

Original article

Super-resolution reconstruction of digital rock CT images based on residual attention mechanism

Liqun Shan^{1,2}, Xueyuan Bai¹, Chengqian Liu¹, Yin Feng², Yanchang Liu¹*, Yanyan Qi³

¹School of physics and electronic engineering, Northeast Petroleum University, Daqing 163318, P. R. China

²Petroleum engineering department, University of Louisiana at Lafayette, Lafayette 70504, USA

³Xiangyang Branch of China Telecom Co. LTD, Xiangyang 441011, P. R. China

Keywords:

Digital rock
super resolution
reconstruction
residual network
attentional mechanism

Cited as:

Shan, L., Bai, X., Liu, C., Feng, Y., Liu, Y., Qi, Y. Super-resolution reconstruction of digital rock CT images based on residual attention mechanism. *Advances in Geo-Energy Research*, 2022, 6(2): 157-168.

<https://doi.org/10.46690/ager.2022.02.07>

Abstract:

Computer tomography technology is widely used in geological exploration because it is a nondestructive and three-dimensional imaging method that can be integrated with computer simulation. However, the large-scale application of the computer tomography technique is limited by economic costs and time consumption. Therefore, it is challenging and intractable to indicate the pore structure characteristics of rock. To address this issue, a super-resolution reconstruction algorithm based on convolutional neural networks, residual learning, and attention mechanism was proposed to generate super-resolution images in this study. This algorithm was applied to the reconstruction of carbonate rock and sandstone. The performance of two-dimensional image reconstruction was evaluated by quantitative extraction and qualitative visualization. The results from experiments indicate that the built model performs well on different upscaling factors and is superior to the existing super-resolution approaches based on convolutional neural network.

1. Introduction

Using digital rock technology to study the physical properties of reservoir rocks, pore-throat network structure, and the micro-seepage mechanism of fluid in porous media provides a new perspective for the accurate characterization of rock pores (Lin et al., 2018; Ji et al., 2018). Digital rock technology can establish three-dimensional (3D) rock data volume by mathematics and computer technology, which can be used for rock numerical mini-calculation and simulation (Andhroumoudine et al., 2021). Among them, computer tomography (CT) technology plays the most important role in the field of studying digital rocks. CT technology is a nondestructive and 3D imaging method, which is considered to be the most effective way to obtain the rich internal structure of the rock (Liu et al., 2017; Engelmann and Lessmann, 2021). However, owing to the inherent limitations of devices, the difference of

the external environment, the choice of the image degradation model, the network transmission medium, and the broadband, collected rock images by CT technology are destructed to some degrees. Meanwhile, it is impossible to directly obtain ideal high-resolution (HR) images with clear details, textures, and sharp edge information (Shi et al., 2016).

Super-resolution (SR) methods aim to generate HR images with rich structure information from low-resolution (LR) images (Zeng et al., 2021). Images resolution can be converted from low to high through algorithms such as digital image processing and computer vision. At present, there are four kinds of image SR algorithms: interpolation-based, reconstruction-based, shallow-level learning-based, and deep learning-based algorithms. Among them, the first three methods belong to frequency domain algorithms, which have dominated the reconstruction methods in the past few decades. Deep learning belongs to the spatial reconstruction algorithm. Due to the

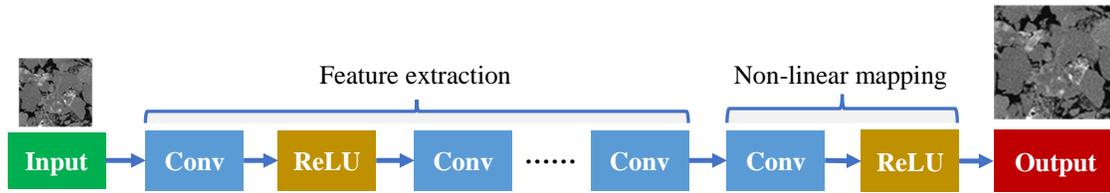


Fig. 1. Structure diagram of CNN.

inherent limitations of frequency-domain algorithms, deep learning-based algorithms have become the mainstream technology in the field of image reconstruction (Wang et al., 2020b).

In order to obtain HR images, many scholars have devoted themselves to proposing a large number of practical algorithms. Dong et al. (2015) first introduced the convolutional neural network (CNN) to the image SR and proposed an SR convolutional neural network. Then, Dong et al. (2016) optimized the network based on the Super Resolution Convolutional Neural Network (SRCNN) and proposed the fast SR convolution neural network model. Shi et al. (2016) proposed a network model based on sub-pixel convolutional layer convolution using an efficient sub-pixel convolution neural network, which shows a significant promotion of the reconstructed image quality, but the network structure is still too rudimentary. Zuo et al. (2021) constructed a mixed deep convolutional network to shrink the LR image to the specified size in upsampling phase and extract features from LR images. The extracted initial features are fed into a encode-decode structure in high-dimensions for capturing SR features. Qiu et al. (2021) employed multiple improved residual networks to improve the reconstruction performance. However, deeper networks for image SR are more difficult to train, and equally treating low-frequency information across channels hinders the representational ability of SRCNNs.

Currently, a great many scholars replace the above network weaknesses with more efficient algorithms and these changes are applied in the field of geo-energy research (Shan et al., 2018; Yang et al., 2021; Liu et al., 2022). Wang et al. (2021) constructed a dual closed-loop learning structure with a primary face super-resolution network to generate additional prior constraints of the primary branch for guiding the essential facial features reconstruction. Theoretically, this can further improve the model training efficiency and reduce parameters without damaging the model performance. However, the approach still has a very large space of the possible mappings, which makes it difficult to learn a good solution. Liu et al. (2020) presented a residual feature aggregation network to aggregate multiple residual blocks and make full use of the hierarchical features learned. But its training promotion outcome is inapparent. Geng et al. (2021) introduced shearlet transform into a deep residual learning network to extract high-frequency details of images. Chen et al. (2022) employed attention augmented multi-scale residual blocks in CNN, which increases the proportion of useful information. Compared with convolution operations, attention mechanism focuses more on the entirety. It simulates the human habit of observing the

environment, which only pays attention to the most important parts and ignores unnecessary circumstances. It was originally used in machine translation (Popel et al., 2020). However, it is not a hot-point issue until Google proposed “attention is all you need” (Vaswani et al., 2017), which provides a new idea to completely replace the traditional CNN and Recurrent Neural Network structure with the attention mechanism. After that, there are a large number of attention mechanism applications (Song et al., 2022).

To date, deep learning-based SR models have caught the attention of researchers in the oil and gas field (Zha et al., 2020). Wang et al. (2020a) developed enhanced deep super-resolution generative adversarial networks to generate SR images for sandstone, coal, and carbonate samples. Their results show that SRCNN restores large-scale edge features, while the method regenerates perceptually indistinguishable high-frequency textures and shows excellent visual similarity in texture regeneration. To the authors’ best knowledge, the existing studies in the context of digital rock imaging haven’t introduced attention mechanism for the generation of SR rock CT images. In this study, a convolutional neural network integrated with residual learning and attention mechanism is designed to perform digital rock SR image reconstruction. The remainder of this paper is organized as follows. Section 2 introduces the related model concept of deep learning and describes the proposed residual channel attention mechanism super-resolution reconstruction network (CA-SRResNet) in detail. Section 3 outlines the experimental steps and model training process. Further, it presents tests and comparisons of our CA-SRResNet network with other methods. Section 4 discusses future research directions. Finally, Section 5 presents the conclusions of this study.

2. Methodology

2.1 Super-resolution CNN

The structure diagram of the improved SRCNN is shown in Fig. 1. The SRCNN model learns the resolution mapping relationship between LR and HR images by a 22-layer CNN to form an SRCNN network for the digital rock reconstruction (Wang et al., 2019). Based on the relationship between deep learning and traditional sparse coding, the SRCNN network first uses bicubic interpolation to construct LR images into a specified size, which is transmitted to the network, and then the network is divided by multi-layer convolution for image block extraction, feature nonlinear mapping, and final reconstruction.

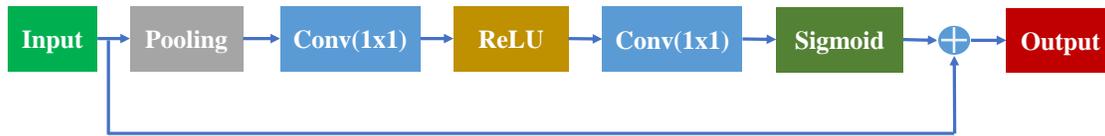


Fig. 2. Channel attention mechanism.

2.2 Attention mechanism

In the SRCNN, image SR reconstruction is realized by combining input information with the weights of the training model to efficiently reconstruct images with high quality. However, in the process of feature learning, only nonlinear mapping, such as convolution operation and activation function, cannot make full use of high-frequency information, which makes it impossible to acquire detailed images. Attention mechanism can search and transfer high-frequency information. Adding attention mechanism to the network can make better use of high-frequency information, improve the network's ability to extract high-frequency information, and make a significant performance improvement for image recognition. Soft attention mechanism and hard attention mechanism are widely used to calculate attention in recent years. In this research, soft attention mechanism is used to improve reconstruction accuracy and reduce training time.

Soft attention mechanism is mainly used in channel attention mechanism and spatial attention mechanism. Zhang et al. (2018) proposed the concept of channel attention, which is used to capture the correlation among the feature maps of different channels of the model. The model with channel attention can automatically learn the importance of each feature channel and assign different weight coefficients to each feature channel. If it is important to feature information, it will be assigned more weight; if not, weight is less. In the CNN, three channels (R, G, B) are used to represent each input color pixel. After passing through different convolution kernels, each channel will generate new information. The convolution kernel information is decomposed into multiple channel information components. The attention mechanism adds each weight value learned to the corresponding channel information. The architecture of the channel attention mechanism is shown in Fig. 2.

2.3 Residual network

The traditional CNN can extract the characteristics of low-frequency information, intermediate-frequency information, and high-frequency information of the data. The more layers of the network, the richer the features that can be extracted (Wang et al., 2020b). Moreover, as the number of network layers deepens, the more abstract the extracted features, the richer the semantic information. However, when the number of layers is increased greatly, it is difficult for the network to converge, information of the input image feature is lost in the deeper convolution process, resulting in fuzzy images. If only the number of network layers is increased, gradient explosion will be a tricky problem (Khan et al., 2020). If regularization initialization and batch normalization is carried out, a multi-

layer network can be trained, but the network will degrade. In order to solve the above issues, He et al. (2016) proposed a residual network, making it effective to build model. To address the problem of gradient dispersion in deep networks, the residual learning technique is employed to the traditional CNN, which can achieve a trade-off between the accuracy and speed while the network can get deeper and deeper.

Shallow neural network models can steadily train an identity mapping function $y = f(x)$. Using this feature function to represent the original x information, the function value y mapped during the learning process will produce different errors. As the number of network layers increases, the error will gradually accumulate. In the process of backpropagation, the network will appear under-fitting, and the gradient will become more and more divergent. With the addition of the batch normalization (BN) layer, it is difficult for the network to learn completely when the depth of the network model is large enough. A direct connection channel in the network is added to solve the above problem. Residual learning changes the mapping relationship of each layer in the network, adding the output value of each layer to the original input value as the input value of the next layer, namely $y = f(x) + x$, as shown in Fig. 3. The introduction of residual learning into the traditional CNN not only avoids the problem of gradient dispersion in the deep network but also addresses the issue of reduced accuracy (Zhang et al., 2020). On the one hand, residual learning improves accuracy; on the other hand, it speeds up. The input data of the residual block needs to go through two paths, one path is a straight path, through multiple convolutional layers for feature extraction and output feature matrix; another one is a shortcut path, directly adding the characteristic matrix output of the input data to the characteristic matrix of the straight path, passing the added result into an activation function, and finally outputting the result. However, due to the need for matrix addition, it is necessary to make the attributes of the feature matrix output by the straight path and the feature matrix output by the shortcut path the same.

The residual block used in this study is shown in Fig. 3(b). Each layer contains a convolutional layer with a convolution kernel size of 3 and a stride size of 2, the number of channels is 64, and the activation function ELU. Since BN normalizes the features and eliminates the flexibility of the network, the first BN layer is deleted from the original residual block shown in Fig. 3(a), and the second BN layer is retained, which not only greatly improves the reconstruction effect, and it saves a lot of memory and reduces the complexity of calculation during the training of the network.

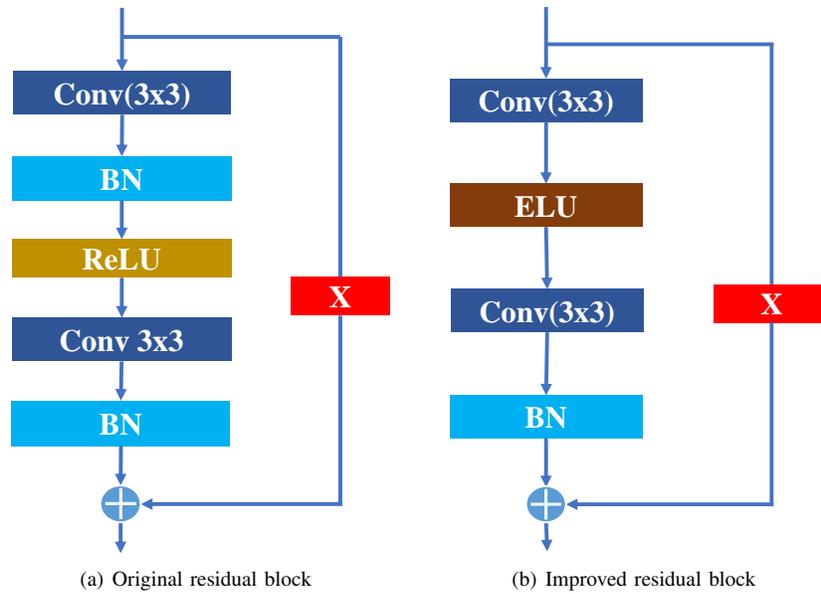


Fig. 3. Residual learning.

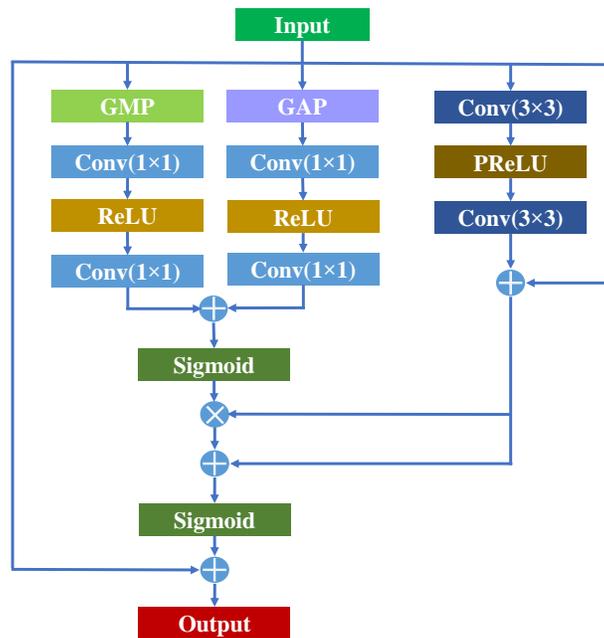


Fig. 4. Residual channel attention module.

2.4 Super-resolution network based on residual attention mechanism

2.4.1 Residual attention module

The introduction of the attention mechanism is of great help to the recovery of image details and textures. However, embedding the attention module directly in the backbone network will greatly weaken the network’s ability to extract features. Based on residual learning, the sum of the input of the network and the weighted feature of the attention mechanism acts as the output of the network. Not only solves the above problems, but also enhances the ability of feature extraction.

The BN layer directly normalizes each batch feature of the input information and restores the original input by stretching, scaling, and transformation. It not only can accelerate the model convergence but also has a certain regularization effect, which can solve the problem of gradient explosion during the training process of the model. However, during the training process, the BN layer destroys the state of the original information. After the image passes through the BN layer, the color, contrast, and brightness will be normalized, which affects the image reconstruction efficiency and makes the model more complicated. Therefore, it is desirable to devise a new block to address this issue. In this study, the attention

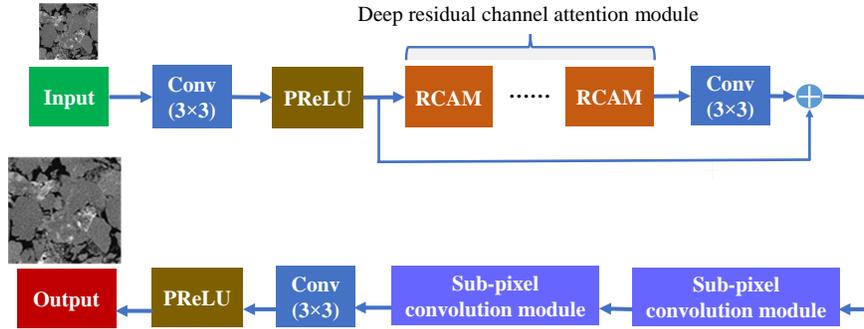


Fig. 5. Super-resolution reconstruction based on residual channel attention.

mechanism unit is introduced to the residual block, and the residual channel attention module is constructed, as shown in Fig. 4. Both convolution kernel and linear rectification function (ReLU) activation layers are used as residual blocks to replace the BN layers. The attention unit is accessed by skip links, but partial skip connections are still added to spread the input information through mapping, which solves the problem of the disappearance of features caused by the addition of the attention mechanism. The dimensionality reduction of the input image can effectively calculate the attention mechanism. For the aggregation of spatial information, the method used most often is average pooling. Some researchers believe that maximum pooling is as convenient as average pooling. In this paper, the average pooling and maximum pooling features were used together to acquire more specific high-frequency information.

2.4.2 Sub-pixel convolution module

After the deep residual module, two sub-pixel convolution modules are added. In the SR algorithm, common methods for expanding the size include direct up-sampling, bilinear interpolation, and deconvolution. In this study, the sub-pixel convolution method proposed in the efficient sub-pixel convolution neural network algorithm was used to better fit the relationship between pixels (Shi et al., 2016). Sub-pixel convolution is an ingenious image and feature map enlargement method, also called pixel shuffle. The sub-pixel convolution module first expands the number of channels of the original feature map through convolution. The number of expanded channels is determined by the enlarged size, which is the square of the enlarged size. After the feature map is convolved and arranged in a specific format, a large image can be obtained, which is the so-called pixel cleaning. Through pixel cleaning, the number of feature channels is restored to the original input size, but the size of each feature map becomes larger.

2.4.3 The proposed network model

A new CNN architecture based on the residual learning network and channel attention mechanism was proposed, as shown in Fig. 5. The residual channel attention module is introduced into the feature learning layer to accelerate the model's feature extraction. In this paper, 16 residual channel attention modules, a 3×3 convolutional layer, and a randomized parametric ReLU (PReLU) function pair are

used. To extract shallow the input image feature extraction, multiple residual channel attention modules are connected in series to transmit information through a global jump, which saves important shallow feature sets and improves gradient scaling. Meanwhile, the deconvolution layer at the end of the network has been replaced by sub-pixel convolution, which can capture richer pixels' information. 3×3 convolution layers are used to enlarge the SR images, making it possible to image reconstruction and high-dimensional feature restoration. The attention module is added to the side branch, which means, the dimension of the last channel passing through the attention unit must be consistent with the number of channels of the convolutional layer of the backbone network. The architecture of CA-SRResNet is shown in Fig. 5.

2.4.4 Loss function

In the process of reconstructing SR images, the most considered loss algorithm is mean square error (MSE), but the MSE loss function will cause the reconstructed image to be too smooth and lack realism. In order to break through the limitations of being excessively smooth, the VGG19 model is introduced to transform the CT image into the deep feature space. The features are extracted from the original HR images and reconstructed SR images by the VGG19 model and are expressed as ϕ_{SR_i} and ϕ_{HR_i} , respectively. The perceptual losses are calculated from ϕ_{SR_i} and ϕ_{HR_i} according to the mathematical equation expressed as Eq. (1).

$$L_V = \sum_{i=0}^H \sum_{j=0}^W \frac{(\phi_{SR_i} - \phi_{HR_i})^2}{H \times W} \quad (1)$$

where L_V denotes the two-dimensional (2D)-image loss of VGG19 model, SR_i is the SR reconstructed image, HR_i represents the original HR image, H and W represent the height and width of the image, respectively. The model loss L_1 is expressed as Eq. (2):

$$L_1 = \sum_{i=0}^H \sum_{j=0}^W \frac{|\phi_{SR_i} - \phi_{HR_i}|}{H \times W} \quad (2)$$

The perceptual loss and the model loss are combined to obtain the total loss. The mathematical expression is as in Eq. (3):

$$L = L_1 + \alpha L_V \quad (3)$$

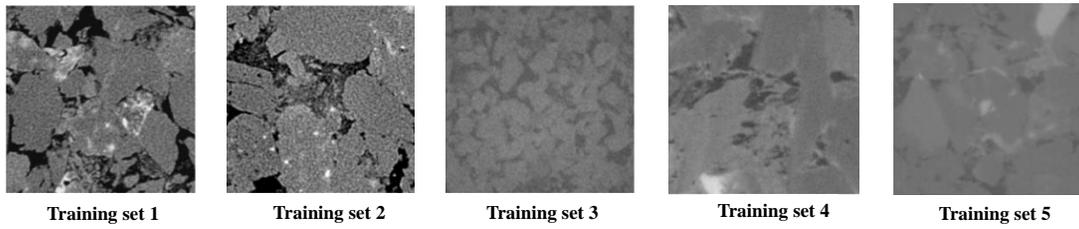


Fig. 6. Training samples.

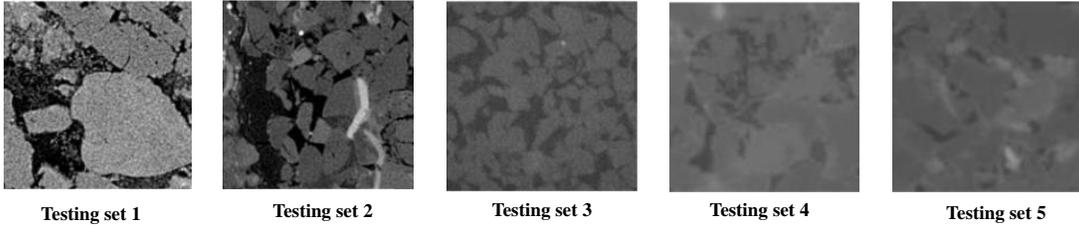


Fig. 7. Testing samples.

Table 1. Parameters of CA-SRResNet network.

Dataset parameters	Crop size	96×96
	Kernel size	3×3
Model parameters	Activation function	ELU
	Number of middle channels	64
	Residual channel attention module	16
	Training epochs	1000
Learning parameters	Initial learning rate	10 ⁻⁴
	Loss function	L ₁ loss+perceptual loss

where α represents the scaling factor. In this paper, $\alpha = e^{-3}$.

3. Experiment

3.1 Experimental setup

The experimental hardware environment is Intel Core i9-9900KF@3.60GHz, equipped with two NVIDIA GeForce RTX 2080Ti GPUs, configured to 32 GB. The software environment consists of a 64-bit Windows 10 operating system, CUDA Toolkit 10.2, and the Pytorch 2.0 framework.

The used dataset in this study is from the research proposed by Wang et al. (2019). Five sets of original CT images of rock are used as training CT samples. Samples (1), (2), and (3) are sandstone with a resolution of 3.8 μm . Sample (4) is carbonate rock with a resolution of 1.07 μm . Sample (5) is sandstone with a resolution of 1.07 μm . Using 2000 HR CT images of the rock. Among them, 1000 pictures constitute the training set, including sandstone and carbonate. Each rock type is represented by 400 images, which are used for uniformly distributed rock geometry and training network model parameters; another 1000 pictures in testing sets also include five rock types. In the process of evaluation, each rock type is represented by 400 images to verify the quality

and generalization ability of the network model. The images included in the training set and testing set are shown in Figs. 6 and 7.

In order to generate labels for input pictures, HR images will be degraded to generate corresponding LR images. According to different magnification scales, this study uses bicubic interpolation down-sampling to perform $\times 2$ down-sampling, $\times 4$ down-sampling, and $\times 8$ down-sampling to generate LR images. The HR and LR images form an effective image pair for later model training. To reduce the training time, the pictures in the training set are randomly cropped, and the LR pictures are cropped to 96×96 pixels.

3.2 Model training setting

In this study, Pytorch 2.0 deep learning platform is used to conduct the training and testing process. In experiments, the used parameters are listed in Table 1. The images were cropped into 96×96. Although the training and testing data sets are divided into small blocks with 96×96, the trained convolutional network can be used for any image of any size without cropping. The size of each other convolution kernel is 3×3. In the residual channel attention module, the number of

cascaded modules is 16, which not only avoids the problem of introducing too many training parameters but also ensures rich structure information with high quality. Adam optimizer is chosen to optimize the network performance. The initial learning rate is set to 1×10^{-4} . The exponential decay rate β_1 is set to 0.9, and β_2 is set to 0.999.

3.3 Experimental results

Both qualitative and quantitative assessments are conducted to verify the proposed method. During qualitative comparisons of CA-SRResNet, SRResNet, SRCNN, and Bicubic interpolation process, testing data are composed of five sets of samples, each of which has a resolution of 400×400 pixels. The resolution of the reconstructed image is 400×400 pixels. In addition, we introduce quantitative evaluation by the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), which are widely used to assess image quality.

PSNR is one of the most popular reconstruction quality measurements of lossy transformation. For image SR, PSNR is defined via the maximum pixel value and the MSE between HR and SR images. Given the HR image f with $H \times W$ pixels and the reconstruction SR image \hat{f} with $H \times W$ pixels, the PSNR between f and \hat{f} are defined as follows:

$$M = \sum_{i=0}^H \sum_{j=0}^W \frac{[\hat{f}(i, j) - f(i, j)]^2}{H \times W} \quad (4)$$

$$P = 10 \lg \frac{MAX_i^2}{M} \quad (5)$$

where M denotes the MSE between SR and HR images, P represents the PSNR between SR and HR images, and MAX_i is the maximum possible pixel value of the image. The higher the value of PSNR, the better the reconstructed CT image. PSNR is based on the error between pixels, but it does not take into account the visual characteristics of the human vision. SSIM can characterize the structural information of the image from many aspects such as brightness, contrast, and structure. For two images x and y , SSIM is expressed as:

$$S(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_x\sigma_y + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (6)$$

where $S(x, y)$ denotes the SSIM of two images x and y , μ_x is the mean of x , μ_y is the mean of y , μ_x^2 is the variance of x , μ_y^2 is the variance of y , $\sigma_x\sigma_y$ is the covariance of x and y , C_1 and C_2 are constants used to maintain stability. C_1 and C_2 are defined as:

$$C_1 = (k_1L)^2 \quad (7)$$

$$C_2 = (k_2L)^2 \quad (8)$$

where L is the dynamic range of pixel values, $k_1 = 0.01$, and $k_2 = 0.03$.

The value range of SSIM is [0,1]. The closer the SSIM value is to 1, the closer the SR image is to the real HR image.

The results are compared with the traditional bicubic interpolation method, SRCNN and SRResNet at scale factors of 2, 4, and 8. Note that SRResNet is obtained by introducing the residual learning into the SRCNN. The comparison results of

PSNR and SSIM are shown in Tables 2 and 3. The PSNR and SSIM values in the table are the average of all image pixels. As can be seen from the tables, the PSNR and SSIM values obtained by CA-SRResNet are the highest, which indicates that the performance of CA-SRResNet is the best compared to other algorithms. At the same time, it can be seen that the PSNR and SSIM values obtained by the bicubic interpolation method are the lowest, which are quite different from other algorithms. In terms of PSNR analysis, when the magnification scale is 2, compared with bicubic interpolation, the PSNR obtained by CA-SRResNet is increased by 16.523 dB, 18.618 dB, 6.051 dB, 5.991 dB, and 4.388 dB, respectively; compared with SRCNN, it is increased by 11.920 dB, 6.565 dB, 1.732 dB, 5.309 dB, 4.355 dB, respectively. compared with SRResNet, it is increased by 1.006 dB, 2.173 dB, 0.169 dB, 0.307 dB, 0.613 dB, respectively. The best values of PSNR and SSIM for each testing set in Tables 2 and 3 are marked in bold.

Since the magnification scale is 2, the difference between SR and HR is not obvious for the human vision. When the scale factor is 8, the image is blurred. Therefore, in this section, a scale of 4 was chosen to visualize qualitative results. The results are shown in Figs. 8-11.

These figures show that the artifacts and blurring of the bicubic interpolation algorithm are more serious, and the reconstruction effect is the most distorted. But with the use of deep learning technology to reconstruct the image, the reconstructed image effect has been greatly improved. When the SRCNN is adjusted to the 22-layer SRResNet, the reconstruction effect has been significantly improved. From these figures, it can be seen that the SRCNN and SRResNet reconstruction results are relatively clear. CA-SRResNet can make the detailed texture information more abundant so that the image reconstruction effect is the best.

To clearly clarify the distinguished texture features obtained by the CA-SRResNet approach, the local binary patterns (LBP) model was used to extract uniform texture information from rock CT images. LBP is the particular case of the texture spectrum model and can encode local texture information of images (He and Wang, 1990). The smaller squared error between LBPs of SR and HR, the more details are captured. Squared errors obtained from SRCNN, SRResNet and CA-SRResNet in different testing sets are shown in Figs. 12-15.

As can be seen that the squared errors from the CA-SRResNet are the smallest in each testing set, which indicates that the CA-SRResNet is able to capture more texture features than other networks.

4. Discussion and future work

In the previous research, attention-enhanced multi-scale residual blocks are used to enhance the proportion of useful information obtained. In the attention-enhanced multi-scale residual block, 3×3 , 5×5 , and 7×7 convolution kernels are utilized to extract multi-scale feature information. Using smaller convolution kernels is one of the current trends to reduce parameters while ensuring network accuracy. In our study, 3×3 convolution kernels are used instead of 7×7

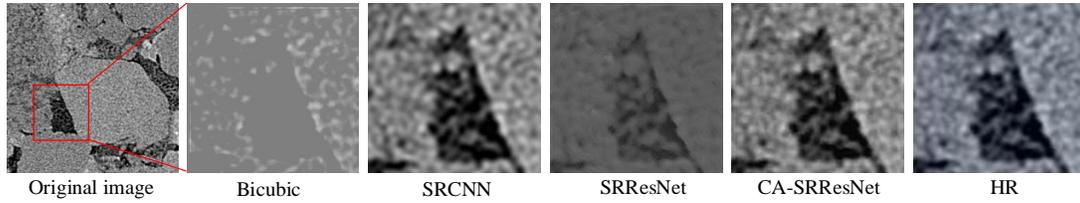


Fig. 8. Enlarged detail comparison of the part from testing set 1.

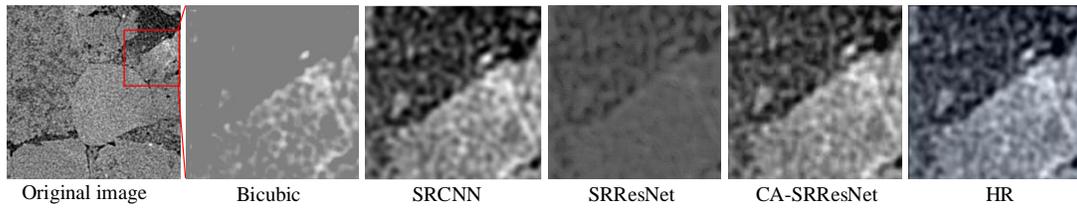


Fig. 9. Enlarged detail comparison of the part from testing set 2 .

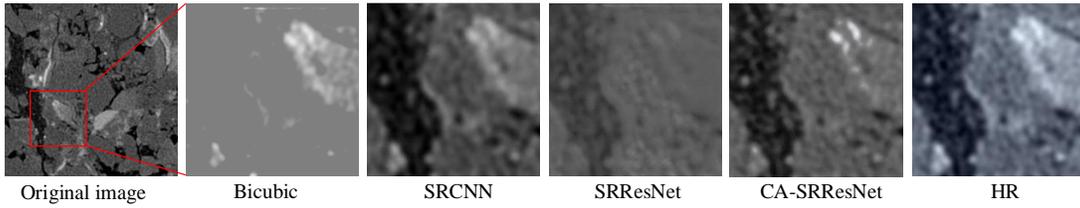


Fig. 10. Enlarged detail comparison of the part from testing set 3.

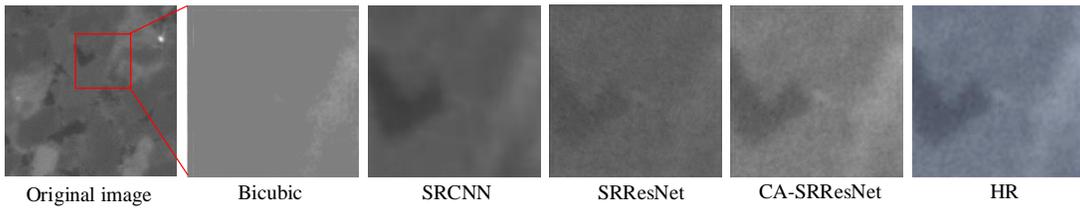


Fig. 11. Enlarged detail comparison of the part from testing set 4.

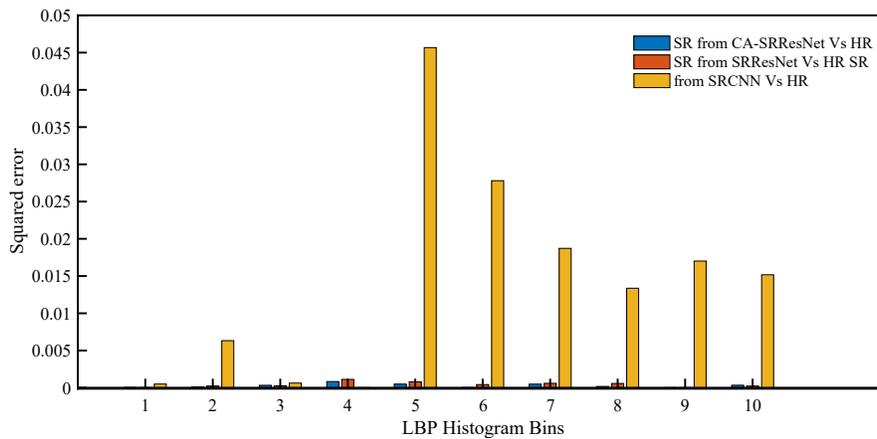


Fig. 12. Squared error between LBP of SR from three methods vs HR in testing set 1.

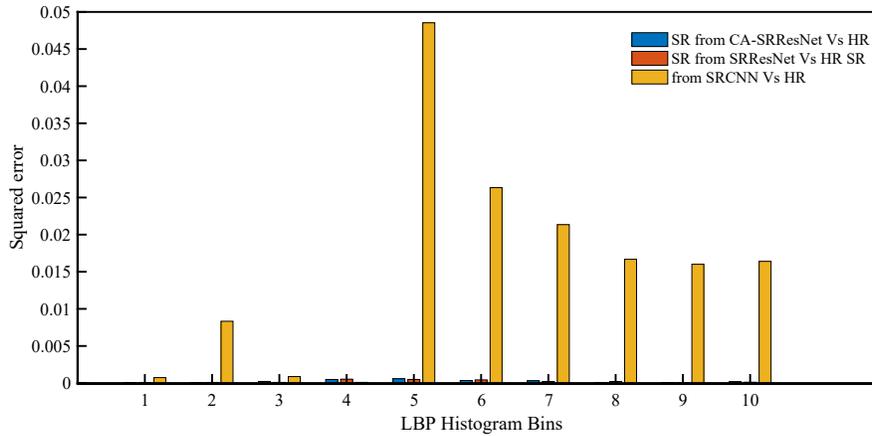


Fig. 13. Squared error between LBPs of SR from three methods vs HR in testing set 2.

convolution kernels, and 1×1 convolution kernels replace the 3×3 and 5×5 convolution kernels. Furthermore, the average pooling and maximum pooling features are used together to acquire more specific high-frequency information. The above figures show that the performance of the proposed CA-SRResNet is superior to those of previous methods. But our experiments mainly focus on 2D images. Accurate 3D images of rock can provide more accurate structure information that helps geologists analyze the physical properties. In a further study, 3D image SR will be considered for the finer digital rock reconstruction. Furthermore, the multiscale fusion of images has attracted much attention in the field of computer vision. It is its superiority of fusing different scale features that provide a superior visual effect. In future work, the rock samples will

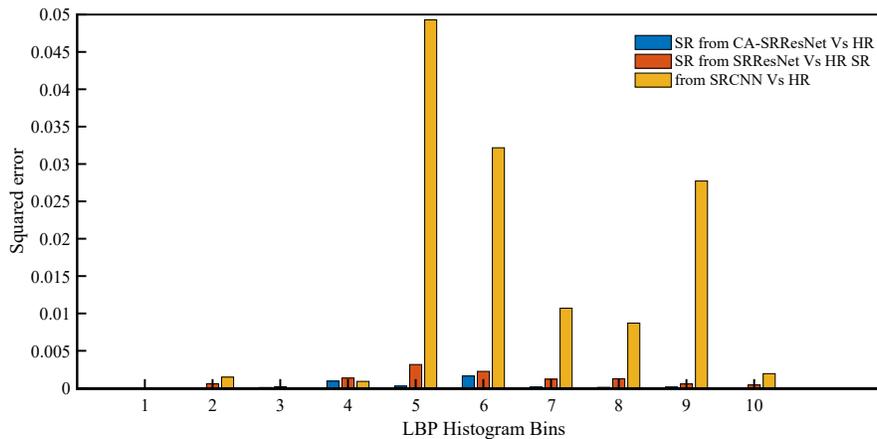
be collected and imaged to generate LR and HR images. The influence of resolution on pore network properties and single-phase, unsaturated, and two-phase flow will be analyzed to verify that pores, pore throats, average and surface area vary with resolution. These properties of the reconstructed images will be analyzed, which can be used to evaluate whether the reconstructed CT images with more details are correct and good enough to characterize the real rocks. Our study will explore efficient multi-scale fusion SR rather than perform SR by using single-scale images. $\times 2$, $\times 4$, and $\times 8$ down-sampling images will be used as model inputs. Then up-sampling and feature fusion are carried out at the corresponding scales. Because the features extracted from images of different resolutions will be richer, the reconstruction effect produced by the multi-scale

Table 2. PSNR comparisons.

Testing set	Scale factor	Bicubic	SRCNN	SRResNet	CA-SRResNet
1	2	30.707	35.31	46.224	47.230
	4	29.164	30.69	30.775	30.920
	8	23.725	24.28	25.471	25.551
2	2	27.447	39.50	43.892	46.065
	4	25.812	26.50	27.548	27.723
	8	21.237	21.23	22.402	22.471
3	2	33.871	38.19	39.753	39.922
	4	25.812	29.95	33.821	33.916
	8	21.237	24.28	33.821	31.492
4	2	38.698	39.38	44.382	44.689
	4	38.179	37.75	38.551	38.752
	8	34.496	33.81	35.594	35.705
5	2	46.127	46.16	49.902	50.515
	4	45.515	41.72	45.601	46.253
	8	41.355	31.41	42.163	42.458

Table 3. SSIM comparisons.

Testing set	Scale factor	Bicubic	SRCNN	SRResNet	CA-SRResNet
1	2	0.816	0.871	0.995	0.996
	4	0.752	0.714	0.817	0.821
	8	0.640	0.510	0.511	0.515
2	2	0.777	0.988	0.995	0.996
	4	0.688	0.752	0.791	0.788
	8	0.376	0.396	0.394	0.397
3	2	0.793	0.939	0.951	0.954
	4	0.762	0.786	0.791	0.795
	8	0.627	0.568	0.650	0.654
4	2	0.903	0.913	0.974	0.974
	4	0.891	0.887	0.900	0.904
	8	0.819	0.831	0.833	0.834
5	2	0.977	0.911	0.990	0.991
	4	0.970	0.888	0.974	0.977
	8	0.949	0.687	0.954	0.956

**Fig. 14.** Squared error between LBPs of SR from three methods vs HR in testing set 3.

fusion model will be more in line with the human visual effect.

5. Conclusion

In order to deal with the distortion of rock CT images caused by factors such as acquisition equipment, environment, network transmission media, a reconstruction model based on CNN, residual learning, and residual channel attention algorithm was proposed to convert a single LR image to an HR image. Directly embedding the attention module into the CNN network will weaken the ability of the network to extract features, both the average pooling and maximum pooling together are introduced to construct the channel attention module and then connect it to the residual module by the skip connection. Meanwhile, the activation function

ELU is used, and the location of a BN layer is adjusted, which not only gets superior SR performance but also reduces the training time. The experimental evaluations on sandstone and carbonate samples have verified the effectiveness of our method towards improving detail texture features of digital rock, which is of importance in the field of geological and petroleum exploration.

Acknowledgement

This work was supported by Fund Projects: Natural Science Foundation of Hebei Province (E2021107005), Northeast Petroleum University Cultivation Fund (2018GP2D-04).

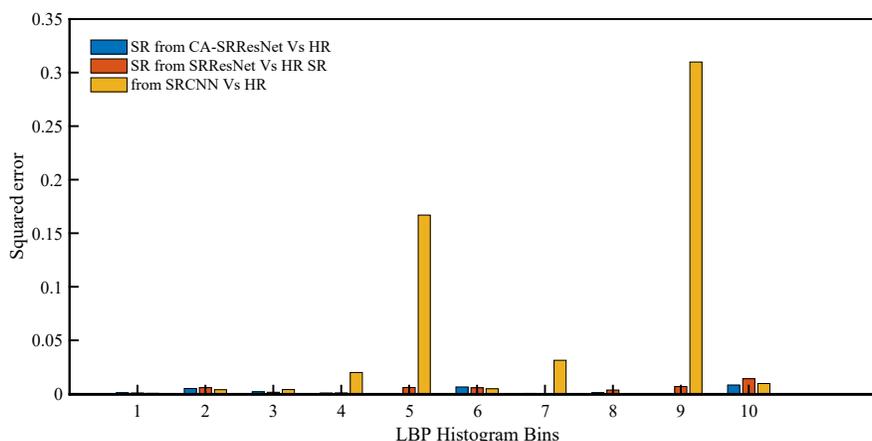


Fig. 15. Squared error between LBPs of SR from three methods vs HR in testing set 4.

Conflict of interest

The authors declare no competing interest.

Open Access This article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC-ND) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

References

- Andhumoudine, A. B., Nie, X., Zhou, Q., et al. Investigation of coal elastic properties based on digital core technology and finite element method. *Advances in Geo-Energy Research*, 2021, 5(1): 53-63.
- Chen, R., Zhang, H., Liu, J. Multi-attention augmented network for single image super-resolution. *Pattern Recognition*, 2022, 122: 108349.
- Dong, C., Loy, C.C., He, K., et al. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 38(2): 295-307.
- Dong, C., Loy, C. C., Tang, X. Accelerating the super-resolution convolutional neural network. Paper Presented in Proceedings of the Computer Vision-ECCV European Conference, Amsterdam, The Netherlands, 11-14 October, 2016.
- Engelmann, J. Lessmann, S. Conditional Wasserstein GAN-based oversampling of tabular data for imbalanced learning. *Expert Systems with Applications*, 2021, 174: 114582.
- Geng, T., Liu, X., Wang, X., et al. Deep shearlet residual learning network for single image super-resolution. *IEEE Transactions on Image Processing*, 2021, 30: 4129-4142.
- He, D., Wang, L. Texture unit, texture spectrum, and texture analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 1990, 28(4): 509-512.
- He, K., Zhang, X., Ren, S., et al. Deep residual learning for image recognition. Paper Presented in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 26 June-1 July, 2016.
- Ji, L., Lin, M., Jiang, W., et al. An improved method for reconstructing the digital core model of heterogeneous porous media. *Transport in Porous Media*, 2018, 121(2): 389-406.
- Khan, A., Sohail, A., Zahoor, U. A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, 2020, 53(8): 5455-5516.
- Lin, C., Wu, Y., Ren, L., et al. Review of digital core modeling methods. *Progress in Geophysics*, 2018, 33(2): 679-689. (in Chinese)
- Liu, J., Zhang, W., Tang, Y., et al. Residual feature aggregation network for image super-resolution. Paper Presented in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Washington, USA, 13-19 June, 2020.
- Liu, S., Sang, S., Wang, G., et al. FIB-SEM and X-ray CT characterization of interconnected pores in high-rank coal formed from regional metamorphism. *Journal of Petroleum Science and Engineering*, 2017, 148: 21-31.
- Liu, Y., Shan, L., Yu, D. An echo state network with attention mechanism for production prediction in reservoirs. *Journal of Petroleum Science and Engineering*, 2022, 209: 109920.
- Popel, M., Tomkova, M., Tomek, J., et al. Transforming machine translation: A deep learning system reaches news translation quality comparable to human professionals. *Nature Communications*, 2020, 11(1): 4381.
- Qiu, D., Zheng, L., Zhu, J., et al. Multiple improved residual networks for medical image super-resolution. *Future Generation Computer Systems*, 2021, 116: 200-208.
- Shan, L., Cao, L., Guo, B. Identification of flow units using the joint of WT and LSSVM based on FZI in a heterogeneous carbonate reservoir. *Journal of Petroleum Science and Engineering*, 2018, 161: 219-230.
- Shi, W., Caballero, J., Huszár, F., et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. Paper Presented in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 21-26 July, 2016.
- Song, H., Jin, Y., Cheng, et al. Learning interlaced sparse Sinkhorn matching network for video super-resolution.

- Pattern Recognition, 2022, 124: 108475.
- Vaswani, A., Shazeer, N., Parmar, N., et al. Attention is all you need. Paper Presented in Proceedings of the Neural Information Processing Systems, California, USA, 4-10 December, 2017.
- Wang, H., Hu, Q., Wu, C., et al. Dcnnet: Dual closed-loop networks for face super-resolution. *Knowledge-Based Systems*, 2021, 222: 106987.
- Wang, Y. D., Armstrong, R. T., Mostaghimi, P. Boosting resolution and recovering texture of 2D and 3D micro-CT images with deep learning. *Water Resources Research*, 2020a, 56(1): e 2019WR026052.
- Wang, Y., Teng, Q., He, X., et al. CT-image of rock samples super resolution using 3D convolutional neural network. *Computers & Geosciences*, 2019, 133: 104314.
- Wang, Z., Chen, J., Hoi, S. C. Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020b, 43(10): 3365-3387.
- Yang, Y., Zhou, Y., Blunt, M. J., et al. Advances in multiscale numerical and experimental approaches for multiphysics problems in porous media. *Advances in Geo-Energy Research*, 2021, 5(3): 233-238.
- Zeng, Y., Xiao, Z., Hung, K.W., et al. Real-time video super resolution network using recurrent multi-branch dilated convolutions. *Signal Processing: Image Communication*, 2021, 93: 116167.
- Zha, W., Li, X., Xing, Y., et al. Reconstruction of shale image based on Wasserstein Generative Adversarial Networks with gradient penalty. *Advances in Geo-Energy Research*, 2020, 4(1): 107-114.
- Zhang, Y., Li, K., Li, K., et al. Image super-resolution using very deep residual channel attention networks. Paper Presented in Proceedings of the Computer Vision-ECCV European Conference, Munich, Germany, 8-14 September, 2018.
- Zhang, Y., Tian, Y., Kong, Y., et al. Residual dense network for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 43(7): 2480-2495.
- Zuo, J., Wang, Z., Zhang, Y., et al. Research on image super-resolution algorithm based on mixed deep convolutional networks. *Computers & Electrical Engineering*, 2021, 95: 107422.