Low-Light Image Enhancement using Keras MIRnet

Team Members:

SATHIYAMATHI V	20MIS1159
NITHISH KUMAR A N	20MIS1091
EYOHESHWARAN V	20MIS1030

ABSTRACT:

Enhancement of low-light images is a challenging task due to the impact of low brightness, low contrast, and high noise. The inability to collect natural labeled data intensifies this problem further. Many researchers have attempted to solve this problem using learning-based approaches; however, most models ignore the impact of noise in low-lit images. In this paper, an encoder-decoder architecture, made up of separable convolution layers that solve the issues encountered in low-light image enhancement, is proposed. The architecture is trained end-to-end on a custom low-light image dataset (LID), comprising both clean and noisy images. We introduce a unique multi-context feature extraction module (MC-FEM) where the input first passes through a feature pyramid of dilated separable convolutions for hierarchical-context feature extraction followed by separable convolutions for feature compression. The model is optimized using a novel three-part loss function that focuses on high-level contextual features, structural similarity, and patch-wise local information. We conducted several ablation studies to determine the optimal model for low-light image enhancement under noisy and noiseless conditions. We have used performance metrics like peak-signal-to-noise ratio, structural similarity index matrix, visual information fidelity, and average brightness to demonstrate the superiority of the proposed work against the state-of-the-art algorithms. Qualitative results presented in this paper prove the strength and suitability of our model for real-time applications.

INDEX TERMS:

Encoder-decoder architecture, separable convolution, dilated convolution, ASPP, perceptual loss, low-light image enhancement.

I. INTRODUCTION

Low light image enhancement is an active area of research that enables the acquisition system to capture superior quality images even under low-light conditions. Its applications include autonomous driving, photography, military, object detection, and surveillance. Low-light image enhancement (LIE) algorithms consider several factors like color contrast, brightness, image resolution, and dynamic range. Several researchers use image processing techniques to enhance low-light images, such as histogram equalization (HE) [1] that tend to equalize the dynamic range of intensities. Also, histogram equalization methods focus only on increasing the image contrast, whereas they fail to address the actual illumination issues. As mentioned in [2], linear and non-linear low-light image enhancement methods have simple yet fast implementation but don't consider the image distribution, leading to limited enhancement ability. Spatial filters and other image processing techniques have shown improvement in the low-lit image quality, but these methods fail to work under noisy context. Noise, typically, is amplified by the filters, as they rely on a small neighborhood. These reasons justify the need for deep-learning in enhancing low-light images, as these methods can dynamically learn and handle noisy as well as clean images.

The associate editor coordinating the review of this manuscript and approving it for publication was Varuna De Silva.

A. LIE ALGORITHMS USING RETINEX

One category of LIE algorithms utilizes Retinex theory formulated by Edward H. Land, which accounts for color constancy and human perception [3], [4]. This theory assumes that the image intensity is a product of reflectance and illumination coefficients. Algorithms based on Retinex theory calculate the illumination map by removing the reflectance component and utilizing it for enhancing the image. Methods like single-scale and multi-scale using a surround function [5], [6]. However, these methods have a trade-off between dynamic range compression and rendition factors. Adaptive weighting based MSR that combines color constancy with local enhancement to transform the images is suggested, in [7]. Low-light image enhancement (LIME) [8], is another Retinex based enhancement technique that estimates the illumination of each pixel by finding the strongest response among the red, green, and blue channels. A structure prior is also imposed on the illumination map to refine the initial map into a well constructed, enhanced map. Another method [9], which supersedes LIME, uses a low-light enhancement component followed by noise removal. The enhancement component takes care of luminance estimation and image restoration, but it does not yield superior results, as noise in the image still prevails.

B. LIE ALGORITHMS USING DEEP LEARNING

Deep learning has yielded promising results in image processing tasks such as denoising, de-hazing, super-resolution, and other computer vision problems. One such deep learning network for low-light enhancement is the LLNet, which enhances the images using a stacked auto-encoder without convolutional layers [10]. LLCNN avoids the vanishing gradient problem using a module to extract the multi-scale feature maps [11]. In, multi-branch low-light enhancement network (MBLLEN) [13], enhancements applied at multiple subnets are fused to generate the desired output. A unique loss function involving structural, contextual, and regional information aided MBLLEN to produce superior results in noise-free images. GLADNet is another CNN based model that employs an encoder-decoder for extracting the global, prior information about the illumination and a CNN for detail reconstruction [14].

Attention guided low-light image enhancement proposed in [15] uses two attention maps for low-light enhancement. The first map distinguishes the under-exposed from well-exposed ones, and the second map identifies real textures from noise. A novel four-part loss function, including attention and enhancement loss, is also introduced in this work, thus covering all factors required for enhancing the low-light images. However, its performance degrades on images with large black regions and compressed images.

Apart from CNN based approaches, Generative Adversarial Networks (GANs) [16] have also become very popular in image processing [17]–[20]. EnlightenGAN [21] is a highly effective unsupervised GAN trained without low or regular lightimagepairs.ItenhancestheimagebyemployingaUNet, followed by two discriminators. The performance of this Retinex (MSR) aim to estimate illumination from the image algorithm in terms of quantitative metrics is inferior though its visual perception is superior.

C. LIE ALGORITHMS USING RETINEX AND DEEPLEARNING

Most techniques ignore the possibility of the noise, and in some cases, they even amplify the presence of noise.
This gap led to the evolution of low-light image enhancement algorithms based on deep learning fused with the Retinex theory. In [22], a deep Retinex-Net, comprising of decomposition and the enhancement net, is proposed. Here, a decomposition net decomposes the image into reflectance and illumination, followed by an enhancement network for adjusting the illumination component. But this method fails to improve the color, contrast, and brightness of the image. Recently, a CNN based progressive Retinex model, to

address the noise and the over-illumination issues, is proposed in [23]. Two point-wise convolutional networks are employed to determine the regularities behind light and noise. Though progressive Retinex facilitates noise suppression, it fails to capture the structural properties in the image. An approach for combining Retinex with GANs is proposed, in [24]. Here, regularization loss helps to prevent the local-optimal solution. Retinex-based theory using a two-stage model for low-light enhancement, employing a separate stage for denoising, is proposed in [25]. In [26], a pseudo-Retinex based method, using a hybrid network to combine content and edge features in two paths, is proposed to learn the global context. This method also uses RNNs, thus making it a complex network.

D. SEPARABLE AND DILATED CONVOLUTION

Most models mentioned above have a common characteristic that they use convolutional layers. But, the computational complexity of complex architecture using convolutional layers is quite expensive. In general, deep networks for low-light enhancement require more layers and are highly complex. For example, an architecture like attention-guided network [15] outperforms simpler models like LLCNN [11]. Separable convolution instead of the traditional convolutional layers helps to achieve a trade-off between computational complexity and performance.

As observed in [12], separable convolutions can reduce the number of parametersrequiredper layer. A simplified form of U-Net, using convolutional layers, can have up to 30 million parameters, whereas the same architecture with separable convolution requires 5 million parameters only. We have also observed that most of the existing deep architectures for low-light image enhancement are not context-sensitive. The inclusion of dilated convolutions using atrous spatial pyramid pooling (ASPP) for semantic segmentation has enabled networks to learn context-sensitive information [27]. For an input x, the output y of atrous convolution using the filter w is given by [27],

$$y_{(i,j)} = \sum_{s,t,u}^{S,T,U} x[i + r \cdot s, j + r \cdot t, u] w[s,t,u]$$

Here r is the dilation rate. where as the traditional convolution is given by,

$$y_{(i,j)} = \sum_{s,t,u}^{S,T,U} w[s,t,u] \cdot x(i + s, j + t, u)$$

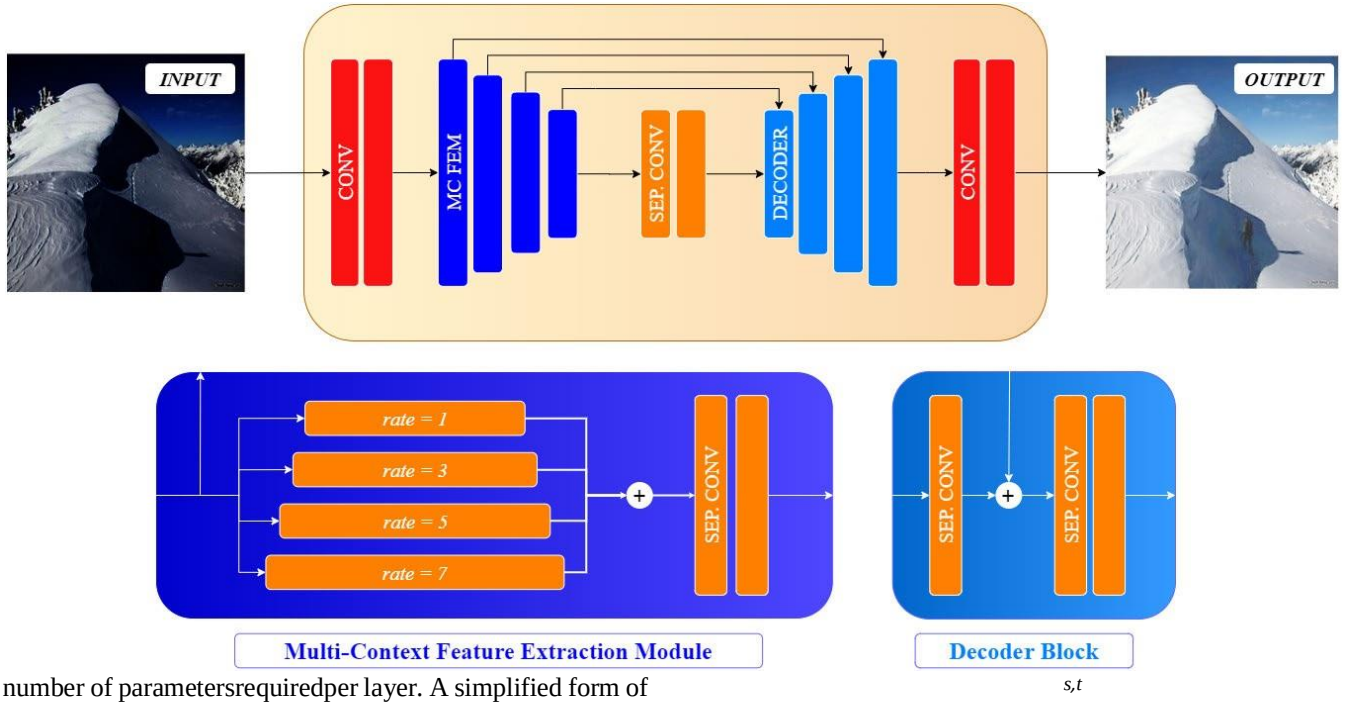

FIGURE 1. Illustrative Diagram of the Context LIE-Net Architecture.

Mathematical formulation for separable convolution is as follows:

$$PointwiseConv(w,x)_{(i,j)} = \sum_{u}^{U} w_u \cdot x_{(i,j,u)}$$

$$DepthwiseConv(w,x)_{(i,j)} = \sum_{s,t}^{S,T} w(s,t)\ x_{(i+s,j+t)}$$

$$SepConv(w_p, w_d, y)_{(i,j)} = PointwiseConv_{(i,j)}(w_p, DepthwiseConv_{(i,j)}(w_d, x)) \quad (1)$$

In the proposed work, dilated, separable convolutional layers, using ASPP capture the context-sensitive information. Simultaneously, it reduces the computational complexity of the network. We have also proposed a novel loss function that addresses the issues mentioned above and enhances the low-light images. The organization of the paper is as follows: Section II elaborates on the proposed context-LIE architecture and the proposed three-part loss function, in detail. In section III, we have discussed the motivation behind the dataset used in our experiments. The quantitative analysis of various experiments and visual observations is presented in Section IV. Section V sums up with the conclusion and further extensions.

II. PROPOSED METHOD

In this work, we have proposed an architecture based on the standard encoder-decoder that is a natural choice for low-light image enhancement. The encoder extracts enhancement related information such asbrightness, contrast, or noise, whereas the decoder utilizes the features required to enhance the image. Success of UNet [28] and UNet-like architectures [13], [14], [21] for low-light image enhancement further validated our choice. In this section, we have presented a detailed description of the proposed context LIE-Net architecture along with the implementation details. Our proposed unique three-part loss function, specific for low-light image enhancement, is discussed, with its impact on the overall learning is highlighted through qualitative analysis.

A. CONTEXT LIE-NET ARCHITECTURE

In Fig.1, an illustrative view of the proposed context LIE-Net architecture, is presented. At the encoder, instead of the conventional convolution layers, we have used a multi-context feature extraction module (MC-FEM). At the decoder, we have utilized depth-wise separable convolution to improve the generalization ability of the network. Generalization is the vital aspect of applications like low-light image enhancement that lack naturally labeled data. Speed up is achieved by performing convolution on the channels separately and then concatenating the intermediate output, thereby resulting in fewer parameters than the traditional convolution layers, making the network less prone to overfitting. Skip connections are employed from the encoder to their corresponding decoder blocks to facilitate the reconstruction of features and image information otherwise lost during the encoding or down-sampling stages. Convolution layers in the beginning and at the end help the network to capture complex mapping between the image and its features.

1) MULTI-CONTEXT FEATURE EXTRACTION MODULE (MC-FEM)

The proposed MC-FEM extracts multi-contextual features followed by feature compression. We have used parallel dilated depth-wise separable convolutions to extract the multi-context features using the feature pyramid, inspired by Atrous Spatial Pyramid Pooling (ASPP) block. Feature pyramid is constructed by concatenating the output of all the parallel dilated depth-wise separable layers. These parallel-layers have progressive wide-contexts owing to the increasing dilation rates. The dilation rate is varied from 1 to 7 in increments of 2 with a fixed filter size, 3 × 3. These characteristics allow the network to learn multi-contextual or spatial-hierarchical features with minimal increase in the network complexity. Feature compression achieved using the stack of separable convolution layers takes a concatenated feature pyramid as input and discards redundant information accumulated in the feature pyramid. The combined effect of these two operations enables MC-FEM to extract richer and more relevant features from a spatially-wide neighbourhood.

Specifically, this module is useful in noisy context due to its ability to capture a wide-spatial neighbourhood, using the stacked dilated layers that have increasing dilation rates, thus capturing a larger area. Outputs from these modules are sent as skip connections to later parts of the model to enhance reconstruction. Particularly under noisy context, the skip connections prove to be very useful as it combines the neighbourhood from encoder module output with the corresponding decoder block.

B. LOSS FUNCTION

In this section, a three-part loss function proposed to include vital factors of low-light image enhancement, such as noise removal, feature, and structure preservation is presented. Our novel loss function, can be represented as:

$$L_{Total} = w_{per}L_{per} + w_{ssim}L_{ssim} + w_{wpw}L_{wpw} \quad (2)$$

In Eq. (2), L_{per}, L_{ssim} and L_{wpw} corresponds to perceptual loss, loss based on structural similarity and weighted patch-wise Euclidean loss, and w_{per}, w_{ssim}, and w_{wpw} correspond to their respective weights.

1) PERCEPTUAL LOSS

Per-pixel based loss functions consider two similar images as different even when they differ by one-pixel intensity. As a result, per-pixel based loss functions like Euclidean fail to capture the difference in high-level features between the ground-truth and the predicted image effectively [29]. The difference in high-level features, extracted using pre-trained CNNs, could be minimized using a perceptual loss function. It leads us to the first component in Eq. 2, which extracts the high-level features between the enhanced image and the ground truth from the convolution layers of the VGG16 network pre-trained on the ImageNet dataset. It enables the network to extract a mixture of low, mid, and high-level features. Subsequently, the average mean square error difference between these features determine the perceptual loss. The proposed perceptual loss can be computed as,

$$L_{Per} = \frac{1}{MN \times (2n+1-1)} \sum_{k=0}^{n} \sum_{i=1}^{M} \sum_{j=1}^{N} 2^{k} (Q_k(Y(i,j)) - Q_k(\hat{Y}(i,j)))^2 \quad (3)$$

Here, Y and $\hat{Y}$ denotes the $M \times N$ ground truth and predicted images, respectively. n is the number of blocks consideredintheVGGnetwork,and$n$issetto3toavoidoverfittin g. $Q_k(:)$ denotes the output of the k^{th} block of VGG16 and k = 0 refers to the input. The difference in outputs of the later VGG16 blocks is weighted more by scaling the difference using 2^k.

For image enhancement applications, the ground truth serves as a guiding factor for the network to learn the enhanced image as there is no defined, exact output. Perceptual loss enhances the hierarchical features to minimize the difference between the predicted image and the ground truth. Perceptual loss has also shown benefits in applications apart from low light enhancement like image inpainting [44].

2) STRUCTURAL LOSS

From [30], we observed that illumination does not affect the structural information in a scene. The structural similarity index matrix (SSIM) compares the similarity between two images in terms of luminance, contrast, and structure. The structural information remains intact independent of the changes in the luminance component or contrast. The SSIM loss is thus included as a part of the proposed loss function as the human visual system is more susceptible to capture the structural information from images.

The SSIM loss component is given by

$$L_{ssim} = 1 - \frac{(2\mu_y\mu_{\hat{y}} + C_1) + (2\sigma_{y\hat{y}} + C_2)}{(\mu_{2y} + \mu_{\hat{y}}^2 + C_1)(\sigma_y^2 + \sigma_{\hat{y}}^2 + {}_{y\ y}\ C_2)} \quad (4)$$

where μ_y, $\mu_{\hat{y}}$ denote pixel mean, σ_y, $\sigma_{\hat{y}}$ denote the variance, $\sigma_{y\hat{y}}$ refers to the covariance between the ground truth, Y and the predicted image, $\hat{Y}$ respectively. Here, C_1 and C_2 are constants to prevent divide by zero error.

3) WEIGHTED PATCH-WISE EUCLIDEAN LOSS

Euclidean loss [23] and $L1$−norm [14] are the most commonly used loss functions in low-light image enhancement. MBLLEN [13] improved the Euclidean based loss function by providing more weightage to the lower intensity levels by utilizing the 25^{th} percentile of the histogram as a threshold. However, MBLLEN fails to capture the contextual information in the image.

We have proposed a weighted patch-wise (WPW), Euclidean loss function as a replacement to the traditional Euclidean loss. In this method, Euclidean loss is computed by segmenting the predicted image into patches and assigning larger weights to those patches with lower average intensities. Patches with a lower mean average in the predicted image signify that they are low lit even after enhancement. Thus, the low-light image enhancer is accurate as these patches are closer to the ground truth. We have sorted the mean average intensity of the patches in increasing order and took the n^{th} percentile to determine the threshold, T. The weighted patch-wise loss function is normalized to avoid its dominance in the total loss function in Eq. (2).

TABLE 1. Optimal Parameter Setting Selection.

Parameter	Min	Max	Step Size	Optimal Value
α	0	1	0.25	0.25
p_x/p_y	8	64	x2	16
w	1	16	x2	4
n	1	6	1	3
w_{per}	0	2	0.25	1
w_{ssim}	0	2	0.25	1
w_{wpw}	0	2	0.1	0.1

The weighted patch-wise loss function is given by:

$$L_{wpw} = \frac{1}{\lambda MN}\left(w \times \sum_{\forall \mu_{\hat{p}} \le T} ||\hat{P}(i,j) - P(i,j)||_2^2 + \sum_{\forall \mu_{\hat{p}} > T} ||\hat{P}(i,j) - P(i,j)||_2^2\right) \quad (5)$$

Here, the weight w, is used to assign more weightage to the patches with lower average intensities. Patches, in the ground truth and predicted image, are represented as $P(:)$ and $\hat{P}(:)$ respectively. $\mu_{\hat{p}}$ denotes the average intensity of the patch, and T denotes the average threshold intensity. The normalization factor λ is given by,

$$\lambda = p_x p_y [\alpha(w-1) + 1] * MN \quad (6)$$

where p_x and p_y denote the number of patches along the rows and columns, respectively. α denotes the percentage of total patches considered for adding weights.

C. IMPLEMENTATION DETAILS

The proposed encoder design has two convolutional layers and four MC-FEM blocks. The input to the MC-FEM blocks is down-sampled using 2×2 max-pooling layer. The convolutional layers and the first MC-FEM block uses 64 filters each. The no. of filters in the subsequent blocks is doubled, making the fourth MC-FEM block with 512 filters. The decoder network has separable convolution blocks

interleaved with up-sampling layers. The no. of filters is halved after each upsampling layer, making the last separable layer and the two subsequent convolutional layers with 64 filters. For an image of size 256 × 256, various loss parameters are set as follows: $\alpha = 0.25$, $p_x = p_y = 16$, $w = 4$,$n = 3$, $w_{per} = 1$, $w_{ssim} = 1$, and $w_{wpw} = 0.1$. Table 1 describes the different experimental parameters along with its minimum and maximum values, as well as the step size, used to decide the optimal value for each.

III. DATASET

Collecting natural low-light scenes along with their corresponding normal/bright image pair is a challenging task. Several authors collected a set of raw images to generate a syntheticdataset [13]–[15],[22].However,mostofthesedatasets consider images taken in a carefully monitored environment, and therefore they do not ideally represent the natural scene. Moreover, it is challenging to generate a synthetic dataset as there is no consensus over the low-light image generation method. Some works have used software like Adobe Photoshop Lightroom [14], while others have devised more elaborateimageselectionandtransformationtechniques[15].

We utilized the MIT outdoor and indoor scenes dataset [31] to generate synthetic, low-light images. Gamma correction, used in [13], is incorporated to simulate low illumination effects. We have used 5000 images from the MIT dataset to generate a low-light image dataset (LID). LID comprises both clean and noisy images to simulate real-world scenarios. Using Matlab, Poisson-Gaussian Noise [15], [32] is added to simulate noise in real-time. Initially, random Gaussian noise, whose variance varied between 0 and 0.005, is added to the low-light image, followed by Poisson noise.

We have used 4500 and 500 image pairs for training and testing, respectively. Fig. 2 portrays few sample pairs from LID. The top row consists of synthetic, low lit images, and the row beneath consists of their corresponding ground truth. Two images from the right are noisy, and two images from the left are noiseless.

IV. EXPERIMENTAL RESULTS

This section summarizes the experiments conducted to verify the quantitative and qualitative performance of synthetic and real low-light images, respectively. We conducted separate experiments to evaluate the performance of the proposed Context LIE-Net on clean as well as noisy images from LID. As mentioned before, we have used 500 image pairs from the LID dataset for testing and comparative analysis. The performance is evaluated using commonly used metrics like Peak Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM) [30], Visual Information Fidelity (VIF) [34] and Average Brightness (AB) [33]. The model is trained with 4500 image pairs using a batch size, 8, and Adam optimizer with a learning-rate of $1e{-}4$. The proposed model converges in 50 epochs under these specifications.

The experimental analysis is subdivided into four sections: ablation on architecture that demonstrates its efficiency under noisy and noiseless conditions, ablation on the loss function that validates the three-part loss, a quantitative and qualitative comparison using synthetic LID, and qualitative comparison using natural low-light images.

Technology used: The Technology we have used for training and testing our project are TensorFlow and keras, Mirenet.

A. ABLATION STUDY ON ARCHITECTURE

The use of dilated convolutions in a pyramid structure alongside separable convolutions is the first of its kind for low-light image enhancement. We conducted experiments on architecture ablation to verify the impact of dilated and separable

	Model	PSNR	SSIM	VIF	AB
Clean	Input	9.144	0.2276	0.255	81.063
	U-Net [28]	22.281	0.7564	0.382	9.249
	U-Net w/ Sep. Conv.	22.367	0.7609	0.371	8.664
	C-LIENet w/o Skip Conn.	15.712	0.3610	0.046	10.551
	C-LIENet w/ Skip Conn.	**22.786**	**0.7821**	**0.395**	**8.422**
Noisy	Input	9.318	0.2202	0.149	78.391
	U-Net [28]	19.215	0.5879	0.241	10.538
	U-Net w/ Sep. Conv.	19.405	0.5975	0.235	9.912
	C-LIENet w/o Skip Conn.	14.952	0.3647	0.053	15.835
	C-LIENet with Skip Conn.	**19.886**	**0.6139**	**0.241**	**9.648**

	Loss Function	PSNR	SSIM	VIF	AB
Clean	Input	9.144	0.2276	0.255	81.063
	SSIM + Weighted Euclidean	23.812	0.815	**0.464**	7.082
	Weighted Euclidean + Perceptual	23.834	0.804	0.421	7.222
	SSIM + Perceptual	24.238	0.819	0.456	7.366
	3 Part loss	**24.438**	**0.820**	0.459	**6.894**
Noisy	Input	9.318	0.2202	0.149	78.391
	SSIM + Weighted Euclidean	20.490	0.655	0.266	7.474
	Weighted Euclidean + Perceptual	20.658	0.623	**0.276**	7.572
	SSIM + Perceptual	20.913	0.663	0.261	7.945
	3 Part loss .	**21.152**	**0.674**	0.273	**7.302**

convolutionlayers.AsseeninTable2,thesignificanceofseparable convolution on U-Net and the proposed architecture, named project, is depicted clearly. All these architectures are trained, for the same number of epochs, using the widely used loss function, the mean squared error [23].

From Table 2, we can see that the use of separable convolutions in U-Net improves the performance in terms of PSNR, SSIM, and AB, but VIF decreases slightly both under noisy as well as noise-less conditions. We can verify that the proposed projectwith MCFEM blocks and skip connections outperforms all the other variants.

B. ABLATION STUDY ON LOSS FUNCTION

This section is to establish the significance of the three-part loss incorporated in project. We trained the projectwith a combination of perceptual, SSIM, and weighted patch-wise Euclidean (WPW Euclidean) losses. We can observe from Table 3 that the quantitative evaluations of these combinations are very similar, however from Fig. 3, notable visual differences are discernible on both clean (rows 1 and 2) as well as noisy (rows 3 and 4) images.

Inclusion of structural loss function without per-pixel or perceptual loss will be ideal only for edge detection or segmentation applications, hence ruled out from the ablation study. Including per-pixel loss without the other components may suit only noise removal application. Similarly, perceptual is only for feature detection, and hence, we have excluded these variants from the experimental analysis. First and the last column in Fig. 3 denotes the input and the ground truth, respectively, whereas the fourth column portrays the results of the projectwith the proposed three-part loss function. The second and third columns display outcomes from a combination of different loss functions.

For each row, in Fig. 3, a reference region is marked, which shows the efficacy of the three-part loss function. The texture of cloud and sky in the first image or the bush and area around the license plate in the second or light post and color contrast in the third noisy image or structural information of the

TABLE 3. Comparison of Performance Metrics for Loss Ablation.

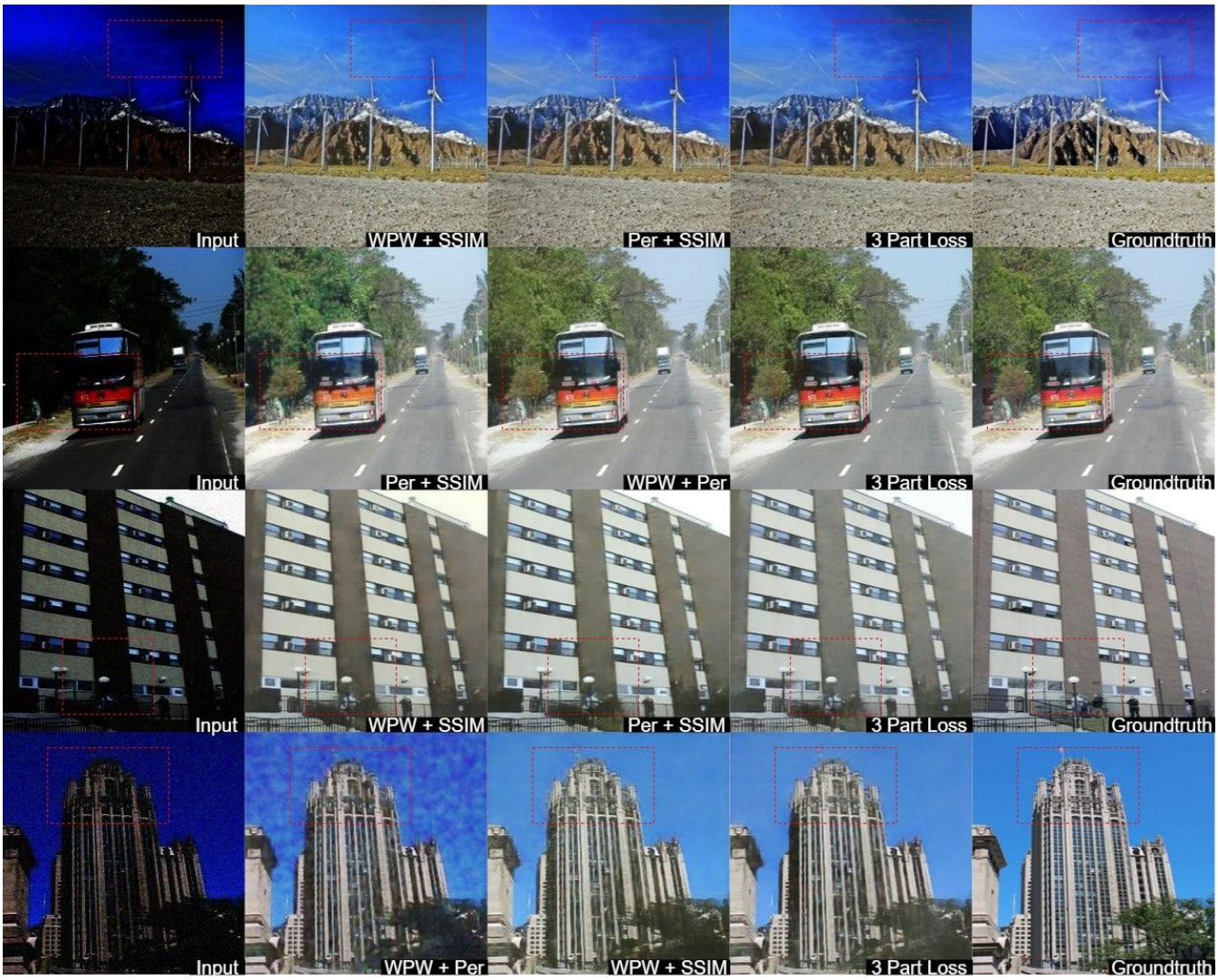

FIGURE 3. Ablation Study on Loss Function using Visual Evaluation.

TABLE 4. Quantitative Comparison on Clean and Noisy Images.

	Clean				Noisy			
Model	PSNR	SSIM	VIF	AB	PSNR	SSIM	VIF	AB
Input	9.144	0.2275	0.255	81.063	9.318	0.2202	0.149	78.391
SRIE [35]	10.308	0.3042	0.255	70.634	10.575	0.2677	0.161	65.493
Dong [36]	11.736	0.3730	0.243	57.963	11.930	0.2605	0.120	45.864
NPE [37]	11.606	0.3979	0.288	60.142	11.782	0.2847	0.149	46.688
MF [38]	11.770	0.3987	0.299	59.349	12.229	0.2996	0.151	47.971
LIME [8]	12.836	0.4136	0.314	43.638	12.847	0.4195	0.305	43.774
MSRCR [6]	18.580	0.7289	0.368	11.255	14.788	0.4156	0.161	21.116
Auto-MSRCR [40]	15.912	0.5833	0.349	18.381	13.331	0.3112	0.155	19.884
MSRCP [41]	14.868	0.5966	0.262	12.794	12.751	0.3342	0.117	20.362
BIMEF [39]	12.365	0.4519	0.324	55.390	13.118	0.3459	0.165	42.146
Exposure [42]	13.878	0.5372	0.217	29.613	12.200	0.4048	0.133	47.547
Deep Retinex [22]	15.255	0.5152	0.189	20.435	14.207	0.2812	0.104	19.628
Distort [43]	15.968	0.6048	0.258	21.316	14.083	0.4333	0.172	32.311
GLADNet [14]	24.176	0.8127	0.440	7.179	21.080	0.6552	0.259	7.7054
MBLLEN [13]	24.210	0.8132	0.446	7.322	20.554	0.6601	0.262	9.6339
C-LIENet (Ours)	24.438	0.8203	0.459	6.894	21.152	0.6739	0.273	7.302

building, color and smoothness of the sky in the fourth image *C. COMPARISON ON CLEAN AND NOISY LID IMAGES* are few notable examples to demonstrate the advantages of In this section, we have compared the qualitative and quanthe three-part loss

function. titative performance of projectagainst other traditional in conjunction with methods based on deep-learning. Meth- are trained on LID to the same extent as projectfor a ods based on deep learning such as Exposure [42], Deep fair comparison. We trained these models for 50 epochs, Retinex [22],Distort [43], MBLLEN[13] and GLADNet[14] with default settings, as mentioned by the respective authors.

Official code repositories, mentioned in the literature, are used for the experimental analysis.

In Table 4, blue, green, and red colors denote the top three algorithms, respectively. For the evaluation metrics, the proposed projectranks among the top 2 both on noisy and clean images. Contextual learning offered by MC-FEM blocks attributes towards the prominence of projecton noisy images. The high VIF of LIME [8] is due to the efficiency of the denoising algorithm on noisy images.

Visual comparison of the proposed approach, on noise-free and noisy images, is shown in Fig. 4 and Fig. 5, respectively. The visual quality of the proposed projectis superior compared to the other benchmarking algorithms, especially in noisy contexts. We observe that the methods like LIME [8] and Dong *et al.* [36] still yield dark images even upon enhancement. Several deep learning methods show comparable results, but they fail to retain the sharpness and overall hue of the image, specifically in the noisy context. The usefulness of the proposed algorithm is due to the increased spatial coverage provided by the MCFEM module. Thus, enabling improved learning under noisy conditions around the neighborhood.

D. COMPARISON ON NATURAL IMAGES

Lastly, we verified the performance of the algorithm on real low-light images. Regions marked, in Fig. 6 confirms the superiority of the proposed approach. These results ensure the suitability of projectfor real-time applications like autonomous navigation.

V. CONCLUSION

In this paper, we have proposed a novel architecture called Context-LIENet using dilated and separable convolution. Then, we have developed an innovative three-part loss function exploring the contextual and structural information. projectconsists of a unique multi-context feature extraction module in the encoder, depth-wise separable convolutions in the decoder, and skip connections from the encoder to the decoder. The multi-context feature extraction module enables the network to learn complex features such as brightness, contrast, and noise from a wider-context. Quantitative and qualitative experimental analysis on simulated and natural low-light images prove the ability of projectto outperform the benchmarking algorithms. The proposed three-part loss function using perceptual, structural, and weighted patch-wise loss components yields an enhanced image with the improved objective quality compared to the standard loss functions. An extensive study on component-wise hyper-parameter tuning needs focus. Using a unified model for both noisy and clean image enhancement might lack generalisability across domains. We have left these limitations open to the researchers working on low-light image enhancement. Several ideas presented in this work can be further extended to other image enhancement applications like image denoising, demosaicing, deraining, etc.

REFERENCES

[1] M. Abdullah-Al-Wadud, M. H. Kabir, M. A. A. Dewan, and O. Chae, ''A dynamic histogram equalization for image contrast enhancement,'' *IEEE Trans. Consum. Electron.*, vol. 53, no. 2, pp. 593–600, May 2007.

[2] W. Wang, X. Wu, X. Yuan, and Z. Gao, ''An experiment-based review of low-light image enhancement methods,'' *IEEE Access*, vol. 8, pp. 87884–87917, 2020.

[3] E. H. Land, ''The Retinex,'' *Amer. Sci.*, vol. 52, no. 2, pp. 247–264, 1964.

[4] E. H. Land, ''The Retinex theory of color vision,'' *Sci. Amer.*, vol. 237, no. 6, pp. 108–128, Dec. 1977.

[5] D. J. Jobson, Z. Rahman, and G. A. Woodell, ''Properties and performance of a center/surround Retinex,'' *IEEE Trans. Image Process.*, vol. 6, no. 3, pp. 451–462, Mar. 1997.

[6] D. J. Jobson, Z. Rahman, and G. A. Woodell, ''A multiscale Retinex for bridging the gap between color images and the human observation of scenes,'' *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, Jul. 1997.

[7] C.-H. Lee, J.-L. Shih, C.-C. Lien, and C.-C. Han, ''Adaptive multiscale Retinex for image contrast enhancement,'' in *Proc. Int. Conf. Signal-Image Technol. Internet-Based Syst.*, Dec. 2013, pp. 43–50.

[8] X. Guo, Y. Li, and H. Ling, ''LIME: Low-light image enhancement via illumination map estimation,'' *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, Feb. 2017.

[9] M. Yang, X. Nie, and R. W. Liu, ''Coarse-to-fine luminance estimation for low-light image enhancement in maritime video surveillance,'' in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Auckland, New Zealand, Oct. 2019, pp. 299–304.