



# Multi image super resolution of MRI images using generative adversarial network

U. Nimitha<sup>1</sup> · P. M. Ameer<sup>1</sup>

Received: 11 May 2022 / Accepted: 8 January 2024 / Published online: 17 February 2024  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

## Abstract

In recent decades, computer-aided medical image analysis has become a popular techniques for disease detection and diagnosis. Deep learning-based image processing techniques, have gained popularity in areas such as remote sensing, computer vision, and healthcare, compared to conventional techniques. However, hardware limitations, acquisition time, low radiation dose, and patient motion are factors that can limit the quality of medical images. High-resolution medical images are more accurate in localizing disease regions than low-resolution images. Hardware limitations, patient motion, radiation dose etc. can result in low-resolution (LR) medical images. To enhance the quality of LR medical images, we propose a multi-image super-resolution architecture using a generative adversarial network (GAN) with a generator architecture that employs multi-stage feature extraction, incorporating both residual blocks and an attention network and a discriminator having fewer convolutional layers to reduce computational complexity. The method enhances the resolution of low-resolution (LR) prostate cancer MRI images by combining multiple MRI slices with slight spatial shifts, utilizing shared weights for feature extraction for each MRI image. Unlike super-resolution techniques in literature, the network uses perceptual loss, which is computed by fine-tuning the VGG19 network with sparse categorical cross entropy loss. The features to compute perceptual loss are extracted from the final dense layer, instead of the convolutional block in VGG19 the literature. Our experiments were conducted on MRI images having a resolution of  $80 \times 80$  for lower resolution and  $320 \times 320$  for high resolution achieving an upscaling of x4. The experimental analysis shows that the proposed model outperforms the existing deep learning architectures for super-resolution with an average peak-signal-to-noise ratio (PSNR) of  $30.58 \pm 0.76$  dB and average structural similarity index measure (SSIM) of  $0.8105 \pm 0.0656$  for prostate MRI images. The application of a CNN-based SVM classifier confirmed that enhancing the resolution of normal LR brain MRI images using super-resolution techniques did not result in any false positive cases. This same architecture has the potential to be extended to other medical imaging modalities as well.

**Keywords** Image super resolution magnetic resonance images (MRI) · Prostate cancer · Deep learning · Generative adversarial network (GAN) · Super resolution GAN

## 1 Introduction

Medical imaging is a technique that creates images of the anatomical structures of the human body for clinical procedures like the study of anatomy and disease diagnosis. Image processing techniques can improve the quality of medical images, thereby enhancing the disease diagnosis (Ganguly et al. 2010). High-resolution images are beneficial

in applications like medical imaging, surveillance, forensics, and institutions like the military. Medical images are important for disease diagnosis, hence the better resolution of images improves disease diagnosis, automatic disease detection, and segmentation. Even though the imaging techniques and reconstruction algorithm performances have improved, images of the desired resolution are not that easy to obtain because of the imaging environments, image acquisition time, limitations of the medical imaging systems, and matters affecting image quality like noise and blurring (Yi et al. 2019).

Magnetic resonance imaging (MRI) is one of the medical imaging techniques for disease diagnosis and image-guided

✉ U. Nimitha  
nimithau@gmail.com

<sup>1</sup> Department of Electronics Communication Engineering,  
National Institute of Technology, Calicut, India

therapies. High-resolution MRI is important as it unfolds anatomical, physiological, cellular, and molecular details. Hardware and physiological constraints cause motion artifacts and low spatial resolution of MRI images; also, high-resolution MRI image acquisition take a longer scanning time at a higher operational cost (Lyu et al. 2018). MRI image quality depends on 3 factors—its resolution, signal-to-noise ratio (SNR), and acquisition time, and these features are interdependent. High-resolution MRI images take longer imaging time, and this longer scanning time makes the patients uncomfortable; causing motion artifacts in the MRI images (Plenge et al. 2012).

A super-resolution image reconstruction problem can be formulated by an observation model that associates the LR and HR images. For such an observation model aliasing, translation and rotational motion, blurring, and effect of noise are to be taken into consideration (Irani and Peleg 1991). HR medical images are essential for exact disease diagnosis, so medical imaging techniques are advanced imaging environments. Obstructions in the medical imaging equipment may lead to low resolution (LR) images (Trinh et al. 2014). Super-resolution (SR) is a classical computer vision problem, that involves deriving an HR image from a single LR image known as single image SR (SISR), or from a set of LR images known as multi-image SR (MISR) (Park et al. 2003). Super-resolution techniques initially focused on frequency domain methods but have disadvantages such as sensitivity to errors and difficulty in modeling the complex motion conditions. Spatial domain techniques overcome the drawbacks of frequency domain techniques. Among the spatial domain techniques, regularized frameworks are more popular (Yue et al. 2016). It is common practice to capture a set of 2-D MRI slices in situations where complete 3-D image acquisition is neither effective nor practicable (Greenspan et al. 2002). T2-weighted imaging, diffusion-weighted imaging, and MR angiography are examples of where this can arise. All of these imaging modalities are critical for early medical diagnosis and visualization, and they typically need considerable resolution of the scanned subject. A super-resolution technique is proposed, in which two spatially shifted MRI slices are combined to produce a high-resolution MRI slice. Challenges associated with SR are resolution enhancement under complex motion conditions, the absence of multiple LR images, and the speed and efficiency of reconstruction. There exist numerous approaches both in the frequency domain (Bose et al. 1993; Tom et al. 1994) and in the spatial domain for super-resolution of images (Elad and Feuer 1997; Stark and Oskouie 1989). Conventional methods lack clarity in mapping from LR space to HR space and are incapable of modeling complex high-dimensional mapping. However, deep learning (DL)—a subset of Artificial Intelligence (AI)—based techniques for super-resolution yield promising performances

both quantitatively and qualitatively. DL-based techniques have massive computations compared to the conventional ones (Yang et al. 2019).

Deep learning-based super-resolution algorithms have attained state-of-the-art performance and the Super-Resolution Convolutional Neural Network (SRCNN) architecture is considered to be the benchmark architecture for the SR problems. SRCNN is a DL-based technique that learns the mapping between the LR and HR images, and it has better performance over the existing methods (Dong et al. 2016). It is a lightweight architecture with only three layers—Patch extraction & representation, Nonlinear Mapping, and Reconstruction. Besides a simple and robust architecture, SRCNN finds applications in image de-blurring and image denoising also. Very deep super-resolution (VDSR) is a 20 layers architecture that is deep compared to SRCNN (Kim et al. 2016), and this deep architecture increases the accuracy of reconstruction. A residual image, which is the difference between LR and HR images, is modeled using residual learning CNN of VDSR. Since the high learning rate leads to exploding gradient problems, the VDSR uses residual learning and gradient clipping to solve the exploding gradient problem. Faster MRI acquisitions can be achieved with the deep neural network E2E-VarNet (Zhang et al. 2023) combining the techniques of Compressed sensing and multi coil MRI imaging. Iterative optimization is done in the k-space and reconstruction of MRI images is achieved with the CNN. Another method to fasten the MRI scanning process is by acquiring under-sampled images resulting in low resolution noisy images. ReconResnet (Chatterjee et al. 2022) network have regularized residual learning for reconstructing under-sampled radial MRI data. Artifact removal from the under-sampled images is accomplished by the modified regularized version of ResNet network and the data consistency steps fuses the network output to improve the reconstructed image quality. The Residual Channel Attention Network (RCAN)(Zhang et al. 2018) is capable of super resolving images using channel attention networks. Though with a huge number of layers, the RCAN model suffer from under-fitting problem not over-fitting. With proper training strategies, the under fitting problem is alleviated in RCAN-it (Lin et al. 2022). The network uses a smooth, non-monotonic function Sigmoid Linear Unit (SiLU) as activation function with consistent improvement. Another CNN based model ShuffleUnet (Chatterjee et al. 2021) uses modified tight-frame UNet architecture with pixel shuffle and unshuffle units for patch based single image super resolution. The model achieves super resolution of diffusion-weighted MRIs in all dimensions with factor of 8.

The super-resolution generative adversarial network (SRGAN) (Ledig et al. 2017) is the very first work on GAN for super resolution. The SRGAN network make use of Residual Network (ResNet), with skip connection in the

generator, to generate high resolution images, and a classifier in the discriminator, to classify the original HR image and the super-resolved image. Both the generator and the discriminator are deep neural networks that compete with each other. Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) (Wang et al. 2018a) is having complex architecture with dense residual layers. It uses relativistic discriminator that can predict the probability that an image is relatively realistic than another image. Cycle GAN based model (Kim et al. 2020) uses multi-generator GAN architecture to super-resolve HR image from its degraded low resolution image and uses domain transfer learning for LR image. The cycle GAN-based architecture uses three generators  $G_1, G_2$  &  $G_3$  and three discriminators  $D_N, D_Y$  &  $D_C$ .  $D_N$  is the discriminator for noise,  $D_C$  is the discriminator for color and  $D_Y$  for texture. Domain transfer learning is used for training all the three generators  $G_1, G_2, G_3$  and the discriminators  $D_N, D_Y, D_C$ . The generators  $G_1$  and  $G_2$  are for noise removal and image enhancement, and  $G_3$  is for super resolution. The architecture uses ESRGAN for super resolving the clean, noise free LR images. Prostate cancer is the second most common cancer worldwide. The deep learning architecture utilizing MSG-GAN and CapsNet (Molahasani et al. 2022) improves the MRI quality in high up-sampling scales for super-resolution of prostate MRI images. The network extracts features using CheXNet and lung scan is used for training the dataset. Super resolution optimized using perceptual-tuned Generative Adversarial Network (GAN) (Zhang et al. 2022) is a 3D SR interpolation technique to generate anti-aliased, deblurred thinner thinner HR MRI slices using 3D perceptual loss. This model is generalized to different imaging modalities, MRI and CT and also to different organs. GAN based EnlightenGAN (Jiang et al. 2021b) network is the first architecture for unpaired low-light image enhancement model. The network improves the image quality of the low light images by learning from the normal light images and achieves superior performance by using the global descriptors and self-regularization. TWIST-GAN (Dharejo et al. 2021) is frequency domain spatio-temporal network that combines wavelet transform and transfer GAN architecture to generate super-resolved remote sensing images. Wavelet packet transformation and Transfer GAN helps in reconstructing more realistic high resolution images with better texture details. Local fusion feature block and feature extraction at different scales are incorporated in Fused Attentive Generative Adversarial Networks(FA-GAN) (Jiang et al. 2021a) thus enhancing the performance of the super resolution of MRI images. The network has two attention modules—self attention and channel attention modules integrated in SRGAN to extract the best features. Demanding thousands of training images, a formidable task, particularly in biological samples, we have showcased the effectiveness of fully convolutional networks

(FCNs) architectures in conjunction with the SRDenseED (Mannam and Howard 2023) method. This strategy integrates Dense Encoder-Decoder blocks, enabling the effective training of SR FCNs despite having a constrained training dataset. Current super-resolution architectures commonly utilize a pre-trained VGG19 loss as their perceptual loss. Since the pre-trained VGG-19 network is originally trained on the general-purpose ImageNet dataset, there is an opportunity to enhance the perceptual quality significantly. This can be achieved by optimizing the perceptual loss through fine-tuning the VGG-19 network specifically using MRI images. Consequently, the computed perceptual loss will be tailored to the domain of MRI images, leading to improved results. Furthermore, the incorporation of spatially shifted MRI images can bolster the extraction of robust features, thereby enhancing the quality of super-resolution reconstructions. Additionally, the utilization of a multi-stage residual feature extraction network in conjunction with a spatial attention network further augments the extraction of robust features.

In this work, GAN based multi image MRI super-resolution that utilizes the correlated features from multiple MRI images is proposed. The contributions are:

- Novel GAN based architecture for super resolution of spatially shifted LR MRI slices.
- Multi stage feature extraction network employing residual blocks and attention network.
- Use of shared weights between the feature extraction network for the multiple MRI images.
- Fine-tuned Perceptual loss using VGG 19 network with sparse categorical cross entropy as the loss function.
- A modified discriminator with lesser number of convolutional layers compared to SRGAN (Ledig et al. 2017) discriminator to reduce computational complexity.

The proposed multi-image super-resolution architecture produces high-resolution MRI images of prostate cancer from its low-resolution counterparts. For the experimentation, we have synthetically down-sampled an HR image to generate distinct LR images with small spatial variations. Spatial variation between the LR images is introduced using sub-pixel shifts. The proposed architecture is compared against the existing super-resolution architectures using peak signal-to-noise ratio (PSNR), Mean Absolute Error (MAE) and structural similarity index measure (SSIM). The architecture has been trained and tested using a publicly available prostate cancer MRI images dataset.

The rest of the paper is organized as follows: Sect. 2 details the GAN and its Related Architectures. Section 3 presents the proposed architecture for the multi-image super-resolution describing in detail the generator network, the discriminator network, and the losses. Section 4 details the

Experimental analysis and Results of the proposed architecture and comparison between different super-resolution architectures. The conclusion of the paper is presented in Sect. 5.

## 2 Related architectures

GAN is generative and adversarial architecture that gained popularity due to its data generation capability from noise. GAN architecture can be modified for different image processing applications and many variations of GAN are present in the literature. For medical applications, GANs are used in two different modes—the generative aspect that produces new medical images from the training data and the discriminative aspect that classifies normal and abnormal images (Yi et al. 2019).

Excessive CT usage raises concerns due to high radiation doses, while reducing doses can lead to noise and artifacts, impacting disease diagnosis. Addressing this, a modified GAN, employing Wasserstein distance and perceptual similarity, mitigates noise in CT images (Yang et al. 2018). This framework adeptly balances noise reduction without exacerbating excessive smoothening, enhancing lesion detection precision. Variations of GAN network have been used to nullify noise in MRI (Ran et al. 2019) and PET images (Wang et al. 2018b). Image-to-image translation can be done using GAN networks that learn the mapping between retinal vessel tree images and the corresponding eye fundus images (Costa et al. 2018). But dependence on pre-existing vessel tree images is the limitation of the model. Patch-based 3D GAN architecture with Auto-Context Model for refinement can synthesize accurate and Morphological features of nuclei and the openness of the chromatin are the cell-level visual attributes which are useful for cell classification and segmentation. WGAN-GP and InfoGAN have been combined to research upon cell-level features of bone marrow and to perform unsupervised categorization as well (Hu et al. 2019).

EEG signals are helpful in Brain-Computer Interaction and Human Machine interaction for identifying human activities. The problem of unavailability of the large annotated dataset is solved by the semi-supervised deep learning GAN (SSL GAN) (Lecouat et al. 2018) after combining a small labeled dataset with a larger unlabelled dataset. SSL GAN makes use of patch level annotations for the small labeled dataset and the patch level predictions of an image are aggregated to arrive at the classification of the image. SegAN (Xue et al. 2018) is an adversarial critic network for segmentation that applies a multi-scale L1 loss function for both segmentor and critic. The segmentor and critic have trained alternately as in the classic GAN. Brain tumor segmentation is done using SegAN and its performance is superior when compared to that of the networks with single scale loss

function or softmax loss function. Adaptive GAN (AGAN) (Wu and Tian 2020) framework employs an adaptive learning algorithm to perform cardiac segmentation from chest X-rays. Here, the network activates multiple generators to extract features and determine the discriminator scores of the extracted features.

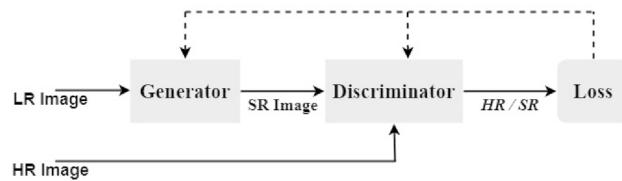
Conditional Generative Adversarial Nets (cGAN) are modified GAN architecture that can generate images based on conditions (Mirza and Osindero 2014). The generator and the discriminator are conditioned with some auxiliary information such as class labels. An additional input layer in the generator and discriminator is fed with the auxiliary information. Cycle consistency GAN (CycleGAN) (Zhu et al. 2017) is an extension of GAN for the image to image translation without paired images. This architecture uses two generator models and two discriminators. One generator is for translation and the other is for reconstruction; one discriminator is for translation and the other is for reverse translation. Super-resolution generative adversarial network (SRGAN) (Ledig et al. 2017) is another class of GAN for image super-resolution.

### 2.1 GAN for image super resolution

SRGAN is one of the variations of GAN that generate super-resolved images from its LR counterpart. The architecture of SRGAN is in Fig. 1. It is a single image super-resolution deep neural network architecture which upscales the LR image to the HR image with improved quality. The generator, G, generates super-resolved images, and the discriminator, D classifies them as SR or HR images.

## 3 Proposed architecture

The proposed MISR model stands apart by employing a single generator, in contrast to the multi-generator approach of comparable models. This distinctive design efficiently extracts features from a set of LR images, streamlining the process to singularly generate the HR image for enhanced efficiency. Multiple LR images are stacked together and these stacked images are the input to the generator. The HR image is down-sampled to produce an LR image by an up-scaling factor, the parameter that decides by how many

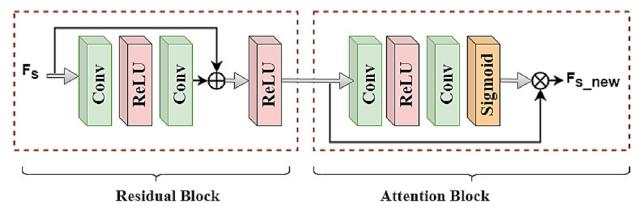


**Fig. 1** SRGAN block diagram

times the resolution of LR images has to be enhanced. To obtain distinct LR images by down-sampling the HR image, we use different sub-pixel shifts. These LR images have more features compared to the single LR image, so the proposed model is expected to outperform the existing SR techniques. The sub-pixel shift can only represent the translational motion between the LR images. The proposed architecture for the multi-image SR using GAN is shown in Fig. 2. The proposed architecture assumes that there is no relative motion, both translational and rotational, between the multiple low-resolution images. More details on the generator, discriminator, and the associated losses of the proposed architecture are in the following section.

### 3.1 Generator

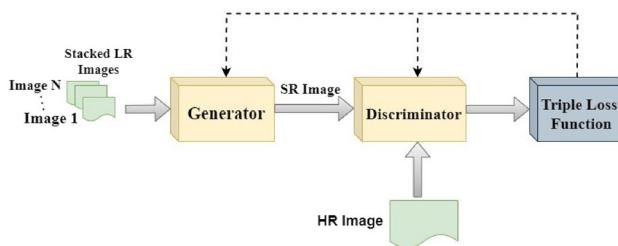
The generator is a deep neural network that generates the super-resolved image from the stacked LR images. The generator shown in Fig. 3 employs a dual-level feature extraction process: the first level extracts features from the individual MRI images, and the second level extracts features from the combined MRI features of these individual images. The initial stage of feature extraction uses a residual connection and spatial attention network as shown in Fig. 4. Features from both MRI images are extracted using a weight-sharing technique. This means that the same set of weights or parameters is applied to both MRI images during the feature extraction process. By sharing weights, the model ensures that the feature extraction process is consistent and coherent for both input MRI images. This technique helps in capturing and aligning essential information from both images. The residual block



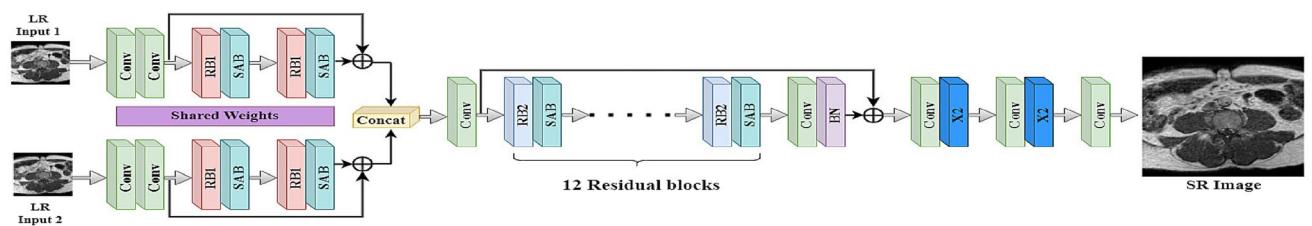
**Fig. 4** Residual block (RB1) and spatial attention block

(RB1) has convolutional layers with 32 feature maps, kernel size 3 and stride 1. The residual network is followed by a spatial attention block (SAB), that enhances the capability of a neural network to identify and emphasize important features. The network learns to assign higher attention weights to regions that contain relevant features while downplaying less informative areas. The attention mechanism is responsible for calculating these attention weights, which is implemented using convolutional layers and sigmoid function. This attention map is applied element-wise to the feature maps to scale the features based on the attention weights. SAB has single feature map in each layer to compute the spatial attention, with kernel size 3 and stride 1.

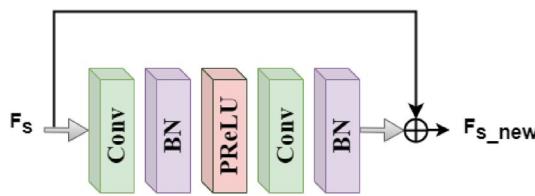
A convolutional layer with 64 feature maps extracts features from the feature maps of first stage of feature extraction and pass it to the second stage. The second feature extraction network within the generator consists of 12 feature extraction blocks, incorporating residual blocks (RB2) and SAB. Each residual block comprises two convolutional layers (Conv) with 64 feature maps, two batch normalization layers (BN), a single ReLU unit, and a summation operation same as the SRGAN, is depicted in Fig. 5. The features are subsequently enhanced through the application of two upscaling blocks. Each block comprises a convolutional layer with 64 features, employs leaky ReLU as its activation function, and incorporates an up-sampling layer. The up-scaling factor for the reconstruction decides the number of up-sampling layers in the generator. Finally, the resulting super-resolved image serves as input to the discriminator.



**Fig. 2** Multi image SRGAN block diagram



**Fig. 3** Multi image SRGAN generator



**Fig. 5** Residual block (RB2)

### 3.2 Discriminator

The discriminator is a binary classifier; that discriminates between the real and fake images by extracting features from the HR image and the SR image, as can be seen in Fig. 6. It uses five convolutional layers with a number of filters ranging from 64 to 256. Feature maps are extracted using the convolutional layers, followed by batch normalization layers for equalizing the inputs, then Leaky ReLU / ReLU as activation functions. Further, the features are passed through dense layers and the softmax function as the final layer, which functions to compute the probability of the extracted features being in the two classes—HR and SR classes. The first layer in the discriminator is a convolutional layer with 64 filters with kernels of size 3 having a stride of 1. This layer is followed by a block with a convolutional layer, a batch-normalization layer, and a leaky ReLU. Four similar blocks with distinct convolutional layers are used sequentially.

Comparing with the SRGAN (Ledig et al. 2017) discriminator, the proposed discriminator uses only 5 convolutional layers, whereas SRGAN uses 8 layers in the discriminator. With the proposed discriminator, number of parameters for the network is reduced to 2,847,617 than the SRGAN discriminator parameters of 5,417,665. Also the proposed discriminator uses lesser number (512) of dense layers.

### 3.3 Loss functions

Loss functions are important parameters that improve the image reconstruction quality. The losses associated with the proposed architecture are content loss, adversarial loss, perceptual loss, and binary cross-entropy loss to optimize

network parameters. The generator loss and discriminator loss of the proposed architecture is given by:

$$L_{Generator} = L_{Content} + \lambda_1 L_{Perceptual} + \lambda_2 L_{Adversarial} \quad (1)$$

where  $\lambda_1$  and  $\lambda_2$  are the coefficients to stabilize the loss function.

$$L_{Discriminator} = \log D_{\theta_D}(I_{HR}) + \log(1 - D_{\theta_D}(I_{SR})) \quad (2)$$

The loss function for the generator is beset with content loss, perceptual loss, and adversarial loss. The discriminator is associated with binary cross-entropy Loss. The generator tries to minimize its triple losses.

#### 3.3.1 Content loss

Content loss is the pixel-wise mean square error (MSE) between the HR and SR images. The content loss ensures the low frequency details between the HR and SR images. Reducing the content loss ensures the accuracy of the SR image, which in turn improves the peak signal-to-noise ratio (PSNR).

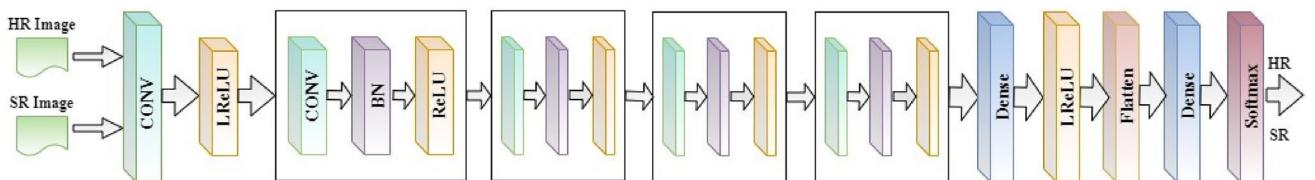
$$L_{Content} = MSE(I_{HR}, I_{SR}) \quad (3)$$

$$L_{Content} = \frac{1}{N} \sum_{i=1}^N \|I_i^{HR} - I_i^{SR}\|^2 \quad (4)$$

$I_{HR}$  is the HR image,  $I_{SR}$  is the SR image, N is the number of images in the training dataset

#### 3.3.2 Perceptual loss

To counter over-smoothing, the approach integrates VGG loss utilizing the VGG 19 neural network (Simonyan and Zisserman 2015). Sole reliance on MSE loss is insufficient for preserving high-frequency details in the image. VGG-19 is a deep convolutional neural network that contains 19 layers, and its pre-trained model has been trained using the ImageNet database, which consists of more than 1 million RGB images. However, it should be noted that MRI images are single-channel gray scale images. Therefore, since the pre-trained RGB model is not suitable for gray scale images, the VGG19 model need be trained from scratch with gray scale images



**Fig. 6** Multi image SRGAN discriminator

or fine-tuned with stacked gray scale images. The model in Fig. 7 is trained with brain MRI images from Kaggle (HEEMALI 2018); having brain MRI images from figshare, SARTAJ and Br35H datasets. The dataset used in this study consists of 7023 brain MRI images categorized into four classes: glioma, meningioma, no tumor, and pituitary. Two different loss functions were used to evaluate the model's performance: categorical cross entropy (CCE) and sparse categorical cross entropy (SCCE). The testing accuracy of the models is given in Table 1.

The classification accuracy of all the four models are similar. So all the four pre-trained models were utilized to compute the perceptual loss, with the feature map from the dense layer containing 1000 neurons selected for this purpose. This layer is capable of extracting fine details from both the HR and SR images. The perceptual loss is computed by comparing the features extracted from the final dense layer of the HR and SR images, using MSE as the measure. By minimizing the perceptual loss, the generator is able to produce SR images that accurately match the high-level feature maps of the HR images. The features from the dense layer (Dense 1000) is used to evaluate the perceptual loss.

$$L_{Perceptual} = \frac{1}{W_i} \sum_{x=1}^{W_i} \left\| \phi_i(I^{HR}) - \phi_i(I^{SR}) \right\|^2 \quad (5)$$

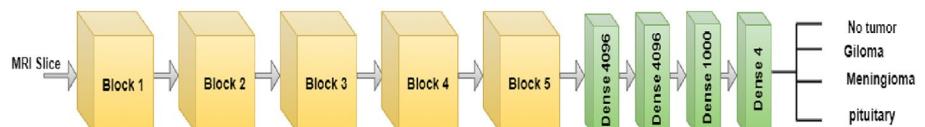
$\phi_i$  is the feature map from the VGG19 network.

### 3.3.3 Adversarial loss

Like any conventional binary classifier, the discriminator has to classify HR images into HR class and SR images into SR class. It tries to maximize the probability of correct classification of HR and SR images into the respective classes, which in turn minimizes its binary cross-entropy loss. The generator works to minimize the inverse probability of the SR image classification into SR class and this loss is known as the adversarial loss. This happens to be a min-max strategy, where the discriminator maximizes the probability of correct classification whereas the generator minimizes the probability of super-resolved images being classified into SR class.

$$L_{adversarial} = \sum_{i=1}^N -\log D_{\theta_D}(I_{SR}) \quad (6)$$

**Fig. 7** VGG 19 classifier



**Table 1** Testing accuracy of VGG 19 classification

Model	Loss	Testing accuracy (%)
VGG 19—fine-tuned	CCE	97.25
	SCCE	98.63
VGG 19—gray scale	CCE	96.72
	SCCE	97.63

By using triple losses for generator and multiple LR images, the architecture yields better performance compared to the existing techniques.

## 4 Experiment and results

### 4.1 Dataset

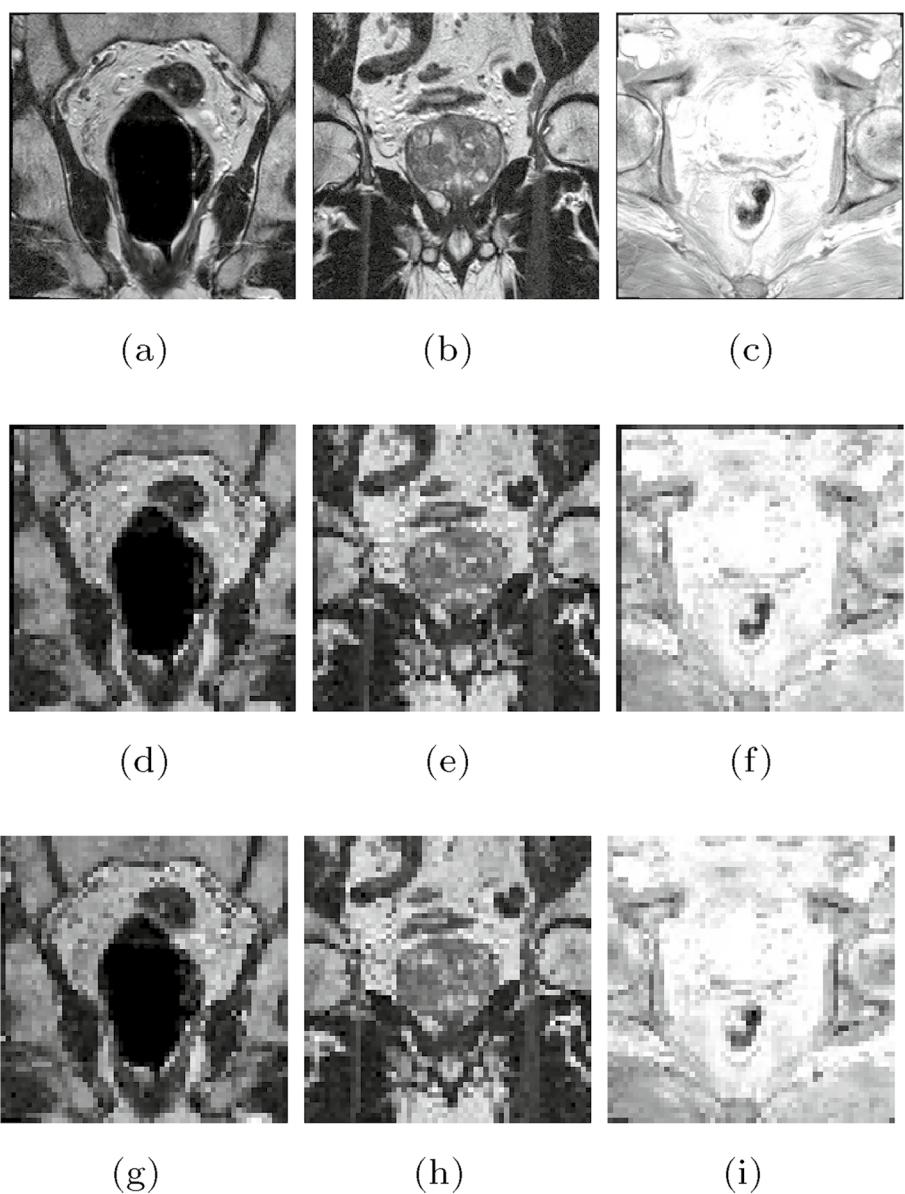
For the analysis, openly available Prostate cancer T1 weighted MRI image dataset is used (Bloch et al. 2015; Clark et al. 2013). The network is trained using the T1 axial MRI images of the 89 patients having spatial resolution  $320 \times 320$ . The splitting of the training and testing images are done at patient level in the ratio of 80:20, MRI slices of 71 patients for training and 18 for testing. The Prostate MRI images are available in Digital Imaging and Communications in Medicine (DICOM) format. The sample set of HR and the corresponding LR images down-scaled by a factor of 4, with different sub-pixel shifts are shown in Fig. 8. HR images hold more details and have good quality compared to the LR images. Practically the proposed model can be used in such a way that, every 2D MRI slice can be split into patches of size  $80 \times 80$  and can be super-resolved to produce HR patches of size  $320 \times 320$ .

### 4.2 Preprocessing

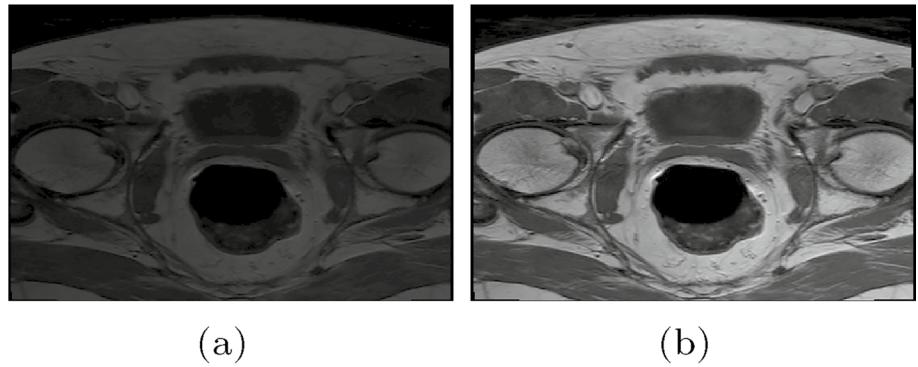
The DICOM MRI images are of 12 bits and the image pixels ranges from 0 to 4095. The visual quality of the MRI images are so poor such that, most of the pixel values lie near to origin, resulting in black. The visual quality is enhanced using contrast stretching and is given in Fig. 9. Contrast stretching is achieved with min–max normalization and then images are re-scaled between  $[-1, 1]$ .

The HR image is then down-sampled by the up-scaling factor to yield an LR image. About multi-image super-resolution (MISR), it is deduced that LR images have spatial

**Fig. 8** HR and LR images with different sub-pixel shifts from training dataset Row 1—HR images, Row 2 and 3—LR images with different sub-pixel shifts



**Fig. 9** DICOM images with and without contrast stretching



shifts. By properly computing the spatial shifts in the LR images, the pixels in the HR image can be approximately derived from the pixels of LR images. To train the network, the HR image is partitioned into sub-images of size  $2 \times 2$ , and the average of the pixel values in each sub-image is computed in raster scan order. The alternating sub-image averages are related to the first LR image, while the remaining averages correspond to the second LR image. Thus, in the pre-processing stage, the HR image is down-sampled by a factor of 4 with different sub-pixel shifts to generate two distinct LR images, each with a size of  $80 \times 80$ . The proposed architecture is capable of up-scaling the LR images by a factor of 4.

#### 4.3 Experimental details

The proposed network is implemented using Tensorflow and Keras. The network has been trained for 30 epochs. Both Generator and the Discriminator use Adam optimizer for optimizing the network parameters. The learning rate used for Adam is  $1e-4$  initially and is depreciated to  $1e-5$  after half number epochs. The model is trained using Google Colab Pro and NVIDIA DGX machine with GPU A100. The generator is trained with two stacked LR images of size  $80 \times 80 \times 2$  to produce a single HR image of size  $320 \times 320$ . The generator uses three losses—content loss, perceptual loss, and adversarial loss. The SR MRI image and the HR image are applied to the Discriminator, which uses binary cross-entropy for optimization.

#### 4.4 Performance metrics

The peak-signal-to-noise ratio (PSNR) is the ratio between the maximum power of an image to the power of the noise that affects the image quality. Given HR and SR images with K number of pixels, the PSNR is given by:

$$PSNR = 10 \log_{10} \left( \frac{MAX_{SR}}{\frac{1}{K} \sum_{i=1}^K (I_{HR}(i) - I_{SR}(i))^2} \right) \quad (7)$$

$MAX_{SR}$  is the maximum pixel value of the SR image and the denominator is the MSE between the HR and SR images.

**Mean Absolute Error (MAE):** It is also known as the L1 loss. It is the average of the absolute error between the HR and SR images.

$$MAE = \|I_{HR}(i) - I_{SR}(i)\| \quad (8)$$

To measure the similarity between the structures of the HR image and reconstructed SR image, the structural similarity index measure (SSIM) is evaluated. SSIM is computed as follows:

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

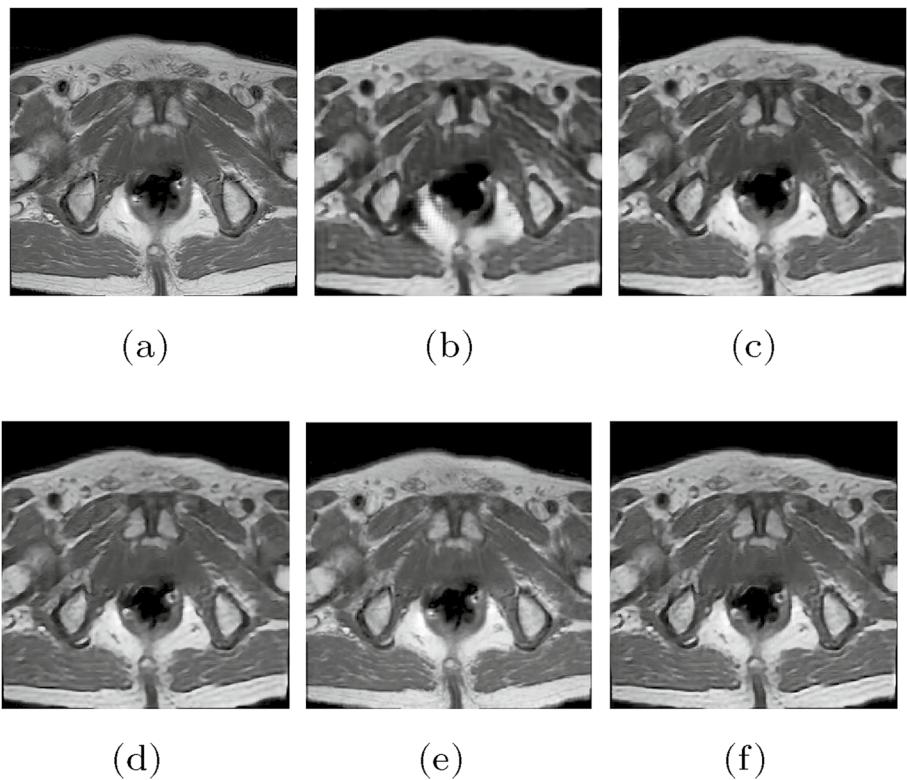
where x denotes the HR image and y denotes the SR image.  $\mu_x$  and  $\mu_y$  are the mean of HR image and SR images,  $\sigma_x$  and  $\sigma_y$  are the respective variances, and  $\sigma_{xy}$  is the covariance of x and y. The variables  $c_1$  and  $c_2$  are stabilizing variables.

#### 4.5 Results and discussion

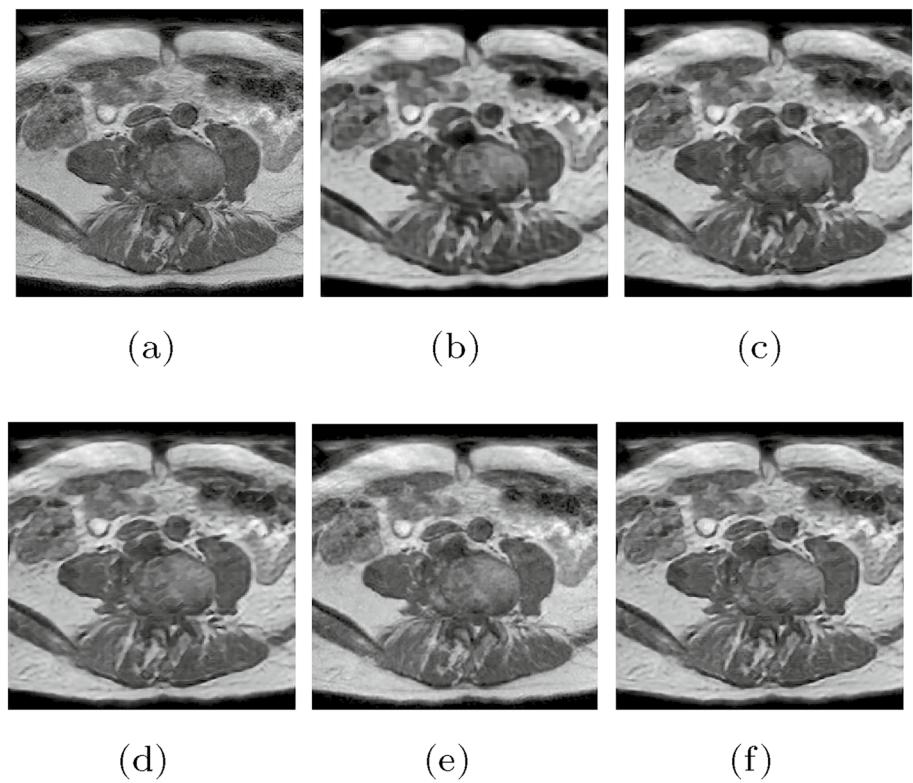
We have evaluated the performance of the proposed super-resolution algorithm using PSNR and SSIM for the Prostate cancer MRI images. The stabilization parameters  $\lambda_1$  and  $\lambda_2$  are tuned with respect to PSNR, MAE and SSIM and obtained as  $\lambda_1 = 2 * 10^{-2}$  and  $\lambda_2 = 3 * 10^{-6}$ . The stabilization parameter  $\lambda_1$  is varied between  $10^{-1}$  and  $10^{-4}$ , while  $\lambda_2$  is varied between  $10^{-2}$  and  $10^{-7}$ . The stabilization parameters need to be tuned to minimize the losses and to converge. The deep learning based super-resolution architectures (Ledig et al. 2017; Wang et al. 2018a; Molahasan et al. 2022) that employs VGG loss as the perceptual loss, uses 4<sub>th</sub> convolutional layer in the 5<sub>th</sub> blocks of the pre-trained VGG19 networks without any fine-tuning. The proposed network uses, fine-tuned VGG networks to compute the perceptual loss. The average PSNR and SSIM for the proposed network using the five VGG19 models are in the Table 2. The fine-tuned VGG19 networks and VGG 19 Gray scale networks perform better than the pre-trained VGG network. Using PSNR and SSIM metrics for evaluation, the fine-tuned VGG-19, incorporating sparse categorical cross-entropy loss and stacked MRI images as input, outperforms the other three pre-trained models. When tested and trained with 6 and 8 layers of convolutional layers, the discriminator architecture demonstrates consistent results in terms of PSNR and SSIM. This achievement is coupled with a reduction in computational complexity. The comparison results for different convolutional layers for the discriminator is in Table 3.

The proposed architecture is compared with different super-resolution techniques—ESRGAN, MSG Capsule GAN, ShuffleUNet, and SRDenseED. The average values of PSNR, MAE and SSIM for these methods are evaluated on the same Prostate Diagnosis dataset and are recorded in Table 4. The evaluation results show that the proposed architecture outweighs the other super-resolution methods both in terms of PSNR and SSIM. For the qualitative analysis, the evaluation results are shown in Figs. 10 and 11. The ESRGAN output has a blurry appearance compared to other techniques. The edges are more visible and sharper in the proposed architecture when compared with all the other techniques. The reconstruction results of the proposed architecture outweigh the other techniques both quantitatively and qualitatively. SSIM is a measure of

**Fig. 10** Super resolution results of various SR algorithms for Test data 1. **a** Ground Truth, **b** ESRGAN, **c** MSG-Cap GAN, **d** ShuffleUNet, **e** SRDenseED, **f** Proposed



**Fig. 11** Super resolution results of various SR algorithms for Test data 2. **a** Ground Truth, **b** SRGAN, **c** MSG-Cap GAN, **d** ShuffleUNet, **e** SRDenseED, **f** Proposed



**Table 2** PSNR and SSIM of proposed architecture for different VGG-19 models

Model	Loss	PSNR (dB)	SSIM
Pre-trained VGG	CCE	28.39	0.7231
VGG 19—RGB (fine tuned)	CCE	29.97	0.7663
	SCCE	30.58	0.8105
VGG 19—gray scale	CCE	29.12	0.7424
	SCCE	29.76	0.7601

**Table 3** PSNR and SSIM of proposed architecture for different convolutional layers in discriminator

Number of convolutional Layers	PSNR	SSIM
6	30.58	0.8105
8	30.51	0.8083

perceptual characteristics, out of all the five techniques, the proposed architecture has better perceptual quality.

To address the possibility of false positive cases in SR MRI images, a deep learning-based classifier is utilized, which combines CNN and SVM techniques. This approach helps to ensure that abnormalities detected in SR images are accurate and not the result of noise or artifacts. By using a combination of both CNN and SVM techniques (Sivadas and Ameer 2021), the classifier is able to learn and identify high-level features in the image, which are then used to make a more accurate classification. This helps to reduce the number of false positive cases and ensures that only valid abnormalities are detected in the SR MRI images. Two separate classifiers are used for this purpose, with one classifying the LR images and the other classifying the SR images. The same dataset that was used to train and fine-tune the VGG 19 network is used for this classification task. The accuracy achieved with LR image classifier is 82% and with SR images it is 86%. Also, with super resolution, there is no case false positive being reported. These results validate the improvement in accuracy of automated disease diagnosis with the SR MRI images over the LR MRI images.

## 5 Conclusion and future scope

We present a Generative Adversarial Network (GAN)-based deep neural network tailored for the super-resolution of multiple MRI images depicting prostate cancer. Two spatially shifted low-resolution MRI images of prostate cancer have been super-resolved to generate a single high-resolution image. The proposed method is evaluated using PSNR and SSIM, and the experimental results show that the proposed architecture is superior to the other existing super-resolution algorithms. It outperforms the other comparison algorithms for an upscaling factor of x4, with an average MAE of  $50.51 \pm 2.04$ , PSNR of  $30.58 \pm 0.76$  dB and SSIM of  $0.8105 \pm 0.0656$ . We have improved upon SRGAN by using multiple LR images as input and by using triple losses to train the generator. The Mean Squared Error (MSE) loss serves as the content loss, capturing low-frequency details, while the VGG loss acts as the perceptual loss, highlighting high-frequency/edge details. The VGG-19 model, fine-tuned for the task and utilizing sparse categorical cross-entropy loss, takes stacked MRI images as input to calculate the VGG loss. The proposed network uses two images as input, furthermore, the number of input images can be used for increased performance. The suggested framework has the versatility to extend its application to various imaging modalities like CT scans and X-Rays. It's worth noting that the same architecture can be employed to achieve super-resolution with even higher upscaling factors by increasing the number of up-sampling layers in the generator. Moreover, there remains untapped potential to achieve super-resolution of blurry and noisy MRI images. Potential avenues for improvement encompass investigating alternative loss functions, integrating supplementary training data, or refining the model's hyper parameters. Additionally, one can explore post-processing techniques like image denoising or sharpening to further augment the visual fidelity of the reconstructed images. The exploration of multi-modality super-resolution holds promise as a potential avenue for future research.

**Table 4** Average PSNR, MAE and SSIM for different architectures using prostate cancer dataset

–	ESRGAN	MSG-Cap GAN	ShuffleUNet	SRDenseED	Proposed
PSNR	$28.39 \pm 0.78$	$29.32 \pm 0.91$	$29.17 \pm 1.67$	$29.75 \pm 0.59$	$30.58 \pm 0.76$
MAE	$57.83 \pm 3.49$	$134.11 \pm 4.18$	$44.42 \pm 1.19$	$51.38 \pm 1.85$	$50.51 \pm 2.04$
SSIM	$0.6538 \pm 0.0539$	$0.7409 \pm 0.0456$	$0.7715 \pm 0.0345$	$0.7884 \pm 0.0497$	$0.8105 \pm 0.0656$

**Data availability** The datasets used for this research are available within the PROSTATE-DIAGNOSIS dataset, accessible at <https://doi.org/10.7937/K9/TCIA.2015.FOQEIJVT>.

## References

- Bloch BN, Jain A, Jaffe CC (2015) Data from prostate-diagnosis. <https://doi.org/10.7937/K9/TCIA.2015.FOQEIJVT>. Accessed 5 Dec 2020
- Bose N, Kim H, Valenzuela H (1993) Recursive implementation of total least squares algorithm for image reconstruction from noisy, undersampled multiframe. In: 1993 IEEE international conference on acoustics, speech, and signal processing, vol 5, pp 269–272. <https://doi.org/10.1109/ICASSP.1993.319799>
- Chatterjee S, Sciarra A, Dünnwald M et al (2021) Shuffleunet: super resolution of diffusion-weighted MRIs using deep learning. In: 2021 29th European Signal Processing Conference (EUSIPCO), pp 940–944. <https://doi.org/10.23919/EUSIPCO54536.2021.9615963>
- Chatterjee S, Breitkopf M, Sarasaen C et al (2022) Reconresnet: regularised residual learning for MR image reconstruction of undersampled cartesian and radial data. Comput Biol Med 143:105321. <https://doi.org/10.1016/j.combiomed.2022.105321>
- Clark K, Vendt B, Smith K et al (2013) The cancer imaging archive (TCIA): maintaining and operating a public information repository. J Digit Imaging 26(6):1045–1057. <https://doi.org/10.1007/s10278-013-9622-7>
- Costa P, Galdran A, Meyer MI et al (2018) End-to-end adversarial retinal image synthesis. IEEE Trans Med Imaging 37(3):781–791. <https://doi.org/10.1109/TMI.2017.2759102>
- Dharejo FA, Deeba F, Zhou Y et al (2021) TWIST-GAN: towards wavelet transform and transferred GAN for spatio-temporal single image super resolution. ACM Trans Intell Syst Technol 12(6):1–20. <https://doi.org/10.1145/3456726>
- Dong C, Loy CC, He K et al (2016) Image super-resolution using deep convolutional networks. IEEE Trans Pattern Anal Mach Intell 38(2):295–307. <https://doi.org/10.1109/TPAMI.2015.2439281>
- Elad M, Feuer A (1997) Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. IEEE Trans Image Process 6(12):1646–1658. <https://doi.org/10.1109/83.650118>
- Ganguly D, Chakraborty S, Balitanas M et al (2010) Medical imaging: a review, vol 78, pp 504–516. [https://doi.org/10.1007/978-3-642-16444-6\\_63](https://doi.org/10.1007/978-3-642-16444-6_63)
- Greenspan H, Oz G, Kiryati N et al (2002) MRI inter-slice reconstruction using super-resolution. Magn Reson Imaging 20(5):437–446. [https://doi.org/10.1016/S0730-725X\(02\)00511-8](https://doi.org/10.1016/S0730-725X(02)00511-8)
- HEEMALI C (2018) Brain MRI images for brain tumor detection dataset. <https://www.kaggle.com/code/heemalichaudhari/brain-tumour-detection-using-deep-learning/input>. Accessed 15 January 2023
- Hu B, Tang Y, Chang EIC et al (2019) Unsupervised learning for cell-level visual representation in histopathology images with generative adversarial networks. IEEE J Biomed Health Inform 23(3):1316–1328. <https://doi.org/10.1109/jbhi.2018.2852639>
- Irani M, Peleg S (1991) Improving resolution by image registration. CVGIP Graph Models Image Process 53:231–239. [https://doi.org/10.1016/1049-9652\(91\)90045-L](https://doi.org/10.1016/1049-9652(91)90045-L)
- Jiang M, Zhi M, Wei L et al (2021a) FA-GAN: fused attentive generative adversarial networks for MRI image super-resolution. Comput Med Imaging Graph 92:101969. <https://doi.org/10.1016/j.compmedim.2021.101969>
- Jiang Y, Gong X, Liu D et al (2021b) Enlightengan: deep light enhancement without paired supervision. IEEE Trans Image Process 30:2340–2349. <https://doi.org/10.1109/TIP.2021.3051462>
- Kim J, Lee JK, Lee KM (2016) Accurate image super-resolution using very deep convolutional networks. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp 1646–1654. <https://doi.org/10.1109/CVPR.2016.182>
- Kim G, Park J, Lee K et al (2020) Unsupervised real-world super resolution with cycle generative adversarial network and domain discriminator. In: 2020 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW), pp 1862–1871. <https://doi.org/10.1109/CVPRW50498.2020.000236>
- Lecouat B, Chang K, Foo CS et al (2018) Semi-supervised deep learning for abnormality classification in retinal images. <https://doi.org/10.48550/arXiv.1812.07832>. arXiv:1812.07832
- Ledig C, Theis L, Huszar F et al (2017) Photo-realistic single image super-resolution using a generative adversarial network. In: 2017 IEEE conference on computer vision and pattern recognition (CVPR). IEEE Computer Society, Los Alamitos, CA, USA, pp 105–114. <https://doi.org/10.1109/CVPR.2017.19>
- Lin Z, Garg P, Banerjee A et al (2022) Revisiting rcan: Improved training for image super-resolution. <https://doi.org/10.48550/arXiv.2201.11279>. arXiv preprint arXiv:2201.11279
- Lyu Q, You C, Shan H et al (2018) Super-resolution MRI through deep learning. <https://doi.org/10.48550/arXiv.1810.06776>
- Mannam V, Howard SS (2023) Small training dataset convolutional neural networks for application-specific super-resolution microscopy. J Biomed Opt 28(3):036501. <https://doi.org/10.1117/1.JBO.28.3.036501>
- Mirza M, Osindero S (2014) Conditional generative adversarial nets. <https://doi.org/10.48550/arXiv.1810.06776>. arXiv:1411.1784
- Molahasani Majdabadi M, Choi Y, Deivalakshmi S et al (2022) Capsule GAN for prostate MRI super-resolution. Multimed Tools Appl 81:4119–4141. <https://doi.org/10.1007/s11042-021-11697-z>
- Park SC, Park MK, Kang MG (2003) Super-resolution image reconstruction: a technical overview. IEEE Signal Process Mag 20(3):21–36. <https://doi.org/10.1109/MSP.2003.1203207>
- Plenge E, Poot D, Bernsen M et al (2012) Super-resolution methods in MRI: can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time? Magn Reson Med 68:1983–1993. <https://doi.org/10.1002/mrm.24187>
- Ran M, Hu J, Chen Y et al (2019) Denoising of 3D magnetic resonance images using a residual encoder–decoder Wasserstein generative adversarial network. Med Image Anal 55:165–180. <https://doi.org/10.1016/j.media.2019.05.001>
- Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. <https://doi.org/10.48550/arXiv.1409.1556>
- Sivadas D, Ameer P (2021) Automated categorization of brain tumor from MRI using CNN features and SVM. J Ambient Intell Humaniz Comput 12:1–13. <https://doi.org/10.1007/s12652-020-02568-w>
- Stark H, Oskoui P (1989) High-resolution image recovery from image-plane arrays, using convex projections. J Opt Soc Am A 6(11):1715–1726. <https://doi.org/10.1364/JOSAA.6.001715>
- Tom B, Katsaggelos A, Galatsanos N (1994) Reconstruction of a high resolution image from registration and restoration of low resolution images. In: Proceedings—international conference on image processing, ICIP, vol 3, pp 553–557. <https://doi.org/10.1109/ICIP.1994.413745>
- Trinh DH, Luong M, Dibos F et al (2014) Novel example-based method for super-resolution and denoising of medical images. IEEE Trans Image Process 23(4):1882–1895. <https://doi.org/10.1109/TIP.2014.2308422>

- Wang X, Yu K, Wu S et al (2018a) Esrgan: enhanced super-resolution generative adversarial networks. In: Proceedings of the European conference on computer vision (ECCV) workshops
- Wang Y, Yu B, Wang L et al (2018) 3d conditional generative adversarial networks for high-quality pet image estimation at low dose. *NeuroImage* 174:550–562. <https://doi.org/10.1016/j.neuroimage.2018.03.045>
- Wu X, Tian X (2020) An adaptive generative adversarial network for cardiac segmentation from x-ray chest radiographs. *Appl Sci* 10(15):5032. <https://doi.org/10.3390/app10155032>
- Xue Y, Xu T, Zhang H et al (2018) Segan: adversarial network with multi-scale L1 loss for medical image segmentation. *Neuroinformatics* 16(3):383–392. <https://doi.org/10.1007/s12021-018-9377-x>
- Yang Q, Yan P, Zhang Y et al (2018) Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Trans Med Imaging* 37(6):1348–1357. <https://doi.org/10.1109/TMI.2018.2827462>
- Yang W, Zhang X, Tian Y et al (2019) Deep learning for single image super-resolution: a brief review. *IEEE Trans Multimed* 21(12):3106–3121. <https://doi.org/10.1109/TMM.2019.2919431>
- Yi X, Walia E, Babyn P (2019) Generative adversarial network in medical imaging: a review. *Med Image Anal* 58:101552. <https://doi.org/10.1016/j.media.2019.101552>
- Yue L, Shen H, Li J et al (2016) Image super-resolution: the techniques, applications, and future. *Signal Process* 128:389–408. <https://doi.org/10.1016/j.sigpro.2016.05.002>
- Zhang K, Hu H, Philbrick K et al (2022) SOUP-GAN: super-resolution MRI using generative adversarial networks. *Tomography* 8(2):905–919. <https://doi.org/10.3390/tomography8020073>
- Zhang Y, Li K, Li K et al (2018) Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European conference on computer vision (ECCV), pp 286–301. [https://doi.org/10.1007/978-3-030-01234-2\\_18](https://doi.org/10.1007/978-3-030-01234-2_18)
- Zhang Z, Du H, Qiu B (2023) FFVN: an explicit feature fusion-based variational network for accelerated multi-coil MRI reconstruction. *Magn Reson Imaging* 97:31–45. <https://doi.org/10.1016/j.mri.2022.12.018>
- Zhu JY, Park T, Isola P et al (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: 2017 IEEE international conference on computer vision (ICCV), pp 2242–2251. <https://doi.org/10.1109/ICCV.2017.244>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.