Applied Computing and Geosciences 22 (2024) 100165

Contents lists available at ScienceDirect

Applied Computing and Geosciences

journal homepage: www.sciencedirect.com/journal/applied-computing-and-geosciences

Single image multi-scale enhancement for rock Micro-CT super-resolution using residual U-Net

Liqun Shan [a], Chengqian Liu [b,*], Yanchang Liu [b], Yazhou Tu [c], Sai Venkatesh Chilukoti [a], Xiali Hei [a]

[a] School of Computing & Informatics, University of Louisiana at Lafayette, Lafayette, LA, 70503, USA
[b] School of Physical and Electrical Engineering, Northeast Petroleum University, Daqing, 163318, China
[c] Computer Science and Software Engineering, Auburn University, Auburn, AL, 36849, USA

ARTICLE INFO

Keywords:
Super-resolution
Micro-CT rock image
Residual block
U-Net
Multi-scale fusion

ABSTRACT

Micro-CT, also known as X-ray micro-computed tomography, has emerged as the primary instrument for pore-scale properties study in geological materials. Several studies have used deep learning to achieve super-resolution reconstruction in order to balance the trade-off between resolution of CT images and field of view. Nevertheless, most existing methods only work with single-scale CT scans, ignoring the possibility of using multi-scale image features for image reconstruction. In this study, we proposed a super-resolution approach via multi-scale fusion using residual U-Net for rock micro-CT image reconstruction (MS-ResUnet). The residual U-Net provides an encoder-decoder structure. In each encoder layer, several residual sequential blocks and improved residual blocks are used. The decoder is composed of convolutional ReLU residual blocks and residual chained pooling blocks. During the encoding-decoding method, information transfers between neighboring multi-resolution images are fused, resulting in richer rock characteristic information. Qualitative and quantitative comparisons of sandstone, carbonate, and coal CT images demonstrate that our proposed algorithm surpasses existing approaches. Our model accurately reconstructed the intricate details of pores in carbonate and sandstone, as well as clearly visible coal cracks.

1. Introduction

To solely rely on the rock structure, rock color, rock thickness and rock water absorption properties of the reservoir is far from sufficient for acquiring large-scale oil and gas fields. Influential oil and gas resources can be discovered by studying the microscopic levels of rocks, such as pore structural features, pore morphology and pore fissure ratio (Bostanabad et al., 2018; Ju et al., 2014, 2017; Hou et al., 2021; Kuang et al., 2021; Zhu et al., 2019). It is essential to accurately simulate three-dimensional (3D) pore structure while researching physical and transport features (Bultreys et al., 2016; Fu et al., 2021, 2022; Li et al., 2018; Tian et al., 2020). Imaging tools like micro-CT can be used to obtain the 3D core pore structure, which was known as "digital core" technology (Feng et al., 2020; Gerke et al., 2017; Karimpouli et al., 2020; Ju et al., 2019; Sanematsu et al., 2019; Schlüter et al., 2014; Tan et al., 2021; Zhang et al., 2019). However, due to the cost of image acquisition equipment and the undesirably external environment, it is not always possible to obtain high-resolution (HR) images of large-sized cores with sharp edges and rich details. Micro-CT technology can reveal microstructure, but we have to admit that zooming in on rock images can lead to distortions, which means seeing finer details in low-resolution (LR) images is challenging. To address these issues, image super-resolution (SR) reconstruction technology has been used in the core CT image to improve the image's resolution (Bizhani et al., 2022; Dong et al., 2014; Tao et al., 2017; Wang et al., 2020; Feng et al., 2022; Zhou et al., 2021). This allows for an in-depth understanding of the rock's specific characteristics and structural analysis. At the same time, it provides a solid scientific basis for the research and exploration of oil or gas reservoirs.

Current single-image SR reconstruction techniques can be divided as are four categories: interpolation-based, shallow learning-based, reconstruction-based, and deep learning-based approaches. The first three techniques have gradually faded out of researchers' field of vision due to their inaccuracy. Many academics have researched deep learning-based SR reconstruction techniques in the past decade. Dong et al. (2014) first introduced deep learning to the field of image SR

* Corresponding author.
E-mail address: 218002090449@stu.nepu.edu.cn (C. Liu).

https://doi.org/10.1016/j.acags.2024.100165
Received 21 September 2023; Received in revised form 10 March 2024; Accepted 15 April 2024
Available online 17 April 2024

reconstruction. They implemented end-to-end learning using a deep convolutional neural network (CNN) with LR images as input and SR images as output (SRCNN). Dong et al. (2016) improved the SRCNN network and put forward a fast SR model (FSRCNN). A very deep convolutional networks (VDSR) approach was soon put forth to accomplish super reconstruction (Kim et al., 2016a). However, a significant downside is that deeper layers increase the number of network parameters, which burdens computation and elevates the risk of vanishing/exploding gradients. Utilizing recurrent supervision and skip-connections, Kim et al. (2016b) employed deeply recursive convolutional networks (DRCN) to achieve SR reconstruction for the first time, avoiding gradient disappearance and gradient explosion while using a high-depth neural network. The SRResNet was developed over deep residual networks to renovate realistic textures from several LR images (Ledig et al., 2017). Tai et al. (2017) built a DRCN-based deep recurrent residual network (DRRN) by adding local residual learning that can leverage residual knowledge to make deep network training easier and implement recursive learning while increasing depth to optimize model parameters. In most of the mentioned studies, the high-frequency features are ignored. Ledig et al. (2017) enhanced the high-frequency detail of LR images by using generative adversarial network (GAN) technology (SRGAN). Wang et al. (2018) proposed an enhanced SRGAN (ESRGAN) using residual dense blocks as the base network without batch normalization (BN). ESRGAN outperforms SRGAN in texture realism, organicity and visual quality.

The CNN depth is critical for SRCNN-based image reconstruction. Lim et al. (2017) designed enhanced deep residual networks (EDSR) based on SRResNet, which have undergone rapid development. The EDSR network has 160 layers and removes several BN layers in SRResNet, which boosts training efficiency and speed. Zhang et al. (2018a) introduced the residual channel attention network (RCAN), a sophisticated 400-layer architecture that leverages residual and channel attention mechanisms. This design enables the network to concentrate on learning high-frequency information, optimizing its focus and efficiency. Haris et al. (2018) proposed deep back-projection networks (DBPN) by constructing an iterative up-sampling method. Different from image interpolation or nonlinear mapping of the above algorithms, DBPN focuses on different depths to improve the sampling rate of SR. Zhang et al. (2018b) developed a residual dense network (RDN) to collect local features from all convolutional operations using residual dense blocks and adopt a holistic approach to fuse global features. Guo et al. (2020) proposed a dual regression network (DRN), which constrains and balances the leading network by fusing shallow and deep features through jump connections mapped by a U-shaped network. Liu et al. (2020) deployed a residual feature aggregation framework (RFA-Net) via an enhanced spatial attention module to make the network focus on critical spatial characteristics. You et al. (2021) introduced a progressive growing GAN approach (PGGAN) for reconstructing 3D digital models of carbonate rocks. This innovative technique allows for the gradual enhancement of network complexity and resolution, resulting in the generation of high-fidelity 3D digital core images.

Most of the aforementioned deep learning-based SR reconstruction techniques might be made up of a shallow feature extraction module, a deep feature extraction module, an up-sampling module and a reconstruction module. However, this kind of technique usually only stacks a lot of residual convolution modules during the feature extraction process. Common challenges include networks being limited by a uniform receptive field due to the use of convolutions of the same size without down-sampling, resulting in the recovery of only same-level features. To overcome these limitations, integrating multi-scale information modeling into super-resolution networks is essential. For exploring the multi-scale SR, a new wave of academic research on multi-scale deep learning algorithms has begun. Gao and Zhuang (2019) presented multi-scale deep neural networks (MsDNN) to resolve the issue that the sampling factor is unknown when the data is input, which makes it challenging to train the super-score if the up-sampling and down-sampling are unclear. We have to admit that later ideas of ensuing multi-scale study are universally generated by corresponding diverse scales fusion. Sun et al. (2021) designed a quick and light framework, stating a better trade-off between reconstructed performance and computational efficiency by stacking residual blocks for multi-scale images. The multi-scale SR reconstruction algorithms mentioned above have become a significant source of inspiration in this study.

Currently, several studies have explored the use of deep learning to reconstruct clear digital rock images. Wang et al. (2020) employed SR techniques for compensating texture recovery, contributing to assessing digital rocks with a high fraction of LR microporous features. Yan et al. (2022) recently unveiled a learning strategy based on several lexicons. The homogeneous HR pore structure's multi-scale reconstruction algorithm generates a more accurate multi-scale pore structure model by introducing the edge mode and micro-pore mode of the homogeneous HR pore structure into the LR pore structure. Liao et al. (2022) verified the effectiveness of decoupled simulations with sandstone and carbonate rock samples, significantly accelerating the computational speed of digital rocks. Liu and Mukerji (2022) put forth a multi-scale imaging approach based on deep GANs to synthesize large-scale HR digital rocks. Shan et al. (2022a) deployed a SR reconstruction model (CA-SRResNet) using CNNs, attention mechanisms and residual technique to produce SR rock images. A GAN-based method (called RDCA-SRGAN) was proposed to enhance resolution of digital rock images by integrating residual blocks and dual-channel attention (Shan et al., 2022b). Cai et al. (2022) investigated the multi-scale theories of rock in unconventional oil and gas reservoirs through numerical simulations. These studies in digital rock motivate us to explore multi-scale digital rock reconstruction using deep learning. In our study, a digital rock CT image SR reconstruction algorithm based on residual U-Net and multi-scale fusion is presented. Our model introduces the following three innovations:1. Utilizing the residual U-Net algorithm, which combines the characteristics of U-shaped networks and residual connections, to enhance the reconstruction quality and accuracy of digital rock CT images. 2. Introducing a multi-scale fusion technique to effectively integrate information from different scales, improving the algorithm's ability to capture details and structures. 3. By leveraging both residual learning and multi-scale information, our model preserves fine features while better restoring the details and clarity of the original image. The remainder of this article is arranged as the following: our proposed multi-scale residual U-Net (MS-ResUnet) SR algorithm is detailed in Section II, along with the pertinent model concepts that are introduced. The experimental procedures and model-training procedures are described in Section III. Additionally, different methods are tested and compared to our proposed approach. Section IV summarizes the findings of this investigation.

2. Methodology

2.1. U-Net model

The multi-scale rock CT reconstruction proposed in this study is the extension of U-Net (Ronneberger et al., 2015). U-Net is an extremely symmetrical architecture made up of an encoder (left side of U-Net structure) and a decoder (right side of U-Net structure). The encoder extracts the image's features hierarchically, while the decoder recovers and reconstructs the SR image. Coordinating the encoder with the decoder does well in extracting high-frequency information. Numerous researchers improved the original U-Net and created several variations (Poudel et al., 2016; Roth et al., 2018; Zhou et al., 2018). Fig. 1 shows the basic U-Net architecture.

2.2. The proposed MS-ResUnet model

Fig. 2 illustrates our proposed framework MS-ResUnet. Be inspired by Lin et al. (2017), we integrate these blocks into multi-scale models. In

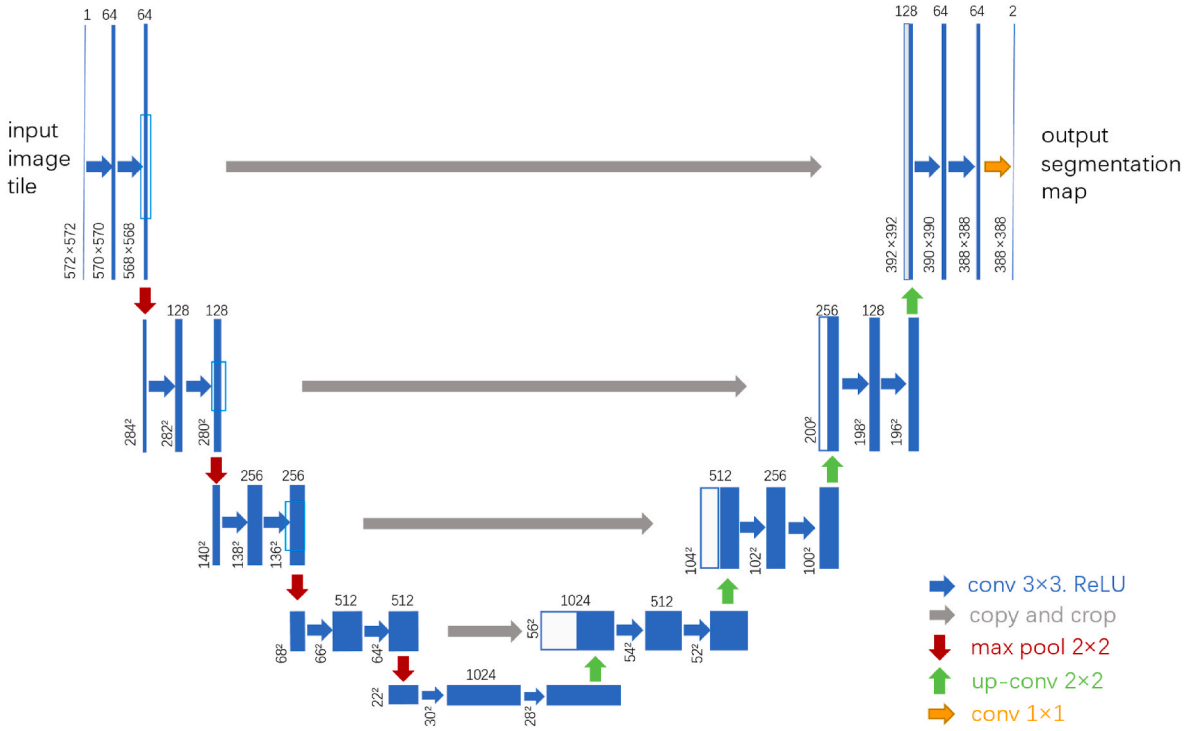

Fig. 1. Basic U-Net architecture.

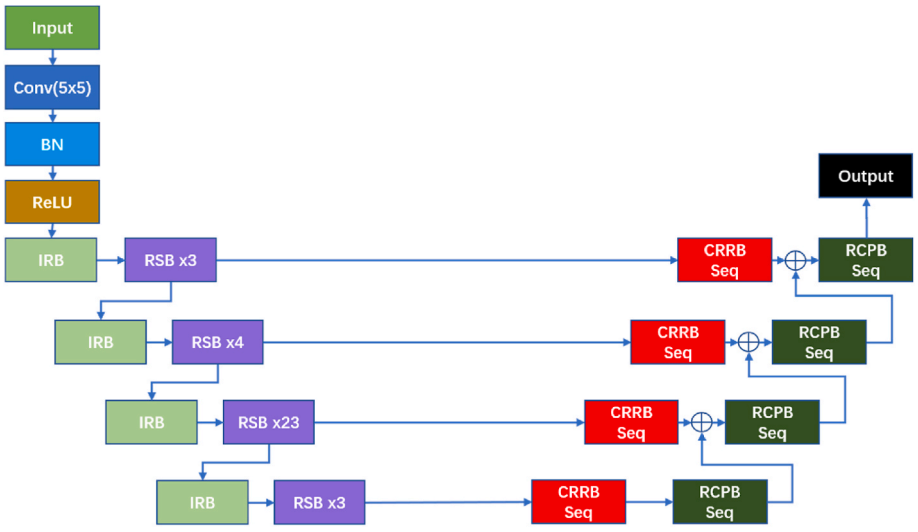

Fig. 2. Framework of MS-ResUnet.

particular, the image up-sampling, information interaction and recurrent fusion between multi-scale features of single image are carried out in the encoding phase. Using the fusion of these long-distance residual connections to transfer the underlying features can refine the rough high-level feature map. In this study, a 4-cascade structure is adopted to further realize the skip connection fusion of the encoding/decoding component of the same resolution image. Gradients can be immediately transferred to early convolutional layers thanks to this long-distance residual link, making end-to-end training effective.

To optimize the training stability, prevent over-fitting and speed up network convergence, the BN layer is utilized before the coding portion. Following that, improved residual block (IRB) and several residual sequential blocks (RSBs) are employed in each layer of the encoder. They are used to downscale images and extract key CT image elements. Convolutional ReLU residual blocks (CRRB) and residual chained pooling blocks (RCPB) make up the decoder. It will be combined with its below layer that was created at a higher scale. Each layer is made up of the fusion of up-sampling sub-levels and skip-connections of the encoding portion. To create HR semantic feature maps, high-level semantic data are combined with low-level detail features using these skip-connections. Up-sampling multi-scale fusion can be utilized in the decoding phase to connect the simultaneous multi-resolution picture sub-networks. After the multi-scale fusion deep network structure is realized, the nonlinear mapping relationship between the input and output images of the model can be simulated as much as possible, and then richer features can be extracted to generate SR images.

2.3. Encoding

1 Residual sequential block (RSB)

The RSB module, which is the main part of deep feature extraction, has two 1 × 1 convolutions that facilitate feature dimension changes and minimize parameter calculations. The first 1 × 1 convolution is capable of reducing the number of channels. To maintain the same number of channels as the original input in the network, the second 1× 1 convolution is used to adjust the channel numbers, ensuring that the output image matches the input image in terms of channels after the 3× 3 convolution. The RSB module is shown in Fig. 3.

2. Improved residual block (IRB)

Each layer has an IRB module configured for scale modification, or dimensionality reduction, being convenient for the output from the layer below. To achieve the dimension reduction and normalization of incoming images in the input stage, the IRB backbone structure is consistent with the main feature extraction part of RSB. Besides, IRB adds a 1 × 1 convolution block and a BN layer as a direct connection in the residual structure part. The structure of IRB is shown in Fig. 4. From Fig. 4, we can see that the data is loaded into the deep RSB module to carry out the encoding section of the data feature extraction phase following the IRB module's dimensionality reduction.

2.4. Decoding

1. Residual chained pooling block (RCPB)

The core CT image area serves as background context information with the proposed residual chain pooling block, which effectively aggregates image features from various window sizes.

The RCPB module comprises several chained pooling blocks, each of which has a max pooling layer and a 3 × 3 convolution layer. The max pooling layer reduces feature map sizes by selecting the maximum value in each 5 × 5 window. The output of the prior pooling block is fed to each subsequent chained pooling block. To retrieve the features of a vast area, each pooling block might reuse the feature information from the preceding pooling operation. Skip connections are employed to merge the output features from all pooling blocks with the input features, promoting gradient propagation throughout the training process. Note that the pooling blocks match up 3 × 3 convolutional layers to obtain the best result after abundant experimental tests. The convolution operation is equivalent to the weighting calculation, working for summation and fusion after each pooling process. The workflow of RCPB is shown in Fig. 5.

2. Convolutional ReLU residual block (CRRB)

Encouraged by a simplified version of the convolutional unit (He et al., 2016), a novel residual block is designed, namely CRRB. It modifies pre-trained weights before the network decodes the output, which consists of two ReLU and 3 × 3 convolutions, as well as the residual sum. In this way, under the same computing resources, more network layers can be stacked or more features can be extracted from each layer, to obtain better performance. A CRRB architecture is shown in Fig. 6.

2.5. Multi-scale fusion

Gradients can be propagated effectively in preceding layers because only linear transformations are implemented in these layers, which is necessary for further fusion processing. We consider multi-scale fusion as the summation of outputs from various residual connections, which is utilized specifically integrates feature maps with different shapes. It is necessary to ensure that the linear feature dimension is adaptive, which enables the feature dimension suitable for fusing the input from each decoding pipeline into an HR feature map. The flow chart of multi-scale fusion is drawn in Fig. 7. The module first employs convolution that acts on the input to produce feature maps; next, it expands all smaller feature maps to generate the same feature dimension; and finally, it fuses all feature maps. The input adaptation of this module aids in correctly tuning feature values along various paths. If there is only one input path, it will travel directly through the module without modification.

2.6. Loss function

The difference between the response estimator and the ground truth is measured by mean square error (MSE). For instance, t represents a population parameter θ derives from a sample estimate. The theoretical likelihood of $(\theta - t)^2$, called the MSE of the estimator t.

MSE is defined as the objective function during the MS-ResUnet network training to improve the features of the reconstructed image. The same dimension images are input. We need to obtain the squares sum of the difference between the predicted and ground truth.

Making a metric by adding the squares of the corresponding differences of the n elements, dividing the result by n, and then taking the MSE standard between the input x (output of the model prediction) and the target y:

$$loss(x, y) = 1 \Big/ N \sum_{i=1}^{n} (x_i - y_i)^2 \tag{1}$$

If the true value is y and the predicted value is $\hat{y}$, where $\hat{y}$ and y can both take on any shape and contain any number of elements, respectively, the loss function is expressed as follows:

$$loss(\hat{y}, y) = 1 \Big/ N \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \tag{2}$$

3. Experiments

We used a GPU server with Intel Core i9-9900KF@3.60GHz and two NVIDIA GeForce GTX 2080ti GPUs configured to 32 GB. All experiments were carried out on a 64-bit Windows10 OS equipped with CUDA Toolkit 10.2 and the Pytorch framework. The state-of-art methods including SRCNN, VDSR, FSRCNN, DRRN, EDSR, SRGAN, ESRGAN, RCAN and our MS-ResUnet model are trained and tested in our experiments.

3.1. Datasets

We used carbonate rock, sandstone and coal data from DeepRock-SR (Wang et al., 2019). The training set consists of 9600 HR rock CT images, while the validation set and testing set are each composed of 1200 images. Rock CT images are characterized by 500 × 500 unsegmented slices, a resolution of 2.7–25 μm, and no overlap between the testing set and training set.

Fig. 3. Residual sequential block (RSB).

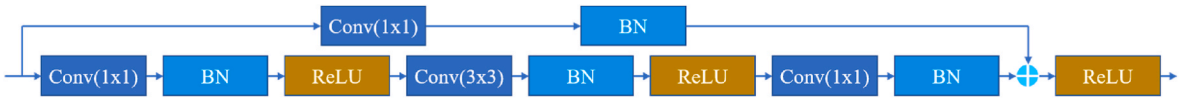

Fig. 4. Improved residual block (IRB).

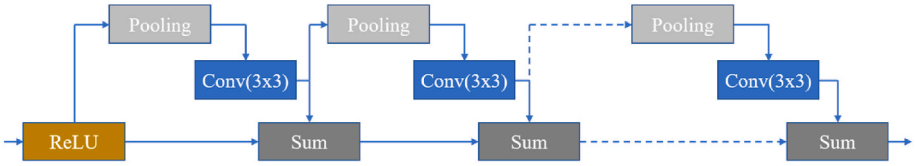

Fig. 5. Residual chained pooling block (RCPB).

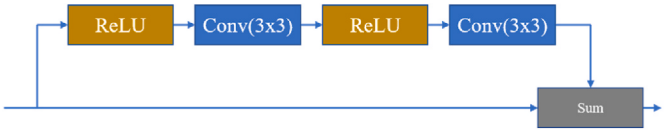

Fig. 6. Convolutional ReLU residual block (CRRB).

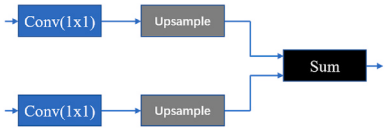

Fig. 7. Structure of multi-scale fusion.

The images are randomly cropped to 96 × 96 size for faster training and testing speed. Fig. 8 shows three clipped training sample. Fig. 9 displays testing images. The training images are augmented with 90-degree rotation, horizontal flips and a random crop. The original HR images serve as the target labels, which are then degraded to corresponding LR images. Notably, in this study, bicubic interpolation down-sampling is used. The initial HR image is scaled by a factor of 2, 4 and 8 to produce noisy and blurred LR pictures, depending on the different magnification scales. Before down-sampling, the input is considered as HR images, subsequently, the ones are changed to LR images. An efficient pair for later model training consists of LR and HR images. As a result, the LR images as the inputs of MS-ResUnet have the following pixel sizes: 48 × 48, 24 × 24 and 12 × 12.

3.2. Training setting

We trained our models using the Adam optimizer during the experiment, with initial learning rates of 1×10^{-4} and exponential decay rates of $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate is halved every 200 epochs throughout the 1000 training epochs.

3.3. Experimental results

1 Quantitative results

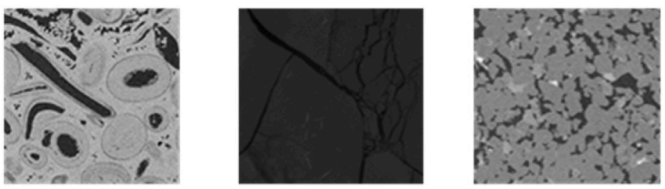

Fig. 8. Training set.

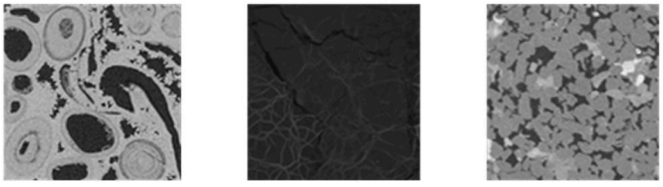

Fig. 9. Testing set.

The structural similarity index (SSIM) and the peak signal-to-noise ratio (PSNR) are two metrics used in the objective evaluation technique. The quality of image translation is frequently measured using PSNR, a full-reference image evaluation metric. Here is how MSE and PSNR are respectively defined in Eq. (3) and Eq. (4):

$$MSE = \sum_{i=0}^{H} \sum_{j=0}^{W} \frac{[\hat{f}(i,j) - f(i,j)]^2}{H \times W} \tag{3}$$

$$PSNR = 10\, log_{10} \frac{MAX_i^2}{MSE} \tag{4}$$

MSE has been referred to in section 2.6, which reacts to the differences between the HR image and the regenerated SR image, *MAX* denotes the highest degree of gray, and PSNR is measured in decibels (dB). The reconstructed CT picture is more faithful to the original image with a higher PSNR value.

Pixel accuracy is taken into account by PSNR, but visual aspects of human eyesight are ignored. Another measure, called SSIM, can describe the structural information of the image. Using the mean (μ), variance and covariance (σ) and stable variables (C). The following Eq. (5) is the SSIM formula:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{5}$$

The value range of SSIM is between 0 and 1. The SSIM is more representative of the real sample when it is nearer 1.

At scale factors of 2, 4 and 8, our MS-ResUnet model is compared with SRCNN, FSRCNN, EDSR, VDSR, DRRN, RCAN, SRGAN and ESRGAN. The testing data consists of three sets of 500 × 500 picture samples, representing images of sandstone, carbonate and coal, respectively. Table 1 demonstrates the average PSNR and SSIM for each benchmark approach. The highest PSNR and SSIM are set in bold. From Table 1, we observe that the deep network models and straightforward structure of SRCNN, FSRCNN, VDSR and DRRN result in lower PSNR and SSIM. The channel attention technique of RCAN also enables the model to acquire a higher evaluation value. EDSR optimizes the deep network topology to conserve space to enlarge the size of the model to boost the model's expressiveness. The PSNR values of the SRGAN and ESRGAN are insufficient, but the texture is more natural with higher SSIM value.

Table 1
Comparison PSNR and SSIM values.

Model	Scale	Average PSNR	Average SSIM
SRCNN	2	32.90	0.791
	4	31.27	0.790
	8	28.32	0.730
FSRCNN	2	33.25	0.800
	4	31.50	0.792
	8	28.40	0.732
VDSR	2	33.72	0.828
	4	32.13	0.814
	8	28.71	0.748
DRRN	2	33.82	0.831
	4	32.33	0.822
	8	28.91	0.750
SRGAN	2	34.11	0.895
	4	32.42	0.862
	8	29.25	0.804
EDSR	2	34.47	0.863
	4	32.98	0.849
	8	29.57	0.766
RCAN	2	34.46	0.888
	4	33.11	0.854
	8	29.68	0.775
ESRGAN	2	34.46	0.900
	4	32.50	**0.879**
	8	29.32	**0.829**
MS-ResUnet (Ours)	2	**36.11**	**0.933**
	4	**33.75**	0.810
	8	**29.84**	0.786

From Table 1, we can summarize that our MS-ResUnet model has a considerable improvement in structural similarity. The MS-ResUnet is capable of capturing the more textural details of rock CT images. It has the optimal PSNR value at all three magnification scales. As the scale increases, the transmission of information within the network may become more challenging in multi-scale fusion. This could result in ineffective transmission and utilization of crucial information at scales 4 and 8, impacting the quality of reconstruction as well as the SSIM scores.

2 Qualitative results

The outcomes of the experiments frequently match the human eye's intuitive reflection. When the scale factor is 2, it is difficult to see the small difference between the HR and SR images by the human eyes. When the scale factor is 8, the feature information is too hazy. For qualitative evaluation, a scale factor of 4 was selected. The HR images and SR images produced from SRCNN, FSRCNN, EDSR, VDSR, DRRN, RCAN, SRGAN, ESRGAN and MS-ResUnet are compared in Figs. 10–12.

As shown in Figs. 10–12, when comparisons of the SR images from the nine algorithms are presented with a scale factor of 4, The SR images generated by SRCNN and FSRCNN are distorted and blurred. The parameters of EDSR, VDSR, DRRN and RCAN are optimized by increasing the number of network layers, which improves the reconstruction effect of rock CT images. However, SRGAN and ESRGAN with the higher theoretical value of structural similarity have a better visual effect on the human eye and form a clearer image, but when the detail texture increases, the edge of the image is not so smooth.

In detail, algorithms such as SRCNN, VDSR, FSRCNN and DRRN are difficult to reconstruct rock fissure edges and particle characteristics in carbonate and sandstone. EDSR and RDCA models cannot extract intragranular information, but they have advantages when intragranular complex information is correlated. SRGAN and ESRGAN algorithms facilitate pixel-by-pixel matching of the original image. Among the nine models, the SRGAN, EDSR, RCAN and ESRGAN algorithms can better analyze the texture and high-frequency information of the reconstructed CT images, resulting in clearer and higher-quality SR images. However, in the enlarged image details, the edges are not sharp enough and the details are not realistic enough.

In contrast to them, the MS-ResUnet utilizes the multi-scale fusion of these long-distance residual connections to transfer low-level features for refining coarse high-level feature maps. The long-distance residual connection enables gradients to be directly propagated to early convolutional layers, which enables efficient end-to-end training. Simultaneously, MS-ResUnet generates richer texture features and less content loss, which are optimal for human vision.

3. Number of parameters and operations

The floating-point operations per second (flops) and number of parameters for all trained models are described in Table 2. Our MS-ResUnet model achieved comparable performance with EDSR, RCAN and ESRGAN with fewer parameters (7.45M). The MS-ResUnet has smaller flops than DRRN, EDSR, RCAN and ESRGAN. Our work incorporated the multi-scale idea into the SR domain, which captures the multi-scale features and reduces network depth, hence the better performance of the network while keeping the parameters and number of operations low.

3.4. Ablation study

1. Number of residual sequential block (RSB) modules

We compare and test the network performance utilizing various

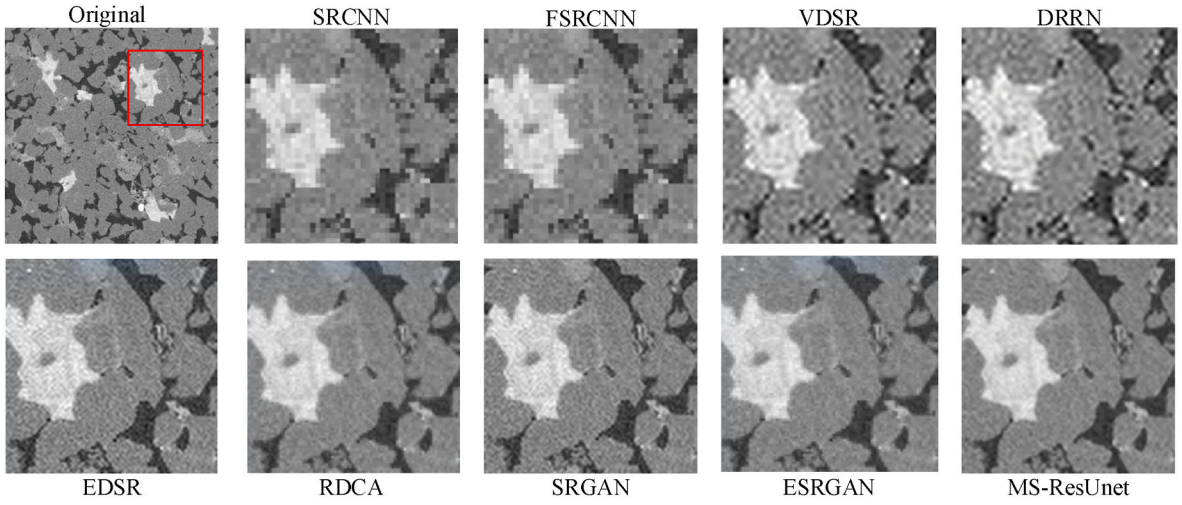

Fig. 10. Comparisons of reconstructed SR images on carbonate using nine models.

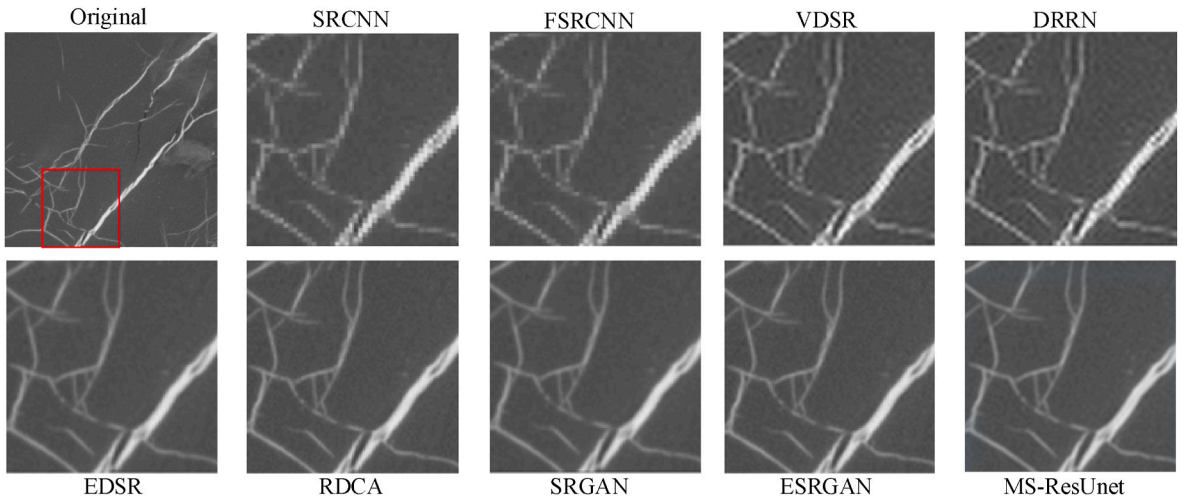

Fig. 11. Comparisons of reconstructed SR images on coal using nine models.

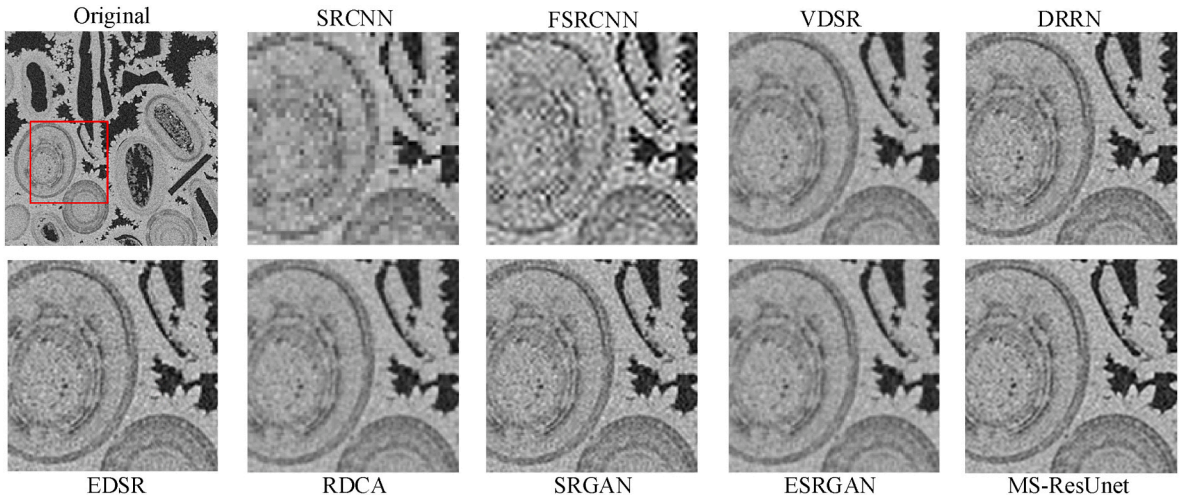

Fig. 12. Comparisons of reconstructed SR images on sandstone using nine models.

Table 2
Number of parameters and operations for nine models.

Model	Parameters	Flops
SRCNN	0.06M	2.87G
FSRCNN	0.01M	1.32G
VDSR	0.66M	33.35G
DRRN	0.30M	59.31G
EDSR	1.37M	68.99G
RCAN	15.44M	770.03G
SRGAN	5.91M	36.51G
ESRGAN	40.26M	904.39G
MS-ResUnet (ours)	7.45M	52.05G

numbers of RSB modules in the encoding stage before coming to the best experimental results. Performance evaluations were independently carried out to compare the impact of shallow versus deep network front ends. The deep network has a better overall impact than the shallow network. On the other hand, an excessively deep network cannot produce better outcomes and has a slower rate of convergence. Table 3 shows that the best PSNR and SSIM values for the 4-layer encoder were obtained with 3, 4, 23 and 3 RSB modules, respectively.

Table 3
Comparisons of PSNR and SSIM for different numbers of RSB modules.

RSB structure	Scale	Average PSNR	Average SSIM
[3, 4, 3, 4]	2	35.77	0.931
	4	33.38	0.806
	8	28.95	0.782
[3, 4, 4, 3]	2	35.21	0.931
	4	33.44	0.805
	8	28.14	0.783
[4, 3, 3, 4]	2	35.13	0.928
	4	33.59	0.805
	8	28.88	0.778
[3, 23, 23, 3]	2	35.90	0.932
	4	33.38	0.804
	8	28.14	0.778
[4, 23, 3, 23]	2	35.78	0.931
	4	33.41	0.805
	8	28.75	0.777
[23, 3, 23, 3]	2	36.03	0.932
	4	33.53	0.806
	8	28.91	0.783
[3, 4, 23, 3]	2	**36.11**	**0.933**
	4	**33.75**	**0.810**
	8	**29.84**	**0.786**

2. Loss analysis

Loss function calculates the pixel-level loss based on the one-to-one correspondence of pixel positions between two images. In an extreme scenario, if two identical images are slightly shifted by one pixel in their corresponding positions, the pixel-level loss function may yield relatively large values, even though the human eye may perceive minimal differences between the images. Common loss functions employed in deep learning-based SR include MSE, mean absolute error (MAE), perceptual loss, and gradient term loss.

MAE is a commonly utilized regression objective function. It is the total of the absolute differences between the anticipated value $\widehat{y}_j$ and the target value y_j. Regardless of the mistake's direction, it stands for the average error margin of the anticipated values. It is expressed as following Eq. (6):

$$loss(x,y)=\frac{1}{n}\sum_{i=1}^{n}|y_i-f(x_i)| \tag{6}$$

The semantic information contained in the image can be understood by the perceptual loss function. The image's feature information can be extracted using a VGG network, and the feature information is then used to calculate the loss. The expression for perceptual loss is as follows in Eq. (7):

$$\ell_{feat}^{\varphi,j}(\widehat{y},y)=\frac{1}{C_jH_jW_j}\left\|\varphi_j(\widehat{y})-\varphi_j(y)\right\|_2^2 \tag{7}$$

Gradient term loss is mainly used for image smoothing, ghost removal, denoising and restoration processing. The total variation of noise-contaminated images is significantly larger than that of noise-free images. Limiting the total variation suppresses noise. For image processing, total variation loss can smooth the image. The principle is that a signal with excessive and possibly spurious detail has a high total variation, that is, the integral of the absolute gradient of the signal is high. According to this principle, the total variation of the signal is reduced so that it closely matches the original signal. This removes unwanted details while preserving important details such as edges. The gradient term loss function is described in Eq. (8):

$$J_{T_0}(u)=\sum_{i,j}\left[\left(x_{i,j}-x_{i,j-1}\right)^2+\left(x_{i,j}x_{i+1,j}-x_{i+1,j}\right)^2\right]^{\frac{\beta}{2}} \tag{8}$$

To find the best loss function for our model, we employ MAE, MSE and the perceptual loss function. Table 4 shows the experimental findings. We can see that the MSE loss function produces the greatest PSNR and SSIM.

4. Discussion

In this study, we presented a novel SR algorithm using U-Net with multi-scale fusion and residual structure (MS-ResUnet). Experimental results show that MS-ResUnet achieves the highest average PSNR at all three magnification scales and the highest average SSIM at × 2 magnification. The comparison in Figs. 10–12 illustrates that our model accurately reconstructs the detailed pores in carbonate and sandstone, as well as clearly visible coal cracks. This approach offers a solution to address the limitations of traditional digital rock reconstruction methods and mitigate the blurring effects induced by hardware equipment and environmental factors. However, it is important to note that the current research is limited to two-dimensional digital rock core images and has not yet considered 3D volumetric rock core images (You et al., 2021). Future work will overcome this challenge and expand the application of digital rock cores to broader scenarios, better supporting research and practices in the geological and petroleum engineering fields.

Table 4
Comparison of MS-ResUet in PSNR and SSIM at different loss functions.

Loss function	Scale	Average PSNR	Average SSIM
Perceptual loss with gradient term	2	30.31	0.933
	4	28.45	0.807
	8	26.12	0.783
Perceptual loss without gradient term	2	31.47	0.933
	4	31.23	0.807
	8	26.59	0.783
MAE loss	2	36.05	0.932
	4	33.03	0.807
	8	28.75	0.783
MSE loss	2	**36.11**	**0.933**
	4	**33.75**	**0.809**
	8	**29.84**	**0.786**

5. Conclusion

HR digital rock core images are essential data sources for geological and petroleum engineering research, offering valuable insights for oil exploration and development. Nevertheless, constraints from physical devices and external conditions often result in the acquisition of LR images. These LR digital rock core images may fail to fully depict the intricate details and textural characteristics of the rock cores, potentially compromising the precision and dependability of geological and petroleum engineering investigations. To address this issue, we introduced a SR approach through multi-scale fusion using residual U-Net for reconstructing rock micro-CT images (MS-ResUnet). The proposed method was tested on image samples of sandstone, carbonate rock, and coal during the experiments. Through comparisons of various loss functions, module numbers, and depths, we identified the optimal network structure. Compared to eight methods, our MS-ResUnet excels in capturing fine details at image edges and restoring high-frequency features by integrating low-level detail traits with high-level pixel classification characteristics. The use of our MS-ResUnet framework results in reconstructed SR images that closely resemble HR images in terms of structure. This method explored in our study contributes to advancing rock physics within the multi-scale digital rock methodology, offering an efficient tool for characterizing unconventional reservoirs.

Funding

This research was funded by the National Science Foundation (NSF), Grant Nos. OIA-1946231 and CNS-1650551.

Code availability section

Name of the code/library: MS-ResUnet.
Contact: chengqian_liu0302@163.com or 19565295403.
Hardware requirements: GPU server with Intel Core i9-9900K F@3.60GHz and two NVIDIA GeForce GTX 2080ti.
GPUs configured to 32 GB.
Program language: Python.
Software required: 64-bit Windows10 OS equipped with CUDA Toolkit 10.2 and the Pytorch framework.
Program size: The floating-point operations per second (flops) is 7.45M and number of parameters is 52.05G.
The source codes are available for downloading at the link: https://github.com/forward-q/MS-ResUnet.

CRediT authorship contribution statement

Liqun Shan: Software, Methodology. **Chengqian Liu:** Writing – original draft, Writing – review & editing. **Yanchang Liu:** Validation, Supervision. **Yazhou Tu:** Visualization, Data curation. **Sai Venkatesh Chilukoti:** Resources. **Xiali Hei:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

Data availability

I have shared the link to my data/code at the Attach File step.

Acknowledgments

We gratefully acknowledge the helpful comments of the editor and anonymous reviewers.

References

Bizhani, M., Ardakani, O.H., Little, E., 2022. Reconstructing high fidelity digital rock images using deep convolutional neural networks. Sci. Rep. 12 (1), 1–14.
Bostanabad, R., Zhang, Y., Li, X., Kearney, T., Catherine Brinson, L., Apley, D.W., Liu, W. K., Chen, W., 2018. Computational microstructure characterization and reconstruction: review of the state-of-the-art techniques. Prog. Mater. Sci. 95, 1–41.
Bultreys, T., De Boever, W., Cnudde, V., 2016. Imaging and image-based fluid transport modeling at the pore scale in geological materials: a practical introduction to the current state-of-the-art. Earth Sci. Rev. 155, 93–128.
Cai, J., Zhao, L., Zhang, F., Wei, W., 2022. Advances in multiscale rock physics for unconventional reservoirs. Advances in Geo-Energy Research 6 (4), 271–275.
Dong, C., Loy, C.C., He, K., Tang, X., 2014. Learning a deep convolutional network for image super-resolution. In: Proc. ECCV. Zurich, Switzerland, pp. 6–12.
Dong, C., Loy, C.C., Tang, X., 2016. Accelerating the Super-resolution Convolutional Neural Network. Proc. ECCV, Amsterdam, The Netherlands, Oct., pp. 11–14
Feng, H., Wang, L., Li, Y., Du, A., 2022. LKASR: large kernel attention for lightweight image super-resolution. Knowl. Base Syst. 252, 109376.
Feng, J., Teng, Q., Li, B., He, X., Chen, H., Li, Y., 2020. An end-to-end three-dimensional reconstruction framework of porous media from a single two-dimensional image based on deep learning. Comput. Methods Appl. Mech. Eng. 368, 113043.
Fu, J., Cui, S., Cen, S., Li, C., 2021. Statistical characterization and reconstruction of heterogeneous microstructures using deep neural network. Comput. Methods Appl. Mech. Eng. 373, 113516.
Fu, J., Xiao, D., Li, D., Thomas, H.R., Li, C., 2022. Stochastic reconstruction of 3D microstructures from 2D cross-sectional images using machine learning-based characterization. Comput. Methods Appl. Mech. Eng. 390, 114532.
Gao, S., Zhuang, X., 2019. Multi-Scale Deep Neural Networks for Real Image Super-resolution. Proc. CVPR, Long Beach, CA, USA, pp. 2006–2013.
Gerke, K.M., Karsanina, M.V., Sizonenko, T.O., Miao, X., Gafurova, D.R., Korost, D.V., 2017. Multi-scale image fusion of X-ray microtomography and SEM data to model flow and transport properties for complex rocks on pore-level. In: Proc. SPE Russian Petroleum Technology Conference. Moscow, Russia.
Guo, Y., Chen, J., Wang, J., Chen, Q., Cao, J., Deng, Z., Xu, Y., Tan, M., 2020. Closed-loop matters: dual regression networks for single image super-resolution. In: Proc. CVPR. Seattle, WA, USA, pp. 5406–5415.
Haris, M., Shakhnarovich, G., Ukita, N., 2018. Deep Back-Projection Networks for Super-resolution. Proc. CVPR, pp. 1664–1673.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition. Proc. CVPR, Las Vegas, NV, USA, pp. 770–778.
Hou, P., Liang, X., Zhang, Y., He, J., Gao, F., Liu, J., 2021. 3D multi-scale reconstruction of fractured shale and influence of fracture morphology on shale gas flow. Nat. Resour. Res. 30 (3), 2463–2481.
Ju, Y., Zheng, J., Epstein, M., Sudak, L., Wang, J., Zhao, X., 2014. 3D numerical reconstruction of well-connected porous structure of rock using fractal algorithms. Comput. Methods Appl. Mech. Eng. 279, 212–226.
Ju, Y., Huang, Y., Zheng, J., Qian, X., Xie, H., Zhao, X., 2017. Multi-thread parallel algorithm for reconstructing 3D large-scale porous structures. Comput. Geosci. 101, 10–20.
Ju, Y., Huang, Y., Gong, W., Zheng, J., Xie, H., Wang, L., Qian, X., 2019. 3-D reconstruction method for complex pore structures of rocks using a small number of 2-D X-ray computed tomography images. IEEE Trans. Geosci. Rem. Sens. 57 (4), 1873–1882.
Karimpouli, S., Faraji, A., Balcewicz, M., Saenger, E.H., 2020. Computing heterogeneous core sample velocity using Digital Rock Physics: a multiscale approach. Comput. Geosci. 135, 104378.
Kim, J., Lee, K.J., Lee, M.K., 2016a. Accurate image super-resolution using very deep convolutional networks. In: Proc. CVPR. Las Vegas, USA, Jun, pp. 27–30.
Kim, J., Lee, J.K., Lee, K.M., 2016b. Deeply-Recursive Convolutional Network for Image Super-resolution. Proc. CVPR, Las Vegas, NV, USA, pp. 1637–1645.
Kuang, L., Liu, H., Ren, Y., Luo, K., Shi, M., Sun, J., Li, X., 2021. Application and development trend of artificial intelligence in petroleum exploration and development. Petrol. Explor. Dev. 48 (1), 1–14.
Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., Shi, W., 2017. Photo-realistic Single Image Super-resolution Using a Generative Adversarial Network. Proc. CVPR, pp. 4681–4690.
Li, X., Zhang, Y., Zhao, H., Burkhart, C., Brinson, L.C., Chen, W., 2018. A transfer learning approach for microstructure reconstruction and structure-property predictions. Sci. Rep. 8 (1), 1–13.
Liao, Q., Xue, L., Wang, B., Lei, G., 2022. A new upscaling method for microscopic fluid flow based on digital rocks. Advances in Geo-Energy Research 6 (4), 357–358.
Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M., 2017. Enhanced Deep Residual Networks for Single Image Super-resolution. Proc. CVPR, Honolulu, HI, USA, pp. 1132–1140.
Lin, G., Milan, A., Shen, C., Reid, I., 2017. Refinenet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation. Proc. CVPR, pp. 1925–1934.
Liu, J., Zhang, W., Tang, Y., Tang, J., Wu, G., 2020. Residual Feature Aggregation Network for Image Super-resolution. Proc. CVPR, Seattle, WA, USA, pp. 2356–2365.
Liu, M., Mukerji, T., 2022. Multiscale fusion of digital rock images based on deep generative adversarial networks. Geophys. Res. Lett. 49 (9).
Poudel, R.P.K., Lamata, P., Montana, G., 2016. Recurrent fully convolutional neural networks for multi-slice MRI cardiac segmentation. Reconstruction, Segmentation, and Analysis of Medical Images, pp. 83–94.
Roth, H.R., Shen, C., Oda, H., Sugino, T., Oda, M., Hayashi, Y., Misawa, Y., Mori, K., 2018. A Multi-Scale Pyramid of 3D Fully Convolutional Networks for Abdominal Multi-Organ Segmentation. Proc. MICCAI, pp. 417–425.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention 234–241.
Sanematsu, P.C., Thompson, K.E., Willson, C.S., 2019. Pore-scale modeling of nanoparticle transport and retention in real porous materials. Comput. Geosci. 127, 65–74.
Schlüter, S., Sheppard, A., Brown, K., Wildenschild, D., 2014. Image processing of multiphase images obtained via X-ray microtomography: a review. Water Resour. Res. 50 (4), 3615–3639.
Shan, L., Bai, X., Liu, C., Feng, Y., Liu, Y., Qi, Y., 2022a. Super-resolution reconstruction of digital rock CT images based on residual attention mechanism. Advances in Geo-Energy Research 6 (2), 157–168.
Shan, L., Liu, C., Liu, Y., Kong, W., Hei, X., 2022b. Rock CT image super-resolution using residual dual-channel attention generative adversarial network. Energies 15 (14), 5115.
Sun, L., Liu, Z., Sun, X., Liu, L., Lan, R., Luo, X., 2021. Lightweight image super-resolution via weighted multi-scale residual network. IEEE/CAA Journal of Automatica Sinica 8 (7), 1271–1280.
Tian, J., Qi, C., Sun, Y., Yaseen, Z.M., 2020. Surrogate permeability modelling of low-permeable rocks using convolutional neural networks. Comput. Methods Appl. Mech. Eng. 366, 113103.
Tai, Y., Yang, J., Liu, X., 2017. Image Super-resolution via Deep Recursive Residual Network. Proc. CVPR, pp. 3147–3155.
Tan, M., Su, M., Liu, W., 2021. Digital core construction of fractured carbonate rocks and pore-scale analysis of acoustic properties. J. Petrol. Sci. Eng. 196, 107771.
Tao, X., Gao, H., Liao, R., Wang, J., Jia, J., 2017. Detail-revealing deep video super-resolution. In: Proc. ICCV. Venice, Italy, pp. 4482–4490.
Wang, Y.D., Armstrong, R.T., Mostaghimi, P., 2020. Boosting resolution and recovering texture of 2D and 3D micro-CT images with deep learning. Water Resour. Res. 56 (1).
Wang, Y.D., Armstrong, R.T., Mostaghimi, P., 2019. Diverse Super Resolution Dataset of Digital Rocks (DeepRock-SR): Sandstone, Carbonate, and Coal. National Science Foundation, Alexandria, VA, USA.
Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Loy, C.C., Qiao, Y., Tang, X., 2018. Esrgan: enhanced super-resolution generative adversarial networks. In: Proc. ECCV Workshops, Munich, Germany, September 8-14, pp. 63–79.
Wang, Z., Chen, J., Hoi, S.C., 2020. Deep learning for image super-resolution: a survey. IEEE Trans. Pattern Anal. Mach. Intell. 43 (10), 3365–3387.
Yan, P., Teng, Q., He, X., Ma, Z., Zhang, N., 2022. Multiscale reconstruction of porous media based on multiple dictionaries learning. Comput. Geosci. 176, 105356.
You, N., Li, Y.E., Cheng, A.C., 2021. 3D carbonate digital rock reconstruction using progressive growing GAN. J. Geophys. Res. Solid Earth 126.
Zhang, L., Jing, W., Yang, Y., Yang, H., Guo, Y., Sun, H., Zhao, J., Yao, J., 2019. The investigation of permeability calculation using digital core simulation technology. Energies 12 (17), 3273.
Zhang, Y., Li, K.P., Li, K., Wang, L., Zhong, B., Fu, Y., 2018a. Image Super-resolution Using Very Deep Residual Channel Attention Networks. Proc. ECCV, Munich, Germany, pp. 8–14 September.
Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y., 2018b. Residual Dense Network for Image Super-resolution. Proc. CVPR, pp. 2472–2481.
Zhou, Y., Wu, G., Fu, Y., Li, K., Liu, Y., 2021. Cross-MPI: Cross-Scale Stereo for Image Super-resolution Using Multiplane Images. Proc. CVPR, pp. 14842–14851.
Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J., 2018. Unet++: a nested u-net architecture for medical image segmentation. In: Proc. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 3–11.
Zhu, L., Zhang, C., Zhang, C.M., Zhou, X., Zhang, Z., Nie, X., Liu, W., Zhu, B., 2019. Challenges and prospects of digital core-reconstruction research. Geofluids 2019 (2), 1–29.