

LumiNet: Multi-Spatial Attention Generative Adversarial Network for Backlit Image Enhancement

Samprit Bose, Sahil Nawale, Dhruv Khut, Maheshkumar H. Kolekar, *Senior Member, IEEE*

Abstract—Backlit image enhancement is a crucial task in improving the quality and visibility of the underexposed regions in an image caused by the difference in illumination between the background and foreground. Traditional methods struggle to effectively handle the dynamic range compression required for backlit image enhancement and fail to properly balance the exposure between the background and foreground areas. In this paper, we propose a novel backlit image enhancement architecture using a modified UNet and a unique 1x1 discriminator-based conditional generative adversarial network. Our approach incorporates custom hyperparameters, a tailored loss function, and bilateral guided upsampling for efficient enhancement of large images. We evaluated our model on the BAID dataset and achieved superior results in various image evaluation metrics like PSNR, NIQE and ΔE° , demonstrating its effectiveness in enhancing both backlit and low-light images. We have also compared the computational efficiency and resources used by our method with other state-of-the-art methods. Extending our work, we have incorporated our model into a pipeline for enhancing backlit traffic images and videos, which is then used to detect the license plates of vehicles with improved accuracy.

Index Terms—Backlit image enhancement, Dynamic range compression, Deep learning, Low-light image enhancement

I. INTRODUCTION

A. Motivation

IMAGE enhancement refers to the action of manipulating an input image to improve specific features or characteristics. It serves a crucial role in enhancing visual perception for tasks such as pattern recognition, computer vision, object tracking and digital image management [1] [2]. Backlit images are photographs or other visual media that are illuminated from behind the subject, with the light source positioned behind the subject rather than in front. However, this type of lighting can have negative effects on the quality of the image [3]. One issue that can arise is underexposure of the subject, where the brightness of the background tricks the camera's metering system and results in the subject appearing darker and with less detail. Additionally, lens flare can occur when shooting directly into the light source, causing unwanted reflections and glare that create a hazy or washed-out appearance. Enhancing backlit images is a crucial step in bringing out the details and aesthetic appeal of the subject, which may have been lost due to the challenges of backlighting. By adjusting exposure, colour balance, and other image parameters, we can effectively highlight the subject and create a visually striking image.

Improving backlit images also helps to counter the negative effects of backlighting, such as underexposure and lens flare. Using techniques such as exposure compensation, fill flash, or reflectors, we can balance the lighting and enhance the overall quality of the image. Current methods for enhancing backlit images have some drawbacks, including issues with colour balance, saturation, and exposure. While these methods have been effective in increasing the detail in backlit images, they often result in over-enhancement of contrast and brightness. To create a single, improved image, several of these techniques also call for combining many images with various exposure settings. These limitations can make it difficult to achieve a natural-looking and visually appealing final result. Enhancing backlit images can be beneficial in various computer vision applications, especially in identifying and locating objects in images that have challenging lighting conditions [4]. Backlit images, with their uneven lighting and loss of detail, can pose challenges to accurate object detection and recognition. However, by enhancing the backlit images, it is possible to enhance the image quality, which can increase the amount of detail and information that can be extracted from them [5]. This can increase object detection precision and recognition, as well as other image processing tasks such as segmentation and feature extraction. Transfer learning technique is also used in several computer vision applications for more accurate feature extraction [6]. Backlit image enhancement can also be useful in critical applications such as video surveillance, autonomous vehicles, sports video enhancement for highlight generation and robotics [7] [8]. Accurate object detection and recognition are crucial in these applications, and by enhancing the quality of the digital images used in these systems, their performance and reliability can be improved [9]. It can also be used for developing IoT-based real-time vehicle detection and classification applications that eliminate the chance of false detection as well as video anomaly detection [10] [11]. This can make them more effective in real-world settings, where uneven and low lighting and loss of detail are common challenges and need to be enhanced in real-time [12] [13]. This can greatly enhance the effectiveness of computer vision applications across various industries and applications. In this paper, we have proposed a backlit enhancement model using modified conditional GAN. A unique architecture is presented in this study for enhancing backlit images, which has been mainly evaluated on a dataset consisting of 3000 backlit images and two public small-scale backlit image datasets. The proposed model has been assessed using various eval-

uation metrics, including PSNR, NIQE, and ΔE° , and has demonstrated superior performance. Additionally, the model has been tested on different benchmark datasets for low-light and general image enhancement, yielding satisfactory results. Notably, the proposed approach not only excels in enhancing backlit images but also exhibits promising capabilities in improving low-light conditions.

B. Related Works

1) *Segmentation and Multi-exposure-based Methods:* In 2019, Liu et al. [14], introduced a method for enhancing underexposed images while preserving details. They proposed a new optimally weighted multi-exposure fusion mechanism to generate a sequence of multi-exposure images from the input underexposed image. However, none of the individual images met the desired quality. To address this, they formulated an energy function to calculate optimal weights based on local contrast, saturation, and exposure. Using a weighted multi-exposure fusion method, they generated the final enhanced image with improved brightness, colour, and preserved details. However, the method may not achieve optimal results when dealing with highly underexposed regions and the computational efficiency is also very low. Ma et al., in 2020 [15], proposed a new technique for improving backlit images. The suggested approach incorporates local colour enhancement of backlit pictures in the R, G and B channels. Additionally, a new function is used to adaptively change the V channel in the locally improved pictures, and the RGB brightness adjustment model is converted into the HSV brightness adjustment model. This method of self-adaptive brightness adjustment is made to work with a variety of backlit pictures. In 2020, Trongtirakul et al. [16], presented an approach for backlit image enhancement using full-piecewise non-linear automated stretching without requiring any input parameters from the user. The suggested method has the ability to boost the local contrast of dark parts, expose hidden intricacies in the dark regions, and retain the characteristics and colour of common and too bright places. The effectiveness of the proposed method is demonstrated through computer simulations using Li's backlit image database and various backlit images from commercial gadgets. A technique for improving the contrast of backlit pictures in each colour channel of the RGB space has been developed by Buades et al. (2020) [17]. To solve the issue of pictures with both very dark and highly saturated parts, they employed gamma and logarithmic tone mapping functions. They have also developed an image fusion algorithm to combine the results obtained from these tone mappings. While certain current methods for enhancing backlit images can be useful, they often lack the ability to adapt to changes in lighting conditions or scene variations since they rely on fixed parameters. Moreover, these methods may lead to undesirable colour distortions and unnatural lighting effects when processing images that contain very dark areas. In 2015, Li et al. [18], proposed a soft binary segmentation process based on a Gaussian mixture model to identify sections of a backlit image that are overexposed or underexposed. Optimal tone mapping is then applied separately to these regions to

enhance visual quality. In a study cited as [19], researchers developed a learning-based method for restoring backlit images. The suggested approach consisted of separating the backlit items from the non-backlit objects, using two different tone mapping functions, and then reassembling the picture. The precision of the segmentation output, however, was crucial to the algorithm's success. Unwanted lighting fluctuations in the output image can occur if the backlit and non-backlit portions are incorrectly classified.

Segmentation-based methods for backlit image enhancement have certain drawbacks that can limit their effectiveness. For instance, these methods may not be able to accurately distinguish between backlit and non-backlit regions, resulting in undesired artifacts in the output. Furthermore, the segmentation task may require manual labelling or pre-trained models, which can take a lot of time and be unnecessary. On the other hand, multi-exposure-based methods for backlit image enhancement rely on capturing multiple images at different exposure levels. However, this approach may not always be feasible or practical and can suffer from alignment issues between the multiple images. If the alignment is not accurate, it can lead to artifacts in the output. Additionally, this method may not be suitable for images with complex lighting conditions, such as scenes with moving objects or changing lighting conditions.

2) *Traditional and Retinex-based Methods:* Ren et al. [20], in 2020, introduced a novel approach called the Low-Rank Regularised Retinex Model (LR3M) for reducing noise in the reflectance map during the retinex decomposition process. Their approach sequentially estimated a piece-wise smoothed illumination and a noise-suppressed reflectance, effectively eliminating noise in both maps. They further enhanced the illumination layer and applied the LR3M to video low-light enhancement by considering inter-frame coherence and leveraging temporal correspondence through reflectance maps of successive frames to establish a low-rank prior. While this is beneficial for certain applications, the complexity of video processing and the presence of motion artifacts can affect the overall performance of the LR3M in video enhancement scenarios. In 2021, Akai et al. [21], developed an efficient method for enhancing backlit images that enhances the visibility of dark regions. The proposed method involves using gamma correction to obtain an intensity-adjusted image and histogram equalization to obtain a contrast-enhanced image. These images are then alpha-blended to produce a blended image, which is fused with the original image using alpha blending to prevent artifacts in bright regions. To determine the weight for fusion, Otsu's method and the guided filter are used. In 2020, a strategy for improving single-backlit photos using histogram specification was suggested by Ueda et al. [22]. The suggested approach is to choose an intensity histogram shape that can improve the bimodal distribution. The approach then uses histogram specifications to change the intensity while keeping the image's hue and saturation intact. Zhao et al. [23], in 2021 proposed a method involving a three-step process for enhancing backlit images. The RGB colour space is first transformed into the HSV colour space. Then, the image is separated into its illumination and reflection components using

the illumination-reflection imaging model. In order to create the improved picture, the original reflection component and the adjusted lighting component are fused. This approach can effectively deal with the issue of inconsistent illumination in backlit images, resulting in improved quality and increased information content. In 2023, Lecca [24] proposed a new Retinex-inspired image enhancer called SuPeR-B that enhances backlight and spotlight photos without the need for multi-scale analysis, segmentation, or smoothing. SuPeR-B analyses each image channel independently and rescales the intensity of each pixel based on a weighted average of intensities from regular sub-windows, in contrast to other techniques that rely on multi-scale analysis, segmentation, and smoothing. SuPeR-B resembles a bilateral filter since its rescaling element accounts for both spatial and intensity data. In 2022, an innovative Adaptive Gamma Correction with Weighting Distribution (AGCWD) technique for backlit picture improvement, based on the Multi-scale Retinex (MSR) approach, was proposed by Chou et al. [25]. The algorithm involves converting the backlit image to HSV colour space, applying a blurred-surrounding luminance filter, and using an MSR-based just-noticeable difference (JND) curve weakening factor in combination with AGCWD. Instead of using a Gaussian filter, the suggested method uses a mean filter to simulate the surrounding luminance in the MSR method, reducing computational complexity. Additionally, the cumulative distribution function curve is subject to a constraint to avoid over-enhancement of high-luminance areas and preserve local detail information.

Traditional methods for backlit image enhancement, such as global and local histogram equalization, have certain limitations. For example, they may lead to over-enhancement or parts of the image that aren't sufficiently enhanced, which can result in an unnatural or unrealistic appearance. Moreover, these techniques might not be effective for images with complex lighting conditions or a high dynamic range. Retinex-based methods for backlit image enhancement can also have drawbacks. These techniques rely on the presumption that the picture may be broken down into components for reflectance and illumination. However, this assumption may not always hold true in practical scenarios, and the resulting output may not always be visually pleasing or realistic. Retinex-based techniques may not be appropriate for real-time applications and can be computationally costly.

3) Deep Learning-based Methods: In 2022, Lv et al. [26], proposed a new backlit image enhancement network called BacklitNet, which uses saliency guidance for natural and robust backlit image restoration. A stacked U-structure and bilateral grids are used in their suggested model to extract multiscale saliency information and enhance images of arbitrary resolution quickly. They also include a loss function that instructs the network to concentrate more on the backlit areas during training and is based on past knowledge of the brightness distribution of backlit pictures. Takao proposed a novel approach for image enhancement in 2023 [27], which incorporates the Laplacian pyramid into the network process. The proposed method increases the detail, sharpness, and contrast of an input picture by using the hierarchical Laplacian

pyramid representation. In addition, Takao introduced a visual prior that is specific to underwater images to the MLE process. This made it possible to develop a model with just seven convolutional layers called the Zero-Shot underwater image improvement model. In 2021, Khan et al. [28], introduced a novel deep-hybrid enhancement technique for low-light images and improved their contrast. Their proposed method involves decomposing input images into reflection and illumination components, followed by a contrast enhancement strategy. They introduced a Divide to Glitter network (D2G-Net), which does not require large quantities of paired training data and can learn from few-shot training samples. The D2G-Net is made up of a Glitter-Net to magnify the illumination map and a multilayer Division-Net for image division. In 2020, Zeng et al. [29], presented a novel approach for efficient and robust photo enhancement using image-adaptive 3D lookup tables (3D LUTs). They used an end-to-end learning approach to concurrently learn numerous base 3D LUTs and a tiny convolutional neural network (CNN). The CNN combined the several base 3D LUTs into an image-adaptive 3D LUT by processing a downsampled version of the input picture and predicting content-dependent weights. The colour and tone of the source image were then effectively transformed using this adaptive 3D LUT.

Deep learning methods are preferred for backlit image enhancement owing to their capacity to automatically learn complex non-linear mappings between input and output images. These methods have the capability to automatically learn relevant features from the input image and utilize them to improve the image in a data-driven manner. Furthermore, deep learning methods can handle images with complex lighting conditions and high dynamic range, which are generally challenging for traditional methods. Another benefit of deep learning methods is their ability to generalize well to unseen data, which makes them suitable for a wide range of applications. These methods can also be trained on large datasets and optimized for specific performance metrics, leading to improved accuracy and robustness.

C. Contribution and Organization of Paper

Backlit images are captured in situations where the primary source of light is behind the subject, such as when taking a photo of a person in front of a bright window. These images frequently have poor contrast, underexposure, and a loss of detail in the subject due to the strong backlight. Enhancing backlit images can provide better information for analysis and decision-making and can lead to better outcomes in various fields like video surveillance, medical imaging and remote sensing. Hence, we have suggested a deep learning approach for the enhancement of backlit images. The following summarises this work's key contribution.

- 1) A modified conditional GAN framework using multi spatial attention-based UNet as a generator and 1x1 patch GAN as a discriminator has been proposed to enhance backlit images.
- 2) Our approach introduces a custom-designed spatial attention gate that enhances inter-spatial features and

performs downsampling simultaneously. We integrate this spatial attention gate into each layer of the encoder in our modified U-Net architecture. Additionally, we make several deviations from the conventional U-Net, including the use of different activation and normalization functions, and the inclusion of convolution, normalization and activation layers in the bottleneck region. We also fine-tune the hyperparameters by experimenting with various kernel sizes, paddings, and strides. These modifications significantly improve the performance and effectiveness of our U-Net model.

- 3) We have developed a customized discriminator that classifies individual pixels, incorporating different activation and normalization layers compared to conventional methods. Furthermore, we have introduced a specialized loss function specifically tailored to enhance backlit images while ensuring high image quality as measured by various metrics.
- 4) To enhance computational efficiency, we have integrated Bilateral Guided Upsampling (BGU) into our model's architecture. This enhancement allows us to effectively enhance large images while maintaining a linear increase in processing time relative to the image resolution.
- 5) We have also extended our work and tested our model on traffic images that were affected by low light and were successfully able to detect the details in the images, including the vehicle number plates, via OCR in python.
- 6) The suggested model was evaluated for low-light and regular image enhancement datasets and produced acceptable results. It also outperformed other cutting-edge approaches on the benchmark BAID dataset in terms of PSNR, NIQE, and ΔE° .

The proposed approach's methodology is described in Section II, which includes the proposed modified c-GAN architecture for backlit image enhancement. Section III presents the experimental analysis along with the obtained results from our approach. Finally, Section IV concludes the research and outlines potential future work.

II. METHODOLOGY

A. Overview

GAN (Generative Adversarial Network) is a particular kind of deep learning model that is able to produce artificial data that closely mimics actual data. A generator and a discriminator are the two networks that make up the model. The generator uses random noise as input to produce synthetic data that is meant to imitate the real data, whereas the discriminator uses both real and synthetic data as input and tries to tell them apart. In a procedure known as adversarial training, the two networks are trained simultaneously while the generator tries to deceive the discriminator and the discriminator tries to accurately identify the input as genuine or synthetic [30].

A conditional GAN (cGAN) is a kind of generative adversarial network (GAN) that can produce samples from a specific category or condition. It is a supervised learning technique where the generator and discriminator are conditioned on additional data including labels, captions, or images, in

addition to the noise vector used in traditional GANs [31]. This conditioning allows the cGAN to generate samples that match a specific condition. GANs are a type of generative model that aims to learn the mapping between a random noise vector z and an output image y , denoted as $G: z \rightarrow y$. On the other hand, conditional GANs learn to map not only the random noise vector z but also an observed image x to an output image y , represented as $G: x, z \rightarrow y$. The primary goal of the generator G is to provide results that are identical to actual images, as determined by a discriminator D that is trained to identify the generator's "fakes". Both the generator and discriminator networks are trained in an adversarial way, where the generator strives to make more realistic pictures to deceive the discriminator while the discriminator seeks to recognise the fake images produced by the generator. [32].

B. Proposed Architecture

We proposed a multi-spatial attention-based UNet as the generator, which takes the input image and a vector of random noise as inputs and produces an output image. To train the generator, a 1x1 patch GAN is used as the discriminator. In order to discriminate between actual and fake patches in the input image, the discriminator is trained, which helps to improve the stability of the training process. The generator is taught to create results that the discriminator cannot tell apart from actual images. By using this modified conditional GAN framework, the resulting enhanced backlit images are able to capture more details and exhibit a more natural appearance compared to traditional methods. Fig. 1 shows the network architecture of the proposed backlit enhancement model. We also adapted local curves for the relationship between the output image and the low-resolution input image. By placing the input into a bilateral grid and solving the 3D array of affine matrices that best matches the input colour to the output colour per x, y intensity bin, the curves are calculated. By analysing these low-resolution curves on high resolution input, the high-resolution output is produced. This approach has shown promising results and has the potential to be further optimized for other image enhancement tasks.

1) Multi-attention based-UNet as Generator: Our proposed method for enhancing backlit images involves using a modified version of the Unet model as the generator, with a multi-spatial attention mechanism added in the encoder part, and a 1x1 patch GAN as the discriminator. In the generator, we have made several modifications to the standard Unet model, such as using a single convolution layer in each layer, adding a spatial attention gate in place of the second convolution layer in all five layers of the encoder, and using InstanceNorm instead of BatchNorm in all layers. Instead of using double convolutions in each layer, we opted for a single convolution. This change significantly improved the runtime and reduced GPU memory usage without significantly affecting the image quality. Additionally, we introduced normalization to the bottleneck layer of the generator, enhancing nonlinearity and improving loss fitting. By using transposed convolutions instead of direct upsampling, we employed trainable kernels, thereby increasing the number of trainable parameters for better fine-tuning. To

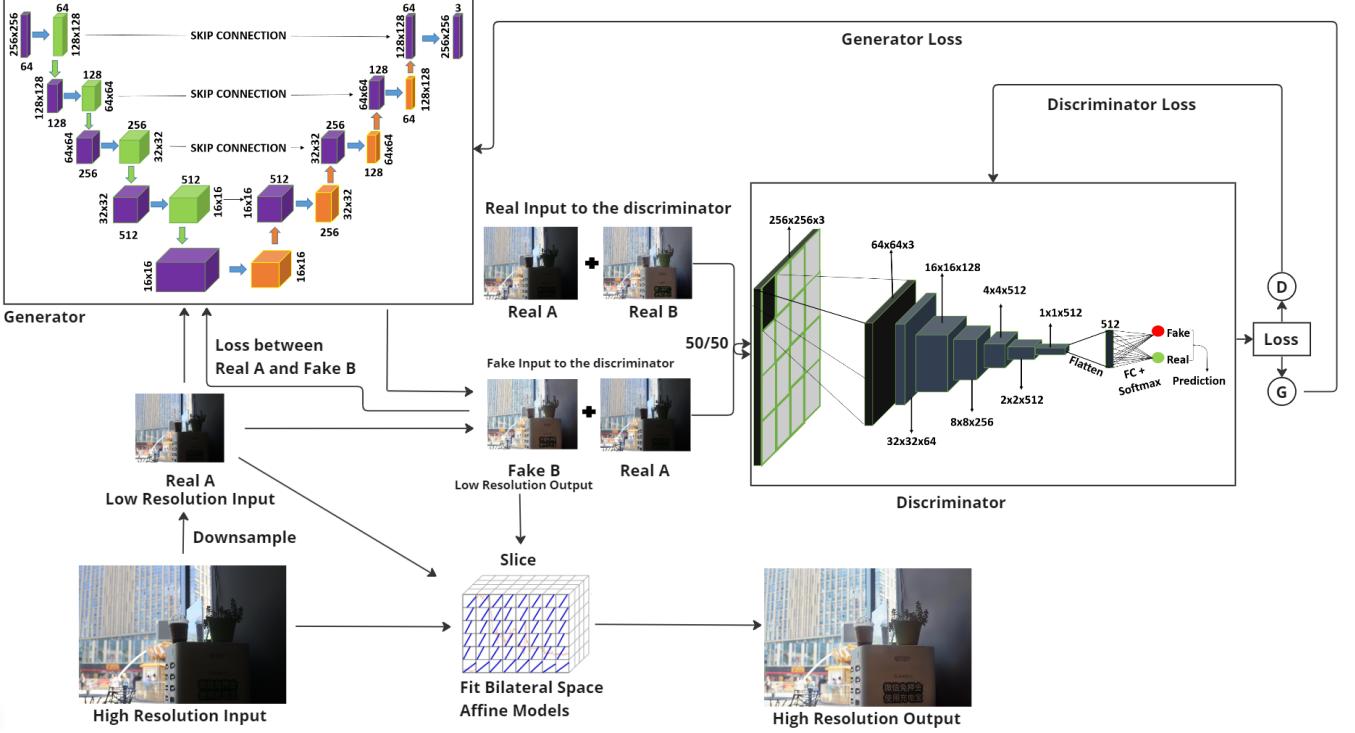


Fig. 1. Network architecture for backlit enhancement model

address issues related to batch normalization, which led to overall whitening and colour distortion, we utilized instance normalization. This modification proved effective in handling images with varying degrees of backlighting. In order to maintain a greater dynamic colour range, we also modified the spatial attention gates to operate with instance normalization and used ReLU activation rather than sigmoid activation. Furthermore, we found that using a combination of max pooling and average pooling helped maintain sharpness and lighting consistency in the background. The convolution layer within the spatial attention gates performed downsampling and channel reduction. Fig. 2 shows the block diagram of the modified UNet architecture used in the generator.

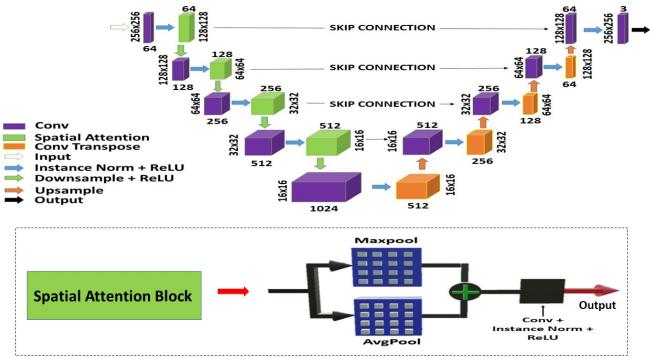


Fig. 2. Block diagram of the spatial attention UNet architecture used in the generator

Attention mechanism is a technique used in deep learning to selectively concentrate on certain sections of the input during

the model's processing. It enables the model to weight the importance of various parts of the input depending on their relevance to the task at hand. In a typical attention mechanism, the model generates an attention weight for each input element based on its relevance to the output [33]. This weight is then used to combine the input elements to form a context vector, which is then fed into the next layer of the model. The attention weight can be generated using various methods, such as dot-product attention, additive attention, and multiplicative attention. The standard sequence-to-sequence (seq2seq) framework struggles when processing long input sequences because it can only use the data generated by the encoder's last hidden state to determine the decoder's output [34]. This limitation is overcome by the attention mechanism, which permits the decoder to access all of the encoder's hidden state data to generate the context vector. The attention mechanism creates a distinct mapping for each time step for all hidden states, enabling the decoder to use the whole hidden sequence rather than just the most recent concealed information. By using this method, the framework is able to concentrate on the input sequence's most important segments as needed.

In this paper, we have utilised spatial attention in all five encoder blocks, as shown in Fig. 1. The spatial attention mechanism typically has two major sections: a pooling layer and a convolutional layer. The pooling layer generates a low-resolution feature map by aggregating information from the input image. This low-resolution map is then passed through a convolutional layer, which learns to delegate weights to each feature in the map. The resulting weight map is then multiplied with the original feature map in order to get the final output, where the important regions of the input are amplified and

the less relevant regions are suppressed. The spatial attention module downsamples the images and generates a spatial map, which is useful for generating a high-quality output. By using ReLU instead of Sigmoid activation function in the spatial attention module, we avoid the issue of the output being restricted to a range of 0 to 1, which is only suitable for image segmentation. The use of a 1x1 patch size in the discriminator helps to reduce the visibility of the patches in the output image. We have also made a modification to the standard kernels used in the Unet architecture by replacing them with larger kernels. By using larger kernels, we aim to capture more contextual information and maintain the minor details of the image that may require knowledge of the background. The larger kernels allow for a broader receptive field, enabling the model to gather more information and make better-informed decisions when enhancing the image. This helps to preserve intricate details that would otherwise be omitted or faded out using smaller kernels. Fig. 3 depicts the fading out of intricate details in an enhanced image due to a smaller kernel size.



Fig. 3. (a) Original Image (b) Fading out of intricate details due to smaller kernel size

2) 1x1 PatchGAN as Discriminator: In GAN, a discriminator is a neural network that has been trained to differentiate between real and fake images. It accepts an image as input and returns a likelihood score stating whether the picture is real or artificial. The generator aims to create visuals that can deceive the discriminator into believing they are authentic. The discriminator is then trained to become more skilled at differentiating between produced and actual pictures. The generator will keep doing this until the final product is indistinguishable from real images. In order to address high-frequency issues, we utilized patchGAN, which only penalizes structures within a specific range of patches [35]. In this paper, we have used 1x1 patchGAN in the discriminator, as shown in Fig. 4. The role of the 1x1 PatchGAN in a GAN is to determine the authenticity of each local patch of a picture. By using a small receptive field, the discriminator can make more fine-grained decisions about the realism of an image. The 1x1 convolutional layer in the PatchGAN aggregates the local decisions into a single score for the entire image. This score is then used to update the generator during the training process. Instance normalization is used in the discriminator in place of batch normalization, and LeakyReLU activation function with a slope of 0.2 is used. The choice of these specific values is based on experimentation and has been shown to produce better results.

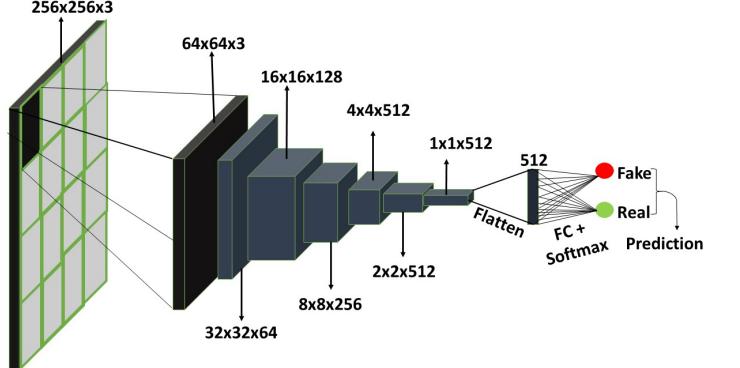


Fig. 4. 1x1 PatchGAN as Discriminator

The proposed network architecture, as depicted in Fig. 1, is designed to address the problem of enhancing backlit images. The generator is a modified version of the UNet model, which utilizes spatial attention to improve its performance. The generator takes in real images affected by backlight and produces a fake backlit enhanced image. The discriminator, on the other hand, is a 1x1 patchGAN, which is used to classify whether the input images are real or fake. In this case, the real part of the discriminator is composed of real backlit images and their corresponding real backlit enhanced images. The fake part, on the other hand, consists of real backlit images and the fake backlit enhanced images produced by the generator. During training, the discriminator's loss is used to adjust its weights, making it more accurate in distinguishing between genuine and false images. The generator's loss is used to modify the generator's weights, improving its ability to produce more realistic and accurate backlit-enhanced images.

C. Loss Function

The loss function used in the proposed GAN network consists of four components. The first component is the standard GAN loss, which is the Binary Cross-Entropy with Logits loss for both the generator and discriminator. This loss is calculated for each 1x1 patch in the image generated by the discriminator. The second component is the L1 Loss, which measures the mean absolute difference between the output generated by the generator and the real image. This loss, along with the GAN loss, constitutes the standard loss for any GAN network. The third component is the Mean Squared Error (MSE) loss, which serves as colour correction for the output image. The fourth component is the Structural Similarity Index (SSIM) loss, which compensates for the reduced SSIM score when the low-resolution output is upscaled using BGU. The SSIM score is a measure of the resemblance of two images based on luminance, contrast, and structure. A conditional GAN's objective may be stated as in equation 1:

$$L_{cGAN}(G, D) = E_{x,y}[\log D(x, y)] + E_{x,z}[\log(1 - D(x, G(x, z))] \quad (1)$$

The discriminator D seeks to maximise the objective function, whereas the generator G seeks to minimise it. It calculates

the difference between the anticipated and actual values, determines the absolute value of the difference, and then arithmetically averages all the absolute differences to determine the final loss amount. L1 loss (eq 2) calculates the difference between the predicted value and the actual value, takes the absolute value of the difference, and then averages all the absolute differences to get the final loss value.

$$L_{L1}(G) = E_{x,y,z}[\| y - G(x, z) \|_1] \quad (2)$$

The ultimate goal of our approach (eq. 4) is to optimize a certain objective function that guides the training of the conditional GAN with the aim of generating realistic and high-quality backlit enhanced images.

$$G^* = \text{argmin}_G \max_D L_{cGAN}(G, D) + \lambda L_{L1}(G) \quad (3)$$

The final formula for the loss function is:

$$\begin{aligned} Loss = & GAN_Loss + L1_Loss * 100 + MSE * 1000 + \\ & (1 - ssim_score) * 20 \end{aligned} \quad (4)$$

In the above equation, ssim_score is given by the formula,

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (5)$$

where:

- μ_x and μ_y are the means of the two images x and y , respectively.
- σ_x^2 and σ_y^2 are the variances of the two images x and y , respectively.
- σ_{xy} is the covariance between the two images x and y .
- c_1 and c_2 are two variables used to stabilize the division with weak denominator.

Also, the mean squared error (MSE) can be formulated as:

$$\text{MSE} = \frac{1}{m \times n} \sum_{i=1}^m \sum_{j=1}^n (x_{ij} - y_{ij})^2 \quad (6)$$

where m and n are the dimensions of the images being compared, x_{ij} and y_{ij} are the values of pixels at position (i, j) in the two images being compared.

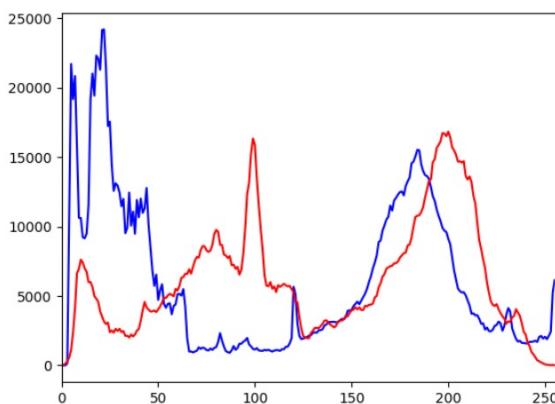


Fig. 5. Intensity Histogram of Backlit (blue) and Backlit enhanced image (red)

Algorithm 1 Algorithm of Proposed Approach

Input : Backlit image

Output : Enhanced image

1. **for** $i = 1 : n_{\text{epochs}}$ **do**:
 2. **for** $j = 1 : n_{\text{images}}$:
 - a. Resize j^{th} image $real_A$ to 256x256 and normalize
 - b. Pass input through net_G to get fake output $fake_B$
 - c. Turn on gradients for disc $net_D_grad = \text{True}$
 - d. Set gradients of disc to 0
 - e. Concat and feed both $real_A$ & $fake_B$ output to disc
 - f. Cal L_d by taking avg of loss of disc on both real input and fake output
 - g. Calculate gradients for disc
 - h. Update W_d by ascending on its gradients L_d
 - i. Turn off disc gradients
 - j. Set gen gradients to 0
 - k. Change shape of $fake_B_{\text{nxchxw}}$ & $real_B_{\text{nxchxw}}$ to $fake_B_{\text{nhwxw}}$ & $Real_B_{\text{nhwxw}}$
 - l. Squeeze $fake_B$ & $real_B$ to $fake_B_{\text{hxwx}}$ & $real_B_{\text{hxwx}}$
 - m. Calculate SSIM loss for net_G as $\text{ssim}(\text{mean}(fake_B_c), \text{mean}(real_B_c))$ with range 1.0
 - n. Cal L_g as $L_g = L_{\text{gan}} + \alpha * L_{L1} + \beta * \text{mse} + \gamma * (1 - \text{SSIM})$
 - o. Calculate the gradients for gen.
 - p. Update W_g by ascending on its gradients L_g
 - q. $W_{g_{\text{new}}} \leftarrow V(L_g)$
 - r. $W_{d_{\text{new}}} \leftarrow V(L_d)$
 3. **end for**
 4. Evaluate performance using state-of-the-art methods in terms of effectiveness indexes.
 5. Store the model and results.
 6. **end for**
-

We have also calculated the intensity histogram as shown in Fig. 5, where blue colour represents a backlit image and red colour signifies backlit enhanced images. From the graph, it is observed that Backlit images tend to exhibit a higher concentration of pixels in the low-intensity range, particularly in the underexposed foreground region.

D. License Plate Recognition

We have broadened the real-life application scope of our model by integrating it into a pipeline designed to achieve precise license plate detection through low-light enhancement. This pipeline encompasses several steps. Initially, the traffic image undergoes enhancement via our model, enhancing the visibility of the license plate. Subsequently, the processed image is subjected to optical character recognition (OCR) to decipher the characters on the license plate. Lastly, the image is adorned with bounding boxes encapsulating the identified number, resulting in an enhanced and legible license plate representation as shown in Fig. 6. This application of our model demonstrates its potential for enhancing the efficiency and accuracy of tasks related to traffic surveillance and monitoring.

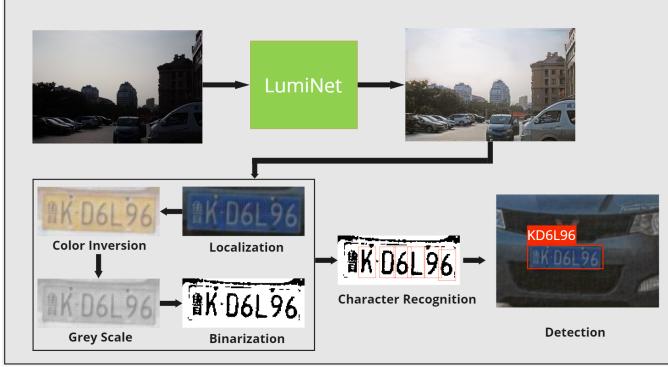


Fig. 6. Automatic Number Plate Recognition from Backlit Enhanced Image

III. EXPERIMENT

A. Dataset

The effectiveness of our suggested model is assessed on three datasets, which include the BAID dataset [26] and two public small-scale backlit image datasets, namely the SICE dataset [36] and the Vonikakis et al. dataset [37]. BAID dataset comprises 3000 backlit images and their respective high-quality ground truth. The SICE dataset comprises 38 images, while the Vonikakis et al. dataset has 23 images. The ground truth images are meticulously created through manual adjustments made by experienced photographers to tailor the backlit images. Additionally, a subset of backlit and ground truth image pairs were meticulously generated by manipulating the camera's exposure settings. This process involved setting the camera to spot metering mode and manually adjusting the exposure to target a specific area, corresponding to the subject, to mitigate the effects of backlighting. Following this approach, the first image captures the original backlit scene, while the second image in the pair captures the ground truth by adapting the camera's exposure settings. To demonstrate the robustness and generalization ability of our model, we additionally performed tests on the LOL (Low-Light Image Enhancement) standard dataset (Wei et al., 2018) [38], consisting of 500 actual pairs of low-light images and 1000 fake pairs of images.

B. Evaluation Metrics

To thoroughly evaluate the effectiveness of our model, we have used four commonly used metrics in terms of Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), ΔE° and Natural Image Quality Evaluator (NIQE). PSNR measures the quality of the reconstructed image in terms of its peak signal level and noise, while SSIM assesses the structural similarity between the original and improved image. Entropy measures the amount of information in an image, which is related to the number of pixels and their distribution. NIQE is a no-reference metric that evaluates the naturalness of an image. It measures the degree of distortion in an image by comparing it to the natural image statistics.

C. Implementation Details

For the first 10 epochs, the learning rate is set at 0.0001, then 10 epochs at 1e-5, and 80 epochs at 1e-6. The learning rate is set to 0.0001 for the first 10 epochs, followed by 10 epochs of 1e-5 and 80 epochs of 1e-6. Input images are scaled to 256x256 and normalized with a mean and standard deviation of 0.5. The entire network is optimized using the Adam optimizer. Our system is equipped with a NVIDIA GeForce GTX 1650 GPU, AMD Ryzen 5 4600H Mobile Processor 3.2GHz, 16GB SO-DIMM DDR4-3200MHz RAM, and 512GB PCIe 3.0 NVMe M.2 SSD storage. The average runtime for image sizes of 512x512 and 1024x1024 is 4ms and 16ms respectively, without bilateral upsampling.

D. Comparison with the State-of-the-Art Methods

1) *Quantitative Analysis:* We have compared our model in terms of evaluation metrics with other state-of-the-art methods. BacklitNet is an image enhancement technique that employs a nested U-structure and bilateral grids to restore backlit images in a robust and natural way [26], backlit image enhancement was also carried out by Buades et al. [17] via means of global tone mappings and image fusion. Li and Wu [39] also proposed backlit image restoration based on learning. Zhang et al., also performed zero-shot enhancement of backlit images using deep internal learning [40] and Liang et al. [41] introduced an initial prompt pair, where they enforced the similarity between the text prompt and the corresponding backlit image using the CLIP latent space. Comparisons were also done on low-light image enhancement methods [38] [42] [43] [44] [45] and five state-of-the-art general image enhancement methods [46] [47] [48] [49] [50]. Table 1 shows a quantitative comparison of our proposed approach on the BAID dataset with the aforementioned state-of-the-art methods.

TABLE I
OUR PROPOSED METHOD WAS QUANTITATIVELY COMPARED WITH OTHER STATE-OF-THE-ART TECHNIQUES ON THE BAID DATASET, WHICH WAS DIVIDED INTO THREE CATEGORIES: (I) METHODS FOR ENHANCING BACKLIT IMAGES, (II) METHODS FOR ENHANCING LOW-LIGHT IMAGES, AND (III) METHODS FOR ENHANCING GENERAL IMAGES. THE BEST RESULTS ARE SHOWN IN BOLD.

Approach	PSNR	SSIM	ΔE°	NIQE
Li and Wu (2018) [39]	17.16	0.82	13.10	4.86
ExCNet (Zhang et al., 2019b) [40]	19.31	0.90	11.41	2.94
Buades et al (2020) [17]	17.47	0.89	13.46	3.67
BacklitNet (Lv et al., 2022) [26]	25.06	0.96	6.45	2.81
CLIP-LIT (Liang et al., 2023) [41]	21.64	0.89	9.52	3.54
RetinexNet (Wei et al., 2018) [38]	21.26	0.90	9.67	3.48
KinD (Zhang et al., 2019a) [42]	22.69	0.91	7.76	3.06
DeepUPE (Wang et al., 2019b) [43]	21.05	0.90	9.73	2.97
Zero-DCE (Guo et al., 2020) [44]	19.74	0.87	10.29	4.47
SNR-Aware-LOLV1 (Xu et al., 2022) [45]	15.64	0.74	10.82	4.75
SNR-Aware-LOLV2real [45]	17.98	0.78	10.52	4.53
SNR-Aware-LOLV2synthetic [45]	18.56	0.79	10.48	4.51
HDRNet (Gharbi et al., 2017) [46]	23.78	0.95	7.54	2.91
DPED (Ignatov et al., 2017) [47]	22.97	0.93	8.24	3.04
Zeng et al. (2020) [48]	23.21	0.93	7.26	3.16
MIRNet (Zamir et al., 2020) [49]	24.21	0.94	7.00	4.69
Chai et al. (2020) [50]	21.39	0.88	9.49	4.72
Proposed Approach	25.77	0.86	6.09	2.48

From the table, it is evident that our proposed approach achieved superior results in terms of PSNR, ΔE° and NIQE

TABLE II

A COMPARATIVE ANALYSIS WAS CONDUCTED ON THE LOL DATASET, WHERE OUR PROPOSED METHOD WAS EVALUATED ALONGSIDE STATE-OF-THE-ART TECHNIQUES. THE BEST AND SECOND BEST RESULTS ARE IN **BOLD** AND UNDERLINE RESPECTIVELY

Approach	HDRNet Ghabi et al. (2017) [46]	DPED Ignatov et al. (2017) [47]	RetinexNet Wei et al. (2018) [38]	KinD Zhang et al. (2019a) [42]	DeepUPE Wang et al. (2019b) [43]	Chai et al. (2020) [50]	Zeng et al. (2020) [48]	MIRNet Zamir et al. (2020) [49]	BacklitNet Lv et al. (2022) [26]	Proposed Approach
PSNR	18.75	19.71	16.77	20.86	19.56	16.56	19.97	24.14	22.79	<u>22.83</u>
SSIM	0.8	0.80	0.55	0.80	0.74	0.79	0.81	<u>0.83</u>	0.85	0.78
Time (ms)	6.78	58.6	94.8	33.2	7.04	11.3	2.50	<u>710</u>	3.86	4.00

and surpasses all the other existing methods. LumiNet not only excels in enhancing backlit images but also yields favourable results for the low-light image enhancement task. Table 2 shows the quantitative comparison between our method and other state-of-the-art techniques on the LOL dataset [38]. We have also tried our model on SICE [36] and Vonikakis et al., [37] backlit datasets. As they do not have ground truth, we utilize the no-reference metric NIQE to conduct quantitative evaluations. Tables 3 and 4 show quantitative analysis in terms of NIQE on the benchmark Vonikakis [37] and SICE [36] datasets. We have obtained superior results in terms of NIQE when compared to the other state-of-the-art methods.

TABLE III
QUANTITATIVE ANALYSIS IN TERMS OF NIQE ON BENCHMARK VONIKAKIS DATASET. THE BEST PERFORMANCE IS HIGHLIGHTED IN **BOLD**

Approach	Vonikakis et al. dataset [37]
Li and Wu (2018) [39]	3.19
ExCNet (Zhang et al., 2019b) [40]	2.08
Buades et al (2020) [17]	2.14
BacklitNet (Lv et al., 2022) [26]	1.96
RetinexNet (Wei et al., 2018) [38]	2.48
KinD (Zhang et al., 2019a) [42]	2.59
DeepUPE (Wang et al., 2019b) [43]	2.05
HDRNet (Gharbi et al., 2017) [46]	1.99
DPED (Ignatov et al., 2017) [47]	2.66
Zeng et al. (2020) [48]	2.54
MIRNet (Zamir et al., 2020) [49]	3.76
Chai et al. (2020) [50]	3.19
Proposed Approach	1.76

TABLE IV
QUANTITATIVE ANALYSIS IN TERMS OF NIQE ON BENCHMARK SICE DATASET. THE BEST PERFORMANCE IS HIGHLIGHTED IN **BOLD**

Approach	SICE dataset
LIME (ACM MM2016) [51]	3.257
BIMEF [52]	5.291
MBLLEN (BMVC2018) [53]	3.756
RetinexNet (BMVC2018) [54]	4.203
SDD (TMM2020) [55]	4.563
Zero-DCE (CVPR2020) [56]	4.000
CSDNet (TNNLS2021) [57]	4.027
RUAS (CVPR2021) [58]	5.158
EnlightenGAN (TIP2021) [59]	3.210
KinD++ (IJCV2021) [42]	3.669
MLLEN-IC (2022) [60]	3.600
Proposed Approach	2.872

2) *Qualitative Analysis:* Qualitative analysis is done to understand the aesthetics of improved images. Qualitative analysis is important as it can reveal the strengths and weaknesses of an enhancement method that may not be evident from quantitative measures alone. In this paper, we have also

conducted a qualitative analysis of our proposed model on the BAID [26] dataset and various state-of-the-art methods. Figs. 7 and 8 display the outcomes of various techniques using several test images from the datasets mentioned above. Our proposed method produces visually pleasing and compelling results in a variety of scenarios. Notably, it significantly enhances the visibility of the images in contrast to the other methods. The substantial emphasis on the Mean Squared Error (MSE) component significantly improves color correction in images, surpassing the capabilities of other methods, as evidenced in Fig. 7. Notably, the proposed method excels in color rendition when compared to state-of-the-art alternatives. In Fig. 7(b), it can be observed that the enhanced outcomes from Li and Wu's (2018) [39] method tend to generate a rim light effect. Furthermore, in Figs. 7(d) and 7(e), the foreground edges produced by and Buades et al.'s (2020) [17] and BacklitNet (Lv et al., 2022) [26] methods are found to be unrealistic. In Fig. 7(e), it is also observed that the backlit enhanced by Lv et al., [26] produces an image that has a warmer image tone and a certain degree of overexposure. This problem is resolved by our suggested method of backlit image enhancement. The results obtained from RetinexNet (Wei et al., 2018) [38] and KinD (Zhang et al., 2019a) [42] exhibit significant colour inconsistencies and irregularities in the results, as given in Figs. 7(f) and 7(g). In Fig. 7(j) DPED (Ignatov et al., 2017) [47] reduces the image contrast. On the other hand, HDRNet (Gharbi et al., 2017) [46] effectively restores backlit images, as shown in Fig. 7(i), but it does not specifically illuminate the dark regions. Our proposed approach produces well-balanced images with optimal exposure, contrast, and intensity. Instance Normalization performs much better than batch normalization, typically in GANs. Batch normalization often results in the blending of different enhancement intensities, leading to a monotonous effect in the shading of certain image regions. This phenomenon is mainly evident in Figs. 7(f), 7(m) (Chai et al., 2020) [50], and 7(k) (Zeng et al., 2020) [48], where the car's lower-left part almost blends into the background trees. Also, in comparison to Fig. 7(k), where a contrast imbalance is evident, and Fig. 7(l), which exhibits a rim light effect, our method avoids these issues. Additionally, Fig. 7(m) is also noticeably underexposed and the enhancement is not satisfactory. In contrast, our method produces an enhanced image that demonstrates improved colour accuracy and uniformity, as shown in Fig. 7(n). When our proposed architecture is tested on a different image, we notice that compared to other state-of-the-art methods like ExCNet [40], SCI-difficult [61], Zero-DCE [56], SNR-aware [62], EnlightenGAN [59] and CLIP-LIT [63] our model has more visually appealing



Fig. 7. Qualitative analysis of state-of-the-art methods on BAID dataset



Fig. 8. Qualitative comparison of the proposed method with the latest state-of-the-art methods

results and completely enhanced backlit areas in the image shown in Fig. 8(h). The spatial attention gate in our model generates spatial maps that effectively preserve the balance between brightness and contrast across various regions in the enhanced image. Clearly, Fig. 8(e) SNR-aware [62] and Fig. 8(f) EnlightenGAN [59] suffer from irregular enhancement in different regions of the image. We can also observe that there is a problem of overexposure in the results obtained from Figs. 8(e) and Fig. 8(f) and the image is underexposed, as can be seen in Fig. 8(b). The presence of a large number of trainable parameters allows for better adaptation to the diverse intensity levels in backlit images, further aided by fine-tuned hyperparameters. Fig. 7, with darker backlit regions compared to Fig. 8, illustrates how the model's capabilities are tied to the

intensity of backlit areas in its training data. This discrepancy is evident in Figs. 8(a), 8(b), and most other images in Fig. 7, which exhibit limited potential to enhance backlit regions, leading to undercorrection. Our method successfully enhances backlit images while avoiding the introduction of artifacts and overexposure. Furthermore, we can also observe that in a particular portion of the image highlighted in the yellow box, the red symbol of a traffic light is clearly visible in our image as compared to the other state-of-the-art methods. Thus, the combination of high-resolution image enhancement capabilities, the use of a 1x1 PatchGAN discriminator, and larger kernels facilitates the preservation of intricate image details. For instance, in Fig. 7, the enhanced number plate of the car and the cycle logo on the red traffic light in Fig. 8

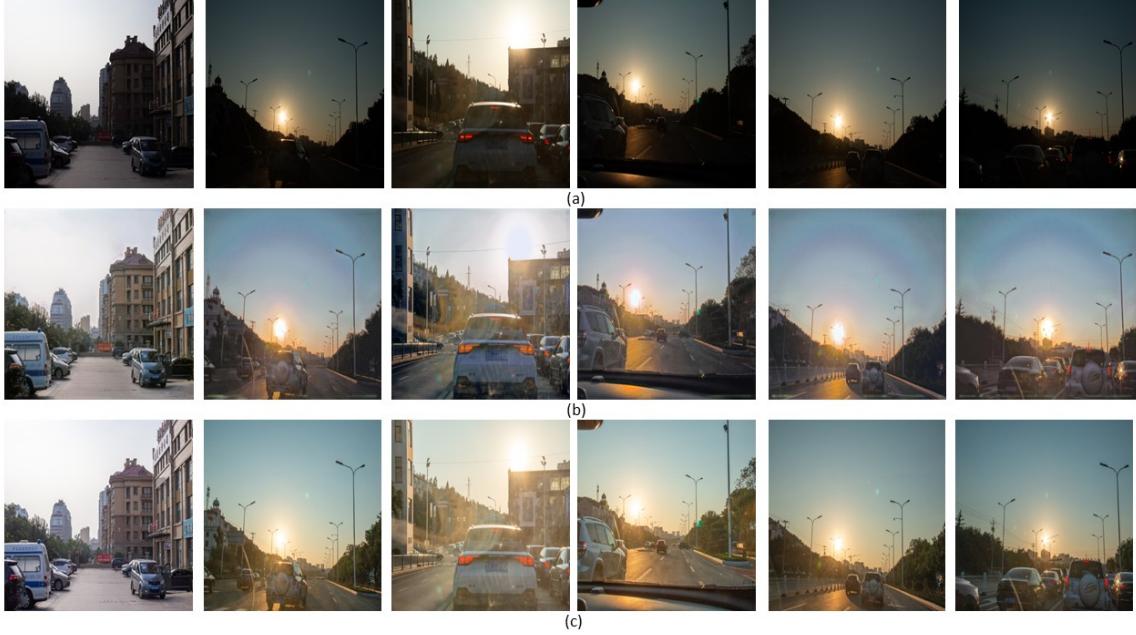


Fig. 9. Backlit enhancement on traffic images (a) Backlit image (b) Backlit enhanced image (c) Groundtruth

testify to the model’s efficacy in maintaining such fine details. The model was also tested on backlit traffic images, as shown in Fig. 9. From the images in Fig. 9, it is observed that the backlit enhanced output is as good as its groundtruth. This is also important when it comes to number plate recognition and vehicle classification tasks, as from the original image it is impossible to recognize number plate and the chances of misclassification of vehicles are also very high. This enhancement task can also be used for anomaly detection in events where the frames are affected by backlight [64]. Overall, our proposed approach offers an effective solution for enhancing backlit images, delivering visually pleasing outcomes with well-balanced luminosity and authentic contrast. It addresses the challenge of preserving image quality while reducing undesirable artifacts, making it suitable for various lighting conditions. By leveraging advanced techniques, our method ensures that the enhanced images maintain their natural appearance without any noticeable distortions or inconsistencies.

3) Computational Efficiency: In addition to evaluating the quantitative comparison of image quality, we also analyze the computational efficiency of the compared techniques across different resolutions. By assessing the running time of these methods, we can gain insights into their speed and efficiency. This analysis is crucial in determining the practical feasibility and real-time applicability of the methods, aiding in selecting the most suitable approach based on specific use cases and computational requirements. To ensure our proposed method is suitable for real-time applications, we have introduced Bilateral Guided Upsampling (BGU) into the architecture. This technique enables us to generate high-resolution enhanced outputs with a comparable runtime to lower-resolution outputs. By using affine matrices to learn the mapping between low-resolution and high-resolution images, we simplify the

generator and remove unnecessary convolution layers that do not significantly contribute to image enhancement. This optimization, combined with the efficiency of BGU, allows our method to outperform many deep learning-based models in terms of memory usage and runtime. Tables 5 and 6 shows the computational efficiency of our approach in terms of runtime and GPU memory usage with and without BGU. We have

TABLE V
COMPUTATIONAL EFFICIENCY OF PROPOSED APPROACH WITHOUT BGU

Image Resolution	512x512	1024x1024	2048x2048
Runtime	4 ms	18 ms	80 ms
GPU Memory Usage	400 MB	1600 MB	6400 MB

TABLE VI
COMPUTATIONAL EFFICIENCY OF PROPOSED APPROACH WITH BGU

Image Resolution	512x512	1024x1024	2048x2048
Runtime	4 ms	6 ms	10 ms
GPU Memory Usage	400 MB	400 MB	400 MB
	+ 20% CPU Utilization	+ 20% CPU Utilization	+ 20% CPU Utilization

also compared the running time of our approach and other state-of-the-art methods, which include Li and Wu (2018) [39], ExCNet (Zhang et al., 2019b [40], Buades et al.(2020) [17] and BacklitNet (Lv et al., 2022) [26] among backlit enhancement methods and other low-light image enhancement methods. RetinexNet (Wei et al., 2018) [38], KinD (Zhang et al., 2019a) [42], DeepUPE (Wang et al., 2019b) [43] and five state-of-the-art general image enhancement methods, namely Chai et al. (2020) [50], Gharbi et al. (2017) [39], Ignatov et al. (2017) [47], Zeng et al. (2020) [48], and Zamir et al. (2020) [49] on BAID dataset and have obtained satisfactory results. We have

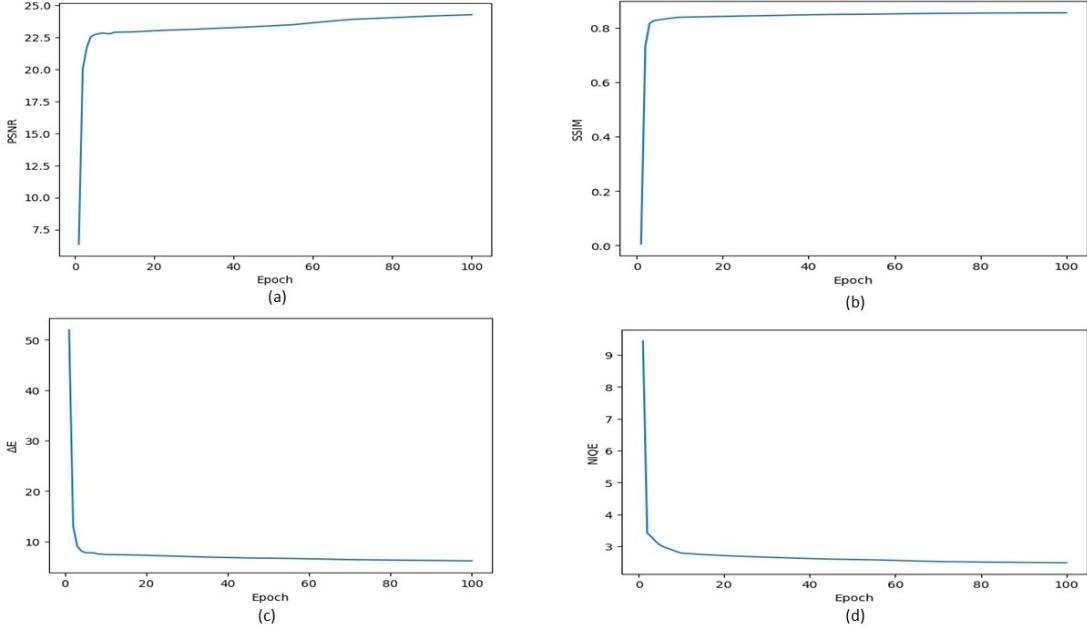


Fig. 10. Graphical representation of evaluation matrices (a) PSNR vs Epoch (b) SSIM vs Epoch (c) ΔE° vs Epoch and NIQE vs Epoch

tested it on 512x512 resolution images. Table 7 shows the running time comparison of our model's performance against other state-of-the-art-methods.

TABLE VII
RUNNING TIME COMPARISON WITH OTHER STATE-OF-THE-ART
APPROACHES

Approach	Resolution (512x512)
Li and Wu (2018) [39]	3.5e4
ExCNet (Zhang et al., 2019b) [40]	9.5e4
Buades et al (2020) [17]	N.A.
BacklitNet (Lv et al., 2022) [26]	3.71
RetinexNet (Wei et al., 2018) [38]	3.84
KinD (Zhang et al., 2019a) [42]	3.76
DeepUPE (Wang et al., 2019b) [43]	7.4
HDRNet (Gharbi et al., 2017) [46]	7.15
DPED (Ignatov et al., 2017) [47]	61.9
Zeng et al. (2020) [48]	2.45
MIRNet (Zamir et al., 2020) [49]	762
Chai et al. (2020) [50]	7.00
Proposed Approach	4.00

E. Ablation Study

Ablation studies are an essential research methodology used to evaluate the individual impact and significance of specific components or factors within a model. By systematically removing or modifying certain elements, researchers can gain insights into the importance of these components in the overall performance of the model. Ablation studies allow for a thorough understanding of the contributions of different sections or features, enabling researchers to make informed decisions about optimization, feature selection, and model variations. Through careful analysis and comparison, ablation studies help researchers refine their models, improve performance, and enhance their overall understanding of the underlying

mechanisms. In this section, we examine the impact of various sections within our method through an ablation study. Specifically, we aim to understand the weight factor's impact on α , β & γ and its significance in the overall effectiveness of the model. In this study, we have varied the values of α , β and γ and obtained different values of PSNR, SSIM, ΔE and NIQE as shown in tables 8, 9 and 10. From table 10, we are able to see that for the values of $\alpha = 100$, $\beta = 1000$ and $\gamma = 20$, we have achieved the highest values in terms of PSNR, ΔE and NIQE of 25.77, 6.09 and 2.48 respectively and these values of weight factor outperform other combinations of values. Furthermore, we have also plotted the graphs of different evaluation matrices with respect to the number of epochs, as shown in Fig. 10.

TABLE VIII
RESULT OF ABLATION STUDY AT DIFFERENT VALUES OF α KEEPING $\beta=1$
AND $\gamma=1$

$\beta = 1,$ $\gamma = 1$	$\alpha = 1$	$\alpha=10$	$\alpha = 500$	$\alpha = 10^3$	$\alpha = 10^4$
PSNR	23.83	25.81	25.82	24.22	23.81
SSIM	0.77	0.76	0.78	0.82	0.79
ΔE° degree	6.71	6.59	6.64	6.98	7.23
NIQE	2.96	2.82	2.63	3.01	3.36

TABLE IX
RESULT OF ABLATION STUDY AT DIFFERENT VALUES OF β KEEPING
 $\alpha=100$ AND $\gamma=1$

$\alpha = 100,$ $\gamma = 1$	$\beta = 1$	$\beta=10^2$	$\beta = 10^3$	$\beta = 5000$	$\beta = 10^4$
PSNR	24.89	25.54	25.23	24.59	23.81
SSIM	0.78	0.79	0.82	0.83	0.85
ΔE° degree	7.45	7.22	6.37	6.28	6.02
NIQE	2.56	2.42	2.35	2.66	3.01

TABLE X
RESULT OF ABLATION STUDY AT DIFFERENT VALUES OF γ KEEPING $\alpha=100$ AND $\beta=1000$

$\alpha = 100$, $\beta = 1000$	$\gamma = 1$	$\gamma=20$	$\gamma = 50$	$\gamma = 100$	$\gamma = 200$
PSNR	25.57	25.77	25.16	24.84	23.25
SSIM	0.79	0.87	0.88	0.89	0.91
$\Delta E \setminus \text{degree}$	6.11	6.09	6.26	6.35	6.39
NIQE	2.56	2.48	2.49	2.76	2.84

IV. CONCLUSION AND FUTURE WORK

In this paper we proposed an approach for image enhancement that shows promising outcomes in comparison to state-of-the-art techniques. Through a comprehensive evaluation, including quantitative metrics and qualitative analysis, we demonstrate that our method achieves improved image quality, enhanced visibility, harmonized brightness, and natural contrast. Additionally, our approach exhibits efficient performance in terms of running time across different resolutions. The ablation study provides valuable insights into the impact of different elements within our approach, further validating the effectiveness of our approach. Overall, our proposed method presents a strong and efficient solution for tasks related to general image enhancement.

Our future work includes executing our model on enhancing backlit-affected videos, which can be further used for several applications like object detection, tracking and surveillance systems.

REFERENCES

- [1] G. Yue, C. Hou, T. Zhou, and X. Zhang, “Effective and efficient blind quality evaluator for contrast distorted images,” *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 8, pp. 2733–2741, 2018.
- [2] N. Singh and A. K. Bhandari, “Principal component analysis-based low-light image enhancement using reflection model,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021.
- [3] K. Xu, H. Chen, X. Tan, Y. Chen, Y. Jin, Y. Kan, and C. Zhu, “Hfmnet: Hierarchical feature mining network for low-light image enhancement,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–14, 2022.
- [4] R. Al Sobbahi and J. Tekli, “Comparing deep learning models for low-light natural scene image enhancement and their impact on object detection and classification: Overview, empirical evaluation, and challenges,” *Signal Processing: Image Communication*, p. 116848, 2022.
- [5] J. Li, X. Feng, and Z. Hua, “Low-light image enhancement via progressive-recursive network,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 11, pp. 4227–4240, 2021.
- [6] D. Ghosal and M. H. Kolekar, “Music genre recognition using deep neural networks and transfer learning,” in *Interspeech*, 2018, pp. 2087–2091.
- [7] M. H. Kolekar, *Intelligent video surveillance systems: an algorithmic approach*. CRC Press, 2018.
- [8] M. H. Kolekar and S. Sengupta, “Bayesian network-based customized highlight generation for broadcast soccer videos,” *IEEE Transactions on Broadcasting*, vol. 61, no. 2, pp. 195–209, 2015.
- [9] G. Thomas, D. Flores-Tapia, and S. Pistorius, “Histogram specification: a fast and flexible method to process digital images,” *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 5, pp. 1565–1578, 2011.
- [10] S. Bose, C. D. Ramesh, and M. H. Kolekar, “Vehicle classification and counting for traffic video monitoring using yolo-v3,” in *2022 International Conference on Connected Systems & Intelligence (CSI)*. IEEE, 2022, pp. 1–8.
- [11] N. Aslam, P. K. Rai, and M. H. Kolekar, “A3n: Attention-based adversarial autoencoder network for detecting anomalies in video sequence,” *Journal of Visual Communication and Image Representation*, vol. 87, p. 103598, 2022.
- [12] H. Cui, J. Li, Z. Hua, and L. Fan, “Progressive dual-branch network for low-light image enhancement,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–18, 2022.
- [13] P. Hambarde, S. Murala, and A. Dhall, “Uw-gan: Single-image depth estimation and image enhancement for underwater images,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.
- [14] S. Liu and Y. Zhang, “Detail-preserving underexposed image enhancement via optimal weighted multi-exposure fusion,” *IEEE Transactions on Consumer Electronics*, vol. 65, no. 3, pp. 303–311, 2019.
- [15] C. Ma, S. Zeng, and D. Li, “A new algorithm for backlight image enhancement,” in *2020 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*. IEEE, 2020, pp. 840–844.
- [16] T. Trongtirakul, W. Chiracharit, and S. S. Agaian, “Single backlit image enhancement,” *IEEE Access*, vol. 8, pp. 71940–71950, 2020.
- [17] A. Buades, J.-L. Lisani, A. B. Petro, and C. Sbert, “Backlit images enhancement using global tone mappings and image fusion,” *IET Image Processing*, vol. 14, no. 2, pp. 211–219, 2020.
- [18] Z. Li, K. Cheng, and X. Wu, “Soft binary segmentation-based backlit image enhancement,” in *2015 IEEE 17th International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2015, pp. 1–5.
- [19] Z. Li and X. Wu, “Learning-based restoration of backlit images,” *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 976–986, 2017.
- [20] X. Ren, W. Yang, W.-H. Cheng, and J. Liu, “Lr3m: Robust low-light enhancement via low-rank regularized retinex model,” *IEEE Transactions on Image Processing*, vol. 29, pp. 5862–5876, 2020.
- [21] M. Akai, Y. Ueda, T. Koga, and N. Suetake, “A single backlit image enhancement method for improvement of visibility of dark part,” in *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021, pp. 1659–1663.
- [22] Y. Ueda, D. Moriyama, T. Koga, and N. Suetake, “Histogram specification-based image enhancement for backlit image,” in *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2020, pp. 958–962.
- [23] M. Zhao, D. Cheng, and L. Wang, “Backlit image enhancement based on foreground extraction,” in *Twelfth International Conference on Graphics and Image Processing (ICGIP 2020)*, vol. 11720. SPIE, 2021, pp. 366–373.
- [24] M. Lecca, “Enhancing backlight and spotlight images by the retinex-inspired bilateral filter super-b,” in *International Joint Conference on Computer Vision, Imaging and Computer Graphics*. Springer, 2021, pp. 328–347.
- [25] H.-S. Chou, H.-Y. Cheng, J.-X. Qiu, T.-K. Chi, T.-Y. Chen, and S.-L. Chen, “Retinex based on weaken factor with truncated agewd for backlight image enhancement,” in *2022 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2022, pp. 1–5.
- [26] X. Lv, S. Zhang, Q. Liu, H. Xie, B. Zhong, and H. Zhou, “Backlitnet: A dataset and network for backlit image enhancement,” *Computer Vision and Image Understanding*, vol. 218, p. 103403, 2022.
- [27] S. Takao, “Zero-shot image enhancement with renovated laplacian pyramid,” in *European Conference on Computer Vision*. Springer, 2022, pp. 721–737.
- [28] R. Khan, Y. Yang, Q. Liu, and Z. H. Qaisar, “Divide and conquer: Ill-light image enhancement via hybrid deep network,” *Expert Systems with Applications*, vol. 182, p. 115034, 2021.
- [29] H. Zeng, J. Cai, L. Li, Z. Cao, and L. Zhang, “Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 4, pp. 2058–2073, 2020.
- [30] W. Shi, F. Tao, and Y. Wen, “Structure-aware deep networks and pixel-level generative adversarial training for single image super-resolution,” *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–14, 2023.
- [31] K. Cheng, R. Tahir, L. K. Eric, and M. Li, “An analysis of generative adversarial networks and variants for image synthesis on mnist dataset,” *Multimedia Tools and Applications*, vol. 79, pp. 13725–13752, 2020.
- [32] L. Ruthotto and E. Haber, “An introduction to deep generative modeling,” *GAMM-Mitteilungen*, vol. 44, no. 2, p. e202100008, 2021.
- [33] Z. Niu, G. Zhong, and H. Yu, “A review on the attention mechanism of deep learning,” *Neurocomputing*, vol. 452, pp. 48–62, 2021.
- [34] C. Xu, J. Feng, P. Zhao, F. Zhuang, D. Wang, Y. Liu, and V. S. Sheng, “Long-and short-term self-attention network for sequential recommendation,” *Neurocomputing*, vol. 423, pp. 580–589, 2021.
- [35] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

- [36] J. Cai, S. Gu, and L. Zhang, "Learning a deep single image contrast enhancer from multi-exposure images," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 2049–2062, 2018.
- [37] V. Vonikakis, R. Kouskouridas, and A. Gasteratos, "On the evaluation of illumination compensation algorithms," *Multimedia Tools and Applications*, vol. 77, pp. 9211–9231, 2018.
- [38] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," *arXiv preprint arXiv:1808.04560*, 2018.
- [39] Z. Li and X. Wu, "Learning-based restoration of backlit images," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 976–986, 2017.
- [40] L. Zhang, L. Zhang, X. Liu, Y. Shen, S. Zhang, and S. Zhao, "Zero-shot restoration of back-lit images using deep internal learning," in *Proceedings of the 27th ACM international conference on multimedia*, 2019, pp. 1623–1631.
- [41] Z. Liang, C. Li, S. Zhou, R. Feng, and C. C. Loy, "Iterative prompt learning for unsupervised backlit image enhancement," *arXiv preprint arXiv:2303.17569*, 2023.
- [42] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proceedings of the 27th ACM international conference on multimedia*, 2019, pp. 1632–1640.
- [43] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 6849–6857.
- [44] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 1780–1789.
- [45] X. Xu, R. Wang, C.-W. Fu, and J. Jia, "Snr-aware low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 17714–17724.
- [46] M. Gharbi, J. Chen, J. T. Barron, S. W. Hasinoff, and F. Durand, "Deep bilateral learning for real-time image enhancement," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, pp. 1–12, 2017.
- [47] A. Ignatov, N. Kobyshev, R. Timofte, K. Vanhoey, and L. Van Gool, "Dslr-quality photos on mobile devices with deep convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 3277–3285.
- [48] H. Zeng, J. Cai, L. Li, Z. Cao, and L. Zhang, "Learning image-adaptive 3d lookup tables for high performance photo enhancement in real-time," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 4, pp. 2058–2073, 2020.
- [49] A. A. K. S. H. M. K. F. Y. M. S. L. Zamir, S.W., "Learning enriched features for real image restoration and enhancement," in *In: ECCV*, 2020, pp. 3277–3285.
- [50] Y. Chai, R. Giryes, and L. Wolf, "Supervised and unsupervised learning of parameterized color enhancement," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 992–1000.
- [51] X. Guo, "Lime: A method for low-light image enhancement," in *Proceedings of the 24th ACM international conference on Multimedia*, 2016, pp. 87–91.
- [52] Z. Ying, G. Li, and W. Gao, "A bio-inspired multi-exposure fusion framework for low-light image enhancement," *arXiv preprint arXiv:1711.00591*, 2017.
- [53] E. P. Bennett and L. McMillan, "Video enhancement using per-pixel virtual exposures," in *ACM SIGGRAPH 2005 Papers*, 2005, pp. 845–852.
- [54] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," *arXiv preprint arXiv:1808.04560*, 2018.
- [55] S. Hao, X. Han, Y. Guo, X. Xu, and M. Wang, "Low-light image enhancement with semi-decoupled decomposition," *IEEE transactions on multimedia*, vol. 22, no. 12, pp. 3025–3038, 2020.
- [56] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 1780–1789.
- [57] L. Ma, R. Liu, J. Zhang, X. Fan, and Z. Luo, "Learning deep context-sensitive decomposition for low-light image enhancement," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 10, pp. 5666–5680, 2021.
- [58] R. Liu, L. Ma, J. Zhang, X. Fan, and Z. Luo, "Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10561–10570.
- [59] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "Enlightengan: Deep light enhancement without paired supervision," *IEEE transactions on image processing*, vol. 30, pp. 2340–2349, 2021.
- [60] G.-D. Fan, B. Fan, M. Gan, G.-Y. Chen, and C. P. Chen, "Multiscale low-light image enhancement network with illumination constraint," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 11, pp. 7403–7417, 2022.
- [61] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward fast, flexible, and robust low-light image enhancement," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5637–5646.
- [62] X. Xu, R. Wang, C.-W. Fu, and J. Jia, "Snr-aware low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 17714–17724.
- [63] Z. Liang, C. Li, S. Zhou, R. Feng, and C. C. Loy, "Iterative prompt learning for unsupervised backlit image enhancement," *arXiv preprint arXiv:2303.17569*, 2023.
- [64] N. Aslam and M. H. Kolekar, "Unsupervised anomalous event detection in videos using spatio-temporal inter-fused autoencoder," *Multimedia Tools and Applications*, vol. 81, no. 29, pp. 42457–42482, 2022.