



MedSRGAN: medical images super-resolution using generative adversarial networks

Yuchong Gu¹ · Zitao Zeng¹ · Haibin Chen¹ · Jun Wei² · Yaqin Zhang³ · Binghui Chen³ · Yingqin Li³ · Yujuan Qin³ · Qing Xie³ · Zhuoren Jiang¹ · Yao Lu^{1,4}

Received: 2 June 2019 / Revised: 12 March 2020 / Accepted: 22 April 2020 /

Published online: 13 May 2020

© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Super-resolution (SR) in medical imaging is an emerging application in medical imaging due to the needs of high quality images acquired with limited radiation dose, such as low dose Computer Tomography (CT), low field magnetic resonance imaging (MRI). However, because of its complexity and higher visual requirements of medical images, SR is still a challenging task in medical imaging. In this study, we developed a deep learning based method called Medical Images SR using Generative Adversarial Networks (MedSRGAN) for SR in medical imaging. A novel convolutional neural network, Residual Whole Map Attention Network (RWMAN) was developed as the generator network for our MedSRGAN in extracting the useful information through different channels, as well as paying more attention on meaningful regions. In addition, a weighted sum of content loss, adversarial loss, and adversarial feature loss were fused to form a multi-task loss function during the MedSRGAN training. 242 thoracic CT scans and 110 brain MRI scans were collected for training and evaluation of MedSRGAN. The results showed that MedSRGAN not only preserves more texture details but also generates more realistic patterns on reconstructed SR images. A mean opinion score (MOS) test on CT slices scored by five experienced radiologists demonstrates the efficiency of our methods.

Keywords Medical images · Super-resolution (SR) · Deep learning · Generative adversarial networks (GAN)

✉ Haibin Chen
chenhb63@mail.sysu.edu.cn

✉ Zhuoren Jiang
jiangzhr3@mail.sysu.edu.cn

✉ Yao Lu
luyao23@mail.sysu.edu.cn

Extended author information available on the last page of the article

1 Introduction

Medical images, including computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), are popular for clinical applications such as non-invasive diseases diagnosis, anatomic imaging and treatment planning [2, 7, 17, 24, 25, 32, 37, 38, 43, 47]. However, there are some limitations in these imaging technologies. For example, the radiation injury is inevitable during CT scanning. Low-dose CT (LDCT) is a clinical recommended technology for reducing the radiation injury to patients, but at the expense of low image resolution and noise contamination. MRI is known as its expressive capacity in providing anatomical, metabolic, and functional information for various regions of human bodies. However, one of the major disadvantages of MRI is the long acquisition duration which makes the image quality is susceptible to patient movement. In order to obtain image with higher single noise ratio (SNR), its spatial resolution tends to be coarse than CT images. Furthermore, low magnetic field of MRI scanner also constrains the spatial resolution of MRI. PET is able to detect the tumor in its early stage accurately by imaging the different absorption capacity of labeled compounds with radioisotope in different parts of the body. The imaging theory of PET is based on detecting the quantity of photon pairs produced by annihilation reaction of radioisotope. Due to the scarcity of photon pairs and noise interference, the low-resolution problem is also a challenge for PET imaging.

To improve the image quality with low radiation dose and technology limitations, one of the solutions is retrieving image details based on low resolution (LR) image to reconstruct high-resolution (HR) [12, 29], such process is widely known as super-resolution (SR). Although clinical obtained LR medical images contain realistic and fundamental information, they lose many details including high-frequency components, certain patterns like tissue background and textures which are extremely necessary for reconstructing HR images. Therefore, SR for medical images is still a big challenge and hardly used in practice. In this work, we propose a novel medical image SR framework based on deep learning and show its effectiveness.

1.1 Related works

In recent years, a large range of successful neural network based algorithms were proposed for SR on natural images. According to different purposes, these methods can be mainly divided into two main categories: high peak signal-to-noise ratio (PSNR) / structural similarity (SSIM) and high perceptual quality. Methods with higher PSNR/SSIM were mainly achieved by network structure innovations such as deeper networks, various feature connections and feature flowing strategies inside networks [8, 26, 27, 33, 41, 49]. With the help of mean square error (MSE) loss, they have got prominent results on PSNR/SSIM metrics, but SR images still contradict with human observation because human eyes are good at catching high frequency components rather than telling pixel-level difference. On the other hand, methods with higher perceptual quality [30, 45] were mainly come out with perceptual loss [22] and GAN [11]. These methods performed well on perceptual index metrics but compromise on PSNR/SSIM metrics.

Deep learning based medical image restoring researches gradually get more attentions in medical imaging research communities in recent years, these researches include image translation between different imaging systems [1, 19], super resolution [3, 4, 31, 34, 35, 39, 48], and image de-noising [1, 10, 20].

1.1.1 High PSNR/SSIM of natural image

Dong et al. [8] proposed Super-Resolution Convolutional Neural Network (SRCNN) for single-image SR (SISR), which was the first convolutional neural network (CNN) research in SISR problem. SRCNN first enlarge input image to target size with bicubic interpolation, then feed it into a simple network with three convolutional layers. The method was already able to achieve higher PSNR/SSIM than any other traditional computer vision algorithms. This research reveals that end-to-end learning methods can effectively learn those necessary nonlinear transformation on pixel level when enlarging an image. Since then, deep learning has become the main stream for SISR on natural image. Many attempts have been done in this field to achieve higher performance on PSNR/SSIM, such as using recursive convolutional network for feature extraction [26], stacking more convolutional filters and adding long residual connections [27], and removing batch normalization layers to preserve range flexibility from networks [33]. Shi et al. [41] proposed an efficient up-sampling module for SR, called pixel shuffle module, which simply rearranges $H \times W \times r^2$ image tensor into $rH \times rW \times 1$ to enlarge the height and width of image features. This process does not contain learnable parameters, and it reduce the channels of feature. Researchers often put an extra convolutional layer before pixel shuffle module to increase the feature channels, which makes up-sampling process learnable and preserves feature channels. Based on these ideas, Zhang et al. [49] designed a very deep CNN with stacked channel attention modules, called Residual Channel Attention Network (RCAN), and achieved state-of-the-art PSNR/SSIM.

All these methods are considered as PSNR-oriented methods because they only take mean square error (MSE) or mean absolute error (MAE) as loss function during the network training, which is an intuitive idea to increase PSNR metric. However, the generated images would still become over-smoothed [30, 45]. Since PSNR often disagrees with visual evaluation by human observers, such results are intolerable in medical imaging applications, which may hinder related clinical applications such as disease diagnosis and lesion detection.

1.1.2 High perceptual quality of natural image

To restore feasible high-frequency components and obtain better visual results in SISR, researchers apply loss on feature level [22] and use GAN [11] to make images as realistic as possible. Loss on feature level, also known as perceptual loss, is widely used on many image reconstruction tasks including pixel-to-pixel image translation (pix2pix) [18], image style transfer [9, 21] and super resolution [30, 45]. It takes a pre-trained neural network to extract features for image outputs and ground truth targets separately, and then calculates loss on these two features, to guide networks to learn semantic features from target images. GAN is initially used in fake image synthesis [11]. It usually contains two independent neural networks with different parameters, one for image generation, and the other for real/fake image discrimination. The target for discrimination network is whether an image is real or produced by generation network, which means fake, while the target for generation network is the adversarial form of that for discriminator network, so it makes fake images get close to real images. In the training process, generation network gradually narrows the gap between produced images and real images until discriminator can no longer tell the differences.

Ledig et al. [30] first proposed a GAN-based SR framework (SRGAN) which contains a generator network for high resolution image generation, a discriminator network for recognizing generated images from real world images, and loss function which includes perceptual loss and GAN loss. SRGAN had got much better visual results than previous PSNR-oriented methods though it compromised some PSNR/SSIM values. In Wang et al.'s work ESRGAN (Enhanced-

SRGAN) [45], they proposed a densely connected [16] Residual in Residual Dense Block (RRDB) network for generator and used relativistic discriminator [23] to determine whether an image is more realistic than the other image, and they won PIRM-SR Challenge (ECCV2018) on high perceptual quality group.

1.1.3 Reconstruction works for medical image

Aforementioned researches have achieved great progress in SISR tasks, however, none of these studies considered the characteristics of medical images nor evaluated their performances on medical images. It is undoubtedly infeasible to directly apply their models in medical images SR due to the totally different data distribution, patterns and textures.

For medical image reconstruction works, Armanious et al. [1] adapted ideas from pix2pix research and proposed a GAN-based framework named MedGAN for PET-CT translation, MRI motion correction and PET de-noising. You et al. [48] proposed a complicated GAN-based framework with ideas from cycle GAN [50] for $2\times$ single-slice CT SR, and it requires independent LR and HR datasets. Due to unnecessary radiation injuries for collecting more CT scans on the same patients, they only used a tibia dataset from fresh-frozen cadaveric ankle specimens and a public abdominal dataset with a small amount of scans for experiments. Furthermore, they only focused on very small local patches instead of the whole CT scans, so the practicality of this approach is limited in clinical applications. As for MRI SR, Chen et al. [3, 4] proposed 3D Densely Connected Super-Resolution Networks (DCSRN) to restore HR features of structural brain MRI and they further developed a GAN framework for guiding DCSRN training in order to further improve the SR quality. Mardani et al. [35] proposed GAN for Compressive Sensing (GANCS) to retrieve higher quality images with improved fine texture details. Although visually reliable results were observed in these methods, they did not evaluate whether these produced images would be acceptable by radiologists nor affect clinical diagnosis.

1.2 Our contribution

In this work, we develop a medical images SR framework using generative adversarial networks (MedSRGAN) for reconstructing reliable and visually realistic SR medical images, which takes a LR medical image as input to generate a $4\times$ SR image. In MedSRGAN, we employ an improved medical image generation network, Residual Whole Map Attention Network (RWMAN) as SR image generator, a new pairwise discriminator to distinguish the pairs of both HR/SR and LR images as well as a novel multi-task loss function combining the content loss, adversarial loss, and adversarial feature loss for guiding the SR image to obtain more reliability and feasibility. Our methods were evaluated on 242 thoracic CT scans and 110 brain MRI scans, and our results would bring marginal influence on disease diagnosis, which reveal that SR based medical imaging systems are possible to be applied in practice.

2 Methods

MedSRGAN is a typical GAN-based architecture. As is shown in Fig. 1, it consists of two neural networks: generator and discriminator, and two independent loss functions for each network. The targets of discriminator are from whether an image is ground truth or generated image, while targets of generator come from discriminator and image contents.

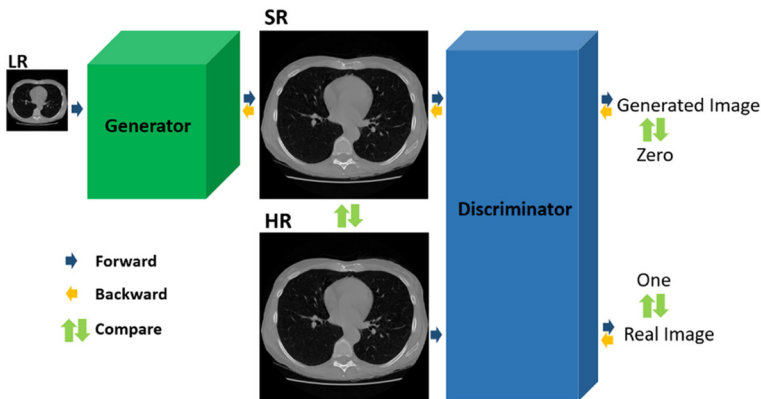


Fig. 1 Basic framework of MedSRGAN. Arrows of different colors defined how data flow forward for calculation, backward for parameters updating, and where to make comparison for loss function

2.1 Generator structure

As shown in Fig. 2, we named our generator as Residual Whole Map Attention Network (RWMAN), which is a modification of RCAN [49] by replacing its Residual Channel Attention Block (RCAB) with Residual Whole Map Attention Block (RWMAB). RWMAN extracts abundant features on LR scale and up-samples after these steps. It consists of an input convolution, RWMAB groups with long and short residual connections, and up-sampling module with two Conv-PixelShuffle blocks.

In RCAN [49], RCAB uses global average pooling to squeeze the whole feature map into a single value through all channels at first, then makes these values learnable in the following layers, and finally treats them as weights for previous channels, such process is regarded as channel attention. Attention mechanism has improved the performance of RNN models such as LSTM and GRU in many NLP tasks [6], and it is now widely used in other different domains including recommending system [5, 13] and computer vision [15, 44, 46]. Generally, this mechanism apply weights on current feature adaptively by adding a sub-network on this feature, and channel attention learns weights through channels of features. This step helps neural network adaptively learn how to use information wisely through different channels, but it neglects attentions on different regions of an image. Generally, the whole map of a natural image should be equally focused on to reconstruct a higher resolution one, since all pixels may reflect the meaningful information in real world. However, for many medical images, only those regions with useful information should be wisely considered, surrounded areas such as air and vacant areas are meaningless. We expect the neural network not only extracts useful information through different channels, but also pays more attention on meaningful regions.

In RWMAB, we use a 1×1 convolutional layer to obtain a tensor having the same shape of the input image tensor, followed by a sigmoid activation to form weights range from 0 to 1 for every pixel through all channels. This sub-network structure helps our model enlarge or lower the effect of each pixel adaptively, to enable our GAN-based loss function make influence. We stack 128 such neural network blocks with long (from the beginning to end) and short (for each 16 blocks) residual connections to reduce the training difficulties [14].

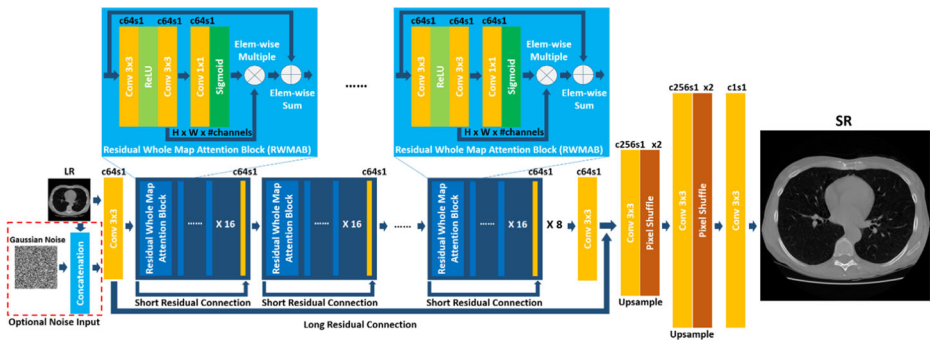


Fig. 2 The generator structure Residual Whole Map Attention Network (RWMAN). For each convolutional layer, channel (c) and strides (s) are pointed out above. ‘ $\times 2$ ’ on pixel shuffle module means it double the height and width of an feature by pixel rearrangements. The red dashed box represents an optional noise input. This network outputs a $4\times$ super-resolution image for each input image

In up-sample module, we use two $2\times$ Conv-PixelShuffle [41] modules to generate a $4\times$ image feature, with the number of feature channels kept, then reduce channels by a 3×3 convolutional layer to make output image.

Additionally, a random Gaussian noise with zero mean and unit variation was introduced as an additional channel for perturbation before input a LR image into neural networks, as shown in Fig. 2. This is an optional operation which is denoted by ‘+’ in this study. For instance, the MedSRGAN with additional noise channel is denoted by MedSRGAN+. GAN framework taking noise and meaningful information as input is regarded as a standard conditional GAN [18, 36]. By doing so, feature maps inside the neural networks are provided with some randomness, and it may help neural networks to be more adaptive on generating more feasible patterns in homogeneous regions, especially in tissue backgrounds, while the major content of images still come from input LR images.

For convenience, HR image generated by our neural network with corresponding LR image is named SR (super-resolution) image in the following descriptions:

$$SR = G(LR) \quad (1)$$

where $G(\cdot)$ is the operation of a given generator. With this notation, HR represents ground truth image while SR means generated image in this study.

2.2 Discriminator structure

Instead of having a single HR/SR image as input of discriminator, we used image pair (LR, HR/SR) as input to discriminate the SR image with a given LR image. Fig. 3 shows the architecture of our discriminator. With this design, the discriminator is specified to learn the pairwise information of both HR/SR and LR images by concatenating feature maps extracted from the LR and HR pathway, and outputs the probability of a (LR, HR) or (LR, SR) pair as a real pair.

With CT slice pair of (LR, HR) as 1 and (LR, SR) as 0, the training stage for discriminator is expressed as:

$$\begin{aligned} D(LR, HR) &\rightarrow 1 \\ D(LR, SR) &\rightarrow 0 \end{aligned} \quad (2)$$

where $D(\cdot, \cdot)$ is the operation of discriminator.

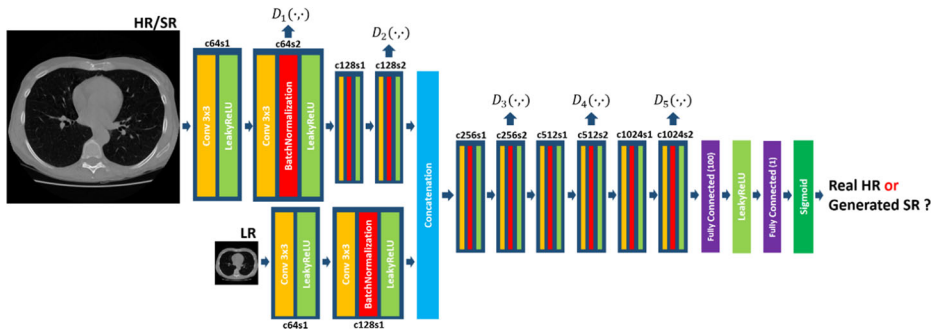


Fig. 3 Discriminator Structure. For each convolutional layer, channel (c) and strides (s) are pointed out above, and all LeakyReLU used $\alpha = 0.2$ on negative axis. $D_i(\cdot, \cdot)$, $i \in \{1, 2, \dots, 5\}$ indicates we output features in after these layers to calculate adversarial feature loss

2.3 Loss function

During the training, either a generator or a discriminator is trained alternately at each iteration with different data to avoid model parameters falling into a local optimum. In this study, Binary Cross Entropy (BCE) loss is used to train discriminator:

$$L_d = -\log(D(LR, HR)) - \log(1 - D(LR, SR)) \quad (3)$$

where $D(\cdot, \cdot)$ is the operation of discriminator.

A weighted sum of content loss, adversarial loss and adversarial feature loss is employed as loss function for generator training:

$$L_g = L_{cont} + \lambda_1 L_{adv} + \lambda_2 L_{advfeat} \quad (4)$$

where λ_1 and λ_2 are tunable hyperparameters for balancing the impacts of each part of loss.

2.3.1 Content loss

This part of loss is mainly responsible for the content restoration of an image. Generally, to obtain higher PSNR, content loss between HR and SR can be simply set as the L1 loss or MSE loss:

$$L1 = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N |HR(i, j) - SR(i, j)|$$

$$MSE = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N (HR(i, j) - SR(i, j))^2 \quad (5)$$

where N represents the size of images, assuming that each image is N by N .

As shown in Eq. (5) and (9), minimizing L1 or MSE is a PSNR-oriented optimization, but it fails to reconstruct the high frequency content and preserve visual realistic characteristics [30, 45]. For medical images, texture features are very important because these high-frequency information are decisive in human visual judgment for the quality of images. Thus, merely using PSNR-oriented loss function is not sufficient for medical image SR.

A well pre-trained neural network, such as VGG-16 or VGG-19 [42], has been used to extract texture features on a hidden layer to calculate content loss in many tasks including image style transfer [9, 21], image translation [18] and image SR [22, 30, 45]. In this work, we used a pre-trained VGG-19 to calculate content loss. Rather than obtaining loss on a single VGG hidden block, we used five hidden blocks to fully extract texture features on each semantic level and combine losses with more weights on shallower blocks and less weights on deeper blocks. This is because deeper features bring more abstract semantic information, while shallow features show relatively more concrete texture features information, as is shown in Fig. 4. With $V_{i,j}$ representing the VGG-19 network operation by the j -th convolution before the i -th max-pooling layer, our content loss is defined as:

$$L_{cont} = \lambda_{L1} \cdot L_1(HR, SR) + \sum_{k=1}^5 w_k \cdot MSE(V_{(i,j)_k}(HR), V_{(i,j)_k}(SR)) \quad (6)$$

where λ_{L1} is a tunable parameter for balancing the direct L1 loss between HR and SR image, weights $w = \{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}\}$ and layer numbers $(i,j) = \{(1, 2), (2, 2), (3, 4), (4, 4), (5, 4)\}$ for 5 hidden blocks, $L_1(\cdot, \cdot)$ and $MSE(\cdot, \cdot)$ represents the calculation of the L1 norm (mean absolute error) and mean square error of two features respectively. Followed by the analysis of VGG feature maps in ESRGAN [45], we used feature maps before ReLU activation for all hidden blocks to avoid the information missing after ReLU activation.

2.3.2 Adversarial loss

Adversarial loss aims to help generator make SR image closer to realistic HR image as much as possible to deceive the discriminator. The adversarial loss L_{adv} has a reciprocal form of L_d , aiming to make $D(LR, SR)$ get closer to 1:

$$L_{adv} = -\log(1-D(LR, HR)) - \log(D(LR, SR)) \quad (7)$$

where $D(\cdot, \cdot)$ is the operation of discriminator.

2.3.3 Adversarial feature loss

Like VGG based content loss, we also view our discriminator as a feature extractor and calculated the loss of hidden blocks between HR and SR images and between (HR, LR) and (SR, LR) pairs. Image examples are shown in Fig. 5. Minimizing the adversarial feature loss is

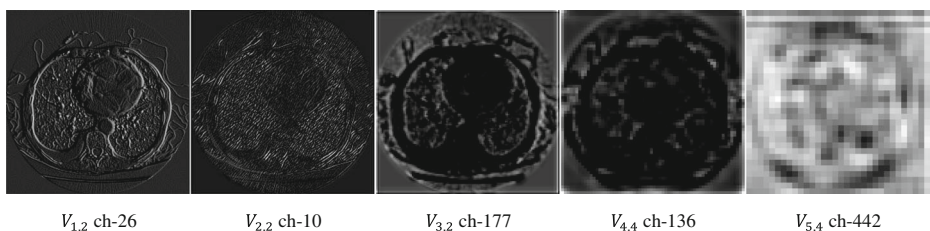


Fig. 4 Output examples of VGG-19 with $V_{i,j}$ denotes the j -th convolution before the i -th maxpooling layer (before ReLU activation) on one CT slice image, which shows that shallow layers get concrete texture information while deep layers get abstract information. ‘ch-’ with a number indicates the channel number of $V_{i,j}$ feature map. **Best viewed on screen with zoom in**

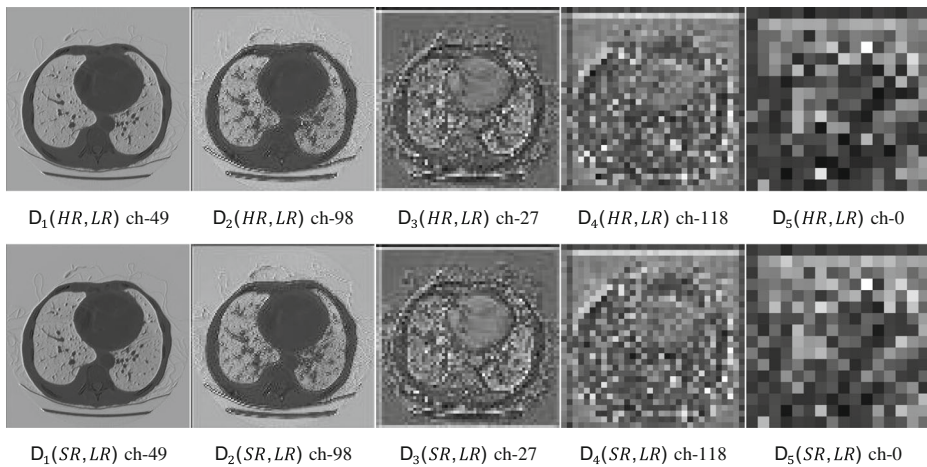


Fig. 5 Output examples of our discriminator with D_i denotes the i -th convolutional block (as is shown in Fig. 3) on one CT slice image. The SR images used in the second row were generated by MedSRGAN+, it can be observed that the feature map between HR and SR by MedSRGAN+ are very similar. ‘ch-’ with a number indicates the channel number of D_i feature map. **Best viewed on screen with zoom in**

to minimizing differences of texture features viewed by discriminator so that it helps generator benefit from discriminator features to make more realistic patterns.

With $D_k(\cdot, \cdot)$ represents the k -th hidden block inside the discriminator as is shown in Fig. 3, our adversarial feature loss is defined as:

$$L_{advfeat} = \sum_{k=1}^5 w_k \cdot MSE(D_k(LR, HR), D_k(LR, SR)) \quad (8)$$

where weights $w = \{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}\}$ for 5 hidden blocks, and $MSE(\cdot, \cdot)$ represents the calculation of the mean square error of two features.

3 Experiments

In this study, MedSRGAN and MedSRGAN+ were compared with bicubic interpolation and CNN based SR frameworks including RCAN [49], ESRGAN [45] and PSNR-oriented RWMAN (denoted as RWMAN (P) in experiments). RCAN is a PSNR-oriented method and has achieved state-of-the-art PSNR and SSIM on several typical SISR datasets such as Set5, Set14 and Urban100, ESRGAN won PIRM-SR Challenge (at ECCV2018 Workshop, a super resolution competition) on high perceptual quality group, which means they did best on high frequency components reconstruction and got highest perceptual index. RWMAN (P) is the RWMAN trained using MSE loss on pixel level.

3.1 Evaluation

PSNR and SSIM are still used for the SR framework evaluation in all methods and the SR community is lack of reliable objective metrics that can imitate judgments by human observation. PSNR and SSIM are calculated as:

$$MSE = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N (x(i,j) - y(i,j))^2$$

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX^2}{MSE} \right)$$

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x + \sigma_y + c_2)} \quad (9)$$

where x , y , N are generated image, ground truth image, and the size of image respectively, MAX is the maximum value of gray scale; μ , σ represent mean and variance, σ_{xy} is the covariance of two images; and two constants $c_1 = (0.01 \cdot MAX)^2$, $c_2 = (0.03 \cdot MAX)^2$ are calculated with the convention of SSIM.

Since PSNR and SSIM often contradict with individual visual judgment, we also perform mean opinion score (MOS) test to qualitatively evaluate visual quality of SR images. Five experienced radiologists are required to give an integer score range from 1 to 5, which represent very annoying, annoying, slightly annoying, perceptible but not annoying, and imperceptible, to an image dataset with hidden tags, which includes 100 randomly selected HR CT slice images and corresponding SR images generated by MedSRGAN and other reference methods.

3.2 CT experiments

3.2.1 Data and training details

242 thoracic CT scans from LUNA 16 Challenge [40] are used in our experiments. 219 randomly selected scans (52,073 slices) are used for training and the remaining 23 scans (5889 slices) for testing. All CT slices are 512×512 and used as HR references. LR slices are obtained by applying 4×4 average pooling on corresponding HR slices and have size of 128×128 . Moreover, 30 additional thoracic CT scans (10,732 slices) collected from one China hospital are used for testing MedSRGAN on clinical data.

All networks take a 128×128 CT slice as input and output a 512×512 CT slice. Our networks are fully convolutional, so they can take CT slices of arbitrary size as input and generate $4 \times$ slices. The intensity of LR and HR CT images is clipped to $[-1024, 1024]$ HU (Hounsfield Unit), and then linearly scaled into range of $[0, 1]$. To train the generator, we set $\lambda_1 = 10^{-2}$, $\lambda_2 = 10^{-4}$ and $\lambda_{L1} = 0$ for loss function (4), and used Adam [28] optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We set the weight of content loss as two orders of magnitude larger than the weight of adversarial losses, such setting is also applied in GAN based methods like SRGAN and ESRGAN, and we tuned λ_1 and λ_2 during experiments but effects were limited on generated images. The initial learning rate was set to 10^{-4} and halved every $50k$ iterations. These experiments were conducted on two NVIDIA Titan Xp GPUs.

3.2.2 MOS test

Medical images constitute a core portion of the information a radiologist utilizes to render diagnostic and treatment decision. The science of medical image perception is dedicated to understanding and improving the clinical interpretation process. In this study, we focus our evaluation on content-based subjective views, which is our MOS test, from five experienced radiologists in simulated clinical environment. 100 CT slices are randomly selected from our test set and to form a MOS test dataset. Among these slice images, half used $[-160, 240]$ HU display window (tissue window) and the other half used $[-1200, 0]$ HU (lung window), which are two significant display windows in thoracic CT viewing. For each slice, the HR image (original image) and SR images obtained by methods including bicubic interpolation, RCAN (P), RWMAN (P), ESRGAN, MedSRGAN and MedSRGAN+ are scored, thus our MOS dataset contains $100 \times (1 + 6) = 700$ CT slice images. These images were randomly shuffled and assured not containing any information that could be implications of which methods to avoid subjective biases before being scored by five experienced radiologists. Results and distribution of their scores are shown in Table 1 and Fig. 6.

3.2.3 Analysis of results

The average PNSR/SSIM for all test images using different methods are shown in the Table 2, and these results were calculated after subjecting the whole $[-1024, 1024]$ HU display window to $[0, 1]$. The result shows that using GAN-based methods can achieve almost the same PSNR/SSIM performance comparing with PSNR-oriented methods in this task.

Fig. 7 shows the image examples as well as PSNR/SSIM values. We observe that PSNR-oriented RCAN generate over-smoothed images though it has higher PSNR/SSIM, while the patterns (the first row of Fig. 7) of images generated by ESRGAN are more visually unnatural, and the textures (the second row of Fig. 7) are less obvious than images generated by MedSRGAN. Moreover, these images of ESRGAN are still confronted with serious chess-board artifacts problem, this reveal that densely connected convolutional blocks are probably not suitable for CT SR. After all, the proposed MedSRGAN performs better in rebuilding these missing information and its generated images are more visually feasible.

Besides, the MOS test of different methods with two display windows are shown in Table 1 and Fig. 6. We observed that under both tissue window and lung window, the average MOS of our methods are better than other methods on visual judgments, no matter using noise channel

Table 1 Results of MOS test for CT. (P) after the title of methods indicates this model was trained on PSNR-oriented loss (MSE loss) and the postfix ‘+’ denotes noise channel was included. **Boldface** and underline indicate the best and the second-best performance, *italics* indicates scores on HR images

Method	Average MOS on tissue window	Average MOS on lung window
Bicubic	1.184	1.008
RCAN (P)	1.520	2.252
RWMAN (P)	1.716	2.960
ESRGAN	2.772	2.936
MedSRGAN	<u>3.284</u>	3.616
MedSRGAN+	3.396	<u>3.544</u>
<i>HR</i>	3.888	<i>4.256</i>

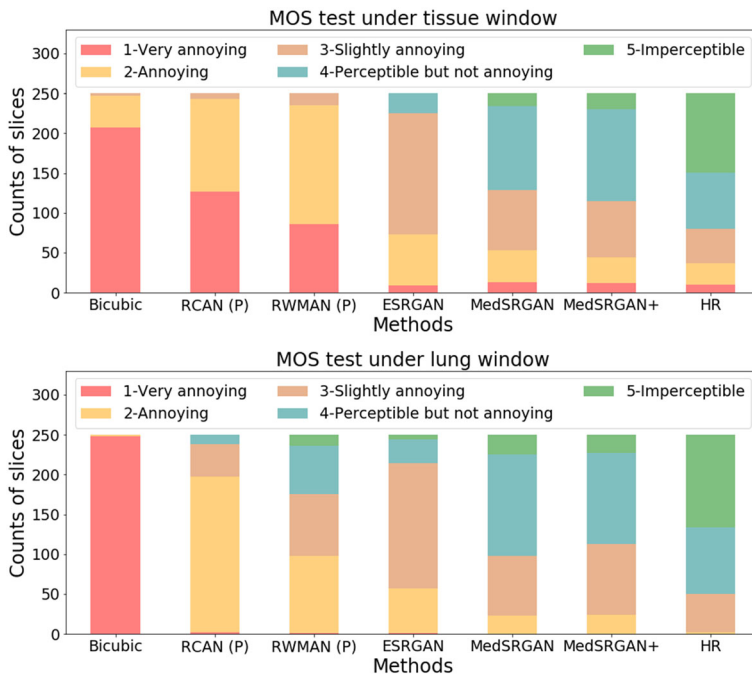


Fig. 6 MOS test for CT on tissue window $[-160, 240]$ and lung window $[-1200, 0]$. (P) after the title of methods indicates this model was trained on PSNR-oriented loss (MSE loss) and the postfix '+' denotes noise channel was included

or not. Specifically, from results shown in the first graph of Fig. 6 we found that models with noise channel input helps generator make more natural artifacts and patterns when displaying on tissue window, thus it got a higher score, but it encountered less visual quality when displaying on lung window (see the second graph of Fig. 6). So there may exist a trade-off between generating natural patterns and restoring textures with using noise channel or not.

Table 2 Average PSNR and SSIM of our experiments on 23 CT scans (5,889 slices) of LUNA 16 test set, additional 30 clinical CT scans (10,732 slices), and 30 brain MRI scans (2,118 slices) for testing. Models in 'LUNA 16 Test Set' and 'Clinical CT Data' were trained with LUNA 16 training set, as stated in Section-3.2.1. Models for 'Brain MRI' were trained with brain MRI training data, as stated in Section-3.3.1. (P) after the title of methods indicates this model was trained on PSNR-oriented loss (MSE loss) and the postfix '+' denotes noise channel was included. **Boldface** and underline in 'PSNR' and 'SSIM' column indicate the best and the second best performance on PSNR and SSIM

Method	CT				MRI	
	LUNA 16 Test Set		Extra Clinical CT Scans		Brain MRI Test Set	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	28.408	0.984	25.251	0.972	25.126	0.984
RCAN (P)	31.803	0.991	27.650	0.983	26.885	0.989
RWMAN (P)	31.353	0.989	27.271	0.982	27.077	0.990
ESRGAN	<u>31.914</u>	<u>0.990</u>	<u>27.573</u>	<u>0.983</u>	26.153	0.987
MedSRGAN	32.004	<u>0.990</u>	27.532	0.983	26.184	0.987
MedSRGAN+	31.700	<u>0.989</u>	27.530	0.983	26.464	0.988

3.2.4 Ablation study

Since our method takes PSNR-oriented RCAN as baseline model, the impacts of our meaningful modifications are worthy of comparative analysis. The progress is shown in Table 3 and a brief discussion is illustrated as follows.

Replace RCAB with RWMAB By comparing to natural images, some regions in medical images may be worthy of attention in generating high resolution images. We used RWMAB for whole image feature map attentions, to enable the network to focus on those meaningful regions. To better show its effect, we trained a PSNR-oriented RWMAN for comparison, which simply used MSE loss on pixel level for training RWMAN. We observe that RWMAN (P) got a marginal decrease in PSNR and SSIM but a higher increase in MOS test.

Use GAN-based framework As illustrated, although PSNR-oriented RWMAN performs not as well as RCAN on PSNR and SSIM, it performs more superior than RCAN in MOS test. Nevertheless, images generated with PSNR-oriented RWMAN are tend to be over-smoothed and lack of high-frequency details. So GAN-based framework with loss function (4) is introduced to allow the generator to gain enhancement from adversarial learning and produce more realistic results.

Add input Noise Channel With adversarial learning, SR images generated by MedSRGAN become closer to the real HR images, but the images are still monotonous in pattern generating, especially when displaying images under the tissue window. In order to deal with the problem, we introduce a noise channel as an additional input for making some randomness in network training. We observe MedSRGAN with noise input (MedSRGAN+) got a marginal increase in MOS test under tissue window displaying.

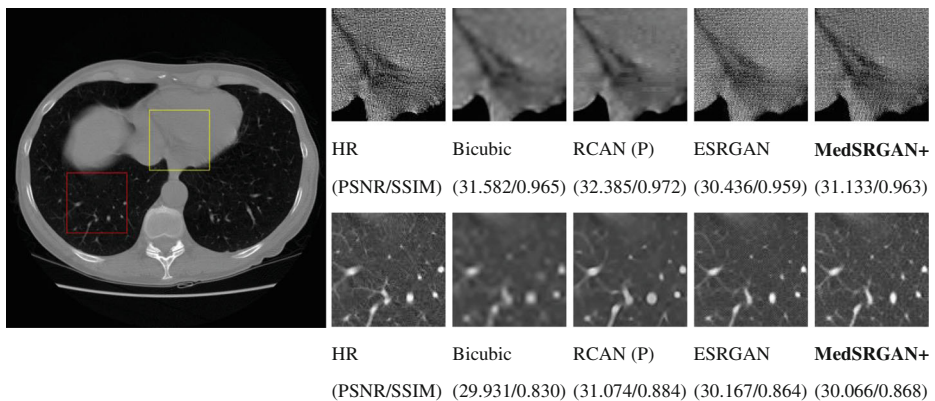


Fig. 7 Results of different methods on 96×96 image patches cropped from 512×512 slices. (P) after the title of methods indicates this model was trained on PSNR-oriented loss (MSE loss) and the postfix '+' denotes noise channel was included. Image patches in yellow box are shown in the first row and its display window is $[-160, 240]$ HU (tissue window), image patches in red box are shown in the second row and its display window is $[-1200, 0]$ HU (lung window). PSNR-oriented RCAN brings over-smoothed images though it has higher PSNR/SSIM. For GAN-based methods, images generated by ESRGAN still have the problem of **chessboard artifacts** (which is difficult to find but extremely noteworthy) while MedSRGAN+ gets rid of these annoying artifacts and looks closer to HR image both on patterns and textures. **Best viewed on screen with zoom in**

Table 3 The ablation study in our experiments. (P) after the title of methods indicates this model was trained on PSNR-oriented loss (MSE loss) and the postfix '+' denotes noise channel was included. **Boldface** and underline indicate the best and second best performance. This table shows the progress on visual quality of our modifications, especially in MOS test

	RCAN (P) (baseline)	RWMAN (P)	MedSRGAN	MedSRGAN+
Change RCAB to RWMAB?		✓	✓	✓
Use GAN-based Framework?			✓	✓
Add Input Noise Channel?				✓
Average PSNR	31.803	31.353	32.004	31.700
Average SSIM	0.991	0.989	<u>0.990</u>	0.989
MOS (Tissue)	1.520	1.716	<u>3.284</u>	3.396
MOS (Lung)	2.252	2.960	3.616	<u>3.544</u>

3.2.5 Nodule detection test

We believe that well rebuilt SR images should not affect real world diagnosis. To better demonstrate the effect on diagnosis, the lung nodule detection test is executed on HR and MedSRGAN+ generated SR images on a sophisticated CAD (computer aided detection) system. The detection system was well trained with reliable lung nodule annotations and has been used in practice. The test set, mentioned in Section-3.2.1, includes 23 scans with 35 annotated lung nodules from LUNA 16. The distribution of nodules diameter and detection results are shown in Fig. 8. In total, CAD reports 165 nodules on HR scans with 31 true positive (TP) nodules, and 156 nodules on SR scans with 32 TPs, which show that rebuilt SR images almost do not affect the nodule detection in this test.

3.2.6 Results on clinical data

MedSRGAN framework was also tested on 30 additional clinical CT scans (10,732 slices) from one China hospital, which were obtained by TOSHIBA Aquilion CT scanner and with size of 512×512 . These data may confront a problem of different data distribution from LUNA16 [40] Dataset because different CT scanners may produce different noise patterns or other artifacts. However, our methods are also able to generate images with high visual quality, as is shown in Fig. 9 with HR reference and results generated by different methods. The first row of Fig. 9 are displayed on $[-160, 240]$ HU (tissue window) and the second row are displayed on $[-1200, 0]$ HU (lung window). The average PNSR/SSIM of SR using different methods for all 10,732 slices are also collected in Table 2. Although there are only marginal differences between MedSRGAN and other methods (except for bicubic interpolation) on the average PSNR and SSIM, MedSRGAN outperforms any other methods in reconstructing patterns and textures information on these data and our results are closer to the real images visually as well, which visual results are shown in Fig. 9.

3.3 MRI experiments

3.3.1 Data and training details

To demonstrate the flexibility of our methods, 110 brain MRI scans from one China hospital were collected for MRI experiments. All MRI scans were obtained by PHILIPS Ingenia 3.0 T

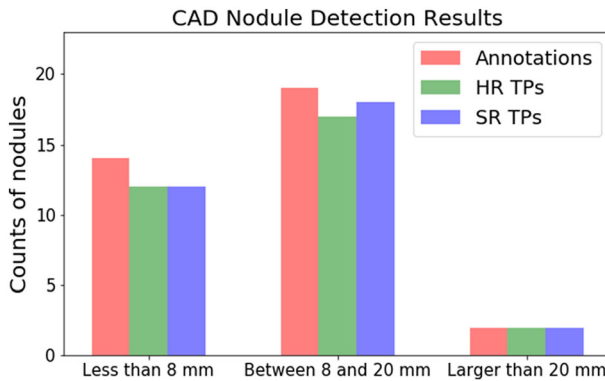


Fig. 8 The nodule detection results on test set. Nodules are divided into three groups according to their diameters to show the effect of SR images on small, medium and large nodules. The SR scans were generated by MedSRGAN+ framework

MR system and with size of 672×672 , 80 scans (7,661 slices) are used for training and the rest 30 scans (2,118 slices) for testing.

The imaging method of MRI is quite different from CT and the value of each voxel in MRI does not have specific physical meaning, so zero-mean normalization (subtract the mean and divided by the standard deviation) was applied to each MRI scan before training and testing. LR MRI slices are obtained by 4×4 average pooling on original HR slices. We set $\lambda_1 = 5 \times 10^{-2}$, $\lambda_2 = 5 \times 10^{-3}$ and $\lambda_{L1} = 10^{-2}$ for loss function (4) in training MRI super-resolution models, and used Adam [28] optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning rate was set to 10^{-4} and halved every 20k iterations.

3.3.2 Results

The average PSNR/SSIM for all brain MRI test images using different methods are shown in the Table 2. All metrics are calculated on cropped images to eliminate the influence of non-body area, which would make PSNR/SSIM very high but meaningless. The quantitative results illustrate that all methods have similar performances in PSNR/SSIM except for bicubic interpolation, though the RWMAN (P) has the highest values.

Fig. 10 shows some patches from original MRI images as well as their PSNR and SSIM on these cropped images. We observe that PSNR-oriented RCAN still result in over-smoothed SR images though it gets high PSNR/SSIM, but in some patches MedSRGAN and MedSRGAN+ can get higher PSNR and SSIM than RCAN. Images generated by ESRGAN still present less feasible artifacts, and MedSRGAN+ performs better in rebuilding these details, which results look closer to the real HR images.

3.4 Efficiency and parameter counts

We performed an experiment on the running speed of these methods, which results are listed in Table 4. Note that RCAN [49] uses 10 residual groups with each contains 20 RCABs in their experiments, so do we use this configuration in training RCAN (P) and RWMAN (P) in our experiments. The counts and storage of parameters are also collected in Table 4.

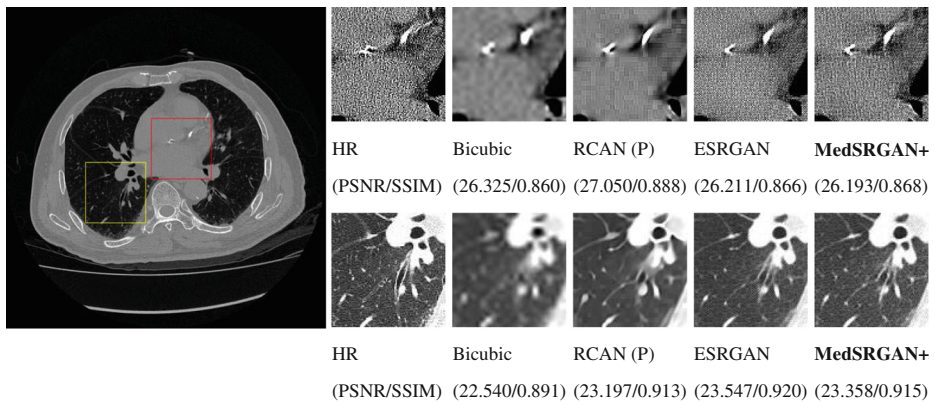


Fig. 9 Results of different methods on 96×96 image patches cropped from 512×512 slices on a slice of clinical data. (P) after the title of methods indicates this model was trained on PSNR-oriented loss (MSE loss) and the postfix '+' denotes noise channel was included. Image patches in red box are shown in the first row and its display window is $[-160, 240]$ HU (tissue window), image patches in yellow box are shown in the second row and its display window is $[-1200, 0]$ HU (lung window). PSNR-oriented RCAN brings over-smoothed images. For GAN-based methods, images generated by ESRGAN still have the problem of **chessboard artifacts** while MedSRGAN+ gets rid of these annoying artifacts and looks closer to HR image both on patterns and textures. These results show that our model performs visually better even though on data set with different distributions. **Best viewed on screen with zoom in**

It was found that, 1) it is almost real-time for MedSRGAN framework and ESRGAN to generate an image on GPU based machines; 2) RCAN has no advantages on running speed, which can be explained with its additional channel decreasing and increasing convolutions in RCAB structure, while our network does not contain such step; 3) ESRGAN slightly performed better than MedSRGAN on speed, but the counts of parameter is larger. In conclusion, our method not only got prominent visual results but also is advantageous when taking speed and parameter counts into consideration.

4 Discussion

In this study, we presented MedSRGAN as an end-to-end framework for medical image super-resolution tasks. MedSRGAN is comprised of a novel neural network for SR medical images generation, a pair-wise discriminator and a GAN-based novel loss function. The framework ensures sufficient contents and details as well as realistic for viewing on rebuilt images to a great extent.

However, we would like to point out that SR images of MedSRGAN are still not perfect because it is impossible for it to rebuild all information in HR based on LR images merely, though MedSRGAN has already got rid of blurring, over-smooth problems and strange artifacts. SR for image generation is actually an underdetermined problem, which means an output SR image incorporates much more information than its corresponding LR image. To interpret where these extra information come from, we reviewed the whole process of our experiments. In the training stage of MedSRGAN framework, each LR-HR pair was treated as a training sample and was fed to the training process independently, and a loss function was employed to make sure generated SR images get close to HR images as far as possible. As a

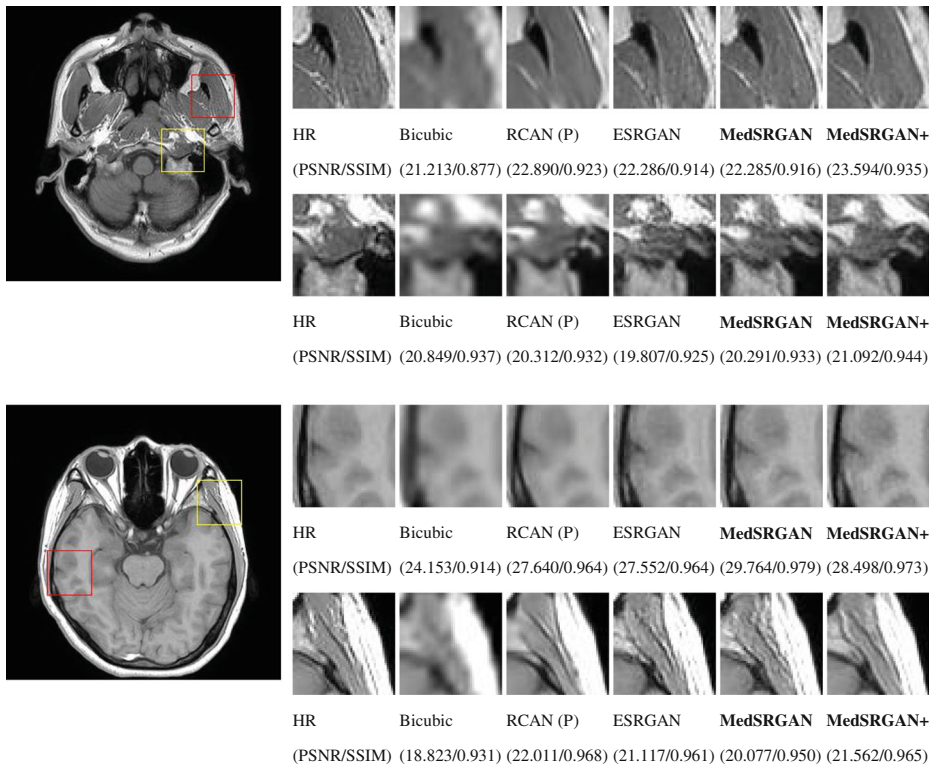


Fig. 10 Results of different methods on 48×48 image patches cropped from original MRI slice. (P) after the title of methods indicates this model was trained on PSNR-oriented loss (MSE loss) and the postfix ‘+’ denotes noise channel was included. The display window is clipped to $[0, 480]$. Image patches in red box are shown in the first row and image patches in yellow box are shown in the second row. PSNR-oriented RCAN brings over-smoothed images though it has high PSNR/SSIM. For GAN-based methods, images generated by our method look closer to HR image both on patterns and textures. These results show that our model performs better on generating feasible patterns. **Best viewed on screen with zoom in**

result, the general missing information were encoded into neural networks. With the assumption that LR images contain sufficient information for radiologists to make diagnosis, these encoded information hidden in neural networks can be illustrated as high frequency edges, specific artifacts and patterns of a certain kind of medical images. In a specific modality of medical images, such high frequency edges, specific artifacts and patterns are actually

Table 4 Speed of SR methods and their counts as well as storage of parameters. (P) after the title of RCAN indicates it was trained on PSNR-oriented loss (MSE loss). RCAN (P) uses 10 residual groups with each contains 20 RCABs, as proposed in [49]. The postfix ‘+’ denotes noise channel was included. **Boldface** and underline indicate the best and second best performance

Method	Speed (slices per second)	Parameter Counts (M for Million)	Parameter Storage (MB)
Bicubic	600.92	—	—
RCAN (P)	11.80	15.59 M	62.64
ESRGAN	27.02	11.88 M	47.64
MedSRGAN	<u>26.49</u>	10.62 M	42.63
MedSRGAN+	25.58	<u>10.62 M</u>	<u>42.63</u>

monotonous and stereotyped when comparing to natural images, which make it relatively easy to learn in machine learning frameworks. But it must be pointed out that, in case of extremely tiny focuses which are smaller than 4×4 pixels in high resolution images, no method is able to generate this imaging because all information of such focuses were unfortunately lost. This situation could happen if we take a CT scan on low-resolution scale.

To find out whether SR images affect diagnosis or not, we performed a simple nodule detection test on both HR CT scans and MedSRGAN generated SR CT scans, as is demonstrated on Section-3.2.5. The results showed that this nodule detection system also performed well in rebuilt SR CT scans, because it obtained almost identical detection results on true positive nodules of different sizes. Also, SR CT scans generated by our MedSRGAN are the closest to real HR scans in human viewing, as was demonstrated in MOS test. Hence, our methods are at least effective and feasible in human viewing and preserving nodules information. However, these evaluations may not cover all diagnosis situations, and we believe more experiments for evaluation should be conducted.

5 Conclusions

In this study, we presented a GAN-based SR framework for common medical images (MedSRGAN), which included Residual Whole Map Attention Network (RWMAN) as generator, discriminator for image pairs and a novel loss function for training generator. The results show that our method can be directly used in SR reconstruction of CT and MRI images, and their results preserve more feasible texture details and generate realistic patterns on HR images. The reliable and effective results of MedSRGAN imply the feasibility of using the SR method for retrieving more image details on clinical obtained low-resolution images, such as LDCT, low magnetic field MR, and MR spectroscopic imaging. In the future, we will conduct more in-depth studies of MedSRGAN, such as investigating a more sophisticated method to evaluate the SR reconstruction performance.

Acknowledgments This work was supported in part by the National Key R&D Program of China under Grant 2018YFC1704206, Grant 2016YFB0200602, in part by the NSFC under Grant 81971691, Grant 81801809, Grant 81830052, Grant 81827802, Grant U1811461, and Grant 11401601, in part by the Science and Technology Program of Guangzhou under Grant 20180420053, in part by the Science and Technology Innovative Project of Guangdong Province under Grant 2016B030307003, Grant 2015B010110003, and Grant 2015B020233008, in part by the Science and Technology Planning Project of Guangdong Province under Key Grant 2017B020210001, in part by the Guangzhou Science and Technology Creative Project under Key Grant 201604020003, in part by the Guangdong Province Key Laboratory of Computational Science Open Grant 2018009, in part by the Construction Project of Shanghai Key Laboratory of Molecular Imaging 18DZ2260400, and in part by China postdoctoral science foundation No.2019M653185.

References

1. K Armanious, C Jiang, M Fischer, T Küstner, K Nikolaou, S Gatidis, B Yang 2018. “MedGAN: Medical Image Translation using GANs”. In: arXiv:1806.06397 [cs.CV]
2. Berrington GA, Kim KP, Berg CD (2008) Low-dose lung CT screening before age 55: estimates of the mortality reduction required to outweigh the radiation-induced cancer risk. In: *J Med Screen* 15(3):153–158. <https://doi.org/10.1258/jms.2008.008052>

3. Chen Y, Xie Y, Zhou Z, Shi F, Christodoulou AG, Li D (2018) Brain MRI super resolution using 3D deep densely connected neural networks. In: In: 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018), Washington, pp 739–742
4. Y Chen, F Shi, A G Christodoulou, Y Xie, Z Zhou, D Li 2018. “Efficient and Accurate MRI Super-Resolution Using a Generative Adversarial Network and 3D Multi-level Densely Connected Network”. In: Frangi A., Schnabel J., Davatzikos C., Alberola-López C., Fichtinger G. (eds) Medical Image Computing and Computer Assisted Intervention - MICCAI. MICCAI 2018. Lecture notes in computer science, vol 11070. Springer, Cham
5. Z Cheng, Y Ding, X He, L Zhu, X Song, M Kankanhalli 2018. “A3NCF: an adaptive aspect attention model for rating prediction”. In: IJCAI’18 Proceedings of the 27th International Joint Conference on Artificial Intelligence, pp. 3748–3754, July
6. K Cho et al. 2014 “Learning phrase representations using RNN encoder-decoder for statistical machine translation”. In: arXiv:1406.1078 [cs.CL]
7. Christe A, Torrente JC, Lin M, Yen A, Hallett R, Roychoudhury K, Schmitzberger F, Vock P, Roos J (2011) CT screening and follow-up of lung nodules: Effects of tube current-time setting and nodule size and density on detectability and of tube current-time setting on apparent size. *AJR am J Roentgenol* 197(3):623–630. <https://doi.org/10.2214/AJR.10.5288>
8. C Dong, C C Loy, K He, X Tang 2014. “Learning a Deep Convolutional Network for Image Super-Resolution”. In: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture notes in computer science, vol 8692. Springer, Cham
9. Gatys LA, Ecker AS, Bethge M (2016) Image Style Transfer Using Convolutional Neural Networks. In: IEEE conference on computer vision and pattern recognition (CVPR), vol 2016, Las Vegas, pp 2414–2423
10. Gondara L (2016) Medical Image Denoising Using Convolutional Denoising Autoencoders. In: IEEE 16th international conference on data mining workshops (ICDMW), vol 2016, Barcelona, pp 241–246
11. I Goodfellow, J Pouget-Abadie, M Mirza, B Xu, D Warde-Farley, S Ozair, A Courville, Y Bengio 2014. “Generative Adversarial Nets”. In: Z. Ghahramani and M. Welling and C. Cortes and N. D. Lawrence and K. Q. Weinberger (eds) Advances in Neural Information Processing Systems 27, , Curran associates, Inc pp 2672–2680.
12. Greenspan H (2008) Super-resolution in medical imaging. In: *Comput J* 52(1):43–63
13. Guo Y, Cheng Z, Nie L, Wang Y, Ma J, Kankanhalli M (2019) Attentive Long Short-Term Preference Modeling for Personalized Product Search. *ACM Transactions on Information Systems (TOIS)* 37(2). <https://doi.org/10.1145/3295822>
14. He K, Zhang X, Ren S, Sun J (2016) Deep Residual Learning for Image Recognition. In: IEEE conference on computer vision and pattern recognition (CVPR), vol 2016, Las Vegas, pp 770–778
15. J Hu, L Shen, G Sun n.d.. “Squeeze-and-excitation networks”. In: IEEE Transactions on Pattern Analysis and Machine Intelligence
16. Huang G, Liu Z, Maaten LVD, Weinberger KQ (2017) Densely Connected Convolutional Networks. In: IEEE conference on computer vision and pattern recognition (CVPR), vol 2017, Honolulu, pp 2261–2269
17. Hunink MGM, Gazelle GS (2003) CT screening: A trade-off of risks, benefits, and costs. *the journal of clinical investigation* 111(11):1612–1619. <https://doi.org/10.1172/JCI200318842>
18. Isola P, Zhu J, Zhou T, Efros AA (2017) Image-to-Image Translation with Conditional Adversarial Networks. In: IEEE conference on computer vision and pattern recognition (CVPR), vol 2017, Honolulu, pp 5967–5976
19. G Jiang, Y Lu, J Wei, Y Xu 2019. “Synthesize mammogram from digital breast tomosynthesis with gradient guided cGANs”. In: Shen D. et al. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI. MICCAI 2019. Lecture notes in computer science, vol 11769. Springer, Cham.
20. Jifara W, Jiang F, Rho S, Cheng M, Liu S (2019) Medical image denoising using convolutional neural network: a residual learning approach. In: *J Supercomput* 75:704–718. <https://doi.org/10.1007/s11227-017-2080-0>
21. Y Jing, Y Yang, Z Feng, J Ye, Y Yu, M Song 2017. “Neural Style Transfer: A Review”. In: arXiv: 1705.04058 [cs.CV]
22. J Johnson, A Alahi, Fei-Fei Li 2016 “Perceptual Losses for Real-Time Style Transfer and Super-Resolution”. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV. ECCV 2016. Lecture notes in computer science, vol 9906. Springer, Cham.
23. A Jolicoeur-Martineau 2018. “The relativistic discriminator: a key element missing from standard GAN”. In: arXiv:1807.00734 [cs.LG]
24. Karadi RL (2006) Radiation risk of screening with low dose CT. *Thorax* 61(6):548
25. Kavanagh J, Liu G, Menezes R, O’Kane GM, McGregor M, Tsao M, Shepherd FA, Schmidt H (2018) Importance of Long-term Low-Dose CT Follow-up after Negative Findings at Previous Lung Cancer Screening. *Radiology* 289(1). <https://doi.org/10.1148/radiol.2018180053>

26. Kim J, Lee JK, Lee KM (2016) Deeply-Recursive Convolutional Network for Image Super-Resolution. In: IEEE conference on computer vision and pattern recognition (CVPR), vol 2016, Las Vegas, pp 1637–1645
27. Kim J, Lee JK, Lee KM (2016) Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In: IEEE conference on computer vision and pattern recognition (CVPR), vol 2016, Las Vegas, pp 1646–1654
28. D P Kingma, J Ba 2014. “Adam: A method for stochastic optimization”. In: arXiv:1412.6980 [cs.LG]
29. D Kouame, M Ploquin 2009. “Super-resolution in medical imaging: an illustrative approach through ultrasound”. In: IEEE International Symposium on Biomedical Imaging pp 249–252
30. Ledig C et al (2017) Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In: IEEE conference on computer vision and pattern recognition (CVPR), vol 2017, Honolulu, pp 105–114
31. D Lee, J Yoo, J C Ye 2017 “Compressed sensing and parallel MRI using deep residual learning”. In: Proc. 25st Annu. Meeting ISMRM
32. Libby DM, Wu N, Lee JJ, Farooqi A, Smith JP, Pasmantier MW, McCauley D, Yankelevitz DF, Henschke CI (2006) CT screening for lung cancer: The value of short-term CT follow-up. *chest* 129(4):1039–1042
33. Lim B, Son S, Kim H, Nah S, Lee KM (2017) Enhanced Deep Residual Networks for Single Image Super-Resolution. In: IEEE conference on computer vision and pattern recognition workshops (CVPRW), vol 2017, Honolulu, pp 1132–1140
34. Lingala SG, Hu Y, DiBella E, Jacob M (2011) Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR. In: IEEE Transactions on Medical Imaging 30(5):1042–1054
35. Mardani M et al (2019) Deep generative adversarial neural networks for compressive sensing MRI. *IEEE Transactions on Medical Imaging* 38(1):167–179
36. M Mirza, S Osindero 2014. “Conditional Generative Adversarial Nets”. In: arXiv:1411.1784 [cs.LG]
37. Mohammed RH, Sabry YY, Nasef AA (2011) Brain MRI screening showing evidences of early central nervous system involvement in patients with systemic sclerosis. *Rheumatol Int.* 31(5):667–671. <https://doi.org/10.1007/s00296-009-1325-5>
38. Rampinelli C, Origi D, Bellomi M (2013) Low-dose CT: technique, reading methods and image interpretation. *Cancer Imaging* 12:548–556. <https://doi.org/10.1102/1470-7330.2012.0049>
39. Ravishanker S, Bresler Y (2011) MR image reconstruction from highly undersampled k-space data by dictionary learning. In: IEEE Transactions on Medical Imaging 30(5):1028–1041
40. Setio AAA, Traverso A, de Bel T, Berens MSN, Bogaard C, Cerello P, Chen H, Dou Q, Fantacci ME, Geurts B, Gugten R, Heng PA, Jansen B, de Kaste MMJ, Kotov V, Lin JYH, Manders JTM, Sónora-Mengana A, García-Naranjo JC, Papavasileiou E, Prokop M, Saletta M, Schaefer-Prokop CM, Scholten ET, Scholten L, Snoeren MM, Torres EL, Vandemeulebroucke J, Walasek N, Zuidhof GCA, Ginneken B, Jacobs C (2017) Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. *Med Image Anal* 42:1–13. <https://doi.org/10.1016/j.media.2017.06.015>
41. Shi W et al (2016) Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. In: IEEE conference on computer vision and pattern recognition (CVPR), vol 2016, Las Vegas, pp 1874–1883
42. K Simonyan, A Zisserman. 2014 “Very Deep Convolutional Networks for Large-Scale Image Recognition”. In: arXiv:1409.1556 [cs.CV]
43. Ulus S, Suleyman E, Ozcan UA, Karaarslan E (2016) Whole-Body MRI Screening in Asymptomatic Subjects: Preliminary Experience and Long-Term Follow-Up Findings. *Polish Journal of Radiology* 81: 407–414. <https://doi.org/10.12659/PJR.897570>
44. Wang F et al (2017) Residual Attention Network for Image Classification. In: IEEE conference on computer vision and pattern recognition (CVPR), vol 2017, Honolulu, pp 6450–6458
45. X Wang et al. 2018 “ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks”. In: Leal-Taixé L., Roth S. (eds) Computer Vision – ECCV Workshops. ECCV 2018. Lecture notes in computer science, vol 11133. Springer, Cham
46. S Woo, J Park, J Y Lee, I S Kweon 2018. “CBAM: Convolutional Block Attention Module”. In: Ferrari V., Hebert M., Sminchisescu C., Weiss Y. (eds) Computer Vision – ECCV 2018. ECCV. Lecture notes in computer science, vol 11211. Springer, Cham.
47. Yasuda S, Ide M (2005) PET and cancer screening. *Annals of nuclear medicine* 19(3):167–177
48. C You et al. 2018 “CT Super-resolution GAN Constrained by the Identical, Residual, and Cycle Learning Ensemble (GAN-CIRCLE)”. In: arXiv:1808.04256 [eess.IV]
49. Y Zhang, K Li, K Li, L Wang, B Zhong, Y Fu 2018. “Image Super-Resolution Using Very Deep Residual Channel Attention Networks”. In: Ferrari V., Hebert M., Sminchisescu C., Weiss Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture notes in computer science, vol 11211. Springer, Cham

50. Zhu J, Park T, Isola P, Efros AA (2017) Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In: 2017 IEEE International Conference on Computer Vision (ICCV), Venice, pp 2242–2251

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Yuchong Gu is a M.E. student major in computer technology in School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. He received the B.E. degree of computer science and technology in Sun Yat-sen University in 2018. His research interests include technologies of artificial intelligence on medical area, developing practical medical image applications, machine learning and deep learning on computer vision.



Zitao Zeng received his B.E. in computer science from Center China Normal University, China in 2016. He is currently working toward his M.E. in computer science at the Sun Yat-Sen University, China. His research interests include Medical Image Processing and Computer Aided Diagnosis.



Haibin Chen is a post-doctor in School of Data and Computer Science at San Yat-sen University. His interests range in various topics in medical image processing, radiotherapy technology and machine learning. During his Ph.D. study period, he was invited as visiting researcher to the Department of Radiotherapy and Oncology at UTSW medical center. His Ph.D. researches focused on the applications of artificial intelligence in radiotherapy.



Jun Wei received his Ph.D. in bio-application and system engineering from Tokyo University of Agriculture and Technology in 2001. He is an associate research scientist in Radiology at University of Michigan, Ann Arbor, Michigan, USA. His research interests include computer vision, pattern recognition, medical image analysis, quantitative imaging biomarker, and artificial intelligence.



Yaqin Zhang is an attending doctor in Radiology Department, the Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, China. She received her M.D. degree in from Sun Yat-sen University, China in 2008. Her research interests include abdominal and uro-radiology.



Binghui Chen is a doctor in The Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, China. He received his M.M. in the Third Affiliated Hospital, Sun Yat-sen University, Guangzhou, China in 2016. His research interests include radiological diagnosis and molecular imaging.



Yingqin Li received her M.D. in Human Anatomy and histoembryology from Zhongshan histoembryology School of Medicine, Sun Yat-sen University, China in 2015. She is currently working as an attending radiologist in the fifth affiliated hospital of Sun Yat-sen University. Her research interests include medical molecular imaging and functional imaging.



Yujuan Qin received the M.D. degree of oncology in Southern Medical University, Guangzhou, China, in 2016. She is currently working in Department of Radiology, the Fifth Affiliated Hospital of Sun Yat-Sen University, Zhuhai, Guangdong Province, China. Her research interests include molecular imaging of stem cells and radiomics.



Qing Xie received her M.E. degree in imaging medicine and nuclear medicine from Dalian Medical University, China in 2018. She is currently working in Department of Radiology, the Fifth Affiliated Hospital of Sun Yat-Sen University, Zhuhai, Guangdong Province, China. Her research interests include brain functional imaging techniques.



Zhuoren Jiang is a Research Associate Professor of the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. He was a Postdoctoral Research Fellow in Institute of Computer Science & Technology, Peking University. His research interests include Medical Intelligent and Data Mining.



Yao Lu is a Professor of the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. He was a Postdoctoral Research Fellow and Research Investigator in Medical School, University of Michigan. His research interests include Inverse Problem, Medical Image Processing and Computer Aided Diagnosis.

Affiliations

Yuchong Gu¹ · **Zitao Zeng**¹ · **Haibin Chen**¹ · **Jun Wei**² · **Yaqin Zhang**³ · **Binghui Chen**³ · **Yingqin Li**³ · **Yujuan Qin**³ · **Qing Xie**³ · **Zhuoren Jiang**¹ · **Yao Lu**^{1,4}

¹ School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China

² Department of Radiology, University of Michigan, Ann Arbor, MI, USA

³ Department of Radiology, The Fifth Affiliated Hospital of Sun Yat-sen University, Zhuhai, China

⁴ Guangdong Province Key Laboratory of Computational Science, Guangzhou, China