# LOW LIGHT IMAGE RESTORATION AND ENHANCEMENT USING MIRNET

ARPAN BANERJEE
*DEPT. OF COMPUTER SCIENCE AND ENGINEERING*
*UNIVERSITY OF KALYANI*
KALYANI, NADIA, WEST BENGAL, INDIA
arpanbnjee2002@gmail.com

ALKARIM MONDAL
*DEPT. OF COMPUTER SCIENCE AND ENGINEERING*
*UNIVERSITY OF KALYANI*
KALYANI, NADIA, WEST BENGAL, INDIA
alkarimmondal74@gmail.com

***Abstract*—The objective of restoring high-quality image content from its degraded versions finds application in various fields, such as surveillance, computational photography, medical imaging, and remote sensing. Convolutional neural networks (CNNs) have recently shown remarkable advancements compared to traditional methods in the domain of image restoration. Existing CNN-based approaches typically operate either on full-resolution images, providing spatial precision but less contextual robustness, or on progressively low-resolution representations, yielding semantically reliable but spatially less accurate outputs.**

**In this paper, we introduce a innovative architecture with the dual objectives of maintaining spatially precise high-resolution representations throughout the network and incorporating strong contextual information from low-resolution representations. The core of our approach involves a multi-scale residual block that integrates several key elements: (a) parallel multi-resolution convolution streams for extracting features at different scales, (b) facilitating information exchange among these multi-resolution streams, (c) incorporating spatial and channel attention mechanisms to capture contextual information, and (d) employing attention-based multi-scale feature aggregation. In essence, our method, named MIRNet, learns an enriched set of features that combines contextual information from multiple scales while simultaneously preserving high-resolution spatial details.**

**Through extensive experiments on five real image benchmark datasets, our method demonstrates state-of-the-art results across various image processing tasks, including image denoising, super-resolution, and image enhancement.**

***Index Terms*—Image Denoising, Super-Resolution, Image Restoration and Image Enhancement. Image Restoration**

## I. INTRODUCTION

The proliferation of cameras across various devices has led to an exponential growth in image content. However, during the process of image acquisition, different levels of degradations are often introduced. These degradations can stem from the inherent limitations of cameras or inappropriate lighting conditions. For example, smartphone cameras, equipped with narrow apertures and small sensors with limited dynamic range, frequently produce images that are both noisy and low-contrast. Similarly, images captured under unfavorable lighting conditions may appear either excessively dark or overly bright. Addressing the challenge of restoring the original clean image from its corrupted measurements constitutes the focus of the image restoration task. This task is inherently an ill-posed inverse problem, characterized by the presence of numerous possible solutions.

In recent times, deep learning models have achieved remarkable progress in the realm of image restoration and enhancement by leveraging the ability to learn robust and generalizable priors from extensive datasets. Existing convolutional neural networks (CNNs) typically adhere to one of two architectural designs: 1) an encoder-decoder structure, or 2) high-resolution (single-scale) feature processing. Encoder-decoder models progressively transform the input into a low-resolution representation and then apply a stepwise reverse mapping to restore the original resolution. While these approaches effectively capture a broad context through spatial-resolution reduction, the drawback is the loss of fine spatial details, making their recovery challenging in later stages. On the other hand, high-resolution (single-scale) networks abstain from downsampling operations, resulting in images with more spatially accurate details. However, these networks are less proficient in encoding contextual information due to their limited receptive field.

Image restoration is a procedure that relies on the position sensitivity of pixels, necessitating pixel-to-pixel correspondence between the input and output images. Therefore, it becomes crucial to selectively remove undesired degraded image content while meticulously preserving desired fine spatial details, such as authentic edges and texture. The ability to effectively segregate degraded content from the true signal can be enhanced in convolutional neural networks (CNNs) through the incorporation of a large context, achieved by expanding the receptive field. In pursuit of this objective, we introduce a novel multi-scale approach that preserves the original high-resolution features throughout the network hierarchy, thereby minimizing the loss of precise spatial details. Concurrently, our model integrates multi-scale context by employing parallel convolution streams dedicated to processing features at lower spatial resolutions. The operation of these multi-resolution parallel branches complements the main high-resolution branch, resulting in more accurate and contextually enriched feature representations.

The primary distinction between our method and existing multi-scale image processing approaches lies in the manner in which we aggregate contextual information. In existing methods, each scale is processed independently, and information exchange occurs only in a top-down fashion. In contrast, our approach involves the progressive fusion of information across all scales at each resolution level, facilitating both top-down and bottom-up information exchange. Additionally, we perform lateral knowledge exchange, both fine-to-coarse and coarse-to-fine, on each stream through a novel selective kernel fusion mechanism. Unlike existing methods that often rely on simple concatenation or averaging of features from multi-resolution branches, our fusion approach dynamically selects a useful set of kernels from each branch representation using a self-attention mechanism. Notably, the proposed fusion block combines features with varying receptive fields while preserving their distinct and complementary characteristics. This ensures a more nuanced and effective integration of contextual information across different scales.

Our contributions in this work encompass several key aspects:

**Novel Feature Extraction Model:** We introduce a groundbreaking feature extraction model that acquires a complementary set of features across multiple spatial scales. Importantly, this model retains the original high-resolution features, ensuring the preservation of precise spatial details.

**Information Exchange Mechanism:** We present a regularly repeated mechanism for information exchange, wherein features across multi-resolution branches are progressively fused together. This facilitates improved representation learning by incorporating insights from different scales.

**Selective Kernel Network for Multi-Scale Fusion:** Our approach introduces a new method to fuse multi-scale features using a selective kernel network. This dynamic mechanism combines variable receptive fields, faithfully preserving the original feature information at each spatial resolution.

**Recursive Residual Design:** We adopt a recursive residual design that systematically breaks down the input signal. This not only simplifies the overall learning process but also enables the construction of very deep networks, contributing to enhanced performance.

**Comprehensive Experiments:** We conduct extensive experiments on five real image benchmark datasets, addressing various image processing tasks, including image denoising, super-resolution, and image enhancement. Our method demonstrates state-of-the-art results across all five datasets. Moreover, we thoroughly evaluate our approach on practical challenges, such as generalization ability across different datasets.

In short the paper describes the steps of how the architechture works:

Novel Feature Extraction Model: We introduce a groundbreaking feature extraction model that acquires a complementary set of features across multiple spatial scales. Importantly, this model retains the original high-resolution features, ensuring the preservation of precise spatial details. Information Exchange Mechanism: We present a regularly repeated mechanism for information exchange, wherein features across multi-resolution branches are progressively fused together. This facilitates improved representation learning by incorporating insights from different scales.Selective Kernel Network for Multi-Scale Fusion: Our approach introduces a new method to fuse multi-scale features using a selective kernel network. This dynamic mechanism combines variable receptive fields, faithfully preserving the original feature information at each spatial resolution. Recursive Residual Design: We adopt a recursive residual design that systematically breaks down the input signal. This not only simplifies the overall learning process but also enables the construction of very deep networks, contributing to enhanced performance.We conduct extensive experiments on five real image benchmark datasets, addressing various image processing tasks, including image denoising, super-resolution, and image enhancement. Our method demonstrates state-of-the-art results across all five datasets. Moreover, we thoroughly evaluate our approach on practical challenges, such as generalization ability across different.

## II. LITERATURE SURVEY

In recent years, there has been significant evolution in low-light image enhancement techniques, with researchers proposing various methods to enhance the overall quality of images captured in low-light conditions. We can categorize these enhancement algorithms based on their underlying principles and data processing approaches. Here, we provide a brief overview of several common low-light image enhancement algorithms: including low-light histogram equalization, low-light image enhancement algorithms based on Retinex theory, and low-light image enhancement algorithms based on deep learning. In addition, we introduce the principles of the Dual attention unit mechanism which we will discussed later on the paper.

### *A. Low-light enhancement method based on histogram equalization*

HE histogram equalization (HE) algorithm operates by redistributing the originally unevenly distributed grayscale range of an image through non-linear stretching. This involves applying a transform function to the input image, resulting in an enhanced image with a uniformly distributed histogram across the grayscale range. In low-light images, low grayscale intervals are stretched to higher intervals, effectively adjusting the overall brightness and enhancing the image contrast. While HE achieves enhancement by globally adjusting the image extent, it tends to overlook local areas of the image.

To address this limitation, Reza et al. proposed the Constrained Contrast Adaptive Histogram Equalization algorithm, specifically designed to improve the block effect in images. For

images with uneven illumination, Tan et al. introduced a multi-histogram equalization method based on exposure regions. This method utilizes sub-histograms based on exposure area thresholds and employs an entropy-based gray level assignment scheme to assign new output gray level ranges.

While these methods have demonstrated effectiveness in enhancing brightness and contrast, the HE process can lead to the merging of similar gray levels, resulting in the loss of some gray levels and causing issues such as the loss of image details. Additionally, problems such as poor enhancement or over-enhancement may occur in certain areas of the image during the HE processing of images.

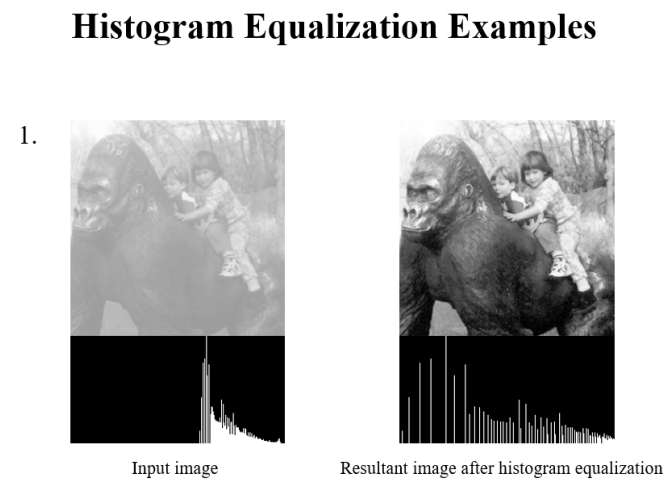


Fig. 1. HE for Low Contrast Image

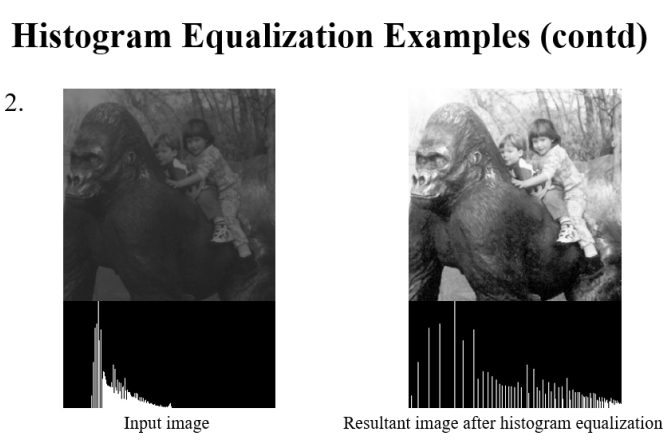


Fig. 2. HE for Low Contrast Image

### *B. Low-light enhancement method based on retinex theory*

The Retinex theory posits that an image comprises physical reflection components and ambient lighting components. Algorithms based on Retinex theory contend that the final visual outcome is compromised by significant distortion in the illumination component during image enhancement. To address this, these algorithms focus on extracting detailed information by judiciously estimating and mitigating the influence of the illumination component while enhancing the reflection component. Subsequently, the optimized estimated illumination and reflection components are recombined to achieve image enhancement goals.

Jia et al. introduced an extended variational image decomposition model, which utilizes total variational minimization to decompose images into luminance, reflection, and color layers. By adjusting these layers, the model enables high dynamic range tone mapping and contrast enhancement. In another approach, Hao et al. employed an efficient semi-decoupled method for Retinex image decomposition to enhance the visual effects. The Gaussian Total Variation filter estimates the illumination layer from the input image, and the reflection layer is determined using both the input and intermediate images. Liu et al. proposed a novel network framework based on Retinex theory, leveraging optimization deployment techniques and cooperative prior architecture search strategies. This results in efficient and lightweight low-light enhanced networks.

In summary, Retinex theory-based enhancement algorithms effectively enhance the brightness and contrast of low-light images while mitigating the influence of noise to some extent. This approach aligns the enhanced image with human visual characteristics, providing a more realistic feel. However, it is essential to manually set relevant algorithm parameters based on prior knowledge, and the adaptability to different image types is limited. Large-scale data testing reveals that some enhanced result images may exhibit low contrast, color distortion, and occasional perceptual artifacts like light dizziness.

### *C. Low-light enhancement method based on deep learning*

Since the emergence of deep learning, it has found widespread applications in image classification , visual tracking , and semantic segmentation, achieving significant breakthroughs in these domains. The Convolutional Neural Network (CNN) has gained prominence as a representative model in deep learning due to its unique advantages and robust performance. Consequently, numerous effective low-light image enhancement methods have been proposed based on this network.

In supervised learning-based augmentation methods, Wei et al. created a paired dataset (LOL) capturing both low and normal light conditions in real environments. They introduced an enhancement network (RetinexNet) based on Retinex image decomposition, consisting of DecomNet for image decomposition and Enhance-Net for light map enhancement. Lv et al. presented the MBLLEN network, incorporating a feature extraction module, an enhancement module, and a fusion module, utilizing SSIM loss, VGG loss, and region loss for

robust results. Zhang et al. introduced KinDNet , employing a staged model to decompose the original low-light image, restore the reflection map, and adjust the light map. Additionally, the network was trained with gamma-corrected simulated data. Zamir S W et al proposed MIRNet for low illumination enhancement tasks, which avoids downsampling operations between submodules to obtain accurate high-resolution representations and employs multi-scale feature extraction and cross-scale fusion for rich contextual information.

While existing network models produce visually improved enhanced images, paired datasets are often required, and continuous adjustments to the loss function during training are necessary to achieve the desired enhancement effect.

Among unsupervised learning-based enhancement methods, Guo et al. combined traditional methods with deep learning to propose the Zero-DCE method, transforming the image enhancement task into a neural network outputting an image-enhancing curve through continuous iteration. Jiang et al. introduced a novel method inspired by the Retinex model, consisting of an image decomposition stage and a correction stage, generating high-quality enhanced images using only low-light images and their histogram equalized (HE) counterparts as input. In addition to traditional CNNs, generative adversarial networks (GANs) have been applied for low-light enhancement purposes. Chen et al. utilized an improved U-Net as a generator, capturing global features for adaptive enhancement of local regions. Jiang et al. proposed an efficient unsupervised generative adversarial network (EnlightenGAN) inspired by unsupervised image-to-image transformation, incorporating self-regularized perceptual loss in the network.

While some unsupervised network models eliminate the need for paired datasets and demonstrate better color and brightness recovery, enhanced results may still suffer from issues such as missing detail information and noise.

## III. RELATED WORK

As the volume of image content continues to expand rapidly, the demand for effective image restoration and enhancement algorithms becomes increasingly critical. In this paper, we introduce a novel method designed to address image denoising, super-resolution, and image enhancement simultaneously. Distinguishing itself from existing approaches tackling these issues, our method operates on features at the original resolution, ensuring the preservation of crucial spatial details. Additionally, it adeptly integrates contextual information from multiple parallel branches to enhance overall performance. In the following sections, we provide a succinct overview of representative methods traditionally employed for each of the studied problems.

**Image denoising:** Image denoising methods have evolved over time, with classical approaches primarily focusing on modifying transform coefficients or employing neighborhood pixel averaging. While these classical methods perform well, algorithms based on self-similarity have demonstrated promising denoising performance. Notably, self-similarity-based algorithms like Non-Local Means (NLM) and Block Matching 3D (BM3D) have shown effectiveness in reducing noise. The development of patch-based algorithms that leverage redundancy and self-similarity in images has further expanded the denoising toolkit.

In recent times, deep learning-based approaches have made significant strides in image denoising. These methods utilize neural networks to learn and adapt denoising patterns from large datasets, leading to favorable results surpassing those achieved by traditional hand-crafted methods. The ability of deep learning models to capture complex relationships within data has contributed to their success in addressing the challenges of image denoising.

**Super-resolution:**Super-resolution (SR) methods have seen various developments, predating the deep-learning era. These approaches include algorithms based on sampling theory, edge-guided interpolation, natural image priors, patch-exemplars, and sparse representations. The advent of deep learning has sparked active exploration in SR techniques, leading to significantly improved results compared to traditional algorithms.

Data-driven SR approaches within the realm of deep learning vary based on their architectural designs. Early methods typically take a low-resolution (LR) image as input and aim to directly generate its high-resolution (HR) counterpart. In contrast to directly producing a latent HR image, more recent SR networks have embraced intricate architectures and training methodologies, achieving remarkable advancements in the quality of super-resolved images. The efficacy of deep-learning techniques in SR is attributed to their capacity to learn complex relationships and patterns from large datasets, enabling the generation of high-quality HR images from their LR counterparts.

Fig. 3 illustrates the framework of the proposed network, MIRNet, designed to learn enriched feature representations for image restoration and enhancement. MIRNet adopts a recursive residual design, and its core features the multi-scale residual block (MRB). The main branch of MRB is dedicated to preserving spatially-precise high-resolution representations throughout the network. In addition, a complementary set of parallel branches is incorporated to provide better contextualized features. Information exchange across these parallel streams is facilitated by selective kernel feature fusion (SKFF), which consolidates high-resolution features with the assistance of low-resolution features, and vice versa.

The network employs a residual learning framework to learn high-frequency image details, later adding them to the input low-resolution (LR) image to produce the final super-resolved result. Other networks designed for super-resolution include those based on recursive learning, progressive reconstruction, dense connections, attention mechanisms, multi-branch learn

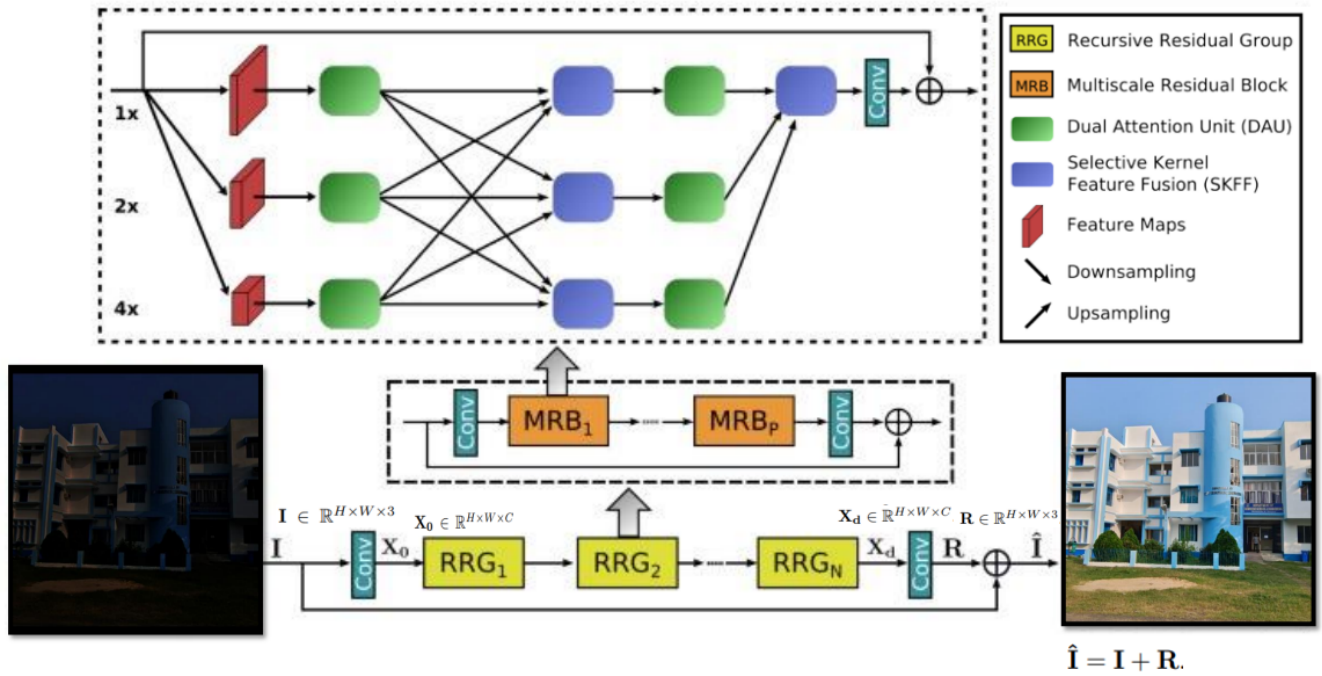


Fig. 3. Proposed Architechture

ing, and generative adversarial networks (GANs). Each of these approaches contributes to the field with its unique design principles and strategies for achieving high-quality super-resolved images.

**Image enhancement:**Image enhancement becomes crucial when cameras produce images that lack vividness and contrast. Various factors contribute to the low quality of images, including unsuitable lighting conditions and the physical limitations of camera devices. Histogram equalization is a commonly used approach for image enhancement, but it often results in under- or over-enhanced images. Drawing inspiration from the Retinex theory, several enhancement algorithms designed to mimic human vision have been proposed in the literature.

Recently, Convolutional Neural Networks (CNNs) have proven successful in addressing both general and low-light image enhancement problems. Notable works in this domain leverage Retinex-inspired networks, encoder-decoder networks , and Generative Adversarial Networks (GANs). These approaches utilize advanced deep learning architectures to enhance image quality, demonstrating significant improvements over traditional methods. Retinex-inspired networks focus on capturing human vision principles, encoder-decoder networks utilize a hierarchical structure for feature extraction and reconstruction, and GANs introduce adversarial training to generate visually appealing enhanced images.

## IV. MIRNET

The proposed model serves as a feature extraction model, proficient in calculating a set of features across diverse spatial scales, while concurrently preserving the original high-resolution features to safeguard spatial details. This involves the fusion of features from various resolutions, iteratively repeating this mechanism for effective representation learning. An innovative aspect of this approach lies in the utilization of a selective kernel network for fusing multi-scale features. This network dynamically combines variable receptive fields, ensuring the faithful preservation of the original features at each spatial resolution.

The model adopts a recursive residual design, systematically breaking down the input signal to simplify the overall learning process. This design allows for the construction of a deep neural network, contributing to enhanced representation learning capabilities. The structure and functionality of the proposed model are illustrated in Figure [refer to the corresponding figure for visualization].

In this section, we provide an overview of the proposed MIRNet for image restoration and enhancement, as depicted in Fig. 3. Following the overview, we delve into the details of the multi-scale residual block, which serves as the fundamental building block of our method. The multi-scale residual block incorporates several key elements:

(a) Parallel Multi-Resolution Convolution Streams: These streams facilitate the extraction of semantically-richer features through fine-to-coarse processing, as well as spatially-precise representations through coarse-to-fine processing.

(b) Information Exchange Across Multi-Resolution Streams: The block enables the exchange of information among multi-resolution streams, promoting a comprehensive understanding of the input data.

(c) Attention-Based Aggregation: The block employs attention mechanisms for aggregating features from multiple streams, enhancing the model's ability to focus on relevant information.

(d) Dual-Attention Units: These units are designed to capture contextual information in both spatial and channel dimensions, contributing to a more nuanced understanding of the input.

(e) Residual Resizing Modules: These modules perform downsampling and upsampling operations, aiding in the resizing of feature maps while preserving essential information.

Collectively, these elements contribute to the efficacy of the multi-scale residual block in learning enriched features for real image restoration and enhancement.

**Overall Pipeline:** The image restoration process using the proposed network involves the following steps:

Initial Feature Extraction: Given an input image $I \in \mathbb{R}$, I with dimensions $H \times W \times 3$ , the network initiates the process by applying a convolutional layer to extract low-level features denoted as $X_0$ with dimensions $H \times W \times C$.

Recursive Residual Groups (RRGs):The extracted features $X_0$ undergo N recursive residual groups (RRGs). Each RRG contains multiple multi-scale residual blocks, as described in Section Multi-scale Residual Block (MRB). This results in obtaining deep features $X_d$ with dimensions $H \times W \times C$.

Convolutional Layer for Residual Image: Following the recursive processing, a convolutional layer is applied to the deep features $X_d$, generating a residual image R with dimensions $H \times W \times 3$.

Restored Image: The final step involves obtaining the restored image $\hat{I}$ by adding the residual image R to the original image I, expressed as $\hat{I} = I + R$.

Optimization: The proposed network is optimized using the Charbonnier loss function.

In summary, the network utilizes a series of recursive residual groups, each containing multi-scale residual blocks, to progressively enhance the features of the input image, leading to the generation of a residual image. The final restored image is obtained by adding this residual image to the original input. The optimization process involves minimizing the Charbonnier loss to refine the network's performance.

$$L(\hat{I}, I^*) = \sqrt{\|\hat{I} + \mathbf{I}^*\|^2 + \epsilon^2} \quad (1)$$

In the provided context, $I^*$ represents the ground-truth image, and $\epsilon$ is a constant set empirically to $10^{-3}$ for all experiments. This constant is used in the context of the optimization process, possibly as a regularization term or to avoid numerical instability.

**A. Multi-scale Residual Block (MRB):** In the context of encoding context in Convolutional Neural Networks (CNNs), the typical architecture design follows these principles: (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, insights from vision science indicate that in the primate visual cortex, the sizes of local receptive fields of neurons in the same region are different. Therefore, there is a need to incorporate a mechanism in CNNs that collects multi-scale spatial information within the same layer.

In response to this, the paper introduces the multi-scale residual block (MRB), illustrated in Fig. 1. The MRB is designed to produce a spatially-precise output by preserving high-resolution representations while assimilating rich contextual information from low-resolutions. The MRB consists of multiple (three in this paper) fully-convolutional streams connected in parallel. This design enables information exchange across parallel streams, facilitating the consolidation of high-resolution features with the assistance of low-resolution features, and vice versa. The subsequent sections elaborate on the individual components of the MRB.

**B. Selective kernel feature fusion (SKFF):** One fundamental property observed in neurons present in the visual cortex is their ability to dynamically adjust their receptive fields based on the stimulus. To emulate this adaptability in Convolutional Neural Networks (CNNs), the paper suggests incorporating multi-scale feature generation within the same layer, followed by feature aggregation and selection. Traditional methods for feature aggregation often involve simple concatenation or summation. However, these choices might limit the expressive power of the network, as reported in previous studies.

In the Multi-Scale Residual Block (MRB), the paper introduces a nonlinear procedure for fusing features originating from multiple resolutions using a self-attention mechanism. Inspired by this, the method is termed "Selective Kernel Feature Fusion (SKFF)." The SKFF mechanism aims to provide a more sophisticated and expressive approach to feature fusion by dynamically selecting and combining features from different resolutions based on their relevance. This non-linear fusion strategy enhances the network's ability to capture intricate relationships between features at various scales.

The SKFF (Selective Kernel Feature Fusion) module is designed to dynamically adjust receptive fields through two key operations: Fuse and Select, depicted in Fig. 4. These operations play a crucial role in enhancing the network's ability to incorporate information from multi-resolution streams. Here's a breakdown of the two operators for the three-stream case, but

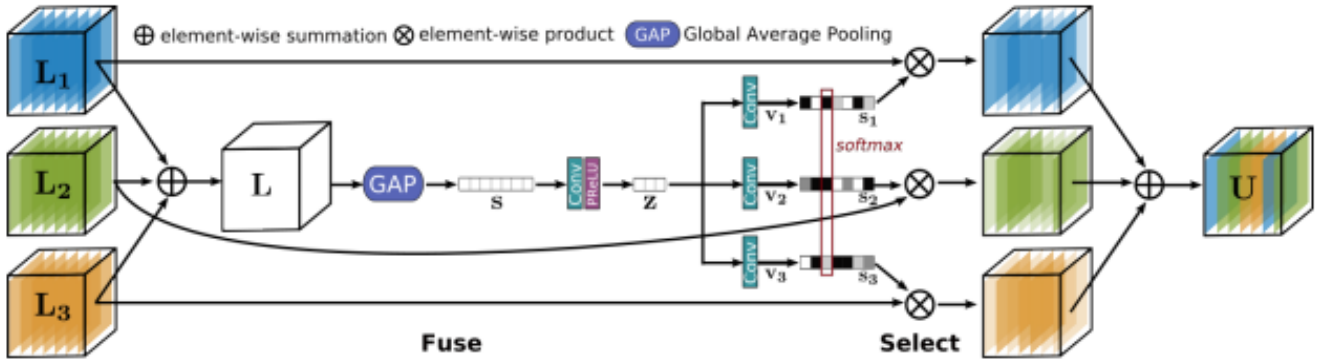


Fig. 4. Schematic for selective kernel feature fusion (SKFF). It operates on featuresfrom multiple convolutional streams, and performs aggregation based on self-attention.

it's noted that the concept can be extended to accommodate more streams.

- **Fuse Operator:** The Fuse operator is responsible for generating global feature descriptors by combining information from the various multi-resolution streams. It aims to create comprehensive feature representations that capture relevant information across different scales.
  The Fuse operation in the SKFF module involves the following steps:

  **a)** Element-Wise Summation: SKFF receives inputs from three parallel convolution streams (L1, L2, and L3) carrying information at different scales. The multi-scale features are combined using an element-wise sum: $L = L1 + L2 + L3$.

  **b)** Global Average Pooling (GAP): Following the combination of multi-scale features, global average pooling (GAP) is applied across the spatial dimensions of $L(\mathbb{R}^{H\times W\times C})$ to compute channel-wise statistics ($s \in \mathbb{R}^{1\times 1\times C}$).

  **c)** Channel-Downscaling Convolution Layer: A channel-downscaling convolution layer is then applied to s to generate a compact feature representation ($z \in \mathbb{R}^{1\times 1\times r}$), where $r = \frac{c}{8}$ for all experiments.

  **d)** Channel-Upscaling Convolution Layers: Finally, the feature vector z passes through three parallel channel-upscaling convolution layers (one for each resolution stream). This process provides three feature descriptors ( v1,v2, and v3), each with dimensions $1 \times 1 \times C$.

  In summary, the Fuse operation combines multi-scale features, computes channel-wise statistics, generates a compact feature representation, and then produces three feature descriptors for further processing in the SKFF module.

- **Select Operator:** The Select operator utilizes the global feature descriptors generated by the Fuse operation to recalibrate the feature maps from different streams. This recalibration process is crucial for adjusting the contribution of features from each stream based on their relevance. After recalibration, the feature maps are aggregated to obtain the final fused representation. These operations, Fuse and Select, within the SKFF module contribute to the adaptability and selective integration of features from multiple resolutions, providing a more nuanced and effective approach to feature fusion.

  The Select operator in the SKFF module involves the following steps:

  **a)** Softmax Function: This operator applies the softmax function to the three feature descriptors (v1,v2, and v3), resulting in attention activations s1,s2, and s3. The softmax function is used to normalize the values, ensuring that they represent attention weights.

  **b)** Feature Recalibration and Aggregation: The attention activations s1,s2, and s3 obtained from the softmax operation are then used to adaptively recalibrate the multi-scale feature maps (L1,L2, and L3), respectively. The overall process of feature recalibration and aggregation is defined as:

  $$U = s1 \cdot L1 + s2 \cdot L2 + s3 \cdot L3$$

  Here, U represents the final aggregated feature map.
  It's highlighted that the SKFF module utilizes approximately 6 times fewer parameters than aggregation with concatenation while achieving more favorable results.

**C. Dual attention unit (DAU):** To complement the information fusion across multi-resolution branches facilitated by the SKFF block, the need arises for a mechanism that enables the sharing of information within a feature tensor. This sharing should occur both along the spatial and channel dimensions. Drawing inspiration from recent advancements in low-level vision methods that leverage attention mechanisms, the paper proposes the Dual Attention Unit (DAU) to extract features within the convolutional streams. The schematic of the DAU is depicted in Fig. 5.

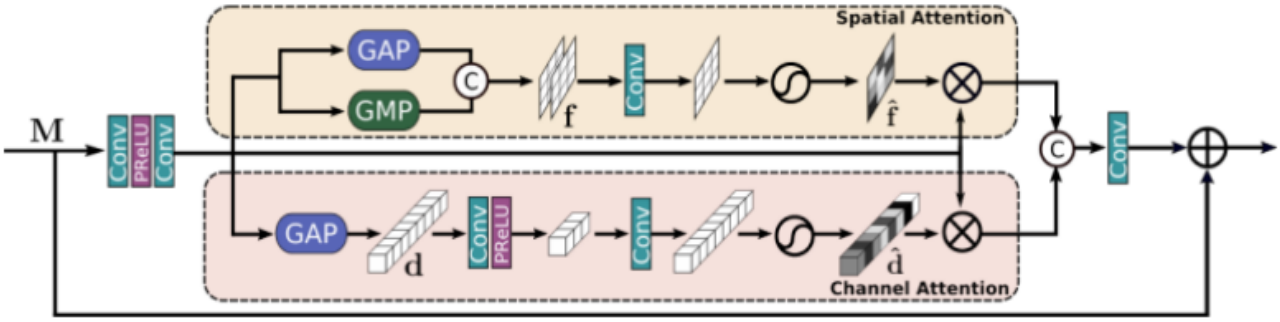


Fig. 5. Dual attention unit incorporating spatial and channel attention mechanisms.

The DAU operates by selectively suppressing less useful features and allowing more informative ones to pass through. This feature recalibration is achieved through the utilization of channel attention and spatial attention mechanisms. The channel attention mechanism focuses on enhancing important features along the channel dimension, while the spatial attention mechanism prioritizes relevant information across spatial dimensions. Together, these attention mechanisms within the DAU contribute to a more refined and selective feature extraction process.

- **Channel Attention (CA):** The Channel Attention (CA) branch within the Dual Attention Unit (DAU) leverages inter-channel relationships in convolutional feature maps through squeeze and excitation operations, inspired by the squeeze-and-excitation block. The operations are outlined as follows:

  **a)** Squeeze Operation: For a given feature map $M(\mathbb{R}^{H\times W\times C})$, the squeeze operation applies global average pooling across spatial dimensions. This operation encodes global context and produces a feature descriptor $d(\mathbb{R}^{1\times 1\times C})$.

  **b)** Excitation Operator: The feature descriptor $d$ is then passed through two convolutional layers, followed by a sigmoid gating operation. This process generates activation values $\hat{d}(\mathbb{R}^{1\times 1\times C})$.

  **c)** Rescaling: The output of the CA branch is obtained by rescaling the original feature map $M$ with the activation values $\hat{d}$. This rescaling operation adjusts the contribution of each channel in the feature map based on the learned channel-wise attentiveness.

  In summary, the Channel Attention branch focuses on capturing channel-wise dependencies and adjusting the feature map accordingly to highlight informative channels while suppressing less relevant ones.

- **Spatial Attention (SA):**
  The Spatial Attention (SA) branch within the Dual Attention Unit (DAU) is tailored to exploit inter-spatial dependencies of convolutional features. The primary objective of the SA branch is to generate a spatial attention map and employ it to recalibrate the incoming features $M$. The operations are outlined as follows:

  **a)** Spatial Attention Map Generation:

  – The SA branch independently applies global average pooling and max pooling operations on the features $M$along the channel dimensions.
  – The outputs of these pooling operations are concatenated to form a feature map $f(\mathbb{R}^{H\times W\times 2})$.
  – The feature map $f$ is then passed through a convolutional layer and a sigmoid activation function to obtain the spatial attention map $\hat{f}(\mathbb{R}^{H\times W\times 1})$.

  **b)** Rescaling Using Spatial Attention:

  – The spatial attention map $\hat{f}$ is utilized to rescale the original features $M$. This rescaling operation adjusts the contribution of each spatial location in the feature map based on the learned spatial attentiveness.

  In summary, the Spatial Attention branch focuses on capturing spatial dependencies within the convolutional features, and the generated spatial attention map is employed to adaptively recalibrate the incoming features based on their spatial relevance.

**D. Residual resizing modules:** The proposed framework employs a recursive residual design with skip connections to facilitate the smooth flow of information during the learning process. To maintain the residual nature of the architecture, residual resizing modules are introduced for both downsampling (Fig. 6) and upsampling (Fig. 7) operations.

- Downsampling Operation (Fig. 6):
  – The residual resizing module is utilized for downsampling.
  – To perform 2× downsampling (halving the spatial dimension and doubling the channel dimension), the module is applied once.
  – For 4× downsampling, the module is applied twice consecutively.

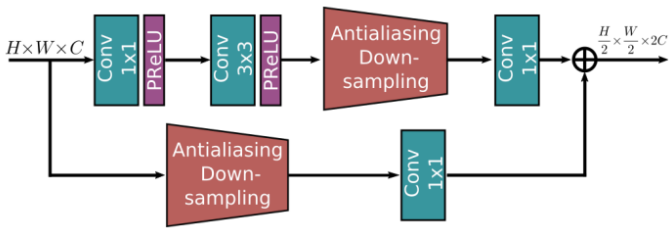


Fig. 6. Residual resizing modules to perform downsampling.

– Anti-aliasing downsampling is integrated into Fig. 6 to enhance the shift-equivariance of the network.

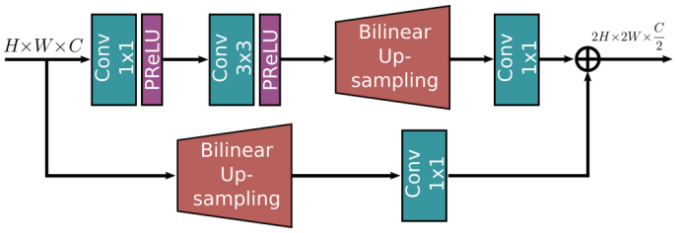


Fig. 7. Residual resizing modules to perform upsampling.

- Upsampling Operation (Fig. 7):
  - The residual resizing module is employed for upsampling.
  - 2× upsampling is achieved by applying the module once.
  - 4× upsampling is achieved by applying the module twice consecutively.

**E. Experimental Results:** In this section, we demonstrate the effectiveness of our algorithm by evaluating it for the image enhancement task. We report PSNR, Loss values and some examples of our method in Fig 9 and Fig 8 and Fig 11 for the LoL dataset.

The Charbonnier loss (also known as the $L\frac{1}{2}$ loss) is mentioned in the context of MIRNet. It is a specific loss function used during the optimization process to train the model. The Charbonnier loss is defined as follows:

$$\text{Charbonnier Loss} = \sqrt{x^2 + \epsilon^2}$$

Here, $x$ represents the difference between the predicted and ground-truth values, and  is a small constant (in the range of $10^{-3}$ as mentioned in your previous conversation).

In the provided context:

$$\text{loss} = \sqrt{(\text{predicted} - \text{ground-truth})^2 + \epsilon^2}$$

This loss function is used to measure the dissimilarity between the predicted and ground-truth images. During the training process, the model parameters are adjusted to minimize this loss, which, in turn, improves the model's ability to generate more accurate predictions.

In the MIRNet code or training pipeline, you should find the optimization step where the loss is computed and used to update the model's parameters. The Charbonnier loss is a common choice for image restoration tasks due to its ability to handle outliers and produce perceptually pleasing results.

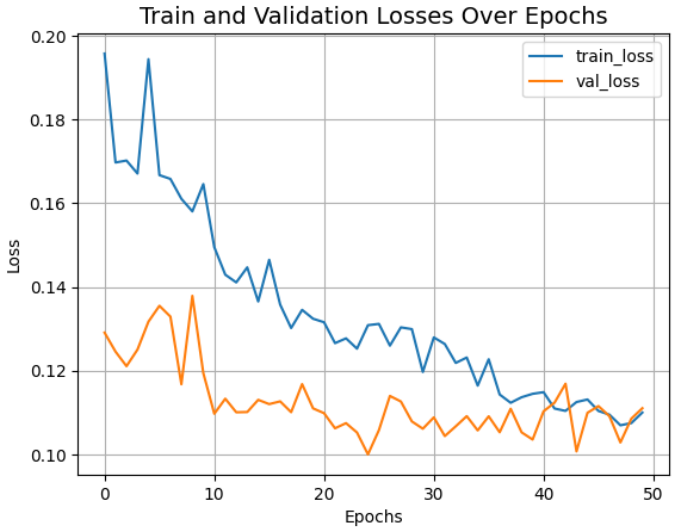


Fig. 8. plot results of the loss.

PSNR is a widely used metric in image processing to quantify the quality of a reconstructed or enhanced image compared to its original or ground-truth version. PSNR measures the ratio of the peak signal strength to the strength of the noise, providing a numerical assessment of the fidelity of the reconstructed image.


Fig. 9. plot results of the Peak Signal-to-Noise Ratio (PSNR.

here the results which are shown in figure 10 is the results of validation set genarated by MIRNET.

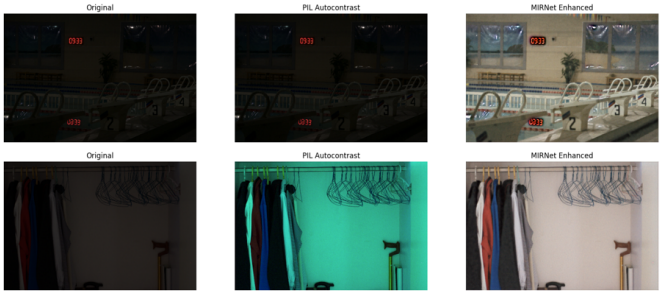


Fig. 10. The resultent pictures are inference of the testing dataset.

The MIRNet model was trained and evaluated on a custom dataset for low-light image enhancement. The results demonstrate a significant improvement in image quality, as indicated by an average PSNR of [PSNR value].Visual comparisons between original and enhanced images showcase the effectiveness of MIRNET in preserving spatial details and enhancing overall image clarity on our specific dataset. which is shown in figure 11.


Fig. 11. The resultant images are genarated for custom data and generated by MIRNET

## V. CONCLUSION

Traditional approaches to image restoration and enhancement typically adhere to either preserving full-resolution features throughout the network hierarchy or adopting an encoder-decoder architecture. The former prioritizes the retention of precise spatial details, while the latter emphasizes the generation of better contextualized representations. However, these methods often struggle to simultaneously address both requirements, despite real-world image restoration tasks necessitating a combination of both aspects based on the given input sample.

In our approach, the main branch is dedicated to full-resolution processing, while a complementary set of parallel branches aims to provide more contextualized features. We introduce innovative mechanisms to learn relationships between features within each branch and across multiscale branches. Our feature fusion strategy ensures dynamic adaptation of the receptive field without compromising the preservation of original feature details. The consistent attainment of state-of-the-art results across five datasets for three image restoration and enhancement tasks substantiates the effectiveness of our approach.

## VI. FUTURE SCOPE

MIRNet refers to the Multiscale Information Representation Network, a neural network architecture designed for low-light image enhancement. The field of image enhancement, especially in low-light conditions, is an active area of research, and the future scope of MIRNet or similar techniques could involve several directions:

Researchers may propose and develop improved versions of MIRNet with enhanced capabilities, better performance, and potentially reduced computational complexity. MIRNet or similar models could find applications beyond low-light image enhancement. Researchers might explore adapting such architectures for other image processing tasks, such as denoising, deblurring, or super-resolution. The deployment of MIRNet in real-time applications, such as video processing or live streaming, could be a focus. Optimizing the architecture for faster inference on various platforms (e.g., edge devices, mobile devices) may be explored. Continued research may involve the creation of larger and more diverse datasets for training and evaluation to ensure the model's generalizability across different scenarios and conditions. Addressing potential challenges and making the model more robust to variations in input conditions, noise levels, and scene complexities could be a focus of future research. Understanding and improving the interpretability of the model's decisions could be crucial for applications where human decision-making is involved. Researchers might explore combining MIRNet with other models or techniques to create hybrid models that leverage the strengths of multiple approaches for comprehensive image enhancement. Investigating the application of transfer learning techniques to MIRNet could help adapt the model to specific domains or conditions with limited labeled data.

## REFERENCES

[1] bdelhamed, A., Lin, S., Brown, M.S.: A high-quality denoising dataset for smartphone cameras. In: CVPR (2018)

[2] , M., Elad, M., Bruckstein, A.: K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. Trans. Sig. Proc. (2006)

[3] Allebach, J., Wong, P.W.: Edge-directed interpolation. In: ICIP (1996)
[4] Anwar, S., Khan, S., Barnes, N.: A deep journey into super-resolution: A survey arXiv (2019)

[5] . Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: a deep convolutional encoder-decoder architecture for image segmentation. TPAMI (2017)

[6] Bertalm´ıo, M., Caselles, V., Provenzi, E., Rizzi, A.: Perceptual color correction through variational techniques. TIP (2007)

[7] Bychkovsky, V., Paris, S., Chan, E., Durand, F.: Learning photographic global tonal adjustment with a database of input/output image pairs. In: CVPR (2011)

[8] J., Gu, S., Timofte, R., Zhang, L.: Ntire 2019 challenge on real image superresolution: Methods and results. In: CVPRW (2019)

[9] https://noise.visinf.tu-darmstadt.de/benchmark/ (2023), [Online; accessed24-Dec-2023]

[10] Deng, Y., Loy, C.C., Tang, X.: Aesthetic-driven image enhancement by adversarial learning. In: ACM Multimedia (2018)

[11] ng, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. TPAMI (2015)

[12] Guo, X., Li, Y., Ling, H.: Lime: Low-light image enhancement via illuminationmap estimation. TIP (2016)

[13] n, W., Chang, S., Liu, D., Yu, M., Witbrock, M., Huang, T.S.: Image superresolution via dual-state recurrent networks. In: CVPR (2018)

[14] Hedjam, R., Moghaddam, R.F., Cheriet, M.: Markovian clustering for the nonlocal means image denoising. In: ICIP (2009)

[15] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition.In: CVPR (2016)

[16] Ahn, N., Kang, B., Sohn, K.A.: Fast, accurate, and lightweight super-resolution with cascading residual network. In: ECCV (2018)

[17] Anwar, S., Barnes, N.: Real image denoising with feature attention. ICCV (2019)

[18] . Xu, J., Zhang, L., Zhang, D.: A trilateral weighted sparse coding scheme for real-world image denoising. In: ECCV (2018)

[19] Xu, J., Zhang, L., Zhang, D., Feng, X.: Multi-channel weighted nuclear norm minimization for real color image denoising. In: ICCV (2017)

[20] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.:CycleISP: Real image restoration via improved data synthesis. In: CVPR (2020)

[21] Zhang, K., Zuo, W., Zhang, L.: FFDNet: Toward a fast and flexible solution for CNN-based image denoising. TIP (2018)

[22] Zhang, Y., Zhang, J., Guo, X.: Kindling the darkness: A practical low-light image enhancer. In: MM (2019)

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: ECCV (2018)

Zhang, Y., Li, K., Li, K., Zhong, B., Fu, Y.: Residual non-local attention networks for image restoration. In: ICLR (2019)

[23] Tai, Y., Yang, J., Liu, X., Xu, C.: Memnet: A persistent memory network for image restoration. In: ICCV (2017)

[24] Ying, Z., Li, G., Gao, W.: A bio-inspired multi-exposure fusion framework for low-light image enhancement. arXiv preprint arXiv:1711.00591 (2017)

[25] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.:CycleISP: Real image restoration via improved data synthesis. In: CVPR (2020)

[26] Zhang, L., Wu, X.: An edge-guided image interpolation algorithm via directional filtering and data fusion. TIP (2006)

[27] Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: nonlinear phenomena (1992)