World Journal of Advanced Engineering Technology and Sciences
eISSN: 2582-8266
Cross Ref DOI: 10.30574/wjaets
Journal homepage: https://wjaets.com/

(RESEARCH ARTICLE)

Check for updates

Dual Stage Bionic Spatial Convolutional Restoration Model for Image De-noising and De-mosaicking

Saranya J [1,*] and E. Sree Devi [2]

[1] *Master of Engineering in Communication Systems, Department of Electronics and Communication Engineering, Rohini College of Engineering and Technology, Palkulam, Variyoor, Kanyakumari District, Tamil Nadu, 629401 India.*
[2] *Professor, Department of Electronics and Communication Engineering, Rohini College of Engineering and Technology, Palkulam, Variyoor, Kanyakumari District, Tamil Nadu, 629401 India.*

World Journal of Advanced Engineering Technology and Sciences, 2024, 12(02), 889–902

Publication history: Received on 04 July 2024; revised on 19 August 2024; accepted on 21 August 2024

Article DOI: https://doi.org/10.30574/wjaets.2024.12.2.0347

Abstract

The first two important steps in the pipeline for processing picture signals are de-noising and de-mosaicking. The joint solution of the highly uncertain inverse problem of de-noising and de-mosaicking has garnered increased attention in research today. It is difficult to restore high-quality images from raw data in low light because of a variety of disturbances brought on by a low photon count and a complex image signal processing scheme. Even if some restoration and improvement techniques have been used, they might not work in harsh situations, including raw data imaging with brief exposure. Therefore, this research focuses on developing a de-mosaicking and de-noising model with effective end to end manner outcomes. Initially, the pre-processing is conducted using Gaussian filtering to eliminate artifacts from the input image, thereby enhancing the image quality. Then, the proposed method incorporates an Enhanced Spatial Convolutional Residual Net (EnConvResNet) for image de-mosaicking and an Adaptive U-net restoration model for image de-noising. An enhanced gazelle optimization (EnGa) algorithm is used to fine-tune the hyper-parameters of the model in order to maximize its performance and improve its generalization capacity. The proposed method accomplished peak signal to noise ratio (PSNR) and structural similarity index measure (SSIM) of 46.65 and 98.89, respectively.

Keywords: Enhanced Gazelle optimization; De-mosaicking; De-noising; Gaussian filtering; Adaptive U-Net; Enhanced Spatial Convolutional Residual Net

1. Introduction

De-mosaicking aims to combine four spatially sparse color channels to produce a full-color image. Actually, color information per pixel can only be captured by digital cameras with monochrome sensors; most of these cameras achieve this by utilizing color filter arrays (CFA), like the Bayer pattern [1]. Two of the four pixels are measured to be green, red, and blue. De-mosaicking is the process of creating a whole color image from the output of partial color samples [2]. In general image processing, the pipeline is assumed to contain the raw data first de-noised and then de-mosaicked. Because de-noising algorithms are usually built on statistical priors, once the raw data is removed, these priors could be seriously disrupted. Additionally, the majority of well-performing common de-mosaicking techniques are created using the essential noise-free condition as their foundation [3].

The main problem with the segmentation de-noising and de-mosaicking processes is that they can conflict with one another. De-mosaicking increases the difficulty of the noise removal process by substituting the interpolation process for the noise distribution. The color models in the raw photos are altered if de-noising is done before, making full color

* Corresponding author: Saranya J.; Email: josesaranya123@gmail.com

recovery from de-mosaicking more challenging [4]. Nevertheless, there has been much less focus on combining the three problems of image de-noising, de-mosaicking, and SR. Despite being an end-to-end network, TENet has a set order of operating activities [5]. Deep learning techniques have recently demonstrated remarkable effectiveness in image identification problems. It is also popular for low-level visual tasks, such as de-mosaicking images [6]. Learning-based approaches contribute significantly to the progress of image reconstruction tasks. Moreover, deep learning has shown the advantages of integrating low-level tasks like de-noising and de-mosaicking [7]. For the purpose of tackling traditional inverse imaging problems, including blurring, super resolution, de-noising, and pixel super-resolution, deep learning presents some intriguing possibilities [8][9].

Deep learning techniques use Convolutional Neural Networks (CNNs) to automatically learn what they desire. The majority of methods employ force-learning mapping networks between noisy images, such as damaged or mosaic images. A four-channel RGGB image and a pure RGB image are useful for intra- and inter-channel communication [10]. However, only a tiny portion of low-visibility research actually works with RAW data, which is a larger scale and more widely available RGB photos for public use [11]. However, the quantity of photons acquired by the camera sensor determines the number of pixels in the RAW field, and the noise in these pixels might be spatial [12]. Additionally, the dataset includes the Bayer and Fuji X Trans, two well-known CFA patterns, allowing for the development and assessment of strategies capable of handling various CFA patterns [13]. It can employ both a big RGB dataset (e.g., 100,000 samples) and a small RAW dataset (e.g., 7,000 pictures) to achieve our goal [14]. Finally, the results of images processed by various ISP pipelines were examined to compare the generalization performance of our model to alternative methods for color photos [15].

The suggested work offers a unique method for improving the quality of images. This model combines two stages to improve both image de-noising and demosaicing processes.

In the first stage, the bionic spatial convolutional restoration is applied to denoise the image. This stage takes inspiration from biological processes in the human visual system, which are highly efficient at filtering out noise. The model uses spatial convolutions to mimic these biological processes and remove noise from the image.

In the second stage, the model focuses on demosaicing, which is the process of converting a color filter array image (CFA) into a full-color image. This is particularly important for digital cameras, which use a CFA to diminish the number of sensors needed. The dual-stage model improves the demosaicing process by incorporating the de-noised image from the first stage, resulting in a higher quality, full-color image.

The organization of the research is as follows: Section 2 details the related works, and Section 3 details the proposed methodology. Section 4 provides an explanation of the findings and discussion, and Section 5 concludes the research.

2. Literature survey

Mykola Ponomarenko et al. [16] introduced CNN for joint images. This system performs five image enhancements, namely de-mosaicking, de-blurring, de-noising, super-resolution, and clarity enhancement. The DRUNet de-noising network is first used to combine the U-Net and ResNet. Transposed convolution and residual blocks are then used to interpolate CFA images. After this procedure was finished, scaling was applied four times to increase and decrease the clarity of the image. Finally, five net CNN improved the processed image clarity and quality. This section has observed that the quality of the images in the training set is low.

The use of CNNs for joint de-noising and de-mosaicking of real-world burst images has been suggested by Shi Guo et al. [17]. The Green Channel Prior network was first implemented in this system in order to produce a good image structure. Afterward, the intra-frame (IntraF), inter-frame (InterF), and merge modules were used to conduct the reconstruction. Finally, the real world burst images are cleaned using a green channel prior to providing high quality and a good sampling rate. However, this section noticed that there is an issue with de-mosaicking.

A machine learning technique for jointly de-noising and de-mosaicking images was presented by Yu Guo et al. [18]. In this system, a noisy CFA picture was first created. Using the gradient based threshold free (GBTF) method, noise-free de-mosaicking is then performed in order to rebuild the image. De-noising is then carried out to eliminate the noise from the image. After pre-processing is complete, the image is reconstructed using the CNN method, yielding a distortion-free result. However, this part observed that artifacts are the reason for low performance.

An Adaptive Convolutional Dictionary Learning Network for Blind De-noising and de-mosaicking was presented by Nikola Janjusevicc et al. [19]. In this system, a natural image was initially taken to de-noise using a convolutional

dictionary learning network, and after that, thresholding was performed. Once this process was completed, blind de-noising was done using CDL net, which provides high performance. However, noise adaptive generalization can be used to improve the sensing.

Hanlin Tan et al. [20] introduced Two-Stage CNN Model for de-mosaicking and de-noising the burst bayer images. In this system, datasets are taken from Kotak. Initially, de-noising was performed in two stages, namely single frame and multi-frame, executed using the DRDD network. One frame is used for first level de-noising in order to map the residual block directly; multiple frames are then used for end-to-end de-noising of the burst of images. Convolution and Rectified Linear Unit (ReLU) activation were used in multi-frame. After that, the input goes via a residual block to produce an intermediate result, and CNN then squeezes and excites the linear RGB image. Nonetheless, this section noted that the image's efficacy was low.

3. Proposed Methodology

The proposed research undergoes two main stages, namely de-mosaicking and de-noising. The initial step is to take the raw image and create the mosaic. A correlated residual noise will be present in the output of a de-mosaicked process that takes a noisy input. It also takes learning to get rid of this noise. Various noises can be introduced during the image capture process, such as shot noise or salt-and-pepper noise. Therefore, in the second stage, for the removal of noise, the de-noising process is performed. Gaussian filtering is used for pre-processing in order to eliminate artifacts from the input image before de-noising and de-mosaicking. The network's workload can be reduced by enhancing the input to the network. The work is incorporated with an Enhanced spatial convolutional residual net for image de-mosaicking and an Adaptive U-net restoration model for image de-noising. The restoration model is adapted with symmetric connections to maintain the recovery information from the raw image. The hyper-parameters of the network model are fine-tuned with an enhanced gazelle optimization model to upsurge the generalization ability of the proposed model. The workflow is presented in Figure 1.

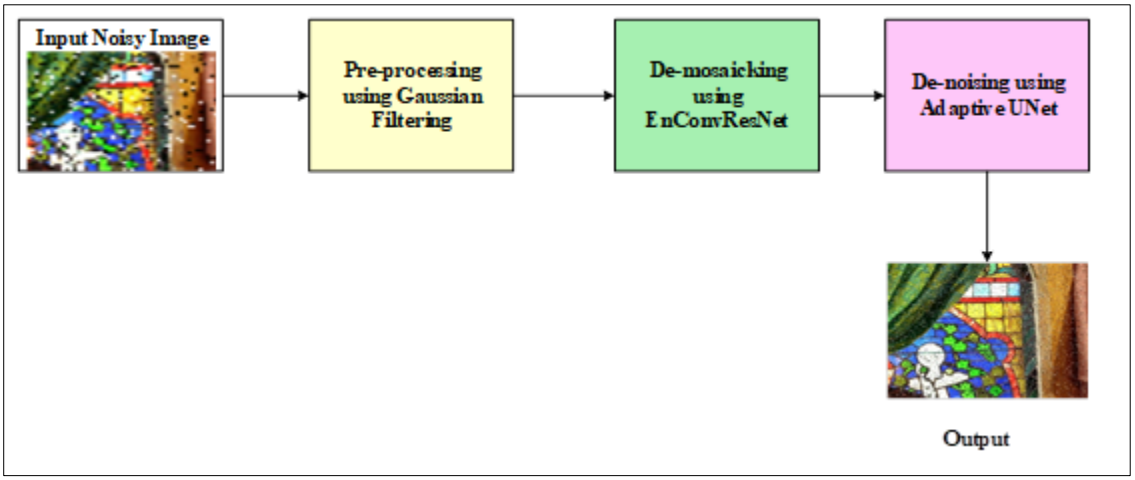


Figure 1 Proposed De-mosaicking and de-noising technique

3.1. Gaussian Filtering

Gaussian filtering is a common method used in image processing to get rid of noise and artifacts. The image and the Gaussian kernel are convolved in the process of Gaussian filtering. Convolution involves sliding the kernel across the image and computing the weighted sum of the pixel values it covers at each location. The resulting image has pixel values that are weighted averages of their surrounding neighbors, with pixels closer to the kernel's center being of higher weight. The formulation for representing the Gaussian filter is expressed as:

$$A(e, f) = \frac{1}{2\pi\sigma^2} e^{-\frac{e^2+f^2}{2\sigma^2}} \quad (1)$$

where, $A(e, f)$ is the value of the Gaussian kernel at position (e, f) and σ is the standard deviation of the Gaussian distribution, which controls the spread or blurring effect of the kernel, e is the base of the natural logarithm. Because of the Gaussian filter's blurring effect, the final image is a smoothed version of the original with less noise and artifacts.

3.2. EnConvResNet based de-mosaicking

The improved spatial convolutional residual net (EnConvResNet) is utilized for image de-mosaicking. The proposed EnConvResNet is designed by integrating the Convolutional layer, ResNet-152, and the attention module to enhance the de-mosaicking performance. Figure 2 shows the proposed EnConvResNet structure.

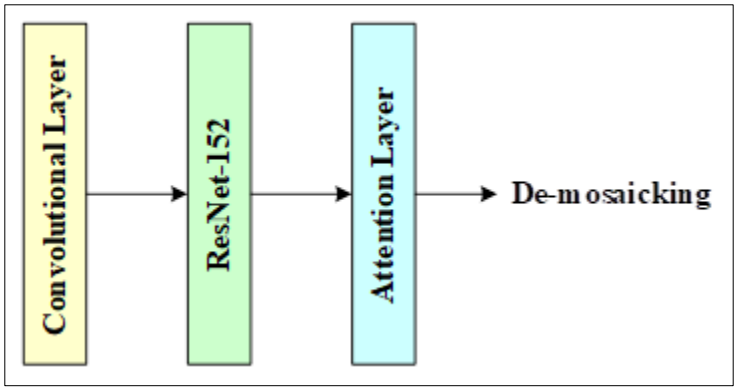


Figure 2 Structure of proposed EnConvResNet

A detailed explanation concerning the EnConvResNet is given below.

3.2.1. Convolutional Layer

The de-mosaicking process begins with a convolutional layer. Feature extraction from the input image is the responsibility of this layer. Important details, such as edges, textures, and patterns, can be extracted from the input image with the help of the convolutional layer. Each neuron in a convolutional layer is associated with a tiny area of the input image due to the layer's use of local receptive fields. The convolutional layer employs these local receptive fields to scan the entire input image in order to learn about the spatial relationships between neighboring pixels. This feature is essential to de-mosaicking because it enables the network to identify correlations between neighboring color channels and precisely fill in color information that is missing. The formulation for performing the convolution operation is given as follows:

$$R(u,w) = (S * M)(u,w) = \sum_{k}\sum_{l} S(u,w) * M(u-k, w-l) \qquad (2)$$

where, the kernel filter is denoted as M, input data is denoted as S, and the convolution operation is specified as $*$. The feature mapped by the convolutional layer is denoted as $R(u,w)$ and the coordinates are indicated as (u,w) and (w,l), respectively. The outcome of the convolutional layer is fed into the ResNet-152 to acquire the best features and make the de-mosaicking process more efficient.

3.2.2. ResNet-152

ResNet-152 is a specific type of deep neural network architecture known for its depth and residual learning framework. In residual learning, the network learns residual functions rather than directly fitting the input to the output by skipping one or more layers. ResNet-152 facilitates training deeper networks more effectively, enabling better feature extraction and representation learning. The proposed EnConvResNet improves the de-mosaicking process by integrating ResNet-152 to improve the feature extraction capabilities. The structure of ResNet-152 is portrayed in Figure 3.

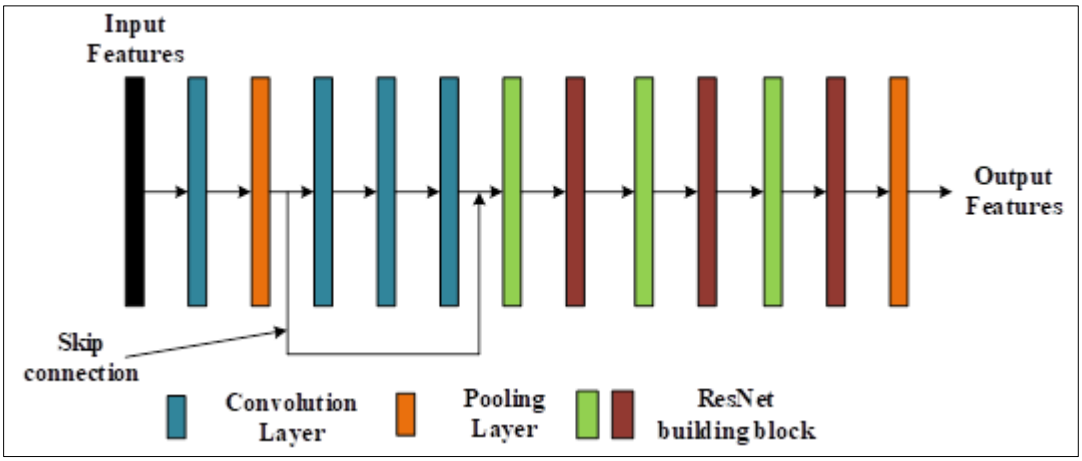


Figure 3 Structure ResNet-152

The ResNet-152 utilizes the skip connection that assists in eliminating the vanishing gradient issue. Let the input fed into the skip connection module be signified as l, the outcome derived by the convolutional network is $P(l)+l$; but the outcome derived through the skip connection is $P(l)$. Hence, using the ResNet-152, the enhanced feature extraction is accomplished with minimal computation burden. The outcome of the ResNet-152 is fed into the Attention module of the proposed EnConvResNet model.

3.2.3. Attention Layer

The proposed EnConvResNet uses attention techniques to suppress extraneous information and concentrate on pertinent portions of the input data. The attention module in EnConvResNet probably enhances the quality of the rebuilt image and the de-mosaicking performance by dynamically adjusting the weights of respective input image segments based on their saliency or significance. The formulation for the attention mechanism is described as:

$$E_h^{'} = soft\max\left(F_E H_h + G_h\right) \tag{3}$$

$$N_h^{''} = E_h^{'} \Theta H_h = \left(E_{1,h}H_{1,h}, E_{2,h}H_{2,h}, \ldots\ldots E_{D,h}H_{D,h}\right) \tag{4}$$

Here, $N_h^{''}$ refers to the outcome of the attention layer, F_E and G_h signifies the weight and bias and $E_h^{'} = \left(E_{1,h}, E_{2,h}, \ldots\ldots E_{D,h}\right)$ signifies the weights. From the attention module outcome, the de-mosaicked image is acquired.

3.3. Adaptive UNet based de-noising

The adaptive UNet is utilized for the de-noising, wherein the loss function of the U-Net is employed using the enhanced Gazelle optimization (EnGa) algorithm.

3.3.1. Structure of UNet:

UNet can learn the precise noise patterns in the training data, in contrast to typical filtering algorithms that rely on predefined assumptions about noise characteristics. This enables it to effectively remove noise while maintaining the underlying image features. UNet architecture is flexible and can accommodate input images of varying sizes without requiring modifications. The adaptability of the UNet makes it suitable for de-noising tasks across different resolutions and aspect ratios. Thus, UNet is considered for the proposed de-noising task. UNet's architecture consists of contracting (encoder) and expanding (decoder) paths, enabling it to extract low-level detailed information from the noisy image, such as textures, and high-level contextual information, such as edges and structures. Furthermore, these features are immediately injected into the up-sampling process through the skip links between the encoder and decoder. The up-sampling process ensures that important details are preserved during the image reconstruction phase, leading to a more accurate de-noised image. The architecture of the UNet is portrayed in Figure 4.

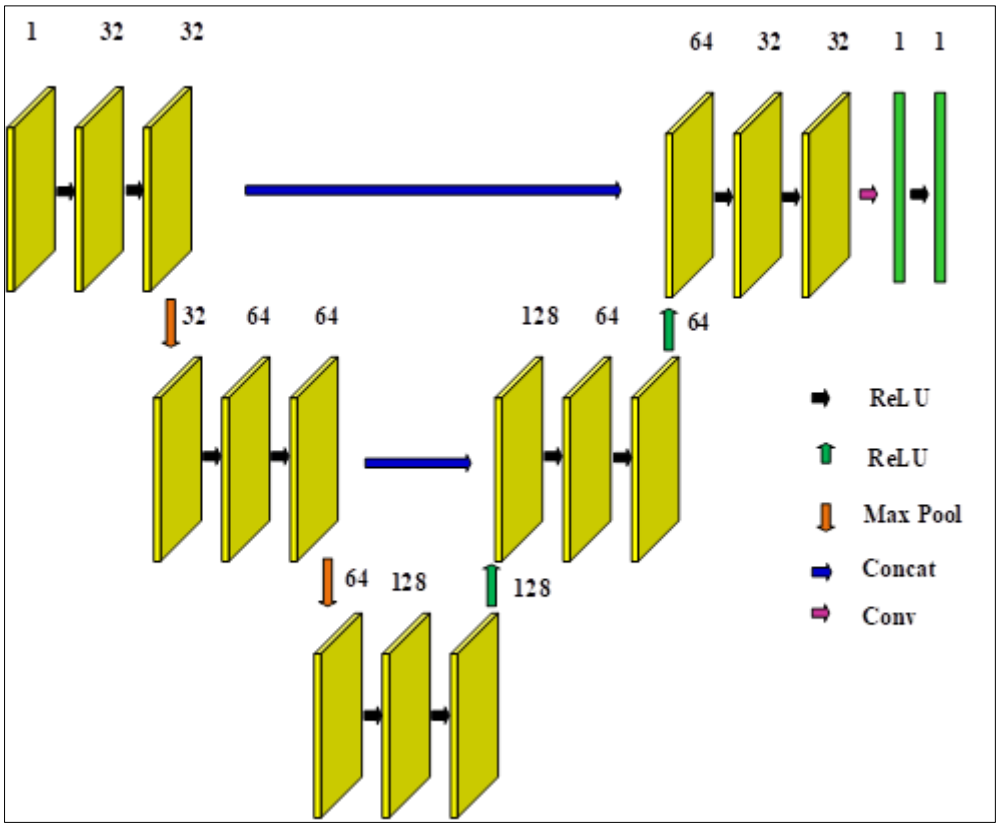


Figure 4 Architecture of the UNet

Contracting Path

The contracting path is tasked with the reduction of the spatial dimensions of the input image and the capture of low-level features. The process involves the iterative implementation of convolutional layers, which are subsequently downsampled via max pooling. Convolutional layers extract features, including edges, textures, and patterns, by capturing local information through the utilization of small receptive fields. The network is supplemented with a non-linear activation function following each convolutional layer. The ReLU, which facilitates the acquisition of complex patterns through the introduction of non-linearity, is a frequent example. To reduce the spatial dimensions of the feature maps and expand the receptive field of subsequent layers, downsampling procedures utilizing maximum pooling reduce the resolution. To extract high-level information from an input image, the dimensionality of the image is progressively decreased along the contracting path.

Expanding Path

The expanding path is responsible for reconstructing the de-noised image from the low-dimensional feature representation obtained from the contracting path. Concatenation with feature maps from the relevant layers in the contracting path and convolutional layers comes after upsampling procedures using transposed convolutions. The resolution lost during downsampling in the contracting path is restored by the upsampling processes, which expand the spatial dimensions of the feature maps. By adding high-resolution features, concatenation with feature maps from the contracted path aids in the preservation of fine details and spatial information. Convolutional layers in the expanding path use small receptive fields to refine the feature maps and generate a detailed reconstruction of the de-noised image. A non-linear activation function called ReLU usually follows each convolutional layer to add non-linearity and improve the de-noising ability. The expanding path gradually refines the low-dimensional feature representation into a high-resolution reconstruction of the de-noised image, leveraging both global context and local details.

Output Layer

The output layer applies a sigmoid activation function with a single filter to a convolutional layer to produce the de-noised image. The de-noised outcome is defined as:

$$Outcome = Sig(Conv(F_{in})) \quad (5)$$

where, the de-noised outcome is denoted as $Outcome$, the sigmoid activation function is denoted as Sig, the input feature to the output layer is signified as F_{in}.

Therefore, the UNet architecture for image de-noising is comprised of numerous modules, with the contracting path being responsible for capturing low-level features and the expanding path being tasked with reconstructing the de-noised image while maintaining critical details. Here, the loss function of the UNet is optimized using the EnGa algorithm.

3.3.2. Enhanced Gazelle Optimization Algorithm:

The Enhanced Gazelle optimization (EnGa) algorithm is created by incorporating self-adaptiveness into the basic Gazelle optimization algorithm to improve unpredictability. Here, the inclusion of self adaptiveness assists in enhancing the randomness in the algorithm, which leads to enhanced exploration for solving the local optimal solution.

Localization of the candidates in the search boundary is the initial step of the algorithm, and the population of the candidate is described as follows:

$$W = \begin{bmatrix} w_{1,1} & w_{1,2} & \cdots w_{1,b-1} & w_{1,n} \\ w_{2,1} & w_{2,2} & \cdots w_{2,b-1} & w_{2,n} \\ \vdots & \vdots & w_{d,e} \vdots & \vdots \\ w_{b,1} & w_{b,2} & \cdots w_{b,n-1} & w_{b,n} \end{bmatrix} \quad (6)$$

The candidates in the search boundary are signified as P, the dimension of the solution is designated as b and n refers to the n^{th} candidate. $w_{d,e}$ refers to the solution estimated by the d^{th} search agent. The solution evaluated by the search agent is described as:

$$w_{b,n} = M \times (J_n - K_n) + D_n \quad (7)$$

Here, the search boundaries of the EnGa algorithm are defined as J_n and K_n that refers to its lower and upper limits, and the arbitrary factor is signified as M. After locating the candidates, the feasibility is evaluated through the fitness and is defined as:

$$TR = \frac{1}{Ts} \sum_{i=1}^{Ts} \left(Pi - \hat{P}i \right)^2 \quad (8)$$

Here, required and estimated solutions by the candidate are signified as Pi and $\hat{P}i$ respectively, and the total solutions are defined as Ts. The solutions acquired by the candidates close to the target occupy the top rank in the Ranking list and are expressed as:

$$TR = \begin{bmatrix} w'_{1,1} & w'_{1,2} & \cdots w'_{1,n-1} & w'_{1,n} \\ w'_{2,1} & w'_{2,2} & \cdots w'_{2,n-1} & w'_{2,n} \\ \vdots & \vdots & w'_{d,e} \vdots & \vdots \\ w'_{b,1} & w'_{b,2} & \cdots w'_{b,n-1} & w'_{b,n} \end{bmatrix} \quad (9)$$

where, w' refers to the ranked candidates based on the solutions acquired in the present iteration. The Brownian movement and levy flight are two different tactics used by the candidates to capture the solution.

Exploitation

The candidates exhibit erratic movements to evade predators, which resemble the Brownian motion observed in physics. Thus, gazelles move randomly in different directions to explore the solution space. Because of its randomness, the algorithm can successfully traverse the whole search space and escape from local optima. It is described as:

$$\overrightarrow{W_{p+1}} = \overrightarrow{W_p} + h.\vec{D} * \vec{D}_S * \left(\overrightarrow{TR_p} - \vec{D}_S * W_p\right) \quad (10)$$

Here, the solution arrived at by the candidates through the exploration phase is denoted as $\overrightarrow{W_{p+1}}$, and the resolution reached during the preceding cycle is denoted as W_p. The candidates move towards the target with the speed of h to capture the target. The Brownian movement based factor indicated arbitrarily is signified as $\vec{D}_S$, and random distribution is denoted as $\vec{D}$ with the range [0, 1]. Self-adaptation is included and defined as a means of augmenting the randomness of the optimization:

$$SA = \overrightarrow{W_{p+1}} + Z\left(\overrightarrow{TR_p} - \overrightarrow{W_p}\right) \quad (11)$$

Here, the self adaptiveness is defined as SA, the control factor is signified as Z, the top ranked solution is defined as $\overrightarrow{TR_p}$, and the solution accomplished by the candidate in the previous iteration is defined as W_p. The solution accomplished by the proposed EnGa is expressed as:

$$\left(\overrightarrow{W_{p+1}}\right)_{EnGa} = SA * \left(\overrightarrow{W_{p+1}}\right)_{Gazelle} \quad (12)$$

Here, through the incorporation of self adaptiveness, randomization is enhanced, and hence, local optimal trapping is eliminated.

Exploration

The identification of the target is captured in this phase, and the solution accomplished is described as:

$$\overrightarrow{W_{p+1}} = \overrightarrow{W_p} + E.\alpha.\vec{D} * \vec{D}_R * \left(\overrightarrow{TR_p} - \vec{D}_R * W_p\right) \quad (13)$$

Here, the candidate with the highest motion capability is referred as E, and the levy flight is denoted as $\vec{D}_R$. The sudden change in direction is denoted as α, and the movement of the predator is described as:

$$\overrightarrow{W_{p+1}} = \overrightarrow{W_p} + E.\alpha.X * \vec{D}_S * \left(\overrightarrow{TR_p} - \vec{D}_R * W_p\right) \quad (14)$$

The parameter utilized for controlling the predator motion is signified as X and is described as:

$$X = \left(1 - \frac{p}{p_{\max}}\right)^{\left(2\frac{p}{p_{\max}}\right)} \quad (15)$$

The solution acquired by the candidate based on randomness is described as:

$$\overrightarrow{W_{p+1}} = \begin{cases} \overrightarrow{W_p} + X\left[\vec{K} + \vec{D} * \left(\vec{J} - \vec{K}\right) * \vec{W}\right] \\ \overrightarrow{W_p} + \left[Y(1-y) + y\right]\left(\overrightarrow{W_{y1}} - \overrightarrow{W_{y2}}\right) \end{cases} \quad (16)$$

Here, the binary factor is signified as $\vec{W} = \begin{cases} 0, & if\ y < 0.2 \\ 1, & Otherwise \end{cases}$ and arbitrary values are indicated as y_1 and y_2.

Finally, for all the candidate solutions, the feasibility is evaluated based on fitness. Algorithm 1 presents the proposed EnGa's pseudo-code.

Algorithm 1 Pseudo-code for EnGa Algorithm

Pseudo-code for EnGa Algorithm	
1	Initialize the parameters: Iteration, population, and boundaries
2	{
3	Locate the search agents
4	Estimate the ranking based on $TR = \frac{1}{Ts}\sum_{i=1}^{Ts}\left(Pi - \hat{P}i\right)^2$
5	While
6	{
7	P < Pmax
8	Evaluate the solution using $\left(\overrightarrow{W_{p+1}}\right)_{EnGa} = SA * \left(\overrightarrow{W_{p+1}}\right)_{Gazelle}$
9	Evaluate the solution using $\overrightarrow{W_{p+1}} = \begin{cases} \overrightarrow{W_p} + X\left[\vec{K} + \vec{D} * \left(\vec{J} - \vec{K}\right) * \vec{W}\right] \\ \overrightarrow{W_p} + \left[Y(1-y) + y\right]\left(\overrightarrow{W_{y1}} - \overrightarrow{W_{y2}}\right) \end{cases}$
10	}
11	Re-check the feasibility using $TR = \frac{1}{Ts}\sum_{i=1}^{Ts}\left(Pi - \hat{P}i\right)^2$
12	p = p + +
13	}
14	End

In order to achieve more precise de-noising, the loss function optimization is created using the discovered global optimal solution.

4. Result and Discussion

The proposed method is implemented in PYTHON and is assessed based on various measures using two datasets. The proposed method is contrasted with current approaches such as ADMM [21], DEEP-JOINT [22], FLEXISP [23], and Deep Residual DRDD [24] to depict the supremacy of the proposed method.

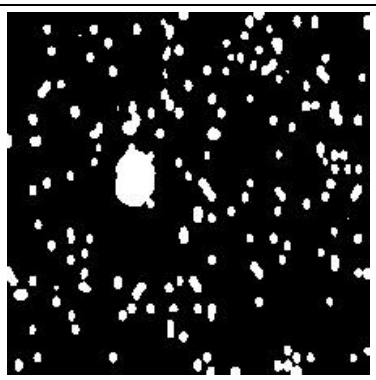
4.1. Dataset Description

McMaster Dataset: The McMaster dataset is a collection of images designed for colour de-mosaicking tasks. This dataset comprises 18 individual images, each of which has been cropped to a size of 500 pixels by 500 pixels.

Kodak Dataset: A popular benchmark dataset in computer vision and image processing is the Kodak dataset. It is made up of a number of photos that have been carefully chosen and calibrated for different image quality evaluation activities.

4.2. Experimental Outcome

The experimental outcome of the proposed de-noising and de-mosaicking technique in terms of input, filtered, de-nosing, and de-mosaicking is portrayed in Figure 5.

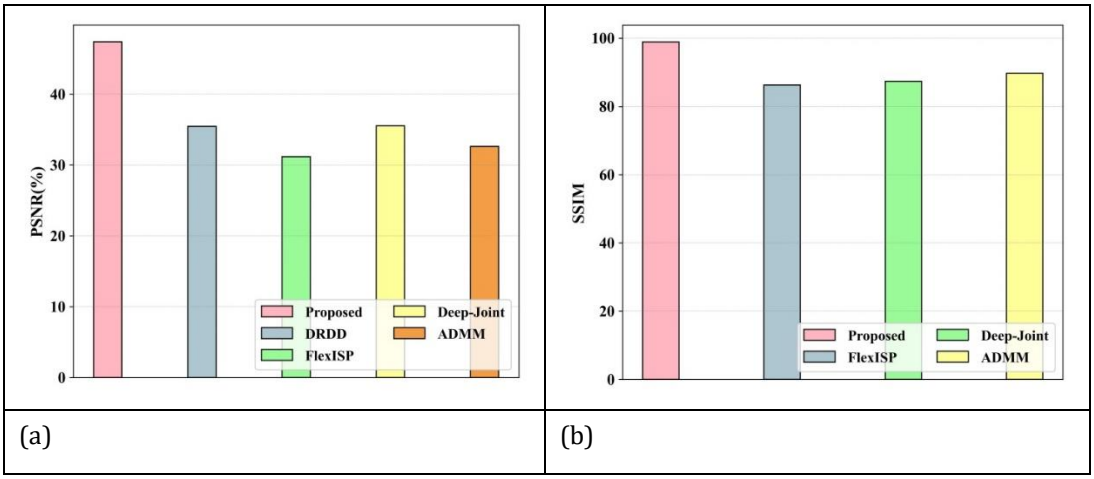
Input		
Filtered Outcome		
De-mosaicked Outcome		

De-noised Outcome		
	(a)	(b)

Figure 5 Experimental Outcome: (a) McMaster Dataset and (b) Kodak dataset

4.3. Analysis based on McMaster Dataset

This section provides an in-depth analysis of the proposed approach using the McMaster Dataset based on the structural similarity index measure (SSIM) and peak signal-to-noise ratio (PSNR) measures. A statistic called PSNR is used to compare an image's quality to an original or reference image. The PSNR is computed by comparing the ratio of a signal's maximal potential power to the amount of corrupting noise that affects the accuracy of the signal's representation. PSNR analysis is helpful when comparing the original image captured by a camera sensor to the quality of the de-mosaicked and de-noised image. Figure 6(a) shows the PSNR analysis of the proposed method performed using the McMaster Dataset. From the analysis, it is clear that the proposed technique produced better results. The SSIM based analysis is presented in Figure 6 (b). SSIM analysis is crucial in image de-mosaicking and de-noising processes. A more comprehensive evaluation of image quality is offered by taking into account both structural and perceptual similarities between the original and reconstructed results. Higher SSIM values indicate that the original image and the reconstructed image have a greater degree of preserved structural similarity. The proposed method acquired higher SSIM using the McMaster Dataset.

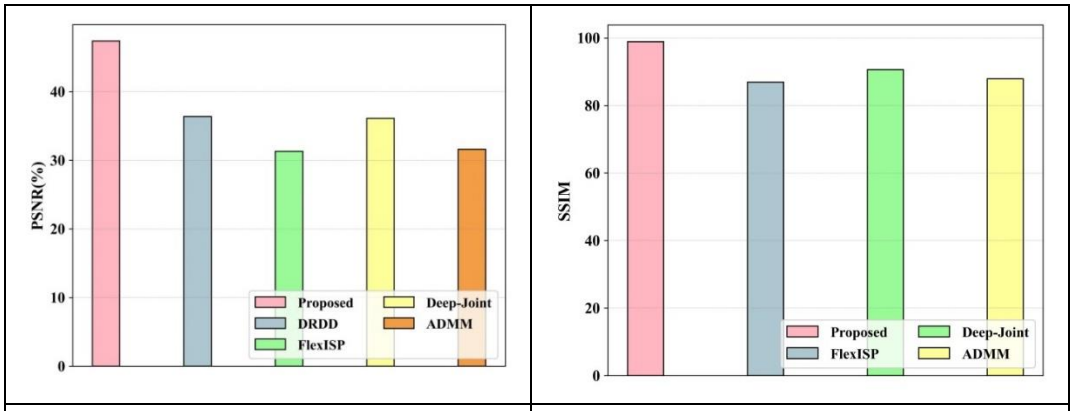


(a) (b)

Figure 6 Analysis based on McMaster Dataset: (a) PSNR and (b) SSIM

4.4. Analysis based on Kodak Dataset

The Kodak dataset based analysis is portrayed in Figure 7, wherein PSNR is presented in Figure 7(a) and SSIM is portrayed in Figure 7(b). The analysis using the Kodak dataset also acquired superior outcomes compared to the existing methods in PSNR and SSIM.



(a) (b)

Figure 7 Analysis based on Kodak Dataset: (a) PSNR and (b) SSIM

4.5. Comparative Discussion

The comparative discussion for the proposed de-mosaicking and de-noising is presented in Table 1. The maximum PSNR calculated by the proposed approach is 46.65%, while the corresponding values for the ADMM, DEEP-JOINT, FLEXISP, and DRDD methods are 32.26%, 22.59%, 32.88%, and 22.02%. The proposed methods yields a maximum SSIM of 98.89% and for the current ADMM, DEEP-JOINT, and FLEXISP methods, which are at 9.29, 11.62, and 12.73%.

Table 1 Comparative Discussion

Methods/ Metrics	**Kodak Dataset**		**McMaster Dataset**	
	PSNR	SSIM	PSNR	SSIM
ADMM	31.6	87.9	32.63	89.7
DEEP-JOINT	36.11	90.6	35.53	87.4
FLEXISP	31.31	86.9	31.17	86.3
DRDD	36.38	-	35.46	-
Proposed	46.65	98.7	42.7	98.89

The analysis based on the PSNR and SSIM portrays the superiority of the proposed image de-mosaicking and de-noising methods. The proposed method divides the image processing task into two stages: de-mosaicking and de-noising. By addressing these two aspects separately, the method can effectively tackle the challenges associated with each stage, leading to better overall performance. In the de-mosaicking stage, the proposed method incorporates a de-noising network specifically designed to remove structured noise resulting from de-mosaicking using the EnConvResNet method. By eliminating this noise, the method improves the quality of the de-mosaicked image. In the de-noising stage, adaptive UNet is utilized, wherein the loss function of the U-Net is employed using the EnGa algorithm. The EnGa optimization technique helps increase the generalization aptitude of the proposed model, allowing it to perform well on a wide range of images and scenarios. Before de-mosaicking and de-noising, the input image undergoes pre-processing using Gaussian filtering. It enhances the quality of the input data fed into the network by assisting in removing artifacts from the input image. The method facilitates more accurate processing and lessens the load on the network by improving the network input. Thus, the enhanced outcome is derived from the proposed model.

5. Conclusion

This research introduced an image reconstruction technique with de-mosaicking and de-noising to address the limitations of existing de-mosaicking and de-noising models, particularly their failure in extreme conditions like imaging short exposure raw data. The proposed method focuses on developing an effective end-to-end solution. First, the image quality is improved by removing artifacts through the use of the Gaussian filtering technique during pre-processing. Subsequently, the method integrates an EnConvResNet for de-mosaicking and an Adaptive U-net restoration model for de-noising. Hyper-parameters are fine-tuned using an EnGa algorithm to enhance the model's generalization

ability. Achieving a PSNR of 46.65 and an SSIM of 98.89 demonstrates the effectiveness of the proposed method. However, challenges remain in extreme conditions, and computational complexity may limit real-time applications. Future research could address these limitations, explore alternative optimization techniques, and extend the application of the method to broader image enhancement tasks.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Ma K, Gharbi M, Adams A, Kamil S, Li TM, Barnes C & Ragan-Kelley J. (2022). Searching for fast demosaicking algorithms. ACM Transactions on Graphics (TOG), 41(5), 1-18.

[2] Rebiere V, Drouot A, Granado B, Bourge A and Pinna A. (2022). Color Pixel Reconstruction for a Monolithic RGB-Z CMOS Imager. Journal of Signal Processing Systems 94(7), 623-644.

[3] Guo Y, Jin Q, Facciolo G, Zeng T and Morel J-M. (2020). Residual learning for effective joint demosaicing-denoising. arXiv preprint arXiv:2009.06205 3.

[4] Elgendy OA, Gnanasambandam A, Chan SH & Ma J. (2021). Low-light demosaicking and de-noising for small pixels using learned frequency selection. IEEE Transactions on Computational Imaging, 7, 137-150.

[5] Xing W and Egiazarian K. (2021). End-to-end learning for joint image demosaicing, de-noising and super-resolution. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 3507-3516.

[6] Liu L, Jia X, Liu J and Tian, Q. (2020). Joint demosaicing and de-noising with self guidance. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2240-2249.

[7] Sharif SMA, Naqvi RA and Biswas M. (2021). Beyond joint demosaicking and de-noising: An image processing pipeline for a pixel-bin image sensor. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 233-242.

[8] de Haan K, Rivenson Y, Wu Y and Ozcan A. (2019). Deep-learning-based image reconstruction and enhancement in optical microscopy. Proceedings of the IEEE 108(1), 30-50.

[9] Li H, Yang M and Yu Z. (2021). Joint image fusion and super-resolution for enhanced visualization via semi-coupled discriminative dictionary learning and advantage embedding. Neurocomputing 422, 62-84.

[10] Zhang T, Fu Y, Zhang J and Yan C. (2024). Deep Guided Attention Network for Joint De-noising and Demosaicing in Real Image. Chinese Journal of Electronics 33(1), 303-312.

[11] Conde MV, Vasluianu F and Timofte R. (2024). BSRAW: Improving Blind RAW Image Super-Resolution. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 8500-8510.

[12] Buades A, Martorell O and Sánchez-Beeckman M. (2024). Joint De-noising and HDR for RAW Image Sequences. IEEE Transactions on Computational Imaging.

[13] Kokkinos F and Lefkimmiatis S. (2019). Iterative joint image demosaicking and de-noising using a residual de-noising network. IEEE Transactions on Image Processing 28(8), 4177-4188.

[14] Li Z, Lu M, Zhang X, Feng X, Asif MS and Ma Z. (2024). Efficient visual computing with camera raw snapshots. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[15] Du GD, Deng H, Su J and Huang Y. (2023). End-to-end Rain Streak Removal with RAW Images. arXiv preprint arXiv:2312.13304.

[16] Ponomarenko M, Marchuk V and Egiazarian K. (2022). FiveNet: joint image demosaicing, de-noising, de-blurring, super-resolution and clarity enhancement. Electronic Imaging 34(14), 218-1.

[17] Guo S, Liang Z and Zhang L. (2021). Joint de-noising and demosaicking with green channel prior for real-world burst images. IEEE Transactions on Image Processing 30, 6930-6942.

[18] Guo Y, Jin Q, Morel J-M, Zeng T and Facciolo G. (2023). Joint demosaicking and de-noising benefits from a two-stage training strategy. Journal of Computational and Applied Mathematics 434, 115330.

[19] Janjušević N, Khalilian-Gourtani A and Wang Y. CDLNet: Noise-adaptive convolutional dictionary learning network for blind de-noising and demosaicing. IEEE Open Journal of Signal Processing 3, 196-211.

[20] Tan H, Xiao H, Liu Y and Zhang M. (2022). Two-stage CNN model for joint Demosaicing and De-noising of burst Bayer images. Computational Intelligence and Neuroscience 2022.

[21] Dong W, Yuan M, Li X & Shi G. (2018). Joint demosaicing and de-noising with perceptual optimization on a generative adversarial network. arXiv preprint arXiv:1802.04723.

[22] Go J, Sohn K & Lee C. (2000). Interpolation using neural networks for digital still cameras. IEEE Transactions on Consumer Electronics, 46(3), 610-616.

[23] Heide F, Steinberger M, Tsai YT, Rouf M, Pająk D, Reddy D & Pulli K. (2014). Flexisp: A flexible camera image processing framework. ACM Transactions on Graphics (ToG), 33(6), 1-13.

[24] He K, Zhang X, Ren S & Sun J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition 770-778.