

# Low Light Image Enhancement Using Mirnet

MD Sana Meharaj<sup>1</sup>, Esampelli Malathi<sup>2</sup>, Chittumalla Keerthana<sup>3</sup>,  
Poosala Harshitha<sup>4</sup>, Dasari Mahesh Kumar<sup>5</sup>

<sup>1,2,3,4,5</sup>Final Year – Department Of Electronics And Communication Engineering, Jyothishmathi Institute Of Technology And Science

## ABSTRACT

With the goal of recovering high-quality image content from its degraded version, image restoration enjoys numerous applications, such as in surveillance, computational photography, medical imaging, and remote sensing. Recently, convolutional neural networks (CNNs) have achieved dramatic improvements over conventional approaches for image restoration task. Existing CNN- based methods typically operate either on full-resolution or on progressively low-resolution representations. In the former case, spatially precise but contextually less robust results are achieved, while in the latter case, semantically reliable but spatially less accurate outputs are generated. In this paper, we present a novel architecture with the collective goals of maintaining spatiallyprecise high- resolution representations through the entire network, and receiving strong contextual information from the low-resolution representations. The core of our approach is a multi-scale residual block containing several key elements: (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. In a nutshell, our approach learns an enriched set of features that combines contextual information from multiple scales, while simultaneously preserving the high-resolution spatial details. Extensive experiments on five real image benchmark datasets demonstrate that our method, named as MIRNet, achieves state-of-the-art results for a variety of image processing tasks, including image denoising, super-resolution and image enhancement.

## 1. INTRODUCTION

Image content is exponentially growing due to the ubiquitous presence of cameras on various devices. During image acquisition, degradations of different severity are often introduced. It is either because of the physical limitations of cameras or due to inappropriate lighting conditions. For instance, smartphone cameras, come with a narrow aperture and have small sensors with limited dynamic range. Consequently, they frequently generate noisy and low-contrast images. Similarly, images captured under the unsuitable lighting are either too dark or too bright. The art of recovering the original clean image from its corrupted measurements is studied under the image restoration task. It is an ill-posed inverse problem, due to the existence of many possible solutions. Recently, deep learning models have made significant advancements for image restoration and enhancement, as they can learn strong (generalizable) priors from large-scale datasets. Existing CNNs typically follow one of the two architecture designs: 1) an encoder-decoder, or 2) high-resolution (single-scale) feature processing. The encoder-decoder models first progressively map the input to a low-resolution representation, and then

apply a gradual reverse mapping to the original resolution. Although these approaches learn a broad context by spatial-resolution reduction, on the downside, the fine spatial details are lost, making it extremely hard to recover them in the later stages. On the other side, the high-resolution (single-scale) networks do not employ any down sampling operation, and thereby produce images with spatially more accurate details. However, these networks are less effective in encoding contextual information due to their limited receptive field.

Image restoration is a position-sensitive procedure, where pixel-to-pixel correspondence from the input image to the output image is needed. Therefore, it is important to remove only the undesired degraded image content, while carefully preserving the desired fine spatial details (such as true edges and texture). Such functionality for segregating the degraded content from the true signal can be better incorporated into CNNs with the help of large context, e.g., by enlarging the receptive field. Towards this goal, we develop a new multi-scale approach that maintains the original high-resolution features along the network hierarchy, thus minimizing the loss of precise spatial details. Simultaneously, our model encodes multi-scale context by using parallel convolution streams that process features at lower spatial resolutions. The multi-resolution parallel branches operate in a manner that is complementary to the main high-resolution branch, thereby providing us more precise and contextually enriched feature representations.

## 2. LITERATURE SURVEY

Image enhancement and denoising are the most prominent topics in computer vision. The two main types of low-light picture enhancement are histogram-based approaches and Retinex-based methods. Histogram Equalization (HE) is the first classification approach image enhancement. The second category is based on the Retinex hypothesis, which contends that light and reflection make up an image. Land introduces the Retinex hypothesis to describe how the human visual system perceives color.

**2.1 PROPOSED SYSTEM:** In this section, we first present an overview of the proposed MIRNet for image restoration and enhancement, illustrated in Fig. 1. We then provide details of the multi-scale residual block, which is the fundamental building block of our Learning Enriched Features for Real Image Restoration and Enhancement 5 method, containing several key elements: (a) parallel multi-resolution convolution streams for extracting (fine-to-coarse) semantically-rich and (coarse-to-fine) spatially-precise feature representations, (b) information exchange across multi-resolution streams, (c) attention-based aggregation of features arriving from multiple streams, (d) dual-attention units to capture contextual information in both spatial and channel dimensions, and (e) residual resizing modules to perform downsampling and upsampling operations. Overall Pipeline. Given an image  $I \in \mathbb{R}^{H \times W \times 3}$ , the network first applies a convolutional layer to extract low-level features  $X_0 \in \mathbb{R}^{H \times W \times C}$ . Next, the feature maps  $X_0$  pass through  $N$  number of recursive residual groups (RRGs), yielding deep features  $X_d \in \mathbb{R}^{H \times W \times C}$ . We note that each RRG contains several multi-scale residual blocks, which is described in Section 3.1. Next, we apply a convolution layer to deep features  $X_d$  and obtain a residual image  $R \in \mathbb{R}^{H \times W \times 3}$ . Finally, the restored image is obtained as  $\hat{I} = I + R$ . We optimize the proposed network using the Charbonnier loss:

$$L(I^{\wedge}, I^{*}) = \sqrt{\|I + I^{\wedge} * \|^2 + \epsilon^2} \dots \dots \dots (1)$$

where  $I^*$  denotes the ground-truth  $image$ , and  $\epsilon$  is a constant which we empirically set to  $10^{-3}$  for all the experiments.

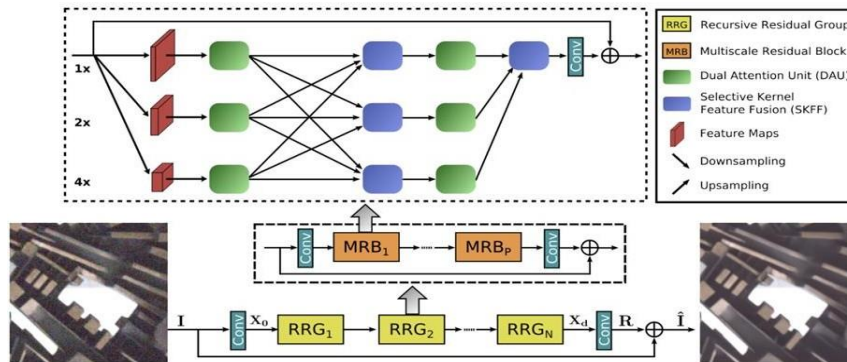


Fig.1: Schematic for MIRNet.

### 3. IMPLEMENTATION IMAGE ENHANCEMENT USING MIRnet Multi-scale Residual Block (MRB):

In order to encode context, existing CNNs typically employ the following architecture design: (a) the receptive field of neurons is fixed in each layer/stage, (b) the spatial size of feature maps is gradually reduced to generate a semantically strong low-resolution representation, and (c) a high-resolution representation is gradually recovered from the low-resolution representation. However, it is well-understood in vision science that in the primate visual cortex, the sizes of the local receptive fields of neurons in the same region are different. Therefore, such a mechanism of collecting multi-scale spatial information in the same layer needs to be incorporated in CNNs. In this paper, we propose the multi-scale residual block (MRB), as shown in Fig. 1. It is capable of generating a spatially-precise output by maintaining high-resolution representations, while receiving rich contextual information from low-resolutions. The MRB consists of multiple (three in this paper) fully-convolutional streams connected in parallel. It allows information exchange across parallel streams in order to consolidate the high-resolution features with the help of low-resolution features, and vice versa. Next, we describe the individual components of MRB.

**Selective kernel feature fusion (SKFF):** One fundamental property of neurons present in the visual cortex is to be able to change their receptive fields according to the stimulus. This mechanism of adaptively adjusting receptive fields can be incorporated in CNNs by using multi-scale feature generation (in the same layer) followed by feature aggregation and selection. The most commonly used approaches for feature aggregation include simple concatenation

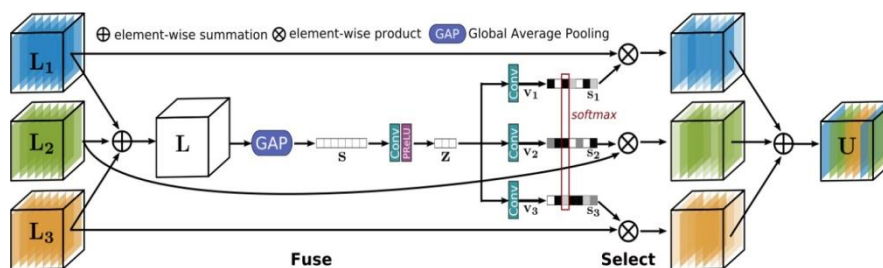


Fig. 2: Schematic for selective kernel feature fusion (SKFF).

It operates on features from multiple convolutional streams, and performs aggregation based on self-attention. or summation. However, these choices provide limited expressive power to the network, as reported in. In MRB, we introduce a nonlinear procedure for fusing features coming from multiple resolutions using a self-attention mechanism. Motivated by, we call it selective kernel feature fusion (SKFF). The SKFF module performs dynamic adjustment of receptive fields via two operations –Fuse and Select. The fuse operator generates global feature descriptors by combining the information from multi-resolution streams. The select operator uses these descriptors to recalibrate the feature maps (of different streams) followed by their aggregation. Next, we provide details of both operators for the three-stream case, but one can easily extend it to more streams. (1) Fuse: SKFF receives inputs from three parallel convolution streams carrying different scales of information. We first combine these multiscale features using an element-wise sum as:  $L = L_1 + L_2 + L_3$ . We then apply global average pooling (GAP) across the spatial dimension of  $L \in \mathbb{R}^{H \times W \times C}$  to compute channel-wise statistics  $s \in \mathbb{R}^{1 \times 1 \times C}$ . Next, we apply a channel downscaling convolution layer to generate a compact feature representation  $z \in \mathbb{R}^{1 \times 1 \times r}$ , where  $r = C/8$  for all our experiments. Finally, the feature vector  $z$  passes through three parallel channel-upscaling convolution layers (one for each resolution stream) and provides us with three feature descriptors  $v_1, v_2$  and  $v_3$ , each with dimensions  $1 \times 1 \times C$ . (2) Select: this operator applies the softmax function to  $v_1, v_2$  and  $v_3$ , yielding attention activations  $s_1, s_2$  and  $s_3$  that we use to adaptively recalibrate multi-scale feature maps  $L_1, L_2$  and  $L_3$ , respectively. The overall process of feature recalibration and aggregation is defined as:  $U = s_1 \cdot L_1 + s_2 \cdot L_2 + s_3 \cdot L_3$ . Note that the SKFF uses  $\sim 6\times$  fewer parameters than aggregation with concatenation but generates more favorable results (an ablation study is provided in the experiments section).

### Dual Attention Unit

The Dual Attention Unit or DAU is used to extract features in the convolutional streams. While the SKFF block fuses information across multi-resolution branches, we also need a mechanism to share information within a feature tensor, both along the spatial and the channel dimensions which is done by the DAU block. The DAU suppresses less useful features and only allows more informative ones to pass further. This feature recalibration is achieved by using Channel Attention and Spatial Attention mechanisms.

The Channel Attention branch exploits the inter-channel relationships of the convolutional feature maps by applying squeeze and excitation operations. Given a feature map, the squeeze operation applies Global Average Pooling across spatial dimensions to encode global context, thus yielding a feature descriptor. The excitation operator passes this feature descriptor through two convolutional layers followed by the sigmoid gating and generates activations. Finally, the output of Channel Attention branch is obtained by rescaling the input feature map with the output activations.

The Spatial Attention branch is designed to exploit the inter-spatial dependencies of convolutional features. The goal of Spatial Attention is to generate a spatial attention map and use it to recalibrate the incoming features. To generate the spatial attention map, the Spatial Attention branch first independently applies Global Average Pooling and Max Pooling operations on input features along the channel dimensions and concatenates the outputs to form a resultant feature map which is then passed through a convolution and sigmoid activation to obtain the spatial attention map. This spatial attention map is then used to rescale the input feature map.

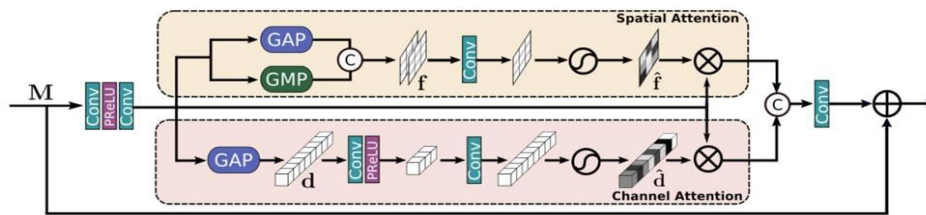


Fig. 3: Schematic for Dual Attention Unit (DAU).

### Multi-Scale Residual Block

The Multi-Scale Residual Block is capable of generating a spatially-precise output by maintaining high-resolution representations, while receiving rich contextual information from low-resolutions. The MRB consists of multiple (three in this paper) fully-convolutional streams connected in parallel. It allows information exchange across parallel streams in order to consolidate the high-resolution features with the help of low-resolution features, and vice versa. The MIRNet employs a recursive residual design (with skip connections) to ease the flow of information during the learning process. In order to maintain the residual nature of our architecture, residual resizing modules are used to perform downsampling and upsampling operations that are used in the Multi-scale Residual Block.

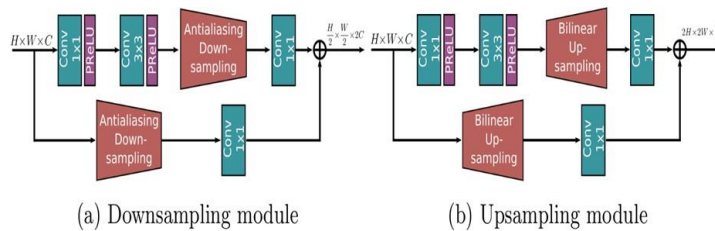


Fig. 4: Schematic for Multi-Scale Residual Block (MSR).

### WORKING:

when low-light, blurred, low quality images are given as input. The proposed architecture is end-to-end trainable and requires no pre-training of sub-modules. We train three different networks for three different restoration tasks. The training parameters, common to all experiments, are the following. We use 3 RRGs, each of which further contains 2 MRBs. The MRB consists of 3 parallel streams with channel dimensions of 64, 128, 256 at resolutions 1, 12,14, respectively. Each stream has 2 DAUs. The models are trained with the Adam optimizer ( $\beta_1 = 0.9$ , and  $\beta_2 = 0.999$ ) for  $7 \times 10^5$  iterations. The initial learning rate is set to  $2 \times 10^{-4}$ . Once trained, MIRNet can be applied to enhance low-light images by passing them through the network. The multiple iterations in the reconstruction network allow the model to refine the image progressively, resulting in improved visibility and enhanced details.

### RESULT

After running the code, we get output as show below.

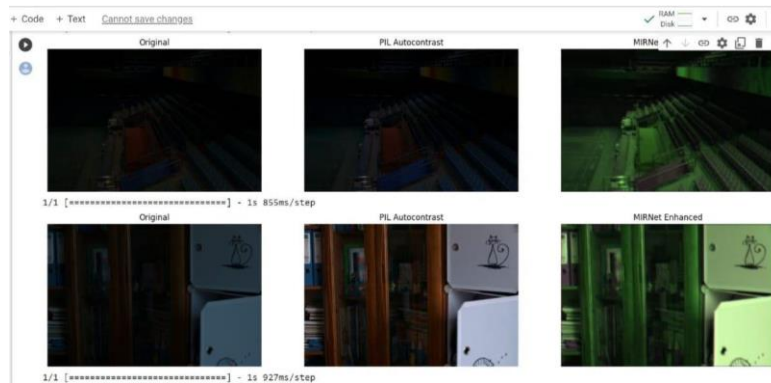


Fig: Result-1

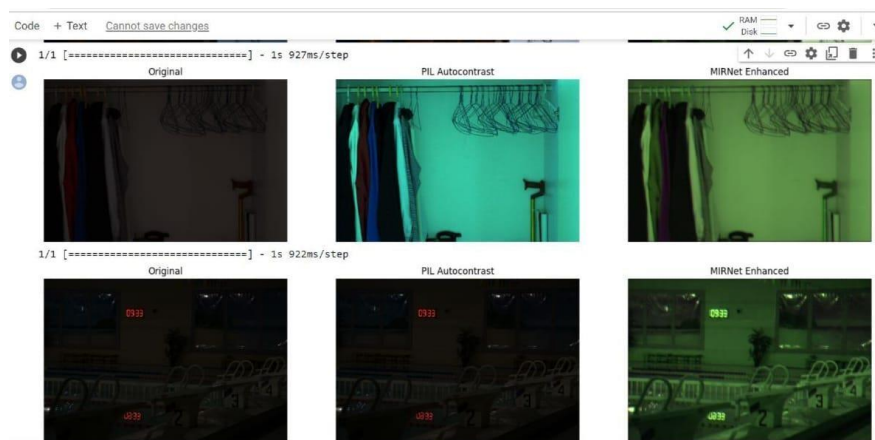


Fig: Result-2

**GRAPHICAL REPRESENTATION:**



Fig: Result -3(Train and validation losses over images)

Above shown Fig: Result-3 tells how much the information is lost, while training the images and and validating the images.

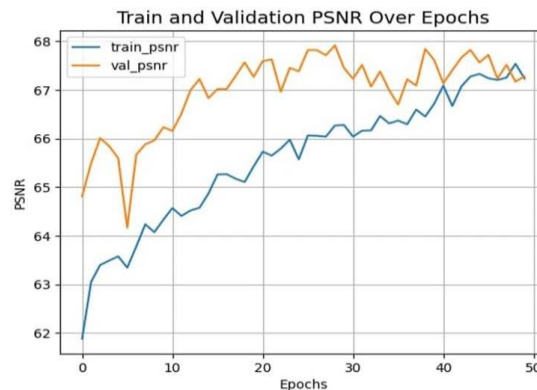


Fig: Result-4(Train and validation PSNR over images)

Above shown Fig: Result-4 gives the PSNR ratio of the images while training and validating.

## CONCLUSION

Conventional image restoration and enhancement pipelines either stick to the full resolution features along the network hierarchy or use an encoder-decoder architecture. The first approach helps retain precise spatial details, while the latter one provides better contextualized representations. However, these methods can satisfy only one of the above two requirements, although real-world image restoration tasks demand a combination of both conditioned on the given input sample. The main branch is dedicated to full-resolution processing and the complementary set of parallel branches provides better contextualized features. We propose novel mechanisms to learn relationships between features within each branch as well as across multiscale branches. Our feature fusion strategy ensures that the receptive field can be dynamically adapted without sacrificing the original feature details. Consistent achievement of state-of-the-art results on five datasets for three image restoration and enhancement tasks corroborates the effectiveness of our approach.

## REFERENCES

1. Abdelhamed, A., Lin, S., Brown, M.S.: A high-quality denoising dataset for smartphone cameras. In: CVPR (2018) 8, 9, 10
2. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *Trans. Sig. Proc.* (2006) 9
3. Suganthi, S.; Umapathi, N.; Mahdal, M.; Ramachandran, M. Multi Swarm Optimization Based Clustering with Tabu Search in Wireless Sensor Network. *Sensors* **2022**, 22, 1736. <https://doi.org/10.3390/s22051736>
4. Ahn, N., Kang, B., Sohn, K.A.: Fast, accurate, and lightweight super-resolution with cascading residual network. In: ECCV (2018) 4.
5. N.Umapathi., N.Ramaraj., (2016) Wireless adhoc telemedicine system: proving networking performance for multimedia data. *Journal of medical imaging and health informatics* 6(8), 1944-1948.
6. Allebach, J., Wong, P.W.: Edge-directed interpolation. In: ICIP (1996) 3
7. Anwar, S., Barnes, N.: Real image denoising with feature attention.
8. Anwar, S., Khan, S., Barnes, N.: A deep journey into super-resolution
9. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *TPAMI* (2017) 5

- 10.N.Umapathi., N.Ramaraj, (2014) Swarm Intelligence based dynamic source routing for improved quality of service, Journal of Theoretical and Applied Information Technology, Pp 604-608.
- 11.Bertalm'io, M., Caselles, V., Provenzi, E., Rizzi, A.: Perceptual color correction through variational techniques. TIP (2007) 4
- 12.Brooks, T., Mildenhall, B., Xue, T., Chen, J., Sharlet, D., Barron, J.T.: Unprocessing images for learned raw denoising. In: CVPR (2019) 3
- 13.Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising.
- 14.A M Viswa Bharathy ; N Umapathi ; S Prabakaran.,(2019) “An Elaborate Comprehensive Survey on Recent Developments in Behaviour Based Intrusion Detection Systems” **IEEE** International Conference on Computational Intelligence in Data Science (ICCIDS)., held on 21-23 Feb. 2019. 10.1109/ICCIDS.2019.8862119
- 15.Burger, H.C., Schuler, C.J., Harmeling, S.: Image denoising: Can plain neural networks compete with BM3D In: CVPR (2012) 3, 9
- 16.Bychkovsky, V., Paris, S., Chan, E., Durand, F.: Learning photographic global tonal adjustment with a database of input/output image pairs.
- 17.Cai, J., Gu, S., Timofte, R., Zhang, L.: Ntire 2019 challenge on real image superresolution: Methods and results. In: CVPRW (2019) 3
- 18.Cai, J., Zeng, H., Yong, H., Cao, Z., Zhang, L.: Toward real-world single image super-resolution: A new benchmark and a new model.
- 19.Chang, H., Yeung, D.Y., Xiong, Y.: Super-resolution through neighbor embedding.