



OPEN Super-resolution of 3D medical images by generative adversarial networks with long and short-term memory and attention

Qiong Zhang[✉], Yiliu Hang, Fang Wu, Shentao Wang & Yue Hong

Since 3D medical imaging data is a string of sequential images, there is a strong correlation between consecutive images. Deep convolutional networks perform well in extracting spatial features, but are less capable for processing sequence data compared to recurrent convolutional networks. Therefore, we propose a long short-term memory and attention based generative adversarial network (LSTMAGAN) to realize super-resolution reconstruction of 3D medical image. Firstly, we use generative adversarial networks as the base model for super-resolution image reconstruction. Secondly, a long and short-term memory network, which specializes in dealing with long-term dependencies in sequential data, was used to process continuous sequential data of 3D medical images based on its ability to remember and forget information efficiently. Next, an attention gate is used to suppress the background noise information and improve the clarity of image features. Finally, the method proposed in this paper is applied on the Luna16 and BraTs2021 datasets. The experimental results show that the proposed method improves the PSNR and SSIM evaluation indexes compared with other comparative methods, respectively. Therefore, it can prove the advancement and effectiveness of the proposed method.

Keywords Generative adversarial networks, Long and short-term memory networks, Attention gate, 3D medical image super-resolution

Medical image data serves as an important aid for clinicians in the diagnosis of diseases and provides important assistance in the assessment, diagnosis and treatment of diseases. However, medical image data has problems such as small lesion features and fuzzy lesion boundaries, which affect the accurate judgment of clinicians. Super-resolution technology can be effectively utilized to address the aforementioned issues¹. Super-resolution technology refers to the technology of obtaining high resolution by increasing the resolution of low-resolution digital image space, which is mainly divided into methods based on deep convolution, wavelet transform and generative adversarial network. Li G et al.², Kang Let al.³ proposed feature extraction of low-resolution image by deep convolutional network and then mapped to super-resolution image size by up-sampling to generate super-resolution image, but this method is prone to lead to the problem of checkerboard effect⁴ in 2D image. Yu Y et al.⁵, Kim H et al.⁶, Hsu W Y et al.⁷, Wang Q et al.⁸ proposed the method of extracting high and low frequency information by wavelet decomposition of low-resolution images and then generating high-resolution images through interpolation and reconstruction process. However, this method requires the selection of wavelet bases suitable for a particular application, and different wavelet bases have a significant effect on the reconstruction results. Guerreiro J et al.⁹, Güven S A et al.¹⁰, Xiao Y et al.¹¹, Guo K et al.¹², Du W et al.¹³ proposed a generative adversarial network which generates super-resolution images from low-resolution images by generator. And the super-resolution image is compared with the real high-resolution image in a discriminator, which determines the super-resolution image closer to the real one. This method generates more realistic detailed features and can handle discrete image data better, but the GAN often suffers from the vanishing/exploding gradient problem, which requires a gradient activation^{14,15} function, and has limited effect on the super-resolution of 3D medical images.

Based on the above problems, this paper proposes a generative adversarial network based on long and short-term memory and attention to realize a super-resolution reconstruction method for 3D medical image. The main contributions of this paper are as follows:

College of Yonyou Digital & Intelligence, Nantong Institute of Technology, Nantong, China. ✉email: zhangq@ntit.edu.cn

(1) We propose a new network structure for super-resolution reconstruction of 3D medical image data, which makes full use of global features and improves the realism of the details of the generated images.

(2) Based on the sequence characteristics of 3D medical images, we propose to introduce the long and short-term memory network into the discriminator structure of generative adversarial network, and replace the convolutional network by the long and short-term memory network to realize the accurate judgment of fake super-resolution images and real high-resolution images.

(3) We improve the generator network structure by incorporating an attention gate structure into the generator network, which is used to suppress the background noise information and improve the quality of the generated super-resolution images.

Related work

3D medical image data is the most important part of medical data, which involves medical data such as X-rays, CT scans, MRIs and other medical data used to assist clinicians in the diagnosis and treatment of diseases. Most of the 3D medical imaging data is stored in sequences, and clinicians use the effect of a particular frame to diagnose a patient's condition. However, this method is easily to neglecting the characteristics of early fine lesions, resulting in cases of missed diagnosis, and there is also a human subjective influence on the selection of images. To address this problem, research scholars have proposed to solve the problem by deep learning methods, which are mainly divided into three directions: lesion feature-based target detection methods¹⁶; image segmentation methods¹⁷ and image super-resolution methods¹⁸. Ragab M G et al.¹⁹ and Yeerjiang A et al.²⁰ list the current status of development of Yolo series of algorithms for detection of medical images, and attempts were made to detect lesions by Yolo series of algorithms but the accuracy of the target detection methods implemented by Yolo was not high. Siddique N et al.²¹, Weng W et al.²², Punn N S et al.²³ list the applications of U-Net and improved algorithms in medical image segmentation, but those algorithms are supervised learning and heavily rely on data labeling. Umirzakova S et al.²⁴, Ahmad W et al.²⁵, Kang L et al.²⁶ have used convolutional neural network algorithm²⁷ and generative adversarial network algorithm²⁸ to realize super-resolution reconstruction of medical images, which provides larger size and clearer medical images to assist clinicians in the diagnosis of diseases, but the performance of this method is poor in the extraction of global features for sequential data. Based on the aforementioned defects, we propose a generative adversarial network based on long and short-term memory and attention to realize a super-resolution reconstruction method for medical image data. The method suppresses the background noise information through the attention gate to improve the quality of super-resolution images generated by the generator, and at the same time extracts global features with the help of long and short-term memory network to discriminate the authenticity of real images from super-resolution images.

Methodologies

In order to better adapt to 3D medical image data, in this paper, this method uses generative adversarial network as the basic model, suppresses the background noise information by attention gate network, strengthens the global feature capability by long and short-term memory network instead of deep convolutional network, prevents the attenuation and explosion of the feature gradient, and improves the robustness of the model. According to the principle of generative adversarial networks²⁹, the generator generates false super-resolution images (Fake SR), which are judged by the discriminator to classify the false super-resolution images and real images. The generator then continues to optimize the network model based on this judgment, so the performance of the discriminator determines the effectiveness of generating super-resolution images.

Network structure

The network's based on generative adversarial networks, which are mainly divided into generators and discriminators. Integrate an attention gate into the generator network to enhance its feature fusion capabilities and mitigate noise interference. Incorporate the long and short-term memory model in the discriminator to improve the global feature extraction ability and reduce the gradient decay and gradient explosion problem of the feature gradient in the propagation process. The robustness of the network can be improved by these two improvement methods, and its overall network structure relationship is shown in Eq. (1).

$$\min_G \max_D V(D, G) = \mathbb{E}_{\Theta_{I^{HR}} \sim P_{train}(I^{HR})} [\log D(I^{HR})] + \mathbb{E}_{\Theta_{I^{LR}} \sim P_G(I^{LR})} [\log(1 - D(G(I^{LR})))] \quad (1)$$

Where D is the discriminator, G is the generator, I^{HR} is the high-resolution image, I^{LR} is the low-resolution image, $P_{train}(I^{HR})$ is the probability distribution of training set I^{HR} , $P_G(I^{LR})$ is the probability distribution of I^{LR} after generator processing, discriminator for maximum accuracy, generator for minimum error.

Generator

The performance of the generator, as a network structure in the generative adversarial network that generates super-resolution images from low-resolution images, determines the quality of the super-resolution images generated. In the process of generating a super-resolution image from a low-resolution image, since the low-resolution image contains a lot of background noise information that affects the quality of the super-resolution image generation, this paper uses the structure of an attention gate (AG³⁰) to process the image and suppress the background noise of the input image. The generator structure uses the basic structure of SRResNet³¹, as shown in Fig. 1. In the basic structure, this network incorporates the attention gate into the dense connectivity module, as shown in Fig. 2.

In order to better accomplish the feature fusion and suppress the background noise information, this network adopts the method of attention dense connectivity module. The features of different network layers are fused by



Fig. 1. Structure of the generator network.

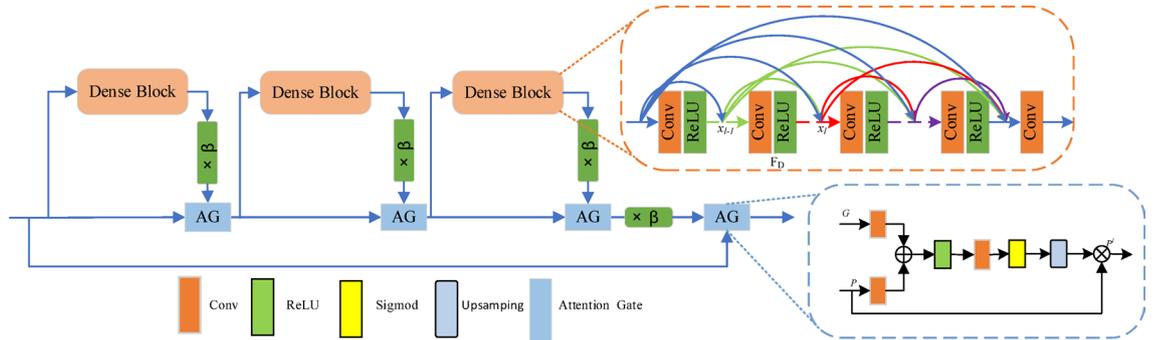


Fig. 2. Attention dense connectivity module (ADC), β is the scaling parameter and $\beta=0.2$ in the paper.

means of dense skip connections, and at the same time, in order to suppress the background noise information of the low-resolution image, the fused feature information and the initial feature information are inputted into the attention gate. The important target features are weighted according to the weight assignment, and the background noise information is suppressed. The attention dense connection module can effectively reduce noise interference and improve the quality of generated high-resolution images.

In the attention dense connection module, the input low resolution image I_{LR} is fed into the dense block, and the F_D represents the nonlinear structure of the current layer's dense connection, containing the Conv and ReLU functions. For the forward propagation network, x_{l-1} , as the output of the current layer $(l-1)^{th}$, is inputted into the next layer l^{th} can be expressed as: $F_D(x_{l-1})$. The dense connectivity module is that the l^{th} layer will concatenate the results obtained from all the previous layers, represented as shown in Eq. (2).

$$x_l = F_D(x_{l-1}) + x_0 + x_1 + \dots + x_{l-1}. \tag{2}$$

The output feature x_l of the dense module is multiplied by the coefficient β and the initial feature as input features G_i and P_i . Then, based on the input features G_i and P_i , after convolution by $1 \times 1 \times 1$, the corresponding position features are added. The negative features are suppressed by the rectified linear unit (ReLU) function, and the features are normalized by Sigmoid function. The processed results are multiplied with the initial feature P_i by channel to obtain the new feature P_i^l after the assigned weight. The process of operating on the input features G_i and P_i can be represented by Eqs. (3) and (4), respectively:

$$Q_{att}^l = \psi^T(\sigma_1(W_p^T P_i + W_g^T G_i + b_g)) + b_\psi \tag{3}$$

$$P_i^l = \sigma_2(Q_{att}^l(P_i, G_i; \Theta_{att})). \tag{4}$$

Where σ_1 is the ReLU function and σ_2 is the Sigmoid activation function, Parameter: $W_p \in \mathbb{R}^{F_l \times F_{int}}$, $W_g \in \mathbb{R}^{F_g \times F_{int}}$, $\psi \in \mathbb{R}^{F_{int} \times 1}$, Offset: $b_g \in \mathbb{R}^{F_{int}}$, $b_\psi \in \mathbb{R}$. \mathbb{R} represents the real number, F_l and F_g represents the matrix dimension.

Discriminator

The discriminator is the quality inspector in the generative adversarial network, according to the super-resolution image generated by the generator and the real high-resolution image comparison, the degree of recognition reaches the threshold is judged to be true, otherwise it is judged to be false. If the discrimination is false then optimization of the generator model is carried out until the image generated by the generator model can pass the discrimination of the discriminator. Thus, the capability of the discriminator determines the quality of the super-resolution image. According to the characteristics of 3D medical image data, 3D medical images are continuous sequential images and the correlation between images is stronger compared to other types of images. In this paper, long and short-term memory networks, which are more capable of processing continuous sequences, are used instead of deep convolutional networks to realize the discrimination of the generated images. Its network structure is shown in Fig. 3, in the structure, according to the input parameters of LSTM, like [batch_size, time_step, input_size]. We take 16 original images as a continuous sequence corresponding to batchsize=16,

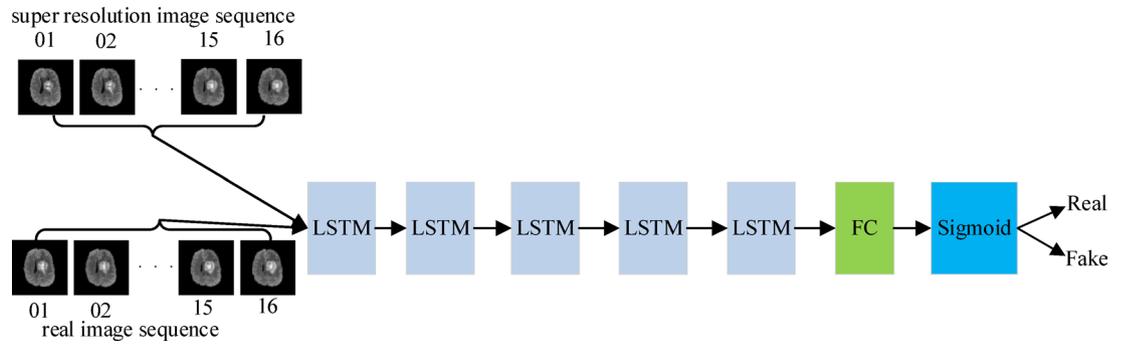


Fig. 3. Structure of the discriminator network, real images are used as training set and super-resolution images are used as test set, and the performance of the discriminator is judged by the final accuracy. LSTM is the long and short-term memory model, FC is the fully connected layer, and Sigmoid is the activation function.

time_step = 1 and input_size = image size, and input the sequence data into the LSTM of discriminator model to discriminate the classification effect of the data.

According to the transmission state of the cells in LSTM, the information of the precursor node is transmitted to the current node to avoid the problems of feature gradient disappearance and gradient explosion during long-term transmission. The LSTM network has four main components which are forget gate, input gate, update gate and output gate. According to the input sequence image features, it works by inputting the image data into the LSTM model according to the time sequence. The forget gate determines how much of the cell state from the previous moment needs to be retained until the current moment, and inputting the current training data x_t and the output value h_{t-1} of the previous moment into the forget gate as shown in Eq. (5).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f). \quad (5)$$

W_f is the weight matrix of the forget gate, $[h_{t-1}, x_t]$ means to splice two vectors into a longer vector, b_f is the bias term of the forget gate, and σ is the Sigmoid function. Information from the previous hidden state h_{t-1} and current input x_t are fed into the Sigmoid function at the same time, and the output value is between 0 and 1, with values closer to 0 meaning that they should be discarded, and closer to 1 meaning that they should be retained. The forget gate determines that the cell state h_{t-1} of the previous moment is partially preserved to the current moment, so the forget gate can partially preserve the state of the previous moment. The input gate determines how much of the input data at the current moment needs to be saved to the cell state, and the input gate is used to calculate the input state of cell based on the previous output and the current input, as shown in Eq. (6):

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i). \quad (6)$$

W_i is the weight matrix of the input gate, $[h_{t-1}, x_t]$ denotes the joining of the two vectors into a longer vector, b_i is the bias term of the input gate, and σ is the Sigmoid function. The update gate determines the update state of the cell at the current moment, and the cell state C_t at the current moment is calculated as shown in Eqs. (7) and (8):

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C). \quad (7)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (8)$$

\tilde{C}_t is the candidate memory cell state, W_C is the weight matrix of the update gate, and b_C is the bias term of the update gate. The cell state C_t at the current moment is the cumulative sum of the previous cell state C_{t-1} multiplied element-wise by the forgetting gate f_t , plus the current candidate memory cell state \tilde{C}_t multiplied element-wise by the input gate i_t . The candidate memory cell state \tilde{C}_t and the previous cell state C_{t-1} to be combined to form a new unitary state C_t . It preserves global feature information due to the control of the forget gate, and it prevents currently irrelevant content from entering memory due to the control of the input gate. And the output gate controls the output of the current state C_t to the output state h_t of the LSTM. The formulas are shown in (9) and (10):

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (9)$$

$$h_t = o_t * \tanh(C_t). \quad (10)$$

W_o is the weight matrix of the update gate and b_o is the bias term of the update gate. Finally, the output state h_t is classified by the results obtained after going through the fully connected layer and the Sigmoid function to determine the model performance of the discriminator.

Experiments Configurations

This experimental environment uses Ubuntu 20.04 Linux operating system, the development language is Python 3.10, the development environment is Stable Diffusion v1-5, the development environment configuration is CUDA 11.7, cuDNN 8, the development framework is Pytorch 2, and the development tool is JupyterLab. Computing power is NVIDIA Tesla T4 GPU – 16GB+ | 8 + TFlops SP CPU – 8 cores | Memory – 32GB.

Loss function

Loss function contains content loss function, perceptual loss function and adversarial loss function.

The content loss function is shown in Eq. (11)

$$L_{MSE}(I_{hr}, I_{sr}) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W [I_{hr}(i, j) - I_{sr}(i, j)]^2. \quad (11)$$

MSE is the mean square error, which is calculated by the squared value of the pixel difference at the corresponding position of the image, where I_{hr} is the real high resolution image and I_{sr} is the generated high resolution image, with an image size of $H \times W$.

Visual geometry group (VGG) based perceptual loss function is shown in Eq. (12)

$$L_{prec}(I_{hr}, I_{sr}) = \frac{1}{W_{x,y} \times H_{x,y}} \sum_{i=1}^{W_{x,y}} \sum_{j=1}^{H_{x,y}} (V_{x,y}(I_{hr}(i, j)) - V_{x,y}(I_{sr}(i, j)))^2. \quad (12)$$

The $V_{x,y}$ is indicated the feature map obtained by the y^{th} convolution (after activation) before the x^{th} max-pooling layer within the VGG19 network. And $W_{x,y}$ and $H_{x,y}$ are described the dimensions of the respective feature maps within the VGG network.

Adversarial loss function

To better prevent the gradient loss that occurs in the GAN, according to the principle of Wasserstein GAN³² loss function, an important trick of Wasserstein GAN (WGAN) is to shear the discriminator's ownership weights to a constant range [-c, c] in order to satisfy the conditions under which the Wasserstein distance can be deduced. However, with the clipping strategy, the weights of the discriminator tend to be either a minimum or a maximum, which causes the discriminator to behave like a binary network, weakening the nonlinear simulation ability of the GAN. Therefore, a gradient penalty³³ has been proposed to replace the shear operation. The new technique restricts the gradient of the discriminator from changing rapidly by adding a new term to the adversarial loss, so in this paper, the loss function of WGAN with gradient penalty (WGANGP) is used as the adversarial loss function. WGAN loss is shown in Eq. (13) and WGANGP loss is shown in Eq. (14).

$$L_{WGAN}(I_{lr}, I_{hr}) = \mathbb{E}_{I_{lr}}[D(G(I_{lr}))] - \mathbb{E}_{I_{hr}}[D(I_{hr})] \quad (13)$$

$$L_{WGANGP}(I_{lr}, I_{hr}) = L_{WGAN}(I_{lr}, I_{hr}) + \mathbb{E}_I[\|\nabla_I D(I)\|_p - 1]^2. \quad (14)$$

Therefore, in this paper, a hybrid loss function is adopted, as shown in Eq. (15).

$$L_{SR}(I_{lr}, I_{hr}, I_{sr}) = L_{MSE}(I_{hr}, I_{sr}) + 6 \times 10^{-3} \times L_{prec}(I_{hr}, I_{sr}) + 1 \times 10^{-3} \times L_{WGANGP}(I_{lr}, I_{hr}). \quad (15)$$

Experimental data

The dataset chosen for this experiment is 3D medical image data, but when the model is computed, the data is processed sequential 2D slice data.

Dataset 1³⁴ is the Luna16 dataset, which is a subset of the largest publicly available lung nodule dataset, LIDC-IDRI. This dataset is deployed by the National Cancer Institute and is widely used to detect and classify early lung cancer. It contains 1186 medical imaging data from low-dose lung CT images of 888 patients in mhd format. The data were processed to obtain 1186 images in png format. Among all image sizes, the smallest value is [300,300], the middle value is [360,360], and the largest value is [400,400], and randomly divide the LUNA16 dataset into training, testing and validation sets according to the ratio of 70%, 20% and 10%. The URL: <https://luna16.grand-challenge.org/Download/>.

Dataset 2³⁵ is the BraTs2021 dataset, which is a large-scale brain multimodal MR glioma segmentation dataset consisting of 8,160 MRI scans from 2,040 patients. Each patient contained MR images of the four modalities T1, T1Gd, T2 and T2-FLAIR. Since this paper only accomplishes the super-resolution reconstruction of 3D medical images, the T2-FLAIR modal data is selected, and the initial data type of this dataset is nii.gz. We convert nii.gz format to png format, a sample of nii.gz data can be divided into 155 png format images of 240 × 240 pixel. After data processing and cleaning, we keep 54,528 images and divide them into training set, testing set and validation set in the ratio of 70%, 20% and 10%. The URL: <https://www.synapse.org/#!Synapse:syn51514105>.

Evaluation indicators

There are two commonly used evaluation indicators for super resolution, the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM³⁶). The two evaluation indexes are the most basic ones used to measure the quality of compressed reconstructed images in super-resolution. Peak Signal-to-Noise Ratio(PSNR(db)) is defined as formula (16):

$$PSNR = 10 \times \log_{10} \frac{(MAX_I)^2}{MSE} = 20 \times \log_{10} \frac{MAX_I}{\sqrt{MSE}}. \quad (16)$$

MSE is the mean square error value. MAX_I is the maximum pixel value in image I . In this paper, the image is converted to YCbCr format, and then the PSNR of the Y component is calculated, and a larger value of PSNR indicates a better result.

Structural Similarity Index Measure (SSIM) assumes that the human visual system is highly adapted to extract image structures, and measures the structural similarity between images based on the comparisons of luminance, contrast, and structures. As shown in formula (17):

$$SSIM(x, y) = \frac{(2 \times \mu_x \times \mu_y + c_1) \times (2 \times \sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}. \quad (17)$$

where x, y denote two images, μ and σ^2 are the mean and variance, σ_{xy} is the covariance between x and y , and c_1 and c_2 are constant relaxation terms. $c_1 = (k_1 \times L)^2$, k_1 is a constant and default value is 0.01. $c_2 = (k_2 \times L)^2$, k_2 is a constant and default value is 0.03. L is the range of pixel values and value is $2^B - 1$.

Ablation studies

To further demonstrate the effectiveness of the proposed method, we use ablation studies to fuse different improved methods to further verify the effectiveness of the improved part. Model 1 is a GAN-based super-resolution method, Model 2 is a super-resolution method that fuses LSTM and GAN, Model 3 is a super-resolution method that fuses attention gate and GAN, and Model 4 is a super-resolution method that fuses LSTM, attention gate and GAN.

Through Fig. 4, we can analyze that Model 1 as the base model has higher variability between the images during the comparison of pixel differences between the generated super-resolution images and the real high-resolution images, and also the metrics of PSNR and SSIM are the lowest. Model 2 incorporates LSTM in generative adversarial networks for super-resolution of 3D medical images. Since LSTM has the ability of global feature extraction, it can better distinguish the real image features from the fake image features in the discriminator, so the model is more robust and more accurate in discriminating the real image from the fake image, so the image effect is better than model 1. Model 3 incorporating attention gate in generative adversarial networks, attention gate can better focus the feature information and suppress the background noise, so that better quality images can be generated in the generator. Model 4 is the method proposed in this paper, which incorporates attention gate in the generator and LSTM in the discriminator to fully utilize the advantages of attention gate and LSTM to enhance the super-resolution reconstruction of 3D medical images, and it is also the most effective method.

Comparison experiment

In order to further verify the performance of the proposed algorithm, we compare the proposed method with other super-resolution methods SRGAN²⁹, MedGAN¹², ESRGAN³⁷, TGAN¹³, DISGAN⁸ methods on dataset 1 and dataset 2, and the experimental results are shown in Figs. 5 and 6.

Through Figs. 5 and 6, we can see that the proposed method has better super-resolution reconstruction effect on Luna16 and BraTs2021 datasets. By image visualization, the proposed method generates clearer quality images with less background noise. Due to the fact that the proposed method uses the attention gate for the allocation of feature weights, suppresses the background noise information and highlights the high-frequency features, so that the generated image effect is of better quality. By comparing different methods on the dataset, we find that the proposed method obtains high evaluation indexes on both PNSR and SSIM.

For 3D medical image data, as it is stored in a continuous sequence, this paper uses LSTM method to replace convolutional neural network in feature extraction in the discriminator. To further validate the classification effectiveness of the discriminators, this paper applies different comparison methods on the Luna16 and BraTs2021 datasets. We measure the classification effect of the discriminator in terms of classification accuracy. Through Fig. 7, we can see that the classification accuracy of the discriminator in the proposed method is gradually improved in the iterative training process, and the highest accuracy value is achieved after 100 rounds of training. Compared to other comparative methods, our proposed LSTM-based discriminator is superior in classification accuracy.

Conclusion

In this paper, we propose a generative adversarial network based on long and short-term memory and attention (LSTMAGAN) for super-resolution of 3D medical images. The method is based on generative adversarial network to generate more realistic super-resolution images, and attention gate is added to the generator network to enhance the feature information and suppress the role of background noise. The LSTM is used as the backbone network in the discriminator network, which utilizes the sequential of 3D medical image data to extract global features and determine the image category. Although the proposed method achieves better experimental results

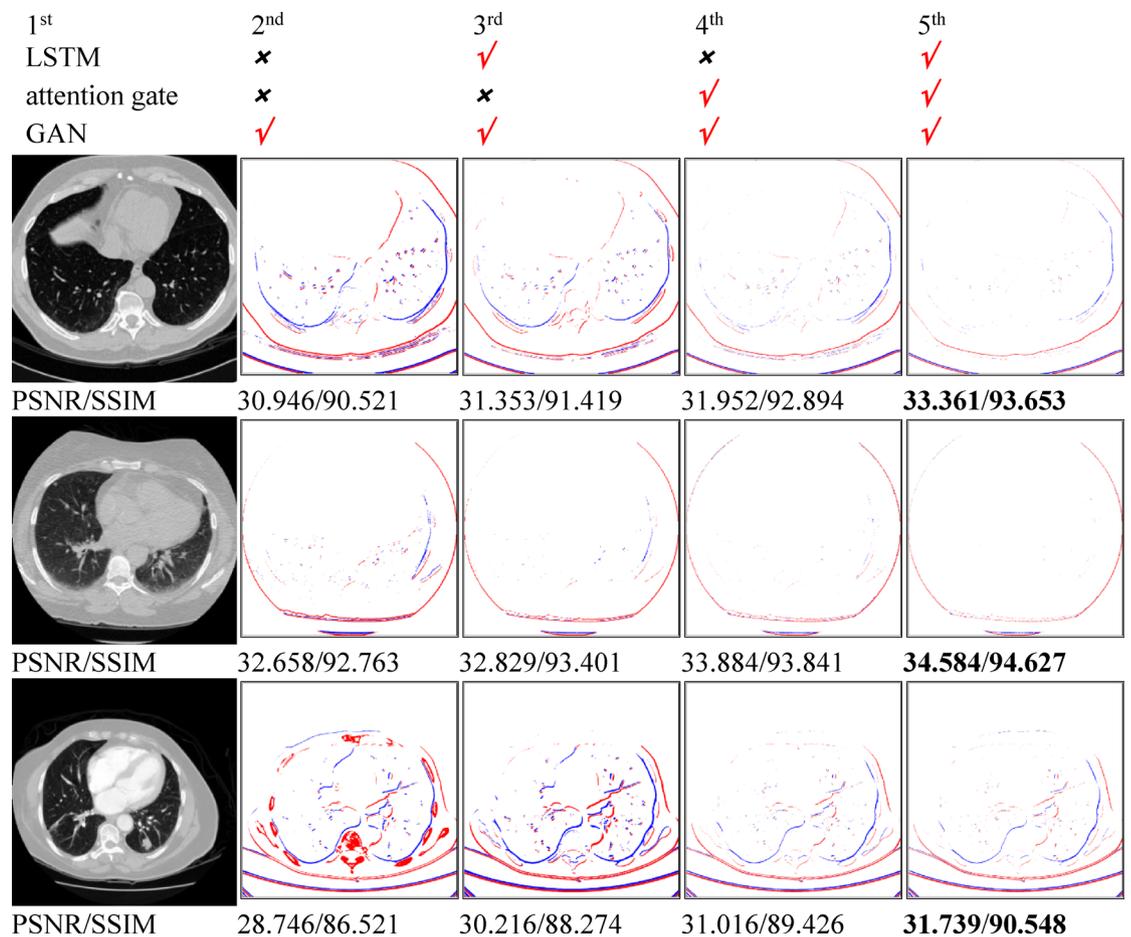


Fig. 4. Ablation studies effect diagram, where 1st indicates the real high-resolution image, and 2nd ~ 5th indicates the pixel difference between the super-resolution image and the real high-resolution image, with red color indicating the positive difference and blue color indicating the negative difference. The more pixel points in the 2nd ~ 5th diagrams, the greater the variability between the generated super-resolution image and the real image, and the poorer the quality of the generated super-resolution image.

on the Luna16 and BraTs2021 datasets, there are still some limitations, such as in the validation data to satisfy the homomorphic distribution with the training data, the inconsistency of probability resolution will lead to poorer quality of the generated images. Therefore, the multimodal super-resolution method will be our next research direction.

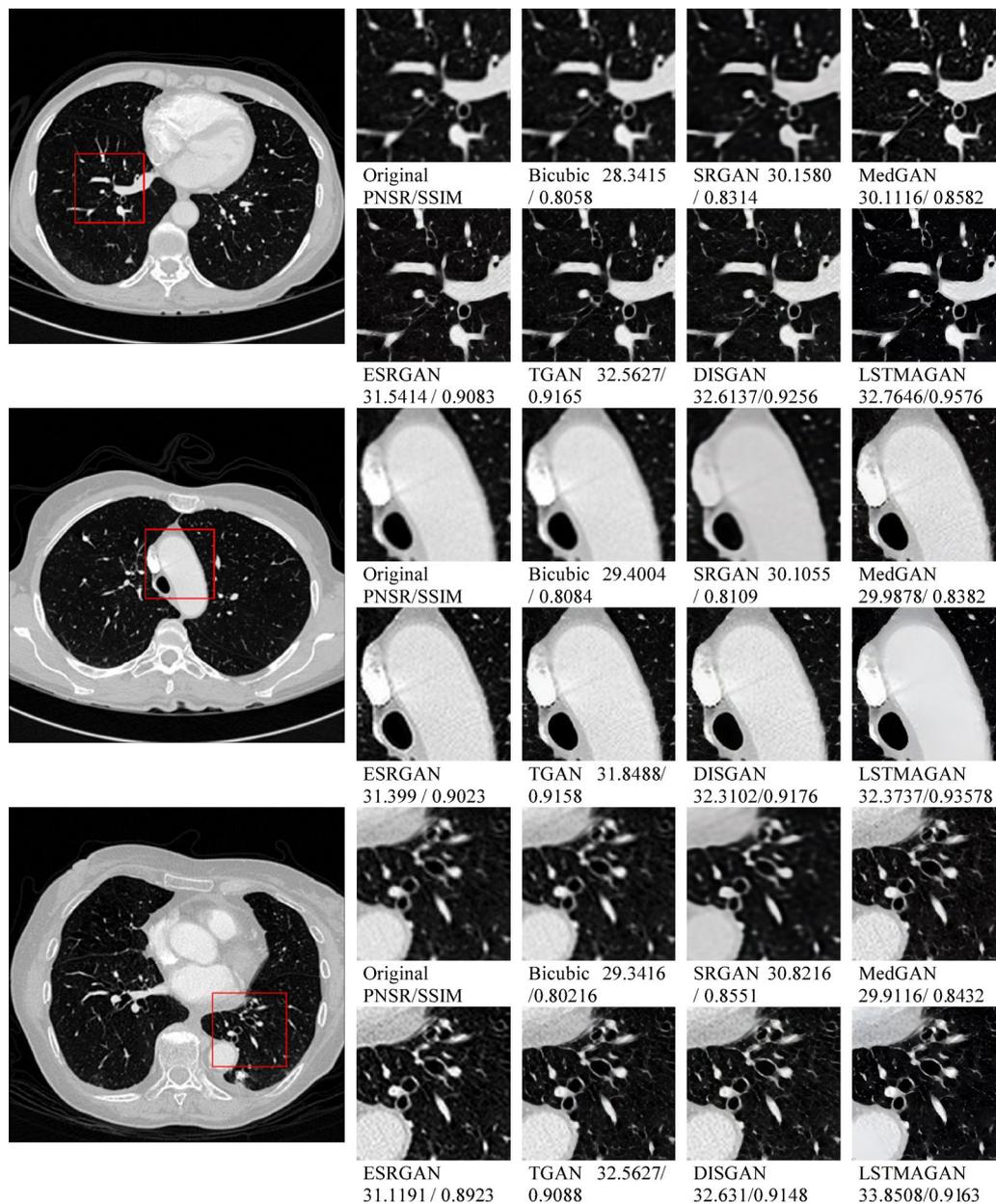


Fig. 5. Super-resolution reconstruction of different methods on Luna16 dataset.

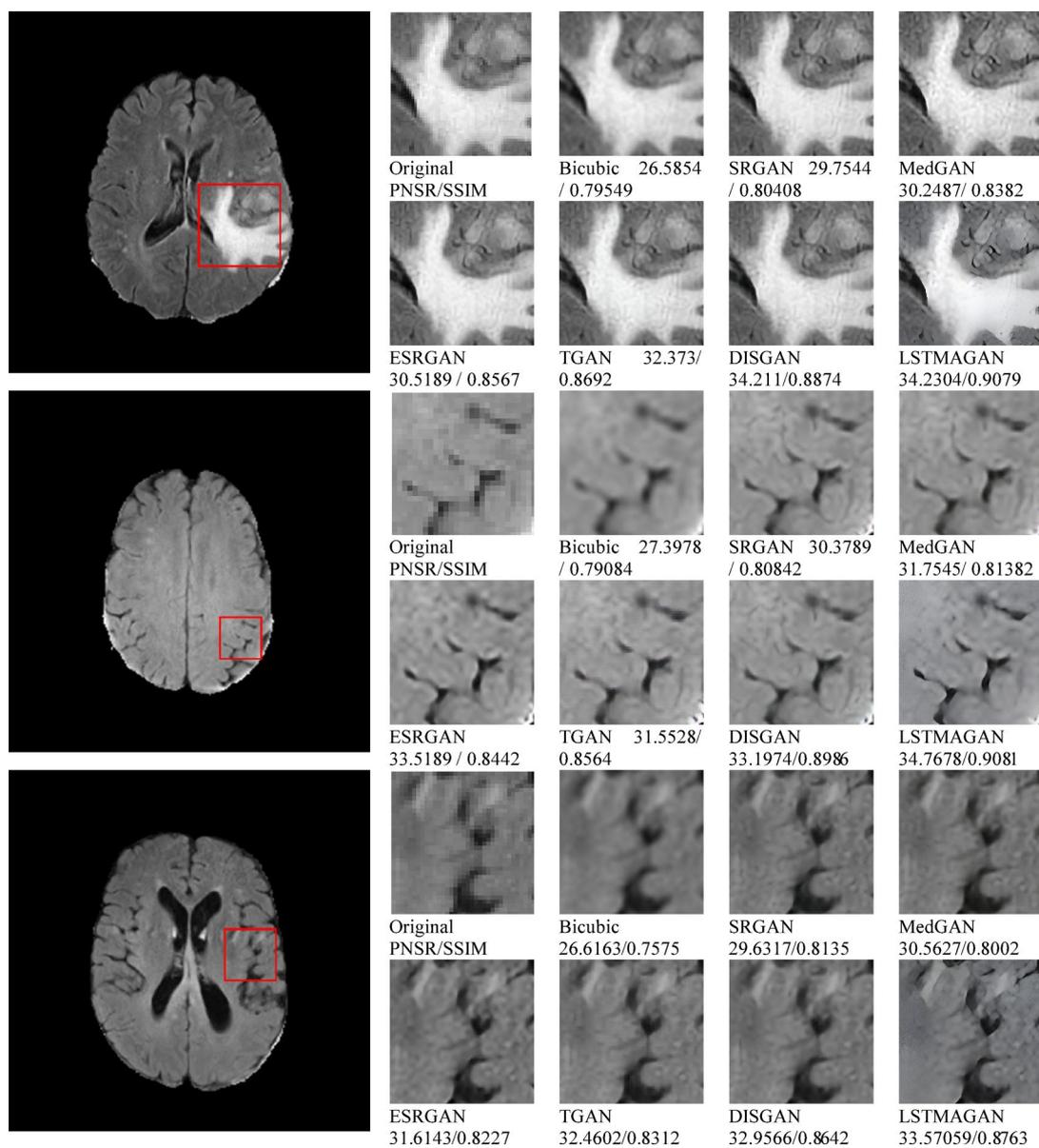


Fig. 6. Super-resolution of different methods on BraTs2021 dataset, we selected the FLAIR modal dataset as the data for this experiment.

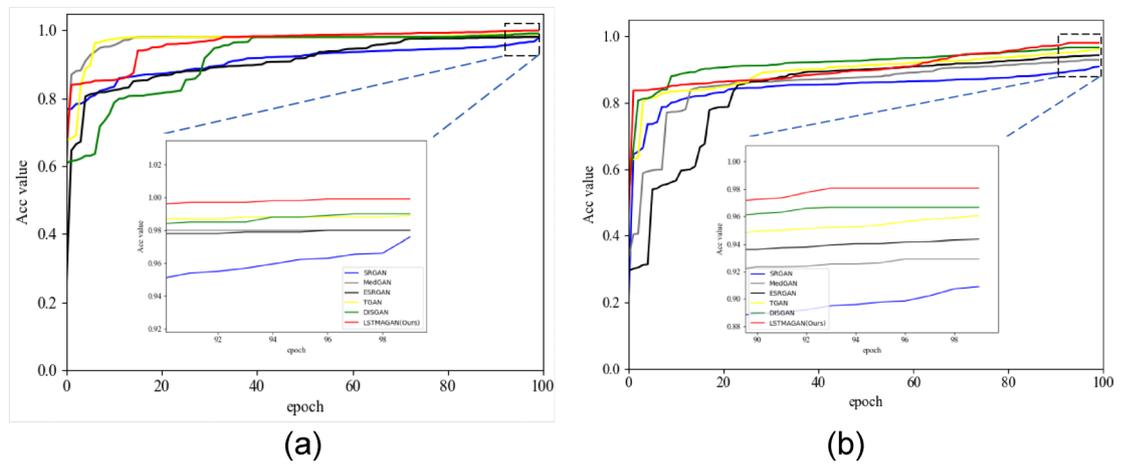


Fig. 7. Accuracy curves of discriminators of different methods on the dataset. **(a)** Accuracy curves of discriminators of different methods on Luna16 dataset. **(b)** Accuracy curves of discriminators of different methods on the BraTs2021 dataset.

Data availability

Data Availability Statement, as follows: The dataset1 download URL: <https://luna16.grand-challenge.org/Download/>. The dataset2 download URL: <https://www.synapse.org/#!Synapse:syn51514105>.

Received: 12 January 2025; Accepted: 4 June 2025

Published online: 01 July 2025

References

- Qiu, D., Cheng, Y. & Wang, X. *Medical Image super-resolution Reconstruction Algorithms Based on Deep Learning: A survey* **238**,107590 (Computer Methods and Programs in Biomedicine, 2023).
- Li, G. et al. Rethinking multi-contrast mri super-resolution: Rectangle-window cross-attention transformer and arbitrary-scale upsampling[C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 21230–21240 (2023).
- Kang, L. et al. *3d-mri super-resolution Reconstruction Using multi-modality Based on multi-resolution cnn* **248**,108110 (Computer Methods and Programs in Biomedicine, 2024).
- Odena, A., Dumoulin, V. & Olah, C. Deconvolution and checkerboard artifacts[J]. *Distill* **1** (10), e3 (2016).
- Yu, Y. et al. A super-resolution network for medical imaging via transformation analysis of wavelet multi-resolution[J]. *Neural Netw.* **166**, 162–173 (2023).
- Kim, H., Lee, H. & Lee, D. Deep learning-based computed tomographic image super-resolution via wavelet embedding[J]. *Radiat. Phys. Chem.* **205**, 110718 (2023).
- Hsu, W. Y. & Jian, P. W. Wavelet pyramid recurrent structure-preserving attention network for single image super-resolution[J]. *IEEE Trans. Neural Networks Learn. Syst.* **35** (11), 15772–15786 (2024).
- Wang, Q. et al. DISGAN: wavelet-informed discriminator guides GAN to MRI super-resolution with noise cleaning[C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2452–2461 (2023).
- Guerreiro, J. et al. *Super-resolution of Magnetic Resonance Images Using Generative Adversarial Networks* **108**,102280 (Computerized Medical Imaging and Graphics, 2023).
- Güven, S. A. & Talu, M. F. Brain MRI high resolution image creation and segmentation with the new GAN method[J]. *Biomed. Signal Process. Control.* **80**, 104246 (2023).
- Xiao, Y. et al. A novel hybrid generative adversarial network for CT and MRI super-resolution reconstruction[J]. *Phys. Med. Biol.* **68** (13), 135007 (2023).
- Guo, K. et al. MedGAN: an adaptive GAN approach for medical image generation[J]. *Comput. Biol. Med.* **163**, 107119 (2023).
- Du, W. & Tian, S. Transformer and GAN-based super-resolution reconstruction network for medical images[J]. *Tsinghua Sci. Technol.* **29** (1), 197–206 (2023).
- Liu, M. et al. Activated gradients for deep neural networks [J]. *IEEE Trans. Neural Networks Learn. Syst.* **34** (4), 2156–2168 (2021).
- Tang, Y., Liu, C. & Zhang, X. Single image super-resolution using Wasserstein generative adversarial network with gradient penalty[J]. *Pattern Recognit. Lett.* **163**, 32–39 (2022).
- Shoebi, A. et al. Applications of deep learning techniques for automated multiple sclerosis detection using magnetic resonance imaging: A review[J]. *Comput. Biol. Med.* **136**, 104697 (2021).
- Zhang, Y., Shen, Z. & Jiao, R. Segment anything model for medical image segmentation: current applications and future directions[J]. *Comput. Biol. Med.*:108238. (2024).
- Lepcha, D. C. et al. Image super-resolution: A comprehensive review, recent trends, challenges and applications[J]. *Inform. Fusion.* **91**, 230–260 (2023).
- Ragab, M. G. et al. A comprehensive systematic review of YOLO for medical object detection (2018 to 2023) [J]. *IEEE Access.* **12**, 57815–57836 (2024).
- Yeerjiang, A. et al. YOLOv1 to YOLOv10: A comprehensive review of YOLO variants and their application in medical image Detection[J]. *J. Artif. Intell. Pract.* **7** (3), 112–122 (2024).
- Siddique, N. et al. U-net and its variants for medical image segmentation: A review of theory and applications[J]. *IEEE Access.* **9**, 82031–82057 (2021).
- Weng, W. & Zhu, X. INet: convolutional networks for biomedical image segmentation[J]. *IEEE Access.* **9**, 16591–16603 (2021).
- Punn, N. S., Agarwal, S. & BT-Unet A self-supervised learning framework for biomedical image segmentation using Barlow twins with U-net models[J]. *Mach. Learn.* **111** (12), 4585–4600 (2022).

24. Umirzakova, S. et al. Medical image super-resolution for smart healthcare applications: A comprehensive survey[J]. *Inform. Fusion*. **103**, 102075 (2024).
25. Ahmad, W. et al. A new generative adversarial network for medical images super resolution[J]. *Sci. Rep.* **12** (1), 9533 (2022).
26. Kang, L. et al. Super-resolution method for MR images based on multi-resolution CNN[J]. *Biomed. Signal Process. Control*. **72**, 103372 (2022).
27. Kshatri, S. S. & Singh, D. Convolutional neural network in medical image analysis: a review[J]. *Arch. Comput. Methods Eng.* **30** (4), 2793–2810 (2023).
28. Zhou, T. et al. GAN review: models and medical image fusion applications[J]. *Inform. Fusion*. **91**, 134–148 (2023).
29. Goodfellow, I. et al. Generative adversarial nets[J]. *Adv. Neural. Inf. Process. Syst.* **27**, 1–9 (2014).
30. Oktay, O. et al. Attention U-Net: learning where to look for the Pancreas[C]. *Medical imaging with deep learning*. 1–10. (2022).
31. Ledig, C. et al. Photo-realistic single image super-resolution using a generative adversarial network[C]. *Proceedings of the IEEE conference on computer vision and pattern recognition*. 4681–4690 (2017).
32. Arjovsky, M., Chintala, S. & Bottou, L. Wasserstein generative adversarial networks[C]. *International conference on machine learning*. PMLR, 214–223 (2017).
33. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V. & Courville, A. C. Improved training of Wasserstein gans[J]. *Adv. Neural. Inf. Process. Syst.* **30**, 1–11 (2017).
34. Setio, A. A. A. et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge[J]. *Med. Image Anal.* **42**, 1–13 (2017).
35. Baid, U. et al. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification[J]. (2021). arxiv preprint arxiv:2107.02314.
36. Palubinskas, G. Image similarity/distance measures: what is really behind MSE and SSIM?[J]. *Int. J. Image Data Fusion*. **8** (1), 32–53 (2017).
37. Wang, X. et al. Esrgan: Enhanced super-resolution generative adversarial networks[C]. *Proceedings of the European conference on computer vision (ECCV) workshops*. 1–16 (2018).

Acknowledgements

This study was funded by the program of the Natural Science Research of Jiangsu Province Higher Education Institutions in 2023 under Grant No.23KJD520011, the General Program of Nantong Science and Technology Bureau in 2024 under Grant No. MSZ2024110, and the Program of the Nantong Institute of Technology Young and Middle-aged Backbone Teacher in 2022 under Grant No. ZQNGGJS202237 and ZQNGGJS202234. We would like to thank everyone who has contributed to this article. We would also like to thank the anonymous reviewers and editors for their helpful suggestions and comments.

Author contributions

Conceptualization, Z.Q. and Y.L.H.; methodology, Z.Q.; software, Z.Q. and F.W.; validation, Z.Q.; formal analysis, Y.H.; investigation, Z.Q.; resources, Y.H.; data curation, S.W.; writing—original draft preparation, Z.Q.; writing—review and editing, Z.Q.; visualization Y.L.H.; supervision, S.W.; project administration, Y.L.H.; All authors have read and agreed to the published version of the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Ethical approval

Luna16 and BraTs2021 belong to public databases. The patients involved in the database have obtained ethical approval. Users can download relevant data for free for research and publish relevant articles. Our study is based on open source data, so there are no ethical issues and other conflicts of interest.

Additional information

Correspondence and requests for materials should be addressed to Q.Z.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025