An efficient U-shaped transformer network for low-light power image denoising

J. Zhang [1], W.X. Huang [1], M.X. Lu [1], L.W. Li [1], X. Wang [1], Y.P. Shen [1], Y.F. Wang [1]
[1] College of Electrical and Information Engineering
Zhengzhou University of Light Industry, Zhengzhou 450002, China

Abstract

Unmanned aerial vehicle (UAV) inspection of transmission lines has been widely applied in recent years. However, in low-light weather conditions, random noise often appears in the captured transmission line images due to the combined effects of brightness, electromagnetic interference, and camera sensor limitations. This noise significantly undermines the quality and accuracy of the inspection. To address this challenge, we propose a novel transformer-based image denoising method called EUformer. First, we propose the Global Feature Compensator (GFC) module, which adaptively captures remote pixel dependencies for improved global image modelling. Second, we designed the Mixed-Gated feed-forward network (MG-FFN), to enhance the aggregation of local contextual information. Finally, the loss function is optimized by introducing a new regular term, effectively addressing negative effects such as artefacts in the reconstructed images. To assess the denoising capabilities of the EUformer model proposed in this study for transmission line images, we developed a benchmark dataset specifically for low-light transmission line image denoising. The results of extensive experiments demonstrate that the EUformer model achieves competitive performance while maintaining low complexity.

Keywords: deep learning, transformer, low-light transmission line, image denoising.

Citation: Zhang J, Huang WX, Lu MX, Li LW, Wang X, Shen YP, Wang YF. An efficient U-shaped transformer network for low-light power image denoising. Computer Optics 2025; 49(5): 775-784. DOI: 10.18287/2412-6179-CO-1629.

Introduction

Thanks to the significant advantages of low cost and high mobility in obtaining high-quality images, UAV inspection plays a crucial role in grid fault detection. The process begins with professionals using specialized equipment to acquire images, which are then analyzed by professionals and relevant algorithms [1]. However, in low-light weather conditions, UAV inspection is affected by factors such as scene brightness, electromagnetic interference from transmission lines, and the impact of camera sensors on the images collected [2]. These factors inevitably introduce random noise, which in turn affects the accuracy and efficiency of the subsequent transmission line inspection analysis. In recent years, various denoising methods have been proposed, which can generally be classified into traditional image-based a priori methods and learning-based methods.

Traditional image prior-based methods [3 – 9] perform denoising by modelling the noise distribution of an image using maximum likelihood estimation or Bayesian inference. These methods commonly utilize prior knowledge such as nonlocal self-similarity [4], sparse representation [5], and total variation [6]. Despite achieving satisfactory denoising results, these methods suffer from the following drawbacks: (1) extreme reliance on manual parameter setting, (2) a highly complex optimization process, and (3) limited generalization ability. Particularly, in the presence of high noise, these methods suffer from a serious degradation in denoising performance.

In recent years, there has been rapid advancement in learning-based denoising methods, which define the mapping relationship between noisy and real images through model training. Several CNN-based denoising methods [10, 11], have achieved excellent performance in denoising additive white Gaussian noise (AWGN). However, real-world noise is dependent on the signal and heavily influenced by the camera imaging pipeline. It is more complex than AWGN [12]. These aforementioned models experience significant performance degradation when applied to real-world noise. In addition, the effective receptive field of convolutional networks impedes the modeling of long-range dependencies in images.

To address these issues, recent works [13 – 17] have introduced the vision transformer (ViT) to the image denoising task. SwinIR [16] employs window multi-head self-attention (W-MSA) to compute self-attention and achieves superior performance compared to CNN. But it has higher computational complexity. Uformer [17] incorporates the Swin transformer into the U-type coding and decoding structure, significantly reducing computational complexity. In low-light images, it is difficult to distinguish target features from noise. It is particularly important to accurately model both the global and local aspects of such images. Therefore, the design of more effective modules for long-range modelling of

images and better fusion of local multi-scale features of images deserve further research.

To this end, we propose an Efficient U-shaped transformer Network for Image Denoising. The main component of the framework is the Efficient Transformer Block, which includes (1) a 16×16 window multi-head self-attention mechanism. This mechanism enhances the global modelling capability of the model and activates a larger number of pixels for image reconstruction. (2) A mixed-gated feed-forward network that extracts aggregated local multi-scale features to improve image denoising. Firstly, the Global Feature Compensator module is designed to use large-scale deformable convolution and simple channel attention, enabling a larger and denser receptive field for the global modelling of the image. Secondly, the mixed-gated feed-forward network is designed to further explore the local multi-scale features and promote the aggregation of useful features. Finally, a new penalty term is introduced in the loss function to further improve the performance of the network, taking into account the characteristics of the model and the loss function.

Generally, our contributions can be summarized as follows:

1) We propose an efficient U-shaped transform network for image denoising, which efficiently captures global representations and local detail features.
2) We have developed a dataset specifically for evaluating and investigating power image denoising algorithms applied to transmission line images. As far as I am aware, this is the first dataset in the existing literature that focuses on denoising low-light transmission line images.
3) Experimental results show that EUformer achieves competitive performance at a low computational cost.

1. Methodology

1.1. Overall pipeline

As shown in Figure 0, EUformer is based on a hierarchical encoder-decoder architecture. given a noisy image $I_0 \in R^{H\times W\times 3}$, EUformer first employs a 3×3 convolutional layer to extract shallow features $X_0 \in R^{C\times H\times W}$:

$$X_0 = \varphi(f_c^{in}(I_0)), \tag{1}$$

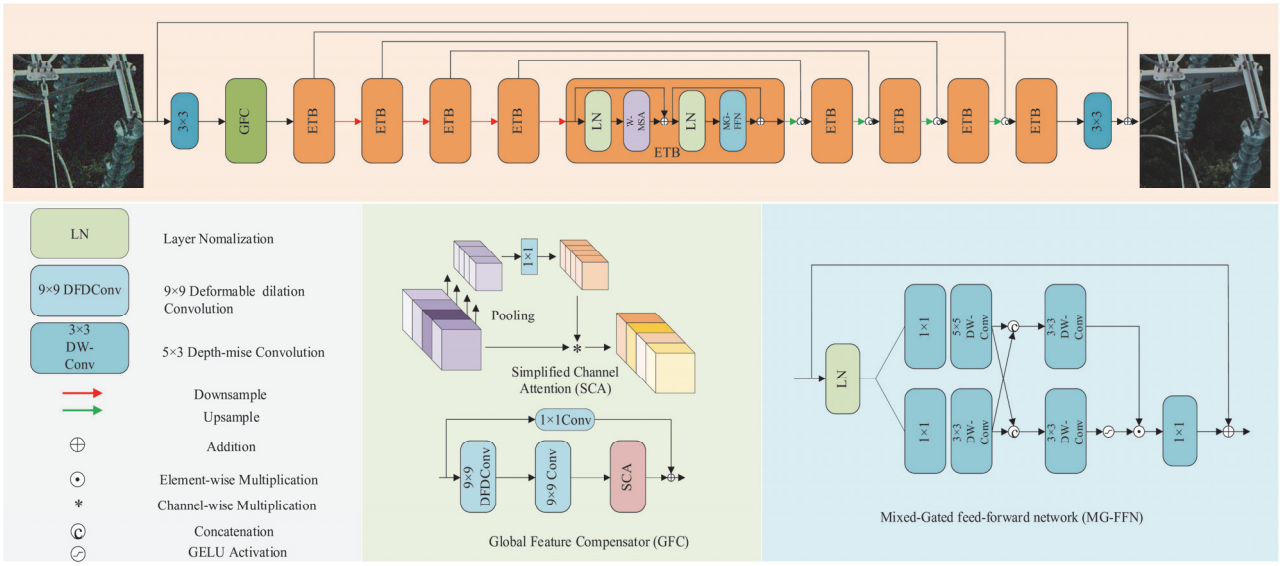


Fig. 1. The overall architecture of the proposed Efficient U-Shaped Transformer network for image denoising (EUformer), which mainly contains efficient transformer block (ETB) and mixed-gated feed-forward network (MG-FFN), and global feature compensator (GFC)

where H and W are the height and width of the noise image, C is the number of channels generated by convolution, f_c^{in} denotes the convolutional layer, φ denotes LeakyReLU activation layer. Next, the feature map X_0 is imported into the Global Feature Compensator (GFC) module, this module conducts global deep feature extraction of the input feature map $X_0 \in R^{C\times H\times W}$, through the utilization of large-size deformable dilation convolution, large-size depth-wise convolution, and simple channel attention. The subsequent output of deep features $X_1 \in R^{C\times H\times W}$:

$$X_1 = H_{GFC}(X_0), \tag{2}$$

where X_1 denotes the global deep feature extracted by the GFC, Then, these deep features X_1 pass through a 4-level U-shaped encoder-decoder and are transformed into deep features $X_d \in R^{2C\times H\times W}$:

$$X_d = H_{ETB(i)}(X_1), \tag{3}$$

To capture the hierarchical multi-scale features of noisy images, this 4-level U-shaped encoder-decoder architecture is implemented, with each level

corresponding to a specific scale. Each scale incorporates a residual connection between down-sampling based downscaling and up-sampling based upscaling. The number of ETBs varies across scales. The encoding stage gradually reduces the spatial size of the high-resolution input image and expands the number of channels for each scale. Conversely, the decoding stage decreases the number of channels of the low-resolution image and restores it to a high-resolution representation. Pixel-unshuffle and pixel-shuffle operations [19] are applied for feature down-sampling and up-sampling, respectively. Following the decoding stage, the depth feature X_d , undergoes a 3×3 convolutional layer to obtain the residual image $R \in R^{H\times W\times 3}$, Finally, the recovered image is obtained by adding the residual image $I = I_0 + R$.

1.2. Efficient transformer block

In existing Transformer models [16 – 17], to reduce the computational overhead of self-attention (SA), many employ non-overlapping shifted-window of size 8×8 to compute SA. However, some studies on vision tasks [20] have indicated that using size 8×8 window for attention computation may lead to the recovery of false textures due to the limited range of pixels involved. Moreover, the naive Feedforward Network (FFN) exhibits limited capability in capturing local information. To address these limitations, we have proposed a novel and efficient transformer block that functions as a feature extraction unit. The computation of an Efficient Transformer block is represented as:

$$X_l' = X_{l-1} + \mathrm{LW-MSA}(LN(X_{l-1})), \quad (4)$$

$$X_l = X_l' + \mathrm{MG-FFN}(LN(X_l')), \quad (5)$$

where LN denotes the layer normalization; X_l' and X_l denote the outputs from the large-window-based multi-head self-attention (LW-MSA) and Mixed-Gated feed-forward network (MG-FFN), which are described below.

Large-window-based multi-head self-attention: We employed the 16×16 window to compute self-attention. Because, many works illustrate large window computation of self-attention significantly expands the use of pixels [20], the impact of 8×8 windows are significantly mitigated. To compute the self-attention module, the input feature of size $H\times W\times C$ is first partitioned into HW/M^2 local windows of size $M\times M$, then the self-attention is computed within each window. By applying linear mappings, the Q, K, and V matrices are obtained for a local window feature $X_W \in R^{M^2\times C}$. Then the window-based self-attention is formulated as:

$$Attention(Q,K,V) = SoftMax(QK^T/\sqrt{d} + B)V, \quad (6)$$

where d represents the dimension of query/key. B denotes the relative position encoding and is calculated as [22].

Mixed-Gated feed-forward network: Traditional FFN typically have only two fully connected layers and GELU for feature transformation [16], with limited ability to extract local information. The local neighbouring pixels of an image are important references in image restoration, and previous work [17] often introduces deep convolution into FFN to improve the ability to express local contextual information, but this development ignores the local multi-scale features of the image. In fact, rich local multi-scale features and high-frequency detail features have been shown to play an important role in image denoising [18,23]. In this work, we introduce two key modifications to FFN to enhance representation learning: (1) multi-scale depth-wise convolution, and (2) gated mechanism. First, the input tensor $X_{l-1} \in R^{H\times W\times C}$ undergoes layer normalization. Then, it is passed through two 1×1 convolutions to expand the channels (expansion factor $\gamma = 2.6$). Subsequently, it is fed into two parallel depth-wise convolutions of size 3×3 and 5×5 , respectively. This allows for the comprehensive exploration of local multi-scale features using two different scales of depth-wise convolution. Afterwards, the output is further processed by cross-fusion, where it is sent into two parallel 3×3 depth-wise convolutions to capture local image structural information. The gate mechanism enriches the fusion of local contextual features and has been widely applied in the field of image restoration [24]. we introduce the gate mechanism after the two parallel 3×3 depth-wise convolutions to further enhance the fusion of local multi-scale features and fine details. In this way, the entire feature fusion procedure of the developed MG-FFN is formulated as:

$$\hat{X}_l = f_{1\times1}^{c}(LN(X_{l-1})), \quad (7)$$

$$X_l^{p1} = \sigma(f_{3\times3}^{dwc}(\hat{X}_l)), X_l^{p2} = \sigma(f_{5\times5}^{dwc}(\hat{X}_l)), \quad (8)$$

$$X_l^{s1} = \sigma(f_{3\times3}^{dwc}[X_l^{p1}, X_l^{p2}]), X_l^{s2} = \sigma(f_{3\times3}^{dwc}[X_l^{p1}, X_l^{p2}]), \quad (9)$$

$$X_l = f_{1\times1}^{c}(\sigma(X_l^{s1}) \odot X_l^{s2}) + X_{l-1}, \quad (10)$$

where $\sigma(\bullet)$ is a GeLU activation, $f_{1\times1}$ represents 1×1 convolution, $f_{3\times3}^{dwc}$ and $f_{5\times5}^{dwc}$ denote 3×3 and 5×5 depth-wise convolutions, $[\bullet]$ is the channel-wise concatenation. $\odot$ is element-wise multiplication, and LN denotes layer normalization.

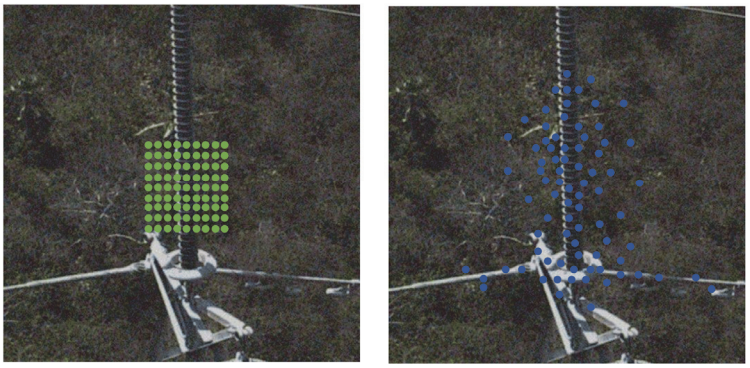
1.3. Global feature compensator

To enhance the ability of the model to capture global information, we developed the global feature compensator module. The detailed structure is shown in Fig. 1*b*. In this module, we aim to capture as much global information as possible by using a larger and denser receptive field. Deformable convolution and large convolution kernels are used in classical CNN networks to accurately capture local spatial information and provide large receptive fields, respectively [26]. Inspired by this, we combine deformable convolution and 9×9 convolution kernel to accurately capture global information while maintaining a larger and denser

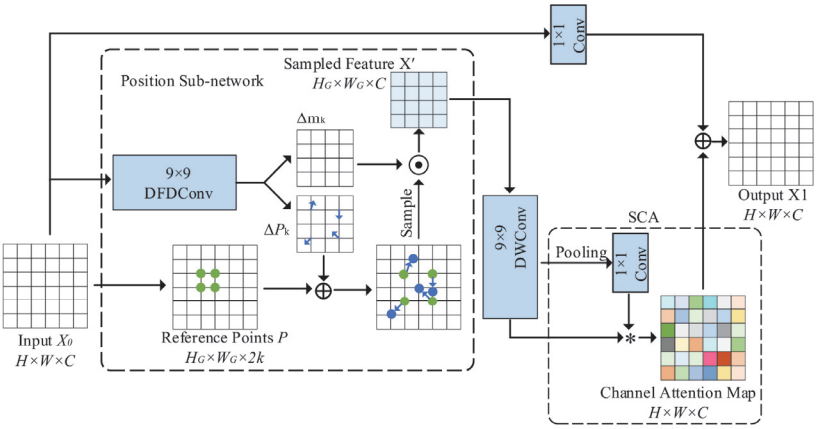
receptive field, as illustrated in Fig. 1*a*. Let x(p) and y(p) represent the features at location p in the input feature maps x and output feature maps y, respectively. The modulated deformable convolution can then be expressed as:

$$y(p)=\sum_{k=1}^{K} w_k \bullet x(p+p_k+\Delta p_k)\bullet \Delta m_k\,, \tag{11}$$

where Δp_k and Δm_k are the learnable offset and modulation scalar for the $k-th$ location, respectively.



(a)



(b)

Fig. 2. a) Comparison of Receptive Field between Regular Convolution and Deformable Dilation Convolution. b) The detailed structure of Global Feature Compensator. A group of offsets Δp are learned from the input x and added on the pre-defined reference points p to get deformable points

However, the use of a large kernel can lead to computational inefficiency. To address this, we enhance computational efficiency by incorporating dilation factors(dilation=2) into the convolution kernel while preserving the original receptive field. First perform a large-size dilation deformable convolution, followed by a 9×9 depth-wise convolution to further extract features. Additionally, we introduce Simplified Channel Attention [27] to weight the global feature information and improve the representational capability of the module. The computation of a global feature compensator is represented as:

$$X_1' = \sigma(f_{9\times9}^{dwc}(\sigma(f_{9\times9}^{dfdc}(X_0)))), \tag{12}$$

$$X_1 = \mathrm{SC}A(X_1') + f_{1\times1}^{c}(X_0), \tag{13}$$

where $\sigma(\bullet)$ is a GeLU activation, $f_{1\times1}$ represents 1×1 convolution, $f_{9\times9}^{dwc}$ and $f_{9\times9}^{dfdc}$ denote 9×9 depth-wise convolutions and 9×9 dilation deformable convolution, and SCA denotes Simplified Channel Attention.

1.4. Loss function

In the training, we aim to (1) facilitate model convergence and (2) restore high-quality denoised images that retain intricate details. The Total Variation Loss effectively mitigates image artifacts while preserving intricate details such as edges and textures, thereby enhancing the visual quality of the image [30]. To this end, we incorporate the Charbonnier loss [31] and introduce the Total Variation Loss as a regularization term. The overall loss with the hype-parameters λ is written as:

$$\ell = \ell_{char} + \lambda \ell_{TV}\,, \tag{14}$$

$$\ell_{char} = \sqrt{\| I_{Denoised} - I_{GT} \|^2 + \varepsilon^2}\,, \tag{15}$$

$$\ell_{TV} = \sum_{i,j} \| I_{i+1,j} - I_{i,j} \| + \| I_{i,j+1} - I_{i,j} \|\,, \tag{16}$$

where I_{GT} represents the ground-truth image, ε is an empirical value and $I_{Denoised}$ is the predicted image

after denoising by the network, $I_{i,j}$ is a pixel point of the input image.

2. Experiments

In this section, we first present the details of the dataset and experimental implementation, then perform extensive experiments and compare them with previous methods, and finally evaluate the effectiveness of each component through ablation experiments.

2.1. Dataset

Real-world denoising: For real-world image denoising, the SIDD [32] dataset was selected as the training set. We resized and cropped 320 image pairs from the SIDD dataset into image patches with a size of 128×128 for training. The performance of the proposed method was then evaluated on the SIDD and DND benchmark datasets [25], respectively, and the qualitative results on the DND dataset, were uploaded to a website for online evaluation.

Low-light transmission line image denoising: We have developed a comprehensive dataset, called TLs100, to benchmark existing image denoising methods and explore new techniques for low-light transmission line image denoising. We employ data synthesis to generate the benchmark dataset: We captured high-quality images under normal lighting conditions using a UAV. To expand more transmission line scenes, we also incorporated some images from the public dataset [28, 29]. For low-light transmission line image denoising, since the SIDD dataset includes numerous scenes with low-light, to conserve training resources and ensure a fair comparison with other methods on the TLs100 benchmark dataset, we choose the pre-training weights trained in Real-world denoising to be tested and evaluated on the TLs100 benchmark dataset.

2.2. Implementation details

Following the common settings of previous work [17], we utilize the adamW [34] optimizer ($\beta 1 = 0.9, \beta 2 = 0.999$) to carry out 300 epochs training on the model. The initial learning rate is set to 2e-4 and gradually reduced to 1e-6 using a cosine decay strategy [35]. To enhance the diversity of the training samples, we randomly rotate the training images by 90°, 180°, and 270°. Additionally, to optimize the utilization of computational resources, we have set the batch size of input at 12 and image patch size at 128×128. Our EUformer design includes a 4-level encoder-decoder, with 2 transformers at each level. All training procedures are executed utilizing 2 NVIDIA GeForce RTX 3090 GPUs.

2.3. Experimental analyses

In this section, we primarily evaluate the proposed EUformer on low-light transmission line image, other low-light image, and real-world image denoising datasets and compare it with other excellent denoising methods both quantitatively and qualitatively. The FLOPs and parameters of the methods listed in this section are computed based on the assumption of an input image patch size of 256×256.

2.3.1. Transmission line image denoising

Since there are no existing works proposing relevant scene datasets, we utilize the benchmark dataset for denoising low-light transmission line images proposed in this paper as a test set to evaluate the denoising performance of the EUformer method.

Table 1 presents the denoising results on transmission line images. As one can see our EUformer achieves optimal in PSNR and SSIM metrics compared to other methods. In comparison to CNN-based methods, our transformer-based approach exhibits a significant performance improvement. Furthermore, compared to MIRNet, our EUformer not only shows improvements in PSNR and SSIM metrics, but also shows significantly lower FLOPs, which are only 8% of MIRNet FLOPs. In comparison to Uformer-B, our EUformer has parameters count that is only 51.28% of Uformer-B, while still achieving a PSNR gain of 0.24dB. These results indicate that the proposed EUformer method effectively removes noise in the transmission line scenario.

Tab. 1. Quantitative results of different denoising methods on TLs100

Method	FLOP, (G)	Parameter, (M)	TLs100	
			PSNR	SSIM
DNCNN [10]	37	0.6	29.60	0.848
DANet [36]	30	63.0	31.18	0.934
MPRNet [21]	573	15.7	31.61	0.938
MIRNet [18]	785	31.8	31.57	0.939
SRMNet [37]	285	37.6	30.87	0.932
DDT [43]	86	18.4	31.42	0.939
Uformer-B [17]	89.5	50.9	31.71	0.938
EUformer(ours)	63.11	22.6	31.95	0.940

Figure 2 and 3 showcase a visual comparison of our EUformer with other existing methods on the TLs100 dataset. In comparison to the other seven methods, our approach not only successfully denoising but also restores finer contour features, and the image as a whole presents a pleasing visual effect. For example, the detailed contours of the transmission line in Figure 2*a* and the angle steel and screw bolt of the transmission line tower in Figure 2*b* are more clearly visible in our EUformer compared to other methods, which indicates that our EUformer can effectively preserve the high-frequency information such as the detailed texture and edge contours of the original images. This further demonstrates that our EUformer has excellent denoising performance and can be effectively applied to low-light transmission line image denoising scenarios. Figure 3 shows that our EUformer recovers images that are overall clearer and sharper, such as the drop-out fuse in Figure 3*a*, and compared to other methods, our EUformer not only recovers fine lines and texture details, but the edge

parts of the fuse porcelain Insulator also do not introduce additional chroma artifacts. The restoration effect of the transmission line connecting fittings in Figure 3*b* and the connecting steel plates of the transmission line tower in Figure 3*c* is more close to the real image. This further demonstrates that our EUformer has excellent denoising performance and can be effectively applied to low-light transmission line image denoising scenarios.

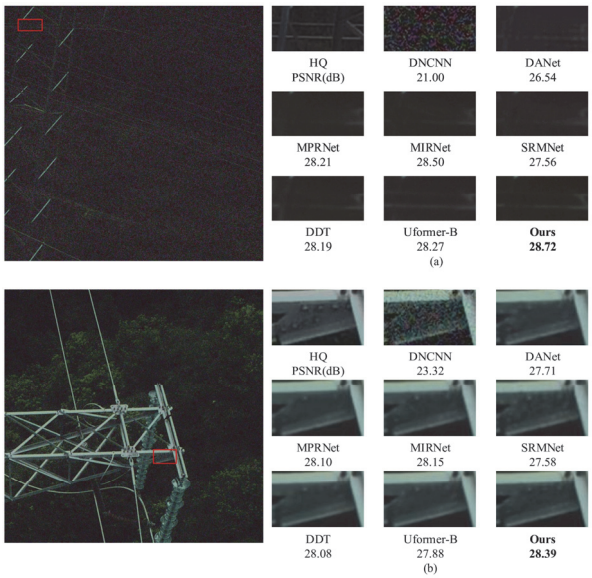


Fig. 3. Visual comparisons of image denoising methods on TLs100

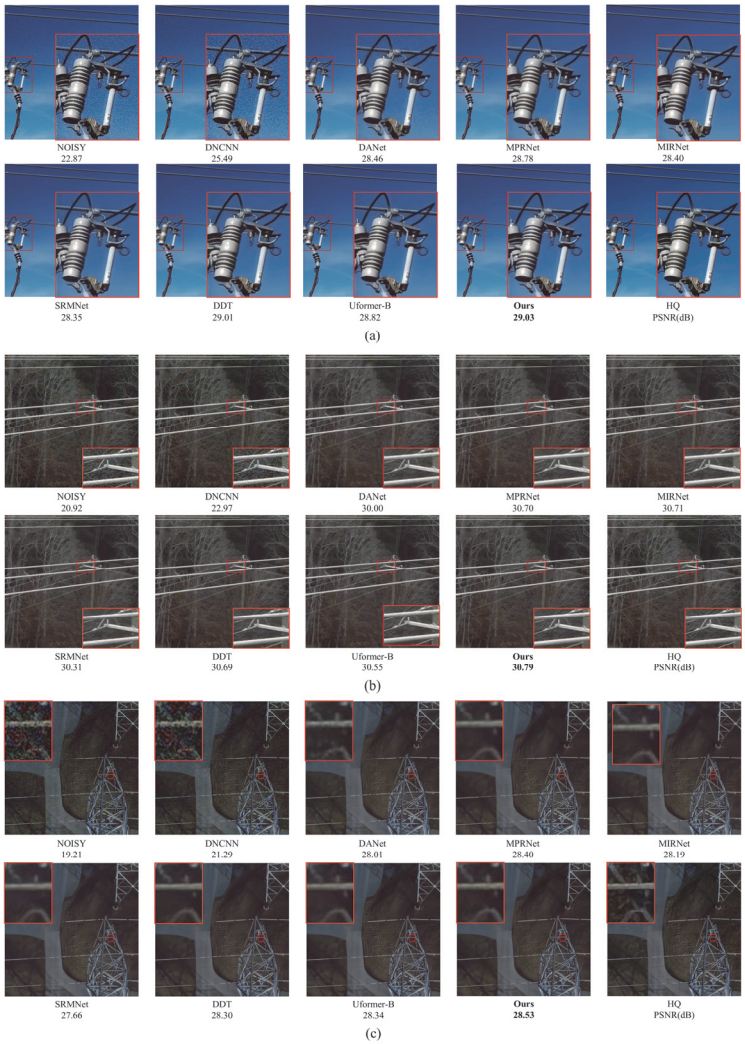


Fig. 4. Visual comparisons of image denoising methods on TLs100

2.3.2. Real-world denoising

To evaluate the denoising performance of the proposed EUformer on real-world image, as shown in Table 2, we show the real-world image denoising results of EUformer and recent representative methods on the SIDD and DND benchmark datasets. All the results are obtained from publicly available data. It can be seen that our EUformer obtains sub-optimal performance on both SIDD and DND benchmark datasets. Compared with all CNN-based methods, our EUformer exhibits a significant improvement in performance. In addition, compared to DDT, the EUformer achieves an average PSNR improvement of approximately 0.09 dB on both datasets, while its FLOPs are only 73.38% of DDT. In particular, the EUformer shows a significant 0.24 dB improvement in PSNR values on the DND dataset. While the average PSNR values of our EUformer on both datasets are the same as SwinIR, our EUformer utilizes only 63.11G FLOPs, which is only 8.14% of SwinIR. Taken together, these experimental results provide further evidence that our proposed EUformer is effective in denoising real-world images.

Tab. 2. Quantitative results of different denoising methods on two real-world datasets

Method	FLOP (G)	Parameter (M)	SIDD [32]		DND [25]	
			PSNR	SSIM	PANR	SSIM
DNCNN [10]	37	0.6	23.66	0.583	32.51	0.851
RIDNet [38]	98	1.5	38.71	0.914	39.26	0.953
IPT [15]	380	115	39.10	0.954	39.62	0.952
VDN [39]	44	7.8	39.28	0.909	39.38	0.952
DANet [36]	30	68.0	39.30	0.916	39.59	0.955
DeamNet [40]	146	2.2	39.47	0.957	39.63	0.953
CycleISP [33]	184	2.8	39.52	0.957	39.56	0.956
MPRNet [21]	573	20.4	39.71	0.958	39.80	0.954
SRMNet [37]	285	37.6	39.72	0.959	39.44	0.951
NBNet [42]	88.8	13.1	39.75	0.959	39.89	0.955
SwinIR [16]	759	11.9	39.77	0.958	40.01	0.958
Uformer-S [17]	43.86	20.6	39.77	0.959	39.96	0.955
DDT [43]	86	18.4	39.83	0.960	39.78	0.954
EUformer(ours)	63.11	22.61	39.82	0.959	40.02	0.956

2.4. Ablation study

To demonstrate the effectiveness of the individual components in our EUformer, we've conducted ablation studies on factors including Global Feature Compensator, Mixed-Gated feed-forward network and Loss Function. We modified the EUformer model by removing the Global Feature Compensator module, replacing the Mixed-Gated Feed-Forward Network with Feed-Forward Network (FFN) (expansion factor = 2.6), (FFN is shown in Fig. 5), replacing the Improved Loss Function with Charbonnier Loss, and using the modified model as the baseline. Calculate FLOPS and Parameters with an input image size of 256×256.

Effects of the Loss Function To explore the effect of the value of in the loss function on the performance of the network, we examine six different values of at four different orders of magnitude. Figure 6 shows that we chose baseline to train 100 epochs on a color Gaussian denoising task with a noise level of 50, evaluated on the urban100 benchmark datasets [41]. It is evident that reducing b from 1e-2 to 2e-4 results in a more pronounced performance improvement. Moreover, the PSNR index remains the same at 29.26 dB when b is taken as 1e-5 and 5e-4, which is the same as when b is 0. However, when b is taken as 2e-4, the PSNR index increases to 29.27 dB. Therefore, we have determined experimentally that the value for b is 2e-4.

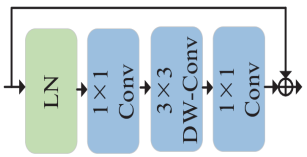


Fig. 5. The structure of feed-forward network

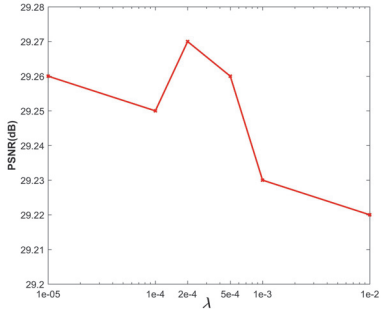


Fig. 6. Denoising results of different scale factor λ with 100 epochs

Ablation studies of the network structure To investigate the effectiveness of each module, we conducted a quantitative analysis on TLs100 benchmark

datasets and Table 7 displays the results. It is evident from the data that the addition of GFC to the baseline model leads to a PSNR gain of 0.19dB on the TLs100 benchmark datasets, which validates the effectiveness of GFC in enhancing the receptive field and improving the model's capacity to representation global information. By incorporating MG-FFN into the baseline, the PSNR increases by 0.15dB with a mere addition of 2.72G FLOPs via the cross-fusion of multiscale features. This provides evidence suggesting that MG-FFN can effectively enhance the model's performance in a practical manner.

Tab. 3. Ablation studies of main components in our model

Baseline	GFC	MG-FFN	LOSS	FLOPS	Parameters	PSNR
✓	-	-	-	50.07	21.98	31.65
✓	✓	-	-	60.38	22.14	31.84
✓	-	✓	-	52.79	22.45	31.80
✓	✓	✓	-	63.11	22.61	31.92
✓	✓	✓	✓	63.11	22.61	31.95

When the GFC and MG-FFN are incorporated into the baseline, the PSNR experiences a 0.27 dB improvement in comparison to the baseline. Furthermore, the overall performance of the network drastically improves in contrast to the addition of each module to the baseline independently. This further substantiates the soundness of the structural design of the network model. The incorporation of the GFC and the MG-FFN into the baseline model, along with the adoption of a modified Loss Function (i.e., the model is EUformer), resulted in a further improvement of the PSNR to 39.95dB. This demonstrates the efficacy of introducing the modified Loss Function as an effective approach to enhance image quality.

Conclusion

In this paper, we propose an efficient transformer network architecture to solve the task of denoising low-light transmission line images. The network is designed with two objectives, i.e., to enhance the global modelling capability and the local multi-scale information aggregation capability. To enhance the global modeling capability, we developed the Global Feature Compensator (GFC) to establish long-range pixel dependencies by utilizing a wide and dense receptive field. To enhance the local features of the aggregated image, we developed the Mixed-Gated Feed-Forward Network (MG-FFN) which leverages the cross-fusion of multi-scale convolutions for a more effective exploration of local multi-scale features. Additionally, we designed a Loss Function to further improve the visual quality of the recovered image. We build the first low-light transmission line image denoising benchmark dataset and perform extensive experiments on this dataset. In the future, we plan to expand to additional scene datasets to further evaluate the model's performance in a wider range of application scenarios.

References

[1] Nguyen VN, Jenssen R, Roverso D. Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning. Int J Electr Power Energy Syst 2018; 99: 107-120. DOI: 10.1016/j.ijepes.2017.12.016.

[2] Liba O, Murthy K, Tsai Y-T, et al. Handheld mobile photography in very low light. ACM Trans Graph 2019; 38(6): 164. DOI: 10.1145/3355089.3356508.

[3] Dabov K, Foi A, Katkovnik V, Egiazarian K. Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans Image Process 2007; 16(8): 2080-2095. DOI: 10.1109/TIP.2007.901238.

[4] Lebrun M, Buades A, Morel J-M. A nonlocal bayesian image denoising algorithm, SIAM J Imaging Sci 2013; 6(3): 1665-1688. DOI: 10.1137/120874989.

[5] Xu J, Zhang L, Zhang D. A trilateral weighted sparse coding scheme for real-world image denoising. In Book: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, eds. Computer vision – ECCV 2018. 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VIII. Cham, Switzerland: Springer Nature Switzerland AG; 2018: 20-36. DOI: 10.1007/978-3-030-01237-3_2.

[6] Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. Phys D: Nonlinear Phenom 1992; 60(1-4): 259-268. DOI: 10.1016/0167-2789(92)90242-F.

[7] Krasheninnikov VR, Kuvayskova YE, Malenova OE, Subbotin AU. Models and filtering of circular images with harmonic components of the covariance function. Procedia Comput Sci 2021; 192: 4047-4054. DOI: 10.1016/j.procs.2021.09.179.

[8] Andriyanov NA, Vasiliev KK, Dementiev VE, Belyanchikov AV. Restoration of spatially inhomogeneous images based on a doubly stochastic model. Optoelectron Instrum Data Process 2022; 58(5): 465-471. DOI: 10.3103/S8756699022050028.

[9] Zhang J, Wang FX, Zhang HL, Shi XL. A Novel CS 2G-starlet denoising method for high noise astronomical image. Opt Laser Technol 2023; 163: 109334. DOI: 10.1016/j.optlastec.2023.109334.

[10] Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a Gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Trans Image Process 2017; 26(7): 3142-3155. DOI: 10.1109/TIP.2017.2662206.

[11] Haghi P, Tan C, Guo A, Wu C, Liu D, Li A, Skjellum A, Geng T, Herbordt M. Smartfuse: Reconfigurable smart switches to accelerate fused collectives in hpc applications. Proc 38th ACM Int Conf on Supercomputing (ICS '24) 2024: 413-425. DOI: 10.1145/3650200.3656616.

[12] Lukas J, Fridrich J, Goljan M. Digital camera identification from sensor pattern noise. IEEE Trans Inf Forensics Secur 2006; 1(2): 205-214. DOI: 10.1109/TIFS.2006.873602.

[13] Yang L, Wang QF, Chi JF, et al. EAVE: Efcient product attribute value extraction via lightweight sparse-layer interaction. arXiv Preprint. 2024. Source: <https://arxiv.org/abs/2406.06839>. DOI: 10.18653/v1/2024.findings-emnlp.80.

[14] Zhang J, Wang F, Zhang H, Shi X. Compressive sensing spatially adaptive total variation method for high-noise astronomical image denoising. Vis Comput 2024; 40(2): 1215-1227. DOI: 10.1007/s00371-023-02842-w.

[15] Chen H, Wang Y, Guo T, Xu C, Deng Y, Liu Z. Pre-trained image processing transformer. 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR) 2021: 12299-12310. DOI: 10.1109/CVPR46437.2021.01212.

[16] Liang J, Cao J, Sun G, Zhang K, Gool LV, Timofte R. SwinIR: Image restoration using swin transformer. 2021 IEEE/CVF Int Conf on Computer Vision Workshops (ICCVW) 2021: 1833-1844. DOI: 10.1109/ICCVW54120.2021.00210.

[17] Wang Z, Cun X, Bao J, Zhou W, Liu J, Li H. Uformer: A general U-shaped transformer for image restoration. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR) 2022: 17662-17672. DOI: 10.1109/CVPR52688.2022.01716.

[18] Zamir SW, Arora A, Khan S, Hayat M, Khan FS, Yang M-H, Shao L. Learning enriched features for real image restoration and enhancement. In Book: Vedaldi A, Bischof H, Brox T, Frahm J-M, eds. Computer vision – ECCV 2020. 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV. Cham, Switzerland: Springer Nature Switzerland AG; 2020: 492-511. DOI: 10.1007/978-3-030-58595-2_30.

[19] Shi W, Caballero J, Huszár F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. 2016 IEEE Conf on Computer Vision and Pattern Recognition (CVPR) 2016: 1874-1883. DOI: 10.1109/CVPR.2016.207.

[20] Chen X, Wang X, Zhou J, Qiao Y. Activating more pixels in image super-resolution transformer. 2023 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR) 2023: 22367-22377. DOI: 10.1109/CVPR52729.2023.02142.

[21] Zamir SW, Arora A, Khan S, Hayat M, Khan FS, Yang M-H. Multi-stage progressive image restoration. 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR) 2021: 14821-14831. DOI: 10.1109/CVPR46437.2021.01458.

[22] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. 31st Conference on Neural Information Processing Systems (NIPS 2017) 2017: 6000-6010.

[23] Chen X, Li H, Li M, Pan J. Learning a sparse transformer network for effective image deraining. 2023 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR) 2023: 5896-5905. DOI: 10.1109/CVPR52729.2023.00571.

[24] Wang Z, Fu Y, Liu J, Zhang Y. LG-BPN: Local and global blind-patch network for self-supervised real-world denoising. 2023 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR) 2023: 18156-18165. DOI: 10.1109/CVPR52729.2023.01741.

[25] Plotz T, Roth S. Benchmarking denoising algorithms with real photographs, 2017 IEEE Conf on Computer Vision and Pattern Recognition (CVPR) 2017: 1586-1595. DOI: 10.1109/CVPR.2017.294.

[26] Ding X, Zhang X, Han J, Ding G. Scaling up your kernels to 31×31: Revisiting large kernel design in CNNs. 2022 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR) 2022: 11963-11975. DOI: 10.1109/CVPR52688.2022.01166.

[27] Chen L, Chu X, Zhang X, Sun J. Simple baselines for image restoration. In Book: Avidan S, Brostow G, Cissé M, Farinella GM, Hassner T, eds. Computer vision – ECCV 2022. 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII. Cham, Switzerland: Springer Nature Switzerland AG; 2022: 17-33. DOI: 10.1007/978-3-031-20071-7_2.

[28] Abdelfattah R, Wang X, Wang S. TTPLA: An aerial-image dataset for detection and segmentation of transmission towers and power lines. In Book: Ishikawa H, Liu C-L, Pajdla T, Shi J, eds. Computer vision – ACCV 2020. 15th Asian Conference on Computer Vision, Kyoto, Japan, November 30 – December 4, 2020, Revised Selected Papers, Part VI. 2020: 601-618. DOI: 10.1007/978-3-030-69544-6_36.

[29] Vieira-e-Silva ALB, de Castro Felix H, de Menezes Chaves T, Simões FPM, Teichrieb V, dos Santos MM. STN PLAD: A dataset for multi-size power line assets detection in high-resolution UAV images. 2021 34th SIBGRAPI Conf on Graphics, Patterns and Images (SIBGRAPI) 2021: 215-222. DOI: 10.1109/SIBGRAPI54419.2021.00037.

[30] Allard WK. Total variation regularization for image denoising, I. Geometric theory. SIAM J Math Anal 2008; 39(4): 1150-1190. DOI: 10.1137/060662617.

[31] Charbonnier P, Blanc-Feraud L, Aubert G, Barlaud M. Two deterministic half-quadratic regularization algorithms for computed imaging. Proc 1st Int Conf on Image Processing 1994; 2: 168-172. DOI: 10.1109/ICIP.1994.413553.

[32] Abdelhamed A, Lin S, Brown MS. A high-quality denoising dataset for smartphone cameras. 2018 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR) 2018: 1692-1700. DOI: 10.1109/CVPR.2018.00182.

[33] Zamir SW, Arora A, Khan S, Hayat M, Khan FS, Yang M-H. CycleISP: Real image restoration via improved data synthesis. 2020 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR) 2020: 2696-2705. DOI: 10.1109/CVPR42600.2020.00277.

[34] Loshchilov I, Hutter F. Decoupled weight decay regularization. arXiv Preprint. 2017. Source: <https://arxiv.org/abs/1711.05101>. DOI: 10.48550/arXiv.1711.05101.

[35] Loshchilov I, Hutter F. Sgdr: Stochastic gradient descent with warm restarts. arXiv Preprint. 2016. Source: <https://arxiv.org/abs/1608.03983>. DOI: 10.48550/arXiv.1608.03983.

[36] Yue Z, Zhao Q, Zhang L, Meng D. Dual adversarial network: Toward real-world noise removal and noise generation. In Book: Vedaldi A, Bischof H, Brox T, Frahm J-M, eds. Computer vision – ECCV 2020. 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part X. Cham, Switzerland: Springer Nature Switzerland AG; 2020: 41-58. DOI: 10.1007/978-3-030-58607-2_3.

[37] Fan C-M, Liu T-J, Liu K-H, Chiu C-H. Selective residual M-net for real image denoising. 2022 30th European Signal Processing Conf (EUSIPCO) 2022: 469-473. DOI: 10.23919/EUSIPCO55093.2022.9909521.

[38] Anwar S, Barnes N. Real image denoising with feature attention. 2019 IEEE/CVF International Conference on Computer Vision (ICCV) 2019: 3155-3164. DOI: 10.1109/ICCV.2019.00325.

[39] Yue Z, Yong H, Zhao Q, Zhang L, Meng D. Variational denoising network: Toward blind noise modeling and removal. Proc 33rd Int Conf on Neural Information Processing Systems 2019: 1690-1701.

[40] Ren C, He X, Wang C, Zhao Z. Adaptive consistency prior based deep network for image denoising. 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR) 2021: 8596-8606. DOI: 10.1109/CVPR46437.2021.00849.
[41] Huang J-B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars. 2015 IEEE Conf on Computer Vision and Pattern Recognition (CVPR) 2015: 5197-5206. DOI: 10.1109/CVPR.2015.7299156.
[42] Cheng S, Wang Y, Huang H, Liu D, Fan H, Liu S. NBNet: Noise basis learning for image denoising with subspace projection. 2021 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR) 2021: 4896-4906. DOI: 10.1109/CVPR46437.2021.00486.
[43] Liu K, Du X, Liu S, Zheng Y. DDT: Dual-branch deformable transformer for image denoising. 2023 IEEE Int Conf on Multimedia and Expo (ICME) 2023: 2765-2770. DOI: 10.1109/ICME55011.2023.00470.

Authors' information

Jie Zhang (b. 1986) graduated from Harbin Institute of Technology in 2018, majoring in Control Science and Engineering. Currently he works as the associate professor at the College of Electrical and Information Engineering, Zhengzhou University of Light Industry, China. Research interests are image processing, deep learning.
E-mail: *zhangjie1234@zzuli.edu.cn*

Wenxiao Huang (b. 2001) received the B.E. degree in Robotics Engineering from the Henan Institute of Technology, Xinxiang, China, in 2022. She is currently pursuing the M.S. degree with the Zhengzhou University of Light Industry, Zhengzhou, China. Her research interests include computer vision, image denoising, image processing and compressed sensing. E-mail: *332201050058@email.zzuli.edu.cn*

Miaoxin Lu (b. 2000) received the B.E. degree in Electrical Engineering and Automation from the Zhengzhou University of Aeronautics, Zhengzhou, China, in 2022. He is currently pursuing the M.S. degree in Electrical Engineering with Zhengzhou University of Light Industry, Zhengzhou, China. His main research interests include compressive sensing, computer vision, and image processing techniques. E-mail: *332201060067@email.zzuli.edu.cn*

Linwei Li (b. 1986) graduated from Beijing Institute of Technology in 2019, majoring in Control Science and Engineering. Currently he works as the lecturer at the College of Electrical and Information Engineering, Zhengzhou University of Light Industry, China. Research interests are image processing, motor servo control.
E-mail: *2019028@zzuli.edu.cn*

Xin Wang (b. 1989) graduated from Institute of Automation of Chinese Academy of Sciences in 2023, majoring in Control Science and Engineering. Currently he works as the lecturer at the College of Electrical and Information Engineering, Zhengzhou University of Light Industry, China. Research interests are image processing, nonlinear control. E-mail: *2023043@zzuli.edu.cn*

Yong Peng Shen (b. 1985) graduated from Hunan University in 2015, majoring in Control Science and Engineering. Currently he works as the associate professor at the College of Electrical and Information Engineering, Zhengzhou University of Light Industry, China. Research interests are image processing, intelligent electric vehicles and energy storage system management. Serve as the corresponding author of this article. E-mail: *shenyongpeng@zzuli.edu.cn*

Yanfegn Wang (b. 1973) works as the professor at the College of Electrical and Information Engineering, Zhengzhou University of Light Industry, China. Research interests are image processing, nonlinear control. Serve as the corresponding author of this article. E-mail: *quhn1234@163.com*

Received October 24, 2024. The final version – December 06, 2024.