Computers and Electrical Engineering 90 (2021) 106973

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Positron emission tomography image enhancement using magnetic resonance images and U-net structure

Farnaz Garehdaghi [a], Saeed Meshgini [b,*], Reza Afrouzian [c]

[a] Master Student, Department of Biomedical Engineering, Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran
[b] Assistant Professor, Department of Biomedical Engineering, Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran
[c] Assistant Professor, Miyaneh Faculty of Engineering, University of Tabriz, Miyaneh, Iran

ARTICLE INFO

Keywords:
Positron Emission Tomography
Convolutional Neural Networks
Residual Blocks
Perceptual Loss
Structural Similarity Index

ABSTRACT

Positron Emission Tomography (PET) has become an important tool for diagnosing abnormalities, but it suffers from low spatial resolution and a high-level noise. In this article, a Convolutional Neural Network (CNN)-based Single Image Super-resolution (SISR) method is used to produce a PET image with a desired quality. The T1-Weighted Magnetic Resonance (MR) images are used to enrich the information applied to the network. A network based on U-Net structure is used and residual blocks are inserted into the network to improve system performance. This article also evaluates the impact of various loss functions, such as Mean Squared Error (MSE) and its combination with a perceptual loss on the efficiency of the proposed method. Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) on two various databases (simulated and clinical data) are 36.78, 0.9927, and 37.36, 0.9714, respectively, indicating good performance of the proposed method compared to previous works.

1. Introduction

Positron Emission Tomography (PET) is a non-invasive nuclear medical imaging technology that determines the biochemical and metabolic functions of tissues. Since detection of chemical abnormalities provides earlier identification of diseases, this method plays a crucial role in the early diagnosis of diseases and disorders, including Alzheimer's in neurology, cancers in oncology, or stenosis in cardiology.

Also, because of the image degrading factors, such as positron range, non-collinearity of photons, patient motions, a tracer dose, and other factors, this imaging method lacks quality and spatial resolution and is impaired by noise. On the other hand, despite the success of PET images in metabolic visualization, this method does not provide any information about human anatomy. Combining PET imaging technology with modalities, such as Computed Tomography (CT) or Magnetic Resonance Imaging (MRI), provides complementary information and allows better localization

The high-quality medical images can lead to a better diagnosis. Therefore, many image-processing methods have been developed to obtain high-quality images. The simplest approach to do this is Super-resolution (SR) whose task is to obtain a high-resolution image from one or multi low-resolution inputs. The main work in this method is to recover the information lost during the acquisition process, which is mostly a high-frequency component. This means, the Nyquist frequency is not achieved while acquiring an image and the

* Corresponding author.
E-mail addresses: f.gharedaghi96@ms.tabrizu.ac.ir (F. Garehdaghi), meshgini@tabrizu.ac.ir (S. Meshgini), afrouzian@tabrizu.ac.ir (R. Afrouzian).

https://doi.org/10.1016/j.compeleceng.2021.106973
Received 22 May 2020; Received in revised form 23 December 2020; Accepted 17 January 2021
Available online 25 January 2021

imaging process is done in low frequency and high-frequency details are discarded. On the other hand, noise and geometric artifacts in medical imaging are other problems that should be considered.

Super-resolution methods can be categorized into two groups. The first one is Multi image super-resolution (MISR), where a series of low-resolution images with sub-pixel shifts are given. In these methods, images with sub-pixel alignments are acquired from different points of view by shifting or rotating the detector or pixel grids, and then, the images are combined, and a high-resolution image is reconstructed. Kennedy's algorithm is an example of this category for PET images [1].

In the second one, a single image is used as input called a single image super-resolution (SISR) [2] that can also be split into two groups. The first group uses only an image, such as interpolation methods that use mathematical formulas and methods, and self-similarities of an image. In the self-similarity-based SR methods, the image is scanned and similar patches are found and utilized to estimate missing details for each patch. This method is mostly used in natural images with repetitive textures. The second one uses a group of images to learn the relation between Low Resolution (LR) and High Resolution (HR) images (external database of LR-HR image pair).

The dictionary learning and linear regression methods are examples of the second group, which use Machine Learning (ML) methods to learn a relationship between LR-HR image pairs [3]. The Convolutional Neural Networks (CNNs)-based super-resolution methods are other examples of the last group, which have been introduced recently.

The rest of the article is organized as follows: related works are given in Section 2. Materials and methods are mentioned in Section 3. In Section 4, the proposed method is explained in detail. Results and discussion are reported in Section 5, and in the last section, conclusions are described.

2. Related works

This article proposes a single image super-resolution-based method that uses a convolution neural network. Therefore, a brief review of SISR and PET image enhancement by convolutional neural networks is given here. The pioneering CNN-based SR method (SRCNN) was introduced by Dong et al. in 2015. It included a shallow network that had three convolutional layers [4]. This method showed better performance than previous works but using a low number of layers was the reason for not performing well. In this article, adding the number of layers did not increase the performance because choosing a low learning rate resulted in low convergence speed. Later, a deeper network with 20 layers, called Very Deep Super-resolution (VDSR), was introduced by Kim et al., which utilized global residual learning [5]. This method utilized a high learning rate to converge the network faster; it therefore can use deeper networks to enhance performance. This method also utilized up-sampled images as a network input, so it had a higher computational cost.

Subsequently, Super-resolution Generative Adversarial Networks (SRGANs) were introduced, including a generator network and a discriminator system [6]. The generator network is a super-resolution ResNet called SRResNet with B residual blocks and skip connections, and solves the gradient vanish problem [7]. The inputs of this method are small down-sampled images that decrease the computational cost. However, this network could not create super-resolution images on all scales in one network that was a defect of this method. Inspired by SRResNet, Lim et al. proposed the Enhanced Deep Residual Super-resolution (EDSR) [8] method that won the NITRE 2017 super-resolution challenge. In this method, batch normalization layers were removed, and as a result, the network used less memory, leading to the increased number of layers. A multi-scale method was later introduced to super-resolution at all scales.

The above-explained methods were examples of natural image super-resolution to overcome the low spatial resolution of PET images, some of which applied to PET images are explained below.

One of the methods to acquire a high-quality PET image is to use high-dose tracer, that can increase the risk of radiation damage. Therefore, there have been some efforts to estimate a high-dose PET image from a low-dose one [9–12]. One of these methods has been proposed by Kang et al. In this work, a tissue-specific regression forest was used to predict the target high-dose PET image from low-dose one and the corresponding MR image. Their work led to an average PSNR of 22.217 on the used clinical dataset [9]. In other work, a mapping-based sparse representation was used by Wang to improve the results on the same database that had a dictionary for LPET and SPET and a mapping between them. Since sparse coding methods are really time-consuming, this method utilized a patch selection-based dictionary that reduced the processing time [10]. Another method has recently de-blurred the PET images by spatially variant de-convolution stabilized by MRI [13].

Owing to the success of CNNs, specially SRCNN in SR, Xiang et al. tried to use this implementation to have a better and faster estimation of high-dose PET images [11]. They concatenated low-dose PET images with T1-weighted MR images and used them as input to a basic four-layer CNN. They repeated this architecture three times to have a deeper auto-context like network and a faster estimation in the testing stage and higher PSNR than previous works. Although this method required more time to train the network, it only took two seconds in the experiment to estimate a high-dose PET image.

Xu et al. even used a lower dose of tracer [12]. They used 200x low-dose PET images as input to estimate a standard one through an encoder-decoder deep network with skip connections called U-Net previously introduced for image segmentation [14].

One of the works done recently to obtain a high-resolution PET image using multi-channel inputs has been done by Song et al. [15]. Inputs are low-resolution PET, high-resolution MRI, and radial and axial coordinate locations, with high-quality PET images as targets and layers of different sizes. The shallow one had three convolutional layers, inspired by SRCNN [4], and the deeper one had 20 layers, the same as for the VDSR [5] method. Residual learning was utilized in both networks where differences between LR and ground-truth were used instead of completely applying the PET image to CNN. In addition, a Rectified Linear Unit (RELU) layer was used after every convolutional layer to speed up training. The results show that the deeper network can estimate better resolution than the shallow one, and additionally, when more information is used at the input layer, results achieve higher resolution.

Another example of using CNNs for PET image enhancement can be found in [16]. In this method, sinograms of PET images with large crystal sizes are used to obtain high-quality images. Malczewski also uses a combination of compressed sensing and super-resolution methods to achieve high performance in enhancing the quality of PET images [17].

This article proposes a method to increase the resolution of PET images. In order to achieve this purpose, a new network is designed based on the capabilities of the state the art CNNs such as U-Net and Res-Net. Both of them are considered successful methods in the field of image classification and super-resolution. Hence, the proposed network tries to use the capabilities and designs a single network based on them. Of course, while designing the network, the purpose of this article has always been considered, which is super-resolution in medical images. So, Changes and adjustments in their structure (U-Net and Res-Net) are made in accordance with the purpose of the proposed method.

Compared to similar works, instead of the MSE criterion, a new method based on MSE and perceptual loss is used for training the proposed network. In this way, this article uses a subjective method for network training that is more in line with the goals of super-resolution methods. Exploiting MRI images (as input to the network) has an important role in intensifying the quality of PET images. In fact, it feeds more detailed information to the network. Hence, this article investigates different methods for simultaneously applying MRI and PET images to the network to find the method that has the best performance for PET images super-resolution.

The main contributions of this article were highlighted in the above two paragraphs.

3. Materials and Methods

3.1. Convolutional Neural Networks (CNNs)

Convolutional neural networks were introduced many years ago, but they recently have shown great performance in many tasks. CNNs are a type of multi-layer perceptron neural networks but use local receptive fields. The typical use of CNNs is in classification

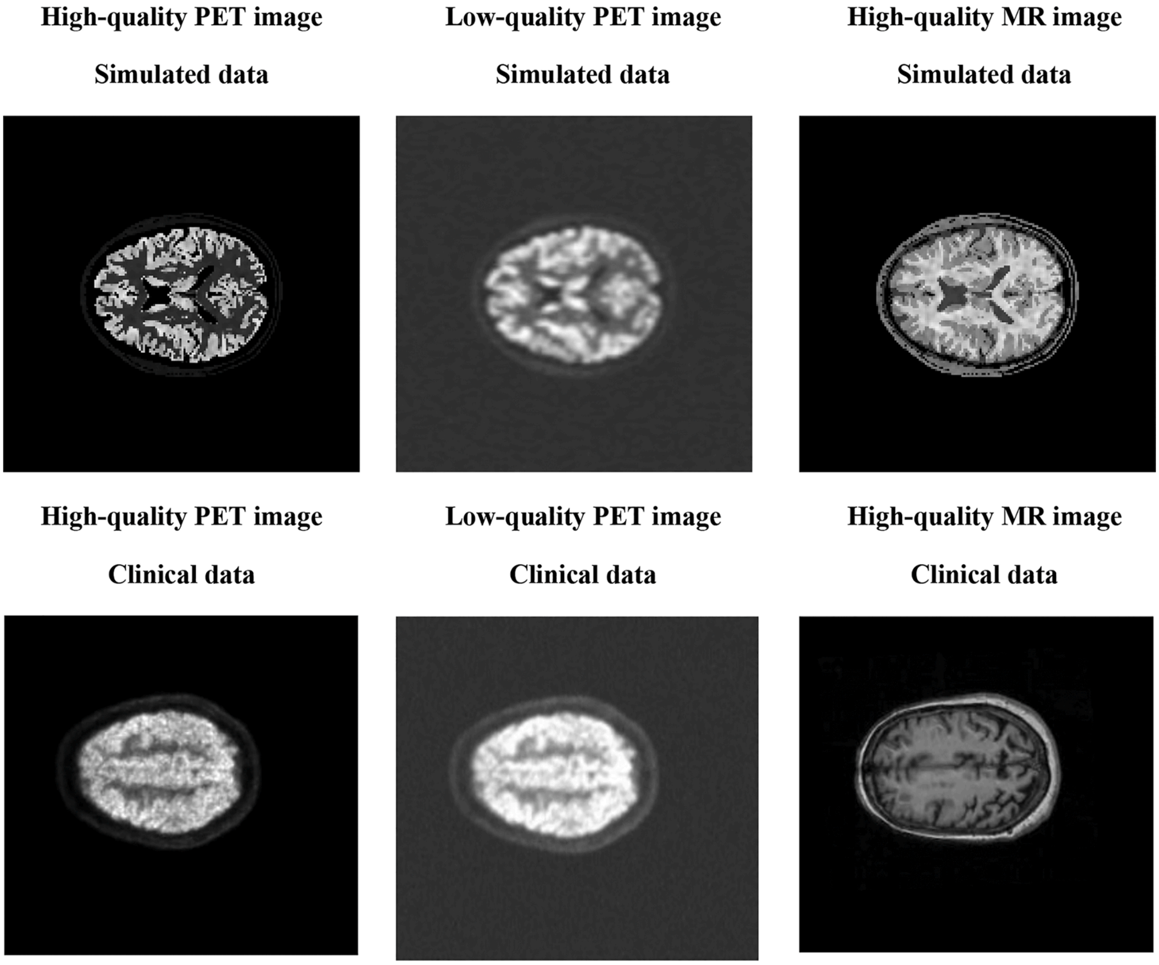

Fig. 1. Examples of high-quality PET images, low-quality PET images and high-quality MR images of the two databases. The upper ones are examples of simulated database and lower ones are examples of clinical database.

tasks, but they have also shown good performance in many other tasks such as object detection, super-resolution, natural language processing, etc. In these methods, the convolution operation is done to extract high-level features of the input data. Equation (1) shows the operation done in a CNN layer with an activation G. Where x and y are the input and target of the network, respectively, and w and b are the weights and biases of that layer, and so "*" is the convolution operator. The aim is to learn the weights and biases.

$$y = G(w * x + b) \tag{1}$$

The progress of CNNs depends on using large datasets, parallel computing using GPUs, efficient structures, and tricks, such as Rectified Linear Unit (ReLU) [18] and batch normalization (BN) [19], which help to speed up convergence.

3.2. Inputs to Network

Since low-quality images and their corresponding high-quality ones are used to train CNNs in image super-resolution and denoising tasks, in this method, high-quality PET images are acquired from the dataset. Since there are no low-quality images (corresponding to the high-quality type) in the dataset, their low-quality images are obtained from them by some degrading factors in Equation (2) [20]. In this Equation, to degrade the PET images, first down-sampling by factor 2 and again, an up-sampling using bicubic interpolation method (to have smoother results) are applied. Then, since Poisson noise is considered as a PET imaging system noise, a Poisson noise with a mean value of 50 is added to the input image [15]. After that, a motion blur with 15 ° angle and 2-pixel length is applied.

$$y = DHx + \varepsilon \tag{2}$$

where, D is the down-sampling operator, H indicates blurring operator, and ε is additive noise, and then, x and y represent the high-quality and low-quality images, respectively, which are used as input and output in network training.

3.2.1. Database 1

In order to enhance the accuracy of the proposed method, MR images are also used. Hence, the BrainWeb dataset is utilized for that purpose [21]. BrainWeb is a simulated brain image database where simulated MR images are publically available, and then PET images are generated from MR images using the BrainWeb library. 20 normal brain images are utilized. 15 brain images from them are used for training, and five images for testing. The degrading process explained above is applied to them, and all database images are cropped so that only the center part of the images, which contains 176 × 176 pixels, remains. This method prevents us from entering useless information into the network. In addition, 55 slices from every brain image in the axial plate are extracted. An example of a low-quality PET image, high-quality MR image, and a high-quality PET image is illustrated in the first raw of Fig. 1.

3.2.2. Database 2

The Alzheimer Disease Neuroimaging Initiative (ADNI) is also utilized to see the network performance on clinical data [22]. The MRI and PET images, which belong to five people, are used for training, and two ones also are used for testing. 100 axial images from every person are selected. For this database, the PET images were taken from the HRRT PET scanner, and MR images for the same data were acquired from the MPRAGE system. PET images are acquired in the axial plate and MR images in the sagittal plate but registered by software. The size of images is 256 × 256. An example of this database is also shown in the second raw of Fig. 1.

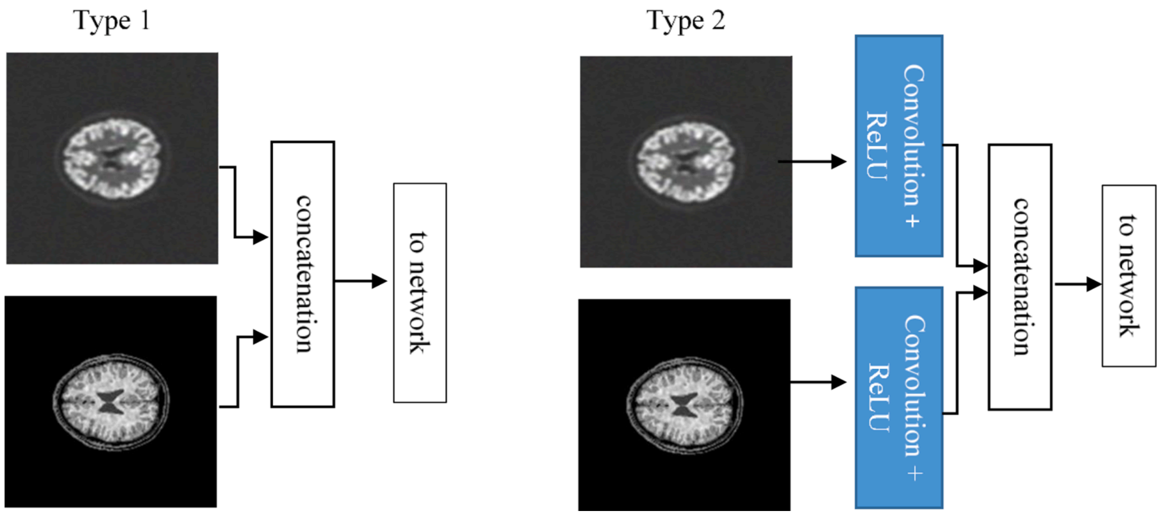

Fig. 2. Two methods for applying MRI and PET images to the network simultaneously. In first type, the images are concatenated and then applied to the network, in the second one, the inputs are first entered to a convolution and ReLU layer and then concatenated and applied to the network.

4. Proposed Method

4.1. Inputs to the Network

In the previous works, which were done in relation to PET image super-resolution and PET image estimation, adding MR images increased the efficiency and accuracy of the results [9-11, 13, 15]. Therefore, this article uses MR images along with PET images. The proposed method evaluates different methods to apply PET and MRI images simultaneously to the network. In the first method, PET and MR images are concatenated and then applied to the network. But in the second one, similar to [15], first, a convolution layer with ReLU is applied for every PET and MRI images individually. Then, the concatenation stage is done, as shown in Fig. 2.

4.2. Network Structure

The proposed method uses an architecture based on U-Net and ResNet architectures [7, 14,23].

A typical U-Net structure has encoder and decoder parts. The inputs are down-sampled in the encoder part and, once again, are up-sampled in the decoder part. Since the down-sampling stage can lead to loss of information, a concatenation is done in the channel dimension with skip connections to access the lost information. This network was first introduced for medical image segmentation [14] and then was used in other tasks such as image reconstruction [12]. For the tasks in which their inputs and outputs are of the same size (e.g., segmentation and image generation), U-Net performs well.

Deep neural networks usually need thousands of images for training, and on the other hand, these numbers of images are not so much possible in medical images. Hence, U-Net uses skip connections and data augmentation, which leads to good performance even in a limited number of images. U-Net also prevents overfitting, which is one of the advantages of this type of network. A defect of the network mentioned above can be the low speed of training.

It is confirmed that an increased number of layers in a network (depth of the network) can improve the network performance, but at the same time, after some layers, the network's accuracy starts to saturate. The mentioned saturation is not due to overfitting, but the main reason is training error. Hence, deeper models have higher training errors than shallower ones, which can be solved using

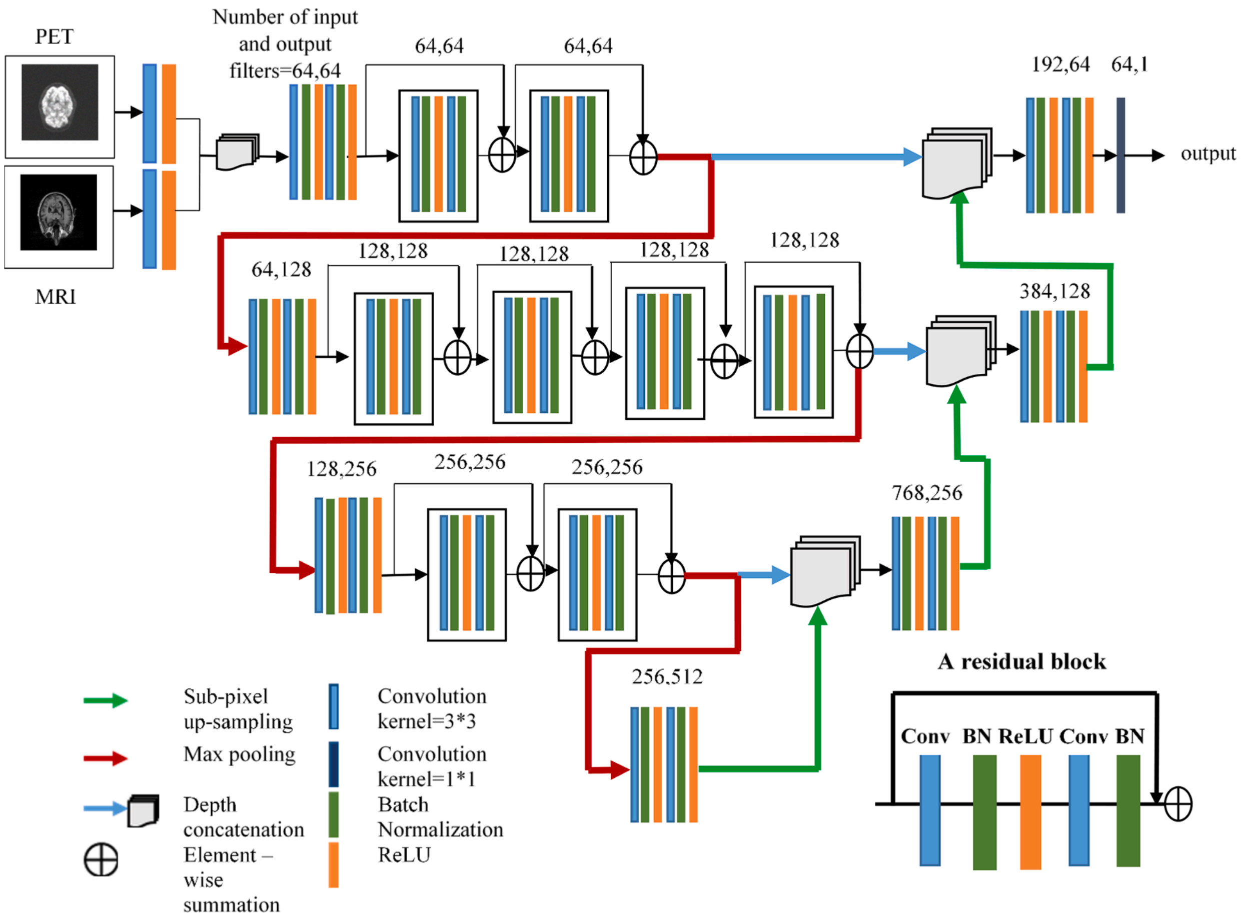

Fig. 3. The block diagram of proposed network and a residual block

residual learning. In such networks, a shortcut connection in a block skips one or more layers and adds the block's input to the output [7]. This connection does not add any extra parameter or complexity, but as an identity map helps the network have higher training error than shallower one. It also avoids gradient vanishes and helps to improve the optimization stage in deep networks.

As mentioned above, the proposed method designs its network based on U-Net and Res-Net. In a typical U-Net structure, the encoder and decoder parts are the same exactly. While in the present study, inspired by [23], residual blocks are added to the encoder part to increase the network depth and, as a result, increment the network performance. Utilizing residual block in the U-Net architecture can increase the speed of the training process and can help to use a deeper network. Every residual block used in this article has two convolution layers, which are followed by Batch Normalization (BN) to speed up convergence [19] and a Rectified Linear Unit (ReLU) layer to avoid gradient vanishes [18].

The U-Net structure, which is utilized in this work, has three down-sampling and up-sampling stages. Down-sampling is done by max-pooling layer, and every time, images are resized to half, but the depth of the feature map is multiplied by 2 to achieve high-level features. Up-sampling is done by the sub-pixel up-sampling method, introduced by Shi et al.[24], to avoid transposed-convolution layer's checker board effect. The architecture of the proposed method is shown in Fig. 3. All of the kernel sizes are 3 × 3, but the last one is 1 × 1 to map the output to the desired depth.

4.3. Loss Function

Mean Squared Error (MSE) loss is one of the famous loss functions used in super-resolution tasks to find the differences between the target images and predicted ones, and it is used for optimization. This pixel-based loss function that attempts to approximate the predicted image to target one pixel per pixel often leads to blurred results. On the other hand, the PSNR ratio, an evaluation metric in super-resolution tasks, would be high for such a predicted image. Equation 3 expresses the MSE loss where y_{ij} and $\widehat{y}_{ij}$ are the value of the ith and jth pixel of the target and predicted image, respectively, and also, n and m are the numbers of image pixels in horizontal and vertical directions.

$$MSE = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left(y_{i,j} - \widehat{y}_{i,j}\right)}{mn} \tag{3}$$

A loss function that is recently utilized and has presented good visual results in super-resolution tasks is the perceptual loss function [25]. This loss function encourages the extracted features of the predicted image and target one to be close to each other. Since this is not a pixel-based loss function, it can lead to better visual results. This loss function computes the Euclidean distance between extracted features (from layers before every max-pooling layer) in a pre-trained network (which usually are VGG networks) and tries to make them approach each other. As this function is not a pixel-based loss function, quantitative super-resolution resulting from it may not be as good as the previous functions such as MSE and PSNR. The equation below shows the perceptual loss.

$$l_{feat}^{\phi,j}(y,\widehat{y}) = \frac{1}{C_jH_jW_j}\| \phi_j(y) - \phi_j(\widehat{y}) \|_2^2 \tag{4}$$

where $l_{feat}^{\phi,j}$ is the perceptual loss at the jth layer of the network φ. Let $C, H,$ and W be the number of channels, height, and width of the input image, respectively, which leads to the feature map of shape $C_jH_jW_j$. $\varphi_j(y)$ and $\phi_j(\widehat{y})$ are the features extracted from layer j of the network φ for the target image y and predicted image $\widehat{y}$.

In this article, to achieve good quantitative results and good visual results, a weighted summation of these two functions is utilized. This loss function was first introduced by the FastAI library called feature loss. The pre-trained network used here is VGG16 with batch normalization layers. Layers 7, 10, and 13 of this network are used for feature extraction. Equation 5 illustrates the loss function used here.

$$loss - function = MSE + 2l^7 + 3l^{10} + l^{13} \tag{5}$$

where l^7,$l^{10}$and $l^{13}$are the perceptual loss computed in layers 7, 10, and 13.

5. Experiments and Results

5.1. Evaluation metrics

Evaluation metrics, including PSNR, Root Mean Squared Error (RMSE), and Structural Similarity Index (SSIM) are used for SR tasks. These metrics are explained in Equations 6,7,and 8.

$$PSNR = 10.\log_{10}\left(\frac{MAX^2{}_I}{MSE}\right) \tag{6}$$

$$RMSE = \sqrt{MSE} \tag{7}$$

$$SSIM(\widehat{y}, y) = \frac{\left(2\mu_{\widehat{y}}\mu_y + c_1\right)\left(2\sigma_{\widehat{y}y} + c_2\right)}{\left(\mu_{\widehat{y}}^2 + \mu_y^2 + c_1\right)\left(\sigma_{\widehat{y}}^2 + \sigma_y^2 + c_2\right)} \tag{8}$$

where, MAX_I is the maximum intensity of the image, $\mu_{\widehat{y}}$and $\mu_y$indicates the average of the predicted and target image, respectively.$\sigma_{\widehat{y}}$, and$\sigma_{\widehat{yy}}$show the variance and covariance of the predicted image $(\widehat{y})$ and target one (y). c_1 and c_2 are parameters for stabilizing the divisions considered equal to $(k_1L)^2$ and $(k_2L)^2$, respectively, where L is the dynamic range of the image ($k_1 = 0.01$and $k_2 = 0.03$).

5.2. Experiments on simulated data

In the first experiment, a comparison is made on how the inputs are applied to the network. In the first one, the PET images are used alone for super-resolution. In the next, both types of PET and MRI images are utilized. For applying them simultaneously, as mentioned previously, two methods are used in this article. The results were summarized in Table 1. It is inferable from Table 1 that applying two types of images simultaneously will improve the system performance.

Three experiments listed in Table 1, are trained and tested separately for 50 epochs. The learning rate and the kind of optimizer for them are the same to have a fair comparison. The learning rate that is a hyper-parameter to control the speed of convergence is set to 1 × 10^{-4} and is multiplied by 0.1 in every 30 epochs. The Adaptive Momentum (ADAM) is chosen as the network optimizer. The numbers of 825 axial slices are used for training in a shuffling way, and 275 slices are chosen for testing the network performance. Data augmentation is also used to avoid overfitting of the network and increase the input number. This comparison is made on the simulated database, called database 1 in this article. Better results are shown in bold type.

As Table 1 shows, PSNR for the proposed method without using MR images is 25.68, and adding them has increased the network performance. The PSNR criterion while using both PET and MRI images (in type 2 of Fig. 2) is increased to 36.78. This is because, as mentioned in the previous section, both types of images (MRI and PET) are applied into separate convolution layers with ReLU before concatenating them. In this case, instead of concatenating the raw images with each other (type 1), the network itself decides how to combine and concatenate the above information, which is the main purpose of convolutional neural networks.

To evaluate the impact of loss function on the network performance, the proposed network is trained and tested separately on both simulated and clinical data with different loss functions. The results are summarized in Tables 2, and 3, and their qualitative results are presented in Figs. 4 and 6. The number of epochs is equal to 100, and the learning rate and optimizer in all experiments are the same. As expected, the MSE loss function leads to better results in PSNR and MSE. On the other hand, the new loss function, which is a combination of MSE and perceptual loss, presents better visual results. As shown in Figs. 4 and 6, the MSE loss function results in smooth edges, but perceptual plus MSE loss results in sharper edges (Of course, this change is not seen in the simulated images).

In the next experiment, the proposed network is compared with the previous works. In order to make a fair comparison, the database used for training and testing of networks, the learning rate, and the degrading stage were considered the same in all networks.

We compare the proposed network with the networks which are presented in [15] with 3 and 20 layers. In both networks, each convolution layer has 64 kernels with a size of 3 × 3 and a RELU unit follows every convolution layer. Also, a U-Net structure, such as the method proposed in [12], is used for comparison with a little modification (the up-sampling is sub-pixel up-sampling). Note that the U-Net structure is the same as the proposed method, but without residual blocks. The SRResNet network, which is utilized in SRGANs, is also implemented with five residual blocks [6]. The results of the mentioned networks and the proposed method are listed in Table 2, and better results of each column are written in bold type. In addition, their qualitative results are shown in Fig. 4.

The numerical results summarized in Table 2 prove the good performance of the proposed network compared to other networks. The PSNR performance measure of the proposed method (while using only MSE loss) is 36.78 that is higher than other methods.

Looking carefully at the results listed in Tables 2 and 3, a fine corollary can be observed. As shown in these tables, the training based on the new loss function (a combination of MSE and perceptual loss) decreases PSNR to 35.02 and increases SSIM. This is justified because SSIM is a human vision-based perceptual metric and higher SSIM means higher visual quality and more similarity to the target image. Since a combination of perceptual loss is used to train the network, the weight updates are done to decrease the perceptual errors and increase the SSIM. On the other hand, given that, the network learning method is not fully compatible with the MSE loss, the PSNR criterion reduction will not be unexpected.

As shown in the tables, two popular methods, SRResNet and U-Net, show higher PSNRs, a convenient reason for using them in the proposed network. As we explained before and experimental results show, designing a network based on the capabilities increases the network's performance compared to each of them alone. Fig. 5 also shows the level of training loss concerning the number of epochs on the simulated data for our proposed network and different methods for comparison.

Table 1
Results of comparing different inputs on simulated data

Method	Input Type	Loss (MSE)	RMSE	PSNR	SSIM
Proposed Method	PET	0.002570	0.0520	25.68	0.9494
Proposed Method	PET/MRI (1)	0.000570	0.02387	34.45	0.9732
Proposed Method	PET/MRI (2)	0.000207	0.01585	36.78	0.9927

Table 2
Numerical results of comparing different methods on simulated data

Method	Loss function	Loss	RMSE	PSNR	SSIM
Original	MSE	0.037362	0.19402	14.24	0.1732
3 layer [15]	MSE	0.000374	0.01933	34.57	0.9917
20 layer [15]	MSE	0.000373	0.01952	34.21	0.9901
U-Net [12]	MSE	0.000268	0.01701	35.39	0.9926
SRResNet [6]	MSE	0.000254	0.01590	36.01	0.9917
The proposed method with MSE loss	MSE	0.000207	0.01585	36.78	0.9927
The proposed method with perceptual + MSE loss	Perceptual + MSE	0.000507	0.01784	35.02	0.9939

Table 3
Numerical results of comparing different methods on clinical data

Method	Loss function	Loss (MSE)	RMSE	PSNR	SSIM
Original	MSE	0.033090	0.1826	14.76	0.2632
3 layer [15]	MSE	0.000310	0.0176	35.62	0.9734
20 layer [15]	MSE	0.000252	0.0159	35.98	0.9716
U-Net [12]	MSE	0.000219	0.0148	36.59	0.9421
SRResNet [6]	MSE	0.000176	0.0135	37.26	0.9713
Proposed Method	MSE	0.000175	0.0132	37.53	0.9714
Proposed Method	Perceptual + MSE	0.00452	0.01511	36.44	0.9771

5.3. *Experiments on clinical data*

In order to evaluate the performance of the proposed network on a real dataset, the clinical data is used for training and testing. The quality reduction stage is done on the real dataset is the same as the simulated one. All of the networks, which are explained previously, are also trained and tested on real data. The results were summarized in Table 3, and their qualitative results are shown in Fig. 6.

Table 3 shows that the RMSE criterion for the proposed network is the minimum value. In addition, the PSNR score for our network is 37.53, which is higher than other networks. This criterion for the proposed network with a new loss function is 36.44. To justify this observation, similar to the simulated data results, it is notable that when a new loss function is used for the training stage, it allows the network to reconstruct fine details and edges better while maintaining the precision of quantitative results with acceptable accuracy.

The similarity between high-quality PET image and the output of the proposed method (new loss function based on the combination of MSE and perceptual loss) can be seen in Fig. 6 and in the SSIM value in Table 3. In addition, the outputs of other methods have been presented in Fig. 6 for comparison. Hence, in the case of using a new loss function (MSE and perceptual loss), it can be stated that the reduction in accuracy of quantitative results is justified by the significant increase in the qualitative results shown in Fig. 6.

Also, the MSE score concerning the number of epochs on the clinical data is shown in Fig. 7 for the proposed network and other methods for comparison. As the figure shows, the proposed network has less training error compared to other networks.

As mentioned in the related works section, in designing the proposed network, the purpose of this article has always been considered, which is PET images super-resolution. In order to accurately design the number of residual blocks in each down-sampling stage, another experiment is performed on clinical data. In this experiment, we test other choices for the number of residual blocks to clarify this selection's effect on the network's accuracy. The results of these experiments were summarized in Table 4, when the number of residual blocks is varied in stages one to three.

It is inferable from Table 4 that the change in the number of residual blocks in stages one and three doses not considerably affect the number of evaluation metrics. However, the network performance improves by increasing the number of blocks in the second stage. Of course, increasing the number of blocks by more than four does not significantly affect PSNR. Hence, as shown in Fig. 3, the number of residual blocks in stages one to three is adjusted to two, four, and two, respectively.

In order to raise the challenge and evaluate the network performance in a newly defined condition, another experiment is done in clinical data. This experiment defines a new condition to produce low-quality images. The down-sampling and up-sampling scale is changed to 4, and the mean of added noise is considered equal to 75 (the motion blur is the same as before). The results of such changes were summarized in Table 5.

By comparing the results in Tables 3 and 5, it can be concluded that further quality reduction in input images leads to the decreased accuracy of the results (MSE increased, and PSNR and SSEM criteria decreased). Of course, such a conclusion is not unexpected. However, it is worth mentioning that the difference between the proposed method's results and similar works in Table 5 is more significant. It can also be interpreted that in a way that in cases where the input image's quality is severely reduced, the proposed network (compared to other networks) has worked better. Also, according to the column of results related to the SSIM criterion in Table 5, the proposed training method's accuracy (MSE+ Perceptual loss) is considerably higher than other methods.

6. Conclusion

PET is an imaging modality that plays a crucial role in diagnosing disorders. Nevertheless, this imaging modality suffers from low

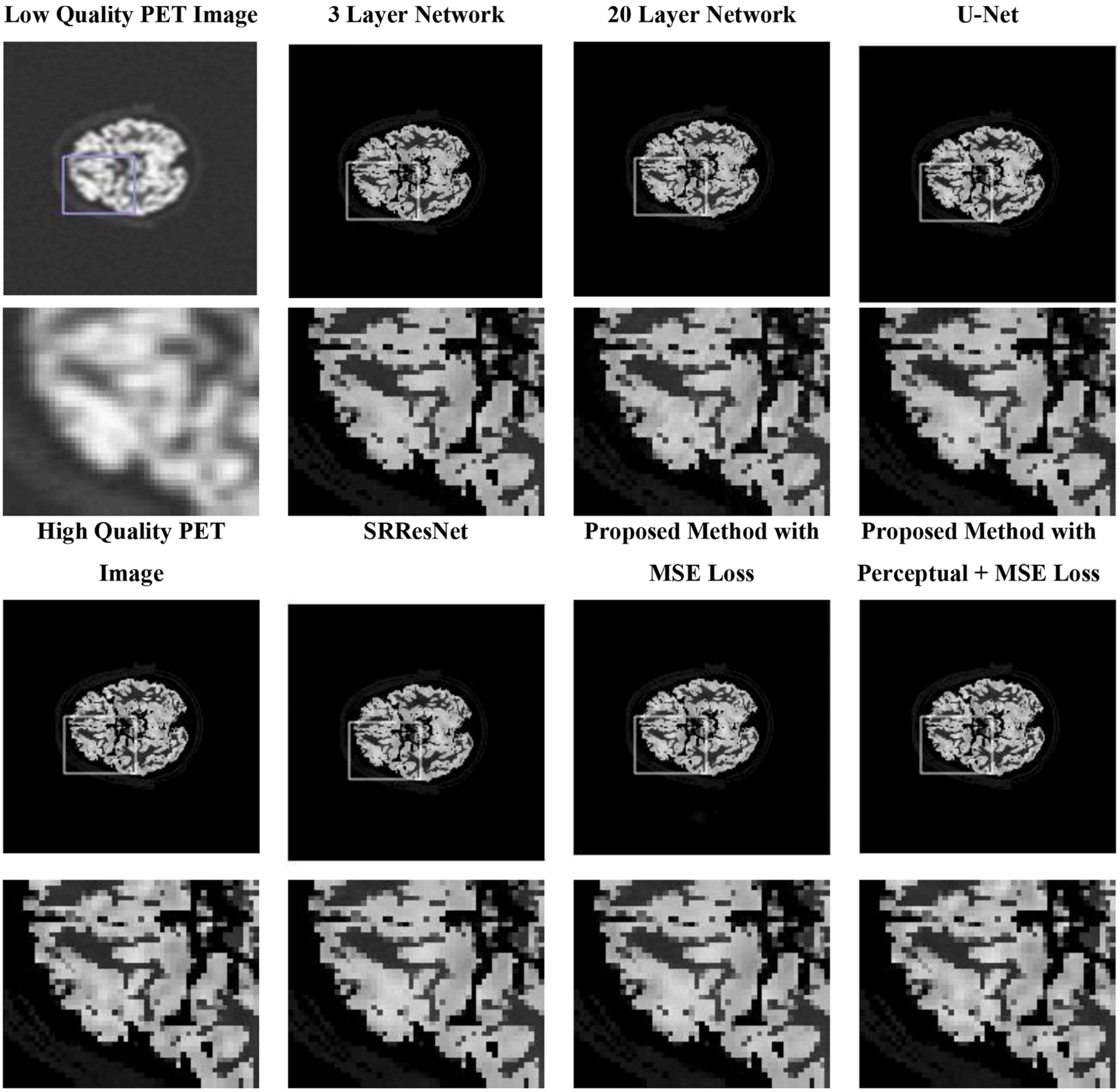

Fig. 4. Qualitative comparison for experimental results of different methods on simulated database

spatial resolution and a high noise level. In order to acquire high-quality PET images, there are some methods. The easiest one is to create super-resolution images after acquiring them. The proposed method is a single image super-resolution system that receives a single image and gives a single high-resolution one. Since the Convolutional Neural Networks recently have achieved good performance on various tasks such as SISR, a CNN-based SR system was proposed here. The proposed method designs a network based on the capabilities of successful CNNs such as U-Net and Res-Net. This allows us to take advantage of the strengths of the two methods simultaneously. Besides, two kinds of datasets were used to evaluate the proposed method's performance on simulated and clinical data. To evaluate the proposed method's performance, this network was compared with previous methods presented recently, which showed better results in terms of PSNR and RMSE performance measures. A new loss function based on MSE and perceptual loss was utilized to increase the quality of visual results and SSIM value in the proposed network. Low-quality images were manually obtained by the down-sampling of high-quality ones. If generative networks can be used for generating low-quality images and some more images to feed as input to the network instead of data augmentation, the results of the network would be more reliable.

CRediT authorship contribution statement

Farnaz Garehdaghi: Conceptualization, Methodology, Software, Writing - original draft. **Saeed Meshgini:** Conceptualization,

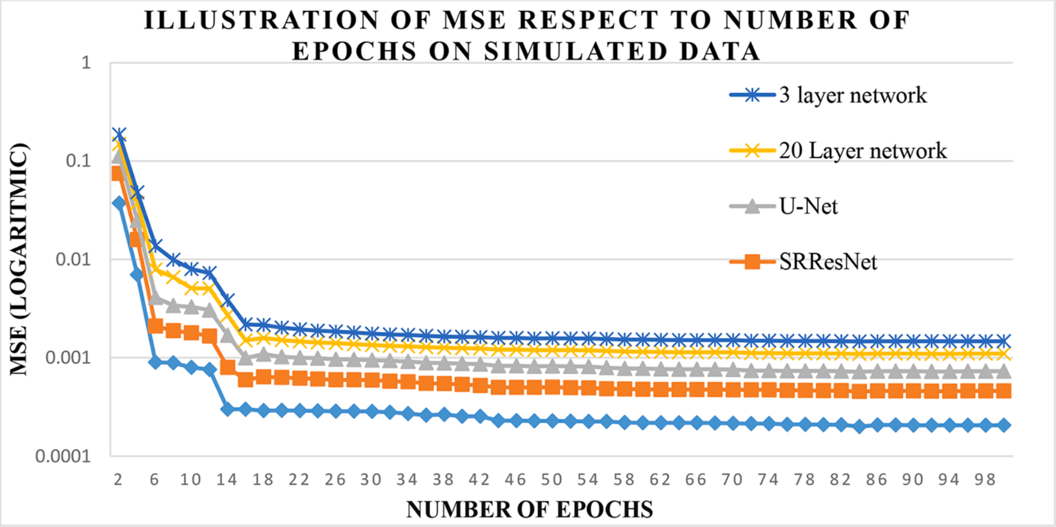

Fig. 5. Training loss based on MSE concerning the number of epochs on simulated data for different methods. The vertical axis is on a logarithmic scale.

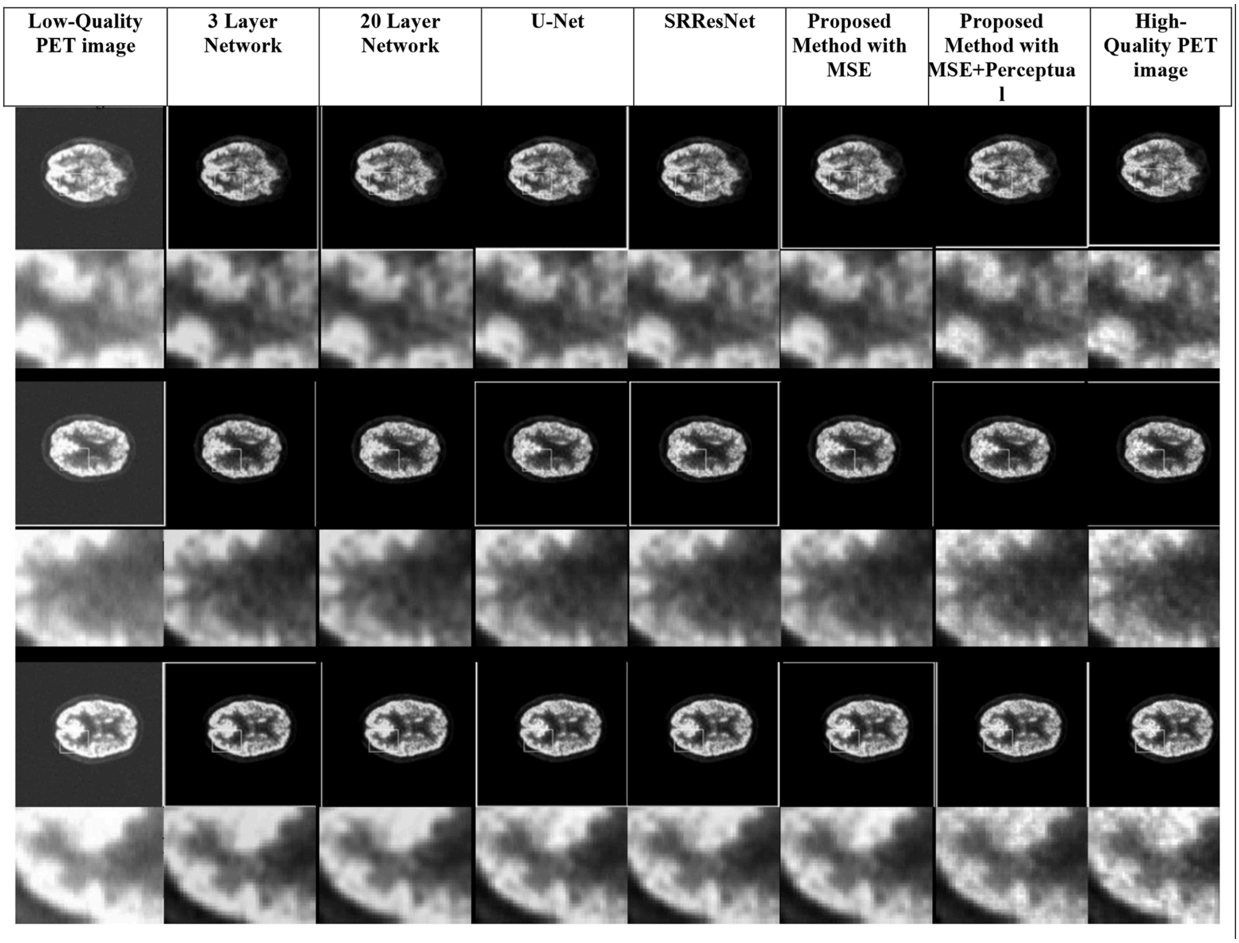

Fig. 6. Three examples of qualitative results for the proposed method with perceptual loss and without perceptual loss criterion. In addition, the results of similar networks for comparison are presented in this figure.

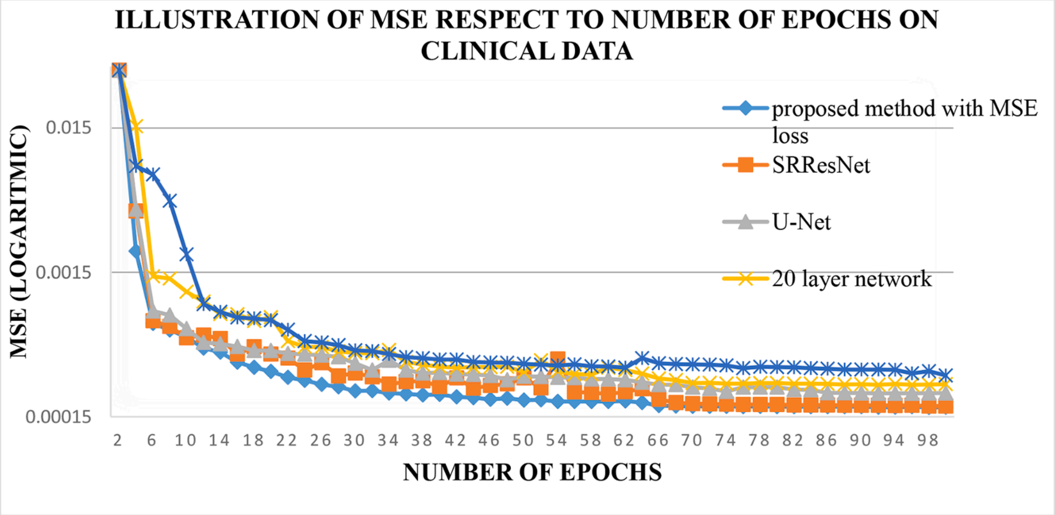

Fig. 7. Training loss concerning the number of iterations on clinical data for different methods. The vertical axis is on a logarithmic scale.

Table 4
Impact of variation in the number of residual blocks on the network performance (based on PSNR, RMSE, and SSIM). The number before "A" shows the number of residual blocks in stage one of down-sampling, and also the numbers before "B" and "C" show the number of residual blocks in stage two and three, respectively.

Method	Loss function	Loss (MSE)	RMSE	PSNR	SSIM
2A2B2C	MSE	0.000439	0.02251	33.04	0.872
3A2B2C	MSE	0.000423	0.02067	33.32	0.881
4A2B2C	MSE	0.000412	0.02029	33.50	0.890
2A3B2C	MSE	0.000395	0.01991	34.85	0.9012
2A4B2C	MSE	0.000351	0.01873	35.41	0.9245
2A5B2C	MSE	0.000352	0.01876	35.34	0.9188
2A2B3C	MSE	0.000437	0.02090	33.12	0.887
2A2B4C	MSE	0.000426	0.02063	33.25	0.891

Table 5
Numerical results of comparing different methods on clinical data with higher noise and down sampling scale of 4

Method	Loss function	Loss (MSE)	RMSE	PSNR	SSIM
Original	MSE	0.07533	0.2744	11.23	0.2058
3 layer [15]	MSE	0.00066	0.0256	31.82	0.9148
20 layer [15]	MSE	0.00062	0.0249	31.96	0.9169
U-Net [14]	MSE	0.00054	0.0234	32.14	0.9214
ResNet [6]	MSE	0.00052	0.0229	32.56	0.9331
Proposed Method	MSE	0.000497	0.0223	33.03	0.9441
Proposed Method	Perceptual + MSE	0.00740	0.0231	32.41	0.9553

Methodology, Supervision, Writing - review & editing. **Reza Afrouzian:** Conceptualization, Methodology, Validation, Writing - review & editing.

Declaration of Competing Interest

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

References

[1] Kennedy JA, Israel O, Frenkel A, Bar-Shalom R, Azhari H. Super-resolution in PET imaging. IEEE Trans. Med. Imaging 2006;25(2):137–47.

[2] Yang C-Y, Ma C, Yang M-H. Single-image super-resolution: a benchmark. In: European Conference on Computer Vision. Springer; 2014. p. 372–86.

[3] Yang J, Wright J, Huang TS, Ma Y. Image super-resolution via sparse representation. IEEE Trans. Image Process. 2010;19(11):2861–73.

[4] Dong C, Loy CC, He K, Tang X. Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. 2015;38(2):295–307.

[5] Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 1646–54.

[6] Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Shi W, Wang Z. In: Photo-realistic single image super-resolution using a generative adversarial network. CVPR; 2017. p. 4681–90.

[7] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–8.

[8] Lim B, Son S, Kim H, Nah S, Mu Lee K. Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops; 2017. p. 136–44.

[9] Kang J, Gao Y, Shi F, Lalush DS, Lin W, Shen D. Prediction of standard-dose brain PET image by using MRI and low-dose brain [18F] FDG PET images. Med. Phys. 2015;42(9):5301–9.

[10] Wang Y, Zhang P, An L, Ma G, Kang J, Shi F, Wu X, Zhou J, Lalush DS, Shen D, Lin W. Predicting standard-dose PET image from low-dose PET and multimodal MR images using mapping-based sparse representation. Phys. Med. Biol. 2016;61(2):791.

[11] Xiang L, Qiao Y, Nie D, An L, Lin W, Wang Q, Shen D. Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. Neurocomputing 2017;267:406–16.

[12] Xu, J., Gong, E., Pauly, J., & Zaharchuk, G., 200x low-dose PET reconstruction using deep learning. arXiv preprint arXiv:1712.04119, 2017.

[13] Song T-A, Yang F, Chowdhury SR, Kim K, Johnson KA, El Fakhri G, Li Q, Dutta J. PET image deblurring and super-resolution with an MR-based joint entropy prior. IEEE Trans. Comput. Imaging 2019;5(4):530–9.

[14] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer; 2015. p. 234–41.

[15] Song T-A, Chowdhury SR, Yang F, Dutta J. Super-resolution PET imaging using convolutional neural networks. IEEE Trans. Comput. Imaging 2020;6:518–28.

[16] Hong X, Zan Y, Weng F, Tao W, Peng Q, Huang Q. Enhancing the image quality via transferred deep residual learning of coarse PET sinograms. IEEE Trans. Med. Imaging 2018;37(10):2322–32.

[17] Malczewski K. Super-Resolution with compressively sensed MR/PET signals at its input. Inform. Med. Unlocked 2020;18:100302.

[18] Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML-10); 2010.

[19] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. ICML 2015.

[20] Zeng K, Zheng H, Cai C, Yang Y, Zhang K, Chen Z. Simultaneous single-and multi-contrast super-resolution for brain MRI images based on a convolutional neural network. Comput. Biol. Med. 2018;99:133–41.

[21] Cocosco CA, Kollokian V, Kwan RK-S, Pike GB, Evans AC. Brainweb: Online interface to a 3D MRI simulated brain database. In: NeuroImage. Citeseer; 1997.

[22] Jr Jack, R. C, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, Whitwell JL, Ward C, Dale AM. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magn. Reson. Imaging: Offi. J. Int. Soc. Magn. Reson. Med. 2008;27(4):685–91.

[23] Hu X, Naiel MA, Wong A, Lamm M, Fieguth P. RUNet: a robust unet architecture for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; 2019. p. 1–3.

[24] Shi W, Caballero J, Huszár F, Totz J, Aitken AP, Bishop R, Rueckert D, Wang Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 1847–83.

[25] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision. Springer; 2016. p. 649–711.

Farnaz Garehdaghi received the M.sc degree in biomedical engineering from University of Tabriz, Tabriz, Iran in 2020. She is currently a Ph.D. student at Faculty of Electrical and Computer Engineering in University of Tabriz, Tabriz, Iran. Her current research interests include deep learning and its applications in medical image processing.

Saeed Meshgini received the Ph.D. in electrical engineering from University of Tabriz, Tabriz, Iran in 2013. He is currently an assistant professor in the Faculty of Electrical and Computer Engineering at University of Tabriz, Tabriz, Iran. His research interests include digital signal processing, machine learning, deep learning and pattern recognition.

Reza Afrouzian received his Ph.D. degree in electrical engineering from University of Tabriz, Tabriz, Iran in 2015. He is currently an assistant professor of Miyaneh Faculty of Engineering in University of Tabriz, Iran. His research interests include image processing, computer vision, pattern recognition and machine learning.