ISSN 2616-5775 Vol. 7, Issue 7: 9-16, DOI: 10.25236/AJCIS.2024.070702

Research on Dual-Channel Image Denoising Algorithm Based on Deformable Convolution

Yuan Chen[1,a], Yun Yang[1,b,*], Palizhati·Wusiman[1,c], Yanan Wu[2,d], Hong Yang[3,e]

[1]School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, Shaanxi, China
[2]Xi'an Institute of Metrology Technology, Quality Management Department, Xi'an, Shaanxi, China
[3]Xi'an Fude Medical Electronics Co., Ltd., Quality Management Department, Xi'an, Shaanxi, China
[a]1337738950@qq.com, [b]yangyun11@163.com, [c]2205439188@qq.com, [d]89716594@qq.com, [e]11679021@qq.com
[*]Corresponding author

Abstract: *In response to the issues of excessive smoothing and detail loss in existing image denoising algorithms, a dual-channel image denoising algorithm based on deformable convolution is proposed. Firstly, a noise estimation network is used to obtain a noise level map, enhancing the detail preservation capability of the edge feature extraction block enhancement network. Then, efficient channel attention is combined to focus on key channel features, effectively capturing noise characteristics. Finally, the deformable convolution with deformable learning kernels is integrated into the dual-channel denoising network to extract offset pixels of feature mappings, avoiding excessive image smoothing. Experimental results show that on the Set12 dataset with a noise level of 25, compared with commonly used algorithms, the proposed algorithm improves the average peak signal-to-noise ratio and structural similarity by 0.08dB to 0.62dB and 0.001 to 0.02, respectively, while removing Gaussian noise and preserving image details as much as possible.*

Keywords: *Image denoising; Dense block; Feature extraction; Attention mechanism*

1. Introduction

Image denoising is a fundamental computer vision task that indirectly influences higher-level visual tasks. The application of image denoising techniques to remove noise is crucial for recovering potential observations from a given degraded image[1]. The goal of image denoising is to eliminate or reduce noise interference in images, making them clearer, with more apparent details, thereby enhancing the visual effects and information representation capabilities of the images.

With the development of deep learning, methods based on Convolutional Neural Networks (CNNs) have gradually become mainstream in the field of denoising. CNNs can automatically learn effective feature representations and relationship mappings from large-scale data, leading to significant breakthroughs in image denoising tasks[2]. The Deep Convolutional Residual Denoising Neural Network (DnCNN) proposed in literature [3] improves denoising performance while accelerating model training. However, increasing depth limits its feature extraction capability, making it difficult to capture context information for larger image structures. The Fast and Flexible Denoising Network (FFDNet) proposed in literature [4] decomposes the input image into 4 sub-images and a horizontally noisy image generated from user input parameters, outputting 4 denoised sub-images. Using Depth-to-Space reduces parameters and enhances network efficiency and flexibility. The Convolutional Blind Denoising Network (CBDNet) from literature [5] considers signal interference from both image signals and device acquisition, generating numerous noisy images for training. It designs a noise estimation network on top of FFDNet to estimate the σ parameter originally inputted by the user, improving the network's generalization ability. The Attention-guided Denoising Convolutional Neural Network (ADNet) from literature [6] uses attention mechanisms to guide denoising, integrating dilated convolutions with conventional convolutions and long-pathways to extract precise noise information in complex backgrounds, enhancing denoising performance and image representation. The Batch-renormalization Denoising Network (BRDNet) proposed in literature [7] uses skip connections to speed up network training, effectively addressing training saturation issues. However, BRDNet trains for a specific noise level, limiting its flexibility and practicality. The Dual Denoising Network

ISSN 2616-5775 Vol. 7, Issue 7: 9-16, DOI: 10.25236/AJCIS.2024.070702

(DudeNet) from literature [8] proposes global and local features separately using two sub-networks through a sparse mechanism, compressing and reconstructing the extracted features finely.

This paper proposes a dual-channel image denoising algorithm based on deformable convolution (DCDN). It designs a noise estimation network to estimate the noise level map, enhances network feature extraction capability through the edge feature extraction module and efficient channel attention; proposes a dual-channel denoising network to expand network width, enhance learning ability, integrate deformable convolution to restore noise distribution based on surrounding pixels, and effectively improve denoising performance. Experiments show that DCDN achieves more robust and efficient denoising performance while reducing the model's parameter count, ensuring image quality.

2. Denoising Network

2.1. Network Architecture

The network architecture of the dual-channel image denoising algorithm based on deformable convolution is shown in Figure 1.

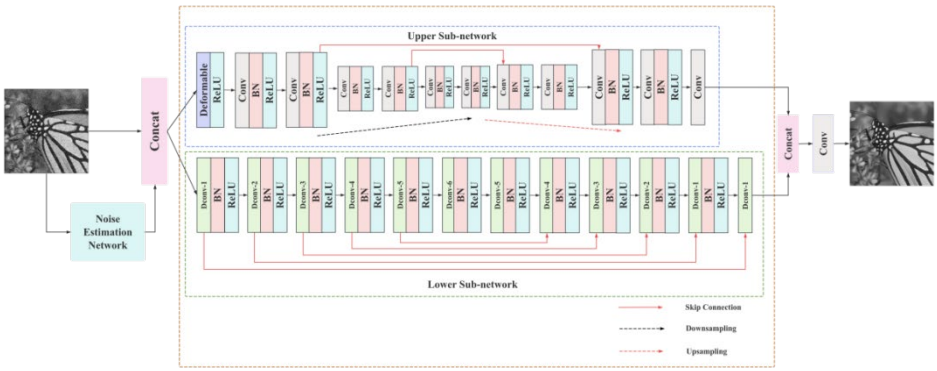

Figure 1: Dual-channel image denoising network structure based on deformable convolution

DCDN consists of a noise estimation network and a dual-channel denoising network. Firstly, the noise estimation network is used to estimate the noise level map in the noisy image. It combines an edge feature extraction block to enhance the network's ability to preserve image details and introduces efficient channel attention to focus on key channel features, effectively capturing noise characteristics. Next, the noise level map and the noisy image are inputted into the dual-channel denoising network. The upper branch of this network introduces deformable convolution with deformable learning kernels, extracting more representative noise features, thus avoiding excessive image smoothing.

2.2. Noise estimation network

The noise estimation network replaces traditional convolution with an edge feature extraction block in a manner similar to residual structures, refining image details. Combined with Effective Channel Attention (ECA)[9], the network focuses more on channels with key information in the image, enhancing efficiency and accuracy in estimating noise levels. The overall structure is illustrated in Figure 2.

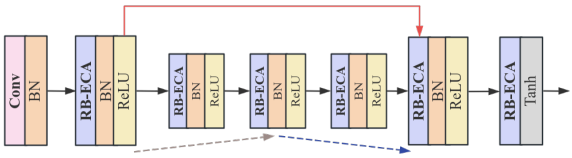

Figure 2: Noise estimation network

2.2.1. Edge feature extraction block

To more fully extract and utilize image feature information, an edge feature extraction module is proposed. It combines channel attention mechanism and residual block (residual block with ECA, RB-ECA) to refine image details. The structure is illustrated in Figure 3.

ISSN 2616-5775 Vol. 7, Issue 7: 9-16, DOI: 10.25236/AJCIS.2024.070702

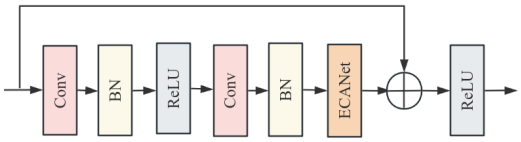

Figure 3: Edge feature extraction block structur

Since noise estimation requires extracting and integrating image features to accurately estimate noise levels in images, introducing attention mechanisms can help the model focus on key channel features, enhancing the model's ability to capture noise characteristics.

ECA achieves non-dimensional reduction local cross-channel interaction through 1D global average pooling, obtaining the global average value for each channel. Then, it calculates channel weights using linear transformation and activation functions, reflecting the importance of each channel in the entire feature map. Finally, using these weights, it weights the original feature tensor to generate an attention-controlled feature representation. ECA effectively captures important channel information in the feature map, enhances the model's perception of key features, and improves the accuracy and robustness of noise estimation. The structure is illustrated in Figure 4.

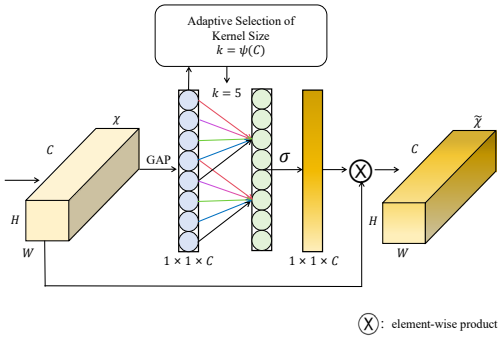

Figure 4: Efficient channel attention structure

2.3. Dual-channel denoising network

The dual-channel denoising network consists of an upper U-shaped subnet and a lower subnet containing dilated convolutions, with each subnet having 12 convolutional layers. The upper U-shaped subnet is designed to learn high-level semantic information and extract image features to better capture the overall structure of the image. On the other hand, the lower subnet is responsible for learning local details and texture information to enhance the capture of fine image features.To promote information flow and gradient propagation, two skip connections are introduced in the upper subnet for feature aggregation. In the lower subnet, symmetric skip connections are used to accelerate network training and improve detail preservation, enabling effective integration and interaction of features from different hierarchical levels, thereby enhancing feature learning and detail capture capabilities and further improving image denoising performance. Additionally, after cascading the subnets, an additional convolutional layer with a kernel size of 1 is added to make the network suitable for both grayscale and color images.

2.3.1. Deformable blocks

The deformable block utilizes deformable convolution to obtain more representative noise features.That is the first layer is deformable Conv + ReLU, which denote a deformable convolution and an activation function of ReLU.A deformable convolution uses relations of surrounding pixels to restore the position of the original pixel to enhance clarity of predicted image. Also,ReLU is used to convert obtained linear features into non- linear features. Input and output channels of this layer are 3 and 64. Specifically, if given images is grey, its input channel is 1. Also, its kernel size is 3×3. The 2nd layer-10th layer are composed of Conv+BN+ReLU, which denotes the com-bination of a convolutional layer, BN and ReLU. Their input and output channels are 64. Also, their kernel sizes are 3×3.BN is used to normalise data to accelerate network speed. To visually express the mentioned process, we define some symbols as follows. DC and R are used to represent a deformable convolution and ReLU, respectively. C and B express functions of a convolution and BN, respectively. Also, RBC denotes a Conv + BN + ReLU. 10 RBC is used to stand for eleven stacked conv + BN + ReLU.

ISSN 2616-5775 Vol. 7, Issue 7: 9-16, DOI: 10.25236/AJCIS.2024.070702

$$O_{DB} = DB(I_n) = 11RBC(R(DC()I_n)) \tag{1}$$

According to analysis of mentioned formulae, we can see that deformable convolution has good performance to obtain more contextual information. Thus, we put a deformable convolutional layer in the front of the whole denoising as Figure 1, where its parameters are input channel of 3, output channel of 64, kernel size of 3×3. Subsequently, a ReLU can convert obtained linear features into non-linear features. To further learn accurate features, a 12-layer the combination of convolutional layer, BN and ReLU (i.e., Conv + BN + ReLU) are embedded in the DB, where their parameters are input channel and output channel of 64, kernel size of 3×3. BN is used to normalise obtained features and ReLU is exploited to convert obtained linear features into non-linearity.

2.4. Loss function

To better preserve high-frequency texture information and reduce blurry and overly smooth visual effects, this paper adopts the Charbonnier loss[10] as the reconstruction loss to optimize DCDN. Additionally, referencing literature [11], edge loss is used to constrain the high-frequency components between the ground truth image and the denoised image. Since DCDN uses a noise estimation network to estimate the noise level map in noisy images, the total variation regularizer from literature [12] is applied to constrain the smoothness of the estimated noise level. Therefore, the loss function of DCDN can be defined as:

$$\mathcal{L} = \mathcal{L}_{char}(\hat{x}, x) + \lambda_{edge}\mathcal{L}_{edge}(\hat{x}, x) + \lambda_{TV}\mathcal{L}_{TV}\big(\sigma(y)\big) \tag{2}$$

Thereinto:$\mathcal{L}_{char}, \lambda_{\text{edge}}, \lambda_{\text{TV}}$, denote Charbonnier loss, edge loss, and total variational regularizer, respectively, and setλ_{edge} and λ_{TV} the sum to 0.1 and 0.05, respectively, as shown in Eq. (3), (4), and (5).

$$\mathcal{L}_{char} = \sqrt{\|\hat{x} - x\|^2 + \epsilon^2} \tag{3}$$

where the constant is ϵ set to10^{-3}.

$$\mathcal{L}_{edge} = \sqrt{\|\Delta(\hat{x}) - \Delta(x)\|^2 + \epsilon^2} \tag{4}$$

where Δdenotes the Laplace operator.

$$\mathcal{L}_{TV} = \|\nabla_h \sigma(y)\|_2^2 + \|\nabla_v \sigma(y)\|_2^2 \tag{5}$$

where$\nabla_h(\nabla_{\text{v}})$denotes the gradient operator in the horizontal (vertical) direction.

2.5. Experimental evaluation indexes

The peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) are used as quantitative indicators to evaluate the denoising effect. A noiseless image of a certain size and a noisy image are defined to obtain MSE, with the calculation formula as follows:

$$\text{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \tag{6}$$

PSNR is a metric used to measure the error between two images, typically the original image and a compressed or restored image, in terms of the errors introduced by compression or restoration[13]. The formula for PSNR is as follows:

$$\text{PSNR} = 10\lg\left[\frac{(2^n-1)^2}{MSE}\right] \tag{7}$$

The Structural Similarity Index (SSIM) is a measure that compares two images based on three components: luminance, contrast, and structure. It provides a similarity score ranging from 0 to 1, where a value closer to 1 indicates greater similarity between the two images. The calculation formula is as follows:

$$\text{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{8}$$

In the formula, μ_{x} and μ_{y} represent the means of images x and y respectively, σ_{x} and σ_{y} represent their respective standard deviations, and σ_{xy} represents the covariance of images x and y. The constants c_1 and c_2 are regularization parameters.

ISSN 2616-5775 Vol. 7, Issue 7: 9-16, DOI: 10.25236/AJCIS.2024.070702

3. Experimental Results and Analysis

3.1. Experimental datasets

The DIV2K dataset is used for network model training and contains high-quality images from fields such as movies, nature, and animations[14]. The Set12 and BSD68 datasets are used for grayscale image denoising evaluation, where Set12 includes 12 grayscale images involving different scenes, while the BSD68 dataset contains 68 grayscale images; the CBSD68 dataset is used for color image denoising evaluation.

When training the denoising model, to ensure consistency in input data, the size of the training images is rescaled to 512×512pixels and converted to grayscale; subsequently, random cropping technique is employed to crop the images into 180×180 pixel patches, increasing data diversity; finally, additive white Gaussian noise with noise levels ranging from 0 to 75 is added to each clean patch to produce noisy image patches.

3.2. Experimental setup

The experiment is run on a Windows system with an Intel(R) Core(TM) i5-8300H CPU @ 2.30GHz and an NVIDIA RTX 4090 Ti GPU with 24GB of memory. The network framework used is PyTorch. During the training process, the batch size is set to 16, and the Adam optimization algorithm is utilized with an initial learning rate of 10^{-4}.

3.3. Analysis of experimental results

3.3.1. Quantitative comparisons

Table 1 compares the average PSNR (dB) results of different image denoising methods on the Set12 and BSD68 datasets. On the Set12 dataset, when σ= 15, 25, 50, the average PSNR values of DCDN increased by 0.11dB, 0.08dB, and 0.08dB compared to DCBDNet, and by 0.02dB, 0.16dB, and 0.36dB compared to DnCNN-S, respectively. Under the same conditions on the BSD68 dataset, the average PSNR values of DCDN increased by 0.10dB, 0.10dB, and 0.08dB compared to DCBDNet, and by 0.03dB, 0.11dB, and 0.22dB compared to DnCNN-S, respectively. DCDN achieved superior PSNR metrics, indicating that it has better denoising performance when dealing with different levels of noise.

Table 1: Comparison of average PSNR/dB for different denoising methods

Datasets		Set12			BSD68	
σ	σ=15	σ=25	σ=50	σ=15	σ=25	σ=50
BM3D	32.37	29.97	26.72	31.07	28.57	25.62
WNNM	32.70	30.26	27.05	31.37	28.83	25.87
DnCNN-S	32.84	30.43	27.18	31.72	29.23	26.23
IRCNN	32.77	30.38	27.14	31.63	29.15	26.19
FFDNet	32.77	30.44	27.32	31.63	29.19	26.29
ADNet	**32.98**	30.58	27.37	31.74	29.25	26.29
DCBDNet	32.75	30.51	27.46	31.65	29.24	26.37
DCDN	32.86	**30.59**	**27.54**	**31.75**	**29.34**	**26.45**

Table 2 shows the average SSIM comparison results of different image denoising methods on the Set12 and BSD68 datasets. It can be seen that DCDN achieved the highest SSIM values.

From Table 2, it can be observed that on the Set12 dataset, when σ=15, 25, 50, the average SSIM values of DCDN increased by 0.002, 0.001, and 0.003 compared to DCBDNet, and by 0.002, 0.005, and 0.014 compared to DnCNN-S, respectively. Under the same conditions on the BSD68 dataset, the average SSIM values of DCDN increased by 0.004, 0.005, and 0.003 compared to DCBDNet, and by 0.002, 0.006, and 0.011 compared to DnCNN-S, respectively. This indicates that DCDN is better at preserving the structural information and visual quality of images in the image denoising task.

Table 3 shows the average PSNR (dB) comparison results of different image denoising methods on the CBSD68 dataset. From the table, it is evident that DCDN achieved the highest PSNR metrics.

ISSN 2616-5775 Vol. 7, Issue 7: 9-16, DOI: 10.25236/AJCIS.2024.070702

Table 2: Comparison of average SSIM for different denoising methods

Datasets		Set12			BSD68	
σ	σ=15	σ=25	σ=50	σ=15	σ=25	σ=50
BM3D	0.896	0.851	0.766	0.872	0.802	0.687
WNNM	0.894	0.846	0.756	0.878	0.810	0.698
DnCNN-S	0.903	0.862	0.783	0.891	0.828	0.719
IRCNN	0.901	0.860	0.780	0.888	0.825	0.717
FFDNet	0.903	0.864	0.791	0.890	0.830	0.726
ADNet	0.905	0.865	0.791	0.892	0.829	0.722
DCBDNet	0.902	0.865	0.794	0.889	0.829	0.727
DCDN	**0.904**	**0.866**	**0.797**	**0.893**	**0.834**	**0.730**

Table 3: Comparison of average PSNR for different denoising methods on CBSD68

Datasets		CBSD68	
σ	σ=15	σ=25	σ=50
CBM3D	33.52	30.71	28.89
DSNetB	33.91	31.28	28.05
CDnCNN-S	33.89	31.23	29.58
IRCNN	33.86	31.16	27.86
FFDNet	33.87	31.21	27.96
BRDNet	34.10	31.43	28.16
ADNet	33.99	31.31	28.04
DCBDNet	34.01	31.41	28.22
DCDN	**34.13**	**31.52**	**28.32**

3.3.2. Qualitative comparison

To more intuitively verify the denoising capability of the network model presented in this paper, DCDN is compared with BM3D, DnCNN-S, ADNet, and DCBDNet. The comparison focuses on the restoration of details in the "Boat.png" image from the Set12 dataset, specifically the text on the boat, and the butterfly's neck details in "Monar.png," with the results being magnified for comparison.When the noise level is set to 25, the denoising results for the two images are shown in Figures 5 and 6.

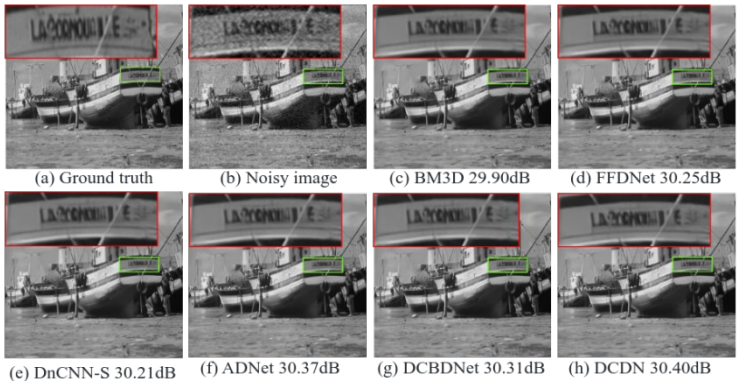

Figure 5: Image of text details on the boat at σ =25

In the restoration of the text details on the boat, DCDN demonstrated significant advantages, producing high clarity in the restored text and achieving the highest PSNR values; whereas BM3D, FFDNet, and DnCNN-S exhibited smearing and loss of details. ADNet and DCBDNet resulted in more blurred text restorations.

Regarding the restoration of the butterfly's neck details, although DCDN did not dominate in terms of PSNR values, it avoided excessive distortion and better preserved the neck feature information, enhancing the overall image quality; whereas BM3D and FFDNet showed noticeable blurring and twisting phenomena, affecting visual quality. DnCNN-S was inaccurate in restoring detail textures, exhibiting local distortions.

ISSN 2616-5775 Vol. 7, Issue 7: 9-16, DOI: 10.25236/AJCIS.2024.070702

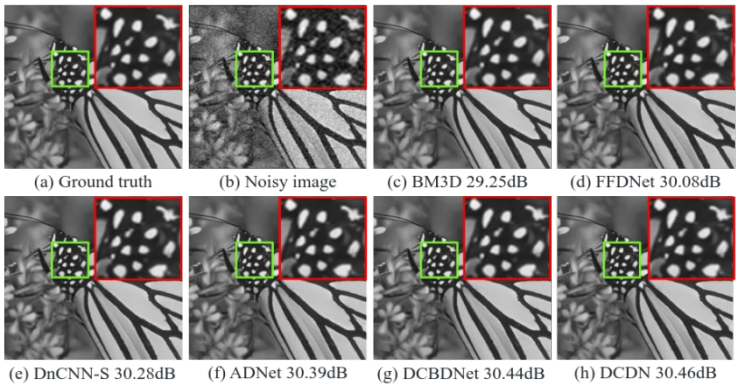

Figure 6: Detailed image of the butterfly's neck at $\sigma = 25$

3.4. Ablation experiments

To validate the effectiveness of the network design, a on networks with and without the ablation experiments were conducted on networks with and without RB-ECA and with and without Deformable Conv. Training was carried out under the same training set conditions at a noise level of 25, and the results are shown in Table 4.

Table 4: The disintegration experiment results on the Set12 dataset

experiments	RB-ECA	Deformzble Conv	PSNR
1	√	×	30.43
2	×	√	30.50
3	×	×	30.32
4	√	√	30.59

From Table 4, it can be seen that without incorporating any modules, the model only achieved a PSNR value of 30.32dB; with the introduction of RB-ECA alone, the model's PSNR value increased by 0.11dB, indicating that RB-ECA enhances the network's feature extraction capability for noisy images, thereby improving the denoising effect; with the introduction of Deformable Conv alone, the model's PSNR value increased by 0.18dB, indicating that the deformable convolution can extract offset pixels from feature maps, enhancing noise identification capabilities; when both RB-ECA and Deformable Conv were introduced, a PSNR value of 30.59dB was obtained, achieving the best denoising performance.

3.5. Network performance analysis

When the noise level is set to 25, the algorithms BM3D, DeamNet, DRUNet, and DCDN are used to process grayscale images of sizes 256x256, 512x512, and 1024x1024, respectively. The average processing time for each algorithm to handle one image is compared, as shown in Table 5.

Table 5: Different algorithms comparison on denoising speed

Methods	Device	256×256	512×512	1024×1024
BM3D	CPU	0.458	2.354	9.782
DeamNet	GPU	0.054	0.121	0.392
DRUNet	GPU	0.068	0.106	0.276
DCDN	GPU	0.047	0.078	0.186

From Table 5, it is evident that DCDN demonstrates a significant advantage in terms of processing speed when compared to the traditional denoising method BM3D. Compared to DeamNet and DRUNet, which also utilize GPU processing, DCDN shows a substantial improvement in processing speed, with only a slight compromise in performance while preserving image details as much as possible.

4. Conclusion

This paper proposes a dual-channel image denoising algorithm based on deformable convolutions. Initially, a noise estimation network is designed to enhance the model's flexibility, incorporating an edge feature extraction block and an efficient channel attention mechanism to extract image features.

ISSN 2616-5775 Vol. 7, Issue 7: 9-16, DOI: 10.25236/AJCIS.2024.070702

Subsequently, a dual-channel denoising network captures image features to improve denoising performance. By integrating deformable convolutions, the algorithm restores the noise distribution based on surrounding pixels, further enhancing the imaging quality of the denoised image. Experimental results demonstrate that DCDN possesses strong robustness in image denoising.

References

[1] C. Tian, L. Fei, W. Zheng, et al. Deep learning on image denoising: An overview[J]. Neural Networks, 2020, 131: 251–275.
[2] Zhang Q, Xiao J, Tian C, et al. A robust deformed convolutional neural network (CNN) for image denoising [J]. CAAI Transactions on Intelligence Technology, 2023, 8(2): 331-342.
[3] K. Zhang, W. Zuo, Y. Chen, et al. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising[J]. IEEE Transactions on Image Processing, 2017, 26 (7): 3142–3155.
[4] Zhang K, Zuo W, Zhang L. FFDNet: Toward a fast and flexible solution for CNN-based image denoising [J]. IEEE Transactions on Image Processing, 2018, 27(9): 4608-4622.
[5] Guo S, Yan Z, Zhang K, et al. Toward convolutional blind denoising of real photo graphs[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 1712-1722.
[6] TIAN C, XU Y, LI Z, et al. Attention-guided CNN for image denoising[J]. Neural Networks, 2020, 124: 117-129.
[7] TIAN C, XU Y, ZUO W. Image denoising using deep CNN with batch renormalization[J]. Neural Networks, 2020, 121:461-473.
[8] TIAN C, XU Y, ZUO W, et al. Designing and trainingofa dual CNN forimage denoising[J]. Knowledge-Based Systems, 2021, 226:106949.
[9] Wang, Q., Wu, B., Zhu, P. et al. Eca-net: Efficient channel attention for deep convolutional neural networks[J]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR),2020.
[10] W. Lai, J. Huang, N. Ahuja, et al. Deep laplacian pyramid networks for fast and accurate super-resolution[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2017, 624–632.
[11] K. Jiang, Z. Wang, P. Yi, et al. Multi-scale progressive fusion network for single image deraining[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2020, 8346–8355.
[12] S. Guo, Z. Yan, K. Zhang, et al. Toward convolutional blind denoising of real photographs[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2019, 1712–1722.
[13] Mahdaoui A E, Ouahabi A, Moulay M S. Image denoising using a compressive sensing approach based on regularization constraints[J]. Sensors, 2022, 22(6): 2199.
[14] AGUSTSSON E, TIMOFTE R. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study[C]. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, 1122–1131.