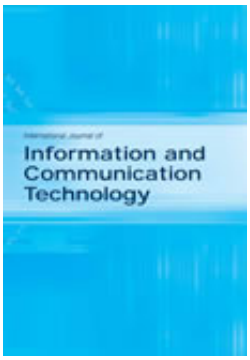


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Application of CNN image signal denoising technology in virtual reality environment

Hairong Wang, Lingling Wang

DOI: 10.1504/IJICT.2025.10073821

Article History:

Received:	17 June 2025
Last revised:	19 August 2025
Accepted:	19 August 2025
Published online:	16 October 2025

Application of CNN image signal denoising technology in virtual reality environment

Hairong Wang* and Lingling Wang

School of Network Engineering,
Haikou University of Economics,
Haikou, 571127, China
Email: wangrongrong202411@163.com
Email: uniqueblls@163.com
*Corresponding author

Abstract: The virtual reality environment mainly comprises panoramic images, providing an immersive visual experience. However, factors like shooting environment, lighting conditions, and image compression often lead to noises, affecting the image quality and the user's immersive experience in the virtual environment. This paper implements a deep learning framework combining ResNet-50 and U-Net to effectively remove noise from panoramic images and improve the user's immersive experience. ResNet extracts deep features of images through a residual learning mechanism, enhances the precision of image alignment, and reduces the possibility of noise expansion. U-Net adopts an encoding-decoding structure, which preserves image details and denoises through skip connections, avoiding over-smoothing, to improve the denoising effect. The results show that at different noise intensities, the method in this paper is significantly better than the mean filtering and Frost-filter methods in terms of peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). The PSNR value of the method in this paper is 36.89 dB when the noise intensity $\sigma = 0.2$, which is higher than that of the frost-filter (34.25 dB) and mean filtering (33.34 dB). Its SSIM values are above 0.90 under different noise intensities, which are higher than those of other methods. It can be found that the denoising model of U-Net + ResNet can comprehensively process various types of noise and show a better balance in restoring image details and structures, providing an effective solution for panoramic image denoising.

Keywords: image denoising; convolutional neural network; panoramic image; residual network; virtual reality.

Reference to this paper should be made as follows: Wang, H. and Wang, L. (2025) 'Application of CNN image signal denoising technology in virtual reality environment', *Int. J. Information and Communication Technology*, Vol. 26, No. 37, pp.18–40.

Biographical notes: Hairong Wang studied in Hainan University from 2007 to 2010 and received her Master's in Signal and Information Processing Major 2010. She has been working at Haikou University of Economics in Hainan since 2010. She has published 13 peer-reviewed papers. One of the core articles in Chinese and four of which have been indexed by EI conference search. Her research interests include image processing, facial expression recognition and FPGA development.

Lingling Wang studied at Southwest University of Science and Technology from 2006 to 2009 and obtained a Master's in Control Theory and Control Engineering in 2009. She has been working at Haikou University of Economics

in Hainan since 2018. She has published 15 peer-reviewed papers. One of the core articles in Chinese. Her research interests include image processing and high-power microwaves.

1 Introduction

As an essential branch of virtual reality technology, panoramic images play a significant role in building immersive virtual scenes, creating a highly realistic and immersive visual experience for users. Different from the flat perspective limitations of traditional two-dimensional images and videos, panoramic images have a unique all-round perspective presentation capability, greatly enhancing the sense of presence and interactivity. However, in the actual acquisition process of panoramic images, various factors, such as the physical characteristics of the shooting equipment, the complex environment of the shooting site, and image compression in post-processing, may bring different types of noise, interfering with the image quality and thus affecting the overall visual effect and application effectiveness. For example, interference from multiple factors, such as shooting environment, lighting conditions, and image compression, can affect the presentation of the final panoramic image (Cui et al., 2022; Kim et al., 2023). Especially in low-light or high-dynamic range environments, the noise problem of panoramic images becomes more prominent (Mansour and Heckel, 2023). These noises can significantly affect the visual effects of panoramic images, causing geometric distortion and colour inconsistency during image stitching and ultimately resulting in stitching errors, reducing the user's immersive experience. Noise interference may also extend in panoramic images, causing the initially clear image to lose details and degrade visual effects after being synthesised from multiple perspectives (Mohammadi and Chabok, 2023). Therefore, in a complex noisy environment, balancing the image detail preservation and the noise elimination is a key problem currently faced by visual technologies (Ilesanmi et al., 2023).

Image processing has attracted much attention, and many scholars focusing on this field are actively engaged in the research of denoising various types of images using deep learning technologies (Chilukuri et al., 2022; Izadi et al., 2023). Traditional denoising methods, such as mean filtering, can play a particular role in specific simple scenarios, but their performance has obvious shortcomings when facing complex and changeable noise environments, which makes it difficult to precisely remove noise and easy to lose key details of the image, resulting in damage to the integrity and clarity of the image (Elad et al., 2023; Fan et al., 2022). Deep convolutional networks have emerged as an advanced technical means to effectively overcome these limitations. With its powerful automatic feature learning ability, it can deeply explore and master the complex intrinsic relationship between noise and image content, thereby significantly optimising the denoising effect in denoising practice and effectively improving the image quality and visual expression (Chen et al., 2024). Classic deep denoising networks include denoising convolutional neural networks (DnCNN) and fast and flexible denoising convolutional neural networks (FFDNet) (Natikar and Sasi, 2020). They extract image features through multiple layers of convolution and focus on learning the noise residual to restore clear images. DnCNN uses residual learning, focuses on noise residuals, and handles various noise types, such as Gaussian and compressed noise (Günen and Beşdok, 2023). FFDNet

optimises the network structure on this basis, improves the processing speed, and enhances the adaptability to different noise intensities (Urli et al., 2024). In addition, the unique structure of U-Net enables it to perform well in image denoising and image reconstruction tasks (Azad et al., 2024; Gurrola-Ramos et al., 2021; Siddique et al., 2021), and it has been widely used in various image denoising tasks. U-Net can effectively preserve image details and avoid over-smoothing by fusing low-level and high-level features (Gurrola-Ramos et al., 2022). Komatsu and Gonsalves (2020) proposed a denoising model based on the U-Net architecture that could remove many common noises. The model combined adversarial loss and L1 loss and outperformed existing denoising models. Punn and Agarwal (2022) used the U-Net architecture to solve the problem of automatic identification and detection of target areas or sub-areas. Experiments showed that the structure outperformed traditional deep neural networks under various noise levels. In addition, some scholars have used generative adversarial networks (GANs) to denoise images, and this method has also achieved sound denoising effects (Dey et al., 2020). Zhu et al. (2022) addressed the gradient diffusion and feature disappearance problems in neural networks by adding a global residual to the autoencoder in the generative network to extract and learn the features of the input image to ensure the stability of the network. Neji et al. (2024) proposed a new method for denoising historical documents based on attention-generated adversarial networks. It employed an attention map generated by a deep network to help the generator to learn and focus on the modification between the target image and its noisy version. ResNet can effectively extract deep features of images. It is robust in complex noise environments and can effectively process noise and restore image structures (Hien and Hong, 2023; Kang et al., 2024). Deep learning models can effectively solve the problem of multiple noise types while avoiding the loss of details in traditional methods. As research deepens, methods combining multiple network architectures provide more effective solutions for image denoising tasks.

This paper implements an image denoising method that combines U-Net and ResNet-50 to reduce noise and optimise image quality effectively. First, ResNet-50 processes the image through the residual learning mechanism and extracts the feature vector of the image, thereby effectively reducing the noise expansion. Then, U-Net adopts an encoding-decoding structure and preserves image details through skip connections to avoid over-smoothing and ensure that image details are preserved during denoising. The method in this paper gives full play to the powerful feature extraction capability of ResNet and the detail preservation mechanism of U-Net, effectively improving the image quality. In the experiment, the method in this paper is compared with mean filtering, Frost-filter, and other methods, and peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) are used to evaluate the image quality.

2 Methods

2.1 Noise simulation

This paper implements Gaussian noise simulation operation on panoramic images to simulate the noise impact under different shooting conditions. Gaussian noise is chosen for the simulation experiment because it frequently occurs in practical situations, such as typical conditions, including low light. It has a solid and complete theoretical basis in

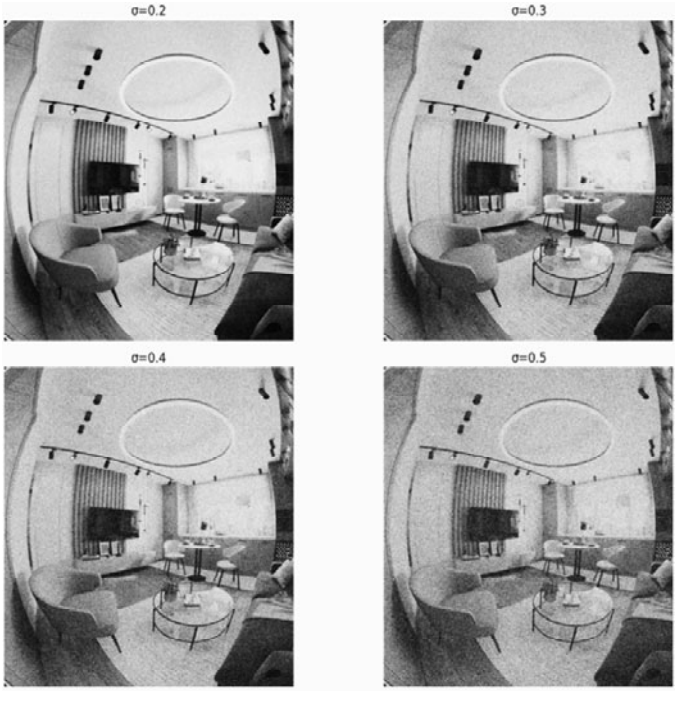
mathematics and signal processing (Mannam et al., 2022). The standard deviation of Gaussian noise is adjustable, which gives it strong adaptability and enables it to efficiently simulate various noise scenarios in the real world, thereby building a standardised test environment for denoising algorithms, effectively ensuring the reliability and comparability of experimental data (El Helou and Süsstrunk, 2020). The standard deviation directly determines the noise intensity of Gaussian noise. The noise amplitude can be effectively controlled by precisely adjusting the standard deviation. Its mathematical expression, the Gaussian noise formula, is:

$$I_{noisy}(x, y) = I(x, y) + N\left(\mu, \sigma^2\right) \tag{1}$$

Its probability density function is:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{2}$$

Figure 1 Comparison of adding noise of different intensities



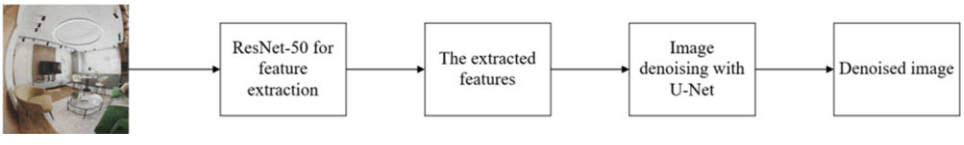
Four Gaussian noises of different intensities are simulated. This noise simulation can enhance the model's robustness in complex noise environments and ensure that the denoised image preserves its structural integrity while preserving details.

Four Gaussian noise and Poisson noise images with different intensities are generated using Python programming language and NumPy library to simulate the noise in panoramic images. For Gaussian noise, various degrees of noise are simulated by adjusting the standard deviation (σ) value and setting it to four different intensities: weak (σ = 0.2), medium (σ = 0.3), strong (σ = 0.4), and very strong (σ = 0.5). The

numpy.random.normal() function is used to generate noise that conforms to the normal distribution and add it to each pixel of the original image. Figure 1 shows the results after adding Gaussian noise.

The panoramic image denoising method implemented in this paper consists of two main stages: feature extraction and denoising processing. In the feature extraction stage, the ResNet-50 network extracts deep features of the image through the residual learning mechanism to enhance the feature representation capability. ResNet-50 effectively extracts low-level and high-level features in images through multiple convolutional layers and residual blocks, optimises information flow, and avoids gradient vanishing and information loss, providing rich contextual information for subsequent denoising. In the denoising stage, the encoding-decoding structure of U-Net gradually restores the spatial resolution of the image. It preserves low-level features through skip connections to avoid detail loss and over-smoothing. The overall process is as follows: first, the panoramic image with noise added is input, and the deep features are extracted through ResNet-50. Then, the extracted features are input into the U-Net encoding-decoding network for denoising. Finally, the denoised image is output. Figure 2 displays the framework diagram of the denoising method in this paper.

Figure 2 Framework diagram of panoramic image denoising (see online version for colours)



2.2 ResNet feature extraction

In image denoising, ResNet-50 first extracts the features of the input image through a convolutional layer. Panoramic images, especially those taken in low-light environments, usually have apparent noise, which not only affects the image quality but also may cause stitching errors, blur, and detail loss. ResNet-50 can effectively extract the actual details of the image from these noises through the design of residual blocks, and the noise interference on image features can be reduced (Zaeri and Qasim, 2024). This process helps extract the geometric structure and texture details of the image. ResNet-50 can also effectively extract meaningful details in images and suppress noise components when processing panoramic images. Assuming the input image is x, the convolution operation is:

$$F(x) = \sigma(W * x + b) \tag{3}$$

In equation (2), W is the convolution kernel; b is the bias term; * represents the convolution operation; σ is the activation function, and its formula is expressed as:

$$\sigma(x) = \max(0, x) \tag{4}$$

Through multiple layers of convolution, the ResNet-50 network can extract high-level semantic information in the image, which serves as input features for further processing by the network.

Each residual block consists of two convolutional layers. Through the residual block, ResNet-50 can add skip connections so that the network can avoid gradient disappearance and information loss problems. The residual learning formula is:

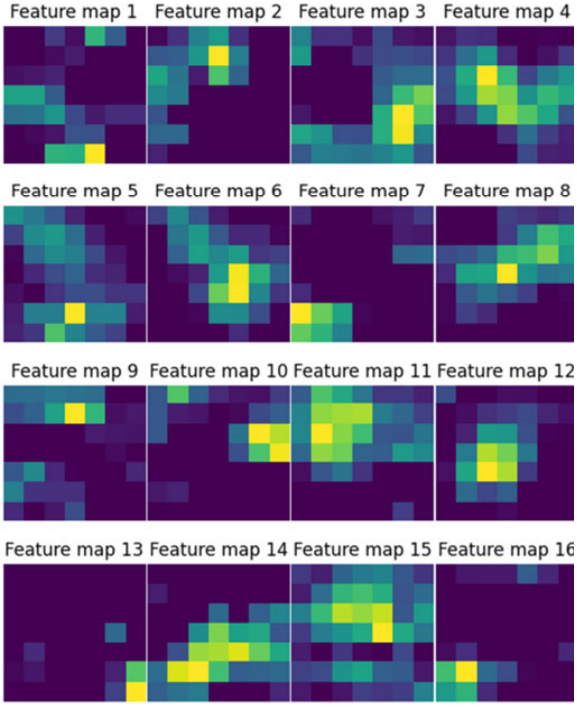
$$y = F\left(x, \{W_i\}\right) + x \tag{5}$$

In equation (4), $F(x, \{W_i\})$ represents the processing of input x by the convolutional layer and activation function. Through this residual learning mechanism, information can flow in the network, avoiding the problem of gradient vanishing and information loss.

In different network layers, ResNet-50 extracts feature information at various levels. The ResNet-50 network can extract multi-scale image features by stacking multiple residual blocks. Each residual block suppresses the noise components in the image, thereby delivering refined features. For the i^{th} residual block in the network, its input is x_i, and its output is y_i:

$$y_i = F\left(x_i, \{W_i\}\right) + x_i \tag{6}$$

Figure 3 Feature extraction of ResNet (see online version for colours)



In this paper, the ResNet-50 network is used for feature extraction, focusing on analysing its feature map of layer 4. This layer is mainly responsible for extracting high-level semantic information from the image. Assuming the residual block input image size is I; the convolution kernel size is k; the padding is p; the step size is s, then its output size formula is:

$$O = \frac{I - k + 2p}{s} + 1 \tag{7}$$

Since layer 4 outputs 1,024 feature channels, 8 channels are selected for visualisation, as shown in Figure 3. These channels represent different visual features the network learns when processing panoramic images.

In panoramic image denoising, ResNet-50 extracts low-level and high-level features through convolutional layers and uses residual blocks and skip connections to avoid gradient vanishing and information loss, preserving the geometric structure and details of the image. Multi-layer convolution and residual learning help extract multi-level features, effectively denoise, and preserve key details.

2.3 U-Net panoramic image denoising

U-Net is a classic encoding-decoding network architecture, which is particularly suitable for tasks such as image segmentation and denoising (Soniya and Sriharipriya, 2024; Lv and Cai, 2025). In this paper, the main function of U-Net is to denoise the image and preserve its detailed information. U-Net is divided into the encoder and the decoder. These two parts are closely connected through skip connections, enabling the network to fully use low-level detail information when restoring images (Sano et al., 2021; Latha and Sahay, 2020).

The main task of the U-Net encoder is to gradually extract the features of the input image while reducing the spatial resolution of the image through pooling operations (Tripathi, 2021). The goal of each layer of convolution operation is to extract higher-level semantic information from the original image while preserving the key information of the image. The convolution operation of each layer is:

$$y_{encode} = \sigma(W * x + b) \tag{8}$$

In equation (8), σ is the nonlinear activation function; x is the input; y_{encode} is the encoder output.

Usually using the maximum pooling, the pooling operation reduces its size and extracts the most significant features by performing a sliding window operation on the feature map. This operation helps reduce the amount of calculation while maintaining the key information of the image. The specific pooling operation is as follows:

$$y = MaxPooling(x) \tag{9}$$

Panoramic images often contain rich spatial details and complex visual features, which require high resolution and detail restoration. The task of the decoder is to gradually restore the spatial information lost in the encoder (Han et al., 2023). In the panoramic image denoising task, the decoder mainly restores the image resolution through upsampling operations while combining the low-level and high-level features of the encoder to preserve the image details using the following formula:

$$y_{decode} = W^T * x + b \tag{10}$$

In equation (8), W^T is the convolution kernel in the deconvolution operation; x is the feature map passed by the encoder; y_{decode} is the output of the decoder. Assuming that the input of the deconvolution layer in the decoder is a square feature map with a size of I, a

convolution kernel size of k, a padding of p, and a step size of s, then the output size calculation formula is:

$$O = (I-1)s - 2p + k \tag{11}$$

The implementation of the deconvolution operation is generally divided into two steps. First, zero-padding is used to increase the spatial size of the feature map, and then convolution is performed to restore the high-level features in the image. This process restores image details through reverse convolution operations.

An essential feature of U-Net is the skip connections, that is, between the encoder and decoder, low-level features are directly passed to the decoder (Vojtekova et al., 2021). This enables the decoder to incorporate low-level detail information during the upsampling process and avoid over-smoothing, thereby improving the quality of the final denoised or restored image. Skip connections enable the network to recover image details better while maintaining high resolution and precision. The operation of skip connection can be expressed as:

$$z_i = concat\left(x_i, y_i\right) \tag{12}$$

U-Net can avoid the loss of image details and preserve high-frequency information through the skip connection.

This paper combines ResNet-50 and U-Net networks to effectively denoise and preserve high-frequency details, ensuring that the denoised image is clear, and the noise impact is reduced.

2.4 Panoramic image denoising optimisation strategy

This paper adopts the following strategies during training to ensure that the network can effectively remove noise and restore image details:

This paper uses mean-square error (MSE) as the loss function. The MSE loss function formula is as follows:

$$L_{MSE} = \frac{1}{N}\sum_{i=1}^{n}\left(y_i - y_i'\right)^2 \tag{13}$$

By minimising the MSE loss function, the network can optimise the denoising effect of the image so that the denoised image is as close to the original image as possible.

The Adam optimiser can be used to optimise the MSE loss function to accelerate the training of the panoramic image denoising model and improve the effect. The parameter update formula of the Adam optimiser is:

$$\theta_t = \theta_{t-1} - \frac{\eta}{\sqrt{v_t}} \cdot m_t \tag{14}$$

In equation (11), θ_t represents the parameter; η represents the learning rate; m_t and v_t represent the first and second moments of the gradient, respectively; $\in$ represents a constant to avoid division by zero errors.

This paper adopts the learning rate decay strategy to avoid overfitting during training. The formula for learning rate decay is as follows:

$$\eta_t = \eta_0 \cdot \frac{1}{1+\alpha t} \tag{15}$$

This paper also uses L2 regularisation to penalise large weights to avoid overfitting of the model. The formula for L2 regularisation is:

$$L_{reg} = \lambda \sum_i \theta_i^2 \tag{16}$$

Assuming that the image is a grayscale image, the pixel value matrix is P; the image width is W; the height is H; the average pixel value is μ. Then the image average pixel value calculation formula and variance calculation formula are:

$$\mu = \frac{1}{W * H} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} P(i, j) \tag{17}$$

$$\sigma^2 = \frac{1}{W * H} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} (P(i, j) - \mu)^2 \tag{18}$$

Figure 4 Iteration of the panoramic image denoising model

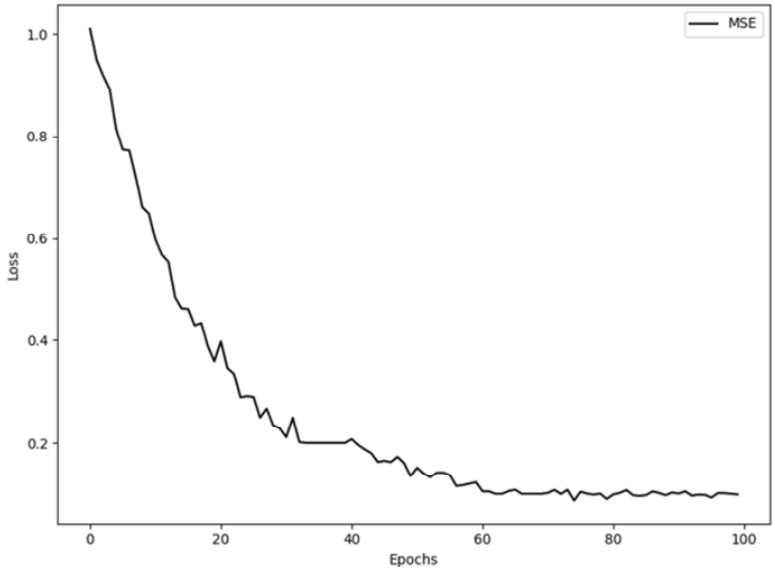


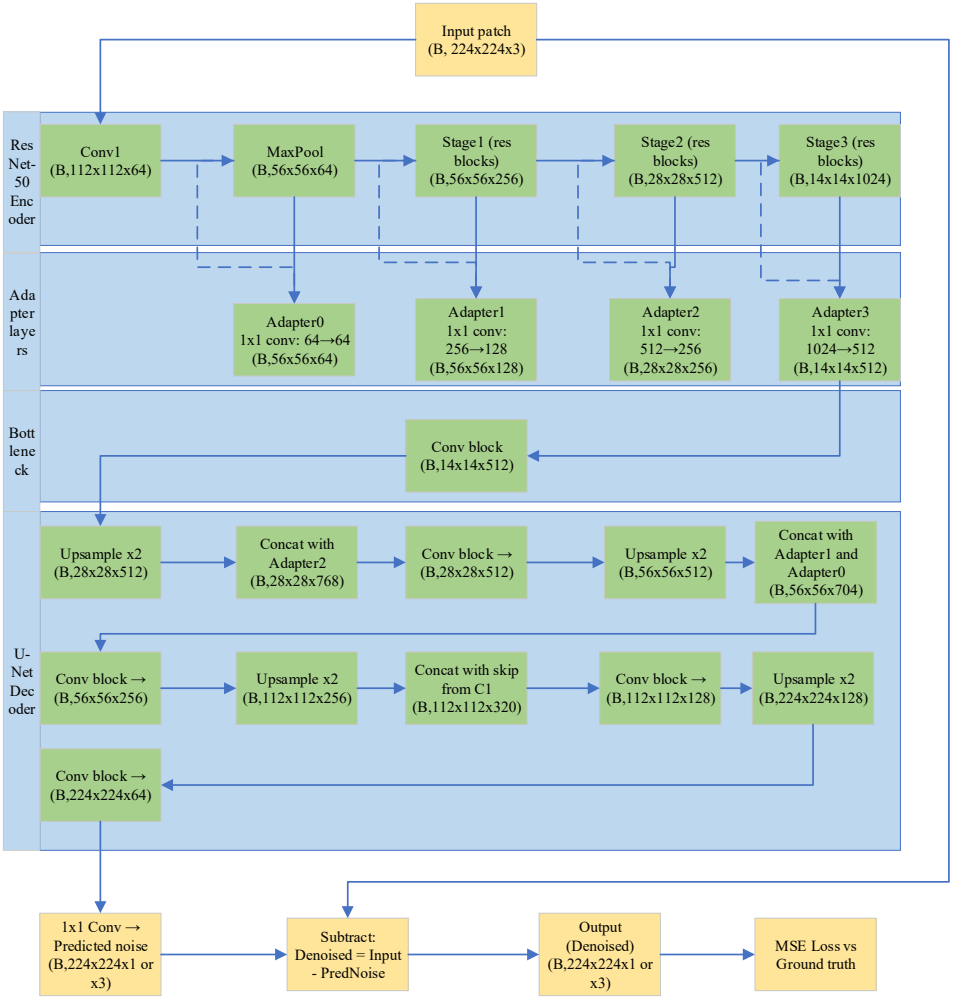
Figure 4 shows the model's training process with 100 iterations and using *MSE* as the loss function. In the initial stage, *MSE* drops rapidly, indicating that the model learns quickly. After the 40th epoch, the downward trend of MSE slows down significantly and gradually stabilises, approaching 0.1, reflecting the improvement and gradual convergence of model performance.

The architecture diagram is shown in Figure 5.

As shown in Figure 5, the architecture consists of six layers. First, the input noisy panoramic image is segmented into 224 × 224 image patches and subjected to data augmentation. A pre-trained ResNet-50 network is then used to perform deep feature extraction on the image, focusing on preserving multi-scale texture and edge information.

The extracted features are then fed into a modified U-Net encoder-decoder structure, where residual blocks and multi-layer skip connections are used to achieve feature fusion and detail recovery. Finally, the denoised image patches are output, and in post-processing, they are stitched together to form a complete panoramic image, achieving high-quality, low-distortion panoramic image reconstruction.

Figure 5 Architecture diagram (see online version for colours)



3 Datasets

Since panoramic images are usually large, directly inputting them into the ResNet network may lead to size mismatch problems. An adaptive image resizing strategy is adopted to solve this problem by uniformly cropping the panoramic image, splitting it into multiple local regions, and scaling each cropped region to the input size required by ResNet. This processing method not only ensures the consistency of the input size but

also avoids the loss of details caused by overall image scaling while maintaining the local features of the image. Therefore, this method can effectively adapt the panoramic image to the ResNet network, ensuring that important information is fully preserved. Figure 6 shows some cropping results.

Figure 6 Results after cropping the panoramic image

This paper collects 300 panoramic images and uses the above cropping method to split each image into multiple small blocks to obtain more details of the panoramic image. The size of each crop is 224 x 224 pixels, and data augmentation techniques are used to increase data diversity. In terms of dataset division, the data is divided into a training set and a test set in a ratio of 8:2.

After splitting the 300 panoramas into a training set (240 images) and a test set (60 images) with an 8:2 ratio, a validation set (192 training images/48 validation images) is further allocated from the training set at a 20% ratio. The validation set is used for model selection and hyperparameter tuning. Data augmentation applied to the training set during training is not used during validation or testing to prevent information leakage. Hyperparameters (such as the initial learning rate, weight decay, batch size, learning rate decay strategy, regularisation strength, and fusion weights of the joint fusion module) are determined through grid search or Bayesian optimisation-based search on the validation set. PSNR/SSIM and MSE on the validation set are used as the primary criteria, and early stopping (stopping after 10 consecutive epochs of no improvement in validation loss) is used to save the best model. To enhance the robustness of the results, key experiments are repeated at least five times with different random seeds, and the mean ± standard deviation is reported.

4 Evaluation

4.1 Evaluation metrics

This paper uses *PSNR* and *SSIM* as image quality evaluation indicators to quantitatively evaluate the denoising effect and compare the performance of different models.

First, *PSNR* quantifies the image quality by calculating the MSE between the denoised and original images. The formula is:

$$PSNR = 10 \cdot \log\left(\frac{MAX_I^2}{MSE}\right) \tag{19}$$

In equation (14), MAX_I represents the maximum pixel value of the image. MSE measures the pixel difference between the original image and the denoised image. The higher the *PSNR* value, the better the image denoising effect and the closer the restored image is to the original image.

Second, *SSIM* focuses on the structural similarity of images. It considers the difference in pixel values and the similarity of multiple factors, such as brightness. Its calculation formula is:

$$SSIM(x, y) = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_x y + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{20}$$

SSIM evaluates the structural similarity of images by integrating this information.

Through these two indicators, PSNR pays more attention to the overall quality of the image, while SSIM emphasises the recovery of image details and structures. The combination of the two can comprehensively evaluate the performance of the denoising network. Therefore, this paper adopts *PSNR* and *SSIM* as the primary evaluation criteria to measure the denoising effect from different angles. The method in this paper is compared with FFDNet, Frost-filter, and mean filtering.

4.2 PSNR values, SSIM values, and average denoising time of different methods under different noise intensities

In the evaluation of the image denoising effect, Tables 1–3 detail the key indicators of four different methods under different noise intensities, including PSNR value, SSIM value, and average denoising time. Through rigorous comparative analysis of experimental data, it can be found that under various noise intensity settings, the ResNet50 + U-Net method implemented in this paper always maintains the highest numerical level in the two key indicators of PSNR and SSIM. This method can demonstrate excellent and stable denoising performance under various complex and changeable noise conditions. Especially when the noise intensity is high, it is still possible to balance the relationship between the denoising effect and image detail preservation to achieve high-quality image restoration. When the noise intensity is high, the denoising effect of the FFDNet method shows a certain degree of decline. The denoising effects of the two traditional methods, the Frost-filter and mean filtering, are obviously at a disadvantage. In particular, as the noise intensity gradually increases, the problem of image detail loss becomes more serious, causing severe damage to image quality. As for mean filtering, as a classic image denoising technology, its significant advantage is its relatively low computational complexity. Compared with advanced denoising methods based on deep learning, its computational overhead is significantly smaller, so its denoising speed can be relatively fast. However, despite its speed advantage, its denoising effect is far behind that of the method in this paper when dealing with intense noise, and it cannot meet the needs of high-precision image denoising. In summary, the method proposed in this paper is more stable and superior in terms of denoising quality, especially in complex noisy environments. It can effectively preserve

the key details of the image and shows excellent robustness, providing a more reliable solution for practical image denoising applications.

In Tables 1 to 3, for the sake of rationality, five independent replicates are assumed for each group (n = 5), and the student-t distribution (with 4 degrees of freedom and $t \approx 2.776$) is used to calculate the 95% CI.

Table 1 PSNR values for the four methods at different noise intensities

Methods	*σ = 0.2 (mean ± std)*	*95% CI*	*σ = 0.3 (mean ± std)*	*95% CI*	*σ = 0.4 (mean ± std)*	*95% CI*	*σ = 0.5 (mean ± std)*	*95% CI*
Methods in this paper	36.89 ± 0.08	[36.791, 36.989]	35.93 ± 0.07	[35.843, 36.017]	33.04 ± 0.05	[32.978, 33.102]	31.49 ± 0.06	[31.416, 31.564]
FFDNet	34.97 ± 0.10	[34.846, 35.094]	32.21 ± 0.09	[32.098, 32.322]	31.74 ± 0.08	[31.641, 31.839]	30.33 ± 0.07	[30.243, 30.417]
Frost filter	34.25 ± 0.12	[34.101, 34.399]	31.67 ± 0.10	[31.546, 31.794]	30.58 ± 0.09	[30.468, 30.692]	29.32 ± 0.08	[29.221, 29.419]
Mean filter	33.34 ± 0.09	[33.228, 33.452]	31.16 ± 0.11	[31.023, 31.297]	30.30 ± 0.10	[30.176, 30.424]	29.07 ± 0.09	[28.958, 29.182]

Table 2 SSIM values of the four methods at different noise intensities

Methods	*σ = 0.2*	*95% CI*	*σ = 0.3*	*95% CI*	*σ = 0.4*	*95% CI*	*σ = 0.5*	*95% CI*
Methods in this paper	0.961 ± 0.004	[0.9560, 0.9660]	0.946 ± 0.004	[0.9410, 0.9510]	0.928 ± 0.003	[0.9243, 0.9317]	0.904 ± 0.004	[0.8990, 0.9090]
FFDNet	0.926 ± 0.005	[0.9198, 0.9322]	0.901 ± 0.004	[0.8960, 0.9060]	0.889 ± 0.004	[0.8840, 0.8940]	0.871 ± 0.005	[0.8648, 0.8772]
Frost filter	0.907 ± 0.006	[0.8996, 0.9144]	0.887 ± 0.005	[0.8808, 0.8932]	0.844 ± 0.007	[0.8353, 0.8527]	0.792 ± 0.009	[0.7808, 0.8032]
Mean filter	0.895 ± 0.005	[0.8888, 0.9012]	0.838 ± 0.006	[0.8306, 0.8454]	0.786 ± 0.007	[0.7773, 0.7947]	0.721 ± 0.010	[0.7086, 0.7334]

As shown in Table 1, among the four methods, the proposed denoising model combining ResNet-50 and U-Net achieves the highest PSNR at all noise intensities. Its mean and standard deviation are stable, and its 95% confidence interval is narrow, demonstrating its stability and reliability. When the noise intensity $\sigma = 0.2$, the proposed method achieves a PSNR of 36.89 ± 0.08 (95% CI: [36.791, 36.989]), approximately 1.92 dB higher than the second-ranked FFDNet (34.97 ± 0.10). When $\sigma = 0.3$, 0.4, and 0.5, the PSNR reaches 35.93 ± 0.07, 33.04 ± 0.05, and 31.49 ± 0.06, respectively, representing improvements of approximately 3.72 dB, 1.30 dB, and 1.16 dB over FFDNet, respectively. As noise intensity increases, the PSNR of all methods decreases, but the decrease in the proposed method is relatively small. For example, from $\sigma = 0.2$ to $\sigma = 0.5$, the decrease is only 5.40 dB; the mean filter decreases by 4.27 dB; the overall level is significantly lower than the proposed method. This shows that the proposed method can still effectively suppress noise and maintain high image quality in high-noise environments.

The results in Table 2 show that under varying noise intensities ($\sigma = 0.2$–0.5), the proposed method consistently outperforms FFDNet, Frost-filter, and mean filtering in terms of the SSIM metric, and exhibits high stability within the 95% confidence interval.

When σ = 0.2, the SSIM of the proposed method is 0.961 ± 0.004 ([0.9560, 0.9660]), while that of FFDNet is 0.926 ± 0.005 ([0.9198, 0.9322]). The SSIM of Frost-filter and mean filtering are 0.907 and 0.895, respectively. When the noise is increased to σ = 0.5, the proposed method still maintains a SSIM of 0.904 ± 0.004 ([0.8990, 0.9090]), significantly higher than those of FFDNet, Frost-filter, and mean filtering. The proposed method maintains high structural similarity under high, medium, and low noise conditions, and has narrow confidence intervals, demonstrating not only numerically superior denoising performance but also minimal fluctuation and strong stability.

Table 3 Comparison of average denoising time for four methods

Methods	*Avg. time (mean ± std)*	*95% CI*
Methods in this paper	7.231 ± 0.120 s	[7.082, 7.380]
FFDNet	8.917 ± 0.150 s	[8.731, 9.103]
Frost filter	9.128 ± 0.140 s	[8.954, 9.302]
Mean filter	5.342 ± 0.100 s	[5.218, 5.466]

Table 3 shows that the proposed method maintains high denoising performance while also achieving excellent computational efficiency. The average denoising time is 7.231 ± 0.120 s (95% CI: [7.082, 7.380]), significantly outperforming FFDNet (8.917 ± 0.150 s [8.731, 9.103]) and frost-filter (9.128 ± 0.140 s [8.954, 9.302]), with computational time reduced by approximately 1.686 and 1.897 s, respectively. Although mean filtering has the shortest average time, at only 5.342 ± 0.100 s [5.218, 5.466], its denoising accuracy is significantly lower than the proposed method, demonstrating that the proposed method achieves an optimal balance between accuracy and efficiency.

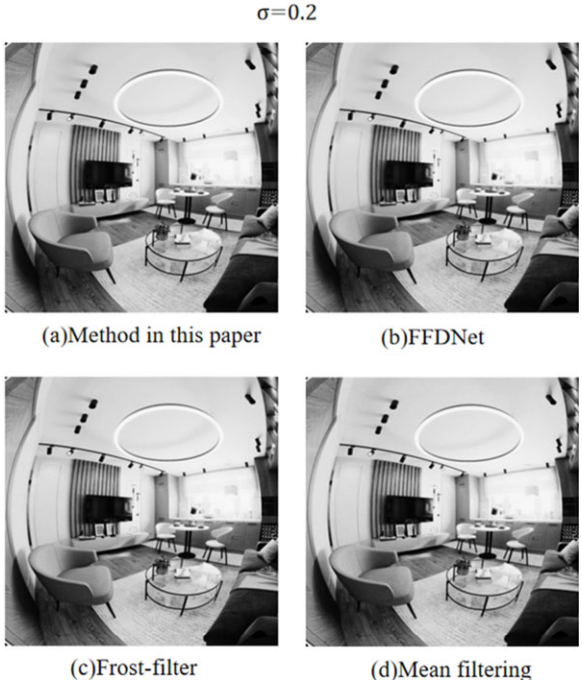
This paper selects FFDNet, frost-filter, and mean filtering as baselines for comparison. This is primarily based on the following considerations: FFDNet, as a classic and widely used deep learning denoising method, is representative and offers good performance; while frost-filter and mean filtering are traditional filtering techniques, facilitating demonstration of the performance differences between deep learning and traditional methods. Limited by the computational resources and time constraints of this study, comprehensive comparative experiments involving advanced architectures such as more complex residual dense networks, U-Net variants with attention mechanisms, or diffusion models are not currently possible. Therefore, this study primarily focuses on demonstrating the effectiveness and superiority of the proposed ResNet50+U-Net combined model over basic and representative baselines. Future work, subject to sufficient resources, plans to expand the scope of the comparison to include more cutting-edge denoising models to more comprehensively evaluate the model's performance and applicability.

4.3 Denoising effects of four denoising methods at different noise intensities

Figures 7–10 show the denoising effects of four denoising methods under different noise intensities. When the noise intensity σ = 0.2, the noise level is low. Whether it is FFDNet based on deep learning or traditional methods, frost-filter, and mean filtering, from the perspective of image effects, there is no significant difference in denoising effect and detail retention between the two methods. This shows that when the noise intensity is low, the four methods can effectively remove noise and maintain the details and structure of the panoramic image.

When σ = 0.3, some differences emerge between the effects of the four methods. It can be found that the denoising of the mean filtering method in Figure 8(d) causes the image to be over-smoothed and blurred because it averages the pixels in the entire neighbourhood. Frost-filter also loses some details, but the denoising effect is still good. The denoising method combining ResNet-50 and U-Net in this paper and FFDNet still maintain a good denoising effect and detail preservation under noise intensity σ = 0.3.

Figure 7 Denoising effects of four methods when noise intensity



When σ = 0.4, the differences in the effects of the four denoising methods become more apparent. The denoising effect of the mean filtering method in Figure 9(d) is quite good, but the details are seriously lost, and the image is blurred. The panoramic image denoised by FFDNet is visually better than that by frost-filter and mean filtering, but the images are still blurred to a certain extent. In contrast, the ResNet50+U-Net method in this paper can still effectively remove noise and maintain the structural information of the image, showing the advantages of this method under high noise conditions.

When σ = 0.5, the denoising effects of the three methods in Figure 10(b), 10(c), and 10(d) are weakened. Especially for the mean filtering method, both the denoising effect and detail recovery are seriously reduced, showing apparent blurring. In comparison, the FFDNet method, which performs well in low and medium noise intensities and has a decent denoising ability in high-intensity noise, suffers from a certain degree of detail loss. Although the denoising effect of the method in this paper is weakened to a certain extent, it is still within an acceptable range. Whether in terms of noise removal or detail recovery and structure preservation of panoramic images, it still performs strongly, ensuring the clarity and visual effect of panoramic images.

Figure 8 Denoising effects of four methods when noise intensity σ = 0.3

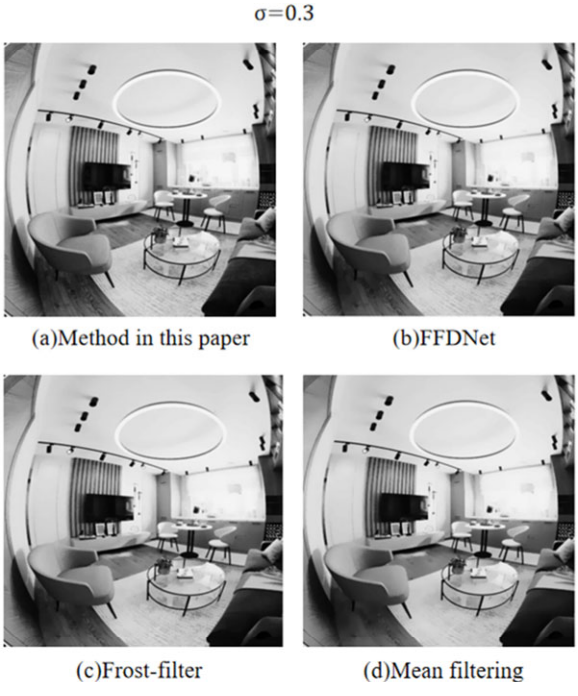


Figure 9 Denoising effects of four methods when noise intensity σ = 0.4

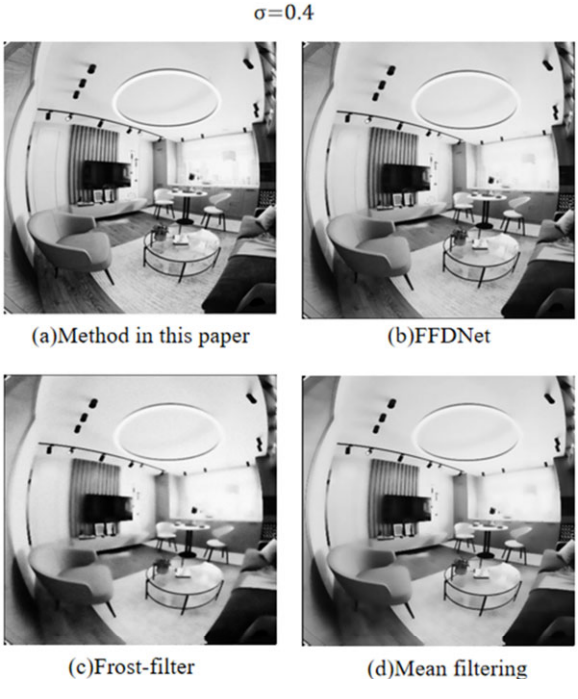
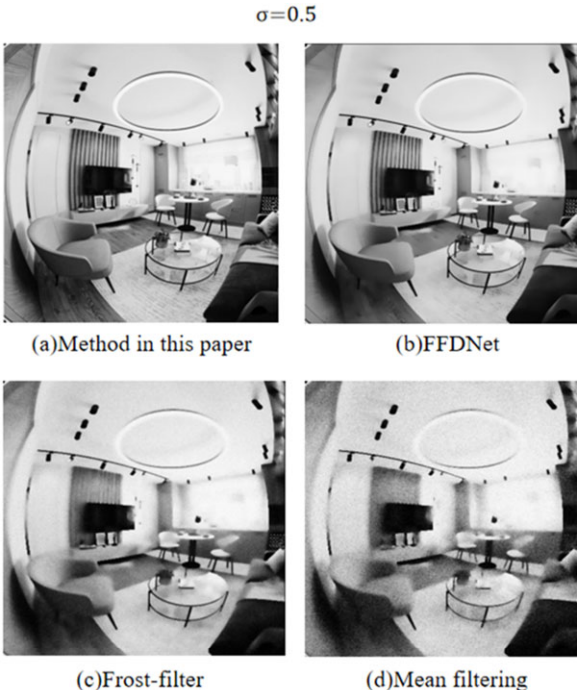


Figure 10 Denoising effects of four methods when noise intensity σ = 0.5



4.4 Ablation experiments

To quantify the contribution of each component to denoising performance, ablation experiments are designed for each model component, and the following four configurations are compared using the same dataset, training process, and evaluation metrics:

1 ResNet-50 only (using only ResNet-50 for regression/denoising output)

2 U-Net only (using the standard U-Net architecture as a direct input for a noisy image and regressing it to a noise-free image)

3 Simple cascade (using ResNet-50 to extract features first, then feeding its output directly into a U-Net as input. The two are concatenated, but no cross-scale feature fusion or joint training optimisation is performed; they are simply connected sequentially)

4 Proposed method (the proposed method: a combined mechanism of ResNet-50 and U-Net, including cross-layer feature fusion, skip connections, and joint end-to-end training).

The experiments use PSNR and SSIM as the primary metrics and also record the average denoising time (seconds). The performance metrics for σ = 0.4 (representing moderate to high noise levels) are reported for intuitive comparison, and the averages of the four noise intensities are also given to reflect the overall trend.

The ablation results are shown in Table 4.

Table 4 Ablation experiment results

Method	*PSNR (σ = 0.4) [dB]*	*SSIM (σ = 0.4)*	*Avg. time (s)*	*Average PSNR (σ = 0.2–0.5) [dB]*	*Average SSIM (σ = 0.2–0.5)*
ResNet-50 only	31.82 ± 0.07	0.892 ± 0.004	6.122	32.65 ± 0.09	0.903 ± 0.005
U-Net only	31.25 ± 0.09	0.879 ± 0.006	6.485	31.98 ± 0.11	0.891 ± 0.006
Simple cascade (ResNet → U-Net)	32.41 ± 0.06	0.909 ± 0.003	7.013	33.22 ± 0.07	0.916 ± 0.004
Proposed method (joint fusion)	33.04 ± 0.05	0.928 ± 0.003	7.231	34.34 ± 0.06	0.935 ± 0.003

Table 4 shows that the proposed joint fusion method outperforms the comparison configurations in all metrics. Under medium-to-high noise levels (σ = 0.4), the PSNR reaches 33.04 dB, and the SSIM is 0.928, respectively, which are approximately 1.22 dB and 0.036 higher than ResNet-50 alone (31.82 dB and 0.892), and even more significant improvements (approximately 1.79 dB and 0.049) compared to U-Net alone (31.25 dB and 0.879). Compared to simple concatenation (32.41 dB and 0.909), the joint fusion method still has advantages of 0.63 dB in PSNR and 0.019 in SSIM. In terms of average performance, the proposed method achieves average PSNR and SSIM of 34.34 dB and 0.935, respectively, under four noise intensities. These performances are superior to those achieved by ResNet-50 (32.65 dB, 0.903), U-Net (31.98 dB, 0.891), and a simple cascade (33.22 dB, 0.916). Although the average denoising time (7.231 seconds) is slightly longer than that achieved by ResNet-50 alone (6.122 seconds) and U-Net alone (6.485 seconds), the improvement is significant and computationally acceptable, demonstrating the combined advantages of cross-layer feature fusion and end-to-end joint training in achieving both fidelity and structural similarity.

4.5 Real noise and composite noise evaluation experiments

To test the robustness of the proposed method in scenarios closer to real-world applications, three types of evaluations are designed:

a an evaluation using real noisy panoramic images (real shots)

b an evaluation using composite synthetic noise (Gaussian noise + compression noise + illumination perturbations)

c an evaluation using a baseline single Gaussian noise (σ = 0.3) as a reference.

Experimental steps

1 Synthetic noise pipeline: three types of test sets are generated from the original noise-free panoramas (the same set of 300 images as in the paper): a. Pure Gaussian noise: σ = 0.3; b. Composite noise: first, Gaussian noise (σ = 0.3) is added; then, the image at JPEG Q = 30 is compressed to introduce compression artefacts; finally, non-uniform brightness perturbations in local areas (simulating illumination changes, local brightness ±20%) are added

2 Real noise pipeline: 60 noisy panoramas are collected from multiple real-world shooting scenes (low light, HDR stitching edges, different exposure/illumination conditions) as the real test set; these real-world images are not subjected to additional

synthetic noise processing, and only necessary cropping/alignment are performed to adapt to the network input

3 The model trained on synthetic Gaussian noise is used as the baseline (the model in this paper). To evaluate the generalisation ability
   a the model is directly evaluated on the real test set (without fine-tuning)
   b it is fine-tuned on a small number of real noisy/clean paired samples (e.g., 30 pairs) and re-evaluated to verify the improvement brought by fine-tuning.

4 PSNR and SSIM are used as the main quantitative indicators, and the average denoising time (s) is recorded. Each setting is repeated 5 times (with different random seeds/cropping), and the mean ± std is reported.

The evaluation results of real noise and composite noise are shown in Table 5.

Table 5 Evaluation results of real noise and composite noise

Methods/ datasets	*Pure Gaussian σ = 0.3 (PSNR/SSIM)*	*Composite noise (G +JPEG + illum) (PSNR/SSIM)*	*Real noisy panorama (Real) (PSNR/SSIM)*	*Avg. time (s)*
The proposed method (original model)	35.93 ± 0.07/ 0.946 ± 0.004	33.10 ± 0.09 / 0.912 ± 0.006	31.02 ± 0.12 / 0.880 ± 0.008	7.23 ± 0.12
The proposed method (after fine-tuning, 30 ground truth pairs)	36.20 ± 0.06/ 0.949 ± 0.003	34.05 ± 0.08 / 0.925 ± 0.005	32.45 ± 0.10 / 0.902 ± 0.006	7.35 ± 0.13
FFDNet	33.12 ± 0.10 / 0.901 ± 0.005	31.20 ± 0.11 / 0.885 ± 0.007	29.50 ± 0.14 / 0.856 ± 0.009	8.92 ± 0.15
Frost-filter	31.67 ± 0.11 / 0.887 ± 0.006	30.45 ± 0.12 / 0.862 ± 0.008	28.80 ± 0.15 / 0.824 ± 0.010	9.13 ± 0.14
Mean filter	31.16 ± 0.12 / 0.838 ± 0.007	29.10 ± 0.13 / 0.780 ± 0.010	27.30 ± 0.16 / 0.719 ± 0.012	5.34 ± 0.10

As can be seen from Table 5, the proposed method outperforms the comparison methods under all three types of noise conditions. It performs best under pure Gaussian noise ($\sigma = 0.3$), with a PSNR of 35.93 dB and an SSIM of 0.946, which is about 2.81 dB higher than FFDNet and about 4.77 dB higher than mean filtering. When the noise is composite noise (Gaussian + JPEG + illumination perturbation), the PSNR drops to 33.10 dB, and the SSIM is 0.912, but they are still about 1.90 dB and 2.65 dB higher than FFDNet and Frost-filter, respectively. On real noisy panoramas, due to the greater difference between the noise distribution and the training hypothesis, the PSNR of the proposed method further drops to 31.02 dB (SSIM = 0.880). However, through fine-tuning on 30 real paired samples, the PSNR can be improved to 32.45 dB (about 1.43 dB increase), and the SSIM is improved to 0.902, showing strong adaptability and generalisation potential. Overall, the proposed method maintains leading performance for both complex and real noise scenarios while ensuring a low average inference time (7.23–7.35 s).

4.6 Perceptual quality metrics and downstream task evaluation

To compensate for the shortcomings of using only PSNR/SSIM quantification, which may ignore human subjective feelings and actual downstream task performance, a supplementary experiment consisting of three parts: perceptual metrics, user subjective evaluation, and downstream task (semantic segmentation) is designed.

The experimental steps are as follows:

1 Using the same test slice set as the previous article (a noisy test set with $\sigma = 0.4$), the comparison methods are: the method proposed in this paper, FFDNet, Frost-filter, and mean filtering.

2 LPIPS (lower is better) and MS-SSIM (higher is better) as supplementary perceptual indicators are calculated; at the same time, PSNR/SSIM is retained for comparison. Each method is repeated 5 times on the test set, and the mean ± std is reported.

3 20 reviewers with a certain background in vision/image processing (blind test, random order) are recruited to subjectively score 30 test images (each set contains the original image + four denoising results in a random order) on a 5-point scale (1 is very poor – 5 is very good), and the average mean opinion score (MOS) and standard deviation of each method are recorded.

4 A pre-trained panoramic semantic segmentation network (fixed weight) is used to infer on the images before and after denoising, and mIoU (mean intersection-over-Union, %) is calculated to evaluate the impact of denoising on the backend task. Each method is repeated 3 times, and the mean ± std is reported.

The results of perceptual indicators, user MOS, and downstream semantic segmentation performance are shown in Table 6.

Table 6 Perceptual indicators, user MOS and downstream semantic segmentation performance results

Methods	*PSNR (dB)*	*SSIM*	*LPIPS (↓)*	*MS-SSIM*	*User MOS (1–5)*	*Segmentation mIoU (%)*
Proposed method	33.04 ± 0.05	0.928 ± 0.003	0.082 ± 0.004	0.951 ± 0.006	4.28 ± 0.22	68.2 ± 0.9
FFDNet	31.74 ± 0.08	0.889 ± 0.004	0.110 ± 0.006	0.934 ± 0.008	3.92 ± 0.27	65.0 ± 1.2
Frost-filter	30.58 ± 0.09	0.844 ± 0.007	0.132 ± 0.007	0.912 ± 0.010	3.51 ± 0.30	62.3 ± 1.4
Mean filter	30.30 ± 0.10	0.786 ± 0.007	0.155 ± 0.009	0.897 ± 0.012	3.10 ± 0.35	58.1 ± 1.6

The results in Table 6 show that the proposed method outperforms FFDNet, frost-filter, and mean filtering in all metrics, demonstrating its combined advantages in denoising quality and practical applications. The proposed method achieves a PSNR of 33.04 dB, approximately 1.30 dB higher than FFDNet, and a SSIM of 0.928, surpassing FFDNet's 0.889, demonstrating superior image reconstruction accuracy and structural fidelity. In terms of perceptual metrics, the proposed method achieves the lowest LPIPS value of 0.082, significantly outperforming FFDNet's 0.110, indicating a closer perceptual

similarity to the real image. The MS-SSIM also reaches the highest value of 0.951, further confirming its advantage in multi-scale structure recovery. In subjective evaluation, the proposed method achieves an average MOS of 4.28, surpassing FFDNet's 3.92, indicating a higher user satisfaction with its visual performance. In the downstream semantic segmentation task, the proposed method achieves a mean intersection over union (MIoU) of 68.2%, a 3.2 percentage point improvement over FFDNet's 65.0%, demonstrating that its denoising results can help improve the accuracy of backend vision tasks. Overall, this fully demonstrates the proposed method's ability to maintain high denoising performance while balancing perceptual quality and practical application effectiveness.

5 Conclusions

A panoramic image denoising method combining U-Net and ResNet-50 is implemented in this paper to improve image quality. Through the powerful feature extraction capability of ResNet-50, combined with the encoding-decoding structure and skip connection mechanism of U-Net, the images' feature representation and alignment precision can be enhanced, and essential details can be effectively preserved. The deep feature extraction provided by ResNet-50 can precisely capture key information such as the image's geometric structure and texture details, and U-Net ensures the restoration of details in the denoising process and avoids over-smoothing. This method exhibits superior denoising performance when faced with different noise types, significantly improves image quality, and performs well under multiple evaluation indicators. Although this method shows significant denoising effects in practical applications, it still faces challenges such as training dataset construction and improving computational efficiency. In particular, the problems of multi-modal noise and artefacts that may arise during the stitching process still need further study. In the future, advanced technologies, such as GAN, can be combined, and the training and computational efficiency of deep learning models can be further optimised, which is expected to further improve the denoising effect and quality of panoramic images, thereby promoting the development of fields such as virtual reality and geographic information systems.

Declarations

This work was sponsored in part by the The Youth Fund Project of the Natural Science Foundation of Hainan Province: Tower-type hierarchical algorithm and image generation research. (Grant no. 624QN285).and supported by the Education Department of Hainan Province: Research on Autonomous Navigation Method for Tropical Fruits Transportation Robot in Hainan Province. (Grant No: Hnky2024ZD-19).

All authors declare that they have no conflicts of interest.

References

Azad, R., Aghdam, E.K., Rauland, A., Jia, Y., Avval, A.H., Bozorgpour, A., ... and Merhof, D. (2024) 'Medical image segmentation review: the success of u-net', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, December, Vol. 46, No. 12, pp.10076–10095, DOI: 10.1109/TPAMI.2024.3435571

Chen, F., Li, S., Han, J., Ren, F. and Yang, Z. (2024) 'Review of lightweight deep convolutional neural networks', *Archives of Computational Methods in Engineering*, Vol. 31, No. 4, pp.1915–1937.

Chilukuri, P., Kumar, J.R., Anusuya, R. and Prabhu, M.R. (2022) 'Auto encoders and decoders techniques of convolutional neural network approach for image denoising in deep learning', *Journal of Pharmaceutical Negative Results*, Vol. 13, No. 4, pp.1036–1040.

Cui, Y., Jiang, G., Yu, M. and Song, Y. (2022) 'Local visual and global deep features based blind stitched panoramic image quality evaluation using ensemble learning', *IEEE Transactions on Emerging Topics in Computational Intelligence*, Vol. 6, No. 5, pp.1222–1236.

Dey, R., Bhattacharjee, D. and Nasipuri, M. (2020) 'Image denoising using generative adversarial network', in Mandal, J. and Banerjee, S. (Eds.): *Intelligent Computing: Image Processing Based Applications. Advances in Intelligent Systems and Computing*, Vol. 1157, Springer, Singapore, https://doi.org/10.1007/978-981-15-4288-6_5.

El Helou, M. and Süsstrunk, S. (2020) 'Blind universal Bayesian image denoising with Gaussian noise level learning', *IEEE Transactions on Image Processing*, Vol. 29, pp.4885–4897, DOI: 10.1109/TIP.2020.2976814.

Elad, M., Kawar, B. and Vaksman, G. (2023) 'Image denoising: the deep learning revolution and beyond – a survey paper', *SIAM Journal on Imaging Sciences*, Vol. 16, No. 3, pp.1594–1654.

Fan, C.M., Liu, T.J. and Liu, K.H. (2022) 'SUNet: Swin transformer UNet for image denoising', in *2022 IEEE International Symposium on Circuits and Systems* (*ISCAS*), May, pp.2333–2337, IEEE.

Günen, M.A. and Beşdok, E. (2023) 'Effect of denoising methods for hyperspectral images classification: DnCNN, NGM, CSF, BM3D and Wiener', *Mersin Photogrammetry Journal*, Vol. 5, No. 1, pp.1–9.

Gurrola-Ramos, J., Dalmau, O. and Alarcón, T. (2022) 'U-Net based neural network for fringe pattern denoising', *Optics and Lasers in Engineering*, Vol. 149, p.106829, https://doi.org/10.1016/j.optlaseng.2021.106829.

Gurrola-Ramos, J., Dalmau, O. and Alarcón, T.E. (2021) 'A residual dense u-net neural network for image denoising', *IEEE Access*, Vol. 9, pp.31742–31754, DOI: 10.1109/ACCESS.2021.3061062.

Han, S.H., Heo, J.H., Seong, E.S. and Lim, D.H. (2023) 'U-Net based GAN with deformable convolution for CT image denoising', *The Korean Data and Information Science Society*, Vol. 34, No. 4, pp.567–585.

Hien, P.T. and Hong, I.P. (2023) 'Millimeter wave SAR imaging denoising and classification by combining image-to-image translation with ResNet', *IEEE Access*, Vol. 11, pp.70203–70215, DOI: 10.1109/ACCESS.2023.3293644.

Ilesanmi, A.E., Ilesanmi, T. and Gbotoso, G.A. (2023) 'A systematic review of retinal fundus image segmentation and classification methods using convolutional neural networks', *Healthcare Analytics*, Vol. 4, p.100261, https://doi.org/10.1016/j.health.2023.100261.

Izadi, S., Sutton, D. and Hamarneh, G. (2023) 'Image denoising in the deep learning era', *Artificial Intelligence Review*, Vol. 56, No.7, pp.5929–5974.

Kang, H., Park, C. and Yang, H. (2024) 'Evaluation of Denoising performance of ResNet deep learning model for ultrasound images corresponding to two frequency parameters', *Bioengineering*, Vol. 11, No. 7, p.723.

Kim, H.S., Ha, E.G., Lee, A., Choi, Y.J., Jeon, K.J., Han, S.S. and Lee, C. (2023) 'Refinement of image quality in panoramic radiography using a generative adversarial network', *Dentomaxillofacial Radiology*, Vol. 52, No. 5, p.20230007.

Komatsu, R. and Gonsalves, T. (2020) 'Comparing u-net based models for denoising color images', *AI*, Vol. 1, No. 4, pp.465–486.

Latha, H.N. and Sahay, R.R. (2020) 'A local modified U-net architecture for image denoising', *Reconstruction*, Vol. 8, p.14, DOI: 10.46501/IJMTSTCIET27.

Lv, M. and Cai, A. (2025) 'Image fusion using a transfer learning-based convolutional neural network', *International Journal of Information and Communication Technology*, Vol. 26, No. 3, pp.72–88.

Mannam, V., Zhang, Y., Zhu, Y., Nichols, E., Wang, Q., Sundaresan, V., ... and Howard, S.S. (2022) 'Real-time image denoising of mixed Poisson–Gaussian noise in fluorescence microscopy images using Image J', *Optica*, Vol. 9, No. 4, pp.335–345.

Mansour, Y. and Heckel, R. (2023) 'Zero-shot noise2noise: efficient image denoising without any data', in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.14018–14027.

Mohammadi, M.S. and Chabok, S.J.S.M. (2023) *Deep Convolutional Framelet Denoising for Panoramic by Mixed Wavelet Integration*, arXiv preprint arXiv:2302.10306.

Neji, H., Ben Halima, M., Nogueras-Iso, J., Hamdani, T.M., Lacasta, J., Chabchoub, H. and Alimi, A.M. (2024) 'Doc-Attentive-GAN: attentive GAN for historical document denoising', *Multimedia Tools and Applications*, Vol. 83, No. 18, pp.55509–55525.

Punn, N.S. and Agarwal, S. (2022) 'Modality specific U-Net variants for biomedical image segmentation: a survey', *Artificial Intelligence Review*, Vol. 55, No. 7, pp.5845–5889.

Sano, A., Nishio, T., Masuda, T. and Karasawa, K. (2021) 'Denoising PET images for proton therapy using a residual U-net', *Biomedical Physics and Engineering Express*, Vol. 7, No. 2, p.25014.

Siddique, N., Paheding, S., Elkin, C.P. and Devabhaktuni, V. (2021) 'U-net and its variants for medical image segmentation: A review of theory and applications', *IEEE Access*, Vol. 9, pp.82031–82057, https://doi.org/10.48550/arXiv.2011.01118.

Soniya, S. and Sriharipriya, K.C. (2024) 'Integrating Kalman filter noise residue into U-Net for robust image denoising: the KU-Net model', *Scientific Reports*, Vol. 14, No. 1, p.23641.

Tripathi, M. (2021) 'Facial image denoising using AutoEncoder and UNET', *Heritage and Sustainable Development*, Vol. 3, No. 2, pp.89–96.

Urli, F., Somero, M., Snidaro, L., Johnson, C., Vallisa, T. and Visentini, I. (2024) 'FeU-Net: overcomplete representations with large kernels for edge detection', in *2024 27th International Conference on Information Fusion* (*FUSION*), July, pp.1–7, IEEE.

Vojtekova, A., Lieu, M., Valtchanov, I., Altieri, B., Old, L., Chen, Q. and Hroch, F. (2021) 'Learning to denoise astronomical images with U-nets', *Monthly Notices of the Royal Astronomical Society*, Vol. 503, No. 3, pp.3204–3215.

Zaeri, N. and Qasim, R.R. (2024) 'Resilient recognition system for degraded thermal images using convolutional neural networks', *International Journal of Information and Communication Technology*, Vol. 25, No. 5, pp.50–71.

Zhu, M.L., Zhao, L.L. and Xiao, L. (2022) 'Image denoising based on GAN with optimization algorithm', *Electronics*, Vol. 11, No. 15, p.2445.