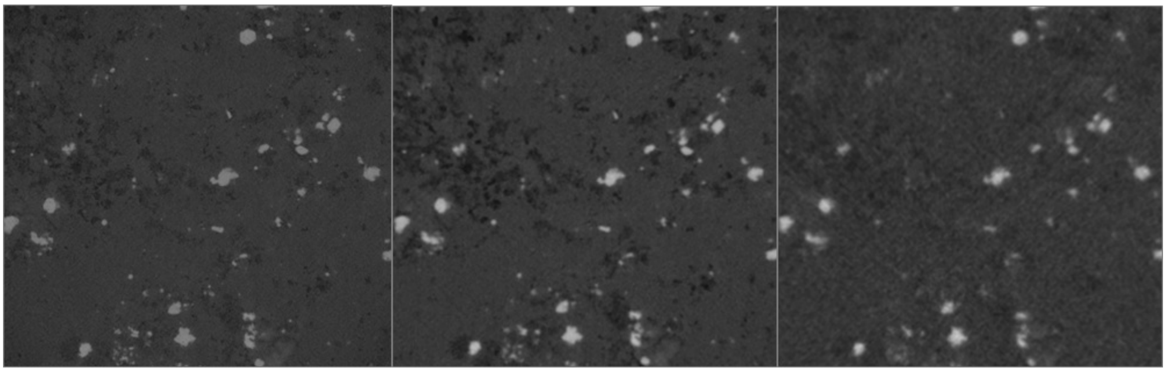
The second part of the architecture is the decoder. The decoder consists of upsampling and concatenation blocks followed by convolution operations. The upsampling procedure expands the feature dimensions and halves the number of feature channels, restoring the feature map to the original size of the input image. It is required to meet the same size with the corresponding concatenation blocks from the encoder part. The 3D U-Net architecture has a similar organisation, but an extra depth dimension is added.

3. Results

The real micro-CT rock images were utilised to enhance the quality of the low-resolution input data using the 3D U-Net architecture. The training parameters are presented in this section, as well as the quality assessment criteria used to analyse the accuracy of the SR processing. A combination of the micro-CT images with different resolutions from two micro-CT scanners was used. Several configurations were tested, and the results were analysed. Conclusions about the most accurate combination were drawn, and different combinations of micro-CT data were analysed to estimate the most optimal model training time.

3.1. Training

The input dataset consisted of micro-CT images obtained at different resolutions, as given in Table 1. Those datasets were aligned and registered on the XRadia 36.0 image set and resampled to correspond to the resolution of the XRadia 5.0 set. Figure 4 shows the area of investigation for the low-resolution input (Scanco 10.3 dataset) and the high-resolution ground truth images (Xradia 5.0 and XRadia 10.3 image sets). The XRadia 36.0 image set was used only for data registration. After registering, the same volume of investigation was clipped from each set for training. The clipped volume consisted of $690 \times 651 \times 875$ voxels. The model was trained with and without a mask. When the mask was used, only one out of four slices was kept for training. The total number of voxels for the input dataset was 393,041,250 without the mask and 98,372,610 with the mask. The volumes contained only the rock material, while the background (air and the sample holder material) was excluded.



(a) (b) (c)

Figure 4: Comparison of the (a) 5 μm XRadia set (ground truth), (b) the 10.3 μm XRadia set (second ground truth), and (c) the 10.3 μm Scanco set (input dataset). The same area of investigation is shown. Each image size is 690×651 voxels, or $3,450 \times 3,255$ voxels.

The next step was to create the U-Net 3D models. This was done using the Deep Learning Tool in the software package, Dragonfly (Object Research Systems Inc.). The default model architecture parameters were used, including a depth level of four, patch size of $32 \times 32 \times 32$, and initial filter count of 32. The models were trained with the Adadelta optimiser using a default learning rate parameter of 1.0. The learning rate was reduced by a decay factor of 0.1 once the learning stagnated, given a patience of 10 epochs. The models were trained for 100 epochs, using mini-batch stochastic gradient descent on a mean square error loss function, with a batch size of 128.

Several U-Net models were created to enhance the resolution of the Scanco images. The input image set was the same for all models (for one of the models, XRadia 10.3 set was added as a second set). The output image set was XRadia 10.3 or XRadia 5.0. These datasets were also used as ground truth, as shown in Figure 4. The models were trained with and without augmentation. When augmentation was used, two times augmentation with default flip, rotate, shear, and scale parameters was chosen. Augmentation was used together with a mask for model training. The models which were used and the configurations specified are summarised in Table 3.

3.2. Validation

The first round of validation is performed during the model training. The training process was controlled using a mean squared error (MSE) loss function, and the output model was validated against 20% of the total training data that had been reserved for this purpose. When the models were trained, they were applied to the whole volume of investigation of the studied samples and several metrics were analysed to estimate the

Table 3: The 3D U-Net models created for comparison of performance in SR processing of Scanco data.

Model	Input dataset	Output dataset	Mask used	Augmentation used	Resulted dataset
#1	Scanco 10.3	XRadia 10.3	no	no	Processed #1
#2	Scanco 10.3	XRadia 5.0	no	no	Processed #2
#3	Scanco 10.3, XRadia 10.3	XRadia 5.0	no	no	Processed #3
#4	Scanco 10.3	XRadia 10.3	yes	yes	Processed #4
#5	Scanco 10.3	XRadia 5.0	yes	yes	Processed #5
#6	Scanco 10.3	XRadia 10.3	yes	no	Processed #6
#7	Scanco 10.3	XRadia 5.0	yes	no	Processed #7

quality of the processed and input image data. This volume of investigation contained the overlapping volumes from each dataset. The first parameter was the peak signal-to-noise ratio (PSNR). PSNR is expressed as the ratio between the maximum possible value (power) of a signal and the power of distorting noise that affects the quality of its representation. The second parameter was the structural similarity index measure (SSIM) which is a perceptual metric that quantifies image quality degradation caused by processing. SSIM measures the perceptual difference between two similar images. In this case, these are the ground truth and processed input image, and the ground truth and resampled input image. In other words, SSIM looks for similarities within pixels (if the pixels in the two images line up and or have similar pixel density values). SSIM is a standardised quality metric and is often considered one of the most objective criteria to estimate the quality of image reconstruction. The analysed parameters are summarised in Table 4 for each U-Net model created, as well as the raw Scanco images. For models which used the XRadia 5.0 image set, solely or together with the XRadia 10.3 set, as an output dataset (Table 3), the Xradia 5.0 set was used as a ground truth (Table 4). For the models trained using the XRadia 10.3 dataset only, the processed images were assessed using the XRadia 10.3 set.

As soon as the quality control parameters are analysed, the models were also applied to the whole volumes of data (for which the ground truth high-resolution data were not available). When the entire input datasets are reprocessed, the images were also analysed visually to estimate the quality of the reconstruction in the areas where ground truth high-quality images were not available.

Table 4: Image quality assessment criteria for datasets compared with the ground truth sets (XRadia 5.0 and XRadia 10.3 image sets).

Processed dataset	Ground truth dataset	PSNR (dB)	SSIM
Processed #1	XRadia 10.3	32.74	0.85
Processed #2	XRadia 5.0	35.35	0.86
Processed #3	XRadia 5.0	35.14	0.87
Processed #4	XRadia 10.3	33.02	0.86
Processed #5	XRadia 5.0	35.31	0.86
Processed #6	XRadia 10.3	21.02	0.78
Processed #7	XRadia 5.0	35.15	0.86
Scanco 10.3	XRadia 10.3	9.28	0.57

Table 4 also contains the quality metrics for the input Scanco 10.3 image set. It can be seen that the PSNR of the Scanco set is much lower than that of the processed datasets. PSNR (in dB) is defined as,

$$PSNR = 20log_{10}(V_{max}) - 10log_{10}(MSE), \tag{2}$$

where V_{max} is the maximum possible voxel value. For the analysed images (16-bit images), it is 65,535.

As the MSE approaches zero, the PSNR tends to infinity. It is usually supposed that the higher the PSNR, the better the quality of the reconstructed images. Wang et al. [10] demonstrated good super-resolution processing results and the reported average PSNR was in the range of approximately $23 - 29\,dB$ depending on the rock type. For the SR processing of the XRadia triplexes, the PSNR was reported in the range of $34 - 36\,dB$ [24].

The SSIM ranges between zero and one. For good reconstruction techniques, the SSIM should be close to one. The reported values of the SSIM for the XRadia triplexes were in the range 0.81-0.91.

Figure 4 supports the conclusion that the input medical image set is characterised by a significant noise level (grain effect). Figure 5 compares the processed datasets with the input Scanco image set. Both Figure 5 and Table 4 show that SR processing can significantly enhance the resolution of the micro-CT data by sharpening the contrast and suppressing background noise. The grain noise which was present on the input images is not observed on the reconstructed images, while the appearance of the image features is improved. The importance of this observation will be discussed in the next section. Figure 6 demonstrates an example of 3D cube and three orthogonal projections for the processed #2 set. In turn, Figure 7 shows an example of the processed masked dataset (processed #7).

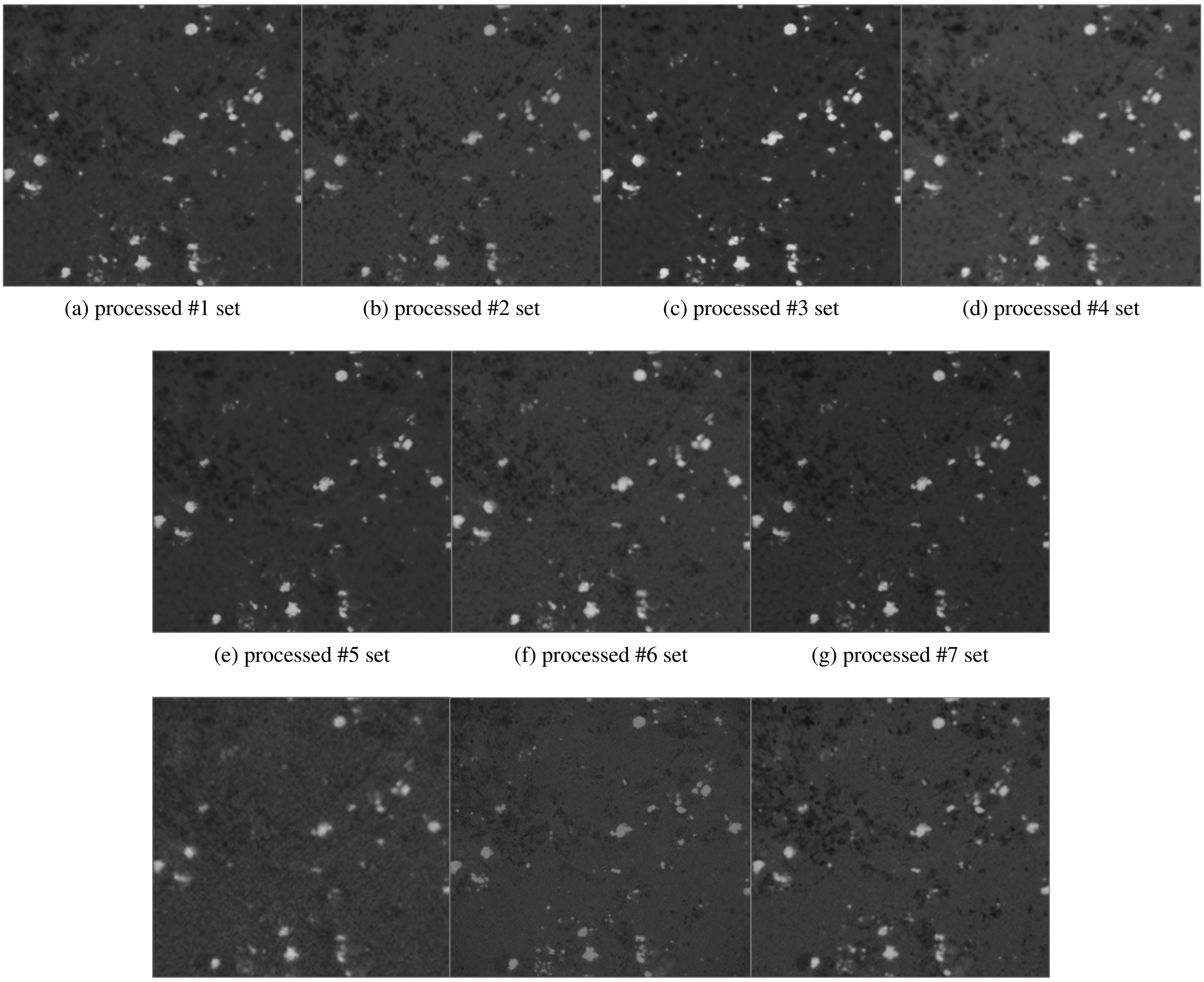
Assessment of all processed images demonstrates that the models that used the XRadia 5.0 dataset performed better than the models trained on the XRadia 10.3 set (see Table 4). The performance of the model built on the XRadia 10.3 set can be slightly improved by adding a second set of XRadia 10.3 images (processed #3 dataset). However, using two image datasets increases the training time (from approximately 18 hours to 26 hours). To optimise the training process, masked datasets can be used instead of the original image datasets. In this study, masked datasets were used with and without augmentation. Augmentation did not influence the performance of the models built on the XRadia 5.0 set, but in the case of the model trained with the XRadia 10.3 set, the absence of augmentation significantly degraded the processed images. These results will be discussed in more detail in the next section.

This paper is primarily focused on combining two different scanners, and the detailed analysis of the influence of the SR on the permeability and porosity estimation. However, it has been done in the previous paper of the authors [24].

4. Discussion

The current research builds on the work of Roslin et al. [24], which investigated the application of SR processing to granodiorite rock samples. In this earlier work, super-resolution techniques were applied to the combination of XRadia micro-CT images obtained with different resolutions using the same minicores and instrument. The performance of the SRCNN models was analysed, as well as the manner in which the enhanced quality of the processed images influenced the segmentation results, and how to choose the resolutions of the image pair to get the optimal improvement of the image quality. In this research, SR processing was employed for a dataset comprised of Scanco and XRadia images of the same rock sample.

SR processing of micro-CT images of rocks is an actively developing area of research. However, the ongoing published research studies have primarily focused on creating and testing the performance of different CNN architectures. The images which are usually utilised for SR processing are the downsampled high-resolution XRadia images. To the best of the authors' knowledge, only a few attempts have been made to combine real rock images obtained from different instruments. However, combining the images from different instruments to



(a) processed #1 set (b) processed #2 set (c) processed #3 set (d) processed #4 set

(e) processed #5 set (f) processed #6 set (g) processed #7 set

(h) input Scanco 10.3 set (i) 5 μm XRadia set (j) 10.3 μm XRadia set

Figure 5: Comparison of the processed and input image sets, showing (a) the processed #1 set, (b) the processed #2 set, (c) the processed #3 set, (d) the processed set #4, (e) the processed set #5, (f) the processed set #6, (g) the processed set #7, (h) the input Scanco 10.3 set, (i) the 5 μm XRadia set (ground truth), and (j) the 10.3 μm XRadia set (second ground truth). The same area of investigation is shown. Each image size is 690×651 voxels, or $3,450 \times 3,255$ voxels.

increase the resolution of low-resolution images may have significant practical value. It could make it possible to revise previously acquired image datasets to enhance their quality, or scan large volumes of cores faster due to the opportunity to reprocess low-resolution images rather than physically scanning entire samples with high resolution.

The research conducted in this study and previously published work [24] was preceded by a literature survey of existing CNN architectures used to enhance image quality. The literature review section demonstrates the overview of the different super-resolution methods which were also considered for image enhancement. U-Net is not the most modern deep learning architecture, but the performance of this architecture is sufficient to make the U-Net acceptable for rock image resolution enhancement. This particular research and some other studies mentioned in the literature review show that the U-Net architecture and its modifications may significantly improve the image resolution, and there are no objective reasons not to use this architecture (as well as any other CNN architecture) for image processing. As a part of the CNN exploratory process, the performance of

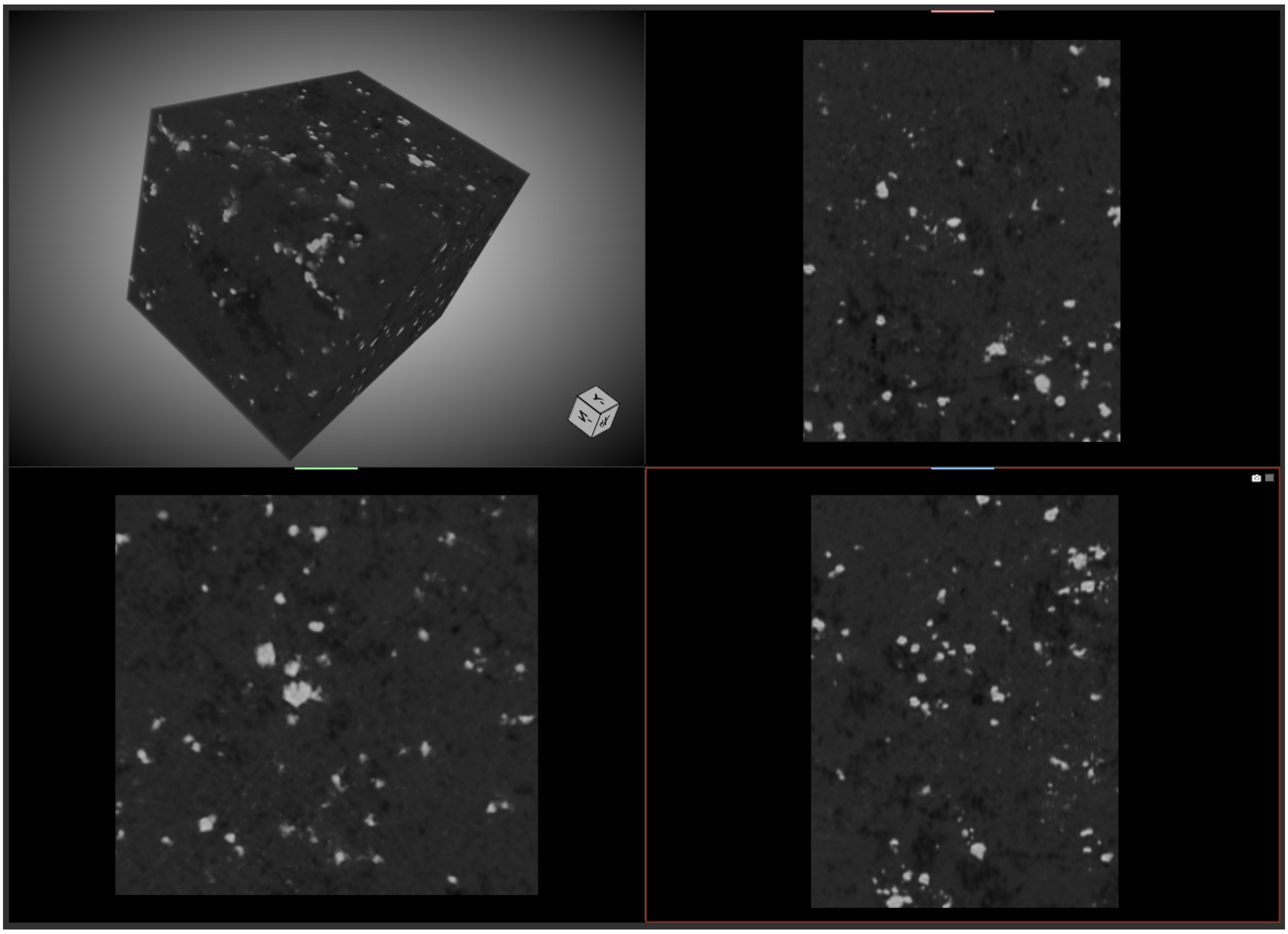


Figure 6: 3D cube and three orthogonal projections for the processed #2 set.

available CNN architectures (2D and 3D U-Net, EDSR, WDSR) was analysed and the outcomes from 2D and 3D architectures was compared using a small image dataset. As demonstrated, the performance of the U-Net is comparable to or better than that of the other most popular CNN models (e.g., WDSR and EDSR). Other CNNs were also tested in a similar manner (the workflow is presented in the paper). Generally, they demonstrated lower PSNR and especially SSIM values for the studied samples, but still comparable to the results published in the literature [9, 10, 17]. The methods which were used to compare with the U-Net architectures are explained in more detail in the papers cited in the literature review (e.g. [9, 10]).

In turn, GAN architectures consist of two neural networks – one of which is a convolutional neural network. The generator creates an image, and the discriminator evaluates it. The GANs may produce images of superior quality, however, the generator is basically not trained to minimise the distance to the ground truth image but rather to create the image which will be recognised by the discriminator as realistic. The GANs are often used to enhance the quality of the images when the purpose is to create realistic textures rather than optimise for a pixel-accurate reproduction of ground truth images during training. The purpose of the authors was to enhance the image quality but to ensure that the resulting images are pixel-accurate and close to the real ground truth since the images are used to calculate porosity and mineral content. Thus, the authors find it essential to improve the quality of the images by focusing on preserving the input images rather than indefinite improvement of texture resolution. So, the GANs were not chosen for the super-resolution processing and the main focus was on the CNN architectures.

The 3D U-Net architecture, which was ultimately chosen for super-resolution processing, is one of the most

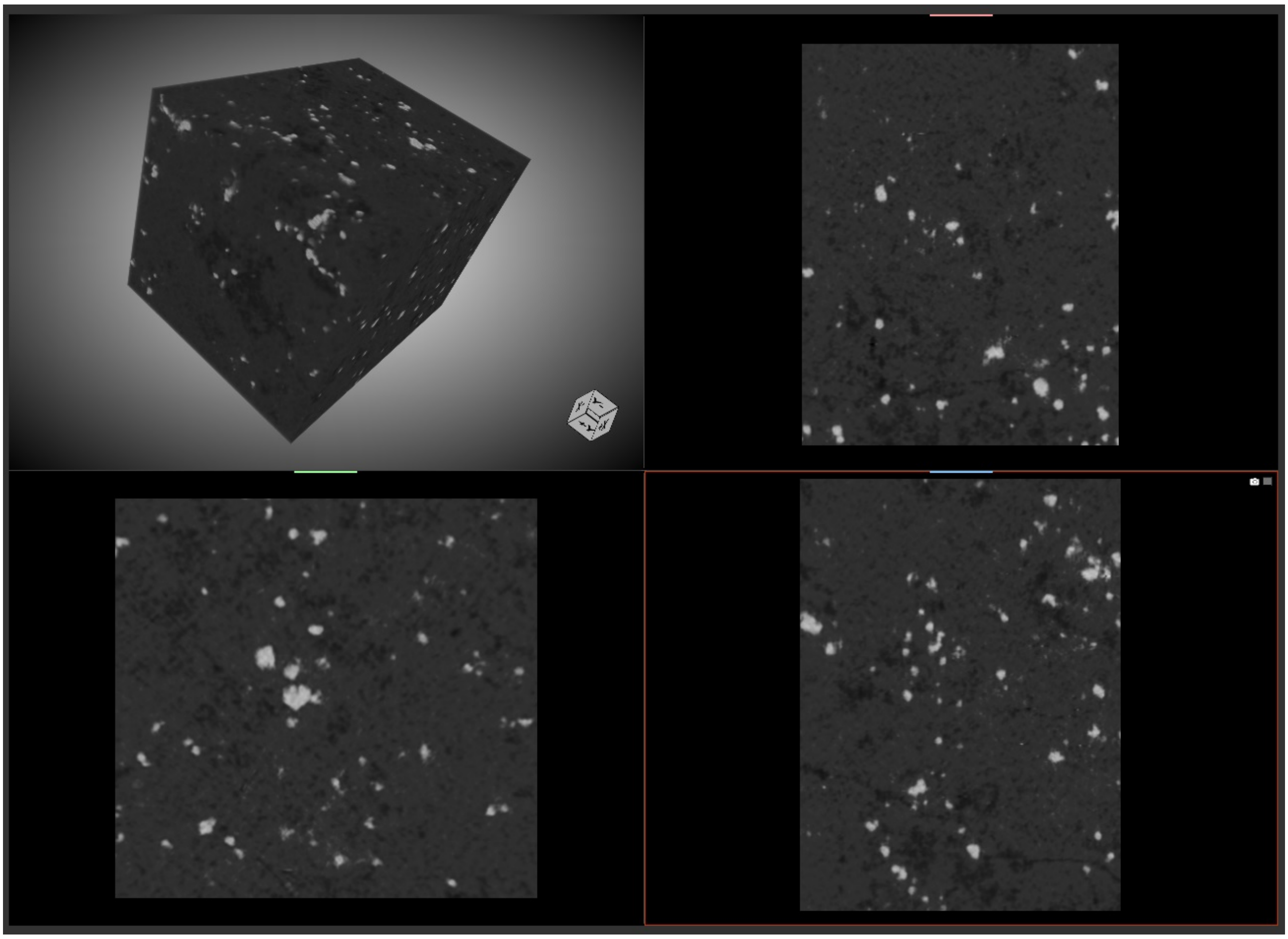


Figure 7: 3D cube and three orthogonal projections for the processed #7 set.

popular architectures for image segmentation. However, it is rarely used for SR processing of rock images even though it was reported that U-shape architectures demonstrated successful outcomes for SR enhancement of image quality [22, 23]. The 3D U-Net demonstrated more accurate detection of boundaries of the rock features (voids and mineral grains) and more robust results when the trained model was applied to unseen image data. This can be explained by the nature of the micro-CT images. The signal from the actual materials within a single voxel is partially distributed across neighbouring voxels as defined by the point spread function (PSF), which can be represented as a Gaussian smoothing kernel. In turn, each micro-CT slide is not a static photo of the object but rather an attenuation of the material spread over the 3D grid, where a grid step is equal to the voxel size. Thus, it is important to take into account the third (depth) dimension for SR processing of micro-CT images.

In the current research, the 3D U-Net CNN architecture was used to train the models and enhance the quality of Scanco images using XRadia images. The XRadia micro-CT tool is generally preferred for digital rock analysis, however geoscientists also use medical scanners, such as Scanco, for rock analysis [26]. The outcomes of different instruments may have the same voxel size, but different spatial resolution. The spatial resolution determines the feature separation (see Figure 8). If the same pair of features is separated by spacing smaller than the resolution of the scanning system, it becomes indistinguishable as a pair on the image [3]. Consequently, poor spatial resolution may influence segmentation accuracy [24]. The comparison of Scanco and XRadia images showed that the spatial resolution of the XRadia system is higher than that of the Scanco system, even when the voxel size of both image sets is similar (10.3 microns). For this research, SRCNN processing was used to enhance the resolution of the Scanco 10.3-micron images using the XRadia 10.3- and 5.0-micron image data. All processed images had quality comparable with the ground truth images (see Table 4 and Figure 5).

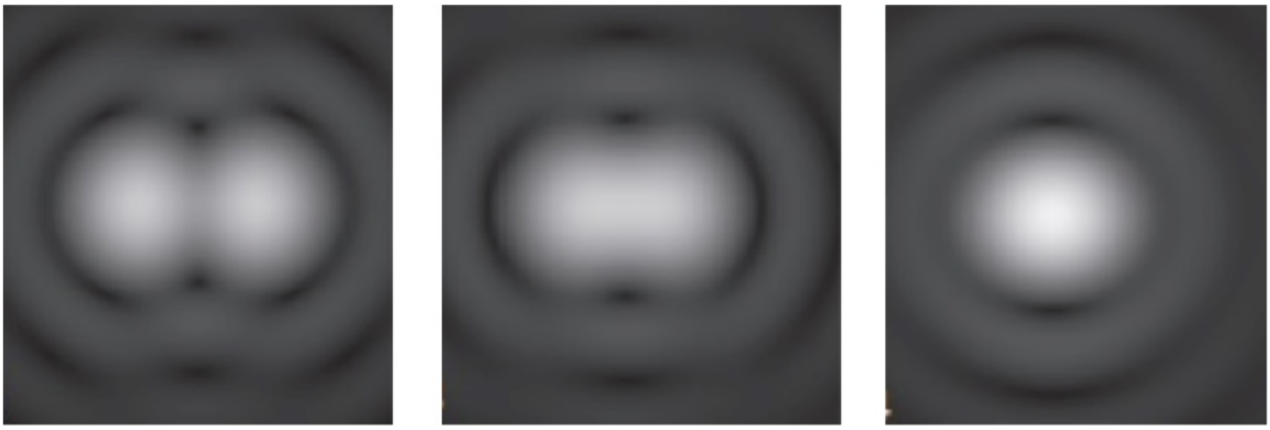


Figure 8: Spatial resolution as a function of feature separation (reproduced from [3]).

The models trained using the XRadia 5.0-micron images outperformed those trained only on the XRadia 10.3-micron data (for comparison, Model #2 and Model #1). However, adding 10.3-micron images as the second training dataset improved the quality of the processed images (see Table 4 Model #3 compared to Model #1). The models built on 5.0-micron data produce nearly similar results for the masked dataset. Augmentation was also found to have minimal impact on the processing results. For the models built only on 10.3-micron data, it is recommended to use the masked dataset with augmentation since the masked dataset without augmentation produced the least accurate, yet still acceptable, results (the PSNR values are comparable to that reported in the literature [10]).

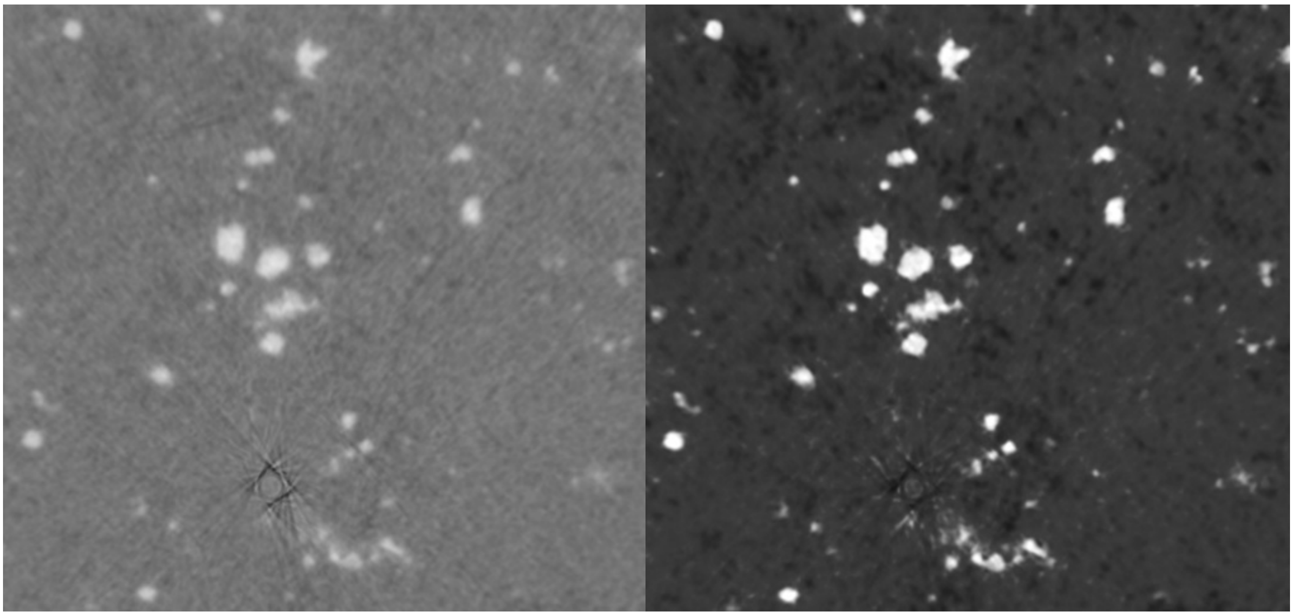
After the comparison of the processed results, it was concluded that the training process could be optimised by masking. For the analysed data, the training dataset was reduced by choosing every fourth slide. The results of SR processing were not degraded, but the training time was significantly reduced (from approximately 18 to eight hours without augmentation and 14 hours with augmentation).

As observed, SRCNN processing performed very well for the combination of images with the exact voxel sizes, and also for the pair of the datasets with aliquant voxel sizes. In the case of image sets with different nominal resolutions, the lower resolution datasets should be resampled to correspond to the higher resolution image set. In this work, a linear interpolation function was employed, but based on the empirical experience in some cases, other interpolation functions (such as the nearest neighbour) can be used for resampling.

Another procedure, which is required before model training, is image registration. The registration was done manually on the fixed 36-micron dataset followed by automatic registration using an initial step of $VS/10$ microns and smallest step of 0.01 micron (for translation) and initial step of 1° and smallest step of 0.2° (for rotation), where VS is the voxel size of the registered image set (in microns). Linear interpolation and sum square differences (SSD) functions were employed for image registration. Registration and resampling are usually not mentioned in literature studies discussing SRCNN, although these are necessary pre-training procedures. For example, even a shift in one voxel significantly degrades the processing results and introduces errors in image resampling.

The images presented in the manuscript were primarily chosen for illustration to visually demonstrate the input quality and processed images. The model was trained on 3D volumes and applied to the large 3D volumes of images. The validation results were also obtained for the whole 3D volumes, and the validation results are presented in Table 4. This approach assures that the quality assessment is more unbiased and repeatable, and the results can be compared to outcomes of the performance of other neural network models. The images presented in the manuscript were chosen randomly, and it should be noted that these are not the only successful examples but rather the typical examples of the model performance.

This paper demonstrated that super-resolution can successfully be used for micro-CT image resolution enhancement. It was shown that SRCNN processing improves the recognition of feature boundaries and suppresses unwanted noise associated with images obtained from medical CT equipment. However, it was observed that some types of noise, particularly ring artefacts, could not be removed solely by SRCNN and may require a deep learning denoising procedure or massaging of the input images before SR model training. This is one of the limitations of the proposed CNN super-resolution model (Figure 9). However, the ring artefacts can removed during the post-scanning stage. The images used in the research were not processed to remove the ring artefacts and some artifacts are still present on the SR processed images. Other limitations are discussed in more detail in another paper by the authors [24], where that question was analysed more thoroughly. The main limitations are related to the size of the features, which resolution should be enhanced. Since the CNN model enhances the resolution of the input images using the real high-resolution images rather than creating high-resolution unpaired images which are not voxel-accurate, the features which are not present on the input images won't appear on the processed images.



(a) input Scanco 10.3 set (b) processed #2 set

Figure 9: Comparison of the processed and input image sets, showing (a) the input Scanco 10.3 set, and (b) the processed #2 set. The same area of investigation is shown. The image artefacts are observed very well on the input image, and slightly removed but still presented on the processed images.

SR processing has been shown to be a powerful tool for digital rock analysis, which has a straightforward, practical application for rock characterisation. If SRCNN processing is appropriately applied, namely, pre-training stages are accurately conducted, it has been shown in this study that it can improve low resolution images to a level comparable to that of high resolution scans. Only a small dataset of the high-resolution images is required for the training to ensure that the model relies on the real data. Having this ground truth dataset can be compared to using the reference standard for calibration. Some research studies demonstrated that unpaired images could be used to train and apply neural network models. Still, the authors' position is that some high-resolution images should always exist for low-resolution datasets to claim that the result can be reliable and not purely mathematical.

The main practical advantage of applying the proposed method is saving time which is required to scan large volumes of rocks with high-resolution settings. Only one high-resolution image set is needed to use a deep learning algorithm to receive quite decent processed images. If the model is trained on a high-resolution dataset, it can be applied to a large volume of the low-resolution data without re-training. Only image normalisation could be required, which is performed easily. In the future, work targeted at SRCNN processing of a combination of micro-CT images with a focus on the practical application of this method for geosciences will be continued.

5. Conclusions

This research investigated the application of SRCNN processing techniques to enhance the resolution of medical micro-CT images of granodiorite rock samples and analysed the processed image quality. The training dataset consisted of Scanco 10.3-micron and XRadia 10.3- and 5.0-micron image sets. The CNN was trained by taking 80% of the images from the ground truth scan, leaving 20% for validation. Several SRCNN models were trained to assess the performance of the 3D U-Net architecture and the influence of different parameters on the resultant image quality. It was concluded that the best models were those trained on the XRadia 5.0-micron images, and the training process could be expedited if masking was utilised. Masking assumed using every fourth slide from the image dataset and reduced the model training time. In turn, augmentation did not significantly improve the quality of the processed images. However, in the case of models trained only on the XRadia 10.3-micron images, masked datasets without augmentation demonstrated the worst results of the processed image quality. It was concluded that augmentation might be necessary for masked datasets of lower image resolution. Another method to improve the results of the XRadia 10.3 models was to add XRadia 5.0-micron images as a second training dataset.

In this research, it was observed that SR processing improves the image quality not only due to better detection of feature boundaries but also by noise suppression which is especially important for medical images. This research demonstrated that the quality of the rock images obtained by the medical scanners could successfully be improved by SRCNN processing which opens new opportunities for practical implementation of deep learning techniques for rock analysis.

6. Acknowledgments

The authors acknowledge the financial support of this work from BHP. Special thanks to Jared Townsend for his help and collaboration. The authors also acknowledge and thank the Centre in Regenerative Medicine at the School of Mechanical, Medical and Process Engineering at the Queensland University of Technology for conducting the micro-CT scanning (Scanco medical) used in this work. In particular, the authors thank Dr. Sinduja Suresh and Dr. Marie-Luise Wille for their assistance with scanning.

7. Data availability

The software used for the research is Dragonfly (Object Research Systems (ORS) Inc.). Object Research Systems (ORS) Inc. provides researchers with a free academic licence for their software. The licence covers all modules used in the research. The software can be found at https://www.theobjects.com/company/products.html

The deep learning architectures used for the research are available in the software, all parameters for model training and the workflow are explained and illustrated in the manuscript.

The datasets used for this research are available here:

https://cloud.rdm.uq.edu.au/index.php/s/RizHetaY4JPLsxE (the password is "Open Access").

References

[1] A. Golab, C. R. Ward, A. Permana, P. Lennox, and P. Botha, "High-resolution three-dimensional imaging of coal using microfocus X-ray computed tomography, with special reference to modes of mineral occurrence," *International Journal of Coal Geology*, vol. 113, pp. 97–108, 2013.

[2] H. L. Ramandi, P. Mostaghimi, R. T. Armstrong, M. Saadatfar, and W. V. Pinczewski, "Porosity and permeability characterization of coal: A micro-computed tomography study," *International Journal of Coal Geology*, vol. 154-155, pp. 57–68, 2016.

[3] G. Carl Zeiss Microscopy GmbH., "Resolution of a 3D X-ray microscope. Defining meaningful resolution parameters for XRM," *Technical note*, 2013.

[4] M. J. Blunt, *Multiphase flow in permeable media: A pore-scale perspective*. Cambridge: Cambridge University Press, 2017.

[5] C. Dong, C. C. Loy, K. He, and X. Tang, "Image Super-Resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, pp. 295–307, 2016.

[6] P. Dong, B. Provencher, N. Basim, N. Piché, and M. Marsh, "Forget about cleaning up your micrographs: Deep learning segmentation is robust to image artifacts," *Microscopy and Microanalysis*, vol. 26, pp. 1468–1469, 2020.

[7] B. Lim, S. Son, H. Kim, and K. M. Lee, "Enhanced deep residual networks for single image Super-Resolution," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 136–144, 2017.

[8] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image Super-Resolution using a Generative Adversarial Network," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 4681–4690, 2017.

[9] Y. D. Wang, R. T. Armstrong, and P. Mostaghimi, "Boosting resolution and recovering texture of 2D and 3D micro-CT images with Deep Learning," *Water Resources Research*, vol. 56, p. e2019WR026052, 2019.

[10] Y. D. Wang, R. T. Armstrong, and P. Mostaghimi, "Enhancing resolution of digital rock images with Super Resolution Convolutional Neural Networks," *Journal of Petroleum Science and Engineering*, vol. 182, p. 106261, 2019.

[11] Y. Tang, P. Yan, Y. Yuan, and W. Liu, "Single-image super-resolution via local learning," *International Journal of Machine Learning and Cybernetics*, vol. 2, pp. 15–23, 2011.

[12] Z. Li, Q. Teng, X. He, G. Yue, and Z. Wang, "Sparse representation-based volumetric super-resolution algorithm for 3D CT images of reservoir rocks," *Journal of Applied Geophysics*, vol. 144, pp. 69–77, 2017.

[13] A. Lukin, A. Krylov, and A. Nasonov, "Image interpolation by Super-Resolution," *In: 16th International Conference Graphicon 2006, Novosibirsk Akademgorodok*, pp. 239–242, 2006.

[14] A. B. Gavade and P. Sane, "Super Resolution image reconstruction by using bicubic interpolation," *In: ATEES 2014 National Conference, Belgaum. India*, pp. 204–209, 2013.

[15] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, p. 85–117, 2015.

[16] H. Chen, X. He, Q. Teng, R. E. Sheriff, J. Feng, and S. Xiong, "Super-resolution of real-world rock micro-computed tomography images using cycle-consistent generative adversarial networks," *Physical Review E*, vol. 101, p. 023305, 2020.

[17] Y. Wang, S. S. Rahman, and C. H. Arns, "Super resolution reconstruction of micro-CT image of rock sample using neighbour embedding algorithm," *Physica A: Statistical Mechanics and its Applications*, vol. 493, pp. 177–188, 2018.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for biomedical image segmentation," *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, 2015.

[19] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d U-Net: Learning dense volumetric segmentation from sparse annotation," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, Proceedings, Part II*, 2016.

[20] Z. Zhou, S. M. M. Rahman, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA 2018, ML-CDS 2018. Lecture Notes in Computer Science*, vol. 11045, 2018.

[21] C. Du, Y. Wang, C. Wang, C. Shi, and B. Xiao, "Selective feature connection mechanism: Concatenating multi-layer CNN features with a feature selector," *In press*, vol. 129, pp. 108–114, 2020.

[22] X. Hu, M. A. Naiel, A. Wong, M. Lamm, and P. Fieguth, "RUNet: A robust UNet architecture for image super-resolution," *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 505–507, 2019.

[23] H. Zang, L. Zhu, Z. Ding, X. Li, and S. Zhan, “Cascaded Dense-UNet for image super-resolution,” *Journal of Circuits, Systems, and Computers*, vol. 29, p. 2050121, 2019.

[24] A. Roslin, M. Marsh, N. Piché, B. Provencher, T. Mitchell, I. Onederra, and C. Leonardi, “Processing of micro-CT images of granodiorite rock samples using convolutional neural networks (CNN). Part I: Super-resolution enhancement using a 3D CNN,” *Minerals Engineering*, 2022.

[25] S. Kamrava, P. Tahmasebi, and M. Sahimi, “Linking morphology of porous media to their macroscopic permeability by deep learning,” *Transport in Porous Media volume*, vol. 131, pp. 427–448, 2019.

[26] L. Falco and M. Burkhart, “Extended characterization of pore structure in sandstone. Application Note,” *SCANCO Medical AG, Brüttisellen, Switzerland*.