**CT-image Super Resolution Using 3D Convolutional Neural Network**

**Yukai Wang,**

Sichuan University, College of Electronics and Information Engineering, No. 24 South Section 1, Yihuan Road, Chengdu 610065, China

**Abstract:** Computed Tomography(CT) imaging technology is widely used in geological exploration, medical diagnosis and other fields. In the practical situation, however, the resolution of CT image is usually limited by scanning devices. SR(Super resolution) methods based on deep learning has obtained surprising performance in 2D(two-dimension) images. Unfortunately, there are few super-resolution algorithms for 3D images. In this paper, we proposed a novel network named as three-dimension Super Resolution Convolutional Neural Network(3DSRCNN) to realize voxel super resolution for CT images. To complete tasks with less number of CT samples, we explored the guidelines of optimizing network architecture to solve practical problems such as slow convergence of network training, insufficient memory, etc. We utilize the deeper network structure and employ adjustable learning rate, residual-learning, gradient clipping, momentum SGD strategies to optimize procedure of training. Finally, the experimental results show that our single model can complete the task of multi-scale interpolation reconstruction and have better performance in terms of PSNR , SSIM and efficiency compared with conventional methods.

**keywords**: super resolution, CT images, 3D-CNN, residual learning

1. **Introduction**

CT is a three-dimension(3D) imaging technique which is widely used to provide detailed information for accurate analysis. Recently, CT, micro-CT and nano-CT have been the popular equipment to display real 3D rock sample images.1 Establishment of accurate 3D image of rock can provide rich structure information helping geological researchers to analyze the physical properties of rocks2,3 and play an important role in the field of geological and petroleum exploration. High resolution CT sequence, however, generally need longer time to scan causing degraded resolution in slice direction(Z axis), shown in Fig. 1. Due to its inherent limitations of CT devices, setting high resolution will not only come at the expense of cost, but also the higher resolution will result in decrease of field of view(FOV).10 Therefore, the use of super-resolution algorithm is an effective method to improve the resolution of CT providing more clear sample data for subsequent medical diagnosis or geological research.



**Fig. 1**  CT images acquisition. 3D image acquisition is inconvenient, usually composed of a series of 2D slice images. The resolution in the slice direction is commonly lower than in the in-plane directions.

Super-resolution(SR) reconstruction, having drawn extensive attention in computer vision field, is effective method to improve quality of image.5 If external examples are given, learning-based SR algorithms are more plausible to acquire good results.

Deep learning techniques11 recently have shown remarkable performance in the tasks of image classification15, object detection16, etc, and it is much better than conventional machine learning algorithms. Chao Dong raised a network SRCNN4 which only contains 3 layers CNN structure but outperform than former method. Subsequently, Jiwon Kim found that deeper network structure show a significant promotion and proposed VDSR12 to resolve issues in Dong’s work.

In current research, scholars mainly focus on single 2D image reconstruction rather than spatial 3D voxel. SR research for 3D images mainly focuses on Magnetic Resonance Imaging(MRI). H. Greenspan archived super resolution of MRI images in slice direction using an iterative algorithm which gives improved resolution7. Manjon proposes a non local MRI upsampling method6 which can recover some of underlying high frequency data. Iwamoto proposes a method based on sparse representation and self-similarity to improve resolution of MRI.8 This method only improves the resolution in the slice direction, and does not have effect in plane direction.

Yuzhu Wang9 using neighbour embedding algorithm to improve resolution of CT image of rock sample while high frequency information is supplemented by high resolution scanning electron microscopy (SEM) image. Li proposed a voxel SR reconstruction algorithm10 based on sparse representation, which improve the resolution in all directions.

Zhang extend A+ ,i.e., Adjusted Anchored Neighborhood Regression algorithm13, to 3D format and proposed high frequency modified 3DA+ algorithm14 , establishing a correlative dictionary between high frequency and low frequency 3D block. The matched dictionary atom and mapping matrix were searched for each input of the 3D block in reconstruction stage.

In view of the fact of CT samples of rock, the following issues remain to be solved: First, the amount of 3D image data to be calculated is far greater than the 2D images, so the method to handle with 2D images can’t be directly transferred to 3D model; Second, CT samples are not as convenient as 2D images to get. It's not easy to obtain substantial alignments of rock CT samples. In addition, CT image of rock has the characteristics of low contrast, single texture, and complex pore structure, which all brings difficulty to task of SR; Third, during training network and reconstruction stage, the space complexity and time complexity have to be taken account to ensure our work can be carried out on the general computing equipment.

We attempt to enhance resolution of CT images of rock from three directions(i.e., x, y ,z). In our work, we propose a novel network, termed as three-dimension super resolution covolutional neural network(3DSRCNN), to promote resolution for volumetric images. To resolve aforementioned issues, we use 12 layers CNN network of which each layer contains 64 convolutional kernels. Deeper network architecture may cause gradient exploding and slow convergence, so we employ residual learning, gradient clip strategy and adjustable learning rate, etc strategies. Our network can be applied to different upscaling factors and performs as well as method separately training network with each upscaling factor. Before training, cropping samples is necessary for a promising result. In order to assess performance of 3DSRCNN, we compare the reconstruction result with LR images by PSNR and SSIM. Using trained model to reconstruct, it is observed that 3DSRCNN has better performance of PSNR and SSIM on different testsets, while ours have faster reconstruction speed on GPU.

In summary, we introduce 3DSRCNN to realize the SR reconstruction of 3D CT images. Given actual sample condition, We have experimentally investigated the influence of network depth on the reconstruction accuracy. Thus we consider that employing a moderate number of network layers is of significance. Subsequently, We find that it is necessary to use the residual network when the number of network layers go deeper. Consequently, we make a corresponding adjustment on network architecture and training strategies, so it have achieved a trade-off between the accuracy and speed. Moreover, We have addressed the problems existing in pre-processing CT images such as consuming too much memory, etc. The proposed 3DSRCNN performs favorably against the state-of-art methods in terms of accuracy, efficiency, and practicability.

The remainder of paper is organized as follows. In Section 2, we firstly introduce the concept of SR and the implementation of deep learning on it. Then, Our proposed network--3DSRCNN is detailedly described in Section 3. In Section 4, we experimentally investigate how to design the network and pursuit better performance between accuracy and speed. Besides, we also test and compare our method to others. In Section 5, analysis and conclusion of our work are given for future studies.

1. **Related work**

In this section, the conception of super resolution and method of using CNN to SR are briefly introduced.

* 1. *Image Super-Resolution*

Single Image Super Resolution(SISR) is an ill-posed problem due to lacking of detailed information. There are two traditional methods to restore low resolution(LR) images to high resolution(HR) images, one is using context correlation in LR image yet has inborn defects that it cannot acquire more specific high frequency information; The second is learning based method that can acquire the prior information through training given images. The process of SISR is that, for a ground truth image (HR images), and LR images .

 (1)

Our goal is to find a function as, which can restore LR to HR to a certain extent. It is a under-determined problem and most of recent state-of-art methods adopted learning based strategy. SR reconstruction based on learning method is to learn the mapping relation of low frequency information to high frequency information by iteratively training. The learning-based SR such as sparse-representation method17,18 is basically composed of three steps: (1)LR features extraction; (2)Learning mapping relation between LR and HR patch; (3) Reconstruct to HR images using learned mapping relation. In A+ algorithm13 , the nonlinear mapping relationship from low resolution space to high resolution space is transformed into mapping matrix, and the super resolution reconstruction operation is transformed to matrix multiplication, which can be archived by deep learning technique.

* 1. *Convolutional Network for Super-Resolution*

Chao Dong consider deep convolutional neural network is equivalent to the aforementioned pipeline, which can directly learns an end-to-end mapping relation. While SRCNN archive good result in 2D image datasets, there are still limitations as following: (1) Its single model works only for single scale, which cannot be applied on different upscaling factors; Second, training of SRCNN converges too slowly.

Jiwon Kim introduce VDSR to approach the limitations in SRCNN and point out stacking more CNN layer lead convolutional filters become increasingly global, which conceptually benefit to learn mapping relation , and utilize deeper network structure with a total of 20 layers to complete whole super-resolution. Jiwon have experimentally validated the viewpoint –‘the deeper, the better’ . The SR technique of single 2D image has been very mature, but they can’t be directly converted into a 3D model. Because the amount of data used to calculate in 3D image is far larger than the two-dimensional image, It’s necessary to redesign network architecture. Furthermore, acquisition of 3D image sample are not as easy as 2D images. We try to use a relatively small samples to complete the training of the network as far as possible.

1. **Three-dimension Super Resolution Covolutional Neural Network**

In this Section, we introduce the structure of the 3DSRCNN that consist of 12 layers of 3D-CNN. Besides, some strategies for optimizing training process are employed to our network 3DSRCNN. Next, We describe production of training data in detail.

* 1. *Network Structure*

We proposed a 3D network structure, named as 3DSRCNN, to archive super resolution for volumetric CT images .



**Fig. 2** Network structure of 3DSRCNN which contains 12 layers 3D-CNN. Each 3D-CNN have 64 filters to capture diverse features. LR images goes through layers and transforms into HR images. The output of the network is actually the prediction of the residual image. We combine residual image and input as the final output to calculate loss function.

For volumetric super resolution, we employ a network composed of 12 layers each of that has 64 channel(covolutional kernel). The first layer is responsible to extract low frequency patch from LR images; The middle 10 layers learns mapping relationship between LR and HR volumetric block; The last layer combine learned mapping relation and initial LR images to finally formulate SR images.

The convolution network actually extracts spatial correlative information which contains diverse pattern features. Recent study 19,21 shows increasing depth using an architecture with very small (3×3) convolution filters, which shows that a significant improvement on Image Recognition,etc. Simultaneously, When the input image continues to pass through the CNN, the extracted feature becomes global and has a larger receptive field. Consequently, the depth of network layers will affect the reconstruction accuracy and training time. Due to original images containing rich texture information, the deeper network structure have better SR ability as Jiwon pointed in their work12. Computation complexity, however, is a non-negligible topic which directly influence the practical application of our algorithm. The whole computation complexity of network can be calculated as

 (2)

whereis the depth of CNN layers,identify the current layer number,is number of channels,is feature map size. It is obvious to find that dense network structure would increase the computational complexity.

SRCNN that have no padding before covolutional operation, causing boundary pixel missed. On the contrary, padding is necessary for our network because that processing 3D image will typically occupy a lot of memory. In order to save memory, we divide the initial CT image to sub-blocks with small size. Given that the layers of network is 12, the input size is relatively small, which will cause majority loss of internal information without padding during forward propagation. Hence, we use zero padding and subsequent experimental have proved the correctness of this scheme.

Because the SRCNN network has only three layers of network, it not only completes learning mapping relation between LR and HR but also remains initial LR feature during forward propagation. When largely increase layers, the information of the input LR feature will be lost in the continuous convolution process, which leads to training unstable and discard initial information. We consider residual learning can be used to solve the above problems. After each CNN, we utilize Rectified Linear Unit(ReLU)27 as activation function on output of last layer.

 (3)

Wheredenote input and weight parameter respectively,is bias.

* 1. *Pre-process of training set*

Before training, we should crop and transform initial CT sample to suitable shape. Specifically, The step of crop original CT samples is introduced as follows.

We separately use factorto downsample ground truth CT datasetswith original size in different samples, then we use bicubic interpolation to upsample them by same size and these are used as LR images . We crop initial CT blocks to sub-3D-blocks to produce training set. HR images{Y} is viewed as label to calculate loss function, and LR images {X} are fed in network. The whole process is shown as Fig 3.



**Fig 3**  Procedure of making training set.

When deal with CT images, crop is significant for training and there are mainly three points as following:

1. In this way, a larger quantities of training samples can be obtained through image cropping under the condition of limited number of CT samples. These sub-blocks are viewed as small size‘images’ rather than ‘patch’.
2. Cropping promises our program running in general computer since training 3D-block will occupy amounts of memory. When large CT block is cropped into small blocks, it enables computing devices to calculate under low load.
3. The sub-blocks are overlapping containing redundant information, in the sense that training set have rich contexture that is advantageous to learn mapping relationship.

Assuming that the input is a cubic block, the specific number of samples after cropping can be counted with following : , Whereis initial size of CT, is sub-block size, is span length when cropping. Setting suitable parameters is of significance for speed and accuracy.

After the cropping , datasets is composed of pairedthat are used as input and label for training. Through experiments, we find setting  as  is appropriate.

* 1. *Training strategies*

Our proposed network is constituted of massive tensors which represent end-to-end mapping relationship. Weight parameters of network is initialized by Gaussian distribution(zero mean and standard deviation 0.001). Through continuous iterative training, is increasingly optimized by Mean Squared Error(MSE) loss function. However, directly using standard SGD(Stochastic Gradient Descent) takes long time to converge. We employ some strategies to optimize our network structure and training data.

1. *Residual Learning*

In spite of larger depth of network have significant effects on SR, We find gradients vanishing/exploding will appear when depth exceed 10 layers and deeper model produce higher training error which makes training process unstable. On the other hand, In Ref. 12, author consider that input detail is discarded after passing covolutional operation in deep layers, which gives birth to that the output only use learned features to generate images. He Kaiming have introduced a deep residual learning framework24 to and got excellent scores in image recognition. Residual-learning strategy is also adopted in our network to solve these problems. We define input as, output as , and residual image, wheredenotes the output of data passing through network . Given training set, and loss function based on MSE is interpreted as following:

 (4)

Whereis the number of training batch samples. One point must to be stressed is that the residual learning is not necessary in all cases. When the number of layers is not deep, the use of the residual network does not have obvious effect, or degrade instead.

1. *Adjustable learning rate*

In SRCNN, it is found that the training with small learning rate converges very slowly. High learning rate help to boost training yet can lead to gradients exploding. We use the adjustable learning rate to speed up the training process. In early Epoch, setting relatively high learning rate will be benefit of accelerating training. As training epoch going on, learning rate is reduced with following rules.

 (5)

Wherecounts current training times, and  is predefined to control decay of learning rate.

1. *Momentum acceleration*

Due to the magnitude of complexity in 3D images, the convergence of using standard SGD is very slow. We employ momentum SGD to accelerate training process.

Momentum is a commonly used acceleration technique in gradient descent. It accumulates the momentum before it replaces the real gradient. The implementation of SGD with Momentum in our work subtly differs from Sutskerver’s work26 . Considering the specific case of Momentum, the gradient update formula is written as a new form:

|  |
| --- |
| Algorithm：Momentum SGD |
| Require：learning rate ，momentum coefficient ，weight parameters , velocity *v*；  while do:  batch sample with size m，label as ；  Update Gradient：；  Update velocity：；  Update weight parameters：  end |
|  |

Whereand denote the weight parameters in network, gradients, velocity, and momentum factor, respectively. In our experiments, We all set momentum factor to 0.9.

1. *Gradient clipping*

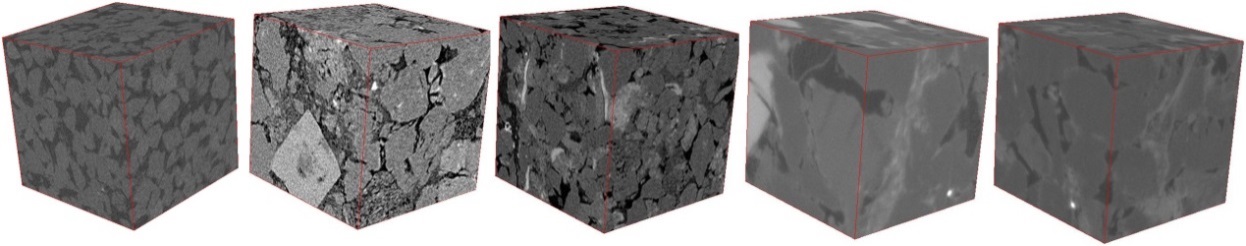
Gradient clipping20 is usually applied in training RNN network in case of gradient exploding/vanishing. One simple way is pre-defining a threshold to clip them whenever they go over a threshold. In VDSR, Jiwon12 use this technique to limit gradients to a certain range. In our work, We directly clip gradients to range, where  is predefined clipping range.

1. **Experiments**

In this Section, we first introduces experimental basis and evaluation metric, Peak Signal to Noise Ratio(PSNR) and Structural Similarity Index(SSIM), which is widely used to assess image quality. Next, We investigate the influence of important parameters on the accuracy of reconstruction, and analyze the reasons. At last, we compare our method with others both accuracy and speed.

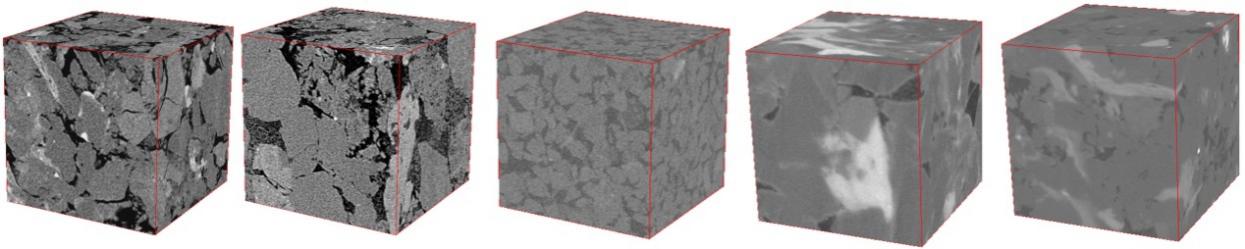
* 1. *Experimental datasets and evaluation Criteria*

Deep learning generally benefits from big data training, considering the actual situation, it is not easy to get rock CT images, we attempt to use a relatively small number of CT samples to making training set. In order to make the experimental results more convincing, we selected a batch of training samples from rock CT samples which come from diverse rock types with different pore characteristics. The displayed in Fig. 4 are scanned CT images of rock samples. The test samples are different from training set which both are consistent with the identical selection rules to guarantee convinced results as Fig. 5 shown.



1. sample1 (b)sample2 (c)sample3 (d)sample4 (e)sample5

**Fig. 4** Five sets of original CT images of reservoirs as training CT samples: (a)(b)(c)are Sandstone with resolution of 3.8 (d) is carbonate rock with resolution of 1.07 (e) is sandstone with resolution of 1.07



1. sample1 (b)sample2 (c)sample3 (d)sample4 (e)sample5

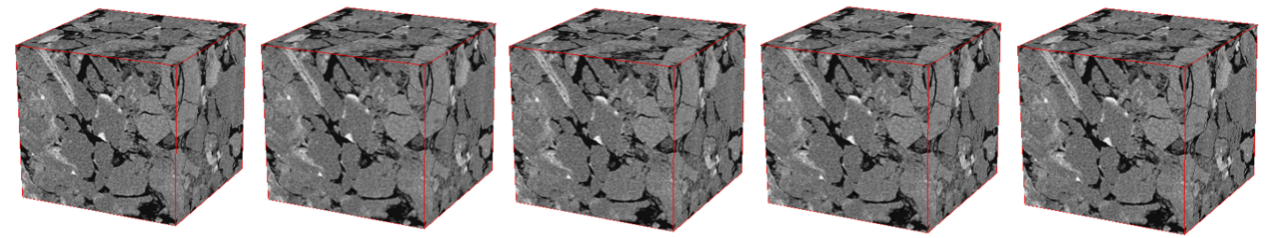
**Fig. 5** Five sets of original CT images of reservoirs as testing CT samples.

Results of reconstruction are validated by PSNR and SSIM. PSNR is widely used to measure the quality of image restoration. For a 3D imagewith a size of , PSNR is calculated as follows:

 (6)

 (7)

Wheredenotes original HR images. The higher PSNR value indicates that reconstructed CT images have better quality. PSNR is a evaluation based on pixel error, however, it does not take into account the visual features of human eyes. Wang use SSIM25 to represent the structure information of the image from the brightness, contrast and structure of the image, and it is more consistent with the sense of the human performance. Range of SSIM value is [0,1], and the higher the SSIM indicates the closer it is to the actual samples.



(a)Original CT image (b)Bicubic interpolation（c)MRI (d)3DA+ (e)3DSRCNN





(f)original slice (g) Bicubic interpolation (h)MRI (i)3DA+ (j)3DSRCNN

**Fig. 6** SR results of different algorithm in scale factortestset of Sandstone rock. We can find that the zoom-in image of 3DSRCNN contains more clear texture than other method.

* 1. *Work platform Details*

Detailed hardware and software are listed in Table 1. In order to ensure program running successfully, computing device must have enough memory at least 8Gb. Owing to huge time consumption by CPU method, a method that utilizes CUDA to invoke GPU resources is adopted to speed up the training process.

**Table1** Hardware and software platform

|  |  |
| --- | --- |
| CPU | Intel i7-6770K 4.0G Hz |
| RAM | DDR4 16GB |
| OS | Ubuntu 16.04 |
| GPU | Nvidia GTX 1080 |
| Framework | Pytorch 0.31 |

We use open source deep learning framework, Pytorch 0.31, to build network and complete reconstruction, and use Matlab to generate training datasets.

* 1. *Multi-scale and single-scale for training*

A single model enabling to be implemented into multi-scale scenarios is critical for practice work. We consider model trained with upscaling factor=(the corresponding trainset/testset is )have better PSNR and SSIM on different samples. We randomly select samples from three scale sets, which is by same proportions, to synthesize the multi-scale training sets and comparing the training process with the multi-scale and single training set, respectively. In training stage, average MSE with multi-scale, is higher than single-scale in the early Epoch, and the convergence process of multi-scale is more slower as well. Nevertheless, From Table 3, after convergence of both network, PSNR of multi-scale surpass counterpart in the case of , 0.32dB and 0.46dB, and almost be equal in. We consider that the use of the multi-scale set is more preferable. The following experiments are conducted by multi-scale training sets.

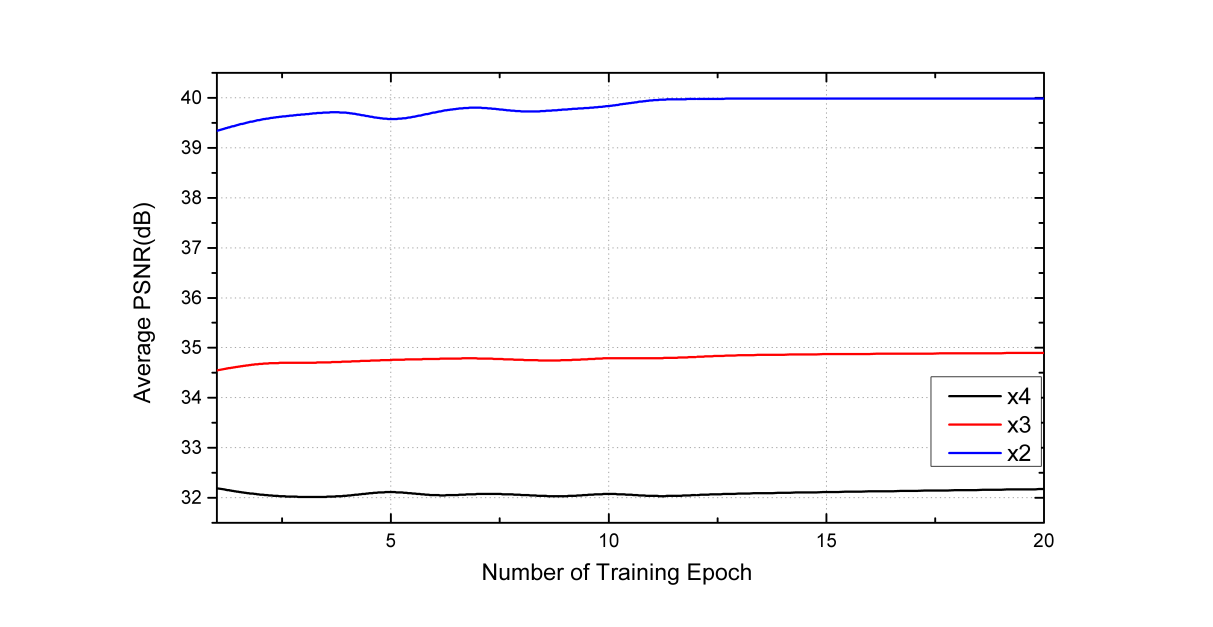
* 1. *Analysis of the process of training and reconstruction*

If training sets is blended up with multi-scale samples, the yielded model can be applied for different interpolation image which save the cost of storing network. In Fig. 7, it is observed that the trend of the curves in different scale factors are basically the same. 3DSRCNN need train about 20 Epochs to converge. In general, with continuous iteration, the reconstruction quality will also increase, but it will eventually converge to a certain number due to limitations of generalization of network and given training sets.

Note that not the low MSE value in training stage means that the reconstruction effect is better. Calculation of MSE is based on one batch that is currently taken out in the training set, which can not be deemed as the criterion of reconstruction quality. It is supposed to use trained model to complete process of reconstruction to compare performance.

As shown in Figure 6, on the training sets of different upscaling factors, the average PSNR increases with the increase of the number of training times Epoch, but it fluctuates in the process of rising, and may decline. This is because the loss function of the network uses the MSE function, and MSE helps to train a higher network of PSNR values, but with SGD optimization, the loss function may fall into a bad local minimum, resulting in a slight drop in the PSNR during the training process. From Table 3, the average PSNR increases 4.72dB, 2.48dB, 2.92dB, and SSIM increases 0.037,0.05, 0.111, respectively. It is evident to find higher gains in . This is because the using a higher scale factor will result in lots of original high-frequency details missed, and the scope of the SR may be limited. Through Fig. 7, we can find that 3DSRCNN converges to a stable PSNR value at about 15 Epoch. Using the computer configuration of Table 1, it takes about 2.29 hours for each training Epoch in our . The time cost and reconstruction precision are considered at the time of setting up the number of Epoch.

In general, with the continuous iteration of the network, the generalization ability of it will be enhanced. As Epoch gradually increase, PSNR is not linearly improved, and there may be a decrease in certain Epoch, which is due to the the loss function into the saddle point in the process of using SGD. As shown in Fig. 7, PSNR curve converges at about 15 Epoch at given setting, andimprove 2.89dB, 2.42dB, 1.50dB than input images, respectively. In higher upscaling factor testsets, the extent of improvement is reduced, which is due to the loss of more high frequency details, learning low frequency to high frequency mapping relationship is more difficult, which brings intrinsic limitation on SR ability.



**Fig. 7**  Average PSNR curves of 3DSRCNN on testsets with ,upscaling factors. The three curves are smooth, and we can hold the point that residual learning can boost convergence.

The trained network model is a set of Tensors storing the weight parameters of each neuron. Although trained CT imagesets are divided to small blocks, there is no requirement about the size of LR images to be SR. The ultimate goal of our work is to get HR images by identical size, but if directly send the images with relatively large pixels to the network, it will consume a lot of memories(e.g. Reconstruction of voxel data with size ofactually takes up about 352G memory, tested in Pytorch 0.31). Therefore, we also crop input CT images to smaller sub block with size of .

* 1. *Training parameters and Trade-offs*

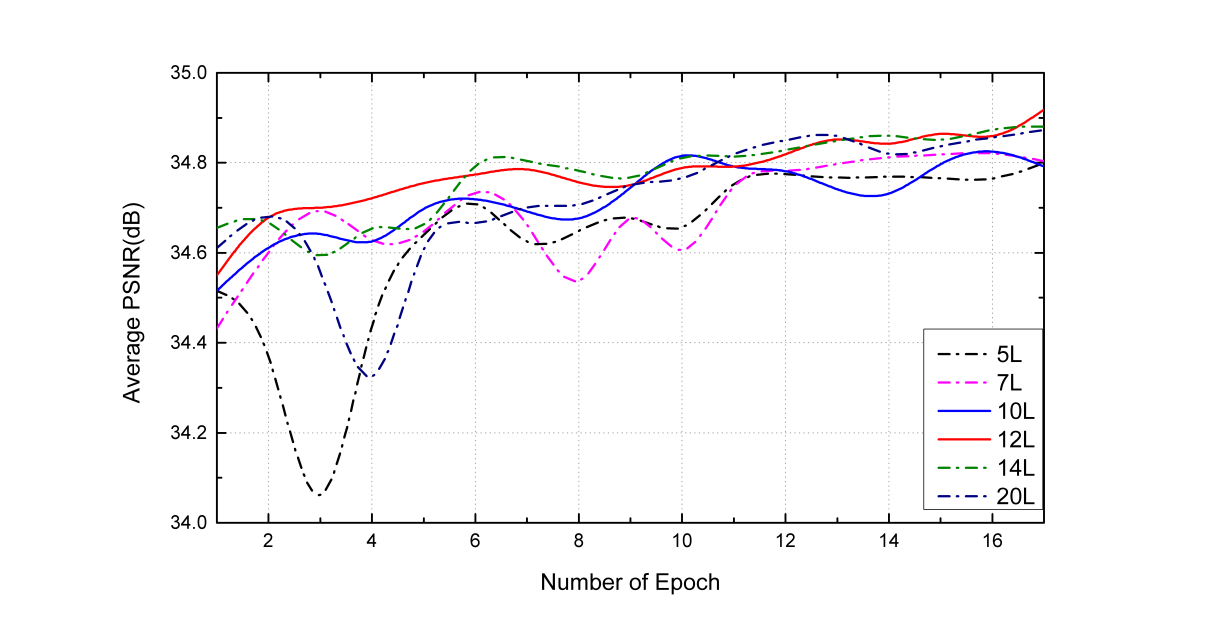
Network parameters including network depth and convolution kernel size will affect the accuracy and time. In most cases, increasing training Epoch is conducive to a better performance. The mapping relationship that a network can learn from a given training set is yet limited to quantities of training data and structure of network. Appropriate increment of the network structure and tune training parameters is of crucial. In this section, we investigate the optimal setting to make a trade-off between performance and speed.

* + 1. *Depth of network*

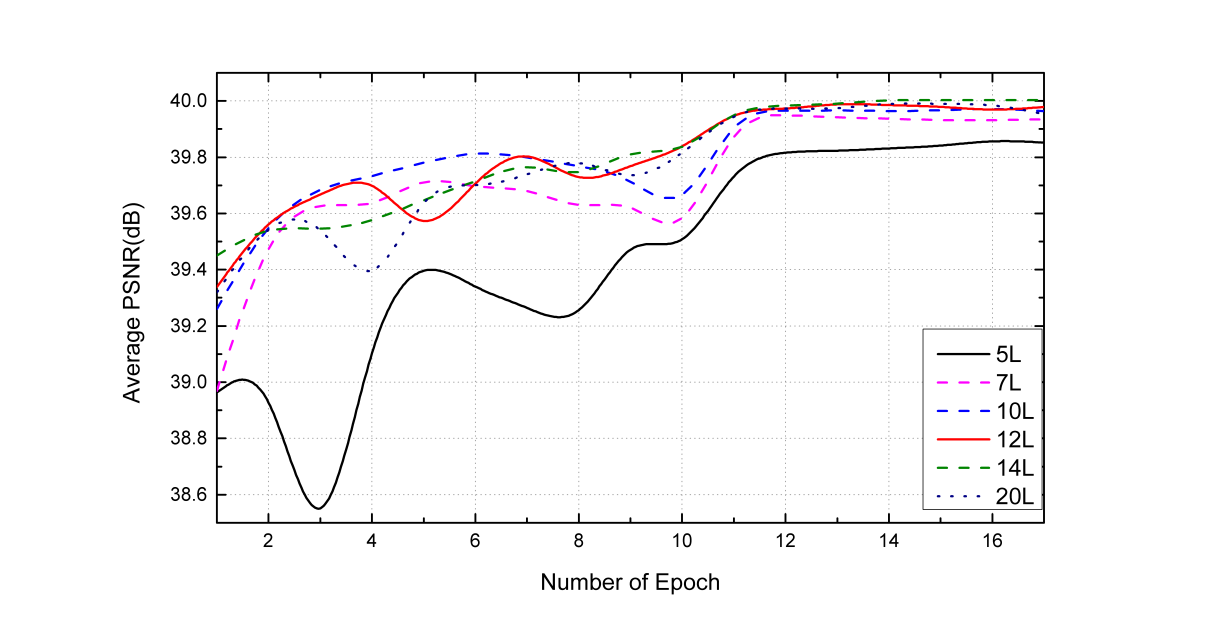
For super resolution in single 2D images, Jiwon Kim demonstrate the large depth help model to capture more contextual information and yield better performances than shallow ones. For the training of 3D samples, the amount of computation and memory occupied is very large. Too many layers could slow down the convergence and exponentially improve the computational complexity. As are shown in Fig. 8 , we found that the training results will be improved until layers are added up to 12. After that, there is no apparent promotion on accuracy of reconstruction. Deepening the network is indeed effective, but we find that the network layer is not ‘the deeper,the better’, which needs to be explored through practical experiments based on given datasets. We try training with small number of imagesets, so that the mapping relation learned from prior information is inherently limited, so an excessive increase of the layers will not substantially improve the accuracy of the reconstruction, Unexpectedly, it will bring about degradation problem due to over-fitting. We observe that the depth of the network will increase the computational complexity resulting in time-consuming. In conclusion, the layers of network should be proportional to the magnitude of training data, and moderate depth will help to improve network ability of SR.

Experiments indicate that increasing the number of network layers can apparently improve performance until 12 layers after that there is no obvious promotion. Considering both of the performance and speed, We think layers of 12 is admissible for our work.

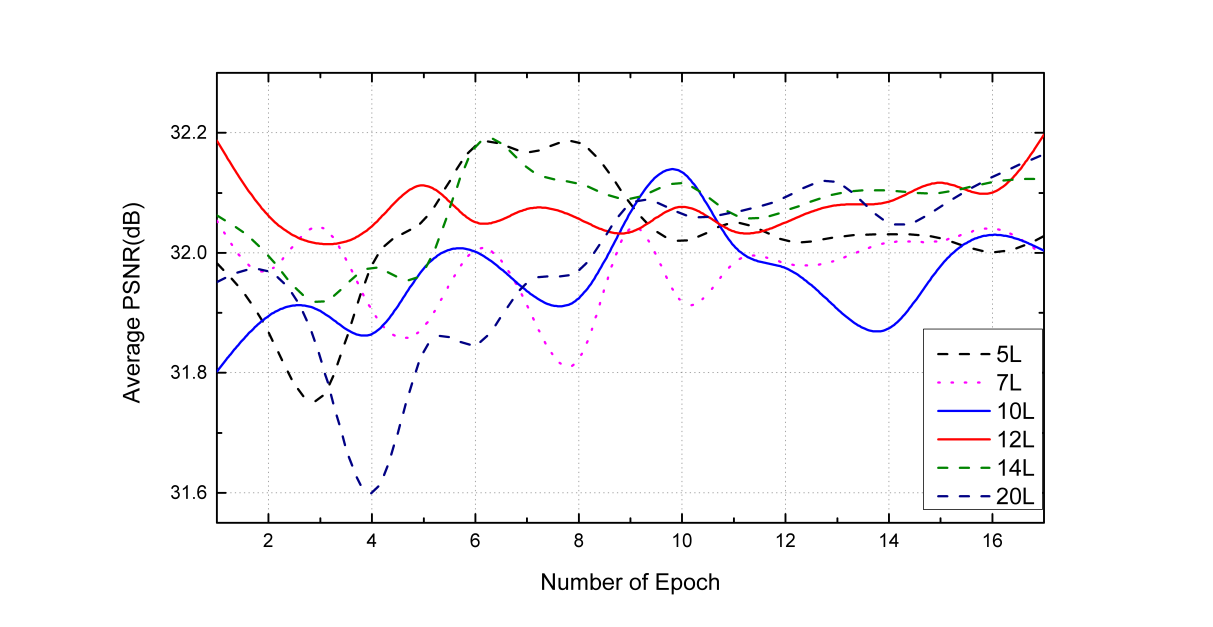
The 12-layers is about 0.15, 0.09 dB higher than 10, 14-layers network inand training curve is also relatively stable. In addition, 20 layer is lower than that of the 12 layer in . This issue is also mentioned in Ref. 4, where improper increase of depth will cause degradation of SR accuracy. This also confirms the above point of view, the number of layers in the network should be set up according to the number of training data and increasing the network by experiment will help to get better results.



1. The average PSNR curves in 



1. The average PSNR curves in 



1. The average PSNR curves in 

**Fig. 8** AveragePSNR of networks with different number of layers on three upscaling factors testsets. Considering the performance in the three scales, we can think that the 12 layers network is preferred

* + 1. *Convolutional kernel size*

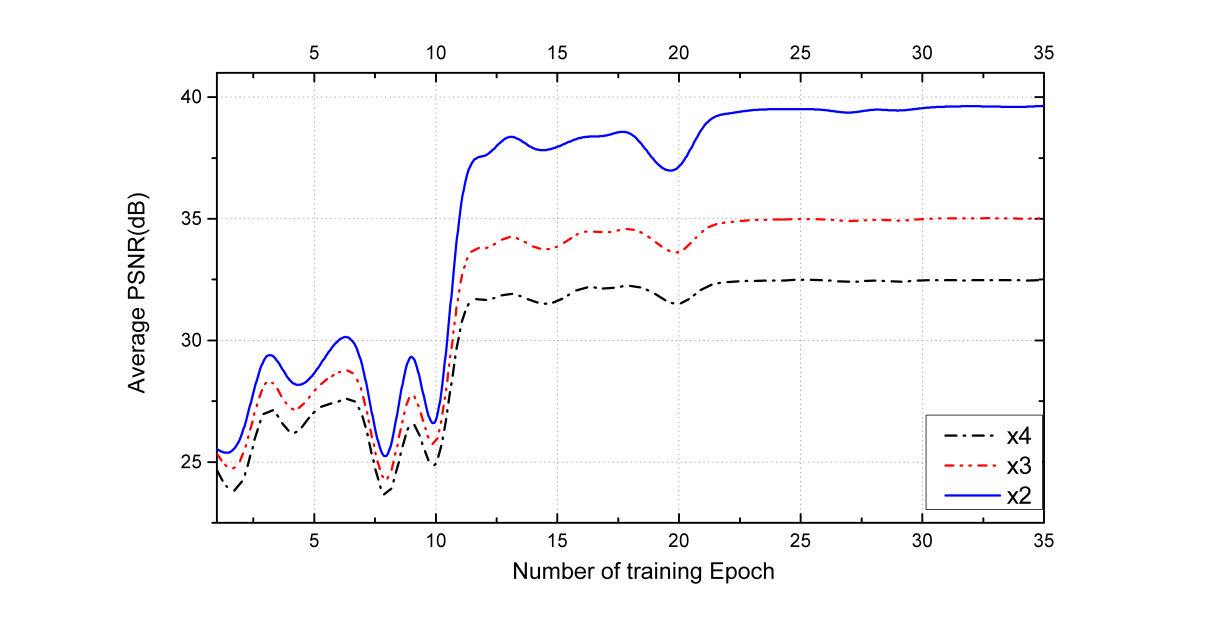
In this section, we explore network sensitivity to different convolutional kernel size. In Ref. 12, all the filters of CNN are set same size of . In Ref. 4, they examine the impact of different filter size, whereas their settings are inapplicable to ours due to the limitation of memory. Besides, the size of our input data is uniformly , that means it is better to set relatively small filter. Based on previous experiment that number of layers is set to 12, we conduct three comparative trials with convolutional size of ,and. The experimental results are shown as Table 2. We find that there is no evident improvement when widening kernel size tothan. Instead, usinghas a large reduction in quality. By using small convolution kernel, the larger receptive field can also be achieved by increasing the network depth. According to Formula (2), increasing the size of the filter incurs higher complexity than that of increasing the network depth. It is obvious that small filter size is preferable.

**Table2** Comparison of time for training a epoch with different convolutinal kernel size

|  |  |  |  |
| --- | --- | --- | --- |
| Size(pixel) | 3x3x3 | 5x5x5 | 7x7x7 |
| Time(min) | 81 | 250 | 428 |
| PSNR(dB) | 40.00 | 39.74 | 36.79 |

* + 1. *Residual vs Non-residual Learning*

Increasing network layer could cause the problem of the gradient explosion/vanishing which bring difficulties to training. Fig. 6 and 8 have intuitively displayed performance between the use and non use of residual learning. In the process of initial training, it is found that the MSE loss is very high and the trained model is used to reconstruct the whole curve. As shown by the Fig 8, there is a dramatic fluctuation in the curve without residual learning. Even worse, PSNR has a large decline during the 7,8 Epoch. Moreover, it will take long time to converge than using residual learning. In the same case, residual learning can make MSE converge to a smaller number in the first 3 epoch, and the network without residual learning needs to converge after training more than 25 epoch. The final result shows that using non-residual learning on andis about 0.2dB, 0.1dB higher ,0.4dB lower than counterpart onyet. Using residual learning can enhance the stability of training. When the network is deeper, the necessity of residual learning will be more prominent.



**Fig. 9**  PSNR curves with non-residual learning. It’s obvious to find that network with non-residual emerge a shock wave in early stage. Curve tend to converge until training 25 epoch which is much time-consuming than using residual learning.

* 1. *Comparison to state-of-the-art*

The performance of our proposed network on the reconstruction quality is experimentally analyzed compared with the benchmark bi-cubic interpolation, MRI, sparse-representation, 3D-A+, in the Table 3. We can see that the 3DSRCNN training model is better than the previous one on three upscaling testsets. Among them, PSNR is 2 times higher than 3DA+ on the test set of 3 times, higher than 0.14dB on the 3 times test set, and higher 0.356dB on 4 times test set. Therefore, it can be concluded that the neural network has a better learning ability than the traditional way of establishing a dictionary. The appropriate setting of the network parameters can fully obtain the redundant information in the training image, thus learning more prior information and finally achieve excellent performance.

The reconstruction time in 3DA+ is related to the size of the feature block. When the test feature block is large, the reconstruction quality is good. When thesize block is selected, reconstructing the size of  is about 22 minutes on CPU. Our work uses GPU to run program, and the reconstruction of the same size image only need 3 minutes and ours have better performance.

**Table 3** The average results of PSNR(dB),SSIM in comparison with other algorithm.3DSRCNN(single) are sperately trained and only works for corresponding upscaling factor. In contrast, 3DSRCNN(multi) can be applied in different upscaling factors.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Eval | Scale | Bicubic | MRI6 | Sparse representation | 3DA+ | 3DSRCNN(single) | 3DSRCNN(multi) |
| PSNR(dB) | 2 | 35.28 | 36.35 | 38.45 | 38.58 | 39.68 | 40.00 |
| 3 | 32.54 | 33.15 | 34.74 | 34.89 | 35.03 | 35.02 |
| 4 | 29.55 | 30.17 | 31.59 | 31.65 | 32.01 | 32.47 |
| SSIM | 2 | 0.950 | 0.965 | 0.983 | 0.983 | 0.986 | 0.987 |
| 3 | 0.879 | 0.881 | 0.924 | 0.930 | 0.931 | 0.929 |
| 4 | 0.738 | 0.799 | 0.817 | 0.828 | 0.850 | 0.849 |

1. **Conclusion**

In our work, we proposed a novel method, 3DSRCNN, based on deep learning to approach SR of voxel images. While using CNN to restore single LR image to high resolution have excellent score than tradition method, there are many challenges previously mentioned to accomplish CT sample super-resolution. Our proposed model employ 3D-convolutional operation to handle CT images, which ensures the spatial continuity in slice direction. Through practical experiments, We introduced empirical guideline in designing network and parameters-tuning in training process. Experiment prove that our single model can work for multi-scale SR processes. Moreover, stacking moderate network layers and setting training strategies will exert on the accuracy and time of reconstruction. Accordingly, we investigate certain parameter setting rules such as learning rate, depth of network, as well as size of covolutional kernel. In the process of experiment, we found that training and reconstruction need to spend a lot of memory, which need to draw more attention. To cope with aforementioned issues, we crop original CT images to small blocks. We have demonstrated our approach surpasses previous methods, which is shown in Table 3. For future research, we want to theoretically explain the effect of network depth on SR. Meanwhile, we will study the better training techniques to deal with the issues of larger quantities of 3D data, thus greatly reducing training time.

*Acknowledgements*

Thanks to Zhang TR , Feng JX, Li Y, Ding K .

*Reference*

1. Fredrich J T. 3D imaging of porous media using laser scanning confocal microscopy with application to microscale transport processes[J]. Physics & Chemistry of the Earth Part A Solid Earth & Geodesy, 1999, 24(7):551-561.
2. Hazlett R D. Simulation of capillary-dominated displacements in microtomographic images of reservoir rocks[J]. Transport in porous media, 1995, 20(1-2): 21-35.
3. Cnudde V, Boone M N. High-resolution X-ray computed tomography in geosciences: A review of the current technology and applications[J]. Earth-Science Reviews, 2013, 123: 1-17.
4. Dong C, Chen C L, He K, et al. Image Super-Resolution Using Deep Convolutional Networks[J].IEEE Transactions on Pattern Analysis & Machine Intelligence 2016, 38(2):295-307.
5. Yang C Y, Ma C, Yang M H. Single-Image Super-Resolution: A Benchmark[J]. Lecture Notes in Computer Science, 2014, 8692:372-386.
6. Manjón J V, Coupé P, Buades A, et al. Non-local MRI upsampling[J]. Medical Image Analysis, 2010, 14(6):784-792.
7. Greenspan H, Oz G, Kiryati N, et al. MRI inter-slice reconstruction using super-resolution[J]. Magnetic Resonance Imaging, 2002, 20(5):437-446.
8. Iwamoto Y, Han X H, Sasatani S, et al. Super-resolution of MR volumetric images using sparse representation and self-similarity[C]// International Conference on Pattern Recognition. IEEE, 2013:3758-3761.
9. Wang Y, Rahman S S, Arns C H. Super resolution reconstruction of μ -CT image of rock sample using neighbour embedding algorithm[J]. Physica A Statistical Mechanics & Its Applications, 2018, 493:177-188.
10. Li Z, Teng Q, He X, et al. Sparse representation-based volumetric super-resolution algorithm for 3D CT images of reservoir rocks[J]. Journal of Applied Geophysics, 2017, 144.
11. Hinton G E. Deep belief networks[J]. Scholarpedia, 2009, 4(5): 5947.
12. Kim J, Lee J K, Lee K M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:1646-1654.
13. Timofte R, Smet V D, Gool L V. A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution[C]// Asian Conference on Computer Vision. Springer, Cham, 2014:111-126.
14. Zhang TR, Teng Q,et al. Super-resolution reconstruction for 3D core CT images[J].Journal of Zhejiang University(Engineering Science),2018,52,7.
15. Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.
16. Ouyang W, Luo P, Zeng X, et al. DeepID-Net: multi-stage and deformable deep convolutional neural networks for object detection[J]. Eprint Arxiv, 2014.
17. Yang J, Wright J, Huang T, et al. Image super-resolution as sparse representation of raw image patches[C]// Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008:1-8.
18. Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation.[J]. IEEE Transactions on Image Processing, 2010, 19(11):2861-2873.
19. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.
20. Pascanu R, Mikolov T, Bengio Y. On the difficulty of training Recurrent Neural Networks[J]. 2012, 52(3):III-1310.
21. Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2015:1-9.
22. Zeiler M D, Fergus R. Visualizing and Understanding Convolutional Networks[C]// European Conference on Computer Vision. Springer, Cham, 2014:818-833.
23. Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE Trans Neural Network, 2002, 5(2):157-166.
24. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. 2015:770-778.
25. Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Trans Image Process, 2004, 13(4):600-612.
26. Sutskever I, Martens J, Dahl G, et al. On the importance of initialization and momentum in deep learning[C]// International Conference on International Conference on Machine Learning. JMLR.org, 2013:III-1139.
27. Nair V, Hinton G E. Rectified linear units improve restricted boltzmann machines[C]// International Conference on International Conference on Machine Learning. Omnipress, 2010:807-814.
28. Timofte R, De V, Gool L V. Anchored Neighborhood Regression for Fast Example-Based Super-Resolution[C]// IEEE International Conference on Computer Vision. IEEE, 2014:1920-1927.Yang J, Wright J, Huang T S, et al. Image super-resolution via sparse representation[J]. IEEE Transactions on Image Processing, 2010, 19(11):2861-2873.
29. Lécun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
30. Dong C, Chen C L, He K, et al. Learning a Deep Convolutional Network for Image Super-Resolution[M]// Computer Vision – ECCV 2014. Springer International Publishing, 2014:184-199.
31. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks[J]. Journal of Machine Learning Research, 2010, 9:249-256.