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

Yukai Wang Tingrong Zhang Junxi Feng

（Institute of electronic information, Sichuan University, Chengdu 610065, China )

**Abstract:**CT imaging technology is widely used in medical diagnosis, geological exploration, material structure analysis and other fields. In the practical application process, however, the resolution of CT image is typically limited by objective condition. Image super-resolution is a research hotspot in the field of computer vision in recent years. Super resolution methods based on deep learning has obtained surprising results. However, there are rare super-resolution algorithms for 3D images. Regarding to volunmetric image, we proposed a novel network based on 3D-CNN(Convolutional Neural Network) to realize super resolution. To cope with limitation of number of CT samples, very slow convergence of network training, as well as insufficient memory , 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 multiscale interpolation reconstruction. Finally, the experimental results show that our single model can complete the duty of multiscale interpolation reconstruction and have better PSNR and SSIM performance than previous methods, moreover, the speed of our reconstruction extremely surpassed previous method.

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

1. **Introduction**

CT(Computed Tomography) are widely used to provide detailed information for making a diagnosis. High spatial resolution CT sequence, however, generally need longer time to scan, which means patients receive more does of radiation and it also causes degraded resolution in slice direction. Establishment of accurate three-dimensional image of rock help the geological researchers to analyze the physical properties of rocks and play an important role in the field of geological and petroleum exploration. Recently, computed tomographies (CT),micro-CT and nano-CT have been the most popular equipment to display real 3D rock sample images13, but setting high resolution will not only increase the cost, but also the higher resolution will result in decrease of field of view(FOV) 12. Therefore, the use of super-resolution technology to improve the resolution of CT images can provide more clear sample data for the next medical diagnosis or geological research.

Currently, 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 super resolution.

Deep learning technology has excellent performance, and it is much better than traditional machine learning algorithms in feature extraction, image segmentation, target detection and other tasks. Chao Dong raised a network SRCNN which only contains 3 layers CNN structure but outperform than traditional method. In his paper, he consider SRCNN is equivalent to sparsing-coding method and directly learns the end-to-end mapping relationship between LR and HR.

However, the structure of SRCNN is too simple to capture more contexture information and trained SRCNN model works only for a single scale interpolation. Jiwon Kim further studied these issues based on SRCNN and he found that deeper network structure had a stronger generalization ability through a large number of experiments and data.

In current research, scholars mainly focus on two-dimension images rather than spatial 3-dimension voxel. At present, more 3D-voxel based super-resolution research focuses on MRI's super resolution reconstruction. Manjon proposes a non local MRI sampling method, and uses non local similarity regularization to iteratively solve MRI images reconstruction. Iwamoto proposes a method based on sparse representation and self similarity to improve the direction resolution of MRI slices. This method only improves the resolution in the slice direction, and does not have super-resolution effect in plane.

Li proposed a three-dimensional super resolution reconstruction algorithm based on sparse representation, which can learn the mapping of low frequency information to high frequency information by training mapping dictionary, so as to combine the low frequency block and complete the reconstruction with the original low frequency information.

Zhang proposed 3DA+(3D Adjusted Anchored Neighborhood Regression) methods ,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 of image x, y ,z in 3D CT image of rock. In real structure of three-dimension rock sample, 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 three-dimension reconstruction

Motivated by 2D SR methods based on deep learning, we propose a novel 3D neural network to practically promote resolution for 3D voxel. To cope with aforementioned issues, we use 12 layers CNN network which contains 64 three-dimension convolutional kernel. At the same time, to cope with a series of problems caused by the deeper network , we employ residual learning and gradient cutting strategy to accelerate the convergence. Previous methods need to separately training dictionaries for different scales. Our trained single model can be applied to multiscale reconstruction.In order to evaluate performance of 3DSRCNN , we evaluate the reconstruction effect by PSNR and SSIM . Before training, segmentation and preprocessing of 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 scale testset, while ours have faster reconstruction speed.

In summary, we introduce 3DSRCNN to solve the super-resolution reconstruction of 3D CT images. In view of the actual sample condition, we make a corresponding adjustment to the network structure, so that it have achieved a trade-off between the accuracy and speed of reconstruction. Moreover, We have experimentally investigated and addressed the problems existing in procedure of super-resolution of CT images. Compared with previous algorithm, our method performs better in time and reconstruction accuracy.

**Related work**

**Image Super-Resolution**

Image super-resolution inherently is a ill-posed problem due to lacking of detailed information in the process of interpolation. There are two traditional methods to restore LR(Low resolution) to HR(High Resolution), one is using context correlation in LR image yet has inborn defects that it cannot obtain more specific high frequency information; The second is learning based method, which learns the mapping relation between LR and HR through substantial training. The sparse-coding method[] is a representative learning based method, which is mainly composed of three steps:(1)LR features extraction;(2)Learning mapping relation in LR features patch and HR patch;(3)Completion super resolution using the aforementioned relation. Chao Dong in his paper proposed a CNN network termed as SRCNN to fulfill aforementioned pipeline and SRCNN outperforms than traditional algorithm.

While SRCNN archive good result in 2D image datasets, there are following limitations in different aspects:(1) SRCNN cannot apply a trained model with specific scale trainset for other scale;(2)Its network structure cannot be well adapted to three-dimension super resolution;(3)We find the performance of SRCNN only have a little improvement than traditional algorithm A+.

Jiwon Kim raise a different mechanism to address the limitations in Chao Dong’s work. Jiwon Kim points that adding CNN layer lead convolutional filters become increasingly global, which conceptually benefit to learn mapping relation , and utilize deeper network structure summed up to 20 layers to complete whole super-resolution. With extensive experiments, Jiwon have validate the view –‘the deeper, the better’.

Sparse representation-based algorithm

As for 3D volumetric super-resolution, Zhengji Li introduce a spase represention-based method to enhance resolution of CT images in xyz direction.

3 Three-dimension Neural Network for CT image super resolution

3.1 Network Structure

The real rock commonly contains complex pore structures as following images show. Aiming at issues that the texture of the interior rock is complex, we proposed a three dimension network structure, termed as 3DSRCNN to archive super resolution for volumetric CT images.

The convolution network actually extracts local related information in space. Each convolution kernel extracts local pattern features. Through the gradual increase of the network, the information extracted by convolution operation has a larger receptive field (receptive field). The deeper structure ensures that the parameters learned by the convolution kernel have a stronger response to the input information.

Consequently, the depth of network layers will affect the reconstruction accuracy and training time. Due to reconstructed objects containing rich texture feature 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 is , where is the depth of CNN layers, is feature map size, identify the current layer number, is number of channels. It is obvious to find that dense network structure would increase the computational complexity. We make a trade-off between complexity and reconstructed accuracy and

we consider 12 layers CNN is optimal. Unlike practice in SRCNN that have no padding in covolutional operation, we use zero padding during convolution in case that the original data is lost in the convolution.

Due to insufficient memory, the size of input blocks sent to the network cannot be too large.

For volumetric super resolution, we employ a network composed of 12 layers CNN each of which has 64 channel(feature maps) .The first layer is responsible to extract low frequency patch from LR images.The middle 10 layers complete function of learning mapping relationship between LR and HR volummetric patch. The last layer combine learned high frequency information from middle layers and initial LR image to finally formulate super resolution image.

Because the SRCNN network has only three layers of network, it also completes the forward propagation of the LR image while the convolution extracts feature blocks. When the network depth is increased, the information of the input LR image will be lost in the continuous convolution process, and the residual learning can be used to solve the above problems. After each CNN, we utilize ReLU(Rectified Linear Unit) as activation function on output of last layer.

Where x,w denote input and weight parameter of last layer respectively, b is bias.

With a set of covolutional operation in deep layers, input detail increasingly is discarded, which results in the output can only use learned features. Meanwhile, gradients vanishing/exploding will appear in this training process. He Kaiming have introduced residual learning framework to address the issues. Consequently, residual-learning strategy is adopted in our network. We define input as x, output as y,and residual image . The loss function is formula  , where denote output of network . In practical process, input additionally is added to output of network as final output to compute loss function. Here we use zeros operation to ensure that the size of input data will not change after convolution.

In order to eliminate the boundary effect, SRCNN have no padding when training, but it would make the output result smaller in size. If network structure is deep, and the input size is relatively small, volume of input data will shrink in the during forward propagation that will cause the loss of detailed information of original data.

3.2 Pre-process of training set

We use the segmentation method to enlarge CT training set. Specifically, the large blocks are divided into sub-blocks through slicing in xyz direction respectively. Setting appropriate stride in slicing CT image leads to dense block overlap , which is favor of supplying large and abundant texture information.

We first use x2,x3,x4 factor to downsample initial CT dataset  with original size 400x400x400 in respective samples. Then we restore the volume to the original size as input datasetusing bicubic interpolation. Before training phase, we crop initial volumetric CT blocks to sub-3D-blocks, which is critical for training and there are mainly three points as following:

1. In this way, a larger sample can be obtained through image segmentation under the condition of limited CT samples.
2. This trick promises practical feasibility in general computer due to Training 3D-block will occupy amounts of memory. A large block is cut into small blocks, enabling computing devices to calculate faster.
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 segmentation can be calculate with following formula:

Where  denote initial size,  is sub-block size,stride is span length when cropping. Setting suitable  and  has important influence on training time and accuracy.

After the segmentation with aforementioned dataset, paired  is used as input and label for training, respectively.Through experimental, we find setting as is appropriate.

3.2 Training

Before training, we n

Our proposed network formed by amounts of tensor inherently is a function which represent end-to-end mapping relationship. Tensors in network is initialized by Gaussian distribution with zero mean and standard deviation 0.001. Through continuous iterative training, weight parameter is increasingly optimized for minimizing loss function L . Given training set , and based on MSE is defined as following:

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

SRCNN minimize the objective function with classic SGD(stochastic gradient descent) in backpropagation.SGD random gradient descent method is a common optimization algorithm. The weight parameters of the whole network are updates by a small scale sample.

The concrete steps are as follows: after each training, a small batch of samples are extracted from the training center evenly, and the number of batch samples is far less than the training sample set, and the parameters of the gradient update network are used. M samples taken out of the batch. Gradient estimation can be expressed as

However, directly using traditional SGD takes long time to converge. We employ some strategies for our network structure and training data.

Adjustable learning rate

In the actual training process, it is found that the loss rate converges very slowly. We use the adjustable learning rate to speed up the training. In previous training Epoch, setting relatively high learning rate help accelerate convergence.As training possess going , learning rate is reduced for optimal solution.

Momentum acceleration

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[](On the importance of initialization and momentum in deep learning). The gradient update formula is written as

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

Gradient clipping

Gradient clipping is usually applied in training RNN network in case of gradient exploding/vanishing. Jiwon[] consider this method either can be applied to CNN to speed up convergence. We clip gradients to range , where is the current learning rate, is predefined clipping range.

Weight normalization

Weight normalization[] is a commonly used method that is widely used in deep for improving the quality of optimization. Thereby, We consider network is trained with weight normalization

Residual Learning

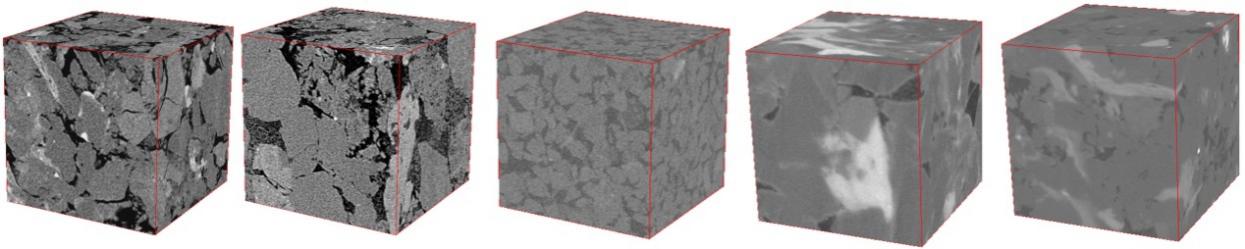
A degradation problem has been exposed when the network depth increasing ,accuracy

Experiments

In this section we first introduces experimental basis and PSNR and SSIM which is widely used to evaluate reconstruction accuracy. Nest, we describe specific steps of making training set, and segmentation trick of CT image in detail. We investigate the influence of important parameters on the accuracy of reconstruction, and analyze the reasons.

**Experimental datasets and evaluation Criteria**

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 different from training set are selected according to the order of PSNR increasing sequentially.



 Results of reconstruction are evaluates by PSNR and SSIM. PSNR is widely used to measure the quality of image restoration. For a three-dimensional image, the calculation of PSNR is as follows:



The higher the PSNR value the smaller the error between the original and the reconstructed result , that is .PSNR is a method of image quality evaluation based on pixel error. 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 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 subjective sense of the human performance than the PSNR. Range of SSIM value is [0,1], and the higher the SSIM indicates the closer it is to the actual samples.

**Workplatform Details**

Detailed Hardware is listed as following Fig. In order to ensure the normal operation of the experiment, we must ensure that the computing device has enough memory at least 8Gb. CPU training is very slow. We can use CUDA to invoke GPU resources to speed up training.

|  |  |
| --- | --- |
| CPU | Intel i7-6770K 4.0G Hz |
| RAM | DDR4 16GB |
| OS | Microsoft windows 7 |
| GPU | Nvidia GTX 1080 |

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

**Reconstruction**

The trained network model is essentially 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 you 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 segment original CT to smaller size of .

In this section, the effect of the proposed network on the reconstruction quality is experimentally analyzed compared with the traditional bicubic 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:

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

Many traditional super-resolution algorithms require different model to be applied to the corresponding samples. A single model enabling to be implemented into multi-scale scenarios is critical for practice work. We consider model trained with mixed scale data has an effect on different samples because it extracts patch from different scale(x2,x3,x4) and establishes mapping relationship. We started the following experiments: comparing the training process with the multiscale and single,respectively. It is found that in the training process with multiscale, the MSE of is higher than counterpart in the previous Epcoh, the overall convergence process is slower as well. Interestingly, average PSNR of reconstructed image surpass counterpart after 35 Epoch training in x2,x4 scale ,0.12dB and 0.21dB, and only a little less 0.1dB in x3 scale. Better than a single model.Through experiments, we think that the directly use of the multiple training set is more beneficial.

**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 modification 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 efficiency.

**Training analysis**

When training sets is blended up with multi-scale sample, the yielded model can be applied for multi-scale work. We observe that required number of training Epoch is not identical in different scale 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.

As shown in Figure 4.2, on the different interpolation multiplying 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 the saddle point, resulting in a slight drop in the PSNR during the training process. With the increase of the scaling ratio, it is obvious that the quality of the three-dimensional superresolution can be reduced significantly. On X2, X3, and X4 times, the average PSNR increases respectively in 2.89dB, 2.42dB, 1.50dB, and SSIM, respectively. This is because the use of a higher proportion of down sampling will result in a large number of original high-frequency details missing, and the quality of the final reconstruction will be limited. Through the specific experiments, we can see that 3DSRCNN converges to a stable PSNR value at about 25 Epoch. Using the computer configuration of table 4.1, 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.

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

When training multiple scale sets, the trend of PSNR in each interpolation scale is not exactly the same. This is because batches randomly taken out from whole training set exerts 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 35 Epoch at given setting, and , X2, X3 and X4 improve 2.89dB,2.42dB,1.50dB respectively.With the increase of scaling factor, the improvement of reconstruction quality is reduced.As the scaling ratio increases, the quality of the reconstruction is reduced, which is due to the loss of more high frequency details, learning low frequency to high frequency mapping relationship is more difficult,thereby limiting the super resolution ability. We investigate parameters setting which take account of both reconstruction quality and speed ,and make a trade-off between them.

**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 networks stack up to slow down the convergence and improve the time complexity. Through experimental verification, we found that the training results will be slightly improved in the early stage of increasing network depth, but the accuracy of reconstruction has not been improved after increasing the number of networks to 20 layers. We found that deepening the network is indeed effective, but we also believe 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 imagesets , so that the mapping relation that can be learned is inherently limited, so an excessive increase of the number of 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 Fig. We can 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 scale of input data, and proper settings will help to improve network generalization ability, but not intuition “the deeper ,the better”.

Experiments indicate that Appropriately increasing the number of network layers can improve performance until 12layers after that there is no apparent promotion. considering both of the performance and speed, We argue that adding layers to 12 is admissible for our work.

**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 25. Based on privious experiment, number of layers is set to 12 and we conduct three compared test with convolutional size of 3x3x3,5x5x5,7x7x7 , and The experimental results are shown as Fig . we

**Residual-Learning**

Increasing network layer will bring about the problem of the gradient explosion/vanishing[]. He Kaiming have introduced residual learning framework to address the issues. Fig 5 dipslay 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. It is found that the whole curve fluctuates greatly. As shown by the picture, PSNR has a large decline in the 7-8 time Epcoh, and the final reconstruction effect is also much worse than the use of residual learning. In the same case, residual learning can make MSE converge to a smaller number in the first 3 epcoh, and the network without residual learning needs to converge after training more than a dozen epoch. The final result shows that the PSNR using the residual network is 3 dB higher than counterpart.

**Comparision to other method**

Through the experimental data in the table, we can see that the 3DSRCNN training model is better than the previous one and is the state of art . 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.

**Conclusion and future work**

In our work , we have proposed a novel method based on deep learning to approach three-dimension super resolution. 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 rather than 2D counterpart to deal with CT rock samples, which ensures the continuity of Z direction. Through practical experiments, We introduced empirical guideline in designing network and parameters-tuning in training process. A single model which was trained by samples mixed with diverse CT sets and it becomes more applicable than previous method. We have demonstrated our approach surpasses previous work such as based on sparse-presentation , A+ methods in experimental results. Stacking different network layers and setting training parameters will have great influence on the accuracy and time of reconstruction. Accordingly, we investigate certain parameter 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, so we should divide the original image into small blocks in whole procedure.

1. 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.
2. 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.
3. Tsai R. Multiframe image restoration and registration[J]. Adv.comput.vis.image Process, 1984, 1.第一个超分辨率
4. Manjón J V, Coupé P, Buades A, et al. Non-local MRI upsampling[J]. Medical Image Analysis, 2010, 14(6):784-792. MRI超分辨