

RESEARCH

Open Access



Super-resolution reconstruction for early cervical cancer magnetic resonance imaging based on deep learning

Chunxia Chen^{1†}, Liu Xiong^{2†}, Yongping Lin^{2*}, Ming Li², Zhiyu Song², Jialin Su² and Wenting Cao¹

[†]Chunxia Chen and Liu Xiong have contributed equally to this work.

*Correspondence: yplin@xmut.edu.cn

¹ Fujian Maternity and Child Health Hospital, College of Clinical Medicine for Obstetrics & Gynecology and Pediatrics, Fujian Medical University, No.18 Daoshan Road, Gulou District, Fuzhou 350001, Fujian, China
² School of Optoelectronic and Communication Engineering, Xiamen University of Technology, No.600 Ligong Road, Jimei District, Xiamen 361024, Fujian, China

Abstract

This study aims to develop a super-resolution (SR) algorithm tailored specifically for enhancing the image quality and resolution of early cervical cancer (CC) magnetic resonance imaging (MRI) images. The proposed method is subjected to both qualitative and quantitative analyses, thoroughly investigating its performance across various upscaling factors and assessing its impact on medical image segmentation tasks. The innovative SR algorithm employed for reconstructing early CC MRI images integrates complex architectures and deep convolutional kernels. Training is conducted on matched pairs of input images through a multi-input model. The research findings highlight the significant advantages of the proposed SR method on two distinct datasets at different upscaling factors. Specifically, at a 2× upscaling factor, the sagittal test set outperforms the state-of-the-art methods in the PSNR index evaluation, second only to the hybrid attention transformer, while the axial test set outperforms the state-of-the-art methods in both PSNR and SSIM index evaluation. At a 4× upscaling factor, both the sagittal test set and the axial test set achieve the best results in the evaluation of PSNR and SSIM indicators. This method not only effectively enhances image quality, but also exhibits superior performance in medical segmentation tasks, thereby providing a more reliable foundation for clinical diagnosis and image analysis.

Keywords: Computer-aided diagnosis, Deep learning, Magnetic resonance imaging, Uterine cavity line

Introduction

Cervical cancer (CC) is the second most common cancer in Chinese women and ranks fourth for both incidence and mortality in women worldwide, and its morbidity and mortality have revealed an upward inclination in recent years [1]. This increase is attributed, in part, to the improvement in people's living standards, resulting in a year-by-year rise in the incidence of CC, affecting an increasing number of young women. Among human cancers, CC uniquely stands as the only curable cancer, underscoring the critical importance of timely interventions for its successful management [2].

Various imaging modalities, including lymphography, ultrasonography, computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography



(PET), PET–CT, and MRI–PET, offer valuable information for both diagnosing and prognosis of CC. Among these, MRI stands out as a crucial tool frequently employed by physicians to comprehend the morphology of the primary CC tumor, assess the extent of parametrial infiltration, and determine the number and location of lymph nodes and metastases [3]. In the development of clinical treatment, surgery and radiotherapy emerge as the preferred modalities for addressing CC. Particularly in radiotherapy, the accurate depiction of the tumor through medical imaging plays a central role. This precise visualization guides pre-operative preparations and post-operative treatments, significantly enhancing the overall detection capability for CC.

Often more than one imaging modality is involved in clinical decision-making, as different modalities often provide complementary insights [4]. Single modalities often fail or are inadequate because they do not adequately subdivide the tumor in the region of interest, and the use of different MRI modalities can effectively compensate for these weaknesses [5]. Diffusion-weighted imaging (DWI) measurements are typically fast, require no administration of exogenous contrast medium, and can be appended to existing imaging protocols without a significant increase in examination time [6]. It can provide both qualitative and quantitative information, with diffusion anisotropy aiding in the identification of tumor infiltration into adjacent structures, thereby proving invaluable in tumor assessment [7]. However, DWI imaging also presents notable drawbacks, such as poor signal-to-noise ratio, more noise, and greater susceptibility to pulsatile and susceptibility artifacts. T2-weighted images (T2WI) are acquired using long repetition time and long signal recovery time, presenting high-intensity signals for tissues with long T2 relaxation time, thus providing distinct contrast and revealing fine anatomical details, particularly useful for inflammation and tumor diagnosis [8]. However, the acquisition time is prolonged, and it requires operators with extensive experience to adjust instrument parameters according to different circumstances, resulting in variable quality of acquired T2WI images [9]. While MRI facilitates comparative observations of various tissue structures from different angles, enhancing the early detection and diagnosis of numerous diseases, its use comes with challenges. Compared to other imaging modalities, MRI acquisitions present significant challenges and are difficult to maintain for long periods of prohibition in enclosed spaces [10]. To enhance acquisition efficiency and alleviate patient discomfort, faster scanning speeds are often required to avoid motion artifacts [11]. It is worth noting that, despite the benefits of expedited scans, this type of acquisition tends to yield low-resolution (LR) images, which may lack the sensitivity required for accurate disease extent determination.

Recent advancements have illustrated the capacity of super-resolution (SR) technology to enhance image quality by algorithms without necessitating any hardware upgrades [12]. The widespread adoption of deep learning has positioned it as a predominant approach for SR imaging, leading to the development of various network-based SR models [12]. To achieve excellent texture information in image super-resolution, the Generative Adversarial Network (GAN) is a key approach. While incorporating GANs produces rich details, it also generates artifacts that can negatively affect the visual experience. Liang et al. [13], proposed a local discriminant learning method to distinguish artifacts generated by GANs from real details. During model training, this method explicitly penalizes artifacts without sacrificing real details through discriminant learning.

Although it does not completely eliminate artifacts, their work represents a significant step forward. Generally, the larger the pixel range utilized, the better the reconstruction effect tends to be. Dong et al. [14]. designed a hybrid attention transformer (HAT) that combines self-attention, channel attention, and overlapping cross-attention to activate more pixels for better reconstruction. This method has achieved very good results in super-resolution tasks. Transformer-based methods perform well in image super-resolution tasks, surpassing traditional neural networks. However, previous work typically limits self-attention calculations to non-overlapping windows to save computational costs, restricting the input information that Transformer-based networks can utilize from a limited spatial range. Chu et al. [15]. Proposed a Hybrid Multi-Axis Aggregation network (HMA) to better utilize feature latent information. HMA consists of Hybrid Transformer Blocks (RHTB) and Grid Attention Blocks (GAB). RHTB combines attention mechanisms to enhance non-local feature fusion, while GAB is used for cross-domain information interaction and jointly modeling similar features. The combination of these two modules allows the utilization of more features, leading to better results.

Not only for natural domain images, but nowadays SR is also increasingly used in the field of medical images. Zhou et al. [16] conducted a comprehensive exploration and evaluation of the magnetic resonance imaging-based SR generative adversarial network for brain tumors (MRBT-SR-GAN), showcasing its potential in enhancing the resolution of magnetic resonance imaging for brain tumors. MRBT-SR-GAN contains only one modality, without including other modalities such as T2WI and T1WI in the study. Qiu et al. [17] developed an efficient deep learning-based medical image SR method for assisting in the examination of knee osteoarthritis. The method uses an efficient sub-pixel convolutional neural network (ESPCN) with three layers of super-resolution convolutional neural network (SRCNN) and one sub-pixel convolutional layer. The efficient sub-pixel convolutional layer is added to the hidden layer and replaced with a small network composed of cascaded convolutions to process low-resolution images. However, the edge reconstruction effect is still different from the original image, and the reconstruction speed is slow. Oktay et al. [18] introduced a method utilizing multiple input data from diverse viewing planes to enhance SR image reconstruction. Experimental results showed its superiority over existing SR methods in image quality and computational efficiency. However, motion-induced inter-slice and stack spatial misalignments pose a challenge, reducing accuracy. Rousseau et al. proposed a method for image super-resolution by utilizing anatomical intermodality priors extracted from a reference image. This method is particularly useful in MRI applications, where it enhances the resolution of T2WI stacks using isotropic HR T1-weighted stacks as a basis [19]. A limitation of this method is its complexity, as it requires balancing the performance of the observation model and the driving reconstruction model. Additionally, it has only been compared to standard interpolation algorithms, which may not fully reflect its performance against more advanced techniques. Yurt et al. [20] proposed a multi-contrast MRI super-resolution method that can simultaneously deblur images of different contrasts. This method relies on a generative adversarial network. GANs can lead to a lack of fidelity in reconstructed images. Cukur et al. [21] noted that the limited prior information of a single-contrast MRI image restricts reconstruction performance. They suggested that using multi-contrast MRI as input data can enhance reconstruction performance and

employ a GAN for image reconstruction. Their work is similar to the previous ones and also fails to address the lack of fidelity of GAN reconstruction. Bhadra et al. [22] introduced an image-adaptive GAN reconstruction method (IAGAN) based on the generative adversarial network, effectively addressing the fidelity issue between reconstructed images and observed data typically caused by using GANs. IAGAN also performs well in noisy image scenarios. However, this study lacks quantitative evaluation. A et al. [23] proposed an adaptive diffusion prior model, AdaDiff. It leverages an efficient diffusion prior trained via adversarial mapping over extensive reverse diffusion steps. It shows strong performance in MRI reconstruction and is suitable for both intra-domain and cross-domain scenarios. However, its adaptive process increases running time, and the memory load during inference is substantial, hindering practical application. Korkmaz et al. [24] proposed a novel unsupervised MRI reconstruction method based on Zero-Shot Learned Adversarial Transformers (SLATER). SLATER uses cross-attention transformers to better capture contextual image features, addressing the challenge convolutional architectures face in capturing long-range relationships. This model achieved good results in MRI image reconstruction. It is worth noting that SLATER demonstrates excellent results for both within-domain and across-domain tasks, primarily due to its high-quality MR prior. The prior is combined with the imaging operator during inference, and the learned prior is used to reconstruct the undersampled acquisition through unsupervised model adaptation. The high-quality MR prior can be flexibly adapted to the test domain, and while the model adaptation procedure in SLATER helps limit potential performance loss, it is very time-consuming.

To address the issues associated with incomplete information caused by single-modality input, the lack of fidelity caused by GAN-based reconstruction, the high complexity of models, and time-consuming inference processes, this paper proposes a multi-input image super-resolution (SR) method to improve the quality of multi-modalities MRI images. The proposed approach uses DWI and T2WI as inputs, leveraging the complementary information provided by these two modalities. A deep convolutional neural network is employed for feature extraction, effectively mitigating image artifacts caused by different imaging angles without significantly increasing computational complexity. Additionally, the study performs a semantic segmentation task on the reconstructed high-resolution images to demonstrate their potential application value in clinical diagnosis.

Results

To test the performance of the proposed networks, deep learning networks LDL [13], HAT[14], HMA [15], ESPCN [25], RDN [26], SRGAN [27] and SwinIR [28] are used as baseline networks for comparison. The two datasets are trained on different methods and LR images obtained by image preprocessing of HR images are used as test images. The results of the quantitative evaluation of each method on two datasets (sagittal and axial) are shown in Table 1. The proposed method outperforms the other methods overall. From Table 1, it can be seen that all the methods achieved better results at an upscale factor of 2 than at an upscale factor of 4. Among all the models, ESPCN, RDN, and SRGAN perform well at $2\times$, while there is a significant drop at $4\times$, and the challenges remain on SR tasks with large upscale factors. SwinIR, LDL, HAT, HMA, and the

Table 1 Quantitative evaluation of the proposed method with state-of-the-art SR algorithms for 2× and 4× on the MRI CC dataset. ↑ means that the larger score is better

	Upscale factors	Sagittal (mean±Std)		Axial (mean±Std)	
		SSIM↑	PSNR↑	SSIM↑	PSNR↑
ESPCN	2×	0.87 ± 0.03	33.28 ± 1.52	0.88 ± 0.04	34.30 ± 1.08
	4×	0.81 ± 0.02	30.16 ± 0.34	0.84 ± 0.02	31.97 ± 0.40
RDN	2×	0.89 ± 0.01	33.61 ± 1.12	0.86 ± 0.02	34.80 ± 1.57
	4×	0.82 ± 0.02	31.46 ± 1.23	0.81 ± 0.06	32.18 ± 1.32
SRGAN	2×	0.94 ± 0.01	32.63 ± 1.74	0.92 ± 0.03	31.18 ± 0.26
	4×	0.86 ± 0.01	24.56 ± 1.50	0.77 ± 0.01	26.47 ± 0.94
SwinIR	2×	0.90 ± 0.02	42.35 ± 1.34	0.97 ± 0.01	43.38 ± 0.45
	4×	0.86 ± 0.05	34.35 ± 1.76	0.92 ± 0.05	33.03 ± 1.09
HAT	2×	0.97 ± 0.02	44.58 ± 0.76	0.96 ± 0.02	43.02 ± 0.89
	4×	0.91 ± 0.01	34.56 ± 0.85	0.91 ± 0.04	35.98 ± 0.71
HAM	2×	0.95 ± 0.03	41.29 ± 0.40	0.97 ± 0.02	43.22 ± 0.57
	4×	0.89 ± 0.03	35.47 ± 0.74	0.91 ± 0.01	35.94 ± 0.64
LDL	2×	0.98 ± 0.07	43.65 ± 1.32	0.97 ± 0.01	43.29 ± 1.54
	4×	0.91 ± 0.01	34.97 ± 0.73	0.92 ± 0.05	35.97 ± 1.78
Ours	2×	0.98 ± 0.01	43.77 ± 0.99	0.98 ± 0.05	43.67 ± 0.62
	4×	0.92 ± 0.01	36.8 ± 0.32	0.93 ± 0.01	38.18 ± 0.74

The best results are highlighted in bold

proposed method maintain high PSNR and SSIM at 2× and 4× sampling factors. At the 2× sampling factor, HAT slightly outperforms our method on the sagittal plane test set. HAT introduces channel attention into the transformer to utilize more input information, activating more pixels for better reconstruction. At the 2× sampling factor, the proposed method outperforms other methods because the network adopts a more complex architecture and employs deeper convolutional kernels, which help to learn more complex features and improve the model's ability to perceive details.

The reconstruction performance of each method was analyzed from a subjective perspective, and the reconstructed images of each method were randomly selected from the sagittal and cross-sectional images for comparison. As can be seen from Fig 1, the image reconstructed by the ESPCN method is still relatively blurry and the enhancement effect is not good; the image reconstructed by the RDN method is significantly improved compared with ESPCN at 2×, but there is a phenomenon of overlapping contour edge shadows; the image reconstructed by the SRGAN method performs well at 2×, and is oversharpened at 4×, which may be due to the fact that the generative adversarial network makes it more difficult for the generator to distinguish between real images and generated images during training, thereby increasing the difficulty of the discriminator, which may cause the generated image to be oversharpened at 4×. In addition, SRGAN usually uses perceptual loss, which may focus too much on high-frequency details and lead to poor subjective perception of the reconstructed image. The clarity of the images reconstructed by HMA, HAT, LDL, SwinIR and the proposed method is greatly improved, the contours are clearer, and closer to HR images. However, compared with the proposed method, the texture details of HMA, HAT, LDL, and SwinIR are weakened at 4×. The proposed method is

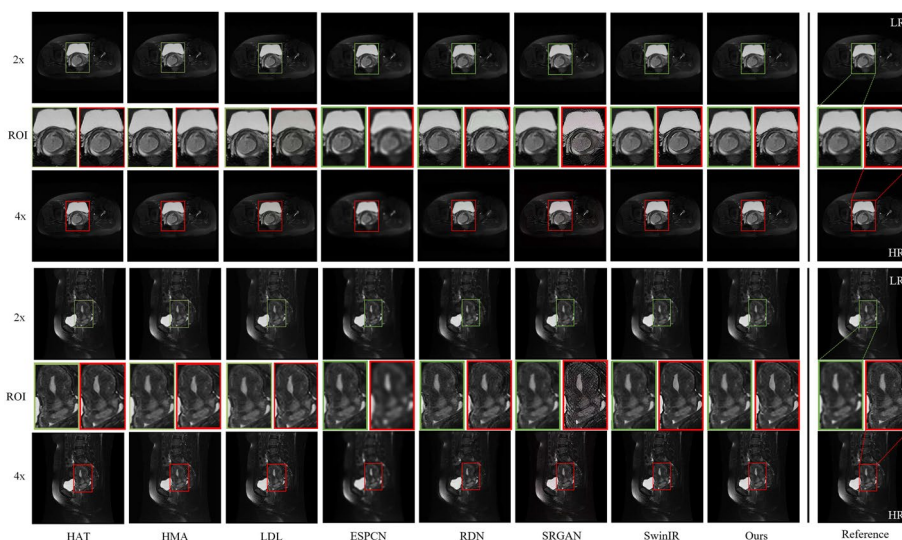


Fig. 1 Comparison of the super-resolution (SR) methods used when upsampling by factors of two and four in example images

comparable to or even better than the SOTA methods in terms of image quality, especially at 4x. This is attributed to the integration of multiple advanced deep learning techniques, which are trained on a large amount of high-quality data to better understand LR images and convert them into HR images. Compared with other SR methods, the proposed method can generate more natural and smoother images without obvious artifacts or distortion, thereby effectively improving image quality and meeting the demand for high-definition images in various fields.

With the continuous advancement of HR MRI and signal processing techniques, histogram analysis in cancer MRI is increasingly gaining prominence [29]. Given that MRI images present grey-scale characteristics, we chose to employ histograms in this study as a means of quantitative assessment to objectively evaluate the differences between the outputs of different models and the HR images for a more comprehensive assessment of model performance. By comparing the histograms of the images generated by each method as shown in Fig 2, we can clearly observe that the histograms of SwinIR and the proposed method are much closer to each other with respect to those of the HR images. Zooming in further on the details of the histograms, we can observe more intuitively the high degree of overlap between the histograms of the images generated by the proposed method and those of the HR images at the higher grey levels, presenting finer and more consistent features. This indicates that the proposed method achieves a significant advantage in preserving image details and contrast when performing SR reconstruction of MRI images, and better mimics the grey-scale distribution of HR images. This histogram similarity not only emphasizes the realism of the model-generated image, but also highlights its better preservation of the fine features of the original image during the reconstruction process. Such an objective assessment method helps to provide a deeper understanding of the performance of different models in medical image SR studies.

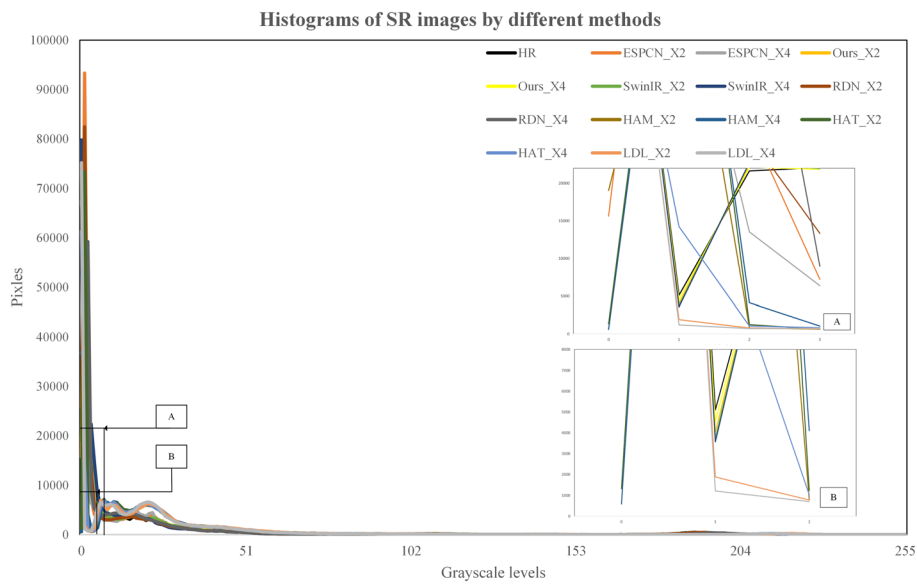


Fig. 2 Histograms of SR images by different methods

The image histograms generated by the proposed method highly overlap with those of HR images at high grey levels, showing finer and more consistent features. This similarity emphasizes the significant advantage of the proposed method in preserving the details and contrast of MRI images. The histograms of the method better simulate the grey-scale distribution of HR images, suggesting that it is more effective in SR reconstruction tasks. This objective evaluation method not only provides an intuitive visual observation, but also provides us with an in-depth understanding of the model performance. Through histogram analysis, we can quantitatively verify the similarity between the model-generated images and the HR images, and the highly overlapping histogram features further demonstrate the superiority of the proposed method in maintaining the features of the MRI images.

The study added eight types of noise to the input images to test the robustness of the proposed method against source image defects caused by the measurement system and calculated the percentage of model performance retention as the robustness score[30]. Fig 3 shows examples of common MRI corruptions, and Table 2 presents the robustness results of our method. The proposed method is generally robust and can maintain performance above 70% even when the image content does not match spatially.

To analyze the performance of the proposed method in task transfer scenarios between training and test data, we conducted an ablation experiment where the model was trained with a $2\times$ upscaling factor and tested with a $4\times$ upscaling factor. Fig 4(a) illustrates the data preprocessing process, where padding is used to ensure the image reaches the target size. Fig 4(b) provides a visualization of the results. The reconstructed image shows significant improvements in the regions of interest, demonstrating that the proposed method performs well in task transfer scenarios.

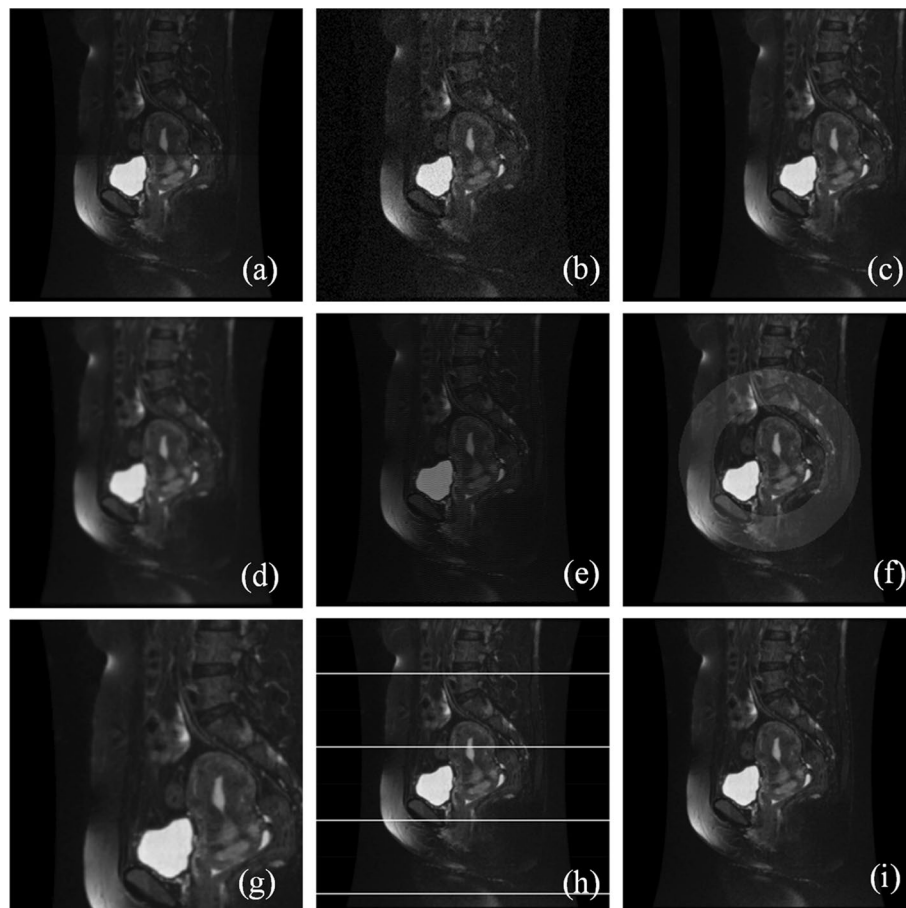


Fig. 3 Examples of corrupted MRI images. **a** Inhomogeneity artifacts, **b** Gaussian noise, **c** ghosting artifacts, **d** Gaussian blur, **e** truncation artifacts, **f** ring artifacts, **g** stretch, **h** RF interference, **i** reference

Table 2 Robustness scores on MRI image corruptions

Corruptions	Score (%)	Corruptions	Score (%)
Inhomogeneity artifacts	83.53	Truncation artifacts	74.09
Gaussian noise	74.57	Ring artifacts	82.08
Ghosting artifacts	71.70	Stretch	70.82
Gaussian blur	87.16	RF interference	90.79

Discussion

In order to verify the validity of the proposed model, ablation experiments were conducted in this study and the results are shown in Table 3. Based on the results of the ablation experiments, it is evident that using a single input leads to the poorest PSNR and SSIM. However, it is worth noting that with only one input, the computational complexity of the model remains relatively low, with parameters totaling only 16.1 M. This suggests that while the performance may be suboptimal with a single input, the computational burden on the model is minimal. We performed statistical analysis of the results using SPSS (version 26.0, SPSS Inc.). The p values were obtained through t-tests, and our model showed p values less than 0.05 when compared with the NAFNet and NAFSSR

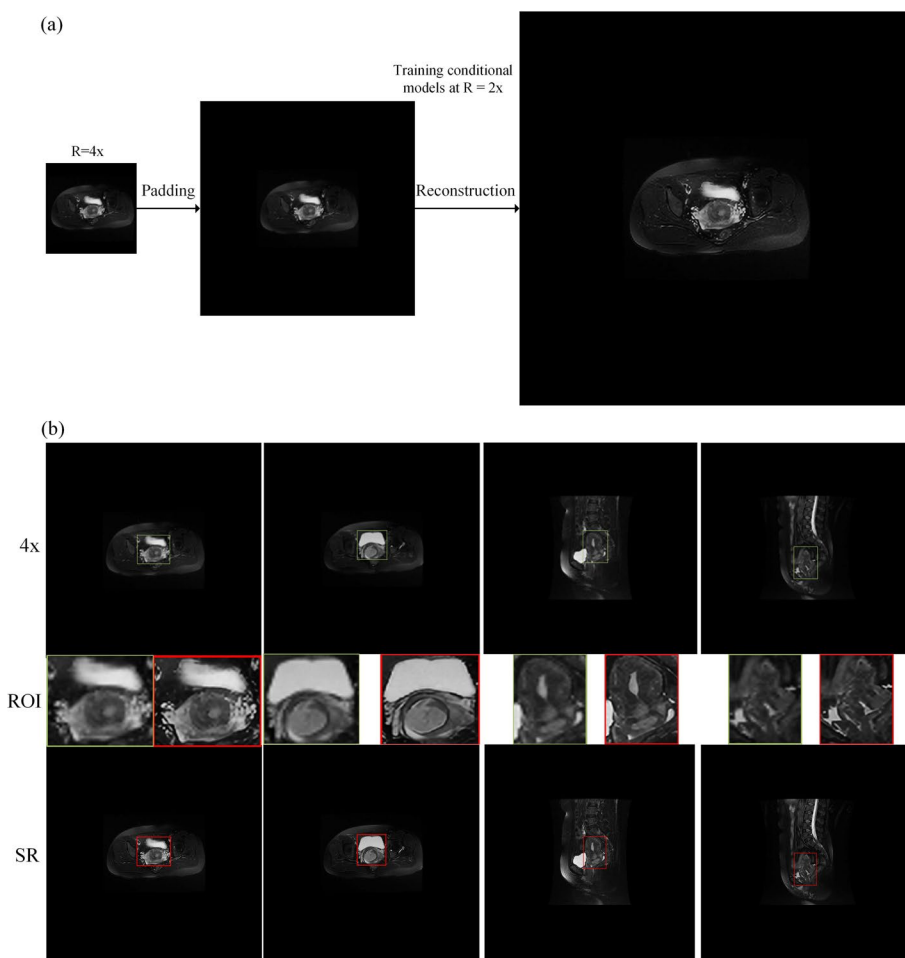


Fig. 4 Performance testing in a task shifting scenario. The model is trained at a 2x upscaling factor and tested at a 4x upscaling factor. a Image resizing: the study applied padding to the images to ensure that the 4x upsampled images can fit the target size. b Visualization of the results: the green and red boxes indicate the regions of interest in the 4x upscaling factor and reconstructed images, respectively

Table 3 Comparative analysis of ablation experiments and model parameter quantities

	Multi-input	DepthConv	PSNR	SSIM	Pararms(<i>M</i>)	<i>p</i> value
NAFNET			33.88	0.886	16.1	< 0.05
NAFSSR	✓		43.24	0.973	23.8	< 0.05
Ours	✓	✓	44.99	0.987	23.8	

models. This indicates a significant difference between our model and the other two models, demonstrating that our model has better reconstruction ability.

On the other hand, when multiple input images are utilized, there is a notable improvement in overall performance. This enhancement underscores the importance of leveraging multiple sources of information for better reconstruction quality. Furthermore, the incorporation of deep convolutional layers into our model has yielded the highest PSNR and SSIM without a significant increase in computational complexity. This indicates that

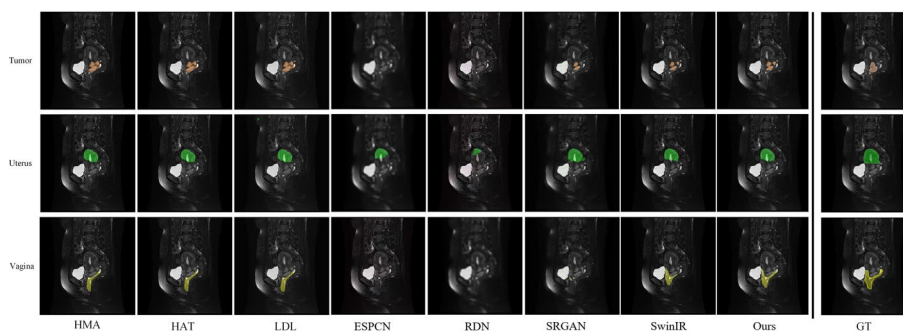


Fig. 5 The results of segmentation of SR images generated by different SR reconstruction methods on different organs using the U-Net network

the additional depth in the model architecture contributes to improved reconstruction quality without imposing a heavier computational burden.

The study of image SR aims to improve the visual quality of medical images and provide doctors with more detailed information to facilitate accurate diagnosis and treatment of diseases. As for U-Net [31], it is a convolutional neural network architecture commonly used for semantic segmentation tasks, particularly in medical image analysis. This study further examines the segmentation results of SR images generated by different SR methods on different organs using the U-Net network and shows the comparison results through Fig 5. This segmentation analysis helps to assess the effectiveness of SR methods in restoring different tissue structures and is directly related to the accurate identification of organs and lesions in clinical medicine.

In the segmentation task, this study employs commonly used segmentation evaluation metrics, including Pixel Accuracy (PA) Dice Similarity Coefficient (DSC) and Hausdorff distance (HD) to assess the segmentation results, as shown in Table 4. When comparing the segmentation results of different organs, we observed that the uterus had the most segmented portion, followed by the tumor, while the vagina was the least segmented. This discrepancy may be due to the morphological differences and different contrasts between organs. ESPCN and RDN are almost impossible to segment correctly or obtain segmentation results at all, which suggests that they have large limitations in reconstructing organ structures. In contrast, HAT, LDL, HMA, SwinIR, and the proposed method performed better in the segmentation of most organs; however, the segmentation results of SwinIR had problems of incompleteness and discontinuity. HAT, LDL, and HMA had incomplete segmentation of the vagina. The proposed method, on the other hand, is able to perform the organ segmentation task with relative accuracy, further validating the potential advantages of the method in practical medical image analysis.

This improved segmentation result is of great significance for doctors to study tumors and the relationship between tissues on MRI images. The proposed method not only achieved significant improvement in image quality, but also demonstrated better performance in real medical image segmentation tasks, providing potentially useful support for clinical practice in the field of medical imaging. This further highlights the potential for practical applications of image SR studies to provide more accurate and reliable tools for medical image analysis. It is reported that using multi-contrast MRI images (T1, T2, PD) as input for super-resolution reconstruction yields better results than using

Table 4 Segmentation results for different tissues on SR images. ↑ and ↓ mean that the larger or smaller score is better, respectively

	Tumor (mean ± Std)			Uterus (mean ± Std)			Vagina (mean ± Std)		
	DSC(%)↑	PA(%)↑	HD↓	DSC(%)↑	PA(%)↑	HD↓	DSC(%)↑	PA(%)↑	HD↓
ESPCN	25.97 ± 2.48	37.01 ± 2.47	46.54 ± 3.62	63.94 ± 3.61	76.88 ± 1.26	40.24 ± 3.98	8.08 ± 2.06	12.08 ± 2.45	36.15 ± 3.17
RDN	61.99 ± 1.50	68.24 ± 1.54	41.10 ± 4.47	44.66 ± 2.46	62.21 ± 1.39	32.62 ± 3.59	17.15 ± 2.56	22.26 ± 2.61	40.59 ± 3.42
SRGAN	84.43 ± 2.56	84.42 ± 2.52	37.42 ± 3.50	87.12 ± 3.24	89.51 ± 2.36	42.93 ± 3.35	25.0 ± 2.35	34.95 ± 1.34	28.45 ± 3.90
SwiniR	76.62 ± 1.48	83.15 ± 2.50	40.34 ± 2.96	85.30 ± 2.42	90.17 ± 2.85	37.46 ± 2.11	55.92 ± 2.45	47.65 ± 1.59	31.11 ± 2.36
HAM	76.03 ± 2.51	76.52 ± 1.32	25.64 ± 1.45	84.02 ± 2.34	84.96 ± 2.42	32.59 ± 2.66	45.57 ± 3.56	35.26 ± 3.49	42.73 ± 2.79
HAT	83.52 ± 2.35	77.24 ± 2.45	30.29 ± 1.15	83.24 ± 3.51	86.64 ± 2.12	25.74 ± 2.15	52.65 ± 2.74	52.65 ± 2.14	34.35 ± 3.78
LDL	79.54 ± 2.45	80.54 ± 2.43	27.58 ± 1.39	79.65 ± 2.49	87.54 ± 2.34	24.15 ± 2.78	53.64 ± 1.15	57.64 ± 2.51	29.97 ± 2.12
Ours	89.47 ± 2.46	83.80 ± 1.34	21.62 ± 1.56	88.55 ± 3.41	91.32 ± 1.95	19.45 ± 2.59	63.23 ± 1.11	69.16 ± 1.59	26.94 ± 2.31

The best results are highlighted in bold

a single image[20]. However, this approach employs GAN for super-resolution reconstruction, which makes it challenging to effectively utilize all feature information. Additionally, GAN is prone to artifacts and often lacks fidelity to real data. [21] is similar to the work of [20], and neither has addressed the defects associated with using GAN. In contrast, the method presented in this paper employs a deep convolutional neural network to extract features from input images. The input images use two modalities, DWI and T2WI, which provide complementary information, allowing deep convolutional networks to effectively utilize all available medical information. While conditional GAN-based image-to-image translation methods also use source images as input, the proposed approach leverages the strengths of deep convolutional networks to more robustly capture and integrate feature information from real images, minimizing artifacts and improving fidelity. In addition, the adversarial training process of GANs can lead to issues such as gradient vanishing and collapse mode. By comparison, deep convolutional neural network training is relatively stable and typically faster because it involves optimizing only one network. GANs require the simultaneous optimization of both the generator and discriminator networks, making parameter adjustments more complex and demanding more computing resources.

Although the proposed SR reconstruction method demonstrates significant advantages in this study, there are still some limitations. While deep separable convolution learns feature representations more efficiently and outperforms standard convolution in terms of parameter efficiency, it still requires learning two sets of convolution kernel parameters: the respective convolution kernels for the deep convolution stage and the point-by-point convolution stage. Therefore, although the size of each convolution kernel is typically smaller than standard convolution, deep separable convolution may result in an increase in the overall number of parameters due to the need for two sets of convolution kernels. In resource-constrained application scenarios, the relationship between computational cost and performance may need to be weighed to ensure that the method can still operate efficiently in real-world deployments. The performance of our method is highly dependent on the quality and diversity of the training data. In cases of insufficient data or poor data quality, the model's performance may be limited. Specifically, when the image content is shifted, our method may perform poorly in restoration. If artifacts appear during the migration process, they can generate unnatural textures that are inconsistent with the statistical properties of the real image. These inconsistent textures not only affect the visual quality of the image, but also pose challenges for subsequent processing and analysis.

In addition, the proposed model is a multi-input model with high matching requirements for input images. When there exists a missing set of images, model training or inference will not be effective, which may lead to data wastage. This strict matching requirement may introduce additional data management and acquisition challenges in practical applications, increasing the complexity of the system. Future research can be carried out in various aspects. First, to address the computational complexity and model size issues of the model, the computational burden and storage requirements of the model can be reduced by modifying the backbone network, adopting a more lightweight network structure, and applying model compression techniques or neural network quantization methods to improve the efficiency and feasibility of the model in practical

applications. Second, when facing the challenge of insufficient data usage, an innovative approach is to introduce an image generation model for generating missing data pairs, which are then used as part of the training set to make full use of all the available image information, thus improving the generalization ability and robustness of the model. In addition, future research could also extend the method to other types of cancer diagnosis. Since the diagnosis of most diseases often requires the integration of information from multiple imaging techniques, the combination of multiple imaging techniques can be explored to utilize multi-modal data for joint analysis, leading to a more comprehensive and accurate disease diagnosis. This integrated analysis approach is expected to lead to new breakthroughs and advances in cancer diagnosis and other disease diagnosis in the future.

Therefore, when promoting and applying the method, we need to consider the computational cost, the feasibility of data management, and the actual usage scenarios of the system. For some specific scenarios, some aspects of the model may need to be targeted and optimized or adjusted to meet the actual requirements and better balance the relationship between performance and resource consumption. Such a comprehensive consideration will help to better understand the applicability and limitations of the proposed approach in different application scenarios.

Conclusions

In this study, an SR algorithm for early CC MRI images is proposed. The results show that the proposed SR method outperforms the four current widely used SR methods under different magnification factors both from the perspective of qualitative and quantitative analyses. The method is able to show the morphology and boundaries of different tissues in early stage CC MRI images more clearly and effectively reduce the appearance of artifacts. Notably, the images after SR present better performance on the segmentation task, further validating the positive impact of the method on medical image analysis.

Materials and methods

Patients and data preparation

The Institutional Review Board (IRB) of Fujian Maternity and Child Health Hospital in China (FMCHH) approved the retrospective study, and the requirement for informed consent was waived. 198 patients who underwent pelvic MRI examination in FMCHH during the period from November 29, 2013, to December 17, 2020, were included in this study after being pathologically diagnosed with early-stage EC. Patients were identified by using information from the hospital's picture archiving and communication system (PACS). The exclusion criteria were as follows: (1) without a final pathologic diagnostic statement; (2) missing MRI data (no corresponding DWI or T2WI sequence). The total number of patients in the study was 99. Subsequently, all the multi-modal sequences selected were chosen with the following exclusion criteria: (1) presence of artifacts; (2) mismatch between DWI slice and T2WI. Finally, the experimental data are 2398 MRI slices. A flow diagram of the cohort selection is presented in Fig 6. Datasets were divided by patients which are shown in Fig 6.

MRI image dataset

In this study, DWI sequences and T2WI sequences of CC patients were used as research objects. Because the DWI image and the T2WI image of the same patients have correspondence, we regarded images of the same layer of slices in different sequences as a pair. One T2 image and one DWI image were guaranteed for each group. The original images were interpolated using bicubic interpolation to generate corresponding LR images with randomly added Gaussian noise. Data were required with a 1.5-T MRI scanner (Optima MRI360, GE Healthcare). Prior to the examination, the completion of bowel preparation (fasting on the morning of the examination) as well as bladder preparation (holding urine appropriately to keep the bladder full) is required. During the examination, the patient is asked to remain as still as possible to avoid motion artifacts, and in special cases, a restraint band may be used to maintain immobilization. Two commonly used MRI modalities (DWI and T2WI) were used in this experiment, and the detailed MR acquisition parameters are shown in Table 5.

Overall framework

Figure 7 shows an overview of the proposed multi-input SR network (ENAFSSR) based on NAFSSR [32]. This network employs a dual-network architecture, where the upper and lower networks are identical and share weights. The network takes LR paired images as input and extracts features of T2WI images I_T^{LR} and DWI images I_R^{LR} through a stack of NAFBlocks and a 3x3 convolutional layer. NAFBlock, an efficient module from NAFNet, simplifies channel attention (SCA) by removing non-linear activation functions and replacing CA/GELU with SimpleGate, maintaining image denoising performance while introducing significant gains in image deblurring. To facilitate interaction between features extracted from different modality images,

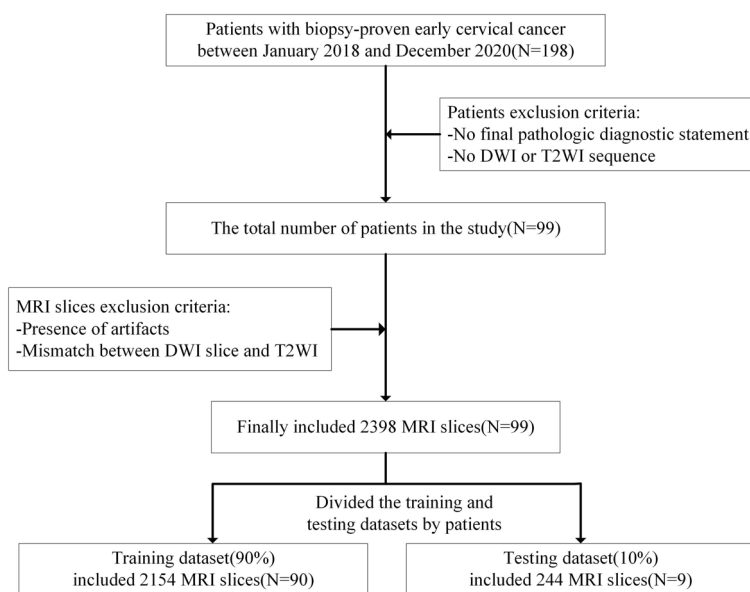


Fig. 6 The flow diagram of the cohort selection. N is the number of patients

Table 5 Details of parameters for MRI protocols

Parameter	Acquisition plane	Repetition / echo time (ms)	Sequence	Bandwidth (Hz)	Thickness (mm)	Flip angle (degrees)
T2WI	Sagittal/axial	3800–5000/40–80	Spin echo	142	5	160
DWI	Sagittal/axial	4500/76	Echo-planar imaging, Spin echo	1953	5	90

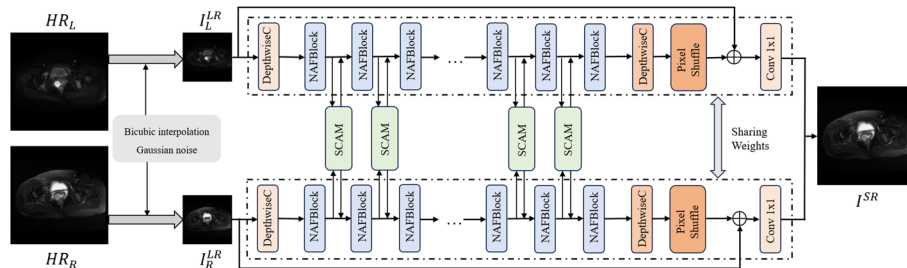


Fig. 7 An overview of the proposed multi-input SR network based on NAFSSR. SCAM (Stereo Cross-Attention Module) fuses the features of the T2WI and DWI views

the SCAM (Stereo Cross-Attention Module) is applied after each NAFBlock. Finally, the features are upsampled using 3×3 depthwise convolutions and pixel shuffle layers according to the scaling factor s . In contrast to the original NAFSSR (Non-Adaptive Feature-based Single Image SR), depthwise convolutions are used here instead of regular convolutions, and global residual learning is not employed. Depthwise convolution improves feature representation while reducing computational cost compared to traditional convolution. Instead, a semi-residual learning approach with additional convolution layers is utilized. This approach differs from traditional residual learning, which only predicts the residual between bilinearly upsampled LR images and ground truth HR images. Instead, it combines the predicted residual and the LR image in a convolutional layer.

Evaluation

Two common quantitative metrics, structural similarity index measure (SSIM), and peak signal-to-noise ratio (PSNR) are used to evaluate the SR performance of the ENAFSSR.

PSNR is the most common and widely used objective evaluation index, which is based on the error between corresponding pixels of two images, and does not take into account the human visual characteristics:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \tag{1}$$

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

where n , $MAX(I)$ represents the number of all data and the theoretical maximum of the pixel value in image I . Using the mean square error loss as loss function:

$$L_{MSE} = \frac{1}{n} \sum_{i=1}^n \|F(X_i : \theta) - Y_i\|^2. \quad (3)$$

The loss is calculated by the mean squared error between the pixels of the HR images and the original images restored by the model, where n represents the number of data, F represents the network mapping, and X_i and Y_i represent the corresponding original images and HR images, respectively.

SSIM attempts to explain the texture change between two images by calculating the similarity from the aspects of luminance, contrast, and structure. In the actual computation of the structural similarity indices of two images, we specify some localized windows, typically small blocks of $N \times N$, and compute the structural similarity indices of the signals within the windows. The window is then shifted one pixel at a time until the localized structural similarity index is computed for each position of the whole image:

$$SSIM_{(I_{SR}, I_{LR})} = l_{(I_{SR}, I_{LR})} \cdot c_{(I_{SR}, I_{LR})} \cdot s_{(I_{SR}, I_{LR})}, \quad (4)$$

where l , c , and s are functions of luminance, contrast, and structure, respectively. They are given as follows:

$$l_{(I_{SR}, I_{LR})} = \frac{2\mu_{I_{SR}}\mu_{I_{LR}} + (K_1L)^2}{\mu_{I_{SR}}^2 + \mu_{I_{LR}}^2 + (K_1L)^2}, \quad (5)$$

$$c_{(I_{SR}, I_{LR})} = \frac{2\sigma_{I_{SR}}\sigma_{I_{LR}} + (K_2L)^2}{\sigma_{I_{SR}}^2 + \sigma_{I_{LR}}^2 + (K_2L)^2}, \quad (6)$$

$$s_{(I_{SR}, I_{LR})} = \frac{\sigma_{I_{SR}I_{LR}} + (K_2L)^2/2}{\sigma_{I_{SR}}\sigma_{I_{LR}} + (K_2L)^2/2}. \quad (7)$$

Loss function and implementation details

This experiment is implemented using the Pytorch deep learning framework and adopts a mean square error loss function (MSE Loss) as the loss function for this model. The primary objective of this loss function is to quantify the disparity between the network-generated output and the actual target, aiming to minimize this difference. The MSE Loss measures the extent of the difference by calculating the squares of the difference between the network-generated image and the HR target image for each pixel value, and then averaging these squared differences to measure the degree of difference. This means that network outputs that are more similar to the target image will have lower MSE Loss values, while outputs that are less similar will result in higher Loss values. This loss function is easy to compute and optimize and encourages the network to generate images as close as possible to the target during training. An Adam optimizer with an initial learning rate of 0.003 was chosen and decreased to 1×10^{-7} by a cosine annealing strategy. The Adam optimizer uses two momentum variables to compute the adaptive learning rate. We set the exponential decay rate of the first momentum to $\beta_1 = 0.9$, and the exponential decay rate of the second momentum to $\beta_2 = 0.9$. The network was

trained on upscale factors of 2× and 4× using machines equipped with an NVIDIA RTX 3090Ti-24 G.

Author contributions

LX, CC, and YL contributed to the conception and design of the study. CC, JS, ZS and WC organized the database. LX and ML performed the statistical analysis. LX wrote the first draft of the manuscript. YL and CC wrote sections of the manuscript. All authors contributed to the manuscript revision, read, and approved the submitted version.

Funding

This work was supported by the Natural Science Foundation of Fujian Province (2021J011216), the Guide Fund for the Development of Local Science and Technology from the Central Government (2023L3019) and Joint Fund Project for Scientific and Technological Innovation of Fujian Province (2021Y9166).

Data availability

All the data used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

The studies involving human participants were reviewed and approved by The Fujian Maternity and Child Health Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Competing interests

The authors have no conflict of interest to declare that are relevant to the content of this article.

Received: 29 December 2023 Accepted: 8 August 2024

Published online: 22 August 2024

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global cancer statistics 2020: globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209–49.
2. Denny L. Cervical cancer: prevention and treatment. *Disc Med.* 2012;14(75):125–31.
3. Kuang F, Yan Z, Li H, Feng H. Diagnostic accuracy of diffusion-weighted mri for differentiation of cervical cancer and benign cervical lesions at 3.0 t: Comparison with routine mri and dynamic contrast-enhanced mri. *J Magn Reson Imaging.* 2015;42(4):1094–9.
4. Liang S, Zhang R, Liang D, Song T, Ai T, Xia C, Xia L, Wang Y. Multimodal 3d densenet for idh genotype prediction in gliomas. *Genes.* 2018;9(8):382.
5. James ML, Gambhir SS. A molecular imaging primer: modalities, imaging agents, and applications. *Phys Rev.* 2012;92(2):897–965.
6. Bammer R. Basic principles of diffusion-weighted imaging. *Eur J Radiol.* 2003;45(3):169–84.
7. Koh D-M, Collins DJ. Diffusion-weighted mri in the body: applications and challenges in oncology. *Am J Roentgenol.* 2007;188(6):1622–35.
8. Chavhan GB, Babyn PS, Thomas B, Shroff MM, Haacke EM. Principles, techniques, and applications of t2*-based mr imaging and its special applications. *Radiographics.* 2009;29(5):1433–49.
9. Katti G, Ara SA, Shireen A. Magnetic resonance imaging (mri)-a review. *Int J Dental Clin.* 2011;3(1):65–70.
10. Lin DJ, Johnson PM, Knoll F, Lui YW. Artificial intelligence for mr image reconstruction: an overview for clinicians. *J Magn Reson Imaging.* 2021;53(4):1015–28.
11. Dong S-Z, Zhu M, Bulas D. Techniques for minimizing sedation in pediatric mri. *J Magnet Reson Imaging.* 2019;50(4):1047–54.
12. Lyu Q, Shan H, Steber C, Helis C, Whitlow C, Chan M, Wang G. Multi-contrast super-resolution mri through a progressive network. *J Magnet Reson Imaging.* 2020;39(9):2738–49.
13. Liang, J., Zeng, H., Zhang, L.: Details or artifacts: A locally discriminative learning approach to realistic image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5657–5666. 2022.
14. 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 (CVPR), pp. 22367–22377. 2023.
15. Chu, S.-C., Dou, Z.-C., Pan, J.-S., Weng, S., Li, J.: Hmanet: Hybrid multi-axis aggregation network for image super-resolution. arXiv preprint [arXiv:2405.05001](https://arxiv.org/abs/2405.05001). 2024.
16. Zhou Z, Ma A, Feng Q, Wang R, Cheng L, Chen X, Yang X, Liao K, Miao Y, Qiu Y. Super-resolution of brain tumor mri images based on deep learning. *J Appl Clin Med Phys.* 2022;23(11):13758.
17. Qiu D, Zhang S, Liu Y, Zhu J, Zheng L. Super-resolution reconstruction of knee magnetic resonance imaging based on deep learning. *Computer Methods Programs Biomed.* 2020;187: 105059.
18. Oktay, O., Bai, W., Lee, M., Guerrero, R., Kamnitsas, K., Caballero, J., Marvao, A., Cook, S., O'Regan, D., Rueckert, D.: Multi-input cardiac image super-resolution using convolutional neural networks. In: Medical Image Computing and

- Computer-Assisted Intervention-MICCAI 2016: 19th International Conference, Athens, Greece, October 17–21, 2016, Proceedings, Part III 19, pp. 246–254. 2016. Springer
19. Rousseau F, Initiative ADN, et al. A non-local approach for image super-resolution using intermodality priors. *Med Image Anal.* 2010;14(4):594–605.
 20. Yurt, M., Çukur, T.: Multi-image super resolution in multi-contrast mri. In: 2020 28th Signal Processing and Communications Applications Conference (SIU), pp. 1–4. 2020. IEEE
 21. Özbey, M., Çukur, T.: Multi-image reconstruction in multi-contrast mri. In: 2021 29th Signal Processing and Communications Applications Conference (SIU), pp. 1–4. 2021. IEEE
 22. Bhadra, S., Zhou, W., Anastasio, M.A.: Medical image reconstruction with image-adaptive priors learned by use of generative adversarial networks. In: *Medical Imaging 2020: Physics of Medical Imaging*, vol. **11312**, pp. 206–213. 2020. SPIE
 23. Güngör A, Dar SU, Öztürk Ş, Korkmaz Y, Bedel HA, Elmas G, Ozbey M, Çukur T. Adaptive diffusion priors for accelerated mri reconstruction. *Med Image Anal.* 2023;88: 102872.
 24. Korkmaz Y, Dar SU, Yurt M, Özbey M, Cukur T. Unsupervised mri reconstruction via zero-shot learned adversarial transformers. *IEEE Trans Med Imaging.* 2022;41(7):1747–63.
 25. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1874–1883. 2016.
 26. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2472–2481. 2018.
 27. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change Loy, C.: Esrgan: Enhanced super-resolution generative adversarial networks. In: *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pp. 0–0. 2018.
 28. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1833–1844. 2021.
 29. Just N. Improving tumour heterogeneity mri assessment with histograms. *Br J Cancer.* 2014;111(12):2205–13.
 30. Gong L, Wang M, Shu L, He J, Qin B, Xu J, Su W, Dong D, Hu H, Tian J, et al. Automatic captioning of early gastric cancer using magnification endoscopy with narrow-band imaging. *Gastrointestinal Endoscopy.* 2022;96(6):929–42.
 31. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-assisted Intervention*, pp. 234–241. 2015. Springer
 32. Chu, X., Chen, L., Yu, W.: Nafsr: Stereo image super-resolution using nafnet. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1239–1248. 2022.

Publisher's Note

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