Liu *et al. BMC Medical Imaging* (2024) 24:259
https://doi.org/10.1186/s12880-024-01437-8

BMC Medical Imaging

RESEARCH **Open Access**

Dual stage MRI image restoration based on blind spot denoising and hybrid attention

Renfeng Liu[1], Songyan Xiao[1], Tianwei Liu[1], Fei Jiang[1], Cao Yuan[1*] and Jianfeng Chen[2*]

Abstract

Background Magnetic Resonance Imaging (MRI) is extensively utilized in clinical diagnostics and medical research, yet the imaging process is often compromised by noise interference. This noise arises from various sources, leading to a reduction in image quality and subsequently hindering the accurate interpretation of image details by clinicians. Traditional denoising methods typically assume that noise follows a Gaussian distribution, thereby neglecting the more complex noise types present in MRI images, such as Rician noise. As a result, denoising remains a challenging and practical task.

Method The main research work of this paper focuses on modifying mask information based on a global mask mapper. The mask mapper samples all blind spot pixels on the denoised image and maps them to the same channel. By incorporating perceptual loss, it utilizes all available information to improve performance while avoiding identity mapping. During the denoising process, the model may mistakenly remove some useful information as noise, resulting in a loss of detail in the denoised image. To address this issue, we train a generative adversarial network (GAN) with adaptive hybrid attention to restore the detailed information in the denoised MRI images.

Result The two-stage model NRAE shows an improvement of nearly 1.4 dB in PSNR and approximately 0.1 in SSIM on clinical datasets compared to other classic models. Specifically, compared to the baseline model, PSNR is increased by about 0.6 dB, and SSIM is only 0.015 lower. From a visual perspective, NRAE more effectively restores the details in the images, resulting in richer and clearer representation of image details.

Conclusion We have developed a deep learning-based two-stage model to address noise issues in medical MRI images. This method not only successfully reduces noise signals but also effectively restores anatomical details. The current results indicate that this is a promising approach. In future work, we plan to replace the current denoising network with more advanced models to further enhance performance.

Keywords Deep learning, Medical imaging, Blind spot denoising, Attention, GANs

Introduction

Medical imaging technologies are crucial for the diagnosis and treatment of patient conditions. However, certain imaging modalities, such as Computed Tomography (CT), Positron Emission Tomography (PET), and Single Photon Emission Computed Tomography (SPECT), expose patients to radiation, posing potential safety risks [1]. In contrast, Magnetic Resonance Imaging (MRI), as a non-invasive technique, is widely used in clinical diagnosis and medical research. However, MRI is often affected by noise during the imaging process. This noise can

*Correspondence:
Cao Yuan
yc@whpu.edu.cn
Jianfeng Chen
fengjianchen2006@163.com
[1] School of Mathematics and Computer Science, Wuhan Polytechnic University, Wuhan 430023, China
[2] Department of Cardiovascular Surgery, Zhongnan Hospital of Wuhan University, Wuhan 430071, China

originate from various sources, including hardware limitations, environmental interference, and patient movement. MRI noise not only reduces image quality and affects the ability of physicians to observe image details but also potentially leads to diagnostic errors and delays.

Denoising is a common research topic in both the medical imaging field and the broader medical community. Tomasi et al. [2] proposed the use of bilateral filtering for denoising and edge preservation. This practical filtering technique achieves effective image denoising and edge retention by considering both the spatial distance between pixels and the similarity of pixel values. However, bilateral filtering involves high computational complexity for each pixel, resulting in slow processing speeds for large-scale images. Additionally, it is highly sensitive to parameter selection, making the algorithm difficult to tune. To address the limitations of adaptability and other issues, Buades et al. [3] proposed a denoising algorithm based on non-local means (NLM). Unlike bilateral filtering, NLM takes into account the similarity between pixels across the entire image rather than just within a local neighborhood. It leverages the redundancy of neighboring pixels to eliminate noise, achieving better denoising performance. However, NLM's dependence on large-scale search leads to high storage requirements, and its performance can be affected when noise levels are high, indicating poor robustness to noise intensity. Dabov et al. [4] introduced a robust algorithm called BM3D, which utilizes block similarity within an image. It first finds similar blocks through block matching to construct a 3D matrix. After filtering and thresholding, the denoised 3D blocks are generated, and noise estimation and processing are performed using the information from neighboring blocks. However, BM3D also has some limitations; its denoising performance is better for Gaussian noise than for other types of noise, making it seem like a method specifically tailored for Gaussian noise.

You et al. [5] explored how deep learning techniques can be utilized to enhance the quality of MRI images. To improve the quality of denoised images, this field is primarily divided into supervised and self-supervised or unsupervised models. Zhang et al. [6] proposed a feedforward denoising neural network structure called DnCNN, which is a classical supervised denoising method. It introduces the concept of residual learning to achieve end-to-end learning without relying on specific image types or noise models. However, this model may lead to excessive smoothing and distortion in regions with strong noise or rich details, and it also requires a large amount of training data. Lehtinen et al. [7] proposed Noise2Noise (N2N), which performs image denoising without the need for noise-free reference images by learning the mapping from noisy images to noisy images. The concept behind this model is quite simple, but it is only suitable for specific noise distributions. If the noise type is already known, traditional denoising models can achieve comparable results. Additionally, this model does not consider spatial correlations, which limits its effectiveness when dealing with structured noise [8]. The aforementioned supervised denoising algorithms have high data requirements, and obtaining medical MRI images is limited. Collecting paired data and heavily relying on clean data is extremely challenging in practical medical applications. Self-supervised approaches are more suitable for the medical imaging field, and many self-supervised models have been proposed based on Noise2Noise. Fadnavis et al. [9] proposed Patch2Self, which is highly adaptable and capable of handling complex noise. It employs self-supervised learning by using different patches within the same image. The model learns the noise patterns of the image by comparing the relationships between different patches. It then uses the learned noise patterns to denoise the entire image without requiring clean data as labels. However, because it processes multiple patches within an image, it can be computationally intensive, especially when dealing with large, high-resolution images that require substantial computational resources. Kim et al. [10] proposed Noise2Score, a model that uses the Tweedie distribution for image noise modeling, which better fits actual noise distributions. Similar to other self-supervised models, this model does not require clean images as labels and learns the parameters of the noise distribution from noisy images through self-supervised learning. Blind-spot denoising is a mainstream approach, as seen in models like Noise2Void [11], Noise2Self [12], and DBSN [13]. These methods assume pixel independence and mask a portion of pixels, allowing the model to estimate and reconstruct clean pixels directly from neighboring noisy pixels. However, since some valuable information is lost due to the masked blind spots, it can affect the model's denoising capability. Huang et al. [14] proposed Neighbor2Neighbor, which constructs two similar sub-images for each noisy image using subsampling and pixel correction. It leverages neighborhood information to predict the values of missing pixels and reconstructs the original image by inferring denoising through a constructed loss function. This model extends the Noise2Noise framework into a new denoising paradigm, improving denoising capability and providing a framework for Blind2Unblind [15]. Blind2Unblind introduces a global mask mapper that enhances the design of blind-spot denoising networks by making the pixels at blind spots visible, thereby avoiding information loss caused by masking. It also provides a theoretical analysis of the re-visible loss, presenting both upper and lower bounds for loss convergence.

In recent years, Generative Adversarial Networks (GANs) have gained widespread application in the field of medical image processing due to their powerful feature learning and data generation capabilities, demonstrating significant advantages in tasks such as image denoising and super-resolution. Li et al. [16] proposed a novel image-domain artifact removal method based on GANs and variable constraints (MARGANVAC). This method introduces a time-varying cost function that relaxes fidelity constraints during the early stages of training and gradually strengthens them as training progresses, thereby significantly improving artifact removal performance. You et al. [17] proposed a semi-supervised model called GAN-CIRCLE, aimed at super-resolution tasks for MRI and CT images. The model enhances cross-domain consistency between source and target domains using residual networks and cycle consistency. Furthermore, You et al. [18, 19] developed an efficient Low-Dose CT (LDCT) image denoising neural network model by integrating CNN, residual learning, and Network in Network techniques. They also proposed a semi-supervised method for restoring CT image resolution, showing that these methods are suitable for various medical image denoising problems. Additionally, a three-dimensional denoising method called Structurally-Sensitive Multi-Scale Deep Neural Network [20] has achieved more effective noise suppression and structural information preservation by combining GANs. These studies collectively demonstrate the broad application prospects of GANs in medical image denoising.

We developed a new denoising and enhancement network, named Noise Reduction and Enhancement (NRAE), by integrating the Blind2Unblind framework with GANs. The core strategy of this network involves dividing each noisy image into several patches and randomly selecting specific pixels within each patch as “blind spots” to form a global mask. Both the noisy image and the masked version are fed into the denoising network for joint training. This design allows the network to utilize more contextual information during the denoising process, thereby preserving more image details. Specifically, for the edges and details within the image, we employ a perceptual loss function, which helps retain important high-frequency information and prevents detail loss due to excessive denoising. To address the issue of blurriness and incomplete local information in the denoised images, we leverage a GAN-based model that reconstructs image details through an encoder with two cascaded adaptive mixed-attention mechanisms. By introducing a cycle consistency loss, we ensure consistency in the images reconstructed by the generator in the reverse process. Through alternating optimization of the generator and discriminator, the NRAE network is capable of generating high-quality, detail-rich images from relatively low-quality images produced during the denoising process, thereby meeting the stringent image quality requirements of medical applications. In the comparative analysis with existing classical denoising techniques, this study utilized a real clinical dataset and also evaluated the model on the publicly available IXI Dataset. The results indicate that the proposed model outperforms existing methods in terms of denoising performance. Notably, the detail enhancement stage of the model significantly improves image quality, resulting in images that are much closer to the original true images compared to those that have undergone only denoising. This aspect is particularly important for clinical applications. The main contributions of this paper can be summarized as follows:

1. This paper proposes an improved self-supervised denoising model that effectively avoids the issue of identity mapping and enhances the clarity of denoised images by adjusting the shape of the masking units and introducing a perceptual loss function.

2. Combining generative adversarial networks (GANs), we propose an Adaptive Hybrid Attention Module (AHM) that focuses on pixel attention coefficients to preserve features and supplement and reconstruct missing pixel information.

3. Extensive evaluation on real clinical datasets demonstrates that compared to current mainstream and advanced denoising techniques, our proposed method exhibits significant performance improvements. This underscores its practical value and potential in clinical medical applications.

Related work

Blind spot denoising

Blind-spot denoising is a unique image processing method that does not rely on prior knowledge of noise types or require clean images as references. Instead, it trains a model to predict the unknown pixels at certain blind spots in the image. It is worth noting that a similar approach is also used in image completion tasks. However, there are three key differences between blind-spot denoising and image completion. First, their target tasks differ: image completion aims to predict the content of missing or occluded parts to generate visually coherent and realistic images. In contrast, blind-spot denoising aims to predict the true value based on surrounding pixel information without using the target pixel itself, thereby removing noise. Second, the nature of the masks used in each task is different. For example, in the image completion task, such as the MAE proposed by Kaiming et al. [21], large-scale masks are used to simulate missing parts of the image. Meanwhile, in denoising tasks, the masks are pixel-to-pixel, discrete, and spaced at certain intervals to ensure that contextual information is fully utilized during embedding. Finally, the content of the predictions also differs: image completion requires generating new content that is consistent with the context and conforms to the distribution of natural images, while denoising focuses on restoring noise-polluted pixels to their original true values, emphasizing the recovery of the original image details rather than creating new content.

Blind spot denoising algorithms, such as blind spot random masking or deletion [12, 22–24], use artificial masking strategies to create noisy pairs with blind spots. This masking approach is not optimal, as it prevents the denoising model from fully learning how to remove noise effectively, thus impairing the denoising performance. Fundamentally, blind spot-based denoising networks are affected by the network design or input data, leading to the loss of valuable information, which can reduce the upper limit of the denoising quality. Ideally, we want to learn how to remove noise directly from the original noisy image while preserving as much valuable information as possible. The key is that the model needs to be taught how to learn denoising effectively. The challenge in this process is the issue of "identity mapping", where the model might simply output the input image unchanged without effectively removing the noise. Wang et al. [15] proposed using a "masked input" strategy to block gradient updates to the model, preventing identity mapping. This masked input approach makes the pixels at blind spot locations visible, fully utilizing all available information to improve performance. Specifically, the method employs a dual-branch processing strategy. First, by introducing a mask, the noisy image is divided into multiple blind spot images, which are then input into the denoising network to allow precise prediction and denoising of the masked regions. The core of this step lies in achieving targeted denoising for specific regions through visually invisible blind spots. Second, a mask mapper is introduced to effectively aggregate the images processed by the blind spot network, generating a complete denoised image. Meanwhile, another branch directly inputs the noisy image into the denoising network. The role of this branch is to compensate for the information loss caused by blind spot processing, thereby enhancing the overall denoising effect. Notably, during the training process of this branch, no gradient backpropagation is performed to avoid identity mapping of the noisy image. This design enables the two branches to complement each other, achieving both targeted processing of specific areas and enhancing the clarity of the entire image, as illustrated in Fig. 1.

Attention mechanisms

Attention mechanisms have garnered widespread attention in the field of computer vision and have demonstrated excellent performance across various tasks. The core idea is to adaptively allocate resources to highlight important features while suppressing irrelevant information, thereby enhancing the model's representation capability and decision-making accuracy [25–28]. Attention mechanisms initially achieved remarkable success in natural language processing, especially with the introduction of attention-based models (such as Transformers [29]) in machine translation tasks.In image processing tasks, attention mechanisms are broadly categorized into channel attention mechanisms and spatial attention mechanisms. Channel attention mechanisms adjust the weights of different channels to emphasize important feature channels. For example, SE-Net [30] involves a channel attention mechanism that uses global average pooling and fully connected layers to learn the weight distribution of each channel, thereby improving image classification performance. On the other hand, spatial attention mechanisms capture important features in the spatial dimension, enhancing the model's focus on details. Attention mechanisms also play a significant role in medical image processing. CASTformer [31] is a 2D medical image segmentation model that combines GANs with attention mechanisms. The model uses a pyramid structure to build multi-scale representations and introduces a Perception Transformer module to more effectively learn discriminative regions with semantic structures. This approach overcomes common challenges faced by traditional Transformer models in medical imaging tasks, such as information loss due to single-scale representations and segmentation results lacking rich semantic

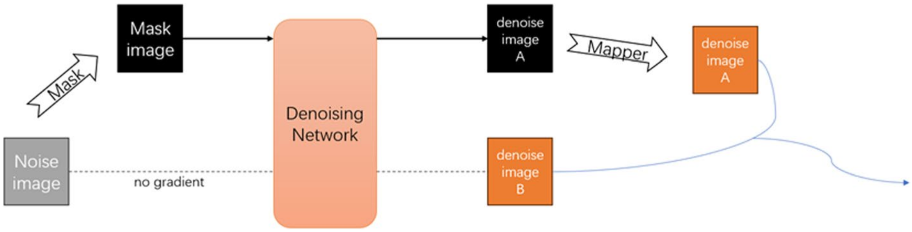

Fig. 1 Blind Spot Denoising Diagram

context and anatomical accuracy. By incorporating attention mechanisms, CASTformer demonstrates significant advantages in multi-scale feature fusion and fine segmentation, leading to substantial improvements in the performance and accuracy of medical image segmentation.

Hybrid attention mechanisms combine multiple types of attention mechanisms to further enhance model performance. Woo et al. [32] proposed the CBAM, a typical hybrid attention mechanism that sequentially processes channel and spatial attention in a cascaded manner, achieving better results in capturing important features. Cui et al. [33] introduced the Dual-Domain Strip Attention, composed of a Spatial Strip Attention (SSA) unit and a Frequency Strip Attention (FSA) unit. The SSA unit learns weights through convolution operations by leveraging contextual information from adjacent positions in the same row or column of each pixel, while the FSA unit refines features in the frequency domain using simple pooling techniques for frequency separation and modulation.Chen et al. [34] embedded spatial and channel hybrid attention modules in the latent layers of the network to capture contextual information of neighboring pixels and refine inter-channel features. They utilized an adaptive control module to dynamically fuse spatial and channel information, allowing the model to emphasize important features adaptively during the restoration process. Zhao et al. [35] proposed Pixel Attention for image super-resolution tasks, which adjusts the weight of each pixel locally and adaptively. Compared to traditional global attention mechanisms, this approach only performs weighting operations on local pixels, resulting in lower computational complexity.Chen et al. [36] argued that the potential of Transformers has not been fully exploited in existing networks. To activate more input pixels for better reconstruction, they proposed a new Hybrid Attention Transformer that combines both channel attention and window-based self-attention schemes. By leveraging the complementary strengths of both, it achieves a balance between global statistics and local fitting capabilities. To better aggregate cross-window information, they introduced an Overlapping Cross-Attention Module to enhance the interaction between adjacent window features. Zafar et al. [37] combined channel and pixel attention to propose a new efficient single-stage adaptive network for image restoration. The adaptive module robustly enhances spatial and contextual feature representations, significantly improving texture information and edge features.

Model ensemble

Model ensemble integrates multiple models in a specific sequence to form a more complex and functionally enriched model, emphasizing the complementarity between different models. This approach is not only theoretically innovative but also demonstrates excellent performance in practical applications. For example, as shown in Fig. 2, Model A is used for denoising, and Model B for enhancing image details. By first reducing noise levels and then enhancing image details, more accurate and clearer medical images can be obtained. Compared to a single model, this method is more effective in image quality processing and improvement, particularly in detail preservation and clarity enhancement.

To further enhance detail preservation and boundary refinement in images, You et al. [38] proposed a method named MORSE, which leverages the Mixture-of-Experts strategy. This method utilizes a stochastic gating mechanism to achieve parallel optimization of multi-scale pixel-level features and adaptively refines boundary regions. In addition, [39] embeds images into a shared content space to capture cross-domain shared feature information while preserving domain-specific appearance spaces. By utilizing the Wasserstein distance, this method better learns complete information representation, thereby improving image processing performance.

Wolterink JM et al. (2017) [40] first utilized a CNN to estimate standard-dose CT images from low-dose, noisy CT images, effectively reducing noise. Then, a well-trained generator of GANs was employed to transform low-dose images into standard-dose images, while a trained discriminator was used to distinguish between the output and real images. By incorporating adversarial network training on top of the CNN, this approach not only improved the similarity between the model's output images and standard-dose images but also validated the feasibility of model fusion for enhancing image quality in medical imaging tasks. To extract more detailed information from low-resolution images, [41] proposed using a GAN-based medical architecture. This architecture includes multi-scale shallow feature extraction, deep feature extraction using ResNet34, and gradual upscaling of feature maps, effectively preserving detail information. It demonstrated superior similarity compared to state-of-the-art models across various modalities of medical images. This also confirmed the excellence of the GAN-based

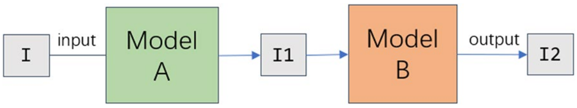

Fig. 2 Model ensemble

medical architecture in MRI image detail restoration tasks.

Methods

Network architecture

Building upon the research in [15], we propose NRAE, a denoising model specifically designed for MRI image denoising and enhancement. The detailed architecture is shown in Fig. 3. This section first provides an overview of NRAE, followed by an in-depth discussion of the model's denoising phase and detail enhancement phase.

NARE Overview Firstly,the input noisy image $y \in \mathbb{R}^{H \times W \times 3}$ undergoes processing where Ω_y and $h(\cdot)$ serve as the mask processing and mask mapper, respectively. These two components are used in conjunction. First, the noisy image is processed through the mask processing operation, and the processed image is then passed to the denoising network. The mask mapper is used to sample global blind spot pixels. Meanwhile, to fully leverage all available information and enhance model performance, another branch of the image y directly passes through the denoising network but does not participate in the backpropagation process. This prevents the denoising network from merely learning an identity mapping. By integrating this dual-branch structure, the problem of information loss in blind-spot networks is addressed, resulting in an effective blind-spot denoising strategy. Furthermore,the generator G consists of two recurrently connected networks,using the denoised images $\hat{y}$ and $\hat{y}^2$ as inputs. The discriminator D is guided by real clean data to train G to generate MRI images with reconstructed details that are indistinguishable from reality.

Image denoising

In Fig. 3, the image denoising part employs a dual-branch strategy. Specifically, the blind-spot denoising described in "Blind spot denoising" section is the core component of the denoising process. It mainly consists of the mask, mask mapper, and visible loss. Here, we will discuss the mask and mask mapper, while the visible loss is covered in "Denoising loss" section.

Mask/Mask mapper The network based on blind spot denoising suffers significant information loss during input or network transmission, which markedly reduces the denoising performance of the network due to the absence of valuable information. To overcome this issue, Blind2UnBlind introduces a globally masking mapping. The approach involves dividing each noisy image into patches and designating specific pixels within each patch as blind spots to create a global mask for input. These global masks are then fed into the network in batches. The global mask mapper samples noise

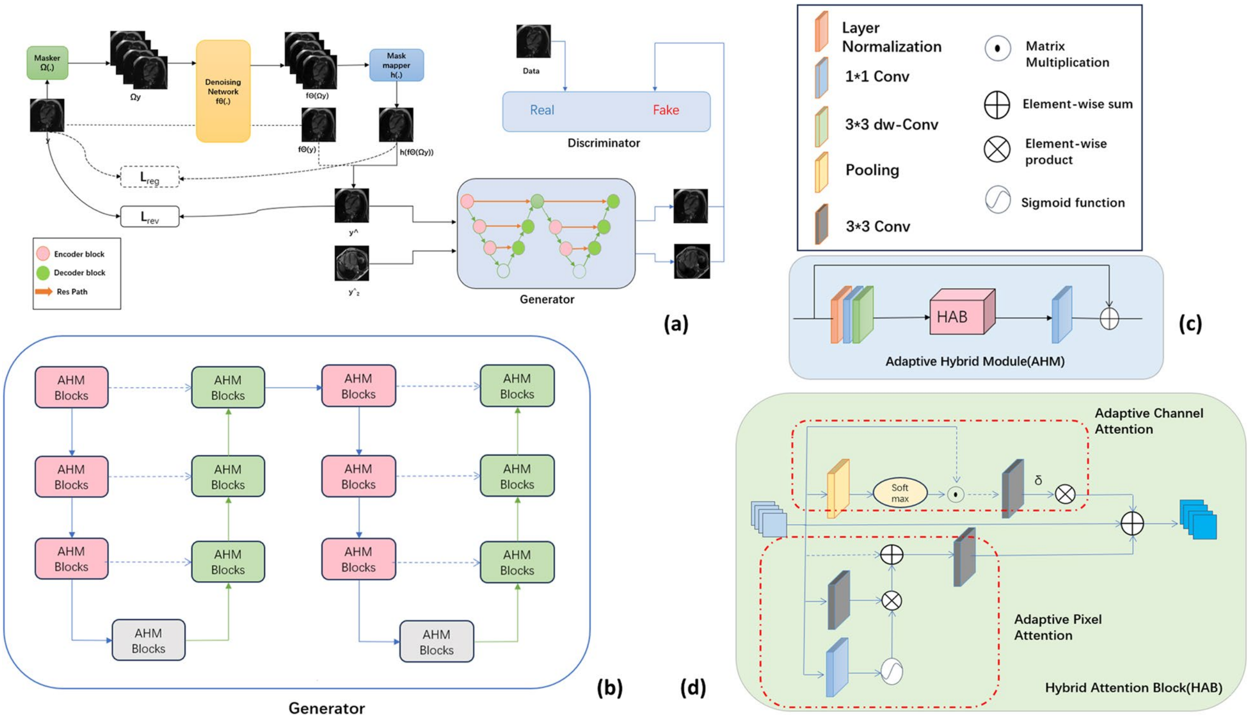

Fig. 3 Overview of the NRAE Architecture. **a** The structure of the overall model. **b** The structure of the generator used in the detail enhancement operation, featuring the Adaptive Hybrid Model (AHM). **c** The structure of the AHM, consisting of the Hybrid Attention Block (HAB) and convolution operations. **d** The HAB, which includes both channel attention and pixel attention mechanisms

at the blind spot locations and projects these samples onto a common plane for the denoising operation. To better leverage the information in the original image y and enhance denoising performance, we optimized the image partitioning strategy. Specifically, we increased the size of the grid from 2 × 2 to 3 × 3 and assigned different mask weights to each of these smaller grids. This adjustment allows for more comprehensive utilization of the information present in the original image. The detailed steps are as follows:

(1) The noisy image $y \in \mathbb{R}^{H\times W\times 3}$ is first divided into $\left[\frac{H}{s}\right] \times \left[\frac{W}{s}\right]$ blocks, where $s = 3$, with each small cell A being a 3×3 grid. The pixels in each cell A are masked in four directions (0°, 45°, 90°, 135°) with black representing masked pixels and different weights are assigned to the remaining pixels.

(2) After masking y, we obtain $\left[\frac{H}{s}\right] \times \left[\frac{W}{s}\right]$ images with blind spots, denoted as Ω_y. Ω_y are then fed into the denoising network f_Θ, resulting in multiple denoised outputs $f_\Theta(\Omega_y)$. In these denoised outputs, the gray areas represent the pixels corresponding to the blind spot locations before denoising. The task of the mask mapper $h(\cdot)$ is to extract all these gray pixels and combine them according to their relative positions in the original image to form a new image, $h(f_\Theta(\Omega_y))$.

(3) On the other hand, another branch directly passes the noisy image y through the denoising network f_Θ to obtain the denoised image $f_\Theta(y)$. During training, both $h(f_\Theta(\Omega_y))$ and $f_\Theta(y)$ are required.

For specific operations, refer to Fig. 4.

Detail enhancement

In Fig. 3, the detail enhancement part employs two networks stacked togethe, resembling the structure of UNet, as the generator G. Specifically, it samples low-resolution images to generate images with complete information. The discriminator D distinguishes between real images and images generated by G,adversarially training the model G and D until reaching a convergence balance. An illustrative diagram of the adversarial process is shown in Fig. 5.

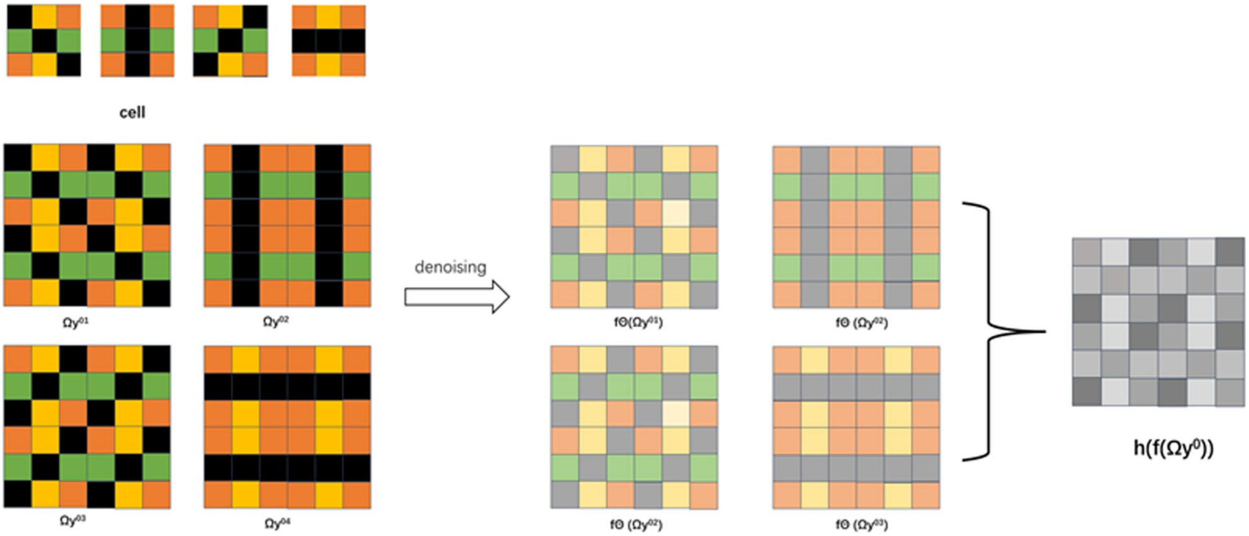

Fig. 4 The masker Ω_y^{ij} hides three points within each 3 × 3 cell,forming a set of nine masked cells Ω_y^{ij} for each image y_i (where i takes values from $\{1, 2, 3, 4\}$). These masked cells represent blind spots in the image. Subsequently, input them into a denoising network, the mask mapper $h(\cdot)$ samples $f_\theta(\Omega_y)$. During the sampling process, the mapper constructs the final denoised image unit $h(f_\theta(\Omega_y))$ based on the position and pixel values of each sampled point. This process enables the model to fully perceive and handle blind spots in the image

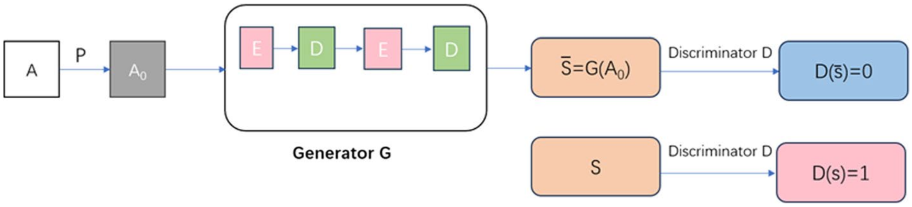

Fig. 5 A_0 is the low-resolution image obtained by blurring operation P from A

Generator It consists of a recurrent U-Net architecture, with each layer being an Adaptive Hybrid Module (AHM), as illustrated in Fig. 3. This module employs a combined channel and pixel attention mechanism. The pixel attention mechanism enhances the efficiency and accuracy of image feature extraction by balancing the processing of local details and global context through localized attention allocation. This significantly boosts the model's perceptual capability and computational efficiency. The channel attention mechanism focuses on adjusting the importance of different channels in the feature map, improving the model's focus on specific features by learning the weights for each channel. By integrating both attention mechanisms, the model retains distant pixel and channel information, effectively capturing both global and local features. The introduction of adaptive weights allows for the effective fusion of long-range channel features, addressing the limitations of traditional fixed attention mechanisms. Through pixel-level correspondence and adaptive weight allocation, the model enhances its sensitivity to edge details and texture variations.

Adaptive Hybrid Module (AHM) As shown in Fig. 3c, The image to be processed is first normalized. Then, a 1×1 convolution is used to extract low-level features F_1, which typically contain basic information and fine structures of the image. Next, a 3×3 depthwise separable convolution encodes the low-level features to capture local features while considering spatial context within the channel dimension. The Hybrid Attention Block (HAB) is then utilized to extract deep features F_2, which better represent the abstract characteristics and high-level structures of the image. These deep features are mapped into a residual image using a 1×1 convolution to aid in restoring the original image. As shown in Fig. 3b, the model employs skip connections to effectively combine features from different levels by element-wise addition, thereby efficiently merging encoder and decoder features. uring downsampling, a 3×3 convolution with a stride of 2 is used to reduce the image resolution. In the upsampling phase, a 1×1 convolution is employed to adjust the channel depth and enhance feature representation, followed by pixel shuffle operations. This rearranges the pixel information in the feature map channels to larger spatial dimensions, improving the feature representation capability.

Hybrid Attention Block (HAB) As shown in Fig. 3d, the Hybrid Attention Block (HAB) consists of two branches: adaptive channel attention and adaptive pixel attention.

As shown in Fig. 3d, the Hybrid Attention Block (HAB) consists of two branches: adaptive channel attention and adaptive pixel attention. In the channel attention branch, the input image first undergoes a pooling operation to reduce the spatial dimensions, encoding the global contextual information. After pooling, the dimensions become $C \times 1 \times 1$, where each dimension represents the information of all channels in the feature map. Next, we analyze the interdependencies between the convolutional feature maps. By using the softmax activation function, we calculate the importance weights for each channel. These weights reflect the contribution of each channel to the overall feature representation, thereby achieving integration of information across channels.

$$\mathrm{W(y{=}j)}=\frac{e^{A_j}}{\sum_{k=1}^{k} e^{A_k}} \tag{1}$$

In Eq. (1) represents the probability that vector A belongs to the jth category, where W dentes the channel attention weights.

$$\mathrm{CA(F)=F} \odot \sigma(\mathrm{F}_{C2}(\mathrm{ReLu}(\mathrm{F}_{c1}(\mathrm{Pool(F)})))) \tag{2}$$

In Eq. (2) represents the weighted feature map obtained after processing the input feature map F. Initially, the pooled vector is passed through the first fully connected layer, resulting in dimensions of $C_1 \times 1 \times 1$ where C_1 denotes the number of channels in the intermediate layer. The output is then subjected to the ReLU function, followed by the second fully connected lay C_2 and its output is further processed using the sigmoid function to obtain channel attention weights. These weights are element-wise multiplied with the original input features, enhancing each channel's features based on their importance and thereby capturing contextual information from distant channels. Next, the feature map F is reshaped into dimensions $C * HW$ to extract global features across the entire image spatial dimension. A 3×3 convolution is applied to the reshaped features to obtain output F_1 These features are multiplied with adaptive weights σ and integrated with pixel features. During training, the network dynamically learns the weight allocation mechanism. The output of the adaptive channel attention module represents the weighted combination of features across channels, allowing the network to focus on significant features while suppressing less relevant ones.

Traditional pixel attention mechanisms [35] capture pixel-level contextual features by calculating the attention weights for each pixel in the feature map. They typically use 1×1 convolutions and sigmoid functions to generate attention maps. However, due to the high redundancy of convolution operations in large-scale networks, the effectiveness of traditional pixel attention can be reduced. Adaptive pixel attention [37] introduces a method to capture contextual information more effectively by generating a three-dimensional matrix and using

1×1 convolutions to produce low-complexity attention features. Our adaptive pixel attention mechanism first encodes the feature map using a 3×3 convolution layer, simultaneously reducing dimensionality by halving the number of channels to lower computational complexity. The output from the 1×1 convolution layer is element-wise multiplied, and the sigmoid activation function is used to compute the attention weights for each pixel, allowing the network to focus on more important pixels. The attention information is reintegrated into the feature map through another 1×1 convolution layer. The attention-weighted feature map is then combined with the original feature map and further processed using a 3×3 convolution to achieve feature fusion. This process, coupled with residual connections that merge the results with the feature map from the previous layer, enhances the attention mechanism's effectiveness while retaining the original features.

$$F_{sum} = F + W_p(F) + \delta[W_1(\sum_{k=1}^{j}(\frac{e^{A_j}}{\sum_{k=1}^{k} e^{A_k}})F_k^j)] \quad (3)$$

In Eq. (3), where F represents the input feature map, W_p denotes the pixel attention output and W_1 signifies the 3×3 convolution output, with the adaptive weight δ effectively fusing the channel features.

Loss function

In this section, our proposed model covers two key stages:initial denoising inference and subsequent image detail reconstruction. Accordingly,the model's loss function is designed into two main parts: the first part is for the denoising loss in the initial stage, aimed at evaluating the effectiveness of the denoising process; the second part is for the reconstruction loss in the subsequent stage, focusing on the ability to recover detail information in the reconstructed image. Next, we will analyze in detail the composition and function of these two loss functions:

Denoising loss

Blind2UnBlind utilizes a blind spot structure for self-supervised denoising, followed by leveraging all information to enhance its performance. Since the visible loss is optimized for blind spots and visible points through a single back-propagatable variable, this optimization process is highly unstable. Therefore, they introduce a regularization term to constrain blind spots and stabilize the training process. The final visible loss is as follows:

$$\begin{aligned} L &= L_{\text{rev}} + \eta L_{\text{reg}} \\ L_{\text{reg}} &= \|h(f_\theta(\Omega_y)) - y\|_2^2 \\ L_{\text{rev}} &= \|h(f_\theta(\Omega_y)) + \lambda\hat{f}_\theta(y) - (\lambda+1)y\|_2^2 \end{aligned} \quad (4)$$

In Eq. (4), Ω_y represents the noise mask,$h()$ is the global perceptual mask mapper. To achieve visualization of blind spot pixels, $f_\Theta(y)$ does not participate in backpropagation, resulting in denoised original noise image, represented as $\Theta(y)$, which indirectly participates in gradient updates.

$$x = \frac{h(f_\theta(\Omega_y)) + \lambda\hat{f}_\theta(y)}{\lambda+1} \quad (5)$$

In Eq. (5), the weighted sum represents the optimal solution for X,where the upper and lower limits correspond to the denoising of the input original image using method $\hat{f}_\theta(y)$ and a denoising method similar to N2V [11] denoted as $h(f_\theta(\Omega_y))$,respectively.

Perceptual loss Research [42–45] has indicated that using perceptual loss to guide tasks such as image denoising helps preserve the original structural details of the image to enhance image quality. Inspired by this research,in this paper, a pretrained VGG-19 network (excluding the last three fully connected layers) is employed as the perceptual feature extractor. Five sets of feature maps are extracted at different levelsand finally combined into a multi-perceptual loss, encouraging the network to focus on perceptual visual quality for restoring noise-free images. It is represented as follows:

$$L_{PL}(y,\hat{y}) = \frac{1}{CHW}\sum_{i=1}^{5}\left\|\emptyset_i(y) - \emptyset_i(\hat{y})\right\|^2 \quad (6)$$

In Eq. (6), Θ_i represents the feature map obtained from block i, where CHW denotes the channels, height and width, respectively. y and $\hat{y}$ denote the ground truth image and the denoised result, respectively.

Overall, in the initial denoising inference stage, the model involves the following loss functions:

$$L_{detotal} = \lambda_1 L + \lambda_2 L_{PL} \quad (7)$$

Reconstruction loss

Generator G aims to reconstruct the image after denoising to restore detailed information, while the discriminator D is tasked with distinguishing between real MRI images and reconstructed images generated by the generator G. The system is optimized through adversarial training until it reaches a state of convergence balance. During this process,adversarial loss and consistency loss are two key components. Next, we will provide a detailed analysis and discussion of these two types of losses:

Adversarial Loss Our training aims to use the generator G to transform the denoised but incomplete image s_0 into a detailed and rich reconstruction image s. To achieve this goal, we introduce the discriminator D, which is tasked with distinguishing whether an image is generated by the generator G (considered fake) or directly reconstructed from s_0 (considered real). This process adopts a strategy based on deception and discrimination, continuously optimizing the generation ability of the generator G while enhancing the discriminator D's accuracy in judging the authenticity of images. During training, the parameters of the generator and discriminator are optimized using the following defined adversarial loss function:

$$\min_G \max_D V(D, G) = \underset{m\in M}{E}[\log D(s)] + \underset{s_0\in S}{E}[1 - \log D(G(s_0))] \tag{8}$$

In Eq. (8), $\log[1 - \log D(G(s_0))]$ encourages the generator G to generate more realistic images that are difficult for the discriminator D to distinguish. $\log D(s)$ enables the discriminator D to better distinguish between real images and fake images generated by the generator G. M is the set of real data and s is the set of images to be recovered and reconstructed.

Consistency Loss To strengthen the connection between s_0 and s, the consistency loss ensures that the images generated by the generator during the learning reconstruction in the cyclic structure, denoted as s_i (images between s_0 and s), maintain a high degree of visual consistency with the original image m. In the experimental setup, this objective is achieved by using distance measurement methods, particularly mean squared error (MSE). This method aims to minimize the difference between s_i and m, ensuring that the generator accurately reproduces the details and structural features of the original image, thereby improving the quality and authenticity of the reconstructed image.

$$L_{cyc}(G) = d(s[j], \bar{s}[j]) \tag{9}$$

In Eq. (9), the consistency loss, is specifically targeted towards the generator G. By making the images generated in each iteration of the loop in Fig. 5 similar to the original image m, it ensures that the model training adheres more closely to the desired criteria.

In summary, during the fine reconstruction stage, we involve two adversarial training operations aimed at minimizing the following losses as much as possible:

$$L_{entotal} = V(D, G) + \alpha L_{cyc}(G) \tag{10}$$

Experiments

Details

Training Details During the training process of our model for denoising in real-world environments, all models are trained using the same settings. Specifically,the batch size is set to 4, Adam optimizer is employed with weight decay set to 1e-8. The initial learning rate of the models is 0.0003 and every 20 training epochs, the learning rate is multiplied by 0.25 to facilitate better convergence. Additionally,images are randomly cropped into patches of size 128x128 pixels and masked as shown in Fig. 4. During the fine-tuning training stage, we set the hyperparameter α to 10, also utilizing the Adam optimizer with an initial learning rate of 1e-4 and the entire training process lasts for 500 epochs. The detailed architecture of the models is illustrated in Fig. 3. All experiments are conducted on a system equipped with NVIDIA Tesla A30 GPU, using Python 3.8.0 and PyTorch 1.2.1 environment for training.

Denoising Dataset In this study, we utilized NIFTI (Neuroimaging Informatics Technology Initiative) format datasets obtained from clinically real cardiac regions. The NIFTI format is a commonly used data format in multidimensional neuroimaging, capable of accurately reflecting metadata including directional information. We converted these multidimensional data into two-dimensional RGB images, with pixel sizes ranging from 256×256 to 512×512 for each image. Among these, 7600 clean images were used for training, while the validation and test sets were approximately divided in a ratio of 70:15:15. To obtain more accurate PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) metrics, we performed 10 calculations on the test set and took the average.

Table 1 Quantitative results of denoising on synthesized datasets. The highest PSNR/SSIM among denoising methods is highlighted in bold, while the second highest is underlined. The last column represents the number of network parameters

Method	PSNR/SSIM	Params (M)
Rician ($\sigma = 25$)		
BM3D	35.93 / 0.805	-
N2N	35.78 / 0.795	1.90
N2V	35.82 / 0.828	1.86
Patch2self	36.04 / 0.809	1.64
NBR2NBR	36.27 / 0.854	1.91
Blind2Unblind	36.54 / **0.908**	1.84
Ours	**37.13** / 0.893	2.07
Rician ($\sigma \in [5, 50]$)		
BM3D	35.69 / 0.837	-
N2N	35.50 / 0.784	1.90
N2V	35.59 / 0.811	1.86
Patch2self	35.87 / 0.872	1.64
NBR2NBR	35.96 / 0.887	1.91
Blind2Unblind	36.11 / **0.896**	1.84
Ours	**36.74** / 0.885	2.07

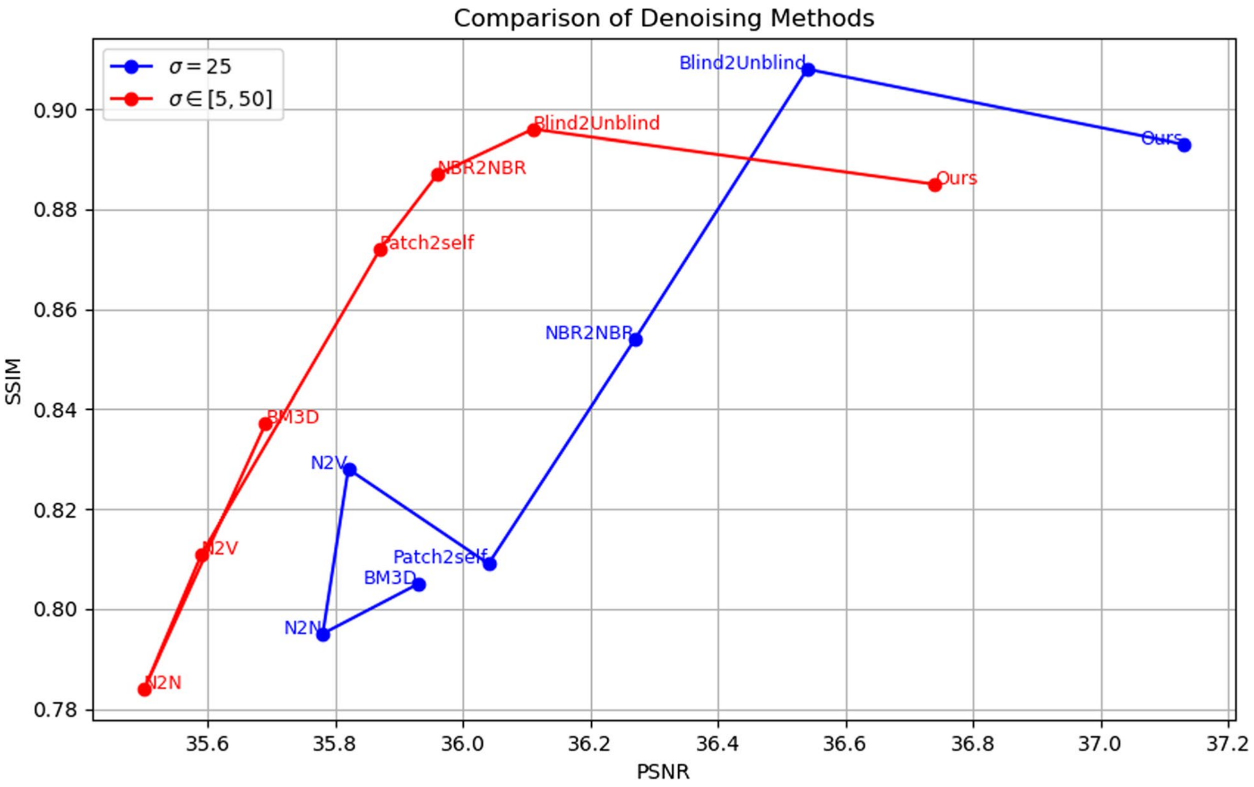

Fig. 6 Comparison of model performance metrics for different noise types

Additionally, this study utilizes the publicly available IXI Dataset, which is a multi-modal, diverse, and high-quality MRI dataset widely used in medical imaging research. We use the T1-weighted MRI brain images from this dataset for our experiments. The dataset comprises 581 sets of MRI images. These multidimensional data were sliced along the coronal plane from multiple angles to enhance the dataset. The remaining preprocessing steps are consistent with the methods described for the previous dataset.

Noise treatment In medical MRI images, the predominant type of noise is Rician noise, which differs from Gaussian noise. Rician noise is a signal-dependent noise, meaning its distribution is influenced not only by the hardware noise of the imaging device but also by the intensity of the original signal. Rician noise can be considered a type of noise formed by the amplitude operation of two independent Gaussian noise components. It describes the distribution characteristics of the amplitude of a complex-valued signal affected by noise. The calculation of Rician noise is given in Eq. (11)

$$R = \sqrt{(S+X)^2 + Y^2} \tag{11}$$

In Eq. (11), X and Y are independent Gaussian noise components with a standard deviation of σ, S is the actual amplitude of the signal, and R represents Rician noise. First, two independent Gaussian noise matrices of the same size as the image are generated. These matrices are then added to the real and imaginary parts of the image, respectively.

The noise settings for this experiment are as follows:

- (1) Rician noise generated with Gaussian noise of $\sigma = 25$;
- (2) Rician noise generated with Gaussian noise where $\sigma \in [5, 50]$.

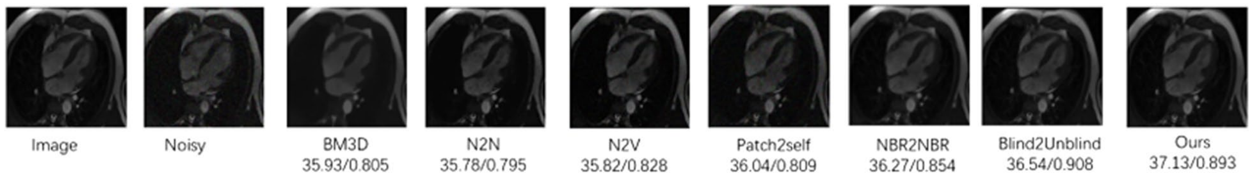

Fig. 7 Visual comparison of denoising MRI images under the setting of $\sigma = 25$

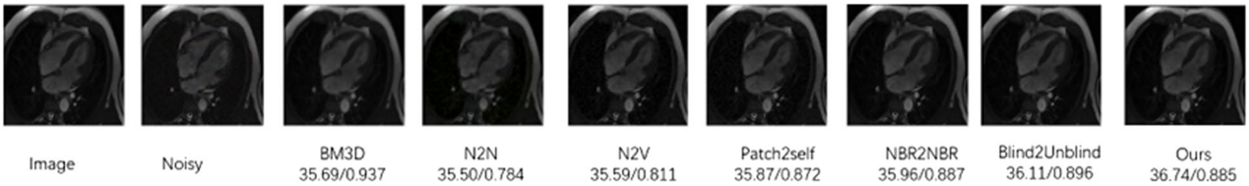

Fig. 8 Visual comparison of denoising MRI images under the setting of $\sigma \in [5, 50]$

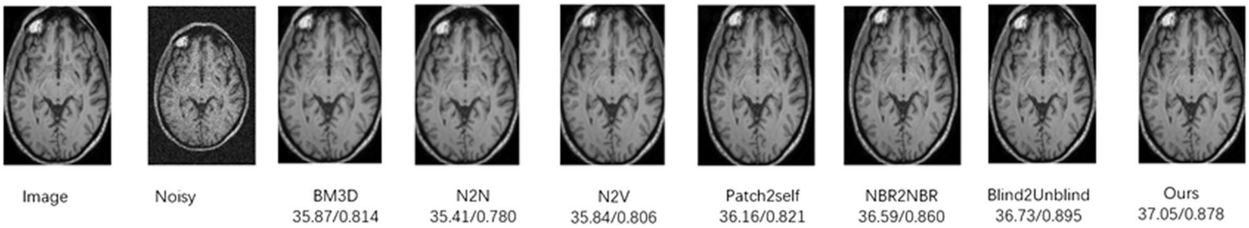

Fig. 9 In the IXI dataset, Visual comparison of denoising MRI images under the setting of $\sigma = 25$

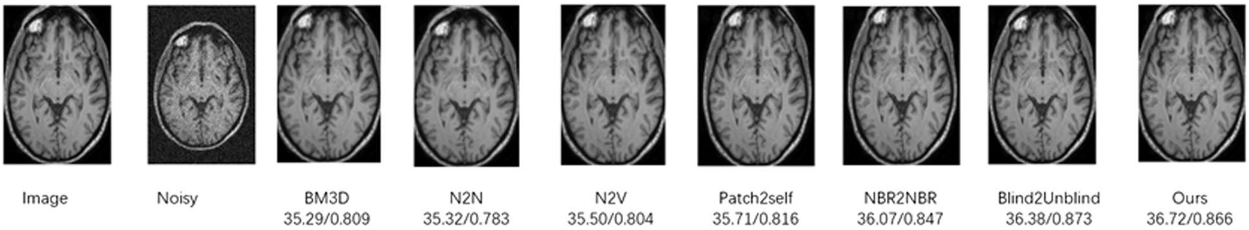

Fig. 10 In the IXI dataset, Visual comparison of denoising MRI images under the setting of $\sigma \in [5, 50]$

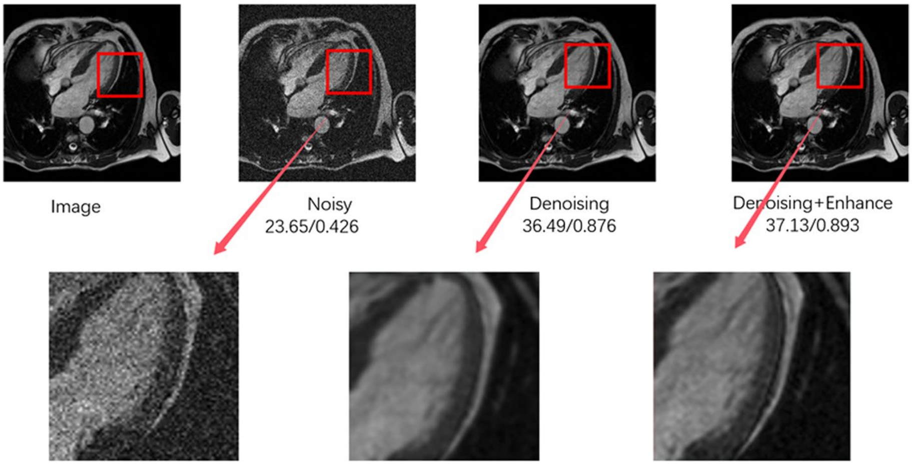

Fig. 11 Visual comparison of whether post-denoising processing with detail enhancement is performed under the setting of $\sigma = 25$

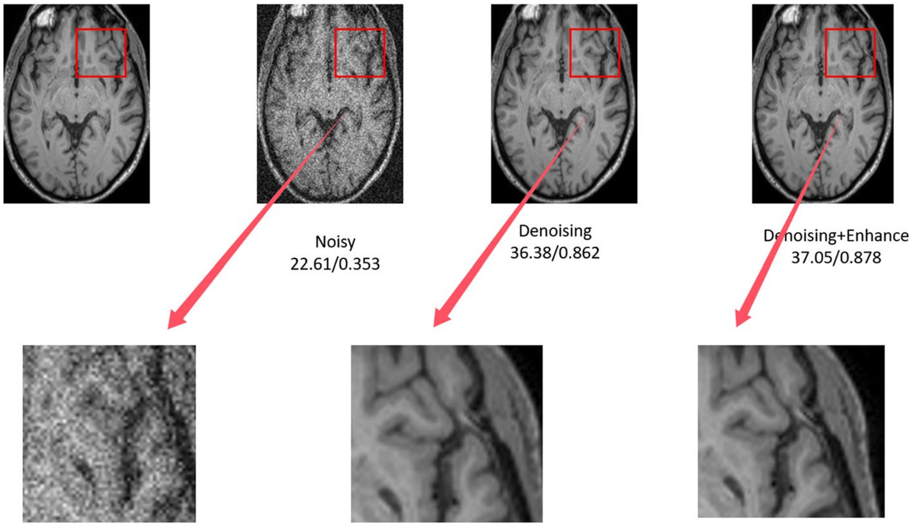

Fig. 12 In the IXI dataset, Visual comparison of whether post-denoising processing with detail enhancement is performed under the setting of $\sigma = 25$

Experimental Details To comprehensively evaluate the denoising method, we use PSNR and SSIM as performance metrics. For fairness, the method is compared with traditional denoising algorithms (such as BM3D [4]) as well as several self-supervised denoising algorithms (including N2N [7], N2V [11], Patch2self [9], NBR2NBR [14], and Blind2Unblind [15]). All comparison methods are implemented using their official versions and retrained on our dataset. Additionally, to provide a more thorough assessment, we also consider the performance of each method at different noise levels and their capability to handle specific medical image features, such as detail and structural preservation.

Results Comparison Quantitative comparisons of Rician noise denoising are shown in Table 1 and Fig. 6. Our denoising method generally outperforms several classic comparison methods for both fixed and variable noise levels. Figures 7, 8, 9, and 10 illustrate the denoising results for two MRI datasets under settings of $\sigma = 25$ and $\sigma \in [5, 50]$, respectively. Objectively, our method demonstrates strong denoising capability. Compared to the baseline model, our SSIM metric is slightly lower than the original model, but it is worth noting that the PSNR improves by approximately 0.6dB. Subjectively, the method offers better restoration of image details. Figures 11 and 12 show the comparison of information loss and recovery in specific detail areas during the denoising process, contrasting between simple denoising and subsequent detailed enhancement operations.

Table 2 Loss Function Ablation Study. LA, LB and LC represent Eqs. 4, 6 and 10, respectively. The highest PSNR/SSIM is highlighted in bold, while the second highest is underlined

Loss Type	$\sigma = 25$	$\sigma \in [5, 50]$
LA	36.59 / **0.901**	36.07 / **0.890**
LA + LB	<u>36.76</u> / 0.870	<u>36.39</u> / 0.853
LA + LB + LC	**37.13** / <u>0.893</u>	**36.74** / <u>0.885</u>

Ablation experiment

Loss Function Ablation Study Table 2 compares the PSNR and SSIM results obtained from training with three different loss functions under two noise-adding conditions. The third method, which incorporates perceptual loss and cyclic loss, achieves the best results. Figure 13 shows a comparison of image detail information obtained from training with the three different loss functions.

Ablation Study on Masking and Attention Table 3 presents the ablation comparison of different masking methods and the improved structure of the Adaptive Hybrid Module (AHM) during the detail enhancement phase.

Discussion

We have demonstrated that the dual-stage approach of NRAE effectively addresses the quality issues of MRI image denoising. However, when compared with classic denoising models such as BM3D, N2N, N2V, and NBR2NBR, we observed some significant differences.

Firstly, supervised models like Noise2Noise typically require paired datasets and have high demands for data scale, which may pose limitations in medical applications. In contrast, self-supervised models like Noise2Void, which use a blind-spot strategy, tend to lose some information, resulting in poorer denoising performance. Although Blind2Unblind effectively avoids the loss of

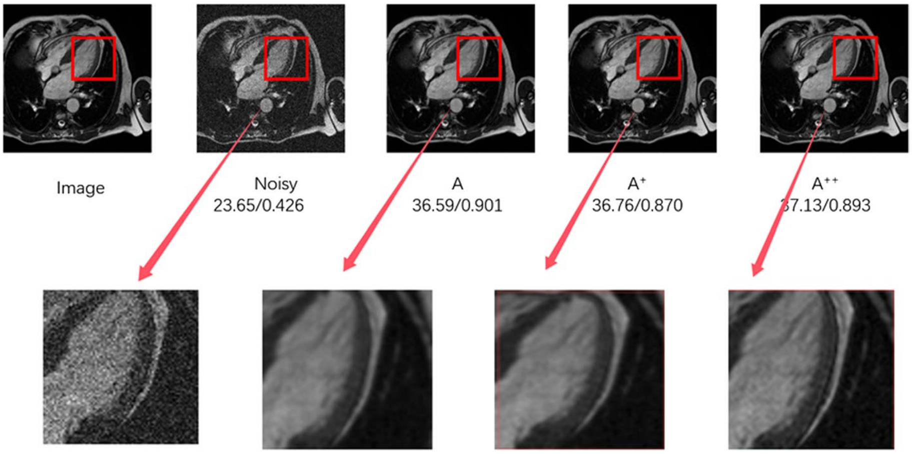

Fig. 13 Comparison of image details obtained from training with three different loss functions under the setting of $\lambda = 25$. Model A, A+ and A++ correspond to loss functions LA, LA+LB and LA+LB+LC, respectively

blind-spot information through masked mappers and dual-branch joint training, there is still room for improvement in detail restoration. In terms of denoising performance, the NRAE network outperforms classic methods like BM3D, which is mainly attributed to the superior denoising capability of our baseline model. For detail enhancement, NRAE combines the advantages of self-supervised denoising models and GANs, inheriting the strengths of the baseline model while leveraging GANs for detail reconstruction. This significantly improves image quality and detail richness during the denoising process.

In comparisons with models such as Blind2Unblind, we observed some performance differences, which may stem from the varying adaptability of the models to data characteristics. Medical MRI images have high demands for contrast and detail, so the NRAE model focuses more on detail information during the denoising process. This emphasis may lead to a decrease in the SSIM metric, suggesting that NRAE may retain some noise to better preserve image details during denoising.

However, we must also honestly acknowledge some shortcomings of the NRAE model. Firstly, our noise addition method may not fully simulate the complex noise characteristics of medical MRI images, especially in contrast to common image noise. Secondly, considering that MRI images come in various sequences applicable for examinations in different parts of the body, further validation of the model's generalization performance across different MRI sequences could be pursued in the future.Lastly, replacing the denoising network with more complex structures might contribute to enhancing the denoising performance of the model, which is one of the directions for our future research to further enhance the practicality and performance of the NRAE model.

Table 3 Comparison under the setting of $\sigma = 25$. Mask-1 refers to the method described in [15], while Mask-2 corresponds to Fig. 4. CA is channel attention, PA is pixel attention, CA+PA is pixel channel hybrid attention, HAB is adaptive hybrid attention

Mask-1	PSNR/SSIM	Mask-2	PSNR/SSIM
CA	34.29/0.768	CA	34.42/0.773
PA	34.14/0.760	PA	34.20/0.764
CA+PA	34.90/0.829	CA+PA	35.36/0.848
HAB	36.49/0.867	HAB	**37.13/0.893**

Conclusion

In order to improve MRI image denoising performance and mitigate information loss, this paper introduces a two-stage denoising enhancement model named NRAE, which is based on Blind2Unblind and incorporates GANs. This method not only successfully reduces noise but also effectively restores anatomical details. It achieves lossless denoising through the use of a masked mapper and dual-branch joint training. By incorporating perceptual loss, cyclic loss and an adaptive hybrid attention mechanism, the model enhances local detail richness during the refinement stage.

Experiments show that in terms of denoising, our method achieves an improvement of nearly 0.7 dB in PSNR and about 0.07 in SSIM compared to classical methods. Notably, our detail information recovery demonstrates a PSNR improvement of approximately 1.4 dB over classical methods and an increase of about 0.6 dB in PSNR compared to the baseline method, with only a 0.015 decrease in SSIM. From a visual perspective, NRAE can more effectively restore the details in the image, resulting in richer and clearer image detail representation.

Clinical trial number

Not applicable.

Authors' contributions

Renfeng Liu: Conceptualization, Data treating, Formal analysis, Writing-original draft, Visualization. Songyan Xiao: Conceptualization, Methodology, Formal analysis, Writing-review & editing. Tianwei Liu,Jianfeng Chen: Formal analysis, Writing-review & editing. Fei Jiang, Cao Yuan: Formal analysis, Visualization.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and materials

The data for this experiment comes from real clinical data and the IXI dataset collected from three different hospitals in London(https://brain-development.org/ixi-dataset/).

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments. The real clinical data used in this study was fully anonymized before being provided to the researchers, ensuring that no identifiable information of the patients was accessible. The public dataset used in this study, the IXI dataset, is freely available and was collected in compliance with ethical standards. No additional ethical approval was required for the use of this dataset. Further details about the IXI dataset can be found at https://brain-development.org/ixi-dataset/.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 17 July 2024 Accepted: 19 September 2024
Published online: 28 September 2024

References

1. Li Y, Zhang K, Shi W, Miao Y, Jiang Z. A novel medical image denoising method based on conditional generative adversarial network. Comput Math Meth Med. 2021;2021:1–11. https://doi.org/10.1155/2021/9974017.
2. Tomasi C, Manduchi R. Bilateral filtering for gray and color images. In: Sixth international conference on computer vision (IEEE Cat. No. 98CH36271). IEEE; 1998. pp. 839–846. https://doi.org/10.1109/ICCV.1998.710815.
3. Buades A, Coll B, Morel JMA, non-local algorithm for image denoising. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol. 2. IEEE; 2005. pp. 60–5. https://doi.org/10.1109/CVPR.2005.38.
4. Dabov K, Foi A, Katkovnik V, Egiazarian K. Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans Image Process. 2007;16(8):2080–95. https://doi.org/10.1109/TIP.2007.901238.
5. Lyu Q, You C, Shan H, Wang G. Super-resolution MRI through deep learning. 2018. arXiv:1810.06776
6. Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Trans Image Process. 2017;26(7):3142–55. https://doi.org/10.1109/TIP.2017.2662206.
7. Lehtinen J, Munkberg J, Hasselgren J, Laine S, Karras T, Aittala M, et al. Noise2Noise: Learning image restoration without clean data.2018. arXiv:1803.04189.
8. Zhussip M, Soltanayev S, Chun SY. Extending stein's unbiased risk estimator to train deep denoisers with correlated pairs of noisy images. Adv Neural Inf Process Syst. 2019;32. https://doi.org/10.48550/arXiv.1902.02452.
9. Fadnavis S, Batson J, Garyfallidis E. Patch2Self: Denoising Diffusion MRI with Self-Supervised Learning. Adv Neural Inf Process Syst. 2020;33:16293–16303. https://doi.org/10.48550/arXiv.2011.01355.
10. Kim K, Ye JC. Noise2score: tweedie's approach to self-supervised image denoising without clean images. Adv Neural Inf Process Syst. 2021;34:864–874. https://doi.org/10.48550/arXiv.2106.07009.
11. Krull A, Buchholz TO, Jug F. Noise2void-learning denoising from single noisy images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019. pp. 2129–2137. https://doi.org/10.48550/arXiv.1811.10980.
12. Batson J, Royer L. Noise2self: Blind denoising by self-supervision. In: International Conference on Machine Learning. PMLR; 2019. pp. 524–533.
13. Deng Z, Luo Y, Zhu J, Zhang B. Measuring Uncertainty through Bayesian Learning of Deep Neural Network Structure. 2019. arXiv:1911.09804.
14. Huang T, Li S, Jia X, Lu H, Liu J. Neighbor2neighbor: Self-supervised denoising from single noisy images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021. pp. 14781–14790. https://doi.org/10.48550/arXiv.2101.02824.
15. Wang Z, Liu J, Li G, Han H. Blind2unblind: Self-supervised image denoising with visible blind spots. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022. pp. 2027–2036. https://doi.org/10.48550/arXiv.2203.06967.
16. Li G, Ji L, You C, Gao S, Zhou L, Bai K, et al. MARGANVAC: metal artifact reduction method based on generative adversarial network with variable constraints. Phys Med Biol. 2023;68(20):205005. https://doi.org/10.1088/1361-6560/acf8ac.
17. Lyu Q, You C, Shan H, Zhang Y, Wang G. Super-resolution MRI and CT through GAN-circle. In: Developments in X-ray tomography XII, vol. 11113. SPIE; 2019. pp. 202–208. https://doi.org/10.1117/12.2530592.
18. You C, Yang L, Zhang Y, Wang G. Low-dose CT via deep CNN with skip connection and network-in-network. In: Developments in X-Ray tomography XII, vol. 11113. SPIE; 2019. pp. 429–434. https://doi.org/10.48550/arXiv.1811.10564.
19. You C, Li Guang YZ. CT Super-resolution GAN Constrained by the Identical, Residual, and Cycle Learning Ensemble. 2018. https://doi.org/10.48550/arXiv.1808.04256.
20. You C, Yang Q, Shan H, Gjesteby L, Li G, Ju S, et al. Structurally-sensitive multi-scale deep neural network for low-dose CT denoising. IEEE Access. 2018;6:41839–55. https://doi.org/10.1109/ACCESS.2018.2858196.
21. He K, Chen X, Xie S, Li Y, Dollár P, Girshick R. Masked autoencoders are scalable vision learners. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022. pp. 16000–16009. https://doi.org/10.48550/arXiv.2111.06377.
22. Krull A, Vičar T, Prakash M, Lalit M, Jug F. Probabilistic noise2void: Unsupervised content-aware denoising. Front Comput Sci. 2020;2:5. https://doi.org/10.3389/fcomp.2020.00005.
23. Broaddus C, Krull A, Weigert M, Schmidt U, Myers G, Removing structured noise with self-supervised blind-spot networks. In,. IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE. 2020;2020:159–63. https://doi.org/10.1109/ISBI45749.2020.9098336.
24. Quan Y, Chen M, Pang T, Ji H. Self2self with dropout: Learning self-supervised denoising from single image. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. pp. 1890–1898. https://doi.org/10.1109/CVPR42600.2020.00196.
25. Kupyn O, Budzan V, Mykhailych M, Mishkin D, Matas J. Deblurgan: Blind motion deblurring using conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018. pp. 8183–8192. https://doi.org/10.48550/arXiv.1711.07064.
26. Thakur RS, Chatterjee S, Yadav RN, Gupta L. Medical image denoising using convolutional neural networks. In: Digital Image Enhancement and Reconstruction. Elsevier; 2023. pp. 115–138. https://doi.org/10.1016/B978-0-32-398370-9.00012-3.
27. Kong L, Dong J, Ge J, Li M, Pan J. Efficient frequency domain-based transformers for high-quality image deblurring. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023. pp. 5886–5895. https://doi.org/10.48550/arXiv.2211.12250.

28. Wang B, Deng F, Jiang P, Wang S, Han X, Zheng H. WiTUnet: A U-Shaped Architecture Integrating CNN and Transformer for Improved Feature Alignment and Local Information Fusion. 2024. arXiv:2404.09533.
29. Ashish V. Attention is all you need. Adv Neural Inf Process Syst. 2017;30:I. https://doi.org/10.48550/arXiv.1706.03762.
30. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. pp. 7132–7141. https://doi.org/10.48550/arXiv.1709.01507.
31. You C, Zhao R, Liu F, Dong S, Chinchali S, Topcu U, et al. Class-aware adversarial transformers for medical image segmentation. Adv Neural Inf Process Syst. 2022;35:29582–29596. https://doi.org/10.48550/arXiv.2201.10737.
32. Woo S, Park J, Lee JY, Kweon IS. Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). 2018. pp. 3–19. https://doi.org/10.48550/arXiv.1807.06521.
33. Cui Y, Knoll A. Dual-domain strip attention for image restoration. Neural Netw. 2024;171:429–39. https://doi.org/10.1016/j.neunet.2023.12.003.
34. Chen S, Ye T, Liu Y, Chen E. Dual-former: Hybrid self-attention transformer for efficient image restoration. Digit Signal Process. 2024;149:104485. https://doi.org/10.48550/arXiv.2210.01069.
35. Zhao H, Kong X, He J, Qiao Y, Dong C. Efficient image super-resolution using pixel attention. In: Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020. Proceedings, Part III 16. Springer; 2020. pp. 56–72. https://doi.org/10.48550/arXiv.2010.01073.
36. Chen X, Wang X, Zhou J, Qiao Y, Dong C. Activating more pixels in image super-resolution transformer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023. pp. 22367–22377. https://doi.org/10.48550/arXiv.2205.04437.
37. Zafar A, Aftab D, Qureshi R, Fan X, Chen P, Wu J, et al. Single Stage Adaptive Multi-Attention Network for Image Restoration. IEEE Trans Image Process. 2024. https://doi.org/10.1109/TIP.2024.3384838.
38. You C, Dai W, Min Y, Staib L, Duncan JS. Implicit anatomical rendering for medical image segmentation with stochastic experts. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2023. pp. 561–571. https://doi.org/10.48550/arXiv.2304.03209.
39. You C, Yang J, Chapiro J, Duncan JS. Unsupervised wasserstein distance guided domain adaptation for 3d multi-domain liver segmentation. In: Interpretable and Annotation-Efficient Learning for Medical Image Computing: Third International Workshop, iMIMIC 2020, Second International Workshop, MIL3ID 2020, and 5th International Workshop, LABELS 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4–8, 2020, Proceedings 3. Springer; 2020. pp. 155–163. https://doi.org/10.48550/arXiv.2009.02831.
40. Wolterink JM, Leiner T, Viergever MA, Išgum I. Generative adversarial networks for noise reduction in low-dose CT. IEEE Trans Med Imaging. 2017;36(12):2536–45. https://doi.org/10.1109/TMI.2017.2708987.
41. Ahmad W, Ali H, Shah Z, Azmat S. A new generative adversarial network for medical images super resolution. Sci Rep. 2022;12(1):9533. https://doi.org/10.1038/s41598-022-13658-4.
42. Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, et al. Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. IEEE Trans Med Imaging. 2018;37(6):1348–1357. https://doi.org/10.48550/arXiv.1708.00961.
43. Uddin AS, Chung T, Bae SH. A perceptually inspired new blind image denoising method using $L_1$and perceptual loss. IEEE Access. 2019;7:90538–49. https://doi.org/10.1109/ACCESS.2019.2926848.
44. Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. Springer; 2016. pp.694–711. https://doi.org/10.48550/arXiv.1603.08155.
45. Yin Z, Xia K, He Z, Zhang J, Wang S, Zu B. Unpaired Image Denoising via Wasserstein GAN in Low-Dose CT Image with Multi-Perceptual Loss and Fidelity Loss. Symmetry. 2021;13:126. https://doi.org/10.3390/sym13010126.

Publisher's Note

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