J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 6421--6437

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Optimization of endometrial tolerance ultrasound image reconstruction algorithm supported by machine vision technology

Yun Wang[1,✉], Yabiao Zeng[2]

[1] *Ultrasound Department, Hunan Maternal and Child Health Hospital, Changsha, Hunan, 123456, China*

[2] *General Surgery Department, Changsha Hospital of Traditional Chinese Medicine, Changsha, Hunan, 123456, China*

ABSTRACT

In today's society, hospitals are treated with images generated by medical examination equipment for disease diagnosis, and high-resolution images can greatly improve the accuracy of doctors' disease diagnosis. The study constructs an ultrasound image dataset US-Dataset suitable for the task of super-resolution reconstruction of ultrasound images. Based on this ultrasound image dataset, a degradation model is proposed, which in turn constructs ultrasound image matching pairs containing high - low resolution images for training the model proposed in this paper. To improve the perceptual quality of endometrial images, a super-resolution reconstruction model UN-SRGAN based on generative adversarial network is proposed in this paper. The network structure of this model consists of a generator and a discriminator. To validate the effectiveness of the model proposed in this paper, it is evaluated on Accuracy, Precision, Recall, Specificity, and F1-score metrics. The proposed model achieves the lead on PSNR and SSIM metrics and subjective quality evaluation, and the UN-SRGAN model has an accuracy of 0.9721, which is better than the other models, verifying the effectiveness of the model.

Keywords: ultrasound image; generative adversarial network; super-resolution reconstruction; assisted diagnosis

1. Introduction

The 3 necessary conditions for success after assisted reproductive technology (ART) embryo transfer are embryos with transfer potential, endometrium in a permissive state and synchronized development of the embryo and endometrium [1]. Among them, endometrial influence accounts for 2/3 of the factors of embryo transfer failure and is crucial for the success of embryo transfer. Some studies have shown that there are many factors affecting endometrial tolerance, including uterine

✉ Corresponding author.
E-mail address: Mier0628@163.com (Y. Wang).
Received 25 March 2024; Revised 05 June 2024; Accepted 15 December 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-352

cavity diseases, other reproductive site diseases and systemic endocrine diseases [2-3]. They directly or indirectly cause, to varying degrees, alterations in endometrial immunity, levels of metabolic substances, and regulation and expression of genes related to the transcriptome, which in turn decreases endometrial tolerance, leading to infertility and adverse pregnancy outcomes [4-7]. Therefore, effective detection and evaluation of endometrial tolerance has become a hot topic in assisted reproductive technology research in recent years.

Diagnosis of endometrial tolerance is challenging, and recent studies have identified multiple markers of endometrial tolerance associated with clinical pregnancy outcomes, with markers assessed primarily by ultrasound, endometrial biopsy, hysteroscopy, and endometrial fluid aspiration [8]. Among them, histologic and transcriptomic markers, such as cytotrophoblast synapses and endometrial tolerance arrays, require endometrial biopsy to obtain the results, which is an invasive and costly procedure, and embryo transfer must be performed in a non-biopsy cycle, so that biopsy results have to be consistent from cycle to cycle, which is poorly accepted by the patients and limited in clinical implementation [9-12]. In contrast, ultrasound can assess endometrial changes during stimulation cycles, and it also has minimal inter- and intra-observer variability [13-14]. In addition, 3D color Doppler ultrasound technology has been developing rapidly in recent years, with increasing resolution, and it also allows the measurement of previously unmeasurable parameters [15-16]. Secondly, transvaginal ultrasound is noninvasive. Ultrasound measurements take a short time and can be performed during the measurement cycle for embryo transfer. Finally, ultrasound measurements are simple, inexpensive, and easily disseminated [17-18]. Based on these advantages, ultrasonography is widely used in research and clinical care in the field of reproductive medicine.

The study proposes a super-resolution reconstruction model for medical ultrasound images based on UN-SRGAN for the perceived quality of endometrial images. The model consists of a generator and a discriminator. The ultrasound image dataset US-Dataset, which is suitable for the task of ultrasound image super-resolution reconstruction, is first constructed and used to train the model proposed in this paper. Then the characteristics of UNet++ and the reason why UNet++ is used as a discriminator are introduced, and various types of mainstream super-resolution reconstruction models are selected to do comparison experiments with UN-SRGAN, which are evaluated on Accuracy, Precision, Recall, Specificity, and F1-score metrics, to validate the validity of the model.

2. Endometrial tolerance ultrasound image dataset construction

2.1. Raw data sets

There are a total of 637 ultrasound images of thyroid nodules in the DDTI public database and a total of 789 endometrial ultrasound images in the BUSI public database, which includes data on normal, benign, and malignant endometrial images, respectively. The two databases totaled 1,426 ultrasound images, which together were constructed into the ultrasound image dataset used in this study: the US-Dataset.

2.2. Benchmarking data sets

In the field of super-resolution research, there are several commonly used classical benchmark test datasets: Set5, Set14, Urban100, BSD100, at present, basically all the super-resolution related research will be tested and validated experiments on these test datasets, but due to the fact that these test datasets are natural images, they can't be applied to the research of this paper. Therefore, this paper constructs several ultrasound image benchmark test datasets applicable to ultrasound image research, in which the ultrasound images are all derived from open-source medical image databases, and are categorized into three types according to the swept parts of the ultrasound images: the thyroid ultrasound image benchmark dataset SetT100, the breast ultrasound image benchmark dataset SetB100, and the endometrial ultrasound image benchmark dataset SetCA100.

2.3. Image Matching Construction

2.3.1. Image cleanup. In order to obtain more high-resolution ultrasound images, this paper tries to generate noise-free and high-quality ultrasound images from existing ultrasound image datasets. Specifically, double triple downsampling can be applied to the low resolution ultrasound images to remove the noise and make the ultrasound images clearer and show more texture details. The downsampling formula is shown below:

$$I_{HR} = (I_{src} * k_{bic}) \downarrow_{sc} \tag{1}$$

where k_{bic} is the ideal double cubic kernel and sc is the reduction factor used for image downsampling.

2.3.2. Classical degradation models. Recovering high-resolution images from low-resolution images with unknown and complex degradations usually requires the use of classical degradation models to synthesize the inputs of low-resolution images. In general, the real image y is first convolved with the fuzzy kernel k. Then a downsampling operation with a reduction factor of r is performed. The low resolution image x is obtained by adding noise n. JPEG compression is also applied to the last step as it is widely used in real world image transmission.

$$x = D(y) = \left[(y \otimes k) \downarrow_r + n \right]_{JPEG} \tag{2}$$

where D denotes the degradation process.

1) Fuzzy. Fuzzy degradation is usually modeled as a convolution operation with a linear fuzzy filter. Isotropic and anisotropic Gaussian filters are common options. For a Gaussian fuzzy kernel k with kernel size $2t+1$, the elements obey a Gaussian distribution of the form:

$$k(i,j) = \frac{1}{N}\exp(-\frac{1}{2}C^T\Sigma^{-1}C), C = [i,j]^T \tag{3}$$

Where $(i,j) \in [-t,t]$; Σ is the covariance matrix; C is the spatial coordinates; and N is the normalization constant. The covariance matrix can be further expressed in the following equation:

$$\Sigma = R\begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}R^T = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \tag{4}$$

where R is the rotation matrix, σ_1 and σ_2 are the standard deviations (i.e., eigenvalues of the covariance matrix) along the two principal axes; and θ is the degree of rotation. When $\sigma_1 = \sigma_2$, k is the isotropic Gaussian fuzzy kernel; when $\sigma_1 \neq \sigma_2$, k is the anisotropic Gaussian fuzzy kernel.

In this paper, further distributions of generalized Gaussian fuzzy kernel and plateau-shaped kernel are considered. Their probability density functions are given in the following equations, respectively:

$$\frac{1}{N}\exp(-\frac{1}{2}(C^T\Sigma^{-1}C)^\beta \tag{5}$$

$$\frac{1}{N}\frac{1}{1+(C^T\Sigma^{-1}C)^\beta} \tag{6}$$

where β is a shape parameter, it has been shown that if these fuzzy kernels are included in the degradation model, it can make some ultrasound image samples produce sharper output images.

2) Noise. Noise is generally divided into two categories: 1) Additive Gaussian noise 2) Poisson noise.

The density function of additive Gaussian noise is Gaussian distribution. The noise intensity is controlled by the standard deviation (i.e. sigma value) of the Gaussian distribution. When each channel of an RGB image has an independent sampling noise, the synthesized noise is colored noise.

Poisson noise follows a Poisson distribution. It can simulate the sensor noise caused by the statistics over the rise and fall, that is, the number of photons can be sensed in the known exposure level of the change.

3) Downsampling. Downsampling is a basic operation for synthesizing low-resolution images in super-resolution problems. Generally both down-sampling and up-sampling operations, i.e., image scaling operations, are considered. Different image scaling algorithms can have different effects, some produce blurry image results while others may output images with overshooting artifacts or sharp edges. Since nearest-neighbor interpolation introduces unaligned problems, it is excluded from this paper, and only the bilinear interpolation operation and the bicubic interpolation operation are considered.

4) JPEG compression. JPEG compression is one of the most common image compression techniques. The first step in the compression process is to transform the image into YCbCr color space, followed by downsampling its chrominance channel, and then dividing the current image into 8×8. All image blocks must first complete the two-dimensional discrete cosine transform, and then quantize the two-dimensional discrete cosine transform coefficients.JPEG compression usually introduces artifacts of interference.

The quality of a compressed image is determined by the quality factor $q \in [0,100]$, where a smaller q indicates a higher compression rate and poorer quality. In this paper, PyTorch deep learning framework is used to implement DiffJPEG.

2.3.3. Higher-order degradation models:

1) Model Composition. When the classical degradation model described above is used to synthesize ultrasound image matching pairs, the trained model can indeed handle some samples. However, it still cannot solve some of the more complex degradation problems in ultrasound images, especially unknown noise and complex artifacts. This is due to the fact that the synthesized low-resolution images are still somewhat different from the real degraded ultrasound images. Therefore, in this paper, we consider extending the classical degradation model to a higher-order degradation model in order to simulate a more realistic image degradation path for ultrasound images. The classical degradation model usually contains only a fixed number of basic degradations and can be regarded as a first-order degradation model.

However, the degradation process of real ultrasound images is diverse, and this complex deterioration process cannot be modeled simply by using only the classical first-order model. Therefore, a higher-order degradation model is proposed in this paper. A n st order model involves n repeated degradation processes, where each degradation process is modeled using the classical degradation model with the same process but different hyperparameters, as in the following equation:

$$x = D^n(y) = (D_n \circ \cdots \circ D_2 \circ D_1)(y) \tag{7}$$

“Higher order” in this context is not the same as the definition of a mathematical function. It mainly refers to the execution time of the same operation. In order to keep the resolution of the image within a reasonable range, the downsampling operation of the classical degradation model is replaced with a random image scaling operation. Based on empirical values, the second-order degradation process is used in this paper because it can resolve most of the samples while maintaining simplicity.

Fig. 1 shows the pipeline of higher-order degradation models described in this subsection.

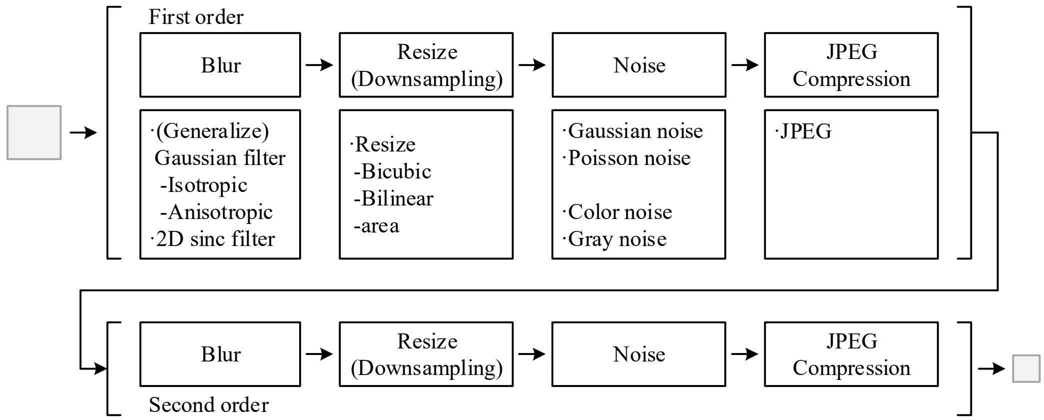


Fig. 1. Higher order degradation model

2) Ringing and overshoot artifacts. In this paper, we choose to use sinc filter, which is an idealized filter that cuts off the high frequencies to synthesize ringing and overshooting artifacts for the image matching pairs. sinc filter kernel can be expressed as:

$$k(i,j)=\frac{\omega_c}{2\pi\sqrt{i^2+j^2}}J_1(\omega_c\sqrt{i^2+j^2}) \tag{8}$$

where (i,j) is the kernel coordinate; ω_c is the cutoff frequency; and J_1 is the first-order Bessel function of the first class.

3. UN-SRGAN based ultrasound image reconstruction

3.1. Generating Adversarial Network Models

The core idea of GAN is to generate data that is similar to the distribution of real data, such as images, text, audio, etc., through the adversarial process of two neural network modules, i.e., the generator and the discriminator. In a generative adversarial network, the generator's task is to reconstruct data as realistic as possible to mislead the discriminator, while the discriminator's task is to accurately discriminate between real and generated data to the best of its ability, and the respective abilities of the generator and the discriminator are gradually improved during the alternating training process [19]. The game between generator G and discriminator D can be described as a well-known adversarial min-max optimization problem, i.e., the original objective function is:

$$\min_G \max_D E_{I^{HQ}\sim P_{HQ}}[\log D(I^{HQ})]+E_{I^{LQ}\sim P_{LQ}}[1-D(G(I^{LQ}))] \tag{9}$$

In the above equation: P_{HQ} and P_{LQ} denote the data distribution of high quality images and low quality images respectively. Both G and D are trained in this process until the discriminator D is unable to determine the generated data and the target data. The whole training process is automatic without any human interference. After the training process, the generator G can reconstruct the new high quality image directly from the low quality image.

The training process of GAN network can be generally divided into the following steps:

1) First, initialize the parameters of the generator G and discriminator D. This step is generally performed by random initialization or pre-training weights.

2) Prepare a set of data x from the target dataset as a positive sample with label 1.

3) Obtain a set of noise vectors z using the noise distribution, and use them as inputs to the generator to generate a batch of fake data G(z) as negative samples with label 0.

4) Mix the positive and negative samples and feed them to the discriminator D. Compute the loss function of the discriminator. The purpose of calculating the loss function is to make the discriminator

distinguish between true and false data as accurately as possible. That is, to maximize the value of D(x) and minimize the value of D(G(z)).

5) Keep the parameters of the generator unchanged and update the parameters of the discriminator in order to reduce the loss function of the discriminator and improve its capability.

6) Sample a batch of noise vectors z from the noise distribution again to generate a batch of fake data G(z).

7) Feed the generated fake data into the generator G. Calculate the loss function of the generator, generally using binary cross entropy or Wasserstein distance, etc., so that the fake data generated by the generator is as close as possible to the distribution of the real data. That is, to maximize D(G(z)).

8) Fix the parameters of the discriminator and adjust the parameters of the generator in order to reduce the loss function of the generator and improve the ability of the generator.

9) Repeat steps (2)-(8) to iterate the training, and stop the training when a balance is reached between the generator and the discriminator or when the desired effect is achieved.

3.2. UNet++ and frequency domain information of ultrasound images

3.2.1. UNet++. $UNet{+}{+}$ is a convolutional neural network improved on $UNet$ with better sensory field and contextual information aggregation capability for medical ultrasound images that require higher accuracy. The network structure of $UNet{+}{+}$ is shown in Fig. 2. The core of $UNet{+}{+}$ is a step-by-step upsampling and downsampling network structure, similar to the encoder-decoder structure of $UNet$. The encoder and decoder of $UNet{+}{+}$ each consist of a number of successive submodules, each of which contains a convolutional layer and a pooling or upsampling layer. Where each submodule of the encoder downsamples the input feature map to a smaller resolution, each submodule of the decoder upsamples the input feature map to a larger resolution [20]. This structure allows $UNet{+}{+}$ to perform feature extraction and feature fusion at multiple scales to better capture the contextual information of the image.

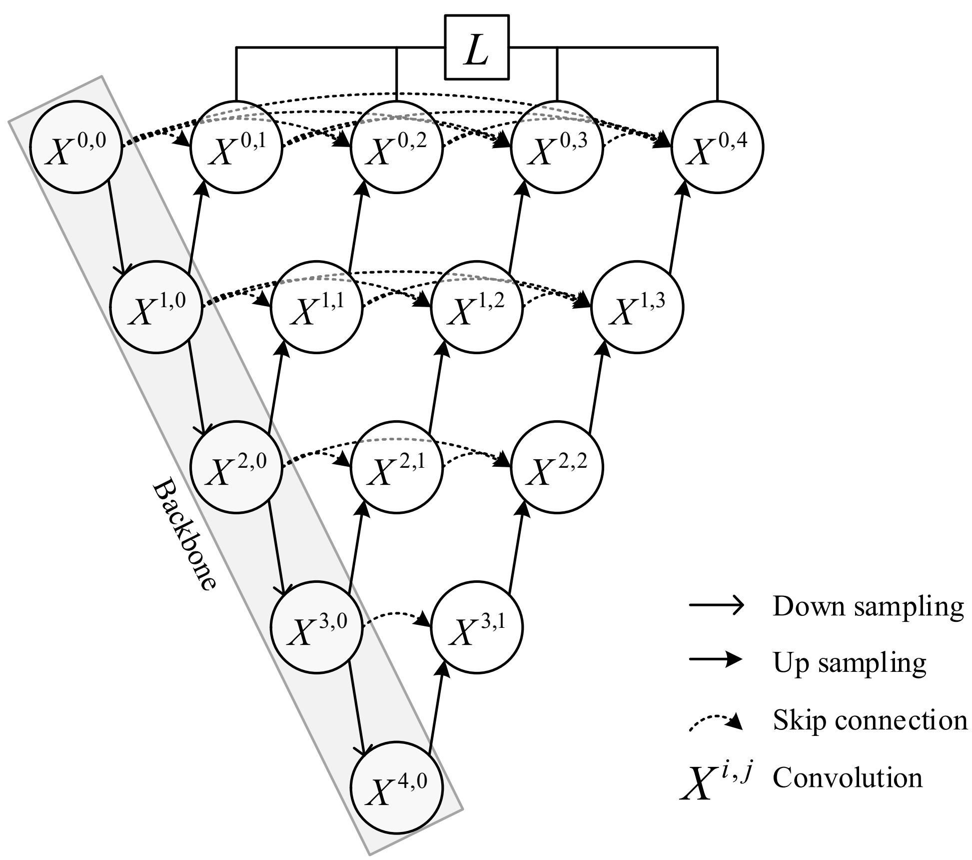


Fig. 2. Unet++ network structur

In addition, $UNet{+}{+}$ introduces a mechanism called Nested U-Net for enhanced feature representation. Nested U-Net adds a number of jump connections to $UNet{+}{+}$, connecting different levels of the feature maps in the encoder and the decoder, allowing the decoder to up-sample and reconstruct using features from different levels of the encoder. This mechanism allows $UNet{+}{+}$ to better capture the detailed information in the image and improve the classification accuracy.

$UNet{+}{+}$ is a network structure that mitigates the gradient vanishing problem, specifically, to prevent the gradient from decreasing as the depth of the network increases, gradient information is passed to all subsequent layers during forward propagation. In addition, $UNet{+}{+}$ introduces a feature sharing mechanism that allows the network to generate a large number of features with only a small number of convolutional kernels, thus reducing the model parameters and making the network easier to train.

It is due to the various excellent features of $UNet{+}{+}$ that this chapter uses the $UNet{+}{+}$ network as the discriminator of the model in the design of UN-SRGAN.

3.2.2. Frequency domain information of ultrasound images. In ultrasound images, acoustic signals reflected, scattered and refracted by different tissues have different frequencies and amplitudes. By performing a Fourier transform on the ultrasound signal, it can be transformed into the frequency domain space. In the frequency-domain space of an ultrasound image, the higher frequency components represent variations in the microstructure of the tissue and various features such as edges and textures. At the same time, the frequency domain space also contains information about disturbances such as noise and artifacts. Therefore, the frequency domain space provides important information about the structure of the tissue, which has a positive impact on the super-resolution reconstruction of ultrasound images and can improve the image quality.

The 2D Fourier transform can be calculated from a 2D ultrasound image for its frequency domain information. The Fourier transform is a linear operation with symmetry mainly used for the analysis and processing of signal-related problems [21]. The 2D Fourier transform formula is given in the following equation:

$$F(u,v)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)e^{j2\pi(ux+vy)}dxdy \tag{10}$$

where u is the spatial frequency domain in the x direction and v is the spatial frequency domain in the y direction. $F(u,v)$ is used to represent the two-dimensional spectrum obtained by two-dimensional Fourier transform of the image $f(x,y)$.

The frequency domain signals obtained from the 2D Fourier transform of the real high resolution image and the generator reconstructed image can be used as additional constraints on the real high resolution and generator reconstructed images, which in turn can be used to construct separate frequency domain loss functions and combine them into the UN-ESRGAN network in order to improve the quality of the image and maintain the details.

3.3. Algorithm Design and Implementation

3.3.1. Degeneration models. Classical degeneracy models contain only a fixed number of basic degeneracies, which can be regarded as first-order degeneracies. As in equation (11) the following operations are performed in sequence.

$$I_{LR}=D(y)=\left[(I_{HR}\otimes\kappa)\downarrow_r+n\right]_{JPEG} \tag{11}$$

Where, I_{LR} represents the degraded image; D represents the degradation function; I_{HR} represents the original image; κ represents the fuzzy kernel; $\downarrow_r$ represents the scaling ratio from the high-resolution I_{HR} downsampling to the low-resolution image I_{LR}, and the commonly used methods are bicubic interpolation, bilinear interpolation, and regional interpolation; n represents the added noise, including Gaussian noise and noise obeying the Poisson distribution, etc.; and JPEG stands for the compression is performed, and the compression quality is usually selected from the range of [0, 100] range to select the compression quality and perform JPEG compression on the image, where 0 represents the worst quality after compression and 100 represents the best quality after compression.

Networks usually suffer from the use of data pairs obtained by degrading the classical degradation model to such an extent that they do not work well with real samples, especially in the presence of unknown noise and complex artifacts. Therefore, in order to better cope with degradation problems in practice, extensions to higher-order degradation models are needed.

The higher-order degradation model contains multiple repeated degradation processes, which are defined as described in equation (12) below:

$$I_{LR} = D^n(I_{HR}) = (D_n \ldots D_1)(I_{HR}) \tag{12}$$

Specifically, for each image a plurality of complete degradations are performed, including blurring, scaling, noise addition and compression operations. Wherein each stage employs the same degradation process but with different degradation hyper-references.

3.3.2. Generator section. UN-ESRGAN refers to the generator structure in ESRGAN, which mainly consists of an input part, a feature extraction part, and an upsampling output part.

The generator of SRGAN adopts the residual block as the basic unit, which consists of two convolutional layers, two BN layers and one activation layer. And ESRGAN improves this residual block by proposing the residual dense block RRDB.RRDB adopts the structure of Residual-in-Residual and introduces dense connections. It has been shown that the BN layer adversely affects the brightness, contrast and other information of the image, so RRDB removes the BN layer. And Residual-in-Residual was originally proposed by ESRGAN, which alleviates the problem of gradient explosion or gradient vanishing in deep residual networks and helps to train deeper networks.

The generator proposed in this chapter similarly removes the BN layer as well as adopts the structure of Residual-in-Residual for better reconstruction. The overall structure of the RRMFFB module used is shown in Fig. 3.The RRMFFB module consists of three MFFB blocks which are connected using Residual-in-Residual and a training strategy of residual scaling is used in order to further deepen the network.

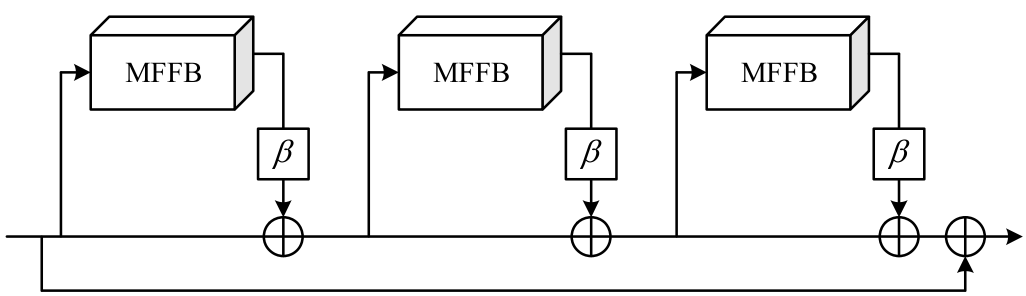


Fig. 3. RRMFFB module structure diagram

3.3.3. Discriminator section. The discriminator in SRGAN judges the probability that the input periimage is real and natural only for the input periimage, whereas the discriminator proposed in the accompanying ESRGAN tries to compute the probability that the real image is more realistic with respect to the reconstructed image, and this tuning will allow the mesh to learn more distinct edges and finer texture details.

In this paper, a $UNet++$ design with jump connections is used to improve the discriminator structure in UN-SRGAN. The $UNet++$ structure is able to output realism values for each pixel and provide detailed feedback to the generator for each pixel.

3.3.4. Loss function:

1) Adversarial loss function. The training process of the generator network, for the adversarial loss function of the generator network, the purpose is to supervise the generator to generate as real as possible output images to deceive the discriminator. The specific formula is defined as follows:

$$L_{adv}^{G}=\frac{1}{N}\sum_{i=1}^{N}\left[-\log D(G(I_L;R;\theta))\right] \tag{13}$$

Similarly, for the loss function of the discriminator, the goal is to ensure that the discriminator network can accurately distinguish and judge the reconstructed HR image from the real HR image. Specifically, the loss function aims to minimize the discriminative loss of the reconstructed image and maximize the discriminative loss of the real image, and consists of the discriminative loss of the reconstructed image and the discriminative loss of the real image. The corresponding formula is defined as follows:

$$L_{adv}^{D}=-\frac{1}{N}\sum_{i=1}^{N}\left[\log D(I_{HR})\right]-\frac{1}{N}\sum_{i=1}^{N}\left[1-\log D(G(I_{LR};\theta))\right] \tag{14}$$

where I_{LR} denotes the low-resolution image input to the generator network, $D(G(I_{LR};\theta))$ the ultrasound image after super-resolution reconstruction $G(I_L R;\theta)$ the probability value of being judged true by the discriminator D. I_{HR} denotes the corresponding true value high resolution map, θ is the parameter of the model, and N is the number of images in a training batch.

2) Content loss function. The content loss function is used to measure the difference between the true image and the reconstructed high-resolution image so that the generated image is as close to the true as possible. We use the mean absolute error as the content loss function to improve the training effect and the quality of the generated images. By minimizing the MAE, the generated high-resolution images maintain better pixel-level consistency and are closer to the real scene, while promoting stable model convergence. The formula for the content loss function is as follows:

$$Loss_{MAE}^{G}=\frac{1}{r^2WH}\sum_{x=1}^{rW}\sum_{y=1}^{rH}\left|I_{HR}-G(I_{LR};\theta)\right| \tag{15}$$

Where, r is the magnification factor, rW is the width of the high resolution image and rH is the height of the high resolution image.

3) Perceptual loss function. The network structure of VGG19 is very simple, with a total of 19 layers of network, consisting of five convolutional modules, which are downsampled using a 2×2 maximal pooling layer between the modules, and a fully connected layer as the last layer. In each convolutional module, multiple 3×3 convolutional kernels are used for convolutional operations. This design allows VGG19 to efficiently extract features at different levels of the image. The perceptual loss function can be expressed as the following equation:

$$Loss_{VGG}^{G}=\frac{1}{hwc}\sqrt{\sum_{i,j,k}(\phi_{i,j,k}^{l}(I_{HR})-\phi_{i,j,k}^{l}(I_{SR}))^2} \tag{16}$$

Where h,w and c denote the length, width and number of channels of the feature map, respectively, and $\phi^{l}(\cdot)$ is the output feature map of layer l in VGG19.

4) Frequency domain loss function. The UN-SRGAN model introduces a frequency domain loss function to better utilize the frequency domain information of ultrasound images. This loss function is used to measure the difference in frequency domain signals between the real high-resolution image and the

high-resolution image reconstructed by the generator, in order to reconstruct a high-resolution image that is more consistent with the frequency domain information of the real image. The frequency domain loss function is calculated as follows:

$$L_f^G = \| F(I_{SR}) - F(I_{HR}) \|_2^2 \tag{17}$$

5) Full variance loss function. The full variational loss function is the absolute value of the difference between neighboring pixels. The full variance loss function is introduced due to the presence of many noises in the HR image reconstructed by the generator. This function is able to achieve a certain degree of smoothing in regions of the image where there are large pixel differences, thus reducing the appearance of noise in the generated high-resolution image, especially in regions with more pronounced noise. The formula for the full variational loss function is as follows:

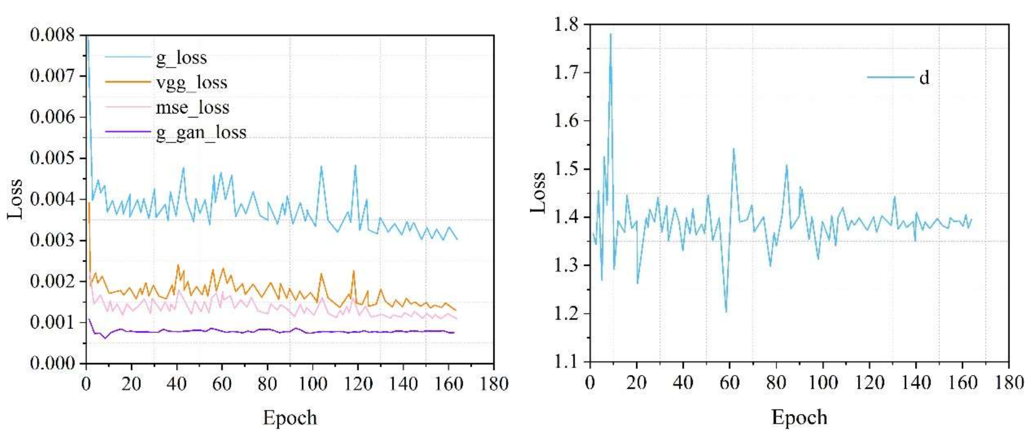
$$Loss_{TV}^G = \sum_{i,j} \sqrt{| G(I_{LR})_{i+1,j} - G(I_{LR})_{i,j} |^2 + | G(I_{LR})_{i,j+1} - G(I_{LR})_{i,j} |^2} \tag{18}$$

where G denotes the generator, and i and j denote the horizontal and vertical coordinates of any point on the reconstructed high resolution, respectively.

4. Experimental results and analysis

4.1. Model training

The generator loss and discriminator loss curves for UN-SRGAN-based ultrasound image training are shown in Fig. 4, with g_loss_curve on the left as the generator loss curve. The pink curve is mse loss mse_loss, the curve oscillates and decreases; the orange curve is vgg loss vgg_loss, the curve oscillates and decreases; the purple curve is the perception loss gan_loss, the relative change is relatively smooth; the final generator loss is the blue curve g_loss, $g\_loss = mse\_loss + vgg\_loss + g\_gan\_loss$. The right figure d_loss_curve is the discriminator loss curve, the oscillation in training is violent, and finally converges to 1.41. The generative adversarial network training is very unstable, the loss function produces violent oscillation in the training process After 160 iterations, the generator loss decreases gently, the discriminator loss gradually stabilizes, and the training ends.



(a) g_loss curve (b) d_loss_curve

Fig. 4. SRGAN loss function convergence curve

4.2. Experimental results and analysis

4.2.1. Evaluation indicators. In this chapter, the effect of increasing the dataset on the classification results of endometrial ultrasound images is evaluated by designing relevant experiments, and the precision, recall, specificity, and confusion matrix are used as the evaluation metrics of the classification results.

The above metrics can be calculated as follows:

$$Precision = \frac{TP}{TP + FP} \tag{19}$$

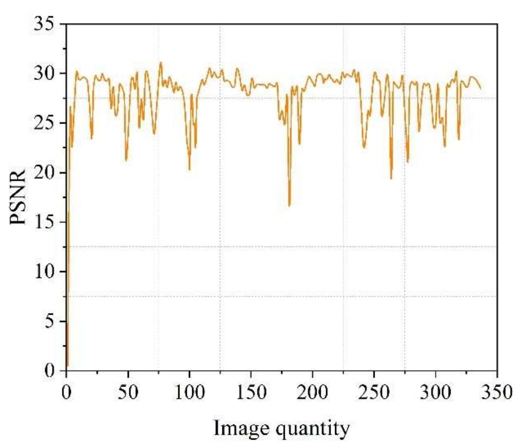
$$Recall = \frac{TP}{TP + FN} \tag{20}$$

$$Specificity = \frac{TN}{TN + FP} \tag{21}$$

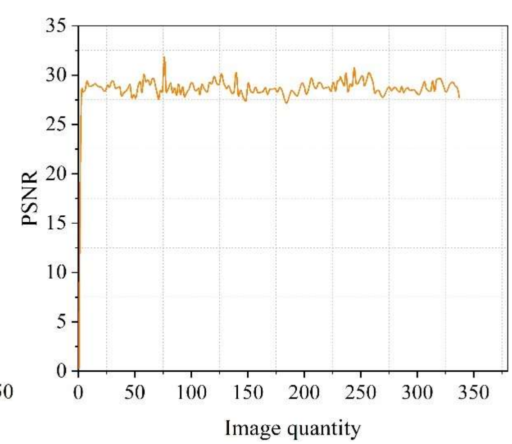
F1 is calculated based on precision and recall, where precision is denoted as P and recall as R. The formula for F1 is:

$$F1 = \frac{2 \times P \times R}{P + R} \tag{22}$$

4.2.2. Ablation experiments. In order to verify the performance improvement of the UNet++SRGAN generator module and GAN module for the network, ablation experiments are conducted in this section to analyze the effect of different modules on the reconstruction results. In order to verify the effectiveness of UN-ESRGAN for improving the network generator, the comparison results of the ablation experiments are shown in Table 1, and the comparison of the PSNR value curves of the experimental results are shown in Fig. 5, and the comparison of the SSIM value curves are shown in Fig. 6. Also combined with Table 1, it can be seen that when the generator adopts the UNet++ network structure, the PSNR value and SSIM value obtained from the experimental results are more stable and the reconstructed image is more realistic. Therefore, finally the super-resolution reconstruction network generator in this chapter uses UNet++ network.

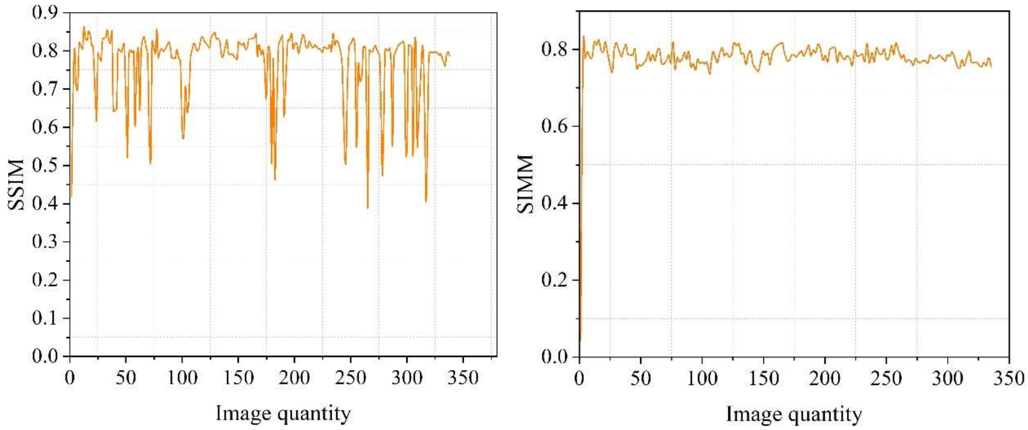


(a) UNet



(b) UNet++

Fig. 5. Generator Contrast (PSNR)



(a) UNet (b) UNet++

Fig. 6. Generator Comparison (SSIM)

In order to verify the effectiveness of the GAN module for improving the network performance, this section is designed by designing experiments in which ResNet34, ResNet50 and ResNet101 are used in the comparison experiments. The results in the table show that as the number of layers of the network deepens, the values of PSNR and SSIM are both improved. At the same time, the standard deviation of the ablation experiment results also decreased, which indicates that (1) the robustness of the model gradually improves, i.e., the model is more resistant to the volatility and disturbance of the data, and the reconstruction is more reliable; and (2) the effectiveness of the model optimization gradually strengthens, i.e., the model's performance becomes more and more excellent in each ablation experiment, and the reconstruction performance of the network gradually improves. Therefore, the GAN module is finally selected as the perceptual loss network in this chapter.

In summary, adding UNet++ module and GAN module are both beneficial to image reconstruction, and the feasibility and effectiveness of this paper's algorithm can be fully proved by the ablation experiments.

Table 1. Comparison of ablation experimental results

Method	PSNR	SSIM
GAN	29.556±0.8288	0.765±0.0241
GAN+UNet++	29.5634±0.8275	0.78±0.028
GAN+ResNet34	29.5126±0.8452	0.78211±0.0253
GAN+ResNet34+UNet++	29.5215±0.8011	0.78509±0.0237
GAN+ResNet50+UNet++	29.5415±0.8175	0.78522±0.0239
GAN+ResNet101+UNet++	29.5579±0.8091	0.78589±0.0238

4.2.3. Comparative analysis of reconstruction effects. In this chapter, 20 endometrial ultrasound images from the in-house dataset were selected to be used as tests and there was no overlap between the datasets used for training and testing.

For the first experiment, this paper tested the performance of the original SRGAN model for super-resolution reconstruction of ultrasound images with 5 residual blocks and 4 residual blocks using the data mentioned above. The VGG19 network trained with natural images contains most of the image features, and for ultrasound images with only one gray level channel, the filter depth in the first convolutional layer of VGG19 is reduced from 3 to 1 by averaging the weights along the depth direction. The average peaks of the reconstructed images comparing the two cases are shown in Fig. 7. In the case of signal-to-noise network convergence, a small number of residual blocks are still able to maintain the super-resolution performance similar to the previous approximation. Therefore, in this paper, the SRGAN network with 4 residual blocks is used for the super-resolution reconstruction work, which reduces the complexity of the network and speeds up the training process at the same time. A small number of residual blocks is important for real-time super-resolution ultrasound imaging.

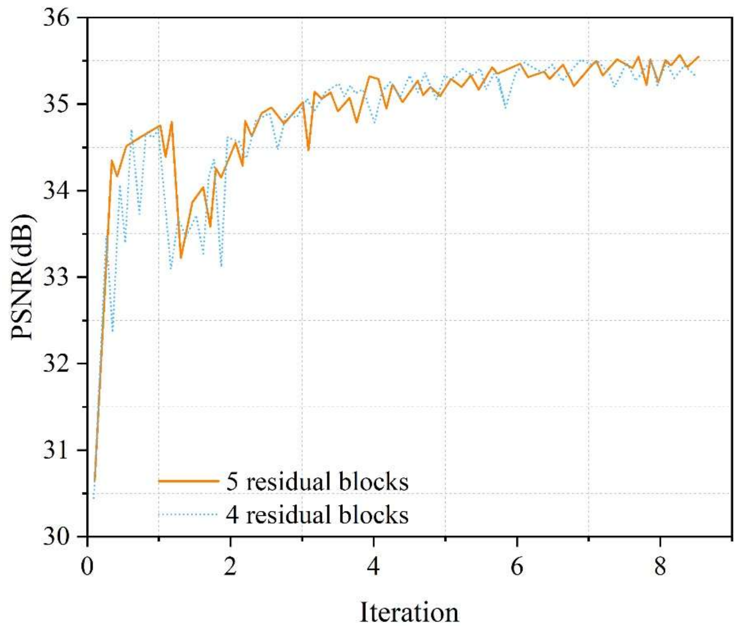


Fig. 7. The value of the PSNR in the SRGAN difference block

Next, the remaining 18 sets of uterine ultrasound images were reconstructed with super-resolution using the bicubic algorithm, the SRGAN model, and the UN-SRGAN model, respectively, and then the PSNR values of the reconstructed images from the three algorithms were calculated, as shown in Fig. 8. From the figure, it can be seen that the UN-SRGAN model produces results with significantly higher PSNR values than the bicubic algorithm and the SRGAN model.

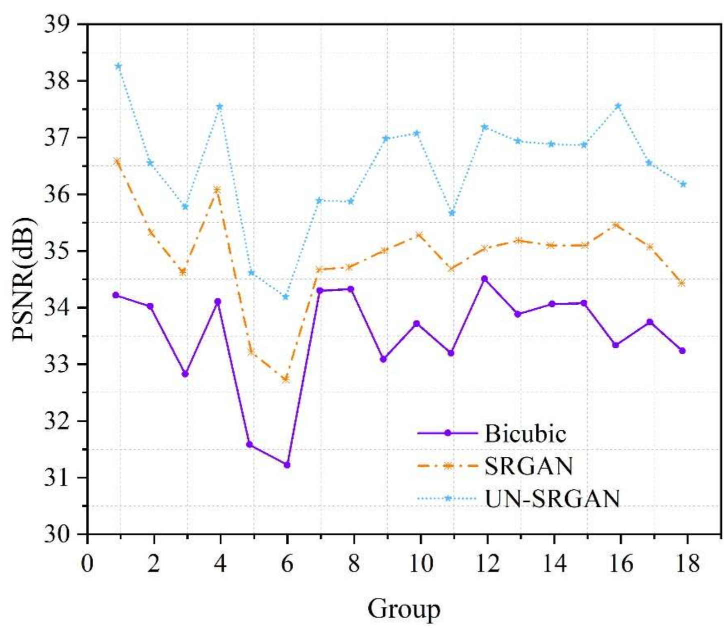


Fig. 8. The three methods of PSNR are compared

The SSIM values of the reconstruction results of these 18 sets of data using the three algorithms are calculated and the results are shown in Figure 9. From the figure, it can be seen that the SSIM values of the results produced by the UN-SRGAN model are significantly higher than those of the bicubic algorithm and the SRGAN model.

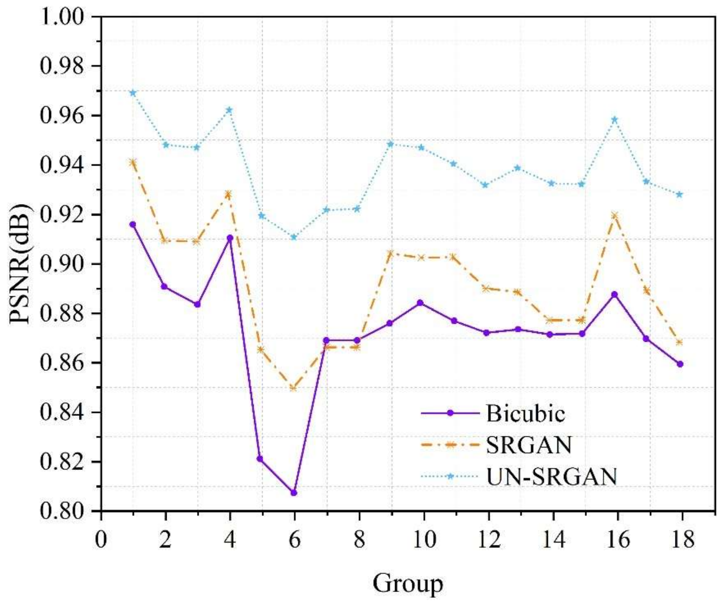


Fig. 9. The three methods of the SSIM value contrast

4.3. Analysis of classification results

In terms of classification, the results of the main classification indexes are shown in Table 2 when compared with the individual models introduced and the distilled ResNet18+LIP model. From the table, it can be seen that in the classifier, the combination of ResNet18+LIP with smaller volume adopted by UN-SRGAN has more obvious improvement in all classification indexes. Among them, the improvement in Accuracy and Specificity is more limited, with 1.8% and 1.1% respectively compared to ResNet18+LIP. On the other hand, the UN-SRGAN classifier part has a significant improvement in Recall and Precision, which are 6.4% and 6.4% compared with ResNet18+LIP, respectively. The accuracy improvement of the model comes from the fact that the generator produces more valid samples to provide data for classification on one hand, and on the other hand, the antagonistic relationship between the generator and the classifier provides new directions and environments to further improve the model performance. As for F1, the UN-SRGAN model has a more refined classification target, and it has a better classification performance compared with most classification models.

Table 2. Classification model index comparison

Method	Accuracy	Precision	Recall	Specificity	F1-Score
ResNet100	0.951	0.957	0.954	0.947	0.955
InceptionV3	0.939	0.951	0.947	0.942	0.948
DenseNet121	0.951	0.946	0.957	0.966	0.951
ResNet50+LIP	0.973	0.978	0.968	0.975	0.963
ResNet18+LIP	0.968	0.924	0.918	0.967	0.951
UN-SRGAN	0.986	0.988	0.982	0.978	0.975

The ROC curves of each classification model are shown in Fig. 10, from which it can be seen that the UN-SRGAN model has an accuracy of 0.9721, which is better than other model constraints for multi-tasking purposes.

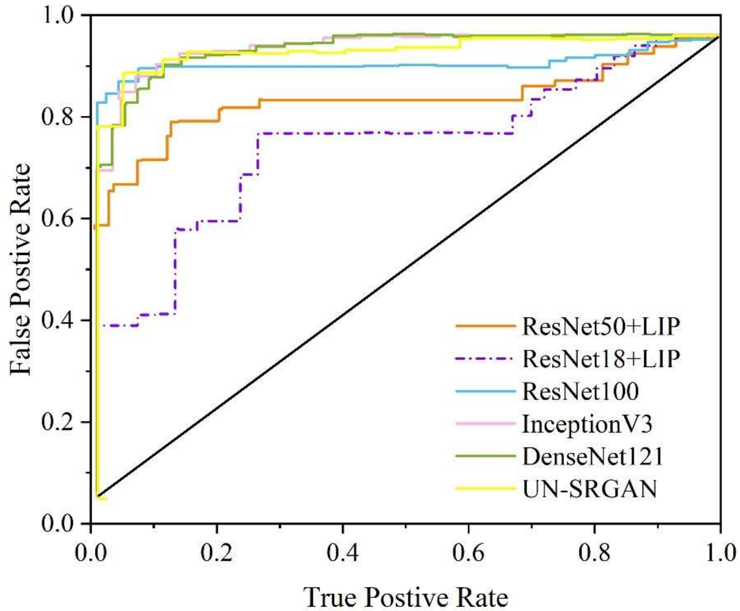


Fig. 10. ROC curve of classification models

5. Conclusion

In this paper, a super-resolution reconstruction algorithm for medical ultrasound images based on UN-SRGAN is proposed, and various types of mainstream super-resolution reconstruction models are selected to do comparison experiments with UN-SRGAN. According to the evaluation results of the private ultrasound dataset in this paper, it is shown that UN-SRGAN not only performs the best in objective qualitative results, but also performs the best in visual effects. Compared with other classical super-resolution reconstruction methods, it improves 1.8% and 1.1% compared with ResNet18+LIP, while it improves 6.4% and 6.4% compared with ResNet18+LIP on Recall and Precision, respectively. The experiment proves that UN-SRGAN can effectively improve the classification accuracy and enhance the reconstruction effect, and has better constraints on both reconstruction and classification tasks in the same batch of training for multi-tasking purposes.

About the Author

Yun Wang was born in Changsha, Hunan, China, in 1978. She obtained a bachelor's degree from NanHua University in China. Her main research direction is Gynecologic ultrasound.

Yabiao Zeng was born in Changsha, Hunan, China, in 1978. He obtained a bachelor's degree from NanHua University in China. His main research direction is general surgery.

References

[1] Silva Martins, R., Helio Oliani, A., Vaz Oliani, D., & Martinez de Oliveira, J. (2019). Subendometrial resistence and pulsatility index assessment of endometrial receptivity in assisted reproductive technology cycles. Reproductive Biology and Endocrinology, 17, 1-7.

[2] Hviid Saxtorph, M., Persson, G., Hallager, T., Birch Petersen, K., Eriksen, J. O., Larsen, L. G., ... & Hviid, T. V. F. (2020). Are different markers of endometrial receptivity telling us different things about endometrial

function?. American Journal of Reproductive Immunology, 84(6), e13323.

[3] Maziotis, E., Kalampokas, T., Giannelou, P., Grigoriadis, S., Rapani, A., Anifantakis, M., ... & Vlahos, N. (2022). Commercially available molecular approaches to evaluate endometrial receptivity: a systematic review and critical analysis of the literature. Diagnostics, 12(11), 2611.

[4] Pirtea, P., de Ziegler, D., & Ayoubi, J. M. (2023). Endometrial receptivity in adenomyosis and/or endometriosis. Fertility and Sterility, 119(5), 741-745.

[5] Ruiz-Alonso, M., Valbuena, D., Gomez, C., Cuzzi, J., & Simon, C. (2021). Endometrial Receptivity Analysis (ERA): data versus opinions. Human Reproduction Open, 2021(2), hoab011.

[6] Haas, J., & Casper, R. F. (2022). Observations on clinical assessment of endometrial receptivity. Fertility and sterility, 118(5), 828-831.

[7] Lessey, B. A., & Young, S. L. (2019). What exactly is endometrial receptivity?. Fertility and sterility, 111(4), 611-617.

[8] Craciunas, L., Gallos, I., Chu, J., Bourne, T., Quenby, S., Brosens, J. J., & Coomarasamy, A. (2019). Conventional and modern markers of endometrial receptivity: a systematic review and meta-analysis. Human reproduction update, 25(2), 202-223.

[9] Bassil, R., Casper, R., Samara, N., Hsieh, T. B., Barzilay, E., Orvieto, R., & Haas, J. (2018). Does the endometrial receptivity array really provide personalized embryo transfer?. Journal of Assisted Reproduction and Genetics, 35, 1301-1305.

[10] Vitale, S. G., Buzzaccarini, G., Riemma, G., Pacheco, L. A., Sardo, A. D. S., Carugno, J., ... & Laganà, A. S. (2023). Endometrial biopsy: Indications, techniques and recommendations. An evidence-based guideline for clinical practice. Journal of gynecology obstetrics and human reproduction, 52(6), 102588.

[11] Labarta, E., Sebastian-Leon, P., Devesa-Peiro, A., Celada, P., Vidal, C., Giles, J., ... & Diaz-Gimeno, P. (2021). Analysis of serum and endometrial progesterone in determining endometrial receptivity. Human Reproduction, 36(11), 2861-2870.

[12] Ben Rafael, Z. (2021). Endometrial Receptivity Analysis (ERA) test: an unproven technology. Human Reproduction Open, 2021(2), hoab010.

[13] Zhang, C. H., Chen, C., Wang, J. R., Wang, Y., Wen, S. X., Cao, Y. P., & Qian, W. P. (2022). An endometrial receptivity scoring system basing on the endometrial thickness, volume, echo, peristalsis, and blood flow evaluated by ultrasonography. Frontiers in endocrinology, 13, 907874.

[14] Shui, X., Yu, C., Li, J., & Jiao, Y. (2021). Development and validation of a pregnancy prediction model based on ultrasonographic features related to endometrial receptivity. American Journal of Translational Research, 13(6), 6156.

[15] Ouyang, Y., Peng, Y., Mao, Y., Zheng, M., Gong, F., Li, Y., & Li, X. (2024). An endometrial receptivity scoring system evaluated by ultrasonography in patients undergoing frozen–thawed embryo transfer: a prospective cohort study. Frontiers in Medicine, 11, 1354363.

[16] Khan, M. S., Shaikh, A., & Ratnani, R. (2016). Ultrasonography and Doppler study to predict uterine receptivity in infertile patients undergoing embryo transfer. The Journal of Obstetrics and Gynecology of India, 66, 377-382.

[17] Wu, J., Sheng, J., Wu, X., & Wu, Q. (2023). Ultrasound‑assessed endometrial receptivity measures for the prediction of in vitro fertilization‑embryo transfer clinical pregnancy outcomes: A meta‑analysis and systematic review. Experimental and Therapeutic Medicine, 26(3), 453.

[18] Tong, R., Zhou, Y., He, Q., Zhuang, Y., Zhou, W., & Xia, F. (2020). Analysis of the guidance value of 3D ultrasound in evaluating endometrial receptivity for frozen-thawed embryo transfer in patients with

repeated implantation failure. Annals of translational medicine, 8(15), 944.

[19] Luoyi Li,Lintao Zheng,Chunlei Yang & Yongsheng Dong. (2025). GAL-GAN: Global styles and local high-frequency learning based generative adversarial network for image cartoonization. Computers and Electrical Engineering,123(PC),110164-110164.

[20] Haoyan Jiang & Ji Zhao. (2024). Multi-lesion Segmentation of Fundus Images using Improved UNet++. IAENG International Journal of Computer Science,51(10),

[21] Jia Chen Wan. (2024). Fast Risk Estimation through Fourier Transform Based Multilevel Monte Carlo Simulation. IAENG International Journal of Applied Mathematics,54(4).