**Abstract**

Deep learning technique has archived excellent performance in 2D-image super resolution. Regarding to 3D-image, such as CT image, 3D-CAD and MRI, also have the necessary to improve the quality of 3D voxel. In this paper, we propose a super resolution convolutional network to enhance 3D resolution for CT images. Aiming at resolving practical problems in 3D datasets, we optimized the network layers and segmentation method of CT images for fitting training of 3D datasets. The trained model represent the mapping relationship of LR and HR. We consider that more larger training set can provide abundant contexture and the deeper network layers can capture more detailed information. Through experiments, we demonstrate our network can complete the duty of multi-scale interpolation reconstruction, which enable a single trained model can be employed in different test sets. Comparing to

Index terms:3D-super resolution, CT images, deep 3D-CNN, Multi-scale learning

**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 Z axis. In the field of geological exploration, CT technique have been widely implemented to display real rock structure.

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. Chao Dong raised a network SRCNN which only contains 3 layers CNN structure but outperform than traditional method. In his paper, he consider the excellent performance of SRCNN that 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. Motivated by Jiwon Kim's work, we propose a novel 3D neural network to practically promote resolution for 3D voxel. 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 as we should pay attention to layers continuity in three axis. To cope with aforementioned issues, we use 22 layers CNN network which contains 72 three-dimension convolutional kernel. Our work proposed the following method to resolve practical issues in CT image super resolution:

Training dataset

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.

Accelerating Convergence

Multi-scale learning

**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 3DCNN 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 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.

With a set of covolutional operation in deep layers, input detail increasingly is discarded, which results in the output can only use learned features. Due to this reason, gradients vanishing/exploding will appear in this training process. Consequently, residual-learning strategy is adopted in 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.

3.2 Pre-process of training set

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 dataset using 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.

Using MSE as the loss function favors high PSNR.

SRCNN minimize the objective function with classic SGD(stochastic gradient descent) in backpropagation. 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 set to 0.4 .

Weight normalization

Experiments

In this chapter, We first demonstrate training and test set which are scanned from real rock sample and crop them into identical size 400x400x400 pixel. Next, we simply handle test CT samples to reconstruct due to limitation of GPU memory. PSNR and SSIM is widely

**Experimental materials and evaluation Criteria**

**Workplatform Details**

**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.

**Comparison**

**conclusion**