Reconstruction of Thin-Slice Medical Images Using Generative Adversarial Network

Zeju Li[1], Yuanyuan Wang[1,2(✉)], and Jinhua Yu[1,2(✉)]

[1] Department of Electronic Engineering, Fudan University, Shanghai, China
{yywang, jhyu}@fudan.edu.cn
[2] Key Laboratory of Medical Imaging Computing and Computer Assisted Intervention of Shanghai, Shanghai, China

Abstract. Slice thickness is a very important parameter for medical imaging such as magnetic resonance (MR) imaging or computed tomography (CT). Thinner slice imaging obviously provides higher spatial resolution and more diagnostic information, however also involves higher imaging cost both in time and expense. For the sake of efficiency, a relatively thick slice interval is usually used in the daily routine medical imaging. A novel generative adversarial network was proposed in this paper to reconstruct medical images with thinner slice thickness from regular thick slice images. A fully convolutional network with three-dimensional convolutional kernels and residual blocks was firstly applied to generate the slices between the imaging intervals. A novel perceptual loss function was proposed to guarantee both the pixel similarity and the spatial coherence in 3D. Moreover, a discriminator network with a sustained adversarial loss was utilized to push the solution to be more realistic. 43 pairs of MR images were used to validate the performance of the proposed method. The presented method is able to recover preoperative t2flair MR images with slice thickness of 2 mm from routine t2flair MR images with thickness of 6 mm. The reconstruction results on two datasets show the superiority of the presented method over other competitive image reconstruction methods.

Keywords: Image reconstruction · Generative adversarial network · Deep learning

1 Introduction

Tomographic medical imaging techniques such as magnetic resonance (MR) imaging and computed tomography (CT) produce tomographic images of cross-sections of human body. Medical instruments pick slices at regular intervals and generalize three-dimensional (3D) volumetric images from a series of slices collected in a certain direction. Medical images with thinner slices preserve higher spatial resolution in coronal and sagittal axis and provide more clinical information. However, thin-slice medical images are not always available because of economic and efficiency issues. For example, glioma patients who get head MR images with slice thickness of 6 mm with routine MRI examination have to get another preoperative t2flair MR images with thickness of 2 mm to obtain more diagnosis information for further treatment. Nevertheless, the reconstruction is also essential for many medical images related

Q. Wang et al. (Eds.): MLMI 2017, LNCS 10541, pp. 325–333, 2017.
DOI: 10.1007/978-3-319-67389-9_38

researches. Therefore, it is of great clinical value to reconstruct thin-slice tomographic medical images.

The task of tomographic images reconstruction is always considered as the image registration problem [1]. 3D information is reproduced by co-registering thicker slices to thinner slices or maps. Registration methods utilize the mutual information between the resource and target volumes and build up transform projection based on affine, spline or radon transform. Registration based reconstruction algorithms largely base on the limited information provided by the paired samples. The information provided by other paired samples are not well utilized. Furthermore, in these algorithms the estimated slices are generated only using neighbor slices, the global information of tomographic images is therefore not included. The performance of these registration based methods is limited by the two constrains.

In this study, the task of image reconstruction for tomographic medical images is seen as single image super-resolution (SISR) problem in 3D. SISR means generating a high-resolution (HR) image from its low-resolution (LR) image for a single image. Actually, reconstruction of tomographic medical images could be viewed as SISR in the coronal plane. Recently, convolutional neural network (CNN) and generative adversarial network (GAN) have made remarkable breakthroughs in SISR [2]. The impressive SSIR results inspire us the reconstruction of tomographic medical images might have better solutions. In this paper, a state-of-the-art SISR method, called super-resolution GAN (SRGAN), is extended to three-dimensional version (3DSRGAN). Specifically, a fully connected CNN with 3D convolutional kernels and residual blocks was proposed to generate tomographic medical images from fewer slices. A perceptual loss function, which consists of four separate losses, is proposed to encourage the network to produce more realistic and reasonable medical images. Pixel-wise loss is the basis of the loss function to encourage pixel-wise image similarity. A 3D total variation loss is proposed to ensure continuity of the deformation in the coronal plane. Traditional regularization loss is presented to avoid overfitting of the deep network. Furthermore, an adversarial loss is added to provide external adjustments which could make the generated results be closer to the original images in the global context. In the experiment, our method is applied to reconstruct preoperative t2flair head MR images of glioma patients from routine t2flair MR images. Preoperative MR images is acquired with thinner slice and preserve slice spacing of 2 mm. On the other side, routine MR images get slice spacing of 6 mm. By comparing the reconstruction results with the original images of two datasets, it shows that 3DSRGAN is more effective and efficient than other popular methods in the task of tomographic medical images reconstruction.

2 Method

It is widely acknowledged that CNN has made great breakthroughs in the field of computer vision. CNN has proved to be a well-performed deep learning model by the use of hierarchical features and convolutional layers. Lately, an adversarial approach, namely GAN, was proposed to learn deep generative models. The idea of GAN is to raise an adversarial network to discriminate the generated samples from the realistic

samples. With the supervision of adversarial network, the generative model is driven to approximate the distribution of the training data and produces more realistic results.

More recently, GAN plays an important role in SISR and recovers photo-realistic textures from heavily down-sampled images. The ultimate results of GAN in SISR motivate us to build a generator network G to generate the information between sampling intervals of tomographic medical images. Concretely, given a tomographic medical images with fewer slices I^{FS} with size $L \times W \times H$, we want to generate the medical images with more slices I^{MS} with size $L \times W \times rH$ of the same examination position.

In CNN, tensor calculation including convolutional kernels is normally in the shape of 2D because natural images are 2D. However, when applied to tomographic medical image, it is more suitable to extend the process to 3D to take fully advantage of the volume data of medical images. 3D convolutional kernels make it possible to calculate the volume as a whole instead of processing slice by slice using 2D convolutional kernels. In this study, SRGAN is extended to 3DSRGAN by the flexible application of 3D convolutional kernels and 3D deconvolutional kernels. The details of the network structures would be described in the following paragraph.

2.1 Generative Adversarial Network

The details of the network structures are described in the following paragraph. The structure of proposed 3DSRGAN is demonstrated in Fig. 1. A 7-layers CNN with 3D convolutional kernels is proposed to generate the reconstructed MRI from MRI with fewer slices (FSMRI). Simultaneously, a discriminator with 6 convolutional layers is trained to discriminate reconstructed MRI from MRI with more slices (MSMRI). The basic block of the discriminator network is convolutional layer + relu + batch normalization rather than convolutional layer + batch normalization + relu to make sure the output of the last layer has negative component. Otherwise, the outputs of relu would be all positive and the classification ability of sigmoid function would be diminished.

FSMRI is up-sampled to be the same size with MSMRI by deconvolutional layers with unbalanced stride of [1,1,3]. In the discriminator network, down-sampling of

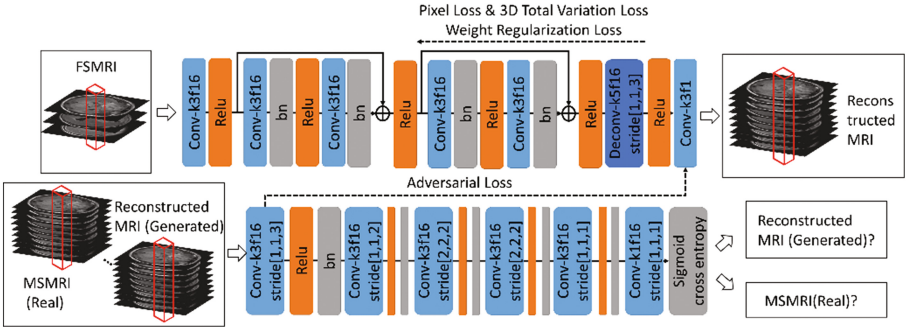

Fig. 1. Flowchart of proposed 3DSRGAN. The size of kernels and the number of filters are illustrated in the blocks. Stride of convolutional kernels is [1,1,1] unless an additional explanation is given.

MSMRI and reconstructed MRI is accomplished by using convolutional kernels with strides instead of pooling, as suggested by DCGAN.

Residual block is a core of the proposed network. The idea of residual blocks was proposed recently to guarantee the integrity of the forward and backward propagated signal. Residual connections help deep network avoid the problems of gradient vanishing and enable the training of networks with very deep structures. Although our generator network is not necessarily deep, it also suffers from gradient vanishing problem for two reasons. First, the use of 3D convolutional kernels would increase the number of trainable network parameters dramatically and make it difficult to pass residual during backward propagation. What's more, a relatively complicated loss function is designed for the generator network and creates difficulties for the global convergence of the network. Residual blocks preserve the flowing signals in the network and would benefit the training of the generator network. Specifically, we add residual blocks between the outputs every two layers. The first convolutional layer is not considered into residual blocks because we want to build to residual network upon high-level features rather than low-level features extracted directly from the images.

In the phase of training, I^{FS} with size $L \times W \times H$ is divided into several volumes with size $15 \times 15 \times H$ to be the input of the network. Similarly, I^{MS} with size $L \times W \times rH$ is divided into several volumes with size $15 \times 15 \times rH$ to be the output of the generator network. In order to quicken the learning speed, both the input and the output volumes are normalized to the range from 0 to 1, separately. Then we normalize the volume blocks on each sequence to have zero mean and unit variance. The outputs of generator network, together with volume of I^{MS}, are concatenated to be the inputs of the discriminator network. The discriminator network is trained to distinguish the volumes from different sources and provide suggestion for improvement back to the generator network. During testing, I^{FS} is input to the generator network with full size with the same preprocessing of zero mean and unit variance. Besides, the mean and variance of batch normalization layers are fixed using the parameters gotten from training.

To demonstrate the effectiveness of our 3D model, a 2D super-resolution CNN (2DSRCNN) is also developed for tomographic medical image reconstruction. 2DSRCNN consists of 12 convolutional layers with small convolutional kernels. 2DSRCNN could be seen as a simplified structure of the 3D generator network and is trained to recover 2D images in the coronal plane. Specifically, for each volume, 2DSRCNN recover L high-resolution images with size $W \times rH$ from L low-resolution images with size $W \times H$.

2.2 Loss Function

In the training phase, we make use of training data to obtain the parameter θ_G of generative convolutional network G_θ. I^{MS} is taken as target and I^{FS} as input:

$$\hat{\theta}_G = \arg\min_{\theta_G} l^G(G_\theta(I^{FS}), I^{MS}) \tag{1}$$

where l^G is the loss function of the generative model. The network is trained by minimizing the loss function l^G which indicated the differences between the generative results and the real images.

In this study, we propose a perceptual loss function designed for the 3D reconstruction problem. The perceptual loss consists of a pixel loss, an adversarial loss, a 3D total variation loss and a weight regulation loss:

$$l^G = \lambda_1 * l^G_{MSE} + \lambda_2 * l^G_{Ad} + \lambda_3 * l^G_{tv} + \lambda_4 * l^G_{\theta} \tag{2}$$

The four components are designed for different optimization purposes and would be discussed in the following paragraphs separately.

Pixel Loss. No one can deny the importance of pixel similarity in the task of reconstruction. l^G_{MSE} is taken as the main body of the present perceptual loss. The pixel-wise MSE loss is calculated as:

$$l^G_{MSE} = \frac{1}{rLWH} \sum_{x=1}^{L} \sum_{y=1}^{W} \sum_{z=1}^{rH} (I^{MS}_{x,y,z} - G(I^{FS})_{x,y,z})^2 \tag{3}$$

Adversarial Loss. Basically, the loss function relied heavily on the mean squared reconstruction error (MSE) in the related study. However, solutions totally based on MSE always appear overly smooth due to the pixel-wise average of all the possible solutions. To address this problem, we proposed a discriminator network to help generator network produce more realistic results.

Typically, the loss function of the discriminator network is calculated as:

$$l^D = \frac{1}{2} \left[l_{bce}\left(D(I^{MS}), 0\right) + l_{bce}\left(D(G(I^{FS})), 1\right) \right] \tag{4}$$

After the discriminator network is settled, generator network would be in turn affected by the feedback from the discriminator:

$$l^G_{Ad} = l_{bce}\left(D(G(I^{FS})), 0\right) \tag{5}$$

This adversarial loss encourages the generator network to confuse the discriminator network. The loss term is large if the discriminator network can discriminate the output of the generator network. The external adjustments brought by the adversarial loss are based on high-order statistics and are accessible by the standard pixel-wise loss function. In our study, the generator network and discriminator network are both updated in each step of a mini batch to provide timely adjustments.

3D Total Variation Loss. Different from the total variation loss which encourages spatial smoothness in output 2D images, we proposed a 3D total variation loss function

to encourage spatially coherent solutions in the coronal plane of output 3D blocks. 3D total variation loss is calculated as:

$$l_{tv}^{G} = \frac{1}{rLWH} \sum_{x=1}^{L} \sum_{y=1}^{W} \sum_{z=1}^{rH-1} (G(I^{FS})_{x,y,z+1} - G(I^{FS})_{x,y,z})^2 \tag{6}$$

The 3D total variation loss is defined based on the sum of MSE between every neighbor slices of the reconstruction images. The variation loss encourages the generator network to estimate the absent data by making use of the neighbor slices. This would strengthen space connectivity of the reconstruction images in 3D and lead to more reasonable results.

Weight Regularization Loss. Structures in different MR images are quite different and networks tend to get poor performance with different dataset. L2 regularization loss is added to prevent overfitting the training data.

3 Experimental Results

Our methods and experiments were all developed by using Tensorflow. Both the generator network and the discriminator network were trained using Adam optimizer with beta1 = 0.9, beta2 = 0.999 and epsilon = 1e−8. The networks were initialized using the Xavier method. The primary learning rates were 1e−2 and 1e−3 for the generator network and discriminator network, separately. The learning rates were decayed in an exponential way with a basis of 0.99. Specifically, we chose the following loss for our experiments:

$$l^{G} = l_{MSE}^{G} + 0.1 * l_{Ad}^{G} + 1e-8 * l_{tv}^{G} + 1e-5 * l_{\theta}^{G} \tag{7}$$

In order to illustrate the effectiveness of adversarial loss, the training of 3DSRCNN and 3DSRGAN were all stop at 25 epoch.

3.1 Material

We validated 3DSRGAN on the reconstruction of brain MR images of glioma patients. T2flair MR images of 43 glioma patients were taken from Huashan Hospital, Shanghai, China. All of these 43 glioma patients have preoperative MR images. The preoperative t2flair MR images were acquired using Siemens scanner with voxel size $0.47 \times 0.47 \times 2$ mm, TR = 9000 ms, TE = 96 ms, TI = 2501 ms, flip angle = 150°. 15 of these glioma patients were found to have routine MR images. The time interval between the two MRI examinations is less than 3 months and no obvious changes were observed in patients' conditions. The routine t2flair MR images were also acquired using Siemens scanner with voxel size $0.47 \times 0.47 \times 6$ mm, TR = 9000 ms, TE = 94 mm, TI = 2501 ms, flip angle = 150°. In this study, preoperative t2flair MR images are down-sampled to slice thickness of 6 mm and used as a simulation dataset. Moreover, 15 cases with both preoperative and routine MR images were taken as a real

images dataset. We take 30 cases from the simulation dataset and 10 cases from the real images dataset for training, separately. The rest of cases in the two dataset were taken as independent test samples. MRI volumes were spatially normalized to the Montreal Neurological Institute (MNI) reference brain to make sure the images were exactly aligned.

3.2 Reconstruction Results

The presented methods were applied to the two datasets mentioned above, separately. To demonstrate the effectiveness of our proposed method, two popular registration methods were used as comparison, namely nearest neighbor (NN) interpolation and 4th B-spline interpolation. The two methods are based on nonlinear deformation and often applied to the reconstruction of thin-slices MR images [1].

To qualitatively compare the reconstruction results of different methods, the reconstruction results of a slice between the sampling intervals are visualized in Fig. 2. Compared with NN and B-spline, 3DSRGAN could provide more realistic results. The error estimation of the skull and brain structure are reduced. Different from reconstruction methods based on interpolation, 3DSRGAN take advantage of a lot of prior knowledge learned from plenty training data. The large number of CNN parameters allows 3DSRGAN to produce more complex mapping relationship between the resource images and the target images. On the other side, the reconstruction results of 3DSRGAN were generated using the global 3D information of the images from fewer slices. 3DSRGAM is able to take advantage of all the given H slices to generate the images between the sampling intervals. Mapping relationship was established from H slices to rH slices. It would benefit the process of image reconstruction with global context. However, registration based methods reconstruct the images with the

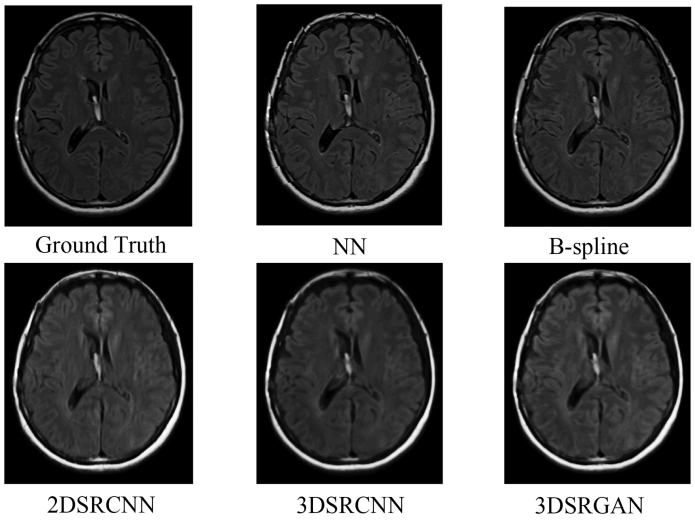

Fig. 2. Visual comparison of original MSMR images and reconstruction result from FSMR images by different methods.

interpolation of neighbor slices. Therefore, the prediction results of 3DSRGAN could be more reasonable. Compared with 2DSRCNN, 3DSRGAN process the images as volumes. Spatially coherent solutions are guaranteed by the use of 3D convolutional kernels. It leads to more accurate and smooth results in the axial plane. Compared with 3DSRCNN, 3DSRGAN benefit a lot from the external adjustment of the discriminator and more details are reproduced. These details make the reconstruction more similar to the original images. It should also be mentioned that the use of residual blocks could make the network converge faster and obtain a better result on both the training data and the test data.

To demonstrate the results of our method in quantitative way, the MSE and peak signal to noise ratio (PSNR) were calculated by taking the images with slice thickness of 2 mm as the ground truth. The global validations are summarized in Table 1. Obviously, the results of 3DSRGAN outperform the results generated by other mentioned methods. In addition, the operation time of 3DSRGAN is shortest because the reconstruction just require one time of the forward propagation of the 3D generator network.

Table 1. Quantitative reconstruction results by different methods. Our methods are validated on two dataset with separate test data.

Dataset	Methods	MSE		PSNR		Operation time(s)
		Mean (std.)	Med.	Mean (std.)	Med.	
Simulation dataset	NN	193.9 (50.7)	182.1	25.4 (1.1)	25.5	322
	B-spline	157.6 (39.4)	149.0	26.3 (1.1)	26.4	202
	2DSRCNN	141.3 (32.7)	140.0	26.7 (1.0)	26.7	4.3
	3DSRCNN	139.5 (36.8)	132.6	26.8 (1.2)	26.9	2.5
	3DSRGAN	**132.4 (34.5)**	**130.5**	**27.1 (1.1)**	**27.0**	**2.5**
Real images dataset	NN	448.8 (140.2)	489.3	21.9 (1.7)	21.2	/
	B-spline	415.8 (132.1)	447.5	22.3 (1.7)	21.6	/
	2DSRCNN	299.8 (90.2)	329.1	23.7 (1.6)	23.0	/
	3DSRCNN	269.5 (84.6)	288.7	24.1 (1.5)	23.5	/
	3DSRGAN	**262.2 (75.9)**	**277.6**	**24.2 (1.4)**	**23.7**	/

4 Conclusion

In this study, we proposed a novel GAN based method for the reconstruction of thin-slice tomographic medical images from images with thick slices. A GAN structure with 3D convolutional kernels and residual blocks is proposed to generate the medical images between the sampling intervals. A novel perceptual loss function which consists of four components is proposed to make the reconstruction results more realistic and reasonable. We applied 3DSRGAN to the reconstruction of thin-slices head MR images of glioma patients. Experiments results demonstrated that 3DSRGAN can provide better reconstruction results than other popular methods.

In the future, other modalities of MR images (such as T1, T2, T1 enhancing) and CT images of the same examination position would be included for the reconstruction of thin-slice t2flair MR images. The effectiveness of enlarging the size of training dataset is also an interesting topic of the future work.

Acknowledgments. This work was supported by the National Basic Research Program of China (2015CB755500), the National Natural Science Foundation of China (11474071).

References

1. Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M., et al.: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. NeuroImage **46**(3), 786–802 (2009)
2. Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., et al.: Photo-realistic single image super-resolution using a generative adversarial network. arXiv:1609.04802 (2016)