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

Yukai Wang Tingrong Zhang Junxi Feng

（College of Electronics and Information Engineering, Sichuan University, Chengdu, 610065, China)

**Abstract:** CT(Computed Tomography) imaging technology is widely used in geological exploration, medical diagnosis and other fields. In the practical situation, however, the resolution of CT image is typically limited by objective condition. SR(Super resolution) methods based on deep learning has obtained surprising performance in 2D images. Unfortunately, there are a few super-resolution algorithms for 3D images. In this paper, we proposed a novel network named 3DSRCNN (3D Super Resolution Convolutional Neural Network) to realize super resolution of volumetric CT images. To complete tasks with less number of CT samples, we explored the rules 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 residual-learning, gradient clipping, momentum SGD strategies to optimize procedure of training. Our single model can complete the duty of multi-scale interpolation reconstruction. Finally, the experimental results show that our single model can complete the duty of multi-scale interpolation reconstruction and have better performance in terms of PSNR , SSIM and efficiency compared with previous methods.

Index terms: super resolution, CT images, 3D-CNN, residual learning

1. **Introduction**

CT are widely used to provide detailed information for accurate post-processing. High spatial resolution CT sequence, however, generally need longer time to scan which causes degraded resolution in slice direction(Z axis). Recently, CT,micro-CT and nano-CT have been the most popular equipment to display real 3D rock sample images[1]. Establishment of accurate three-dimensional image of rock providing rich structure information helping geological researchers to analyze the physical properties of rocks[2][3] and play an important role in the field of geological and petroleum exploration. Due to their inherent limitations of CT devices, setting high resolution will not only increase the cost, but also the higher resolution will result in decrease of field of view(FOV) . Therefore, the use of super-resolution algorithm is an intriguing method to improve the resolution of CT providing more clear sample data for subsequent medical diagnosis or geological research.



Figure 1：CT image acquisition. 3D image are composed of a set of 2D slices. The resolution in the slice direction is commonly much lower than in the in-plane directions.

Currently, SR (super-resolution) reconstruction, having drawn extensive attention in computer vision field, is effective method to improving quality of image. Learning-based algorithm have been typically used to learn the mapping relationship between LR and HR for SR.

Deep learning technology[11]has drawn more attention, and it is much better than conventional machine learning algorithms in task of feature extraction, image segmentation, target detection and others. Chao Dong raised a network SRCNN[4] which only contains 3 layers CNN structure but outperform than former method. However, the structure of SRCNN have limitations: first, single model works only for single scale; second, training of SRCNN converges too slowly. Jiwon Kim found that deeper network structure had a stronger generalization ability than shallow ones and proposed VDSR[12] to resolve these issues.

Although Jiwon Kim have showed that the increasing network depth will enhance the generalization ability of the network[12], the pipeline is not entirely applicable in CT images. First , our training samples are not as convenient as 2D images to get. Moreover, we attempt to fulfill super-resolution reconstruction of rock CT images that commonly contains complex pore structures , which bring difficulties to restore. Aiming at issues that the texture of the interior rock is complex.

The amount of 3D image data is far greater than the two-dimensional image, and the method of two-dimensional method can not be directly transferred to 3D model. In training network and reconstruction, the space complexity and time complexity have to be taken account to ensure our work can be carried out on the general computing equipment. At the same time, we need to consider the problem of gradient exploding during training process.

In current research, scholars mainly focus on single 2D image reconstruction rather than spatial 3D Voxel. Currently, more SR research for 3D images mainly focuses on MRI(Magnetic Resonance Imaging). H. Greenspan archived super resolution in slice direction using an iterative algorithm which gives improved resolution[6]. Manjon proposes a non local MRI sampling method[5] to recover high frequency information by using a data-adaptive patch-based reconstruction in combination with a downsampling coherence constraint. Iwamoto proposes a method based on sparse representation and self-similarity to improve the slice direction resolution of MRI[8]. This method only improves the resolution in the slice direction, and does not have effect in plane direction.

Yuzhu Wang[9] using neighbour embedding algorithm to improve resolution of CT image of rock sample. Li proposed a voxel SR reconstruction algorithm[10]based on sparse representation , which increase the resolution in all directions retaining FOV no decrease. Li also introduce BM4D, eliminating the influence of additive noise , to denoise signal noise when training dictionaries.

Zhang extend A+(Adjusted Anchored Neighborhood Regression) to 3D image and proposed 3D-A+ [13], establishing a correlative dictionary between high frequency and low frequency block. The matched dictionary atom and mapping matrix were searched for each input of the 3D block in reconstruction stage.

We attempt to enhance super resolution from three directions (x, y ,z) for CT images of rock. In real structure of rock samples, there are filled with a number of pore and natural granular siliciclastic. A rich texture can be seen from CT images, which brings difficulty to SR.

Motivated by 2D SR methods based on deep learning, we propose a novel 3D neural network to promote resolution for 3D voxel. To cope with aforementioned issues, we use 12 layers CNN network which contains 64 three-dimension convolutional kernel. To cope with a series of problems caused by the deeper network , we employ residual learning and gradient clip strategy to accelerate the convergence. Our network does not need to train multiple models and single model performs as well as a method using multiple networks trained for each scale. In order to evaluate performance of 3DSRCNN , we evaluate the reconstruction effect by PSNR and SSIM . Before training, cropping samples is necessary for a promising result. Using trained model to reconstruct, it is observed that 3DSRCNN has better average PSNR and SSIM on different testsets, while ours have faster reconstruction speed.

In summary, we introduce 3DSRCNN to solve the super-resolution 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 crucial. We also 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 that it have achieved a trade-off between the accuracy and speed. Moreover, We have addressed the problems existing in procedure of coping with CT images such as problem of insufficient 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. Section2, we primarily introduce the concept of SR and the implementation of deep learning. Our proposed network--3DSRCNN is carefully described in Section 3. In section 4, we experimentally investigate how to design the network and to pursuit better performance between accuracy and speed. Besides, we also test and compare our method to others. In Section 5, analyst and conclusion are given for future studies.

1. **Related work**

**2.1 Image Super-Resolution**

SISR(Single Image Super Resolution) is an ill-posed problem due to lacking of detailed information in the process of downsample. There are two traditional methods to restore LR(Low resolution) images to HR(High Resolution) 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 samples. The process of SISR(Single Image Super Resolution) is that, for a ground truth image (HR images) X, we first downgrade it to the LR images Y.

 (1)

Our goal is to find a function as F(x) , which can restore LR to HR in some certain. SR reconstruction based on learning method is to learn the mapping relation of low frequency information to high frequency information by training mapping dictionary, meanwhile, sparse coefficients are calculated so as to retrieve HR features from LR. The sparse-coding method is a representative learning based method composed of three steps:(1)LR features extraction;(2)Learning mapping relation in LR features patch and HR patch;(3) reconstruct with the mapping relation.

**2.2 Convolutional Newtwork for Super-Resolution**

Chao Dong consider deep convolutional neural network is equivalent to sparse-coding method[4], which can directly learns an end-to-end mapping relation and proposed a novel network termed as SRCNN which outperforms than traditional algorithm.

While SRCNN archive good result in 2D image datasets, there are following limitations in three aspects:(1) SRCNN cannot apply a trained model with specific scale trainset for other scale;(2)Its network structure cannot be applied to spatial 3D super resolution;(3)We find SRCNN have margin improvement than previous algorithm.

Jiwon Kim introduce VDSR to address the limitations in Chao Dong’s work 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 total to 20 layers to complete whole super-resolution. Jiwon have experimentally validated the view –‘the deeper, the better’ in task of 2D images SR. 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 small number of samples to complete the training of the network as far as possible.

1. **3DSRCNN**

In this chapter, we introduce the structure of the network that enhance resolution in plane and slice direction. Besides, some strategies for optimizing training process are employed to our network 3DSRCNN. Next, We describe production of training data in detail.

**3.1 Network Structure**

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



Figure2： 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. Residual image  is label to calculate MSE in pixel wise.

For volumetric super resolution, we employ a network composed of 12 layers CNN 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 patch; The last layer combine learned high frequency information from middle layers and initial LR image to finally formulate super resolution image.

The convolution network actually extracts spatial related information which contains diverse pattern features. When the input image continues to pass through the CNN, the extracted feature becomes global and has a larger receptive field (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 typically have better results as Jiwon pointed in their work. 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. We make a trade-off between complexity and reconstructed accuracy . Unlike SRCNN that have no padding in covolutional operation causing boundary pixel discard. However, padding is necessary for our network because that processing three-dimension image typically comes at the expense of a lot of memory. In order to be able to train in general device, we divide the initial CT image to sub-blocks with size 25x25x25. Given that the layers of network is 12, the input size is relatively small, which will cause majority loss of prior information without padding during forward propagation. To cope with this problem , we use zero padding and subsequent experimental have proved the correctness of this trick.

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 the network depth is increased, the information of the input LR feature will be lost in the continuous convolution process which results in training unstable, and the residual learning can be used to solve the above problems. After each CNN, we utilize ReLU(Rectified Linear Unit)[24] as activation function on output of last layer.

 (3)

Where x,w denote input and weight parameter of last layer respectively, b is bias. In practical process, input is added to output of network as final output to compute loss function.

**3.2 Pre-process of training set**

Before training, we should crop and transform initial CT sample to suitable shape. Specifically, the large blocks are first divided into sub-blocks through slicing in x,y,z direction, respectively.

We first use x2,x3,x4 factor to downsample ground truth CT datasetswith original size 400x400x400 in different samples, then we resize the volume to the original size as input datasets using bicubic interpolation algorithm. We crop initial volumetric CT blocks to sub-3D-blocks to produce training set. HR images{Y} is viewed as label to calculate MSE, and LR images {X} are fed in network.



Figure 3: Procedure of making training set.

With respect to CT images, crop is significant for training and there are mainly three points as following:

1. In this way, a larger quantities of training sample can be obtained through image cropping under the condition of limited number of CT samples. In this way, these sub-blocks are viewed as small size‘images’ rather than ‘patch’
2. Cropping promises practical feasibility in general computer due to 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 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 : , Where  is 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 is used as input and label for training. Through experiments, we find setting  as is appropriate.

**3.3 Training strategies**

Our proposed network is constituted of massive tensors which represent end-to-end mapping relationship. Tensors in network is initialized by Gaussian distribution(zero mean and standard deviation 0.001). Through continuous iterative training, weight parameteris increasingly optimized for minimizing 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. We employ following strategies to accelerate training.

**Residual Learning**

Recent study demonstrate that depth have significant effects on network generalization for image classification[16]. Stacking more layers may help enrich low level feature to high level[19], but degradation problem has been exposed when the network depth increasing. We find gradients vanishing/exploding will appear when depth exceed 20 layers and deeper model produce higher training error than its shallower counterpart. Jinwon Kim consider that a set of covolutional operation in deep layers, input detail increasingly is discarded after passing covolutional operation in deep layers[12], which leads to the output can only use learned features. He Kaiming have introduced a deep residual learning framework[21] and got excellent scores in image recognition. Consequently, residual-learning strategy is adopted in our network. We define input as x, output as y,and residual image , where  denote output of data passing through network .

Given training set , and cost function based on MSE is interpreted as following:

 (4)

Where n is the number of training batch sample, i denote current input data.

One point 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.

**Adjustable learning rate**

In SRCNN, it is found that the loss rate with small learning rate converges very slowly . High learning rate help to boost training but can lead to gradients exploding. We use the adjustable learning rate to speed up the training. In previous Epoch, setting relatively high learning rate to accelerate training. As training process going on, learning rate is reduced by step with following .

 (5)

Where epoch denote current training epoch, and step is predefined to control decay of learning rate.

**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 work[23]. 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 |
|  |

Where p,g,v anddenote the weight parameters in network,gradients, velocity, and momentum respectively.

**Gradient clipping**

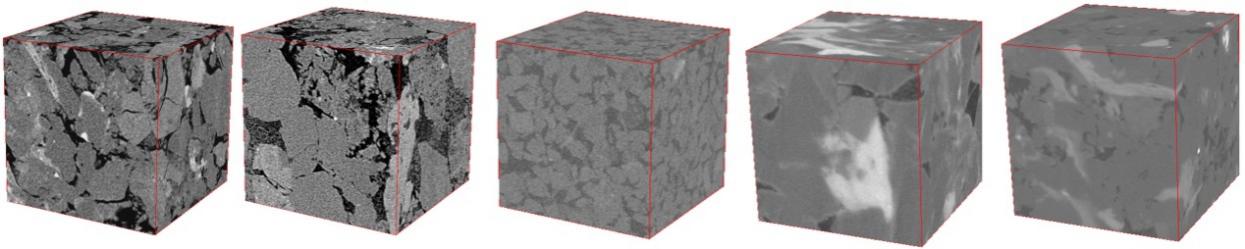
Gradient clipping[17] is usually applied in training RNN network in case of gradient exploding/vanishing. One simple way is predefining a threshold to clip them whenever they go over a threhold. In VDSR, Jiwon[12] use this technique to limit gradients to a certain range. In our work, We directly clip gradients to range, where  is predefined clipping range.

1. **Experiments**

In this section we first introduces experimental basis and evaluation, PSNR and SSIM which is widely used to evaluate image quality. Next, we describe specific steps of producing training set in detail. We investigate the influence of important parameters on the accuracy of reconstruction, and intuitively analyze the reasons.

**4.1 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 test samples are different from training set which are consistent with the identical selection rules.



(a)sample1 (b)sample2 (c)sample3 (d)sample4 (e)sample5

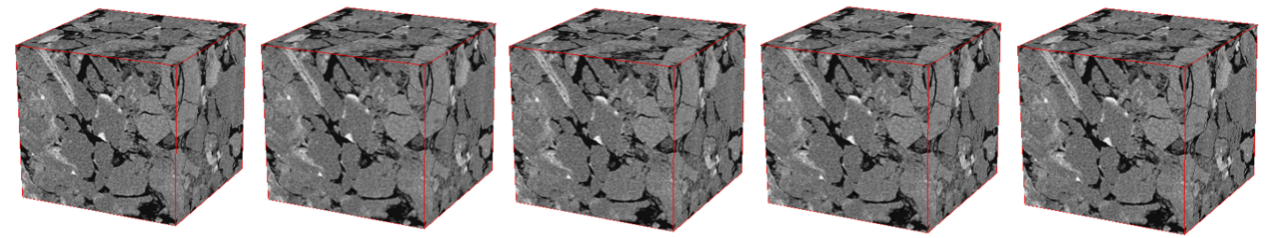
Figure 4: Trainning CT samples (a)(b)(c)are Sandstone with resolution of 3.8 (d) is Dense carbonate rock with resolution of 1.07 (e) is Dense sandstone with resolution of 1.07

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

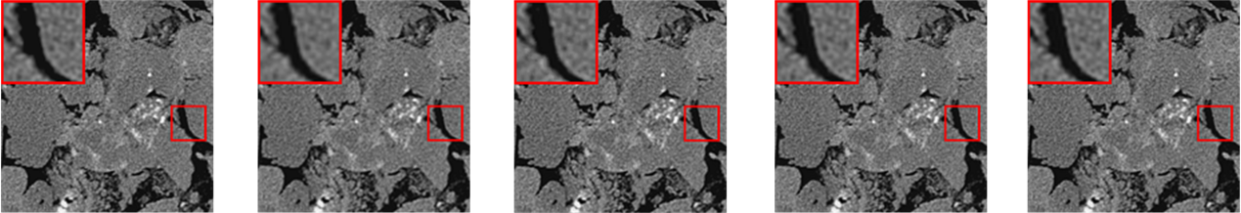
 (6)

 (7)

The higher PSNR value denote reconstructed CT images have better quality. PSNR is a evaluation based on pixel error, whereas, it does not take into account the visual features of human eyes, which may cause that the evaluation results are not consistent with the perception of people. Wang use SSIM[22] to represent the structure information of the image from the brightness, contrast and structure of the image, and the structure similarity model 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 sclice （g）Bicubic interpolation （h）MRI （i）3DA+ （j）3DSRCNN

Figure5: SR results of different algorithm in scale factor x4 testset

**4.2 Work platform Details**

Detailed Hardware is listed as following Table 1. In order to ensure the normal operation of the experiment, It must be guaranteed that computing device has enough memory at least 8Gb. In general, CPU training is very slow, we use CUDA to invoke GPU resources speeding up training.

**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.

**4.4 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 upscalehave better performance on different samples because it extracts patch from different upscale factor(). We started the following experiments: comparing the training process with the multiscale and single,respectively. In training stage, it is found that average MSE with multi-scale, is higher than single upscale in the early Epoch, and the overall convergence process is slower as either. Interestingly, After training 35 epoch, average PSNR of multi-scale surpass counterpart. training in ,scale ,0.12dB and 0.21dB, and only a little less 0.1dB in x3 scale. Better than a single model. We consider that the use of the multi-scale set is more preferable.

**4.5 Analysis of the process of training and reconstruction**

If training sets is blended up with multi-scale sample, the yielded model can be applied for different interpolation image.In Figure 7, it is observed that number of training Epoch to converge is not identical in different upscale testsets. In general, with the increasing number of iterations, the generalization ability of the network will also increase, but it will eventually converge to a certain number.

But not the low MSE value in training stage means that the reconstruction effect is better. MSE is based on one batch randomly selected in the training set, not to the real whole image reconstruction, and not the point that the average MSE small model in training has better reconstruction effect, either. It is supposed to use trained model to complete reconstruction and calculate PSNR to compare.

As shown in Figure 7, on the different upscale factor training sets, 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 in the process of using momentum SGD optimization, the loss function may fall into a bad local minimum, resulting in a slight drop in the PSNR during the training process. In, it is evident that the improved range reduced gradually. On , , and  times, the average PSNR increases respectively in 2.89dB, 2.42dB, 1.50dB, and SSIM, respectively. This is because the use of a higher downsample factor will result in a large number of original high-frequency details missed, and the quality of the final reconstruction will be limited. Through the specific experiments, 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 2.29h for each training Epoch. The time cost and reconstruction precision are considered at the time of setting up the number of Epoch.

When training multiple scale sets, the trend of PSNR curve in each interpolation scale is not simultaneous. This is because batches randomly taken out from whole training set exerting different effect to three interpolation scale. 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 show , PSNR curve converges at about 15 Epoch at given setting, and , X2, X3 and X4 improve 2.89dB, 2.42dB, 1.50dB than input images, respectively. In higher upscale 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.

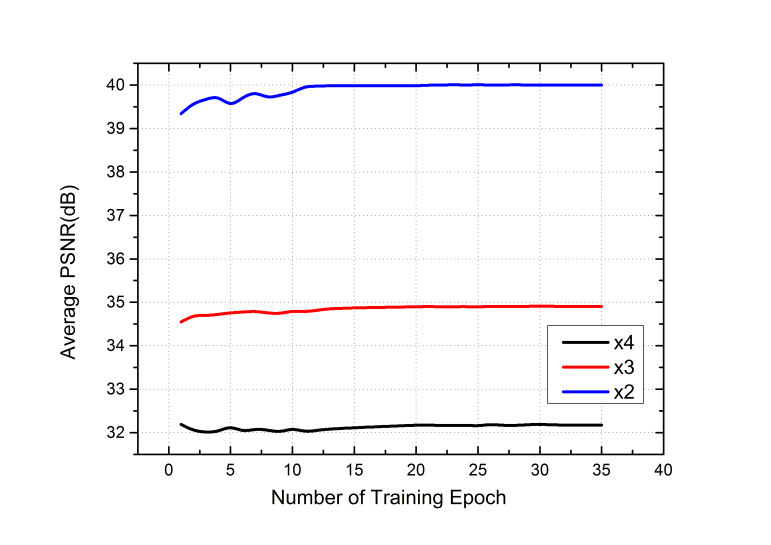


Figure 6: The test PSNR curve of 3DSRCNN on testsets with ,upscale factor

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 on the size of the low resolution image sent to network model. The ultimate goal of our work is to restore low resolution 3D block, but if directly send the corresponding size of the image to the network need to consume a lot of memory about 352Gb, which is difficult to achieve in practice workspace. Therefore, we also crop input CT images to smaller sub block with size of 100x100x100.

the effect of the proposed network on the reconstruction quality is experimentally analyzed compared with the traditional bi-cubic interpolation ,sparse representation based method, 3D-A+. Experimental results have proved that the our algorithm has exceeded previous algorithm in reconstruction speed and accuracy.

The experimental results in this paper are evaluated using PSNR and SSIM. For a three-dimensional image with a size of 100, the calculation is as follows:

**4.6 Training parameters and Trade-offs**

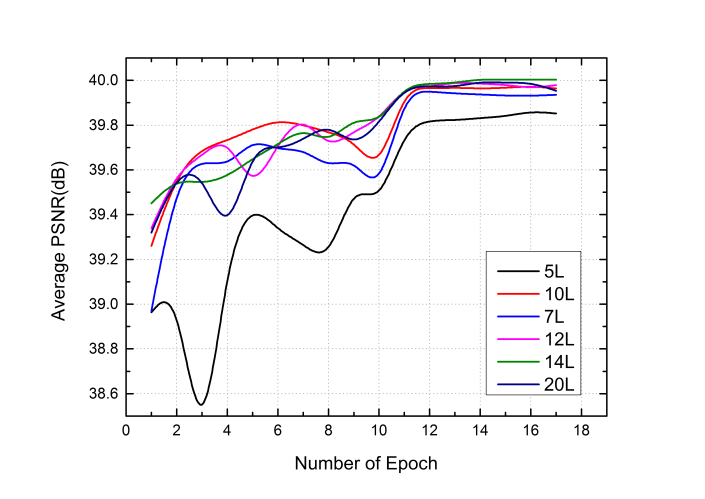
Network parameters including network depth and convolution kernel size will affect the accuracy and training time. In most cases, increasing training Epoch is conducive to training a better performance. The mapping relationship that a network can learn from a given training set is yet limited to data quantity and structure of network. The value of PSNR will converges to a certain number, which is affected by predefined parameters. Appropriate increase of the network structure and training parameters can make the network have better performance. In this section, we investigate the optimal trade-off between performance and speed.

**4.6.1 Depth of network**

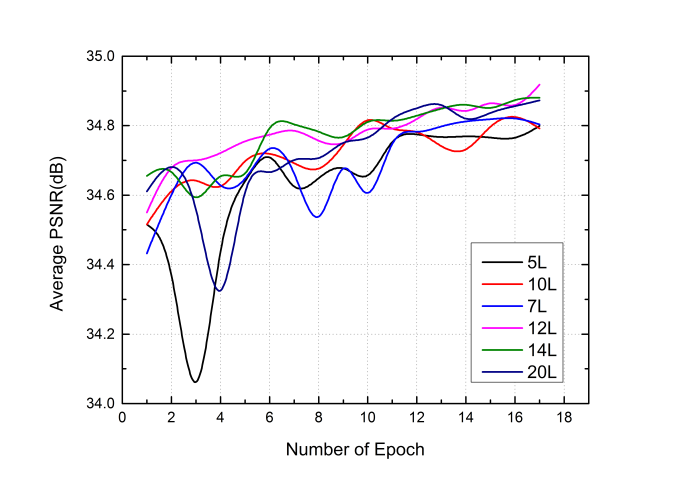
For super resolution in single 2D images, Jiwon Kim demonstrate the large depth help model to capture more contextual information and yield better performances than shallow ones. For the training of 3D samples, the amount of computation and memory occupied is very large. Too many layers could slow down the convergence and improve the time complexity. Through experimental verification, 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 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 overfitting. Time spent on training is shown as Table 2. We observe that the depth of the network will increase the computational complexity resulting in time-consuming. In conclusion, Network complexity should be proportional to the magnitude of training data, and moderate depth will help to improve network generalization ability.

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

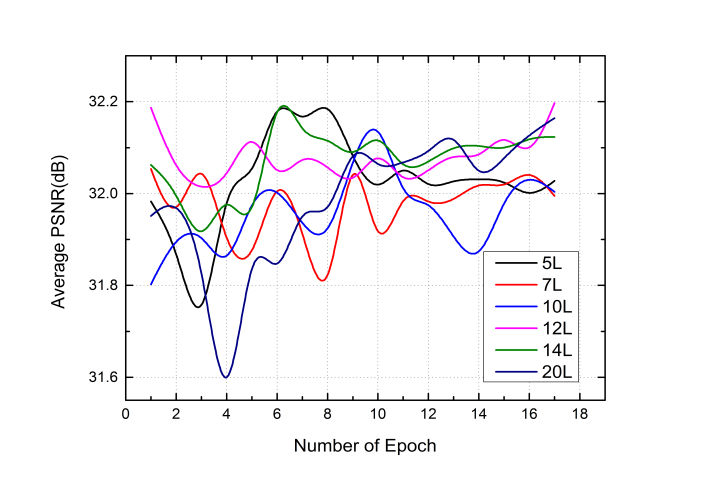
The 12 layer is about 0.1,0.2 dB higher than the shallow ones in x3,x4 factor and training curve is also relatively stable than others. However 20 layer is slightly lower than that of the 12 layer in . This issue is also mentioned in [4], where improper increase of depth will cause degradation of SR 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 data samples and increasing the network by experiment will help to get better results.



1. x2



1. x3



(c)x4

Figure 7: PSNR curve of different layers networks on testsets with ,upscale factor.

**4.6.2 Convolutional kernel size**

In this section, we explore network sensitivity to different convolutional kernel size. Due to the limitation of GPU memory, the size of the input data is set to. Based on previous experiment, number of layers is set to 12 and we conduct three comparative test 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, using has 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 | 3x3x3 | 5x5x5 | 7x7x7 |
| Time(min) | 81 | 250 | 428 |
| PSNR(dB) | 40.00366466 | 39.74091643 | 36.78703898 |

**4.6.3 Residual vs Non-residual Learning**

Increasing network layer could cause the problem of the gradient explosion/vanishing which bring difficulties to training. Fig 6. display difference between the use and non use of residual learning. Convergence will become slower without 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, whole curve fluctuates greatly, PSNR has a large decline in 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 X3 and X4 is about 0.2dB, 0.1dB higher ,0.4dB lower than counterpart on x2 yet. Using residual learning can enhance the stability of training. When the network is deeper, the necessity of residual learning will be more prominent.

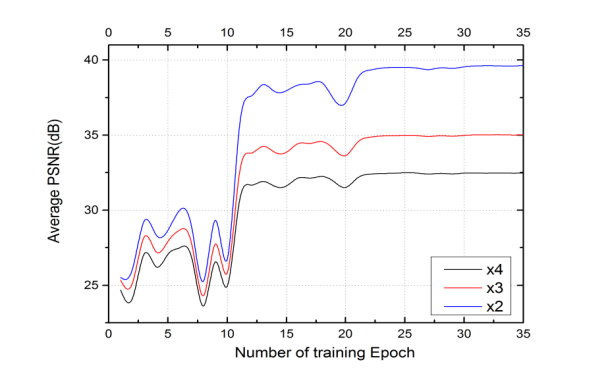


Figure8: PSNR curve with non-residual learning. It’s obvious to find that network with non-residual emerge a shock wave in early stage. Curve tend to converge until training 25 epoch.

**4.7 Comparison to other method**

Through the experimental data in the Table 3, we can see that the 3DSRCNN training model is better than the previous one on three different magnification test sets. 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 the 3x3x3 size 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

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Eval | Scale | Bicubic | MRI | Sparse representation | 3DA+ | 3DSRCNN(single) | 3DSRCNN(multi) |
| PSNR(dB) | 2 | 36.79 | 36.35 | 38.45 | 38.58 | 39.68 | 40.00336 |
| 3 | 32.61 | 33.15 | 34.69 | 34.89 | 35.03 | 35.0179903 |
| 4 | 30.51 | 30.17 | 31.25 | 31.65 | 32.01 | 32.46729689 |
| SSIM | 2 | 0.950 | 0.965 | 0.983 | 0.983 | 0.986 |  |
| 3 | 0.879 | 0.881 | 0.924 | 0.930 | 0.931 |  |
| 4 | 0.738 | 0.799 |  | 0.828 | 0.850 |  |

**5. Conclusion and future work**

In our work, we have proposed a novel method, 3DSRCNN, based on deep learning to approach voxel images. While using CNN to restore single low resolution image to high resolution have gained 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 cope with 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. Our single model trained by samples mixed with diverse CT sets can be applicable to work intestsets. 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 work such as based on sparse-presentation , A+ methods in experimental results.

**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.CT 用于地质领域
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.micro CT用于多孔介质
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.CT 用于展示微观特性
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. SRCNN超分辨
5. Manjón J V, Coupé P, Buades A, et al. Non-local MRI upsampling[J]. Medical Image Analysis, 2010, 14(6):784-792. MRI图像超分辨1
6. Greenspan H, Oz G, Kiryati N, et al. MRI inter-slice reconstruction using super-resolution[J]. Magnetic Resonance Imaging, 2002, 20(5):437-446. MRI图像超分辨率2
7. Wang Y H, Qiao J, Li J B, et al. Sparse representation-based MRI super-resolution reconstruction[J]. Measurement, 2014, 47(1):946-953. MRI稀疏表示超分辨率重建
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. 稀疏标识
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.岩石CT超分辨重建邻域嵌入
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.大师兄SCI引用
11. Hinton G E. Deep belief networks[J]. Scholarpedia, 2009, 4(5): 5947. 深度学习引出
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.VDSR
13. Zhang TR, Teng Q,et al. Super-resolution reconstruction for 3D core CT images[J].Journal of Zhejiang University(Engineering Science),2018,52,7 张廷蓉论文
14. 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.稀疏表示超分辨论文
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. 深度学习广泛应用-图像分类
16. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014. 深层次网络有利于效果1
17. Pascanu R, Mikolov T, Bengio Y. On the difficulty of training Recurrent Neural Networks[J]. 2012, 52(3):III-1310.梯度裁剪
18. 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. 加深网络层数增强效果2
19. Zeiler M D, Fergus R. Visualizing and Understanding Convolutional Networks[C]// European Conference on Computer Vision. Springer, Cham, 2014:818-833. 网络深度增加，特征高级,网络深度效果3
20. 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. 网络层数深的坏处
21. He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. 2015:770-778. 残差网络引入
22. 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. SSIM的介绍
23. 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.动量SGD加速
24. 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.RELU的介绍
25. 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. A+支撑ztr3DA+
26. 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. 稀疏表示支撑李SCI
27. 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. 梯度下降SGD
28. 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. Dongchao深度学习超分辨历程
29. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks[J]. Journal of Machine Learning Research, 2010, 9:249-256. 网络层数加深坏的影响2