IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 25, Issue 4, Ser. 2 (July. – August. 2023), PP 09-18
www.iosrjournals.org

Research On Enhancement Of Natural Images Using An Improved Multi-Scale Residual Neural Network

Paul Shekonya Kanda [1], Kewen Xia [1], Romoke Grace Akindele [1], and Aadel Mohamed [2,3]

[1](School of Electronics and Information Engineering, Hebei University of Technology, Tianjin, China)
[2](School of Mechanical Engineering, Hebei University of Technology, Tianjin, China)
[3](Department of Mechanical Engineering, University of Nyala, Nyala, Sudan)

Abstract:
In botany as well as in agriculture, recent advances in deep learning have led to a myriad of automated applications such as automatic plant classification, disease detection and segmentation – all of which require image processing. This means that images from plant leaves, fruits, bark, and flowers will be needed. In real life scenario, it is a herculean task to capture images in proper lighting conditions due to inherent hardware restrictions and environmental conditions, thus, introducing noise in some instances. The visual quality of images is frequently severely degraded by noise, which is first and foremost an unavoidable part of any image processing application.

To enhance the leaf images as well as other natural images, we propose a network based on a multi-scale residual block (MRB), designed to maintain high-resolution representations across the entire network while receiving strong contextual information from low-resolution features using attention-based multi-scale feature aggregation. To validate, we conducted experiments on six real-world image benchmark datasets and compared the results with various methodologies, ensuring that both noisy and ground truth images were symmetric throughout the experiment. Our network performed better on three of those datasets, with a PSNR value of 40.35 dB, 39.18 dB, 36.72 dB, and 31.99 dB on SIDD, PolyU, CC, and a leaf image dataset respectively. This suggests that our MRB network can learn an enriched set of features by combining contextual information from multiple scales while achieving cutting-edge results in image denoising and low-light image enhancement.

Key Words*: deep learning; image denoising; multi-scale residual block; leaf images.*

Date of Submission: 03-08-2023 Date of Acceptance: 13-08-2023

I. Introduction

The overall goal of image processing is to extract useful information hidden in the pattern of an image so that it can be understood, recognized, and adequately interpreted by the system. Image enhancement plays an important role in digital image processing[1] because many imaging applications require image enhancement. The primary goal of image enhancement is to improve an image's suitability for a certain application.

In real life scenario, it is a herculean task to capture images in proper lighting conditions due to inherent hardware restrictions and environmental conditions[2]. In the past, low-light image enhancement algorithms were developed to improve the features in such images and thereby, improve the overall aesthetics of the image. Low-light image enhancement, despite being a low-level task, acts as an essential pre-processing step for deep learning, eventually aiding in enhancing the performance of other computer vision algorithms such as autonomous navigation, face recognition[4], and biometric security systems.

Low-light images typically suffer from color distortion, contrast loss, and blurred scene details, resulting in poor visibility. It hides several traits that are useful in computer vision applications and can cause poor performance as such. As a result, low-light image enhancement is critical for increasing image quality for deep learning applications. The significance of image denoising cannot be overemphasized. The visual quality of the obtained image is frequently severely degraded by noise corruption, which is first and foremost an unavoidable part of any image processing application. In this paper, we propose a method for restoring plant leaf images, taking low-lit plant leaf images, and noisy images and restoring them. This proposed network is based on a multi-scale residual block which receives strong contextual information from low-resolution features using attention-based multi-scale feature aggregation and enhances the images.

II. Related Work

Image restoration has been extensively studied, and its assessment is usually made given one of these three methods: (a) subjective methods—the assessment is typically made via questionnaire, being subjective to each person; (b) behavioral methods—study the relationship between specific behaviors and cognitive fatigue; (c) physiological-based methods—use physiological sensors to monitor the person and then, by processing the acquiring biosignals, relate them with cognitive fatigue.

Histogram Equalization-based methods have been functional in performing light enhancement through the act of expanding the dynamic range of an image where the histogram distribution of various images is adjusted both globally[5,6] and locally[7,8]. On the other hand, there exist various methods which adopt the Retinex theory[9] which typically works at decomposing an image into illumination and reflectance. Light enhancement is often posed as an illumination estimation problem since the reflectance component is generally considered to be consistent under all lighting situations. Several techniques have been proposed based on the Retinex approach. Wang et al.[8] designed a naturalness-and information-preserving method to enhance images of non-uniform illumination; Fu et al.[10] proposed a model based on a weighted variation that will simultaneously estimate both the reflectance and illumination of a low-light input image; Guo et al.[11] proposed a method which estimates a coarse illumination map through searching the maximum intensity of each pixel in the RGB channels, then refining the coarse illumination map via the application of a structure prior; Li et al.[12] proposed a new Retinex model that considers the presence of noise. The illumination map was estimated by solving an optimization problem.

Currently available CNN-based methods typically operate on one of either full-resolution or progressively lower-resolution representations. In the first case, results are spatially precise but can be contextually less robust, whereas in the second case, the output is semantically reliable but can be spatially less accurate. Contrary to the conventional methods that fortuitously change the distribution of image histogram or that rely on potentially inaccurate physical models, the proposed model in this research aims at maintaining both spatially precise high-resolution representations and at the same time receiving strong contextual information from low-resolution representations through the entire network and produces an enhanced result. Such a strategy enables light enhancement on images without creating unrealistic artifacts.

III. Materials and Methods

Proposed Network Model

The details of the building blocks of our proposed network are described in this section.

The Multi-scale Residual Block

The fundamental building block of this network is based on a multi-scale residual block, which encompasses the following key elements:

1. A couple of multi-resolution convolution streams in parallel for the extraction of feature representations,
2. Exchange of information across multi-resolution streams,
3. Aggregation of features based on attention mechanism, arriving from multiple streams,
4. A combination of dual-attention units that allow for the capture of contextual information in the dual dimensions of space and channel, and
5. Residual modules that perform resizing operations through the implementation of up-sampling and down-sampling.

To encode context, current CNNs take on a typical design in architecture where: (a) there exists a fixed receptive field in each layer, (b) a gradual reduction in the spatial size of the generated feature maps, and (c) a gradual recovery of high-resolution representation from previously generated low-resolution feature maps. However, it is a fact in vision science that in the primate visual cortex, the local receptive fields of neurons in the same region are of varying sizes. For this reason, there's a need for such a mechanism for gathering multi-scale spatial information in the same layer to be incorporated into the architecture of CNNs.

In this network, the proposed MRB, acquires rich contextual information from low resolutions, and is capable of generating a spatially-precise output by maintaining high-resolution representations, thereby, maintaining both useful information. It consists of a stream of multiple fully-convolutional layers connected in parallel which gives room for information exchange across these streams via the selective kernel feature fusion to consolidate the low-res features with the help of high-res features, and vice versa. The individual components that make up the MRB are selective kernel feature fusion, dual attention, residual contextual block, and residual resizing module.

Selective kernel feature fusion

Selective kernel feature fusion (SKFF) is a basic building block for this MRB. One of the essential features of neurons in the visual cortex is the ability to modify their receptive fields in response to a stimulus. With the help of two operations – Fuse and Select, this SKFF module is equipped to perform dynamic adjustments of receptive fields. The *Fuse* and *Select* operations are illustrated in Figure 1. The task of the *fuse* operator is generating global feature descriptors through the combination of information coming from the multi-resolution streams. The *select* operator on the other hand utilizes these descriptors for recalibrating the feature maps followed by their aggregation.

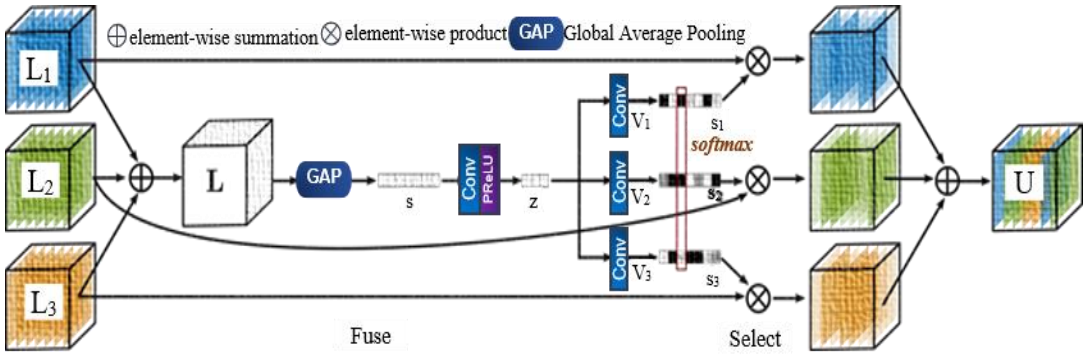

Figure 1. Diagram of the selective kernel feature fusion

1. The Fuse operator

For this network, the input to the SKFF comes from three parallel convolution streams carrying information on different scales. These multi-scale features are first combined by the use of an element-wise sum, $\mathbf{L} = \mathbf{L_1} + \mathbf{L_2} + \mathbf{L_3}$.Global average pooling (GAP) is then applied across the spatial dimensions of $\mathbf{L} \in R^{H\times W\times C}$ for the computation of channel-wise statistics $\mathbf{s} \in R^{1\times 1\times C}$. Then, a channel downscaling convolution layer is applied to generate the compact feature representation $\mathbf{z} \in R^{1\times 1\times r}$, where $r = \frac{C}{8}$.

Finally, the feature vector **z** is passed through three parallel channel-upscaling convolution layers (one for each resolution stream) and provides three feature descriptors $\mathbf{v_1}$,$\mathbf{v_2}$ and $\mathbf{v_3}$, each with dimensions of $1 \times 1 \times C$.

2. The select operator

This operator applies the Softmax function to the three feature descriptors $\mathbf{v_1}$,$\mathbf{v_2}$ and $\mathbf{v_3}$, and yields attention activations $\mathbf{s_1}$,$\mathbf{s_2}$ and $\mathbf{s_3}$ that are used for adaptively recalibrating the multi-scale feature maps $\mathbf{L_1}$,$\mathbf{L_2}$ and $\mathbf{L_3}$, respectively. The whole feature recalibration and aggregation procedure are defined as follows:

$\mathbf{U} = \mathbf{s_1} \cdot \mathbf{L_1} + \mathbf{s_2} \cdot \mathbf{L_2} + \mathbf{s_3} \cdot \mathbf{L_3}$. It aggregates characteristics from numerous convolutional streams using self-attention.

Dual attention unit

The Dual attention unit (DAU) is the next useful component of the MRB design. While the SKFF block fuses information across multi-resolution branches, a mechanism for sharing information inside a feature tensor, both along the spatial and channel dimensions, is also required. Motivated by recent breakthroughs in low-level vision algorithms based on attention mechanisms[13,14], the DAU is proposed to extract features in the convolutional streams. The DAU with the help of its attention mechanism suppresses features that are less useful and only allows those features that are more informative to pass through. This recalibration of features is achieved by the use of both channel and spatial attention mechanisms. These two attention mechanisms are described below:

1. Spatial attention

The spatial attention (SA) branch of the DAU is specifically designed to exploit the inter-spatial dependencies of the convolutional features. It has as its goal the generation of a spatial attention map and using it to fine-tune the incoming features **M**. To achieve the generation of these spatial attention maps, the SA branch independently applies global average pooling and then the max pooling operations on features **M** along the channel dimensions and afterward, concatenates the outputs to form a new feature map $\mathbf{f} \in R^{H\times W\times 2}$. This feature map **f** is passed through a convolution operation and a sigmoid activation to obtain the spatial attention map $\hat{\mathbf{f}} \in R^{H\times W\times 1}$, which is then used in rescaling **M**.

2. Channel attention

By applying *squeeze* and *excitation* operations[13], the Channel Attention (CA) branch of the DAU can exploit the inter-channel relationships among the convolutional feature maps. For a given feature map $\mathbf{M} \in R^{H\times W\times C}$, the squeeze operation applies global average pooling across spatial dimensions to encode global context, therefore, yielding the feature descriptor $\mathbf{d} \in R^{1\times 1\times C}$. The excitation operator, on the other hand, feeds the descriptor $\mathbf{d}$ through two convolutional layers and then applies the sigmoid gating to it, generating activations $\hat{\mathbf{d}} \in R^{1\times 1\times C}$. Finally, the output of the CA branch is obtained through the process of rescaling $\mathbf{M}$ with the generated activations $\hat{\mathbf{d}}$. Figure 2 shows the block diagram of the MRB.

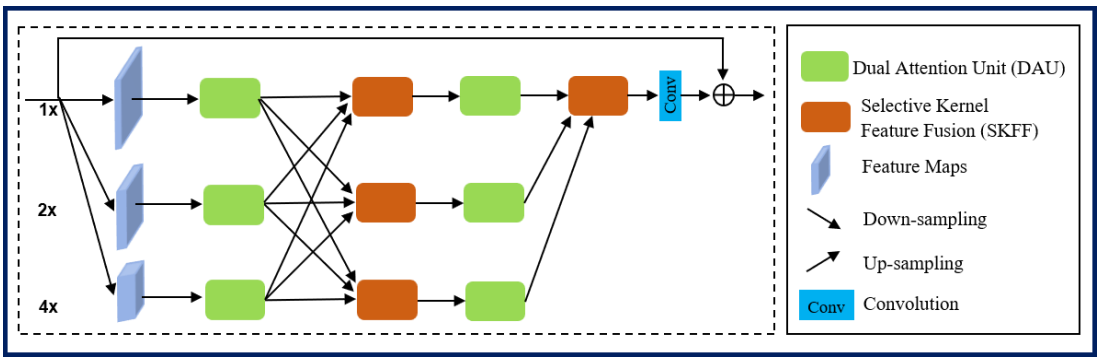

Figure 2. Block diagram of the MRB

End-to-End Architecture of our Proposed Network

Given an input image $\mathbf{I} \in R^{H\times W\times 3}$, with the proposed network aimed at enhancing the distorted input image, it first applies a convolutional operation as the first layer to extract low-level feature maps $\mathbf{X_0} \in R^{H\times W\times C}$. These feature maps $\mathbf{X_0}$ is then passed through an N number of recursive residual groups (RRGs), which further yields deep features $\mathbf{X_d} \in R^{H\times W\times C}$. Each of the RRGs is designed to contain several MRBs.

Immediately after the *Nth* RRG in the network, a convolution layer is applied to the output deep features $\mathbf{X_d}$ obtaining a residual image $\mathbf{R} \in R^{H\times W\times 3}$. Finally, the enhanced image is obtained as $\hat{\mathbf{I}} = \mathbf{I} + \mathbf{R}$. Note that the proposed network is optimized using the Charbonnier loss.

$$\mathcal{L}(\hat{\mathbf{I}}, \mathbf{I}^*) = \sqrt{\left\|\hat{\mathbf{I}} - \mathbf{I}^*\right\|^2 + \varepsilon^2} \qquad (2)$$

where $\mathbf{I}^*$ signifies the ground-truth image, and ε denotes a constant which has been empirically fixed at 10^{-3} throughout the experiments. Figure 3 is the overall framework of our proposed network, showing the block diagram of the overall modules of the proposed network for enhancing distorted images. It shows the arrangement of the recursive residual group (RRG) in the center of the convolution operations performed on the initial input images and before the output image. The RRG module, containing the multi-scale residual block (MRB) is also displayed, and finally, the components that make up the MRB such as the dual attention unit (DAU), and selective kernel feature fusion (SKFF) amid other operations in the process of the network that is shown.

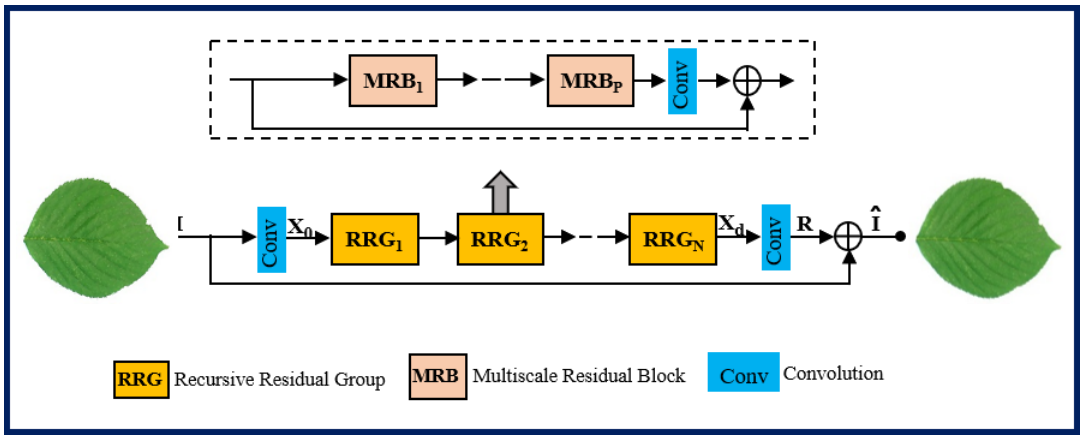

Figure 3. An end-to-end framework of our proposed network

Details of Experimental Dataset

In this research, several image datasets were utilized, coming to a total of about seven different datasets, five of them for denoising, combining both images containing real noise and some containing only Gaussian white noise. Additionally, three of these datasets are low-light images for image enhancement.

1. The Smartphone Image Denoising Dataset (SIDD)[15] is an image denoising dataset that was particularly collected with the use of smartphone cameras. Due to the small sensor and high resolution, the noise levels in images from smartphones are much higher than those of DSLRs. SIDD contains 320 image pairs for training.
2. The PolyU dataset[16] is a real-world noisy image denoising dataset, containing 40 different scenes captured by 5 cameras from three leading camera brands.
3. The CC dataset consists of images that were captured by [17] for training for 11 static scenes, 500 JPEG images per scene, and the mean image of each scene was computed to generate the ground truth noise-free images.
4. CBSD68 dataset[18] is a set of images which is part of the Berkeley Segmentation Dataset and Benchmark. This benchmark dataset is widely used for measuring image denoising algorithms performance and contains 68 images. It includes the original .jpg files, converted to lossless .png, and noisy with additive white gaussian noise (AWGN) of different levels.
5. MIT-Adobe FiveK dataset is the publicly available MIT-Adobe FiveK[19], otherwise referred to as MIT-FiveK, contains 5000 images of various indoor and outdoor scenes captured with DSLR cameras in different lighting conditions. The tonal attributes of all images are manually adjusted by five different trained photographers (labeled as experts A to E). Same as in [20], we also consider the enhanced images of expert C as the ground truth.
6. LOL dataset - The LOL dataset[21] is created for the low-light image enhancement problem. It provides 485 images for training and 15 for testing. Each image pair in this dataset consists of a low-light input image and its corresponding well-exposed reference image.
7. Modified MalayaKew leaf dataset - This is a modified dataset prepared specifically for this research. It was generated from preprocessing the original MalayaKew dataset. This dataset, collected at the Royal Botanic Gardens, Kew, England[22], consists of scan-like images of leaves from 44 species classes. By the use of the Opencv image processing library, the ground-truth leaf images were distorted by the addition of additive white gaussian noise (AWGN), and the brightness was reduced to create a low-light image. Both the generated low-lit and ground truth images were ensured to be symmetric throughout the experiment.

IV. Implementation Details

Implementation Procedure

Now, the implementation details of the proposed network are as follows. Two different networks are trained for the two different tasks in this research – for denoising and low-light image enhancement respectively. Note that the proposed network is end-to-end trainable and does not require pre-training of any of its sub-modules. In training, 3 RRGs are used, each containing 3 MRBs. The MRB has 3 parallel streams with channel dimensions of 64, 128, and 256 at resolutions 1, $\frac{1}{2}$, $\frac{1}{4}$, respectively. Each stream in the MRB has 3 RCBs with shared parameters for image denoising, whereas, the MRB has 3 DAUs for low-light image enhancement.

The network is trained with the Adam optimizer [23] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-3}$ for 7×10^5iterations. An initial learning rate of 2×10^{-4} was adopted, and an implementation of the cosine annealing strategy to steadily decrease the learning rate during training from its initial value to 10^{-6}. Additionally, patches of size 128 × 128 were extracted from images during training, and the batch size was set to 8, with horizontal and vertical flip operations being applied to the input images for data augmentation.

A total of 3200 training images of patches generated from the SSID dataset were used in training the network for the denoising task, while images from the MIT-FiveK dataset were used to train the network for the low-light enhancement task. For each task, other public datasets were used for testing; and records of the PSNR and SSIM [24] were taken to evaluate the results of the network. For the leaf images, the original MalayaKew leaf dataset was pre-processed using the OpenCV framework to distort the images, producing new noisy images by the addition of white gaussian noise, and also new low-light images with the adjustment of the brightness coefficient of the images, while taking the original images as ground truth.

Implementation and testing of the proposed network were carried out on a computer system with the following specifications: Windows 10, 64-bit, Intel Core i7-4720 CPU @ 2.60 GHz, RAM 32 GB, and GPU Nvidia GeForce GTX 1050 4 GB dedicated memory and Python 3.7 on Anaconda. The PC was used on a Linux-based Dell PowerEdge T640 Tower Server with CUDA-based video cards 4X 1080TI, each GPU Video memory is 11GB, with a storage memory of 10TB Hard Drive and 3320GB SSD.

Evaluation Metrics

For proper evaluation of the performance of an image denoiser and low-light enhancing algorithm, there is utmost necessity to utilize multifaceted evaluation criteria. The following metrics were used:

Mean Squared Error

The Mean Squared Error (MSE) is a metric used to describe the average square of the difference between the input and the denoised image. It is a full reference metric and obtaining a lower MSE value often denotes greater image quality. The MSE between two given images $g(x, y)$ and $\hat{g}(x, y)$ is defined as:

$$\text{MSE} = \frac{1}{MN} \sum_{n=0}^{M} \sum_{m=1}^{N} [\hat{g}(n, m) - g(n, m)]^2 \quad (3)$$

Peak Signal to Noise Ratio

Peak Signal to Noise Ratio (PSNR) is used to compute the ratio of the maximum signal strength to the power of the distorted noise that impairs the accuracy of its representation. The decibel form is used to calculate this ratio between the two images. The highest and smallest values that can be achieved within this dynamic range, which can alter depending on their quality, are represented.

PSNR is the most popular method for evaluating the quality of reconstruction in lossy image compression codecs. Obtaining a higher PSNR value often denotes better image quality. It is mathematically expressed as:

$$\text{PSNR} = 10\log_{10}(\text{peakval}^2)/\text{MSE} \quad (4)$$

Structure Similarity Index Method

Structure Similarity Index Method (SSIM) is a metric for comparing the perceived differences between two images that are similar in terms of luminance, contrast, and image structure. Obtaining a higher SSIM value often signifies better image quality. The Structural Similarity Index Method can be expressed as:

$$\text{SSIM}(x, y) = [l(x, y)]^{\alpha} \cdot [c(x, y)]^{\beta} \cdot [s(x, y)]^{\gamma} \quad (5)$$

where *l* is the luminance, *c* is the contrast, *s* is the structure, α, β and γ are positive constants.

V. Results and Discussion

Qualitative and Quantitative Evaluation of the Results

Image Denoising Results

Figures 4–5 show a qualitative analysis of the proposed network for denoising noisy image samples from some datasets, while Figure 6 displays the qualitative visual result of denoising samples from the plant leaf images used for this research.

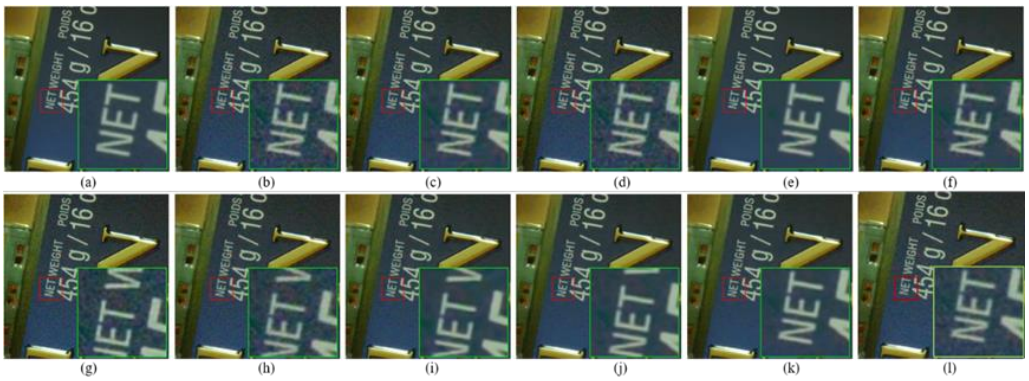

Figure 4. Qualitative comparison using the CC dataset. (a) Ground-Truth, (b) Noisy Image-29.63dB, (c) CBM3D-31.96dB, (d) WNNM-32.97dB, (e) MCWNNM-34.61dB, (f) NC-33.49dB, (g) N2V-single-29.77dB, (h) N2S-single-30.38dB, (i) DIP-33.88dB, (j) S2S-34.43dB, (k) R2R-34.80dB, (l) Proposed-36.37dB.

Figure 5 shows a more qualitative comparison of the proposed network in denoising sample images from the SIDD dataset. For the image at the top, (a) Noisy (18.16dB), (b) RIDNet (29.83dB), (c) VDN (30.31dB), (d) MIRNet (31.36dB), (e) Proposed (31.12dB), (f) Groundtruth. This network effectively removes real noise from very difficult real images. For the image at the bottom, (a) Noisy (18.25dB), (b) RIDNet (35.57dB), (c) VDN (36.39dB), (d) MIRNet (36.97dB), (e) Proposed (36.48dB), (f) Groundtruth. This network effectively removes real noise from very difficult real images while at the same time improving structural content and fine texture recovery.

Figure 5. More qualitative comparison using the SIDD dataset.

Figure 6. Qualitative analysis of denoised leaf images

From Figure 6, the network effectively removes white gaussian noise from the leaf images while at the same time improving structural content and fine texture recovery.

The proposed network is effective at reducing real noise and producing perceptually attractive and sharp images. Furthermore, it is capable of maintaining the spatial smoothness of homogenous regions without generating artifacts. In contrast, most other approaches either produce overly smooth images, sacrificing structural substance and fine textural details, or produce images with chroma artifacts and blotchy texture.

Table no 2. Denoising using SIDD dataset

Method	**DnCNN**	**MLP**	**GLIDE**	**TNRD**	**FoE**	**BM3D**	**WNNM**	**NLM**	**RIDNet**	**VDN**	**MIRNet**	**Ours**
PSNR↑	23.66	24.71	24.71	24.73	25.58	25.65	25.78	26.76	38.71	39.28	39.72	**40.35**
SSIM↑	0.583	0.641	0.774	0.643	0.792	0.685	0.809	0.699	0.914	0.909	0.959	0.97

Quantitative comparisons based on PSNR and SSIM metrics are made summarily in Tables 2–5. The tables show a comparison with the following models: DnCNN[25], MLP[26], GLIDE[27], TNRD[28], FoE[29], BM3D[30], WNNM[31], NLM[32], KSVD[33], EPLL[34], CBDNet[35], RIDNet[36], VDN[37], MIRNet[38]. These tables show that the proposed network performs favourably against both the data-driven and conventional, image denoising algorithms. Specifically, when compared to the MIRNet[38], this network demonstrates a performance gain of 0.63 dB on SIDD and 0.24 dB on PolyU. Furthermore, it is worth noting that CBDNet[35] and RIDNet[36] use additional training data, yet this network provides significantly better results.

Table no 3. Denoising using PolyU dataset

Method	**N2V-single**	**N2S-single**	**DnCNN**	**WNNM**	**NC**	**CBM3D**	**DIP**	**S2S**	**R2R-single**	**MCWNNM**	**MIRNet**	**Ours**
PSNR↑	33.83	35.04	35.60	36.59	36.92	37.40	33.65	34.23	38.51	38.17	38.94	**39.18**
SSIM↑	0.873	0.902	0.964	0.925	0.945	0.953	0.831	0.833	0.967	0.951	0.99	0.995

Table no 4. Denoising using CC dataset

Method	**N2V-single**	**N2S-single**	**DnCNN**	**WNNM**	**NC**	**CBM3D**	**DIP**	**S2S**	**R2R-single**	**MCWNNM**	**MIRNet**	**Ours**
PSNR↑	32.27	33.38	33.47	35.77	36.43	35.19	37.37	37.52	37.78	37.70	36.23	**36.72**
SSIM↑	0.862	0.846	0.932	0.938	0.936	0.858	0.947	0.947	0.951	0.954	0.97	0.98

Image Delighting Results

Figure 7 shows a visual comparison of the proposed network for improving the low-light images on samples from the LOL dataset. Compared to other techniques, the proposed network generates enhanced images that are natural and vivid in appearance and have better global and local contrast.

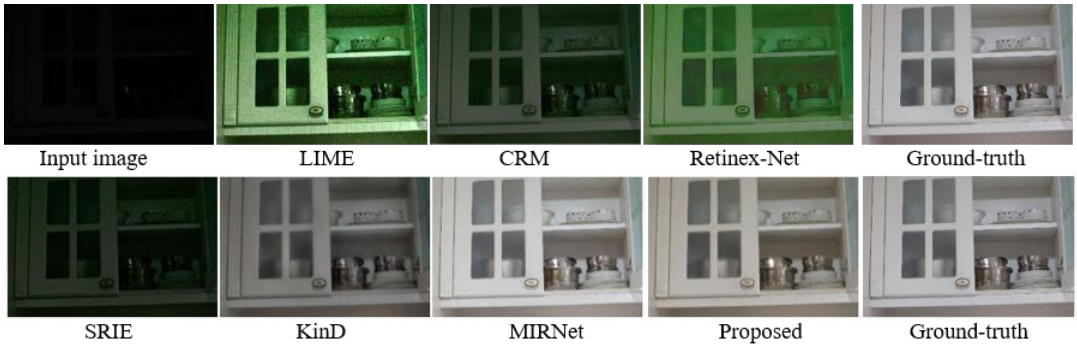

Figure 7. Visual comparison using the LOL dataset.

The proposed method reproduces images that are visually closer to the ground truth in terms of brightness and global contrast. Figure 8 shows the visual results and comparison of the proposed network for improving the low-light images on sample images from the generated MalayaKew leaf dataset.

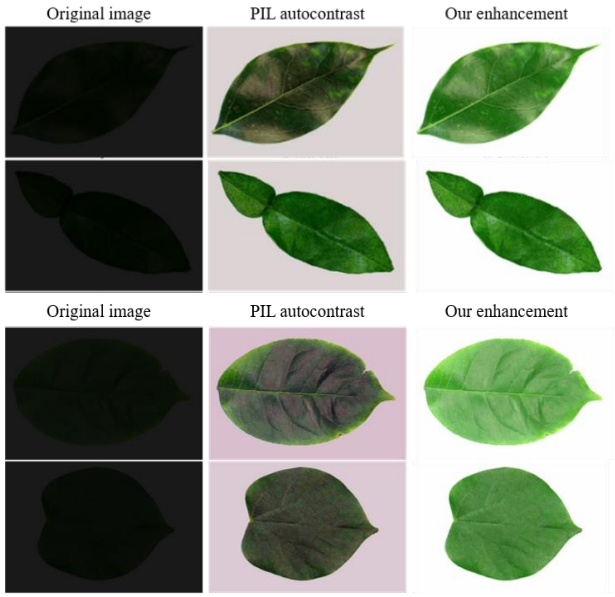

Figure 8. Qualitative analysis of delighted leaf images

With comparably good PSNR and SSIM values, and visually enhanced images on the above datasets, the above results show that the proposed network can achieve successful results for performing low-light image enhancement and image denoising. From Figure 8 above, even the artifacts that the PIL auto contrast deposited on the image was deleted by our network, giving a more realistic image that is close to the ground truth leaf image.

VI. Conclusion

We proposed and conducted research on an image enhancement network based on a multi-scale residual block (MRB) is proposed. The novel multi-scale residual block is designed to achieve the collective goals of maintaining spatially precise, high-resolution representations across the entire network while receiving strong contextual information from low-resolution representations. The multi-scale residual block contains several key components: (a) parallel multi-resolution convolution streams for extracting multi-scale features, (b) information exchange across the multi-resolution streams, (c) spatial and channel attention mechanisms for capturing contextual information, and (d) attention-based multi-scale feature aggregation. Extensive experiments on six real-world image benchmark datasets and the MalayaKew leaf dataset prepared for this research show that this proposed method achieves cutting-edge results for some image processing tasks,

including image denoising, and low-light enhancement. Our network, having a symmetric number of RRGs and MRBs, performed better on three of those datasets, with a PSNR value of 40.35 dB, 39.18 dB, 36.72 dB, and 31.99 dB on SIDD, PolyU, CC, and the leaf image dataset respectively. This suggests that our MRB network can learn an enriched set of features by combining contextual information from multiple scales while achieving cutting-edge results in image denoising and low-light image enhancement.

References

[1] Gonzales R C, Woods R E. Digital Image Processing Second Edition. 2001.
[2] Li Chen, Yingfang Li, Jingquan Tian. An Improved Multi-Frame Integration Technique For Low Light Level Image. In: 2010 International Conference On Computer, Mechatronics, Control And Electronic Engineering. IEEE, 2010.
[3] Zhang C, Jiang H, Jiang C, Et Al. Low Light Level Image De-Noising Algorism Based On Wavelet Transform And Morphology. In: 2009 International Symposium On Computer Network And Multimedia Technology. IEEE, 2009 : 1–4: : 1–4.
[4] Tang F, Wu X, Zhu Z, Et Al. An End-To-End Face Recognition Method With Alignment Learning. Optik, 2020, 205: 164238.
[5] Coltuc D, Bolon P, Chassery J-M. Exact Histogram Specification. IEEE TRANSACTIONS ON IMAGE PROCESSING, 2006, 15(5): 1143.
[6] Ibrahim H, Kong N S P. Brightness Preserving Dynamic Histogram Equalization For Image Contrast Enhancement. IEEE Transactions On Consumer Electronics, 2007, 53(4): 1752–1758.
[7] Lee C, Kim C S, Lee C. Contrast Enhancement Based On Layered Difference Representation Of 2d Histograms. IEEE Transactions On Image Processing, 2013, 22(12): 5372–5384.
[8] Wang S, Zheng J, Hu H M, Et Al. Naturalness Preserved Enhancement Algorithm For Non-Uniform Illumination Images. IEEE Transactions On Image Processing, 2013, 22(9): 3538–3548.
[9] Land E H. The Retinex Theory Of Color Vision. Scientific American, 1977, 237(6): 108–128.
[10] Fu X, Zeng D, Huang Y, Et Al. A Weighted Variational Model For Simultaneous Reflectance And Illumination Estimation. In: Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition (CVPR). 2016 : 2782–2790: : 2782–2790.
[11] Guo X, Li Y, Ling H. LIME: Low-Light Image Enhancement Via Illumination Map Estimation. IEEE Transactions On Image Processing, 2017, 26(2): 982–993.
[12] Li M, Liu J, Yang W, Et Al. Structure-Revealing Low-Light Image Enhancement Via Robust Retinex Model. IEEE Transactions On Image Processing, 2018, 27(6): 2828–2841.
[13] Hu J, Shen L, Sun G. Squeeze-And-Excitation Networks. In: 2018 IEEE/CVF Conference On Computer Vision And Pattern Recognition. IEEE, 2018 : 7132–7141: : 7132–7141.
[14] Wang X, Girshick R, Gupta A, Et Al. Non-Local Neural Networks. In: Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition (CVPR). 2018 : 7794–7803: : 7794–7803.
[15] Abdelhamed A, Lin S, Brown M S. A High-Quality Denoising Dataset For Smartphone Cameras. In: IEEE Computer Vision And Pattern Recognition Workshops (CVPRW). 2018.
[16] Xu J, Li H, Liang Z, Et Al. Real-World Noisy Image Denoising: A New Benchmark. Arxiv Preprint, 2018.
[17] Nam S, Hwang Y, Matsushita Y, Et Al. A Holistic Approach To Cross-Channel Image Noise Modeling And Its Application To Image Denoising. In: 2016 IEEE Conference On Computer Vision And Pattern Recognition (CVPR). IEEE, 2016 : 1683–1691: : 1683–1691.
[18] Martin D, Fowlkes C, Tal D, Et Al. A Database Of Human Segmented Natural Images And Its Application To Evaluating Segmentation Algorithms And Measuring Ecological Statistics. In: Proceedings Eighth IEEE International Conference On Computer Vision. ICCV 2001. IEEE Comput. Soc, 2001 : 416–423: : 416–423.
[19] Bychkovsky V, Paris S, Chan E, Et Al. Learning Photographic Global Tonal Adjustment With A Database Of Input / Output Image Pairs. In: CVPR 2011. IEEE, 2011 : 97–104: : 97–104.
[20] Wang R, Zhang Q, Fu C-W, Et Al. Underexposed Photo Enhancement Using Deep Illumination Estimation. In: Proceedings Of The IEEE/CVF Conference On Computer Vision And Pattern Recognition (CVPR). 2019 : 6849–6857: : 6849–6857.
[21] Wei C, Wang W, Yang W, Et Al. Deep Retinex Decomposition For Low-Light Enhancement. British Machine Vision Conference 2018, (BMV) 2018, 2018,127–136.
[22] Lee S H, Chan C S, Wilkin P, Et Al. Deep-Plant: Plant Identification With Convolutional Neural Networks. In: 2015 IEEE International Conference On Image Processing (ICIP). IEEE, 2015 : 452–456: : 452–456.
[23] Kingma D P, Ba J L. Adam: A Method For Stochastic Optimization. In: 3rd International Conference On Learning Representations, ICLR 2015 - Conference Track Proceedings. San Diego, California, USA: International Conference On Learning Representations, ICLR, 2015 : 1–15: : 1–15.
[24] Wang Z, Bovik A C, Sheikh H R, Et Al. Image Quality Assessment: From Error Visibility To Structural Similarity. IEEE Transactions On Image Processing, 2004, 13(4): 600–612.
[25] Zhang K, Zuo W, Chen Y, Et Al. Beyond A Gaussian Denoiser: Residual Learning Of Deep Cnn For Image Denoising. IEEE Transactions On Image Processing, 2017, 26(7): 3142–3155.
[26] Burger H C, Schuler C J, Harmeling S. Image Denoising: Can Plain Neural Networks Compete With Bm3d?. In: 2012 IEEE Conference On Computer Vision And Pattern Recognition. IEEE, 2012 : 2392–2399: : 2392–2399.
[27] Talebi H, Milanfar P. Global Image Denoising. IEEE Transactions On Image Processing, 2014, 23(2): 755–768.
[28] Chen Y, Yu W, Pock T. On Learning Optimized Reaction Diffusion Processes For Effective Image Restoration. In: Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition (CVPR). 2015 : 5261–5269: : 5261–5269.
[29] Roth S, Black M J. Fields Of Experts. International Journal Of Computer Vision, 2009, 82(2): 205–229.
[30] Dabov K, Foi A, Katkovnik V, Et Al. Image Denoising By Sparse 3-D Transform-Domain Collaborative Filtering. IEEE Transactions On Image Processing, 2007, 16(8): 2080–2095.
[31] Gu S, Zhang L, Zuo W, Et Al. Weighted Nuclear Norm Minimization With Application To Image Denoising. In: Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition (CVPR). 2014 : 2862–2869: : 2862–2869.
[32] Buades A, Coll B, Morel J-M. A Non-Local Algorithm For Image Denoising. In: 2005 IEEE Computer Society Conference On Computer Vision And Pattern Recognition (CVPR'05). IEEE, 2005 : 60–65: : 60–65.
[33] Aharon M, Elad M, Bruckstein A. K-Svd: An Algorithm For Designing Overcomplete Dictionaries For Sparse Representation. IEEE Transactions On Signal Processing, 2006, 54(11): 4311–4322.
[34] Zoran D, Weiss Y. From Learning Models Of Natural Image Patches To Whole Image Restoration. In: 2011 International Conference On Computer Vision. IEEE, 2011 : 479–486: : 479–486.

[35] Guo S, Yan Z, Zhang K, Et Al. Toward Convolutional Blind Denoising Of Real Photographs. In: 2019 IEEE/CVF Conference On Computer Vision And Pattern Recognition (CVPR). IEEE, 2019 : 1712–1722: : 1712–1722.
[36] Anwar S, Barnes N. Real Image Denoising With Feature Attention. In: Proceedings Of The IEEE/CVF International Conference On Computer Vision (ICCV). 2019 : 3155–3164: : 3155–3164.
[37] Yue Z, Yong H, Zhao Q, Et Al. Variational Denoising Network: Toward Blind Noise Modeling And Removal. In: Wallach H, Larochelle H, Beygelzimer A, D\Textquotesingle Alché-Buc F, Fox E, Garnett R, Editors. Advances In Neural Information Processing Systems. Curran Associates, Inc., 2019.
[38] Zamir S W, Arora A, Khan S, Et Al. Learning Enriched Features For Real Image Restoration And Enhancement. In: Lecture Notes In Computer Science (Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics). Springer, Cham, 2020 : 492–511: : 492–511.