2019

Master's thesis

Super-resolution of Multi-contrast MRI Images with Deep Learning

1217003 LIAO Fangyuan

Advisor MATSUZAKI Kiminori

August 01, 2019

Informatics Course

Graduate School of Engineering, Kochi University of Technology

Abstract

Super-resolution of Multi-contrast MRI Images with Deep Learning

LIAO Fangyuan

Magnetic resonance imaging (MRI), one of the medical imaging modalities, uses a strong magnetic field and radio waves to create detailed anatomical and physiological images of the human body without pain and ionizing radiation. Nowadays, MRI plays an indispensable role in our daily life, such as it improves the diagnosis accuracy base on the high-resolution MRI images. In addition, on account of the ability to display high contrast resolution of soft tissue, MRI becomes an essential detection method for the brain neurology.

Recently, the single image super-resolution (SISR) has a growing development in the deep learning field. The previous study (Yuhua Chen, et al.; Brain MRI super resolution using 3d deep densely connected neural networks. CoRR, abs/1801.02728, 2018) showed that the 3D deep densely connected neural networks (DCSRN) [1] achieved state-of-the-art SISR performance in 3D MRI brain image reconstruction.

Inspired by the multi-contrast images of MRI and the work of DCSRN, we propose a 3D multi-contrast super-resolution network. By conducting experiments on 3D multi-contrast MRI images of 32 healthy adult subjects, we confirmed that the multi-contrast input network provides better performance than using only a single-contrast image as input.

key words Super-resolution, MRI, 3D deep convolutional neural networks

Contents

List of Figures

List of Tables

Chapter 1

Introduction

At the beginning of this chapter, we introduce the background of our research, such as medical imaging, magnetic resonance imaging (MRI), and super-resolution (SR). After that, we briefly overview the stage of the current super-resolution studies in deep learning. Then, we explain our work. In the end, we briefly introduce the dissertation structure.

1.1 Background

1.1.1 Medical imaging

With the increasing number of people begin to pay more attention to their health problem, through the help of the medical device's developing and its availability, around a billion medical diagnostic examinations take place all over the world [2]. Among these diagnostic examinations, the clinic value that medical imaging provides is invaluable.

As a technique for creating interior tissue image with the noninvasive, medical imaging is widely used in the medical diagnostic examination, analysis, and treatment purpose. For surgery, a valid medical decision depends on the correct diagnosis. The correct diagnosis usually relies on the accuracy of medical imaging in general. Therefore, an effective and safe medical imaging takes an essential role in medical decision.

According to the status of the patient's interior tissue which shown in the imaging picture, the doctor could accurately locate the lesion, make an assessment of treatment

plan, accurately predict for surgical risk and reduce the area of trauma which caused by the operation. In some case, medical imaging even could help the doctor to reduce some unnecessary surgery base on the patient's physical condition.

The most common types for medical imaging are radiographic imaging, nuclear medicine, ultrasound and Magnetic Resonance Imaging (MRI). In radiographic imaging, base on the ability for identifying the hard tissue in the body, X-ray as a cheap imaging method is widely used for diagnosis of hard tissue like bone and tooth. However, the X-ray only produces a two-dimensional (2D) image. The Computed Tomography (CT) radiographic imaging solves this problem, by producing high-quality digital images in three-dimensional (3D) with the help of X-ray and computer, and these images have a high soft-tissue resolution. However, no matter X-ray or CT, patients have to be exposed under ionizing radiation. The radiographic imaging method might increase the percentage of health risk.

1.1.2 MRI

Unlike radiographic imaging, magnetic resonance imaging is a medical scanning technique which generates medical images by strong magnetic field rather than using ionizing radiation.

Base on a large number of water molecules within the human body, MRI could use the magnetic properties of two hydrogen nucleus which the inside water molecule to create MRI images. According to the fact that the different tissue in the body contains different numbers of the water molecule, the water molecule distribution situation decides the tissue density and shape, which appears in the MRI image.

On account of the MRI's ability for displaying high contrast resolution of soft tissue, such as it could show the vascular structure without the necessity to inject any contrast agent, it becomes an important and common detection method for the brain

in neurology. The superiority of MRI high spatial resolution image helps the doctor by using the abnormal signal in MRI images to identify the abnormalities between healthy and unhealthy tissue.

However, the disadvantage of MRI comes along with its advantage. For keeping the regular operation of the MRI machine, it requires the liquid helium cooling equipment, rebuilding or reinforcement of the shielding room, the electricity and more necessary conditions according to the weight of MRI machine and the strong magnetic field it needed. These practical reasons caused the MRI scanner is costly, and the maintenance fee is high.

Furthermore, if the requirement for higher resolution MRI image is needed, one option is to buy an MRI scanner which can create a stronger magnetic field, but it cost more. For instance, a 3 Tesla MRI scanner can double the field strength than a 1.5 Tesla MRI scanner. It could create high-quality images with shorter time and more tissue details, but a 3 Tesla MRI machine cost approximately twice for 1.5 Tesla MRI machine, it might be unaffordable for some medical institution. Compared with the scan time of CT (approximately 15 minutes), the MRI scan takes more than 30 minutes in general because the acquisition time for collecting all MRI signals is long.

Base on the existing MRI disadvantages mentioned above, the super-resolution method might be a choice to obtaining a higher resolution MRI image by breaking MRI machine physical restriction with computing power.

1.1.3 Super-resolution

A high-resolution image usually has high perceptual quality and more valuable information than the low-resolution image, see in Figure 1.1.

Super-resolution (SR) technique aims to generate a high-resolution (HR) image, which has high perceptual quality and precise image details, from a low-resolution

(a) High-resolution natural image (b) Low-resolution natural image

Fig. 1.1 An example of high-resolution image and low-resolution image

(LR) image. Base on the amount of low-resolution input images, it could simply be grouped into two categories: single image super-resolution (SISR) and multi-image super-resolution (MISR) [3].

With the breakthrough of deep learning in recent years, a class of deep neural networks called convolutional neural network (CNN) has been studied to recognize and analyze the visual pattern in visual imagery. Recently, the convolutional neural network is widely used in many fields, such as image recognition, image classification, natural language processing, etc.

In the super-resolution area, Super-Resolution Convolutional Neural Network (SRCNN) [4] is the pioneering single image super-resolution method with deep learning. It uses only three convolutional layers to producing super-resolution in 2D natural images. By directly extracting image features and creating the end-to-end mapping in CNN, SRCNN achieved the state-of-the-art performance. Whereafter, SRCNN3D [5] extends the work for SRCNN. It demonstrated that 3D fully convolutional neural network has the potential ability to enhance 3D medical images. By fully using all features extracted by previous layers, Dense Convolutional Network (DenseNet) [6] achieved excellent perfor-

mance in image classification. Inspired by this, 3D Densely Connected Super-resolution Networks (DCSRN) [1] adopt one dense block to reconstruct high-resolution details from low-resolution 3D medical images. In addition, comparing with other SISR methods, DCSRN achieved state-of-the-art performance in 4x resolution-reduced 3D MRI images.

1.2 Proposal

In general, MRI sequence decides the appearance and contrast of MRI image. The most common two MRI sequences are T1-weighted scan and T2-weighted scan. Through applying these different type of sequence, the different modality images are generated such as T1-weighted image is obtained from T1-weighted scan.

As mentioned in the last section, most current super-resolution work is based on single-image super-resolution. For example, the super-resolution network of SRCNN3D and DCSRN use only T1-weighted images as network input for reconstructing super-resolution MRI images.

As MRI could provide multi-contrast medical images, it provides a possibility of using multi-image super-resolution technique to produce a high-resolution MRI image. Inspired by the multi-contrast images of MRI and the work of DCSRN, we propose a 3D multi-contrast super-resolution network. In this work, by comparing with using only a single-contrast image as network input, we first confirm the usefulness of multi-contrast input network in super-resolution image generation process. Then, we investigate the performance of the proposed network in the dataset provided by the human connectome project (HCP) [7].

1.3 Dissertation overview

Chapter 2 mainly introduces the deep learning and the convolutional neural network, and the related information at the beginning. Then it presents the previous work of image super-resolution in the convolutional neural network. At the end part of this chapter, it briefly introduces MRI acquisition progress and the relationship between MRI image resolution and k-space. Chapter 3 introduces the proposal 3D multi-contrast super-resolution network and explains its workflow in detail. Chapter 4 first introduces the setting and design of conducted experiments. After that, it presents the experiment results and shows the comparative analysis of the results. Chapter 5 concludes the thesis.

Chapter 2

Related work

In this chapter, we first introduce the deep learning and convolution neural network. Then, we briefly introduce the batch normalization and image augmentation method. After that, we introduce three super-resolution works with deep learning, which are closely related to our work. At the end of this chapter, we introduce the image acquisition progress of MRI and display the relationship between k-space and MRI image resolution.

2.1 Deep Learning

About 60 years ago, artificial intelligence (AI) as a comprehensive field was born for creating an intelligent machine, which could solve the ordinary task automated like a human.[8] Diven by the hardware, big dataset, and improved algorithms, machine learning started to flourish in recent years. As a branch of artificial intelligence, machine learning used a new programming paradigm. In machine learning, compared with classical programming, the model is built by learning the optimal parameters from the given pair of input data and its corresponding answer.

Among all the branches of machine learning, by using multiple representation layers, deep learning achieved better performance than other methods in the computer vision field. Due to its state-of-the-art performance, deep learning has been applied in image recognition, image classification, language generation, and other fields.

As the foundation of deep learning, the neural network consists of the input layer, output layer, and one or more hidden layers in general. For a neural network, a layer is the primary data structure that process data and stores the learned features in weights. As the basic unit of each layer, a neuron receives the inputs from other neurons in the previous layer. Each connection between a neuron and its inputs has independent weights. After passing the weighted sum to the activation function, the output value of the neuron is generated and treated as one of the inputs for the next layer. An example of the neuron and neural networks are shown in Figure 2.1.

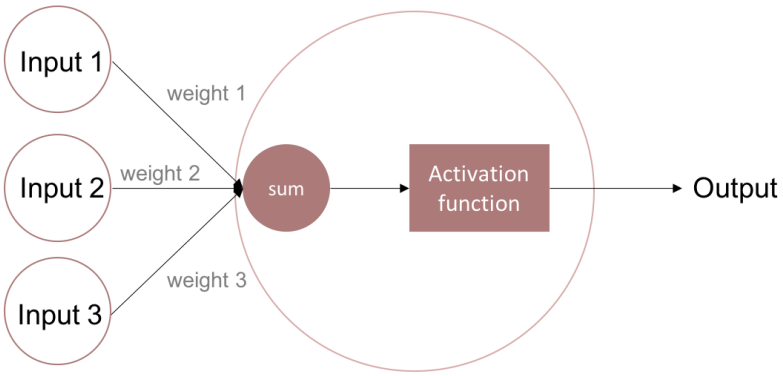
In deep learning, training network actually could treat as making the network approaching the best performance by adjusting the value of each weight. In addition to layers, input data, and its corresponding answer, the loss function and optimizer are needed for obtaining an optimal model. The loss function as the feedback signal shows the gap between the target and network prediction. The optimizer guides the learning progress and minimizes the loss value generated from the loss function.

After all of these components are decided, the weight of each layer will be updated to generate the new prediction more close to the expected target. This kind of data transformation, prediction generation, and model parameter updating will be done for many times until finding the proper value to weights for building the optimal model. A simple example of deep learning progress is shown in Figure 2.2.

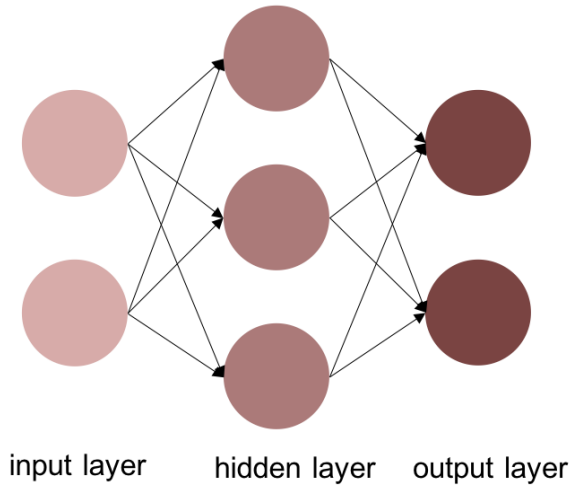
2.2 Convolutional Neural Network

2.2.1 Background

In computer vision, as the most common deep learning model, convolutional neural network (CNN) was designed to detect the visual pattern of its input data. Unlike the fully connected neural network, the convolutional neural network reduces the number of



(a) Neuron



(b) Neural networks

Fig. 2.1 A simple illustration of neuron and neural networks

parameters and learns the local image pattern, which has translation-invariant property.

In general, convolutional neural networks are made by convolution layers, pooling layers, and fully-connected layers. An example of convolutional neural network architecture is shown in Figure 2.3.

At the beginning stage of CNN, the classic LeNet-5 [9] which is shown in Figure 2.4 was born in 1998 as a pioneering work and was used to solve the handwriting number recognition task.With the development of ReLU activation function, dropout

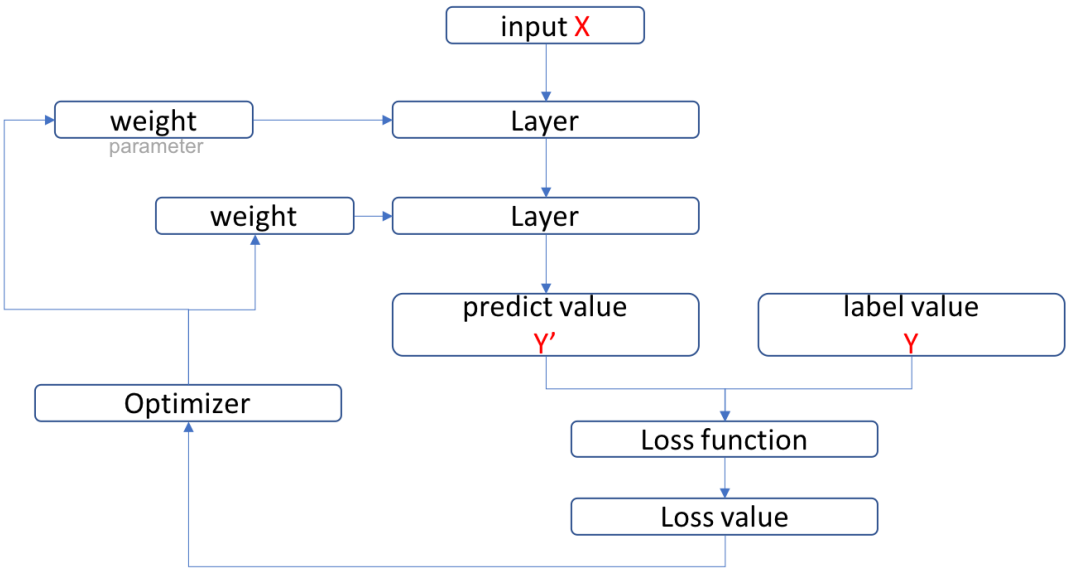


Fig. 2.2 Illustration of a simplified deep learning progress

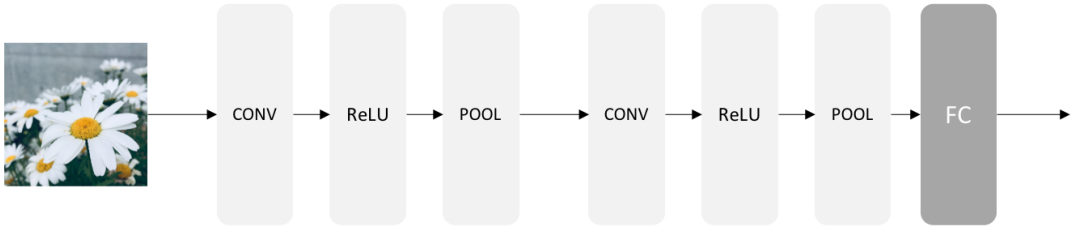


Fig. 2.3 A simple illustration of convolutional nerual network

and historical opportunities which brought from GPUs and the large datasets, AlexNet [10] won the championship in ImageNet competition with an absolute advantage in 2012. Since then, deep learning and the convolutional neural network has explosive development. A simple illustration of AlexNet architecture is shown in Figure 2.5.

Until now, the convolutional neural network has been used for image recognition, image classification, object detection, action recognition, text detection, etc. Also, most deep learning methods are implemented by the convolutional neural network in image recognition competition.

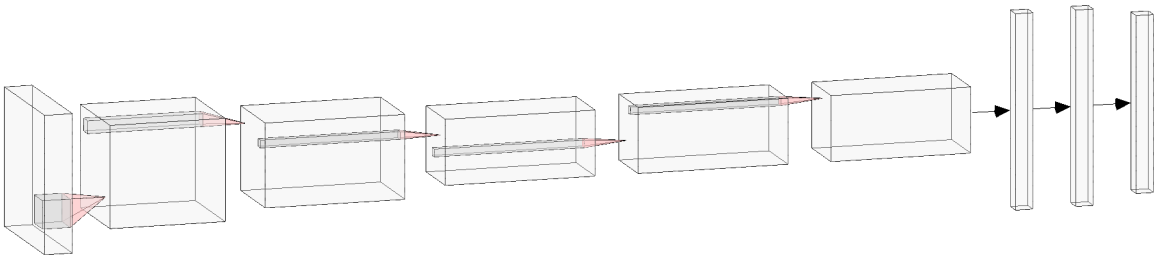


Fig. 2.4 A simple illustration of LeNet

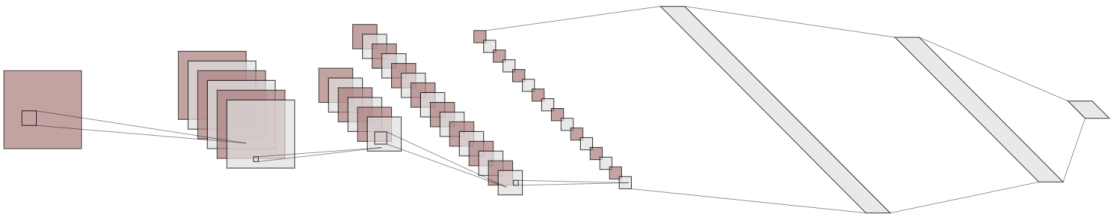


Fig. 2.5 A simple illustration of AlexNet

2.2.2 Convolution layer

As the core building block of the convolutional neural network, to reduce the number of network parameters and help the network learn the local patter from input data, the convolution layer takes the most critical calculation works. In the convolution layer, the convolution filter (or convolution kernel) is treated as the neuron. The feature maps are generated by applying the convolution operation between the convolution filters and the input data from the previous layer.

As we knew, the pixel is the smallest unit of an image. In a grayscale image, image is consist of pixels which has a value range from 0 to 255, among these pixel values, 0 stands for black and 255 stands for white. Furthermore, a color image is made by different color channels. For example, an RGB image, it has three image color channels which red, green, and blue. For a grayscale image, it has only one image color channel.

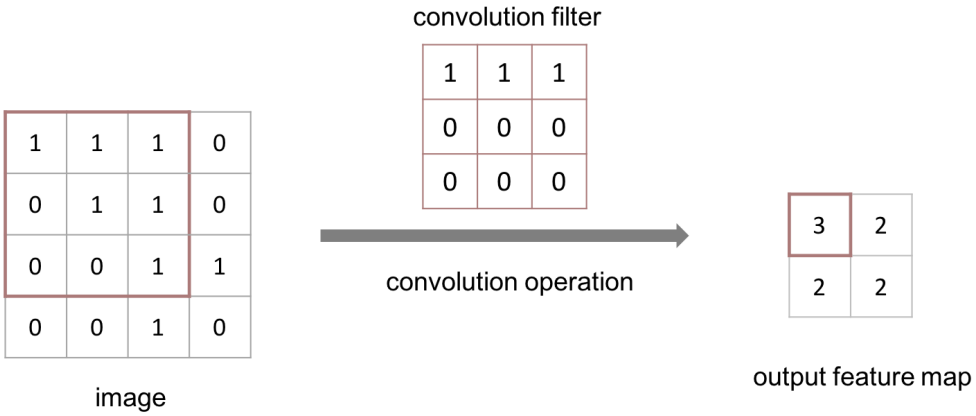


Fig. 2.6 An example of the convolution operation

When a grayscale image as the convolution layer input, image patches are first extracted from the input data by sliding. Besides, each patch has the same shape as the convolution filter. Then, the convolution operation is conducted between the patches and a filter to produce a feature map which contains one type of feature information. By applying above operations for several times, the feature maps are made by combing all the feature map which generated from each filter. Then, the feature maps are treated as the output feature maps of this layer. In addition, the generated feature map has the same width and height as the filter, and it also has the depth, which is the number of feature maps.

For the next convolution layer, the output feature maps from the previous layer are treated as the input feature maps and do the same convolution operation several times to produce its output feature maps. A simple example of convolution layer process is in Figure 2.6.

In general, the output height and width of the feature map is smaller than the input because of the convolution operation. At the convolution layer, by adding the padding into the input feature map and adjust the convolution stride, the output feature map could have the same width and height as the input feature map. An example of a

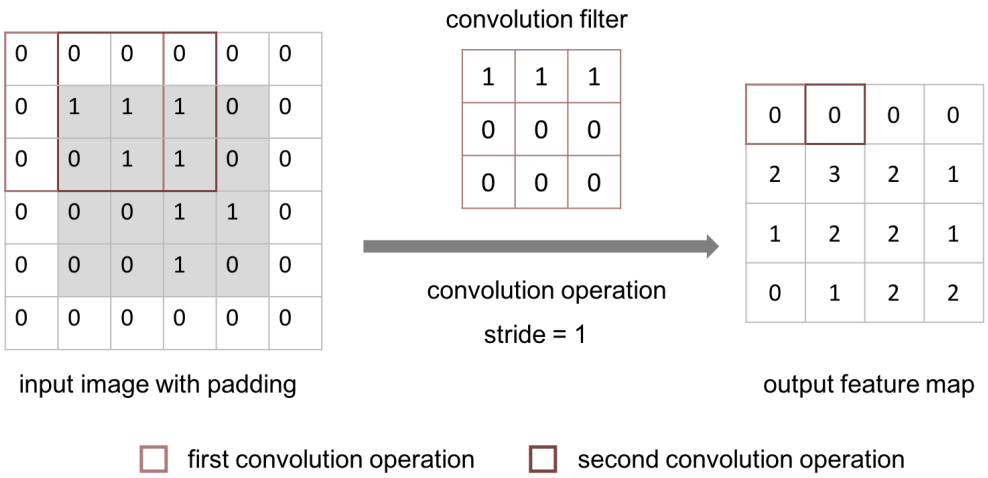


Fig. 2.7 An example of applying the convolution operation on the input image data with padding

convolution operation with padding and a stride of 1 are shown in Figure 2.7.

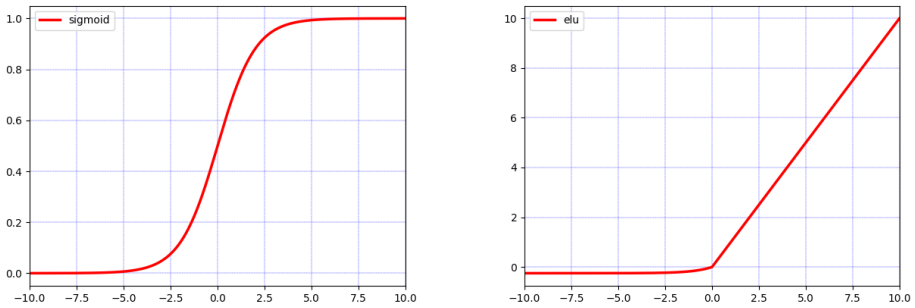
2.2.3 Activation function

The activation function generally attached at the end of layers. It increases the network representation capability by adding the non-linear transform.

The most common activation functions are Sigmoid, Tanh, Rectified Linear Units (ReLU), LeakyReLU, Exponential Linear Unit (ELU). Two example activation functions are shown in Figure 2.8.

2.2.4 Loss function

For measuring the difference between the ground truth image and the super-resolution image that network predicted, the loss function is introduced to guiding the model optimization in the network training stage. The most common loss functions are the mean squared error (MSE) and cross-entropy.



(a) Sigmoid (b) ELU

Fig. 2.8 The function graph of Sigmoid activation function and Exponential Linear Unit (ELU) activation function

2.3 Batch normalization

In general, the model weight initialization will affect the training rate of deep neural network. For example, a bad initialization parameters might leads to gradient propagation problems such as gradient vanishing and explosion, then reduce training speed.

Batch normalization (BN) [11] is proposed to overcome the disadvantages of the training difficult in deep learning. It speedup the convergence and decrease the possibility of overfitting in some way.

By applying batch normalization in the middle layer of the deep neural network, it can somehow reduce the reliance on initialization parameters and set the distribution of the independent features are more closely to the state when the input data feeds to the network. Thus, enhanced the network generalization ability in some extent.

To sum up, batch normalization solves the gradient problem in the backpropagation progress and makes the overall update of the weights of different scales more consistent.

2.4 Image Augmentation

The common reason caused the underfitting problem is the small training dataset. Without enough training examples, the model cannot be optimal and obtain generalization ability. When this underfitting model meets the new input data, it might not output the correct prediction. To avoid this problem, data augmentation is a good option that generates more training data examples from the training dataset. Base on the larger training dataset, the model could observe the data well, then obtain better generalization capability in training stage. The common data augmentation methods are rotation, shift, zoom, flip, etc.

2.5 Image Super-resolution in Deep learning

Base on the current limit of imaging apparatus, the super-resolution technique is widely used in medical imaging, security, space flight, and so on. To achieve super-resolution performance for a single image, there are three categories which are interpolation-based approaches, reconstruction-based approaches, and learning-based approaches.

With the breakthrough of deep learning in recent years, learning-based super-resolution method with deep learning achieved state-of-the-art performance. Among them, three previous studies are closely related to our work, which are SRCNN, SRCNN3D, and DCSRN. We describe the detail of them by sequence.

2.6 SRCNN

In 2014, the super-resolution convolutional neural network (SRCNN) as the pioneering work, introduces the convolutional neural network into the progress of super-resolution image generation. Especially, it achieved state-of-the-art super-resolution

performance with only three convolution layer. The network structure of SRCNN is shown in Figure 2.9.

For generating a low-resolution image, the 32x32 high-resolution image patches are first cropped from the original ground truth image. Then, the 32x32 high-resolution image patches are conducted the downsampling operation. After applied the Gaussian blur, the image patches are conducted the upsampling operation for upscaling the image patch size back to 32x32. The downsampling and upsampling operations are based on bicubic interpolation with the same factor. Then, these ground truth image patches and corresponding low-resolution image patches are treated as the input data and label for the network.

Inspired by sparse-coding, the structure of SRCNN is divided into three parts, which are patch extraction, non-linear mapping, and reconstruction.

The first convolution layer is treated as the patch extraction and representation operation. The kernel size of this convolution layer is 9x9, and the kernel number is 64. Then, its output feature maps have the matrix size 24x24x64. After that, it applied ReLU as the activation function in this layer.

At the non-linear mapping convolution layer, the kernel size is 1x1, and the kernel number is 32. Thus, the output feature maps have 24x24x32 matrix size for this convolution layer. Then, the ReLU activation function applied again to this layer.

In the last convolution layer, which treated as reconstruction operation, it has 5x5 kernel size. For the high-resolution image reconstruction, the number of kernels is the same as the input image channel. Thus, the high-resolution output image has the matrix size 20 x 20 x c, which c is the input image channel.

For the training phase, SRCNN adopted the Mean Squared Error (MSE) as the loss function of the network. In the testing phase, SRCNN achieved better performance than traditional super-resolution method such as sparse-coding method, Anchored Neigh-

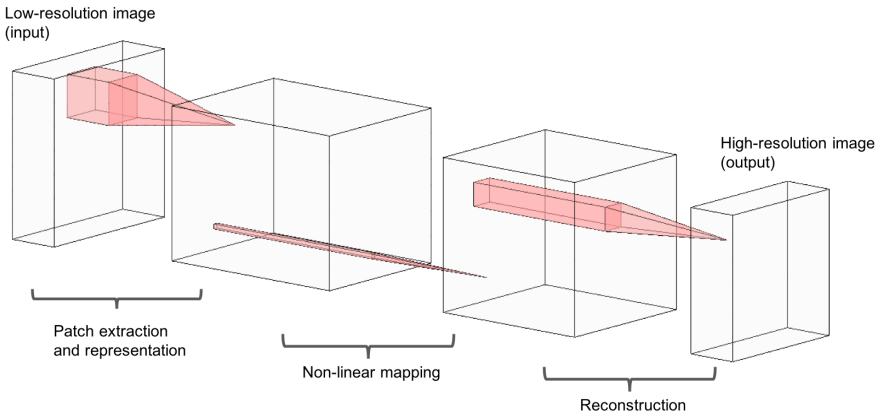


Fig. 2.9 The structure of s uper-resolution convolutional neural network (SRCNN)

bourhood Regression (ANR) method, etc. in natural two-dimensional (2D) images.

2.7 SRCNN3D

Based on the work of SRCNN, C.-H. Pham, et al. proposed a three-dimensional (3D) convolutional neural network (SRCNN3D). Different like SRCNN, which was designed to improve the performance of super-resolution in the 2D natural image, the SRCNN3D extend the dimension of SRCNN architecture to 3D data. Moreover, SRCNN3D investigated network performance in 3D MRI brain data.

With the increasing dimension of the network, the matrix size of the kernel in each convolution layer also extends to 3D. For example, in the first convolution layer, which stands for patch extraction and representation, the kernel size changed from 9x9 to 9x9x9. Besides, the patch size of 3D input data has decreased to 25x25x25 for balancing the computation time and the stability of training.

In SRCNN3D, it used Kirby21 T1-weighted MRI images as network dataset.

Among the dataset, 10 MRI images have been used in the training phase and 5 other MRI images have been used for network performance testing. Same as SRCNN, SRCNN3D used MSE as the loss function to measure the difference between the ground truth image and network prediction.

For measuring the network performance, SRCNN3D used PSRN and SSIM as image metrics. The result of SRCNN3D first shown using the specific image data (MRI) for network training could improve the network performance for specific images than using natural images. Also, it approaches better performance than 2D SRCNN averaging outputs.

2.8 DCSRN

Inspired by the dense convolutional network (DenseNet), the 3D densely connected super-resolution network (DCSRN) was proposed to restore the finer high-resolution image detail for 3D structural brain MRI images. The architecture of DCSRN is shown in Figure 2.10.

For simulating the real MRI image data acquisition progress, the low-resolution MRI image was generated by the following steps: 1) convert the original high-resolution image into k-space 2) zeroing the high-frequency area of k-space as resolution degrading 3) apply inverse FFT to 3D k-space. Then, the network training data were randomly cropped into 64x64x64 cubes from 3D MRI images.

After the training data generation step, the input low-resolution cubes are first fed into a convolutional layer, which has 3x3x3 kernel size and 2x24 filters. Then, the transformed input data go through a four-unit densely connected block. Each unit has a batch normalization layer, which followed by exponential linear units (ELUs) activation function, and a convolution layer which has the kernel number of 24. This four-unit

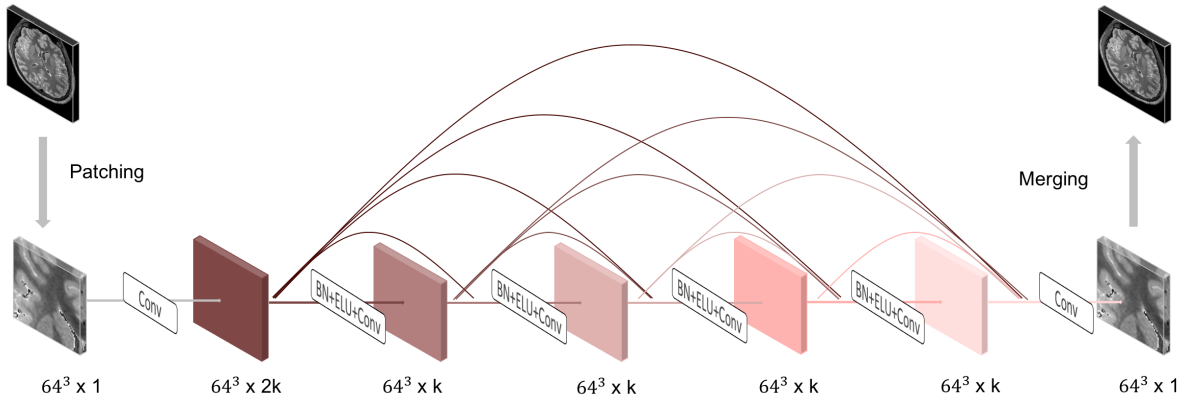


Fig. 2.10 The architecture of 3D Densely Connected Super-resolution networks (DCSRN)

block structure reduce the problem for gradient vanishing and enhanced the feature propagation, supporting feature reuse. Besides, the feature maps of all preceding units are used as inputs, and its own feature maps are used as inputs into all subsequent units. In the end, a convolution layer transforms the data to generate the super-resolution cubes.

The whole super-resolution image was rebuild by averaging the network output cubes, with the 3D sliding window behavior which shifts half cube size each time.

For demonstrating the generalization and performance of the network, DCSRN used the brain MRI database in the human connectome project (HCP), which has 1113 T1-weighted structural images. Same as previous super-resolution works, MSE was treated as the loss function of the network to measure the difference between super-resolution and the high-resolution cubes. For comparing the network performance, peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and normalized root mean square error (NRMSE) were used to measure the image quality.

Compared with simple interpolation method and SRCNN-based method, DCSRN achieved the highest PSNR and SSIM and lowest NRMSE among them.

2.9 MRI and Fourier Transform

2.9.1 MRI image Acquistion

In the MRI image acquisition stage, the different pulse sequence is applied to the MRI scanner. The MRI image generation follows the following steps:

1. placing the subject into a very strong magnetic field
2. using radio frequency (RF) coils to transmit the RF enery to the subject
3. turning off the RF transmitter
4. receiving the radio frequecny energy that emitted from the subject's body
5. From collected MRI signals reconstructing the anatomical and physiological image of internal body structure

2.9.2 MRI in k-space

As a spatial frequency domain, K-space is used to represent the MR image in an array under the spatial Fourier transform. The image resolution in the image domain is related to the number of data points in k-space. By filling more data in k-space, the image resolution will be upgraded in the image domain. For obtaining the MRI image, the MRI signal information will be collected for several times to fill the whole image k-space in the scanning time.

In general, the data point in k-space is corresponding with the part of the information in the image domain rather than a pixel. For example, in the central part of k-space, the basic contrast in the image domain is decided by the low spatial frequencies information. In the border part of k-space, it stores high spatial frequencies information, which decides the details and edges for the image.

In the following Figures 2.11 – 2.16, it shows the relationship between the k-space

and image domain. For Figures 2.11 – 2.13, the k-space data will be filled in a circle area from the center part of k-space as a low-pass filter, and the radius of the circle area will be increased step by step. For Figures 2.14 – 2.16, the k-space data will be filled outside of a circle area as a high-pass filter with the same radius increase.

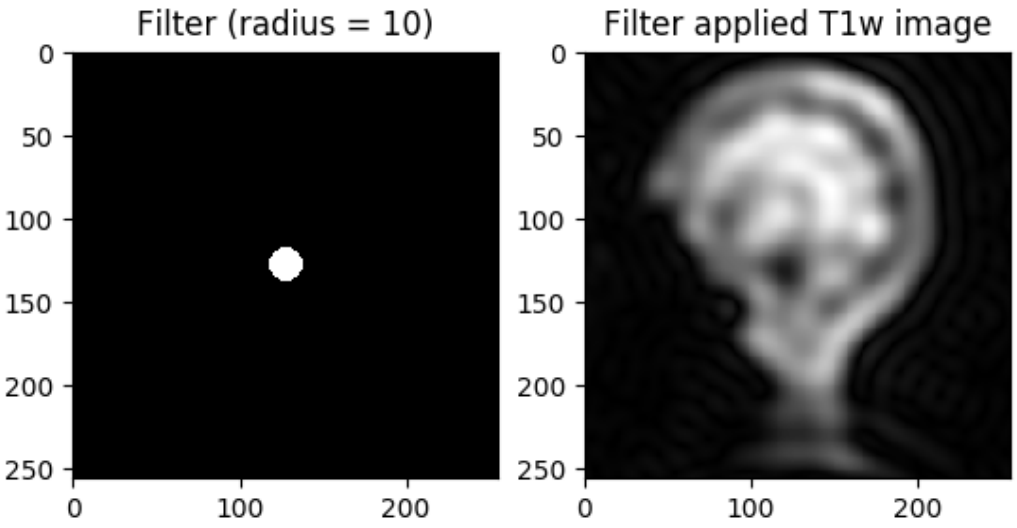


Fig. 2.11 A T1-weighted MRI brain image which applied a low-pass filter with a radius of 10

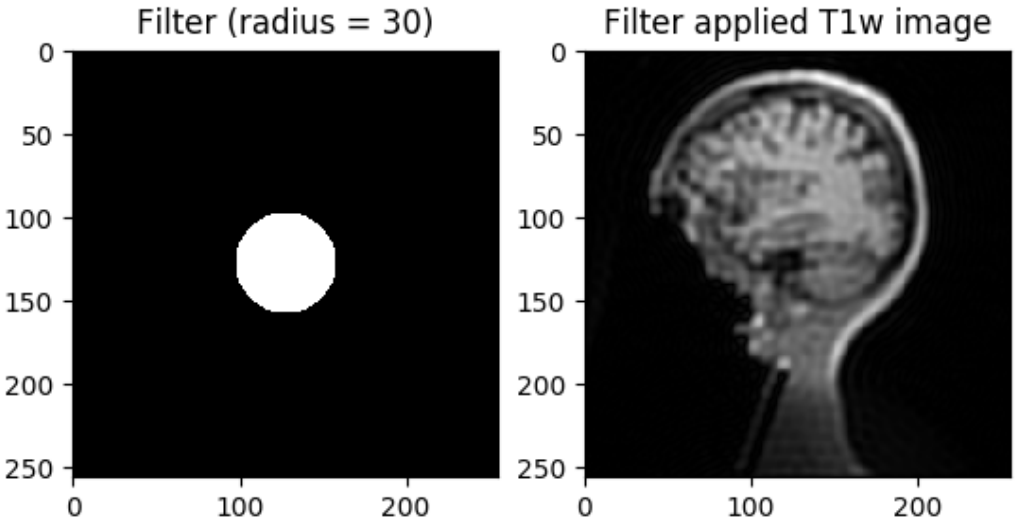


Fig. 2.12 A T1-weighted MRI brain image which applied a low-pass filter with a radius of 30

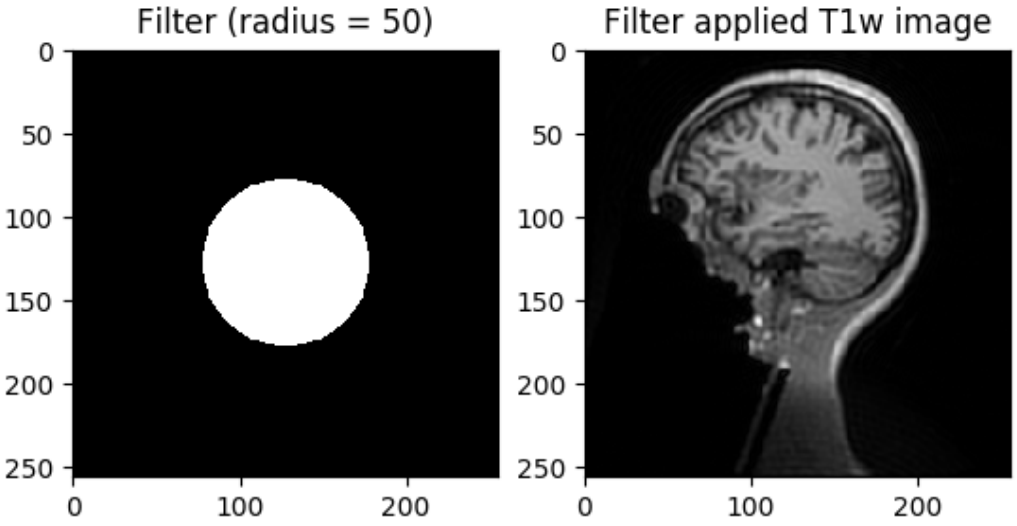


Fig. 2.13 A T1-weighted MRI brain image which applied a low-pass filter with a radius of 50

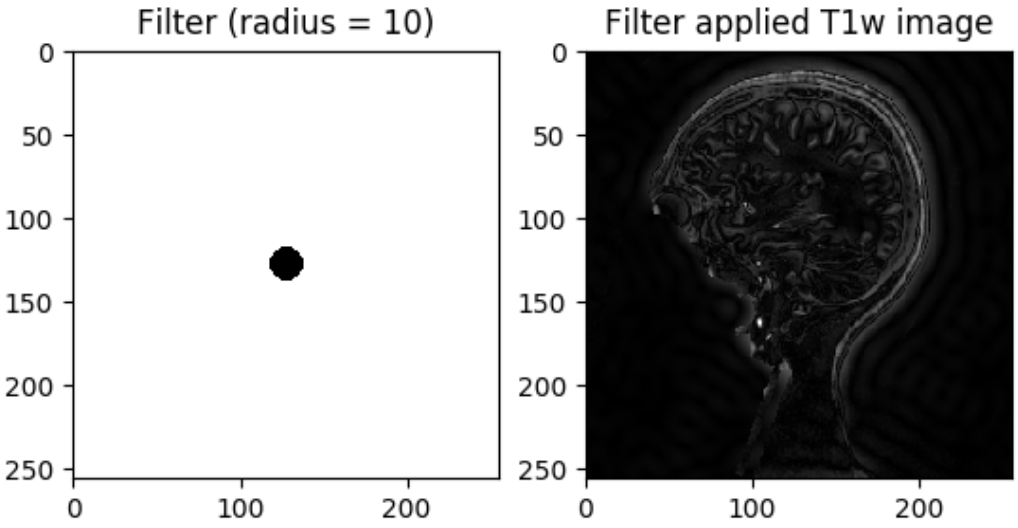


Fig. 2.14 A T1-weighted MRI brain image which applied a high-pass filter with a radius of 10

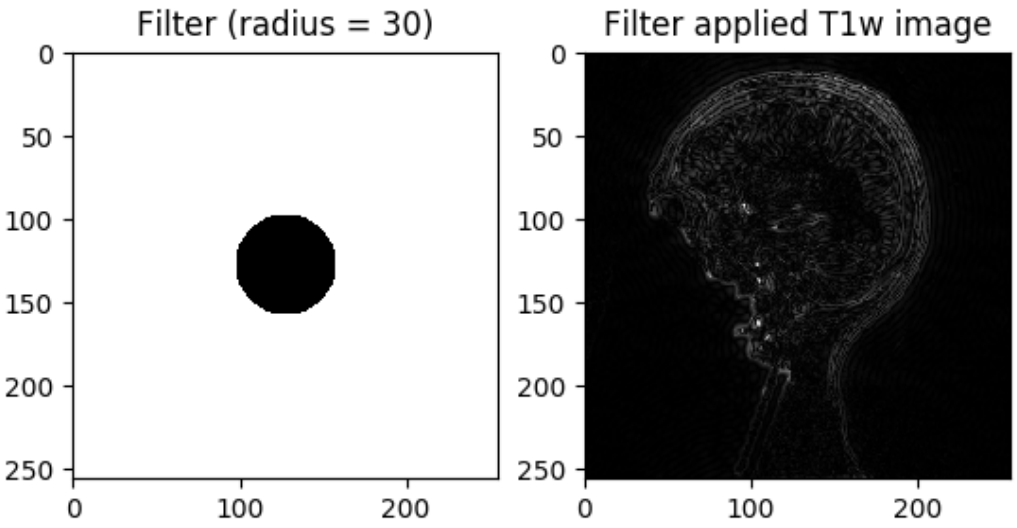


Fig. 2.15 A T1-weighted MRI brain image which applied a high-pass filter with a radius of 30

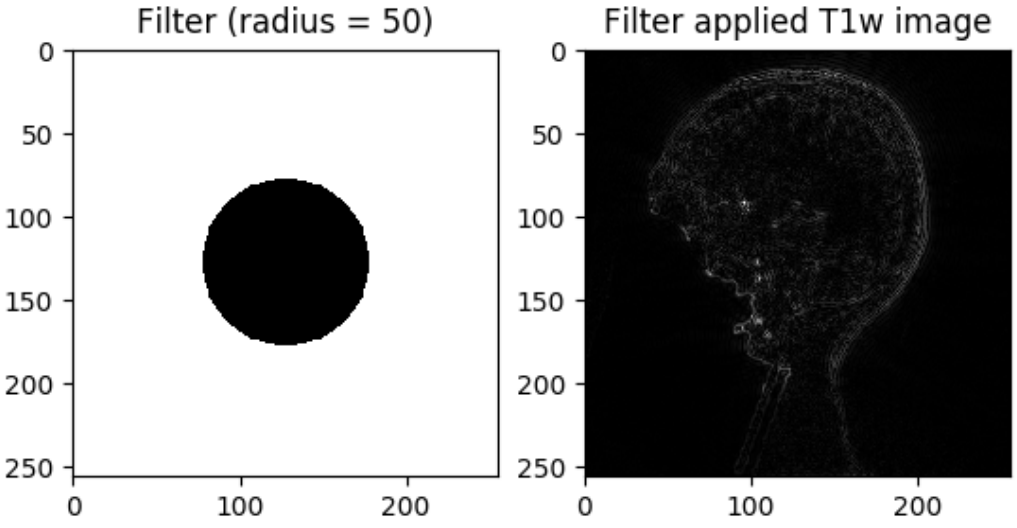


Fig. 2.16 A T1-weighted MRI brain image which applied a high-pass filter with a radius of 50

Chapter 3

Method

In this chapter, Section 3.1 gives an overview of the proposed 3D multi-contrast super-resolution network. Then, Section 3.2 introduces the data pre-processing steps applied before inputting the data into the network. Section 3.3 introduces super-resolution cube generation progress in detail. In the end, Section 3.4 introduces the process of rebuilding the 3D MRI images from network outputs.

3.1 Overview of Proposed Network Architecture

Recent studies unfold the superiority of dense block in Densely Connected Convolutional Network (DenseNet). In medical imaging super-resolution field, the deep learning network which adopts the dense block also achieved state-of-the-art performance, such as DCSRN and mDCSRN.

To achieve a better 3D MRI super-resolution performance, we propose a 3D multi-contrast super-resolution network, which is shown in Figure 3.1. The framework of the proposed network is based on the main architecture of DCSRN, and it is enhanced with multi-contrast MRI images, which are pairs of T1-weighted MRI images and T2-weighted MRI images in this work.

In the proposed network, we use three kind of images, T2wHR, T2wLR, and T1wHR.

T2wHR The high-resolution T2-weighted images as the reference to guide the super-

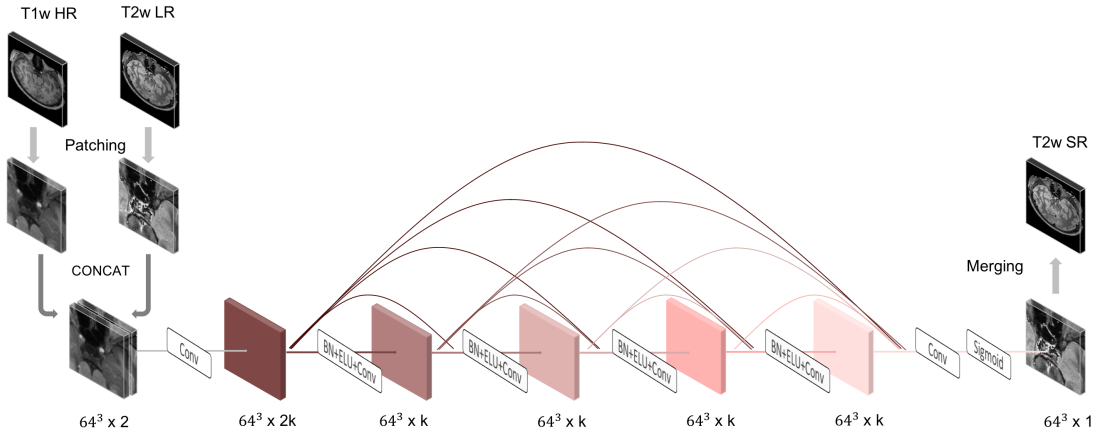
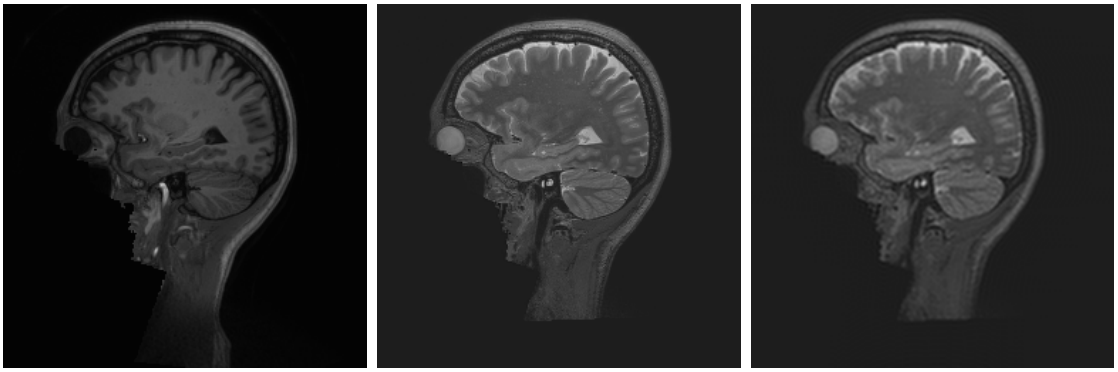


Fig. 3.1 The architecture of proposal 3D multi-contrast super-resolution network



(a) T1-weighted HR image (b) T2-weighted HR image (c) T2-weighted LR image

Fig. 3.2 An example slice of the proposed network input 3D MRI brain image

resolution image generation progress,

T2wLR The low-resolution T2-weighted images as the network input, and

T1wHR The high-resolution T1-weighted images as the hint for improving the network performance.

The example image of T2-weighted high-resolution image, T2-weighted low-resolution image and T1-weighted high-resolution image is shown in Figure 3.2.

The workflow of 3D multi-contrast super-resolution MRI image reconstruction could be divided into three steps:

1. Data pre-processing,
2. Super-resolution cube generation, and
3. MRI image reconstruction.

The detail of each step will be described in the following sections.

3.2 Data pre-processing

3.2.1 Normalization

In the network training stage, the random distribution of input data might slow down the rate of convergence, because the network might be unable to learn similar features from the inputs. Also, in the testing stage, the network might predict the incorrect output in unexpecting.

In our study, all the processing data are 3D MRI images. Since each image has a different maximum and minimum intensity value, the normalization is required to be conducted.

All the T1-weighted MRI images and T2-weighted MRI images are applied the min-max normalization to rescale the value range from 0 to 1. The formula of min-max normalization is:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}},$$

where X is the input image data, $X_{\max}$ is the maximum value in X, $X_{\min}$ is the minimum value of X, and X' is the image data after applied min-max normalization.

3.2.2 Low-resolution image generation

In the central part of k-space, the low-frequency information decides the image contrast . However, the high-frequency information in the outer part of k-space decides

the finer detail information in the image domain.

For simulating the MRI image generation and degrading the image resolution, the low-resolution 3D T2-weighted MRI images are generated by the following steps:

1. First, converting the high-resolution 3D T2-weighted MRI image into k-space by three-dimension Fast Fourier Transform (FFT).
2. Then, along with the phase encoding and frequency encoding direction, the resolution was downgraded by zeroing 75 percentage of the high-frequency information data.
3. In the end, the low-resolution 3D T2-weighted MRI image was generated by applying the inverse Fast Fourier Transform to k-space.

3.3 Super-resolution cubes generation

In this proposed network, the high-resolution T1-weighted MRI image cubes were treated as the hint and expected to help the super-resolution network to produce the better super-resolution T2-weighted MRI image cubes. Before feeding the multi-contrast cubes into the network, the high-resolution T1-weighted and the corresponding low-resolution T2-weighted MRI cubes were concatenated as the network input cube pairs, whose location is matched.

First of all, a convolution layer was applied to the input with kernel size 3x3x3 and filter numbers 2x24. Then, the output of the first convolution layer was treated as the input for a four-unit dense block. Each unit consisted of a batch normalization layer which followed by exponential linear units (ELUs) activation function and a convolution layer. Each convolution layer in the dense block has the same filter size 3x3x3 and filter number 24. However, the input feature map channel of each convolution layer is increased by the number of convolution layer that data passed by.

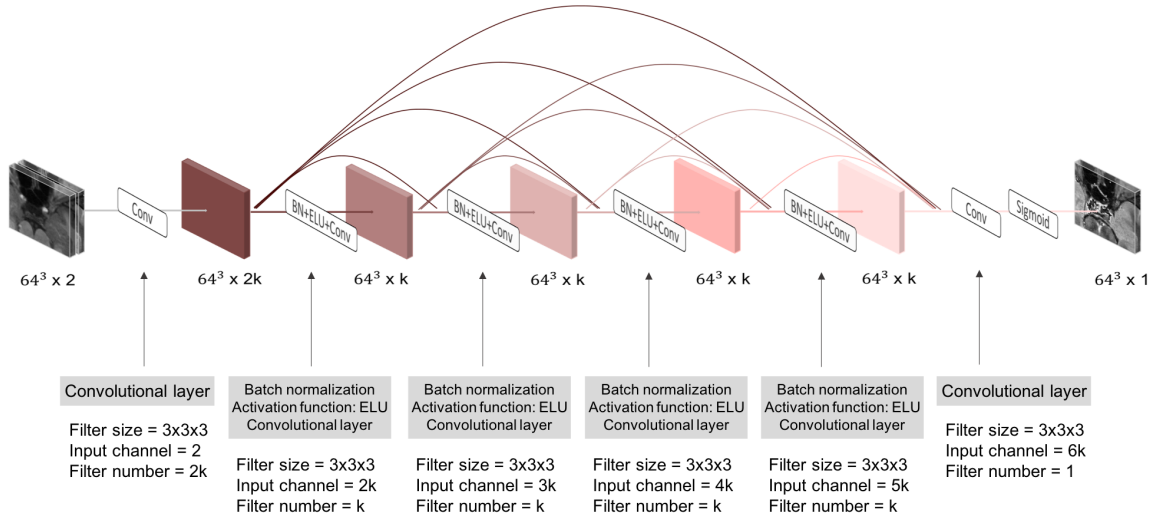


Fig. 3.3 Illustration the central architecture of proposal 3D multi-contrast super-resolution network

After that, the transformed data go through the final convolution layer which has the 3x3x3 filter size and the filter number of 1. Due to all the input data value have been normalization between 0 and 1, Sigmoid activation function usually converts the real value input to another value, which has the range from 0 to 1. Thus, the sigmoid activation layer was treated as the last layer, which helps the network produce more accurate SR image.

The detail of 3D multi-contrast super-resolution network is shown in Figure 3.3.

3.4 MRI image reconstruction

To reconstruct super-resolution 3D MRI image, the T2-weighted low-resolution cubes were first cropped from the 3D T2-weighted low-resolution MRI images with a 3D sliding window behavior where the stride is half of the cube size.

By using the optimal model parameter, which was saved from the training phase, the super-resolution T2-weighted cubes were generated from the proposed network.

After that, the whole super-resolution T2-weighted MRI image was rebuilt by averaging the overlapping outputs.

Chapter 4

Experiments

In this chapter, we first introduce the dataset and the preprocessing detail of experiments. Then, we explain the experiment settings, training details, evaluation settings, and image metrics in order. Finally, we introduce experiments design for evaluating the proposed network, and the result for each experiment.

4.1 Dataset and Preprocessing

4.1.1 Dataset

The dataset was from the Human Connectome Project (HCP)[*1], which contains structural and diffusion imaging data for 35 healthy adult subjects. Among them, 32 subjects matched our network input requirement that they both have T1-weighted and T2-weighted MR images.

These T1-weighted MRI images (3D MPRAGE) and T2-weighted MRI images (3D T2-SPACE) were generated on the customized Siemens 3T Connectom scanner. For the T1-weighted MRI image, the matrix size is 256x256x176, and its spatial resolution is

[*1] Data collection and sharing for this project was provided by the MGH-USC Human Connectome Project (HCP; Principal Investigators: Bruce Rosen, M.D., Ph.D., Arthur W. Toga, Ph.D., Van J. Weeden, MD). HCP funding was provided by the National Institute of Dental and Craniofacial Research (NIDCR), the National Institute of Mental Health (NIMH), and the National Institute of Neurological Disorders and Stroke (NINDS). HCP data are disseminated by the Laboratory of Neuro Imaging at the University of Southern California.

1mm isotropic. The T2-weighted MRI image has 320x320x256 matrix size and 0.7mm isotropic spatial resolution. The whole dataset was split into 29 subjects for training and 3 subjects for testing.

4.1.2 Image co-registration and NaN problem

For the reason that matrix size and spatial resolution differ in our case, the image co-registered was conducted between T1-weighted and T2-weighted MRI image for each subject by SPM12 (Statistical Parametric Mapping) software with the default setting. The example of SPM interface is shown in Figure 4.1.

First, the co-registered T2-weighted MRI image was converted to keep the same matrix size and spatial resolution as the T1-weighted MRI image. Then, the co-registered T2-weighted MRI images treated as the ground-truth HR T2-weighted image because of the dataset limitation.

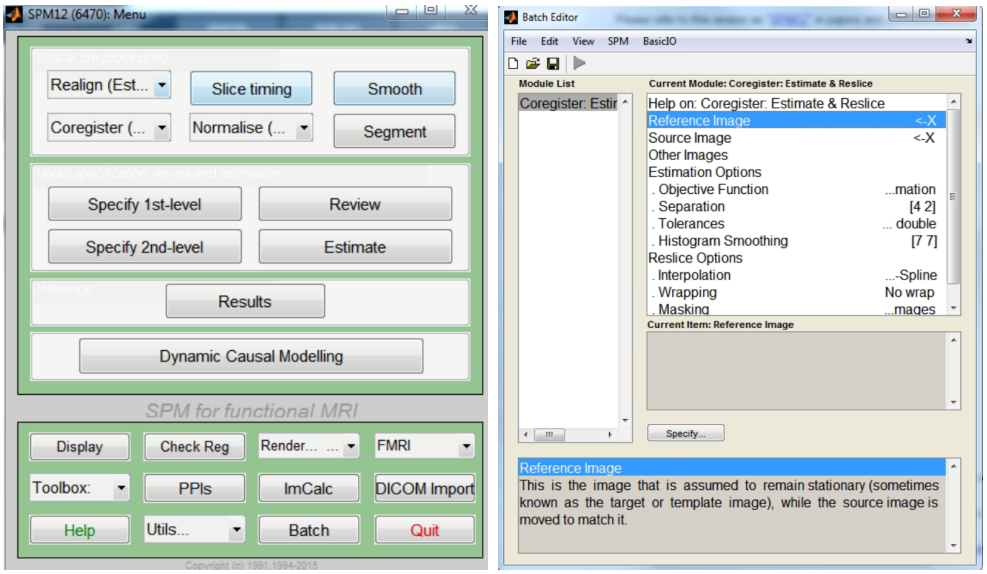
In the data pre-processing progress, the co-registered T2-weighted MRI image has NaN value in the 3D cube outer border part. These NaN values were set as 0.

4.1.3 Patch generation

As mentioned in the dataset section of this chapter, the available T1-weighted and T2-weighted MRI image dataset are small-scaled. However, the precondition of getting success from the deep neural network is large scale dataset.

In this case, we adopted the same data patching process as DCSRN. It could generate similar training samples and then improving the generalization capability of the proposed network.

For high-resolution T1-weighted and T2-weighted image, all the 64x64x64 cubes were randomly cropped from the original MR images after applied min-max normal-



(a) SPM menu interface (b) Coregister operation interface

Fig. 4.1 A example of SPM interface

ization. After the min-max normalization and low-resolution image generation step, the input 64x64x64 low-resolution cubes were cropped from the same position as high-resolution T2-weighted and T1-weighted cubes in the low-resolution T2-weighted images. Each high-resolution T1-weighted, high-resolution T2-weighted and low-resolution T2-weighted cube pairs was cropped from the same position from each image.

4.2 Experiment setting

4.2.1 Experiment environment

All the experiments were implemented by TensorFlow 1.12.0 and conducted on a computer with Intel Core i3-8100 3.6GHz processor, GeForce GTX 1080Ti GPU, and 16G memory. Table 4.1 shows the experiment hardware and software in detail.

Table 4.1 Experiment environment

Processor	Intel Core i3-8100 3.6GHz
GPU	GeFORCEGTX 1080Ti
Memory	16G
Operating System	Ubuntu 18.4.1
Programming Language	Python 3.6.8
Software Library	TensorFlow 1.12.0, Numpy 1.16.4

4.2.2 Training details

We used Adam optimizer [12] to minimize the mean squared error loss between the ground truth MRI images and network prediction MRI images. The setting of Adam optimizer are learning rate = 0.00001, $\beta 1 = 0.9$, $\beta 2 = 0.999$,ϵ = 1e-8.

4.2.3 Evaluation setting

For the multi-contrast super-resolution network, we shaved the 3D MRI image matrix size from 256x256x176 to 256x256x160 for HR T1-weighted, LR T2-weighted, and HR T2-weighted images after resolution degrading step. Then, all the network input 64x64x64 cubes for generating test SR T2-weighted images were produced by sliding windows with stride 32 for each direction. According to this, the final super-resolution T2-weighted image was reconstructed by overlapping and averaging 196 super-resolution T2-weighted output cubes.

Besides, preventing the checkboard artifacts, we shaved the 8 pixels from the edges of each super-resolution T2-weighted cubes. Then using these 48x48x48 T2-weighted cubes, we rebuilt the T2-weighted super-resolution image, and thus matrix size turns from 256x256x160 to 240x240x144.

The low-resolution image generation method for nearest-neighbor interpolation and bicubic interpolation comparison methods is different from multi-contrast super-resolution network. After the high-resolution T2-weighted 3D image applied min-max normalization, the normalized 3D image data convert to K-space data which has matrix size as 256x256x176 by applying FFT.

Let the direction size is 176 as the z-axis. Then, we deleted the same high-frequency area along the z-axis, it instead of setting them as zero. For deleting the same 75 zero percentage in K-space, Then, we combine the remaining part as low-resolution image k-space data which has 128x128x176 matrix size. In the end, the 3D low-resolution T2-weighted image was obtained from applying iFFT 128x128x176 k-space data. A simple illustration of deleting k-space data for generating low-resolution image is shown in Figure 4.2.

In this case, we upscaled the low-resolution 3D image along the z-axis. Then, we obtained 256x256x176 3D T2-weighted super-resolution image by applying nearest-neighbor interpolation or bicubic interpolation. Finally, we shaved the 256x256x176 3D T2-weighted super-resolution image into 240x240x144 matrix size.

For a fair comparison, the measurement of image quality is evaluated along the z-axis. It means the comparison image matrix size is 244x244. Besides, the center 128 slices in z-direction took part in the final image quality evaluation because these central part images are more close to the reality super-resolution application objects.

4.2.4 Image metrics

We used three common image metrics in this work for objective image quality assessment. The assessment was conducted by peak signal-to-noise ratio (PSNR), normalized root mean square error (RMSE) and structural similarity index (SSIM) to measure the differences between the ground truth image and super-resolution image.

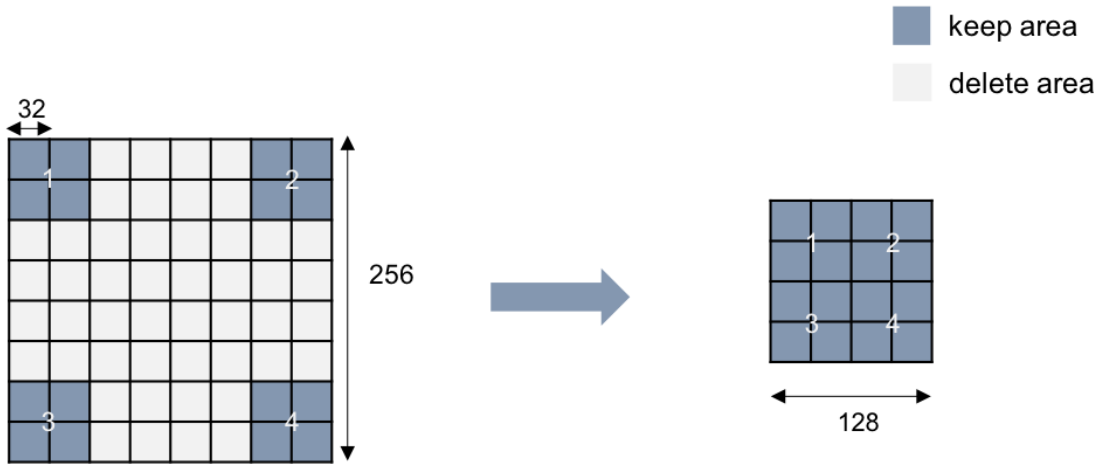


Fig. 4.2 The example of generating k-space data for the low-resolution image in bicubic interpolation and nearest-neighbor (NN) interpolation method

4.3 Experiment design

In our work, we designed three experiments to demonstrate the performance of multi-contrast MRI image network.

In experiment 1, the input MRI data types are T1-weighted, and T2-weighted MRI images that each of them generates 2900 cubes. The aim of experiment 1 is displaying the network performance in a small multi-contrast image cube training dataset.

Compared with experiment 1, experiment 2 increases the number of training cubes from 100 to 400 for each T1-weighted and T2-weighted image. The target for experiment2 is investigating the performance improvement that multi-contrast image network trained in a relatively larger dataset.

Experiment 3 was designed to demonstrate the proposed network performance using single contrast MRI images as input. For confirming the usefulness of using multi-contrast MRI image as network input rather than the single contrast MRI image, the input data of experiment 3 is generated from T2-weighted MRI images. For comparison, all the T2-weighted input cubes are the same as experiment 1.

Table 4.2 The input data detail for experiment 1 (MC-small), experiment 2 (MC-large), experiment 3 (SC-small), and experiment 4 (DCSRN)

Experiment	MC-small	MC-large	SC-small	DCSRN
MRI image type	T1w and T2w	T1w and T2w	T2w	T2w
Input cube number	29x100x2	29x400x2	29x100	29x100

For comparing the network performance with the interpolation method, we conducted nearest-neighbor (NN) interpolation and bicubic interpolation as another two reference experiment.

In addition to demonstrating the performance improvement of the proposed network, experiment 4 is treated as an additional experiment which has the same MRI input type and the same number of input numbers as experiment 3. The network architecture of experiment 4 is DCSRN. The input detail of each experiment is shown in Table 4.2.

4.4 Reconstruction results in 200 epochs

As the results in Table 4.3 displayed, experiment 2 achieved the highest average PSNR and SSIM than other experiments. In experiment 1 and 2, they have the same input MRI image type, but they have different numbers of input cubes. Within 200 training epochs, experiment2 achieved better quantitative performance than experiment1 in PSNR and SSIM. Also, experiment 2 has smaller NRMSE than experiment1. It means, with the number of input cubes increasing, the proposed network could achieve better super-resolution performance within the same training epochs.

As the same number of training epochs and the same amount of T2-weighted training cubes, comparing experiment 1 and 3, experiment 1 achieve better performance than

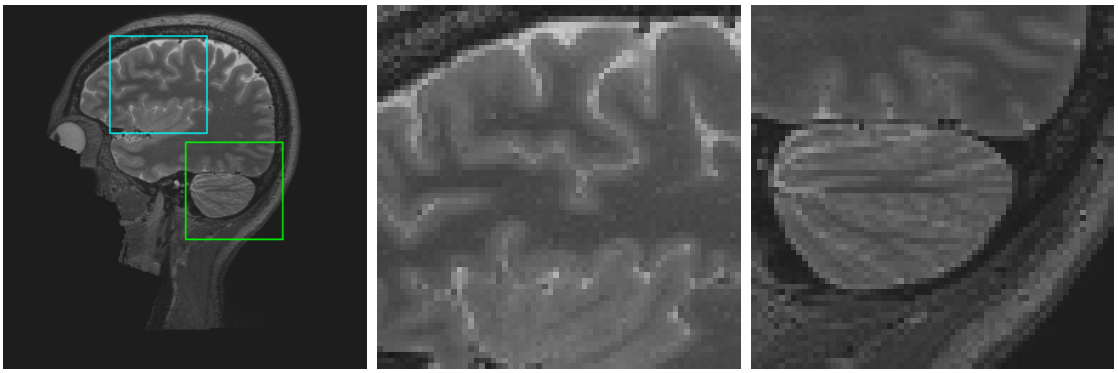
Table 4.3 The averge 240x240x128 super-resolution result of PSNR, NRMSE and SSIM for nearest-neighbor (NN) interpolation, bicubic interpolation, experiment 1 (MC-small), experiment 2 (MC-large), experiment 3 (SC-small), and experiment 4 (DCSRN)

Method	PSNR	NRMSE	SSIM
NN	31.2101	0.1313	0.8928
Bicubic	32.5267	0.1128	0.9153
MC-small	35.2543	0.0822	0.9451
MC-large	35.9975	0.0755	0.9532
SC-small	34.7401	0.0874	0.9438
DCSRN	33.8114	0.0972	0.9173

experiment 3 in all image metric. It shows that the proposed network could achieve higher performance with the help of the multi-contrast image (T1-weighted image).

Under the same input number, the same MRI image type, and the same training epochs, by comparing experiment 3 and 4, experiment 3 achieved better performance in all image metric than experiment 4. It shows that the proposed network achieved better network performance than the DCSRN network by adding the sigmoid activation function in the last convolution layer.Compared with nearest-neighbor interpolation and Bicubic interpolation, all the deep learning-based network achieved better super-resolution performance.

The randomly selected sample of ground truth image, and super-resolution image result with Nearest-neighbor (NN) interpolation method, bicubic interpolation method, experiment 1 – 4 are shown in the follow Figure 4.3 to Figure 4.9. The loss and SSIM value of experiment 1 – 4 in the training stage are shown in the Figure 4.10. The loss and SSIM value of experiment 1 – 4 in the testing stage are shown in the Figure 4.11.

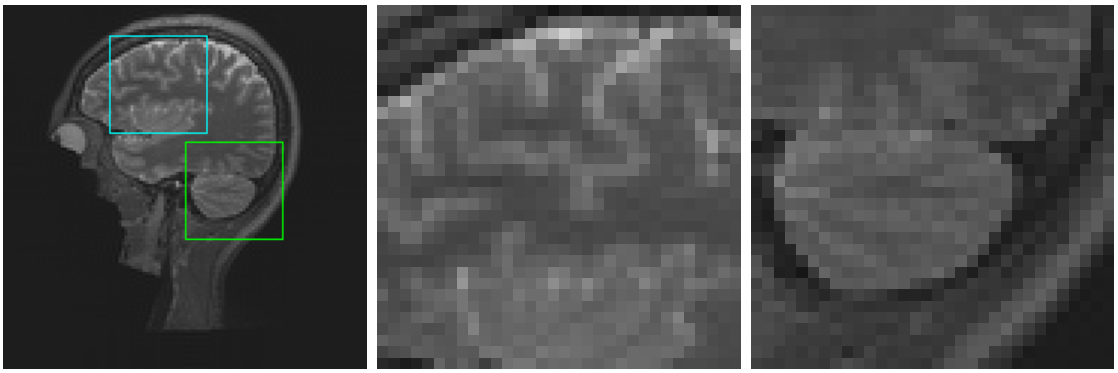


(a) Randomly selected sample ground truth image

(b) Magnified image of cerebellum area

(c) Magnified image of the juncture area of white matter and grey matter

Fig. 4.3 Illustration of randomly selected sample for the ground truth image

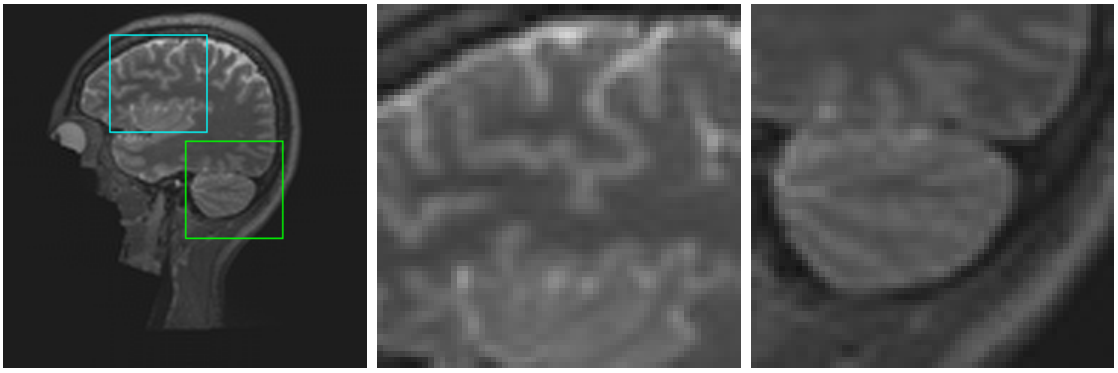


(a) Randomly selected sample super-resolution image

(b) Magnified image of cerebellum area

(c) Magnified image of the juncture area of white matter and grey matter

Fig. 4.4 Illustration of randomly selected sample for the super-resolution result in Nearest-neighbor (NN) interpolation method

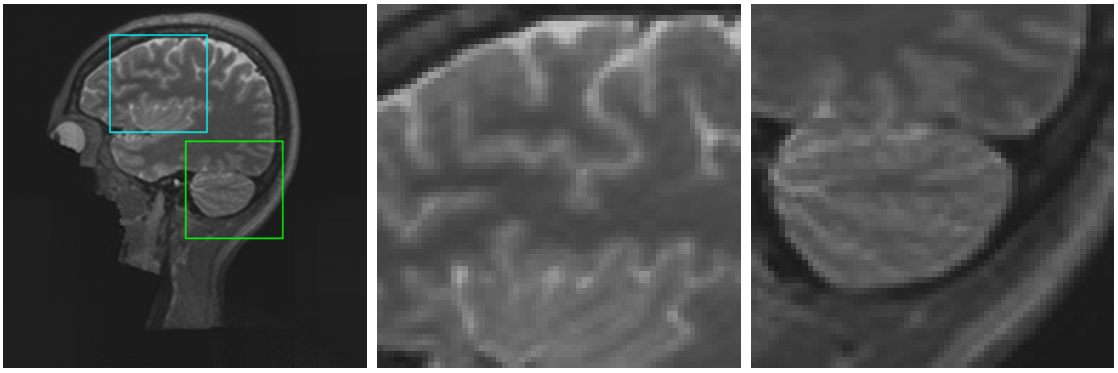


(a) Randomly selected sample super-resolution image

(b) Magnified image of cerebellum area

(c) Magnified image of the juncture area of white matter and grey matter

Fig. 4.5 Illustration of randomly selected sample for the super-resolution result in the bicubic interpolation method

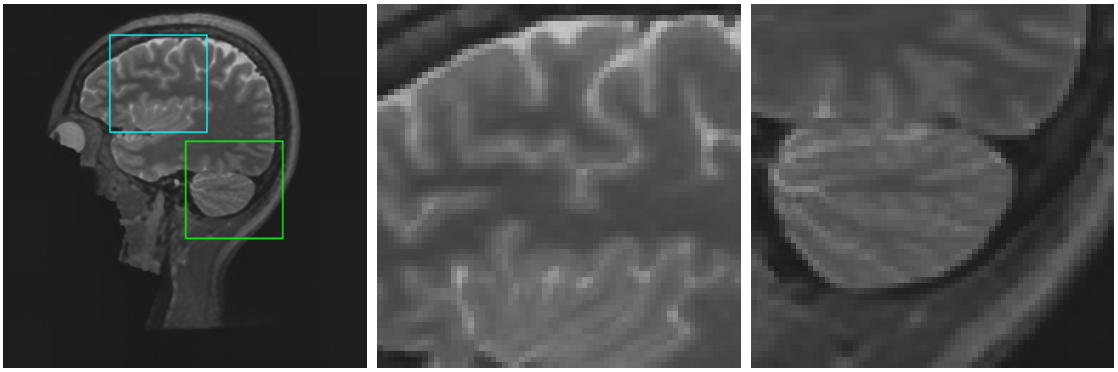


(a) Randomly selected sample super-resolution image

(b) Magnified image of cerebellum area

(c) Magnified image of the juncture area of white matter and grey matter

Fig. 4.6 Illustration of randomly selected sample for the super-resolution result in experiment 1 which based on multi-contrast super-resolution architecture with a small T1-weighted and T2-weighted dataset

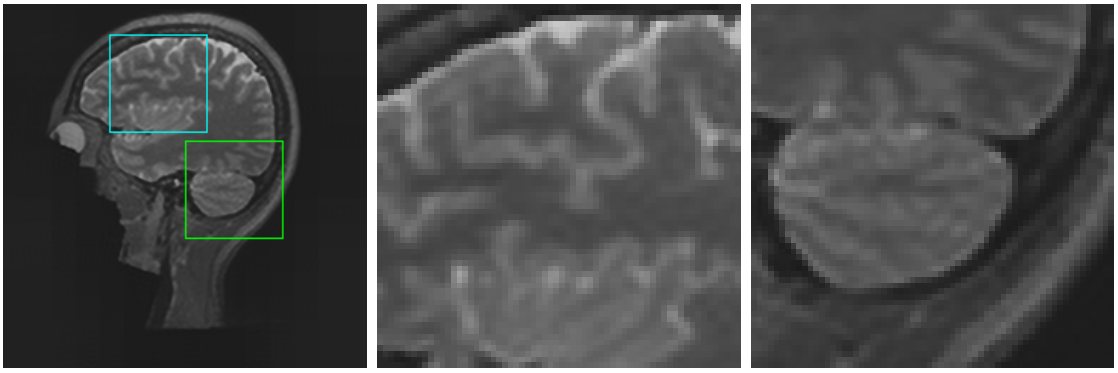


(a) Randomly selected sample super-resolution image

(b) Magnified image of cerebellum area

(c) Magnified image of the juncture area of white matter and grey matter

Fig. 4.7 Illustration of randomly selected sample for the super-resolution result in experiment 2 which based on multi-contrast super-resolution architecture with a large T1-weighted and T2-weighted dataset

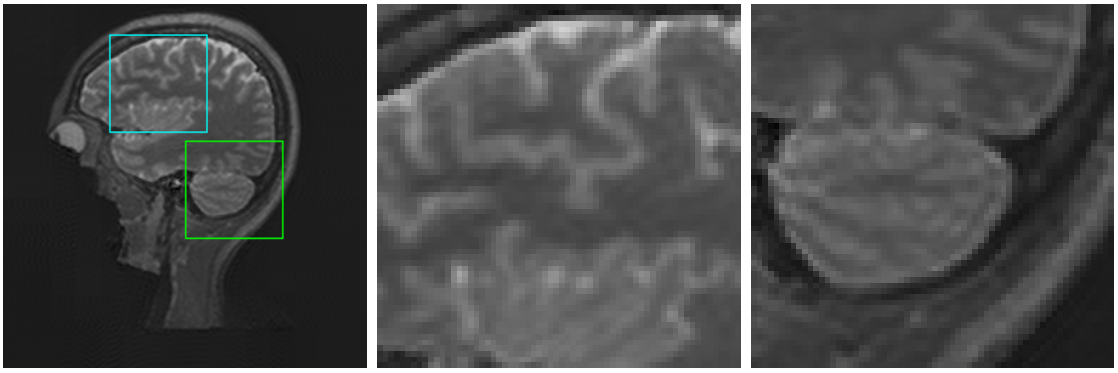


(a) Randomly selected sample super-resolution image

(b) Magnified image of cerebellum area

(c) Magnified image of the juncture area of white matter and grey matter

Fig. 4.8 Illustration of randomly selected sample for the super-resolution result in experiment 3 which based on proposed network architecture with a small T2-weighted only dataset

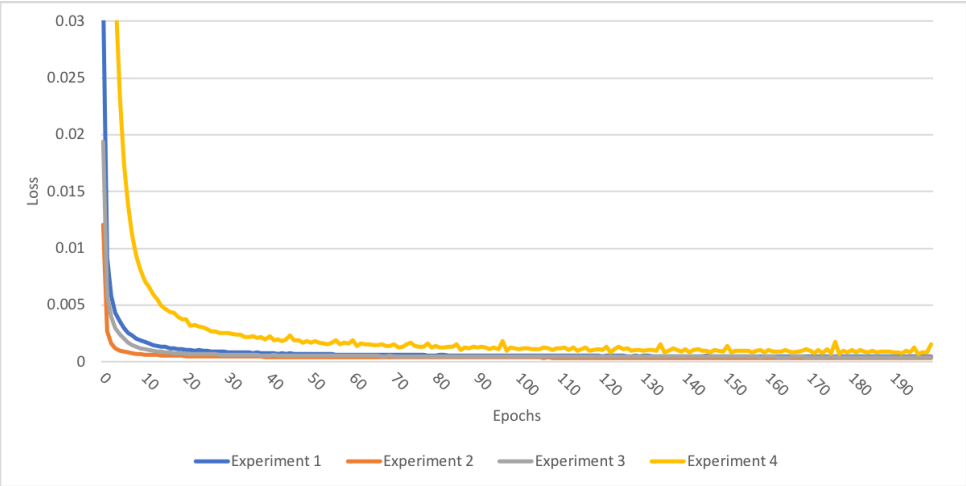


(a) Randomly selected sample super-resolution image

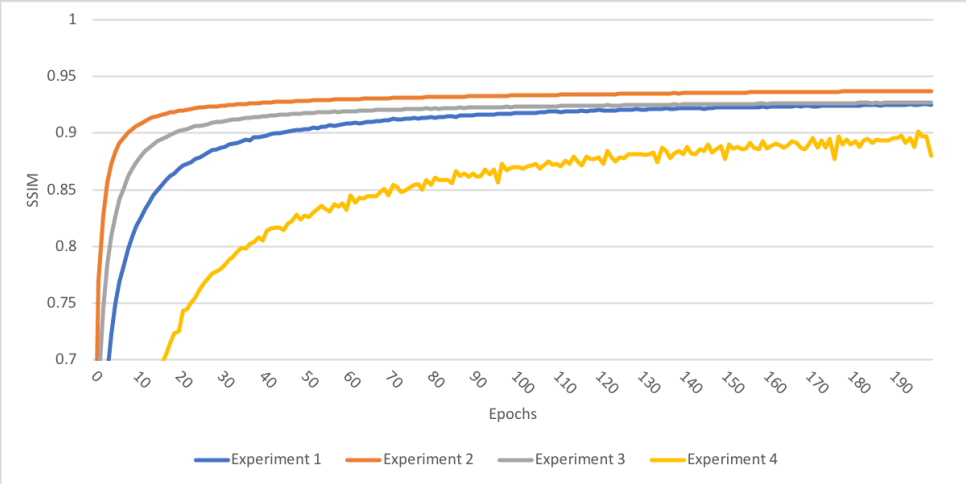
(b) Magnified image of cerebellum area

(c) Magnified image of the juncture area of white matter and grey matter

Fig. 4.9 Illustration of randomly selected sample for the super-resolution result in experiment 4 which based on DCSRN architecture with a small T2-weighted only dataset

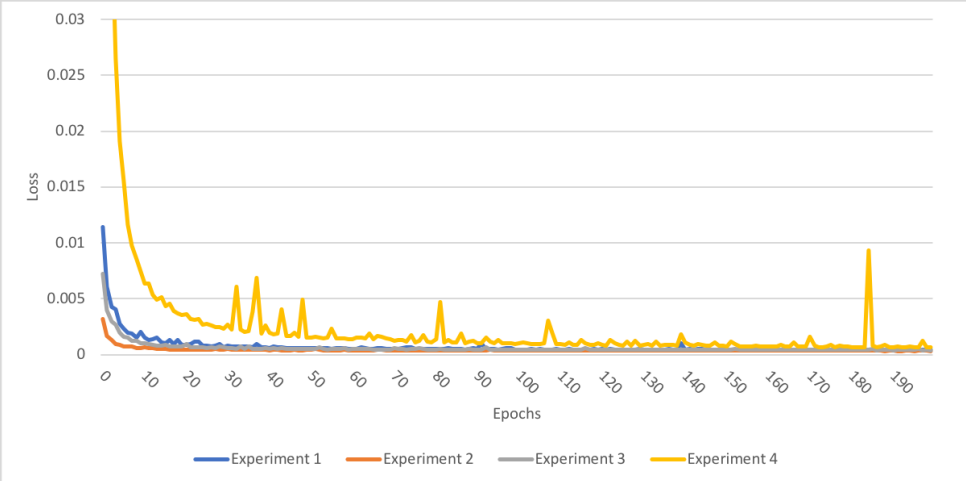


(a) Training loss value trend

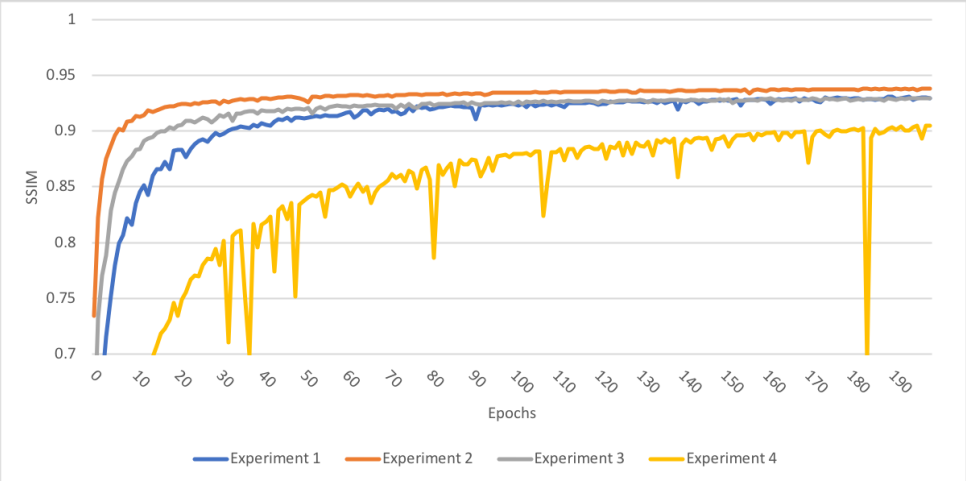


(b) Training SSIM value trend

Fig. 4.10 Experiment 1 - 4 training loss and SSIM trend under 200 epochs



(a) Testing loss value trend



(b) Testing SSIM value trend

Fig. 4.11 Experiment 1 - 4 testing loss and SSIM trend under 200 epochs

4.5 Acquisition time

The acquisition time for 256x256x176 matrix size high-resolution T1w MRI image takes 6 minutes 2 seconds. The acquisition time for 320x320x256 matrix size high-resolution T2w MRI image takes 6 minutes 48 seconds.

We assume the acquisition time for 256x256x160 matrix size high-resolution T2w MRI image approximately takes 2 minutes 43 seconds. According to the 75 percent of zero areas in k-space, we presume the 256x256x160 low-resolution T2w MRI image takes around 41 seconds.

In the testing phase, the average generation time of 256x256x160 super-resolution T2w MRI image takes 1 minute 18 seconds. In this case, taking a low-resolution T2w MRI image and using proposal network to generated super-resolution T2w MRI image used 1 minute 59 seconds. Compared with directly taking high-resolution 256x256x160 T1w and T2w MRI images, the proposal network saves 44 seconds for each subject.

Chapter 5

Conclusion

In this thesis, inspired by the multi-contrast images in MRI, we proposed a 3D multi-contrast super-resolution network that was based on the current state-of-the-art single image super-resolution network DCSRN. For demonstrating the network performance, we conducted several experiments in the dataset of the Human Connectome Project (HCP).

By comparing with different super-resolution methods, such as nearest-neighbor interpolation and bicubic interpolation, the result first confirmed the idea that is using T1-weighted high-resolution MRI image as a hint to improve the T2-weighted low-resolution MRI image quality works. Further, it confirms that the multi-contrast input network provides better performance than using only a single-contrast image as input in super-resolution image reconstruction. Though enlarge the proposed network input dataset size, the network could produce higher super-resolution performance.

Base on the current work, there is a possibility to extend our work in the following directions as future work:

1. using the structural MRI image as a hint to improve the resolution of diffusion-weighted imaging.
2. using the structural MRI image as a hint to enhance the resolution of the fMRI image.
3. implement the 3D multi-contrast network with Generative Adversary Networks [13].

Acknowledgement

In here, I would like to extend the most sincere appreciation and heartfelt thanks to my supervisor Prof. Matsuzaki. In the eight months exchange student life and two years of master student life, when I met a problem in academic, he will explain it several times using different method until I completely understand it. For researching, he always will point out the mistake that I made immediately and urge me to think about the researching problem deeply. His serious scientific attitude and enthusiastic teaching attitude had a profound impact on me. Among this period, I touched many different fields in computer science under his guiding. He always helped me as possible as he can. Further, in this master thesis writing, he gave many valuable opinions from the topic selection, the related materials collection, and the completion of the thesis writing. Without his help, this work might not be done.

In addition, I would like to say thanks to Prof. Iwata and Assoc. Prof. Yoshida as the sub-chief examiner of my master thesis. Also, I would like to say thanks to everyone in Matsuzaki laboratory.

Data collection and sharing for this project was provided by the MGH-USC Human Connectome Project (HCP; Principal Investigators: Bruce Rosen, M.D., Ph.D., Arthur W. Toga, Ph.D., Van J. Weeden, MD). HCP funding was provided by the National Institute of Dental and Craniofacial Research (NIDCR), the National Institute of Mental Health (NIMH), and the National Institute of Neurological Disorders and Stroke (NINDS). HCP data are disseminated by the Laboratory of Neuro Imaging at the University of Southern California.

References

[1] Yuhua Chen, Yibin Xie, Zhengwei Zhou, Feng Shi, Anthony G. Christodoulou, and Debiao Li. Brain MRI super resolution using 3d deep densely connected neural networks. *CoRR*, abs/1801.02728, 2018.

[2] Elizabeth A Krupinski. Current perspectives in medical image perception. *Attention, Perception, & Psychophysics*, 72(5):1205–1217, 2010.

[3] Wenming Yang, Xuechen Zhang, Yapeng Tian, Wei Wang, Jing-Hao Xue, and Qingmin Liao. Deep learning for single image super-resolution: A brief review. *IEEE Transactions on Multimedia*, 2019.

[4] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *CoRR*, abs/1501.00092, 2015.

[5] Chi-Hieu Pham, Aurélien Ducournau, Ronan Fablet, and François Rousseau. Brain mri super-resolution using deep 3d convolutional networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 197–200. IEEE, 2017.

[6] Gao Huang, Zhuang Liu, and Kilian Q. Weinberger. Densely connected convolutional networks. *CoRR*, abs/1608.06993, 2016.

[7] David C Van Essen, Kamil Ugurbil, E Auerbach, D Barch, TEJ Behrens, R Bucholz, Acer Chang, Liyong Chen, Maurizio Corbetta, Sandra W Curtiss, et al. The human connectome project: a data acquisition perspective. *Neuroimage*, 62(4):2222–2231, 2012.

[8] François Chollet. *Deep Learning with Python*. Manning, November 2017.

[9] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, Nov 1998.

[10] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[11] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, abs/1502.03167, 2015.

[12] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.