IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 21, 2022, accepted May 19, 2022, date of publication May 30, 2022, date of current version June 6, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3178698

D2BGAN: A Dark to Bright Image Conversion Model for Quality Enhancement and Analysis Tasks Without Paired Supervision

JHILIK BHATTACHARYA[1], (Member, IEEE), SHATRUGHAN MODI[1], LEONARDO GREGORAT[2], AND GIOVANNI RAMPONI[2], (Life Senior Member, IEEE)
[1]Thapar Institute of Engineering and Technology, Patiala 147004, India
[2]Department of Engineering and Architecture, University of Trieste, 34127 Trieste, Italy

Corresponding author: Jhilik Bhattacharya (jhilik@thapar.edu)

This work was supported in part by the Department of Science and Technology (DST), India, Tide Grant.

What is the loss that you have used? (contextual loss criterion?

ABSTRACT This paper presents an image enhancement model, D2BGAN (Dark to Bright Generative Adversarial Network), to translate low light images to bright images without a paired supervision. We introduce the use of geometric and lighting consistency along with a contextual loss criterion. These when combined with multiscale color, texture and edge discriminators prove to provide competitive results. We performed extensive experiments using benchmark datasets to visually and objectively compare our results. We observed the performance of D2BGAN on real-time driving datasets that are subject to motion blur, noise, and other artifacts. We further demonstrated that our enhanced images can be profitably used in image-understanding tasks. Images processed using our technique obtain the best or second best average scores for three different image quality evaluation methods on the Naturalness Preserved Enhancement (NPE), Low Light Image Enhancement (LIME), Multi-Exposure Image Fusion (MEF) benchmark datasets. Best scores are also obtained on the LOw-Light (LOL) test set and on Berkeley Driving Dataset (BDD) images processed with D2BGAN. Face detection tasks on the DarkFace benchmark dataset show an mAP (mean Average Precision) improvement from 0.209 to 0.301 when images are processed using D2BGAN. mAP further improves to 0.525 when finetuning techniques are adopted.

INDEX TERMS Image enhancement, generative adversarial network, unpaired supervision.

I. INTRODUCTION

Image enhancement is a prerequisite for many computer vision-based image understanding tasks. In particular, it is crucial to enhance low light or dark images to obtain images which not only have better image aesthetic quality but can also be suitably processed for object detection and face detection. The Dark to Bright (D2B) task by itself is one of the earliest studied domains in computer vision, but the continuous evolution of computational resources and deep learning architectures has generated a paradigm shift towards the latter approach. A literature review shows three main ways in which the problem is approached: histogram equalization [1]–[6], Retinex-based [7]–[12] and machine learning-based [13]–[19].

The associate editor coordinating the review of this manuscript and approving it for publication was Sudhakar Radhakrishnan.

The primary challenge in dealing with low-light image enhancement tasks is that there are many noise sources in the acquisition of poorly lit scenes. These include readout, photon shot, dark current, and fixed pattern noises, in addition to photon response non-uniformities. The noise level increases while treating lightness and contrast of low-light images, more so in the case of compressed-dynamics images. Applying a denoising filter prior to the light enhancement results in blurring, while the reverse causes noise amplification as shown in Fig. 1. Hence, addressing denoising and low-light enhancement problems simultaneously using a learning-based approach appears to be the optimal choice.

The success of deep-learning enhancement models depends on the availability of large-scale annotated data. For the present problem, there is a requirement for paired images such that a low-light image serves as the input while its brighter counterpart serves as the target image. The Adobe 5k

FIGURE 1. (a) Original dark image, (b) processed image with LIME (noise amplified), and (c) processed image with MBLLEN (artifacts due to smoothing).

dataset [20] served as a benchmark used by many researchers for this task. This dataset provides an original low-light image; retouched by five different photographic experts, one of which was used as the target of supervised training. Some researchers have used a synthetic paired set for the same by intentionally transforming a bright image to generate its dark counterpart. This transformation is global, and its suitability for challenging real-time images is debatable.

Ideally, a training procedure that does not require paired supervision is the most appropriate. Very few studies have reported on low-light image enhancement using unpaired images [19], [21]. Moreover, even though image-to-image style translation without paired supervision is a popular approach [22]–[24], few have considered the mapping of low-light images to bright-light images as a style transfer problem.

A. MOTIVATION

In this study, we consider the low-light image enhancement task as an image style transfer problem. We trained a deep neural architecture using unsupervised image pairs. This promotes the use of available real-time dark and bright images without the need for pairwise annotations or synthetic treatments. We resort to Generative Adversarial Networks(GANs), using two encoders and two decoders. The use of 2-way GANs is preferred for two main reasons: first, the image generation principle via adversarial losses allows the use of an unsupervised training framework for deep-learning tasks, without annotated data; and second, even though the current work reports a low-to-bright light image transfer problem, the developed architecture can also be used to generate synthetic low-light images that can be used for fine-tuning in image-understanding tasks (e.g., face/object detection and segmentation). Moreover, it is essential to have a reference frame for the loss computation without the need for paired supervision to ensure that the model can be generalized well over any input distribution. However, the unsupervised image-to-image translation problem is challenging because it aims to recover a joint probability given a marginal probability. For this purpose, one of the most common approaches is to add a cycle-consistency constraint along with adversarial losses. This bijective constraint becomes restrictive in cases such as the task at hand, where the low light image domain may contain less information than its bright counterpart. To handle this situation we use a geometric consistency constraint which ensures that an image from one domain and its geometrically transformed version generates the same image in the other domain. Additionally, a contextual loss constraint ensures that the context and semantics of the images are preserved. We used three discriminators to learn color, texture, and edges separately. Together these ensure that the generated images have consistent edges and textures without false structure addition during the translation. Another important factor to be considered is that, contrary to most low-light image enhancement tasks which improve the image quality visually by touching the image, we aim towards enhancing images captured at very poor lighting with onboard moving cameras such that the images are suitable for image-understanding tasks. This highlights the need for edge and texture discriminators, because color-only discriminators are not suitable for preventing the introduction of false structures while enhancing these very low-light images. This study provides the following main contributions: (i) it poses the low-to-bright light enhancement problem as an unpaired image translation problem; (ii) it introduces the use of a geometric and an illumination consistency constraint in the images to prevent the generator from creating false structures, as is typical in cycle-gan-based architectures; (iii) it uses a contextual loss to maintain semantic similarity between the low-light image and its generated bright counterpart; (iv) it uses multiscale color, texture and edge discriminators per domain to ensure that the discriminator is able to force the generator to improve edge and texture information in the generated image. This is demonstrated by the fact that D2BGAN is able to generate images that obtain better quality scores as compared to other GAN-based image enhancement methods; (v) unlike typical image enhancement algorithms which improve the overall image quality but fail to preserve features important for face/object detection tasks, the D2BGAN enhancement proves to be effective for image-understanding tasks. Face detection tasks on D2BGAN-enhanced DarkFace images show an mAP improvement of 10% without finetuning and 32% after fine-tuning.

The remainder of this paper is organized as follows. We provide a study of related work in Sec. II. The proposed technique is discussed in detail in Sec. III. Experimental results and conclusions are provided in Secs. IV and V respectively.

II. RELATED WORK

Histogram based enhancement aims to improve the image quality by modifying the image histogram. This can

result in an overstretched contrast that lacks naturalness. Various techniques have been developed using this framework. Brightness preserving dynamic histogram equalization (BPDHE) [25] uses a histogram equalization (HE) on a dynamic range expanded version of sub-histograms obtained by the local maxima of the input image histogram. Flattest histogram specification with accurate brightness preservation (FHSABP) [1] solves a convex optimization problem to obtain the flattest target histogram with the brightness preservation constraint. Because of their brightness preservation, BPDHE and FHSABP are not suitable for low-light image enhancement, but prevent overstretching. Moreover, these methods do not consider the relationships between adjacent pixels, whereas methods such as Histogram Modification Framework (HMF) [2], Contrast Enhancement Based on Genetic Algorithm (CEBGA) [3] and Differential HE for Color Images (DHECI) [4] do. In HMF [2] the target histogram is obtained via a parametric optimization problem, involving the local variance of the pixels. In CEBGA [3] a genetic algorithm is adopted to enhance the contrast of the images by modifying the histogram, using the number of edges in the enhanced image as a fitness function. DHECI [4] performs histogram equalization on the differential intensity histogram and differential saturation histogram from the HSI color space. More advanced techniques that embed contextual information include, Contextual and Variational Contrast enhancement (CVC) [5] and Layered Difference Representation (LDR) [6], which use 2D histograms. In the former, a target 2D histogram is obtained through an optimization problem involving the uniform histogram, weighted input histogram and smoothing term. In the latter, a 2D histogram of gray-level differences between neighboring pixels is used to generate a mapping function that increases the difference of frequently adjacent pixel values.

Retinex based methods assume that an image can be pixel-wise decomposed into reflectance and illumination. In [26] a classic approach was proposed, where the lightness was first decomposed in reflex lightness and ambient illumination. Reflectance was then extracted from reflex lightness, while ambient illumination was log-transformed and used for the output image. A widely used framework for the Retinex based enhancement methods is the fusion-based framework, as in Fusion-based Enhancing Method for Weakly Illuminated images (FEMWII) [7] and Single Backlit Image Enhancement (FMSBIE) [8]. Both methods estimate the illumination using the pixel-wise maximum in the RGB color space, which is used to obtain the reflectance. Three different modified illuminations were then obtained using different techniques to improve the brightness and contrast. The illumination maps were then fused with a multiscale approach. The multiscale techniques are also used to extract reflectance and illumination: [27] used three guided filters with different radius are used to get the reflectance and the illumination on the intensity layer of the HSI color space. Another approach for illumination estimation is proposed in Low-light IMage Enhancement (LIME) [9], where an optimization problem is used to obtain a smooth, structure-preserving illumination map, which is then enhanced through a gamma correction. As noise is critical in poorly illuminated images in LIME, a denoising technique is also proposed where the output reflectance is the linear combination of the original reflectance and a block-matching and 3D filtered (BM3D) method, using illumination as the weight to prevent oversmoothing of bright areas. In [10], a hybrid regularized variational model is proposed to extract the illumination map, which is later enhanced through an adaptive gamma correction, while the reflectance map gets optimized thanks to a guided filter; such enhanced image is also refined through a deep learning-based denoising. In [26], [7], [8], [27], [9], and [10], reflectance was obtained by estimated illumination. More advanced techniques, such as Simultaneous Reflectance and Illumination Estimation (SRIE) [11] and Joint Intrinsic-Extrinsic Prior Model for Retinex (JIEPMR) [12], jointly decompose reflectance and illumination. The former uses a variational model and the latter uses an iterative approach with the derivatives of the reflectance and illumination. This allows to obtain both a structure-preserving smooth illumination and a shape-and-texture-preserving reflectance. While the Retinex theory has proven to be suitable for low-light image enhancement, this approach suffers from color distortion and hand-crafted illumination manipulation. Learning-based methods can overcome these problems, while providing better denoising, which is crucial in low-light images.

Data learning methods include a variety of methods from deep learning, such as autoencoders, deep Convolutionl Neural Networks (CNNs) and GANs, all of which have proved to be suitable for enhancing image quality and brightness.

In Low-Light Network (LLNet) [13], Global iLlumination-Aware and Detail-preserving Network (GLADNet) [14] and Low-Light Encoder-Decoder Network (LLED-Net) [15], denoising *autoencoders* are used. LLED-Net employs only a single autoencoder, which embeds some residual modules, but uses the SSIM image quality index as a loss. LLNet also uses only an autoencoder, but is composed of a five-layer stacked sparse denoising autoencoder (SSDA). In contrast, in GLADNet the autoencoder is not directly applied to the original image and is followed by a convolutional network, but is trained using only the absolute difference from the target. Despite good results, the simplicity of these methods limits their capability compared to more complex learning methods. In [28] multiple autoencoders were used to work separately on the content and edges of the image while also using CNN, a Recurrent Neural Network (RNN) and a more advanced loss function, including a discriminator and a VGG-16 net. An autoencoder was used for DEEP Underexposed Photo Enhancement DeepUPE [29] and Low-Light Image Enhancement Network (LLIE-net) [30]. In the former an image-to-illumination mapping used for a Retinex-fashioned enhancement is learned, while considering reconstruction loss, color loss and smoothness of the illuminance. In the latter an autoencoder is used with a discrete

wavelet transorm (DWT) and a multiscale approach. In [31], an end-to-end U-Net was used, with a 3D convolution to deal with videos, preventing flickering in consecutive frames, and using the GRBG components from the camera sensor as an input. In Multi-Branch Low-Light Enhancement Network (MBLLEN) [32], a CNN was applied to the input image. Subsequently, the output of each layer was fed to the independent autoencoders. The obtained output image is a weighted sum with learnable weights of the autoencoders output,considering structural, contextual and regional losses for the training. While autoencoders and U-Net are convenient for image-to-image nets, methods such as those in [33] and [34] use only deep neural networks. In particular, they use residual *convolutional neural networks* to mitigate the vanishing gradient problem in deep networks. In [34], the CNN enhanced images from low-end devices to professional Digital Single Lens Reflex (DSLR) quality, using a dataset with multiple acquisitions of the same scene from different devices for training, which is difficult to obtain. [33] uses special-designed residual CNN modules to create a deep CNN, but was trained using only the Structural Similarity Index Measure (SSIM). Apart from [29], RetinexNet [35], Kindling the Darkness (KinD) [36], his successor KinD++ [16] and Retinex-inspired unrolling with architecture search (RUAS) [37] augment Retinex theory, outperforming classic Retinex based methods. [35], [36] and [16] used a CNN to decompose reflectance and illumination components. In both cases, denoising takes advantage of the illumination map, as in [9], to improve the results over plain denoising. In [35], BM3D was used for denoising whereas in [36] a specific CNN was adopted, which was later improved in [16] with multiscale illumination consideration. An encoder-decoder structure was used for illumination adjustment [35] while [36] and [16] adopted a CNN with a free parameter to modify the adjustment ratio. Recently, normalizing flows have been used as an alternative to GANs to address the low-light image enhancement problem, as in low-light flow-based image enhancement (LLFlow) [17]. The normalizing flows permit to model the conditional distribution of normally exposed images and then exploit the network invertibility to enhance low-light images. Despite the promising results of this approach, normalizing flows are typically inefficient, with a high-dimensionality latent space and not expressive when compared to other architectures, making them complex to manipulate and take advantage of.

A common critical aspect of these methods is *data availability*. Since these methods are supervised, they require paired low/normal light images of the same scene. A widely used way to obtain them is to artificially globally darken normal light images, or to locally darken bright images to perform better in partially illuminated images [38], however this artificial manipulation affects the naturalness of the results. Recently, methods such as [35] and [16] have used datasets with images of the same scenes with different exposures. This proved to be effective, but such datasets were scarce and tedious to create. To overcome the issue of data availability, several unsupervised [18], [37], [39] and unpaired supervised [19], [21] methods have been developed. Zero-reference deep curve estimation (ZDCE) [18] is an unsupervised low-light image enhancer that uses a CNN to obtain pixel-wise enhancement curves to adjust the input image, using non-reference loss functions for training. RUAS [37] and [39] are promising unsupervised enhancers as well: the former uses reference-free losses on a Retinex-inspired model and a neural architecture search (NAS) to generate the architecture from a simpler heuristical design, whereas the latter uses a simple encoder-decoder network, with a novel unsupervised loss function based on the bright channel prior. Both EnlightenGAN (EGAN) [21] and Unsupervised image Enhancement GAN (UEGAN) [19] used GANs to obtain a light enhancement with unpaired supervision, using a decoder-encoder for generation. In EGAN a local and global discriminator were used whereas in UEGAN only one multiscale discriminator was used. The promising results of EGAN and UEGAN, together with the availability of unpaired low/normal-light images, make GANs an appealing approach for low light image enhancement techniques. [40] contributed towards increasing nighttime visibility without using paired supervision.

In general, learning-based approaches provide better local illumination enhancement with lower noise impacts than both Retinex-based and Histogram-based approaches. The effects of color distortions are also significantly smaller in learning-based techniques. While different learning-based approaches differ in terms of a) supervision (unsupervised vs. supervised), b) base architecture (CNN/autoencoder/hybrid), c) model structure (Retinex-based, GAN-based, image-image based), (Fig 2), the following main observations are made: (i) the unsupervised or unpaired-supervised learning approach is preferred to address the data availability issue; (ii) for very poorly illuminated and low-resolution images, the Retinex decomposition becomes quite challenging; (iii) a deep CNN architecture with skip connections better resolves vanishing gradient issues, particularly for sparse contents as may be the case for poor quality low-light images; (iv) no reference losses can lead to poor generalization and oversaturation in some cases. These factors guide our decision to choosing a 2-domain GAN model which will provide a reference for loss computation even with unpaired supervision. The underlying encoder/decoder architecture in the model promotes the use of residual connections to ensure improved model convergence. Our contribution has some similarities with EGAN and UEGAN, because it uses a GAN architecture with unpaired supervision. However we focused towards presenting a 2-way architecture with better generalization across different distributions, which is comparatively difficult with no reference losses. To ensure better image generation even from insufficient representations, and to prevent the generation of false structures, we add consistency constraints to the generators. Unlike EGAN and UEGAN, we did not use global attention. We used separate discriminators for color, texture and edge. The use of discriminators for content learning,

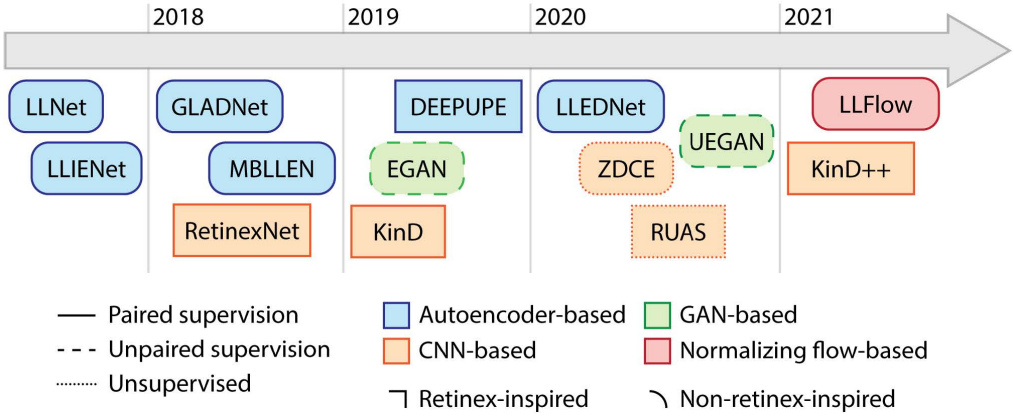


FIGURE 2. A diagrammatic representation of different low light enhancement techniques.

instead of separate edge and content CNNs, is motivated by the fact that the image may be too poor to generate appropriate edge information for learning. We used cycle consistency, geometric and illumination consistency with contextual loss instead of perceptual losses. It has been observed via extensive ablation studies that the domain inconsistency problem and the inability of cycle consistency alone to address the problem is solved to a great extent by using texture and edge discriminators along with geometric consistency losses.

III. METHODOLOGY

CycleGAN [41] was one of the first techniques used for unsupervised image translation, in which a cycle consistency loss was used to learn the semantic dissimilarity between the two transfer domains. This replaced the direct paired loss. Although this technique has been used for image style transfers, to the best of our knowledge ours is the first attempt to use it for an image enhancement problem. This translation is an open-ended problem, as a single image in a domain may result in multiple images in the other domain. Although there are techniques that deal with the multimodal nature of the problem [24] by random sampling or by using more than one target, we focus on generating a single translation at a time. In addition to cycle consistency, many regularizations are used on the generators to enable the creation of real-generated images while using unpaired learning. These include penalizing the distance in the latent space, exploiting a perceptual loss, and forcing the generators to be close to the identity function.

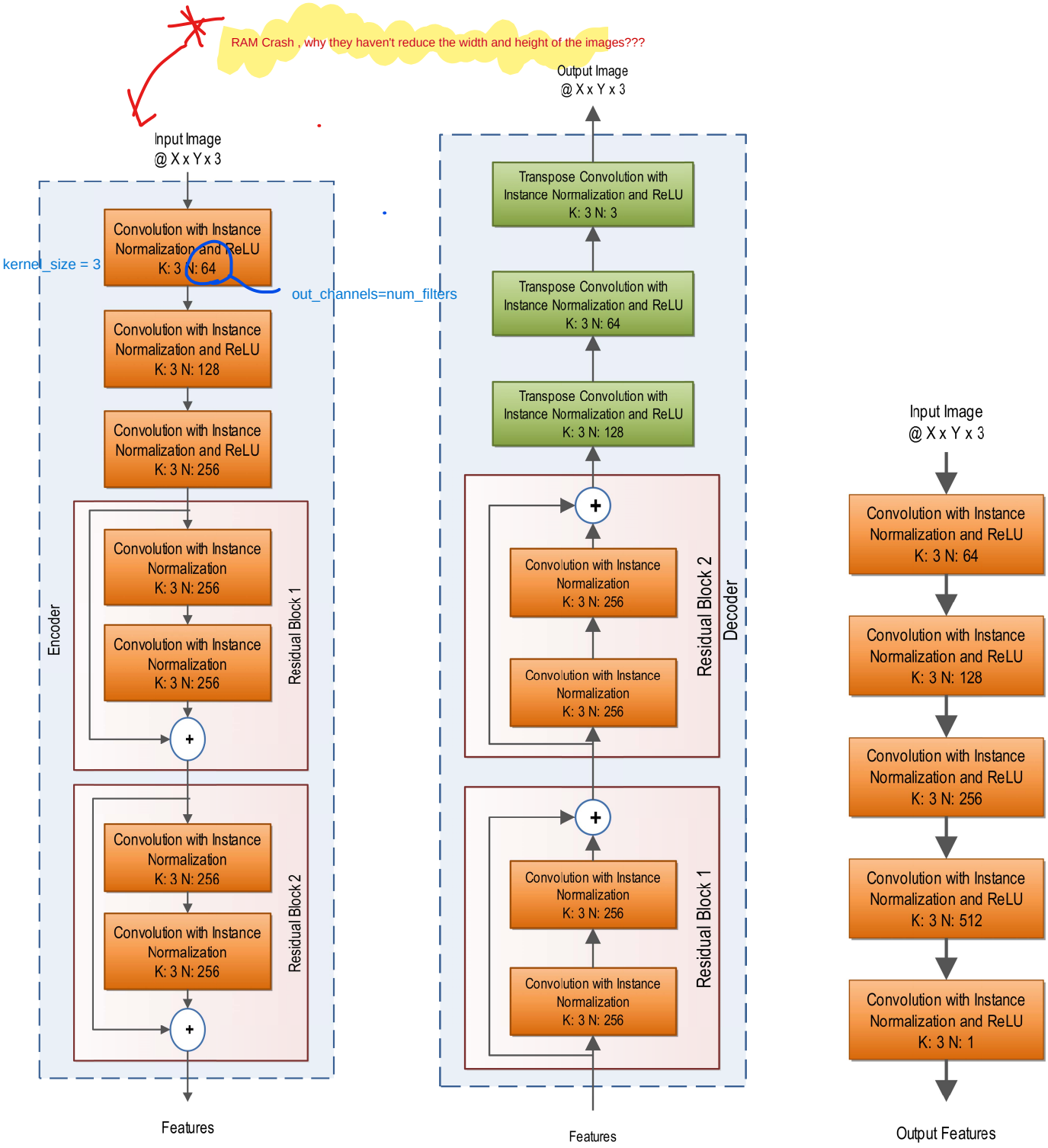
In this study we relax the number of regularizations on the generator while ensuring that the generator is able to learn inverse mappings. We aim to map images in a low-light domain, L_x, to bright images B_y. Both L_x and B_y consist of a finite number of samples. An image X_{real} belonging to L_x can be mapped to Y_{fake} in domain B_y using an encoder E_x and a decoder D_y, $Y_{fake} \rightarrow D_y(E_x(X_{real}))$. Similarly, $X_{fake} \rightarrow D_x(E_y(Y_{real}))$. The cycle consistency can be computed using Y_{fake} to regenerate X_{recon} using $D_x(E_y(Y_{fake}))$.

meaning???

maybe try to add this to the Autoencoder as well?

We used standard encoder and decoder architectures made up of sequentially stacked convolution blocks with instance normalization and rectification linear units. Multiple residual blocks were used in the study. A sample structure is shown in Fig 3. As the cycle consistency alone is not suitable for dealing with the inconsistency between the low light and bright light domains and can introduce false structures, we impose a lighting and geometric consistency. This adds extra leverage to learning the mapping between an information-rich bright image and its low light counterpart. To differentiate the general cycle consistency from the lighting and geometric consistencies used, we refer to the cycle consistency as *cycle reconstruction loss* whereas the geometric and illumination consistencies are referred to as *consistency loss*. The idea of illumination consistency is supported by the fact that two same low light images with slightly different illumination can generate the same enhanced image. This served three purposes. The network is guided to a single translation at a time, thereby controlling the open-ended nature of the problem to some extent. in addition, the generalization capability of the network is improved, and false structure generation is reduced. The geometric consistency acts in a similar fashion and ensures that the false data generation tendency of the generator is checked. These two consistencies consider the structure and illumination factors of the image, improving the overall illumination without saturation while preserving the structure at the same time. To set up the geometric and lighting consistency constraints, we transform X into X_g and X_l, where X_g is a 90 degrees rotation of X and X_l is a gamma transformation of X. When Y_{fake} is generated, it should be similar to Y_l_{fake} and $Y\_g^{-1}_{fake}$. The inverse mapping refers to the inverse rotation of the reconstructed image. This ensures that the generator does not add further artifacts to the image during transformation.

Two discriminators Dc^L and Dc^B are used for the two domains. Furthermore, each discriminator is divided into three parts: Dc_{xc}, Dc_{xt} and Dc_{xe}. These aim to discriminate the color (xc), texture (xt) and edge (xe) differences between



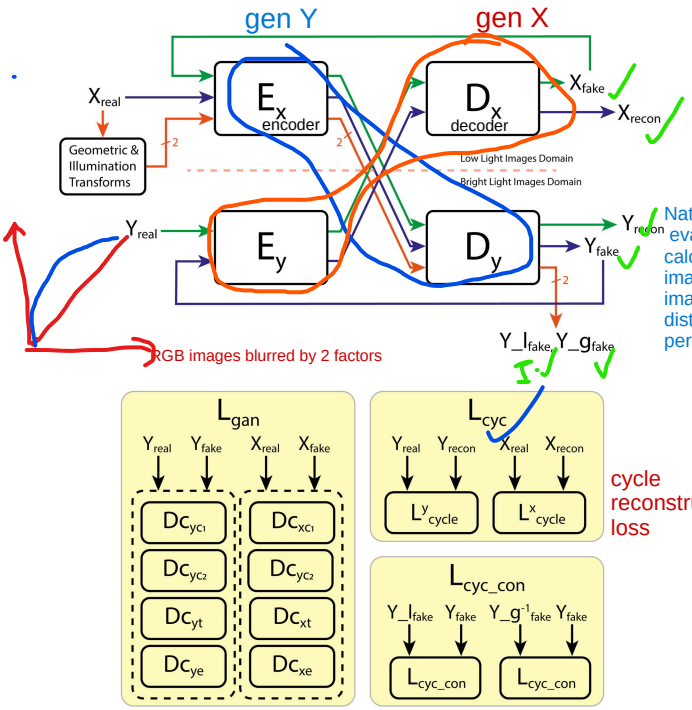
(a) Encoder Architecture of Generator (b) Decoder Architecture of Generator (c) Discriminator Architecture

FIGURE 3. Architecture of generator and discriminator of the GAN. Numbers after @, K and N represents the vector dimensions, kernel size and the number of kernels, respectively.

the real and generated images in L_x domain. The inclusion of these discriminators facilitates the learning of color, texture and edge distributions from unpaired images. The color discriminator uses blurred RGB images. It was later observed via experiments that learning the color distributions via a blurred RGB image, and separate edge and texture patterns from the gradient image and grayscale image provides better edge and texture learning than the use of a single color image.

Here, does the edge and texture both corresponds to the grayscale image & gradient image or both of these images?

a table where you systematically remove parts of the input to see which parts of the input are relevant to the networks output



cycle reconstruction loss

FIGURE 4. Block diagram of D2BGAN: in the upper half the encoder and decoder interconnection is presented with a color-coded dataflow, starting from a low light image X_{real} and a bright image Y_{real}. In the lower half the three contributions to the loss functions are represented.

TABLE 1. Ablation study using different consistency losses: cycle (C), geometric (G), illumination (I), and contextual (Cn) loss along with the three discriminators color (cd), texture (td), edge (ed).

Experiment	NIQE	BRISQUE
cd+C	13.06	41.52
cd+C+GI	5.57	33.22
cd+C+GI+Cn	4.33	25.44
cd+ed	4.33	29.83
cd+ed+C+GI	5.03	32.49
cd+ed+C+GI+Cn	6.68	27.52
cd+ed+td	3.65	22.62

Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE

Natural image quality evaluator (NIQE) calculates the no-reference image quality score for images that may be distorted or of low perceptual quality

The main difference here is that we use three discriminators instead of a single discriminator in each domain. For the 2-domain problem we accumulated losses from six discriminators in total. Hence, X will be replaced by $\{xc, xt, xe\}$ and Y will be replaced by $\{yc, yt, ye\}$. We have $L_{gan}(X, Y)$ and $L_{gan}(Y, X)$ for the 2-domain problem. L_{gan} in Eq. 1 hence is used as shown in Eq. 3. In addition, D_c represents three instead of a single discriminator, i.e. $D_c \in \{Dc_{xc}, Dc_{xt}$ and $Dc_{xe}\}$. Input to Dc_i is i; for example, input to Dc_{xc} is xc.

The cycle reconstruction loss is specific to image style transfer problems.

$$L_{cyc}(X, X_{recon}, Y, Y_{recon}) = \mathbb{E}_a[||X_{recon} - X||_1] + \mathbb{E}_b[||Y_{recon} - Y||_1] \quad (2)$$

cycle reconstruction loss : done

Along with these two we included the illumination and geometric consistency losses referred to as L_{cyc_conc}, as discussed above. While the adversarial loss and the cycle reconstruction loss are computed over both domains, the consistency losses are computed only in the target domain. Further, instead of using an L_1 loss a contextual loss [42] is used in case of geometric and lighting consistency measurements. There are significant differences in the network performance with variations in the loss functions.

The training process was split into four stages. The stages mainly reflect the effectiveness of each step (multiple discriminators/multiple consistencies/contextual loss) towards the entire process: (i) First, we train the network with adversarial loss $L_{gan}(X, Y)$ and cycle reconstruction loss $L_{cyc}(X, X_{recon}, Y, Y_{recon})$ using color, texture and edge discriminators. We call this version D2B_base;

$$L_{d2b\_base} = L_{gan}(\hat{X}, \hat{Y}) + L_{gan}(\hat{Y}, \hat{X}) + L_{cyc}(X, X_{recon}, Y, Y_{recon}) \quad (3)$$

here $\hat{X} \in \{xc, xt, xe\}$ and $\hat{Y} \in \{yc, yt, ye\}$

(ii) we next used geometric consistency $L_{cyc\_con}(Y_{fake}, Y\_g^{-1}_{fake}, X_{fake}, X\_g^{-1}_{fake})$ along with adversarial loss and cycle-reconstruction. We used only color images in the discriminator and used L_1 loss criterion for consistencies. In this case we determine the geometric consistency loss in both domains. We denote this as GC;

$$L_{GC} = L_{gan}(X, Y) + L_{gan}(Y, X) + L_{cyc\_con}(Y_{fake}, Y\_g^{-1}_{fake}, X_{fake}, X\_g^{-1}_{fake}) + L_{cyc}(X, X_{recon}, Y, Y_{recon}) \quad (4)$$

How to differentiate the 4 discriminators they have used here? I used the same architecture for all the discriminators. for example; for color do I have to pass them through blurred version of X instead of original X?

Using separate discriminators again ensures that the Generator can learn color, edges and texture independently. These characteristics are often imbibed in an image-image network by separately processing the content/edges. The objective of using this in the discriminator stage rather than the encoder stems from the fact that a very poor quality image may generate weak edges which fail to transfer suitably in its enhanced counterpart. More precisely, because we observed an increased performance using multiscale discriminators, we use four discriminators: Dc_{xc1}, Dc_{xc2}, Dc_{xt} and Dc_{xe} instead of three. Here Dc_{xc1} and Dc_{xc2} denote RGB images blurred by two factors. A grayscale version of the image was used as the texture image, whereas the edge image was obtained using a Prewitt operator. Any edge filter can be used instead of the Prewitt filter. Alternatively, we also decomposed a grayscale image into its texture and structure images. This adds an extra overhead for computing the structure decomposition in each epoch. We however noticed that there was no significant improvement in terms of epochs convergence or visual quality of the generated images; hence we dropped this line of experiments and reported all experiments with grayscale images and the Prewitt operator.

did not quiet understand this point .....

The overall objective function is shown in Eq. 8. It includes the general adversarial loss L_{gan} along with the cycle reconstruction loss L_{cyc} and the consistency loss L_{cyc_conc}. The general adversarial loss is commonly used for all GAN structures.

$$L_{gan}(X, Y) = \mathbb{E}_b[(Dc(Y) - 1)^2] + \mathbb{E}_a[(Dc(D_y(E_x(X))))^2] \quad (1)$$

Discriminator Y gen Y

geometry + illumination

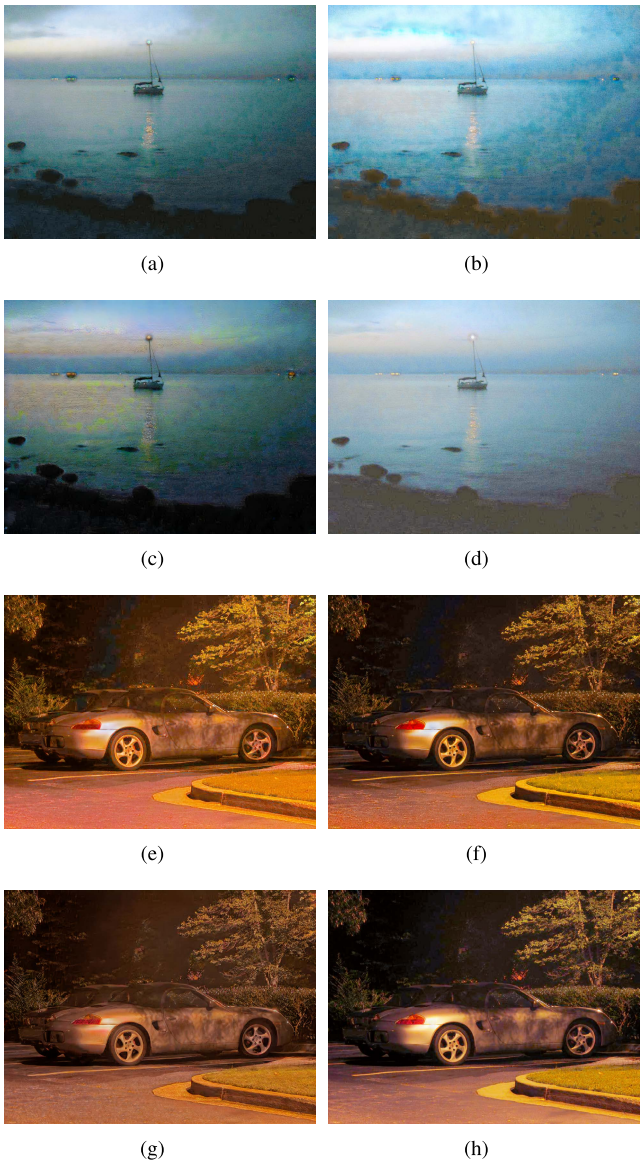


FIGURE 5. (a) and (e) Geometric consistency; (b) and (f) geometric consistency with contextual loss; (c) and (g) edge, texture, and color discriminators; (d) and (h) D2BGAN edge, color, texture discriminators, geometric consistency, and contextual loss. Viewing the images in a zoomed mode on a good quality monitor will provide better understanding of the differences between them.

$$\begin{aligned} & L_{cyc\_con}(Y_{fake}, Y\_g_{fake}^{-1}, X_{fake}, X\_g_{fake}^{-1}) \\ &= \mathbb{E}_a[||Y_{fake} - Y\_g_{fake}^{-1}||_1] \\ &\quad + \mathbb{E}_b[||X_{fake} - X\_g_{fake}^{-1}||_1] \end{aligned} \tag{5}$$

(iii) in the third phase we replaced L_1 loss criterion in (ii) with contextual loss. We name the result GC_con;

$$\begin{aligned} L_{GC\_con} &= L_{gan}(X, Y) + L_{gan}(Y, X) \\ &\quad + L_{cyc\_conc}(Y_{fake}, Y\_g_{fake}^{-1}, X_{fake}, X\_g_{fake}^{-1}) \\ &\quad + L_{cyc}(X, X_{recon}, Y, Y_{recon}) \\ &\quad L_{cyc\_conc}(Y_{fake}, Y\_g_{fake}^{-1}, X_{fake}, X\_g_{fake}^{-1}) \end{aligned} \tag{6}$$

FIGURE 6. (a) and (b) uses only color discriminator; (c) and (d) uses color and edge discriminator; (e) and (f) uses all the discriminators; (b), (d), and (f) uses all the consistency and contextual losses; and (a), (c), and (e) uses only cycle consistency loss.

$$\begin{aligned} &= \mathbb{E}_a[||Y_{fake} - Y\_g_{fake}^{-1}||_c] \\ &\quad + \mathbb{E}_b[||X_{fake} - X\_g_{fake}^{-1}||_c] \end{aligned} \tag{7}$$

(iv) to obtain the final network version we add L_{cyc_con} loss to D2B_base. Unlike GC we only compute the consistencies in the target domain. We used both geometric as well as lighting consistencies for this purpose. Also, the L_1 loss criterion for consistency measurements was replaced with contextual loss. We call the final result D2BGAN. This is shown in Fig 4.

$$\begin{aligned} L_{d2bgan} &= L_{gan}(\hat{X}, \hat{Y}) + L_{gan}(\hat{Y}, \hat{X}) \\ &\quad + L_{cyc}(X, X_{recon}, Y, Y_{recon}) \\ &\quad + L_{cyc\_conc}(Y_{fake}, Y\_g_{fake}^{-1}, Y_{fake}, Y\_l_{fake}) \end{aligned} \tag{8}$$

It should be noted that while the instability of cycle consistency to solve the transfer problem is dealt with successfully by using the geometric consistencies and contextual losses, the two-way GAN requires longer training time compared to any one-path GAN. Although for the dark to bright image enhancement task it is not necessary to train a two-way GAN, we prefer to make the network multipurpose and use it for both way transfers for different types of experiments. Hence we ignore the additional training time.

IV. EXPERIMENTAL RESULTS

We used a mixture of images from two datasets to train our network. We included images from the Adobe and the

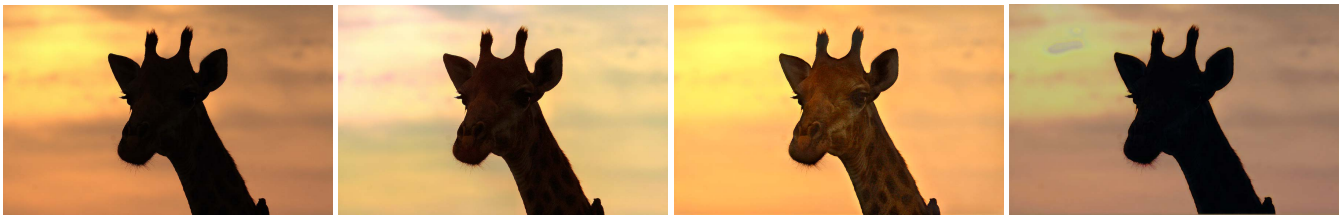


FIGURE 7. Original backlit image, processed images with DeepUPE, EGAN and D2BGAN from left to right.

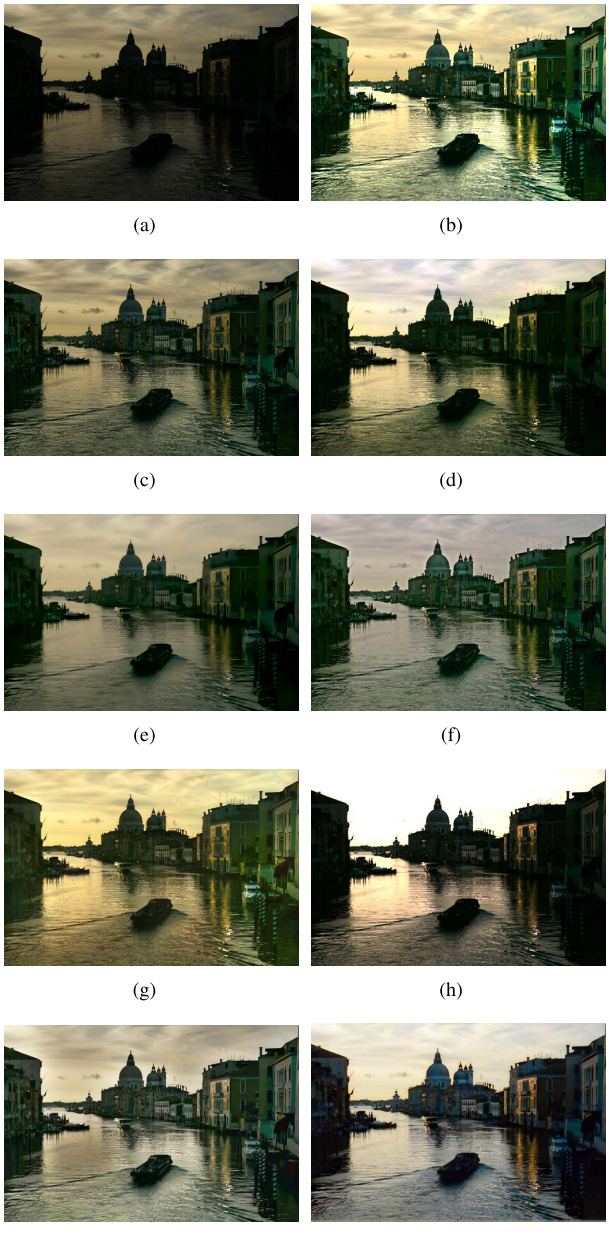


FIGURE 8. (a) Sample original MEF image and its processed versions using (b) Lime, (c) FMSBIE, (d) DeepUPE, (e) MBLLEN, (f) ZDCE, (g) EGAN, (h) RUAS_UPE, (i) LLFLOW, and (j) D2BGAN.

BrighteningTrain dataset. A total of 3000 images were randomly selected for training. The network was trained for 150 epochs with a learning rate of 0.0002. The images were randomly cropped to 256 × 256 pixels. Training was performed on a workstation with an RTX6000 24GB GPU. Each training session lasted approximately 36 hours. We present ablation studies in Sec. IV-A. For testing we used the publicly available benchmark datasets DICM (42 images, mean intensity 63) [43], LIME (9 images, mean intensity 35) [9], MEF (16 images, mean intensity 38) [44], NPE (8 images, mean intensity 90) [45], and a set of backlit images (2 images, mean intensity 90) [8]. We named this entire group of images Dataset-A.

We compared the proposed method with state-of-the-art techniques such as EGAN, MBLLEN, ZDCE, DeepUPE, LIME, FMSBIE, RUAS, LLFLOW. For all these methods, software provided by the authors was utilized. Additionally, we present image understanding results on BDD and on challenging low-light datasets such as DarkFace and ExDark. In Sec. IV-B we present visual and objective evaluations of our technique. We used no-reference image quality assessment (IQA) tools such as NIQE [46], PIQE [47], BRISQUE [48], and UNIQUE [49] for objective evaluations, as the ground-truth for these data is not available. In Sec. IV-C we evaluated the LOL dataset using PSNR and SSIM scores. In Sec. IV-D we present a visual and objective evaluation of the real-time Berkeley Driving dataset. Finally, in Sec. IV-E we provide some evaluations of our technique on the DarkFace and Exdark datasets. The purpose of using different datasets for evaluation was to verify the generalization ability of the proposed approach. For example, Dataset-A used here is a benchmark for comparing different low-light enhancement tasks. The experiments performed on Berkeley Driving datasets verify whether the algorithm can enhance images captured from a moving vehicle in dark roads, in the presence of headlights, street lamps, etc. The appropriate enhancement of these images is important for detection related tasks. Furthermore, the enhancement of typical low-light object/face detection datasets is also shown.

A. ABLATION STUDY

Fig. 5 show the performance of each network discussed in Section III. We observed that the processed images have visible blocking artifacts and color saturation for (ii) and (iii). Hence, the use of (a) color, texture and edge discriminators instead of a single color discriminator, (b) geometric and lighting consistency on the target domain, and (c) contextual loss in place of L_1 loss prove to be significant parts of the D2BGAN. In table 1 we show IQA scores using images of

(a) Original, 42.6789 (b)

(c) DeepUPE, 44.0169 (d)

(e) EGAN, 41.5877 (f)

(g) MBLLEN, 62.3508 (h)

(i) ZDCE, 50.2867 (j)

(k) D2BGAN, 31.8539 (l)

FIGURE 9. Test image and its different processing results (left); enlarged detail for red bounding box in the left Fig is shown on the right. It is observed that comparatively less distortions are seen in the rail tracks region for D2BGAN, hence providing a better PIQE quality score.

(a) Original (b)

(c) DeepUPE (d)

(e) EGAN (f)

(g) MBLLEN (h)

(i) ZDCE (j)

(k) D2BGAN (l)

FIGURE 10. Test image and its different processing results (left); enlarged detail (right). The ceiling region of the processed images for EGAN and ZDCE show significant noise addition which is the probable reason for a poor PIQE score. MBLLEN on the otherhand provides an oversmoothed image and lower color intensity, hence lowering the PIQE score. Results of DeepUPE and D2BGAN are comparatively better both visually as well as in terms of score.

(a) Original, 61.7018 (b) DeepUPE, 61.3802 (c) EGAN, 61.4834

(d) MBLLEN, 79.564 (e) ZDCE, 58.4526 (f) D2BGAN, 53.5719

FIGURE 11. A typical color difference is observed in the different processed images. The color distortion results in poor PIQE scores for the other techniques.

the Exdark dataset (12000 total images). We demonstrate the utility of all three discriminators. The results obtained with only a color discriminator are better when geometric and contextual losses are used. The best results are obtained with all the discriminators. This is illustrated in Fig. 6. The color discriminator alone in Fig. 6a has inconsistent patterns all over the image. This effect is eliminated in Fig. 6b when geometric and contextual losses are used. In Fig. 6d the image appears pixelated after zooming, as it fails to learn the appropriate texture. The use of all the discriminators provides better feature learning. The use of geometric and contextual losses along with all the discriminators reduces the noise in the image. This is clear when comparing Figs. 6e (using all discriminators and cycle consistency) and 6f (all discriminators, cycle consistency and geometric and contextual loss). At the same time, the use of all three discriminators prevents over-smoothing, as can be noted by comparing Figs. 6f, 6b and 6d.

B. EVALUATION ON LOW LIGHT BENCHMARK DATASET-A

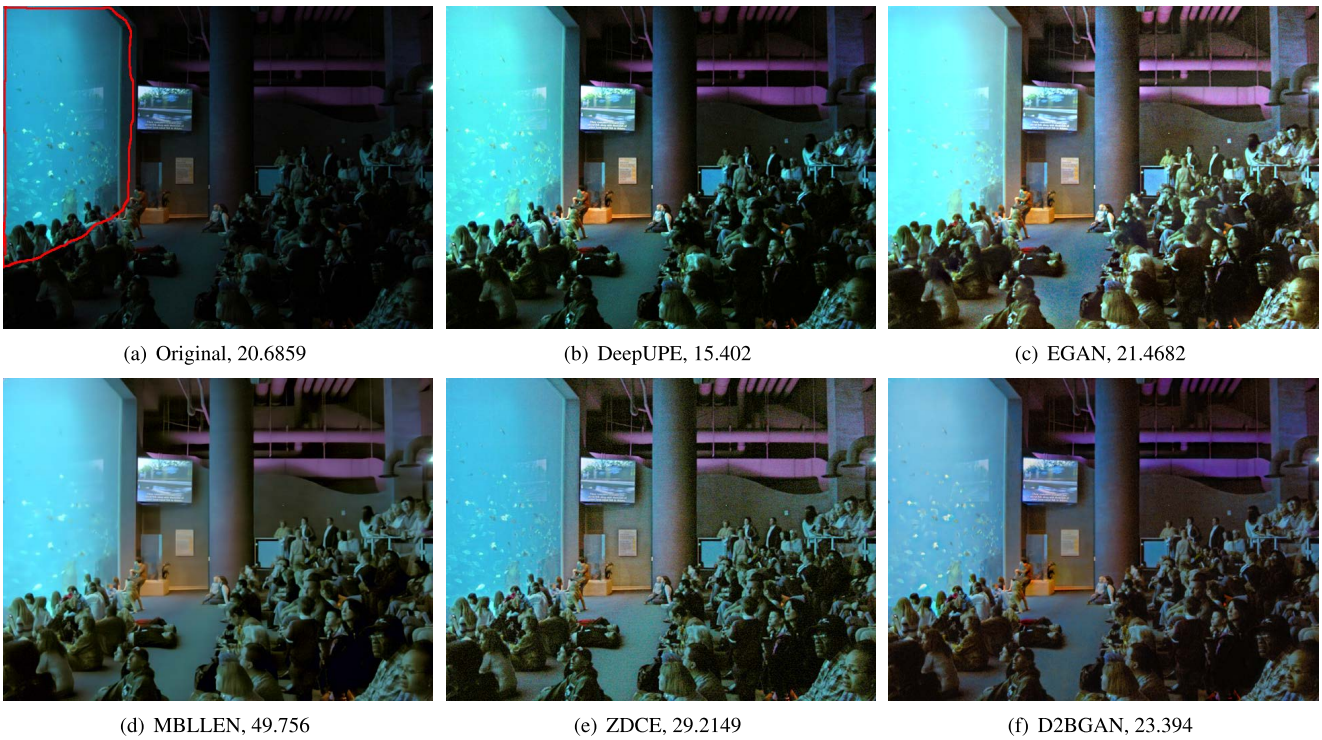
We demonstrate the IQA scores obtained after preprocessing the original dark images using our method and the other state-of-the-art methods indicated above. The scores shown are the average scores for each set forming Dataset-A. We also show the average scores for Dataset-A, together with the corresponding standard deviations. Thus we can verify the flexible of the different techniques, by evaluating the changes in their response to images with different characteristics.

From Table 2 it can be observed that the proposed technique D2BGAN obtains competitive results. The NIQE tool provides the best results for DICM, LIME and MEF datasets when D2BGAN was applied. For NPE, D2BGAN provides the second-best score. The LIME and NPE datasets also showed highest performance with D2BGAN when the PIQE image quality tool was used. However, D2BGAN does not perform well on backlit images. A sample case is shown in Fig 7. It should be noted that most of the techniques (with a couple of exceptions) reported here fail to deal with backlit cases. The dataset used for training D2BGAN does not explicitly use any backlit cases, and no constraint is used to address backlit cases. The most promising results for backlit images are obtained using EGAN. The fact that EGAN uses a global/local discriminator can be a contributing factor. A trivial approach to improve the backlit image processing ability of D2BGAN could be to check whether the network responds

(a) Original, 17.9657 (b) DeepUPE, 18.2862 (c) EGAN, 22.9831

(d) MBLLEN, 33.8 (e) ZDCE, 24.4544 (f) D2BGAN, 29.5288

FIGURE 12. A strong contrast in the hill regions of the image (shown with red border in the original image) is observed in the DeepUPE and EGAN processing. The considerably poor contrast results in a poor PIQE score for D2BGAN.



(a) Original, 20.6859 (b) DeepUPE, 15.402 (c) EGAN, 21.4682

(d) MBLLEN, 49.756 (e) ZDCE, 29.2149 (f) D2BGAN, 23.394

FIGURE 13. A strong contrast in the aquarium region(shown with red border in the original image) is observed in the DeepUPE and EGAN images. The considerably poor contrast results in a poor PIQE score for D2BGAN.

better if it has backlit priors. A more effective method can be the use of blur maps [50] to train the network. These maps will act as attention enforcement, differentiating in-focus areas from background. In the example backlit image, the giraffe

TABLE 2. Objective no-reference image quality scores on different benchmark datasets (values in red and blue are best and second best, respectively).

		Orig	DeepUPE [29]	EGAN [21]	LIME [9]	MBLLEN [32]	ZDCE [18]	FMSBIE [8]	RUAS_LOL [37]	RUAS_UPE [37]	LL_FLOW [17]	D2BGAN
PIQE	DICM	29.28	29.7	**28.04**	34.27	40.18	**25.94**	29.19	44.75	33.29	32.94	28.27
	LIME	**34.19**	35.19	34.75	40.76	52.3	37.26	38.63	43.22	37.99	47.51	**33.25**
	MEF	43	35.3	**32.22**	38.76	54.59	34.63	34.95	41.54	35.63	53.83	**32.96**
	NPE	35.56	**33.2**	33.96	38.87	45.02	37.62	40.98	46.96	41.33	39.46	**29.42**
	backlit	30.78	26.42	**22.8**	27.64	42.9	**21.17**	29.71	46.88	27.02	33.54	30.31
NIQE	DICM	3.03	3.14	2.73	2.91	2.91	2.69	2.66	4.34	3.87	**2.58**	**2.5**
	LIME	3.74	3.76	**3.52**	4.44	3.89	3.97	4.21	4.45	4.24	4.84	**3.26**
	MEF	3.29	3.16	**2.89**	3.67	3.52	3.33	3.39	4.04	3.50	3.45	**2.78**
	NPE	3.42	**3.29**	3.34	3.9	3.46	3.94	4.01	5.91	5.06	3.46	**3.31**
	backlit	3.01	2.97	**2.48**	2.83	3.21	**2.61**	2.99	4.98	3.29	2.72	3.09
BRISQUE	DICM	23.92	**20.84**	21.88	24.72	27.87	24.87	21.63	38.34	36.43	**20.20**	20.89
	LIME	25.41	26.23	**20.87**	22.09	30.6	23.75	27.99	31.64	26.50	32.91	**21.66**
	MEF	29.41	**17.86**	23.26	23.36	32.22	25.27	23.54	33.00	27.54	26.79	**20.94**
	NPE	**25.14**	**21.5**	27.34	27.1	31.34	29.1	28.72	44.51	41.73	26.42	27.21
	backlit	32.28	**19.24**	**25.44**	34.93	37.88	39.64	28.47	41.74	38.25	27.54	28.29
UNIQUE	DICM	0.84	0.93	0.81	0.76	1.05	1.02	**1.18**	-0.34	0.38	**1.29**	0.88
	LIME	0.66	0.69	0.53	0.33	**0.86**	0.68	0.7	-0.26	0.25	**1.00**	0.71
	MEF	0.7	1.02	1.01	0.96	1.03	**1.19**	1.12	0.23	0.71	**1.37**	1.16
	NPE	1.17	1.06	0.92	0.81	**1.25**	1.1	1.19	-0.70	0.22	**1.32**	1.07
	backlit	0.43	0.51	0.71	0.86	0.66	**0.96**	**0.92**	0.31	0.52	**0.92**	0.72

TABLE 3. Quality gain values averaged (avg) on the entire dataset-A, and their corresponding standard deviations (std). Values in red and blue are best and second best, respectively.

		DeepUPE	EGAN	LIME	MBLLEN	ZDCE	FMSBIE	RUAS_LOL	RUAS_UPE	LL_FLOW	D2BGAN
PIQE	avg	6.87	**11.62**	-5.10	-36.62	9.47	-1.14	-32.04	-2.37	-19.32	**9.67**
	std	9.22	12.91	14.30	10.84	16.96	13.76	23.08	15.73	12.68	9.98
NIQE	avg	0.95	**9.55**	-6.91	-3.01	0.35	-4.03	-44.65	-20.95	-2.19	**9.24**
	std	3.17	5.82	11.14	4.58	11.94	11.55	24.31	17.18	17.17	8.63
BRISQUE	avg	**20.76**	11.95	2.86	-17.70	-4.38	3.39	-40.67	-26.94	0.91	**12.07**
	std	18.74	12.67	13.16	5.57	15.24	14.82	26.96	31.07	18.91	13.24
UNIQUE	avg	13.76	12.64	9.08	31.79	41.93	**43.96**	-106.8	-35.22	**65.14**	27.07
	std	20.30	39.23	59.63	17.88	53.28	44.77	56.67	43.3	39.34	36.06

TABLE 4. Quality gain values averaged (avg) on dataset-A without backlit images, and their corresponding standard deviations (std). Values in red and blue are best and second best, respectively.

		DeepUPE	EGAN	LIME	MBLLEN	ZDCE	FMSBIE	RUAS_LOL	RUAS_UPE	LL_FLOW	D2BGAN
PIQE	avg	5.05	**8.04**	-8.92	-35.94	4.03	-2.30	-26.98	-5.97	-21.9	**11.71**
	std	9.55	11.70	13.23	12.39	13.65	15.61	23.22	15.55	13.03	10.25
NIQE	avg	0.87	**7.55**	-10.12	-2.08	-2.86	-5.19	-39.46	-23.86	-5.15	**12.24**
	std	3.66	4.31	9.85	4.71	11.00	13.00	24.65	18.36	18.3	6.29
BRISQUE	avg	**15.85**	9.64	5.63	-17.78	0.22	1.29	-43.51	-29.06	-2.54	**12.00**
	std	17.54	13.35	13.42	6.43	12.98	16.23	30.26	35.46	19.94	15.28
UNIQUE	avg	12.83	-0.32	-13.16	27.05	22.11	**27.16**	-126.57	-48.8	**53.59**	17.19
	std	23.32	30.54	37.97	16.62	34.15	28.12	40.94	35.64	34.25	32.93

will be labelled as an in-focus object and the network will be guided to improve the illumination of that region.

Because no-reference IQA tools provide non-uniform scores for the same datasets, it is reasonable to study the achieved Quality Gain (QG) which we define as the increase in the scores after processing, with respect to the score of the original dark image. The QG values were calculated as the average percent IQA score relative change. Table 3 lists the QG values averaged over Dataset-A, and their corresponding standard deviations. We observed that D2BGAN ranks second according to PIQE, NIQE, and BRISQUE; EGAN ranks first for PIQE and NIQE, whereas DeepUPE ranks first for BRISQUE. The s.d. of D2BGAN was generally low, indicating the robustness of the proposed method.

If the backlit images are excluded from the evaluation, it can be verified (Table 4) that D2BGAN ranks first for PIQE and NIQE, and second for BRISQUE. The results for UNIQUE are less satisfactory: this IQA tool seems to be sensitive to quality characteristics that other indices tend to disregard. Indeed, it can be observed that the ranking of the different methods provided by UNIQUE is almost reversed with respect to those provided by the other indices. We also present visual comparisons in Fig. 8.

Fig 9-13 demonstrate some sample cases, highlighting the reasons for the success and failure of D2BGAN. It is observed that EGAN/ZDCE often produces strong illumination in low-resolution patches, resulting in a strong noise in those regions. This reduces the image quality score. Noise amplification in the illuminated regions is controlled in most cases by D2BGAN, which results in better image scores. MBLLEN processing does not generate noise in illuminated image parts; however this is mainly because of the smoothing effect which again provides poor image score. DeepUpe provides good results particularly for moderately low-light indoor and natural scenes. Cases of failure other than backlit images have also been reported. In such cases D2BGAN provides

FIGURE 14. Original low light, ground truth, processed images with MBLLEN and D2BGAN from left to right.

TABLE 5. Objective no-reference image quality scores on the LOL dataset.

METHOD	PIQE	NIQE	BRISQUE
DeepUPE	34.16	10.63	33.09
EGAN	41.77	7.67	30.04
FMSBIE	50.85	12.04	38.34
LIME	55.83	11.81	39.23
MBLLEN	42.33	5.41	28.48
ZDCE	47.27	11.21	36.86
D2BGAN	26.13	3.15	22.13

TABLE 6. Objective SSIM and PSNR scores on LOL dataset.

METHOD	DATASET	PSNR	SSIM
DeepUPE	DPAIR	10.70	0.3233
EGAN	DPAIR	15.31	0.4918
FMSBIE	DPAIR	12.16	0.3995
LIME	DPAIR	14.45	0.3901
MBLLEN	DPAIR	16.59	0.5134
ZDCE	DPAIR	13.42	0.4252
D2BGAN	DPAIR	16.53	0.5742

a noise-free illuminated image, but some regions have lower contrast compared to the top scoring methods for that image. This can be seen in Fig 12 and 13.

C. EVALUATION ON BENCHMARK LOW LIGHT DATASET-B

In Table 5 we demonstrate the image quality scores obtained on LOL dataset (789 images, mean intensity 15) after pre-processing the original dark images in each case using the respective techniques. As ground-truth data are available for LOL, we also show PSNR and SSIM scores for the same in Table 6. It is observed that the proposed technique D2BGAN obtains the best scores for both cases and for all IQA tools.

A sample Low-light image, its ground-truth (GT), and the processed version using *D2BGAN* are shown in Fig. 14. It was observed that PSNR (16.53) and SSIM (0.57) for the processed images were quite low. However, the original PSNR and SSIM scores are 8.18 and 0.17 respectively (averaged on 789 test images). Although the scores of the processed images are still low, they are largely improved compared to the original ones as well as those obtained using other techniques. It should be noted that these are real-world (not synthetic) images; i.e. they represent an actual problem a user may face.

D. EVALUATION ON REAL-TIME DRIVING DATASET

In Table 7 we demonstrate the image quality scores obtained on the BDD after preprocessing the original dark images in each case using the respective techniques. The Berkeley Driving Dataset consisted of road scenes captured during day and night. We selected dark images by filtering those images with a global mean of less than 30. We chose 69 such images

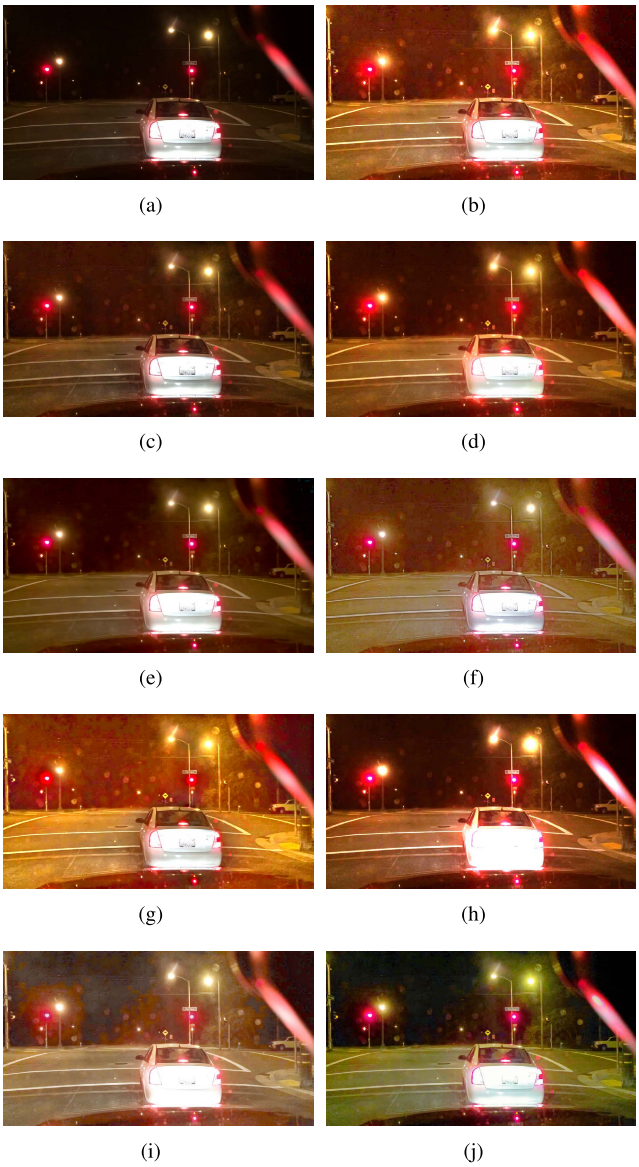


FIGURE 15. (a) Sample original BDD image and its processed versions using (b) Lime, (c) FMSBIE, (d) DeepUPE, (e) MBLLEN, (f) ZDCE, (g) EGAN, (h) RUAS_UPE, (i) LLFLOW, and (j) D2BGAN.

(a) (b) (c) (d)

FIGURE 16. Face detection performances on original (a) and (c) and preprocessed image (b) and (d) using D2BGAN from DarkFace (a) and (b) and ExDark (c) and (d) datasets.

(a) (b)

FIGURE 17. Face detection performances on original and preprocessed images from DarkFace dataset. (a) is original while (b) is the processed version using D2BGAN.

to generate the test results. The mean intensity of these images was 29. The scores shown are the average scores for the entire dataset. It is observed that the proposed technique D2BGAN obtains the best scores for all the IQA tools. The images

TABLE 7. Objective no-reference image quality scores on BDD.

METHOD	DATASET	PIQE	NIQE	BRISQUE
DeepUPE	BDD	67.70	3.62	46.56
EGAN	BDD	68.75	3.21	36.53
FMSBIE	BDD	72.50	4.25	46.89
LIME	BDD	73.65	4.13	47.85
MBLLEN	BDD	64.84	3.58	46.91
ZDCE	BDD	72.01	4.29	49.02
RUAS_LOL	BDD	69.12	4.07	47.33
RUAS_UPE	BDD	71.90	4.58	50.18
LL_FLOW	BDD	75.78	3.65	41.86
D2BGAN	BDD	61.17	3.12	36.03

in Fig. 15 show that blocking artifacts and light reflections are enhanced for EGAN and LIME, providing poor images. These problems were not visible for FMSBIE, MBLLEN and D2BGAN. In addition, the enhancement provided by D2BGAN is smoother and brighter in the road regions without color saturation and changes in the sky region.

E. EVALUATION OF IMAGE UNDERSTANDING TASKS

One of the most important purposes of low-light enhancement is to improve the performance of image-understanding operators. As shown in case of Berkeley Driving dataset, a brighter and noise free image of the road enables more accurate object detection results. Exdark (12 object classes) and DarkFace (2 classes) are benchmark datasets for object and face detection respectively. The DarkFace dataset contains 6000 training and validation images and is particularly challenging because it presents an extremely dark environment with strong noise. Enhancing the images amplifies this noise making face detection very difficult. We randomly selected two test sets of 500 images each from the 6000 image dataset and used the remaining 5000 for training. We obtain a mAP of 0.209 and 0.218 on the original DarkFace images for the two test sets when a pretrained Retinaface (trained on Widerface dataset) detector is applied. After preprocessing with D2BGAN the mAP was 0.301 and 0.331 respectively. We also tested the mAP after applying EGAN, and obtained values of 0.285 and 0.331 respectively. Fine-tuning the pretrained network with preprocessed DarkFace images yields a mAP of 0.525 and 0.407 on D2BGAN and EGAN respectively. Fig. 16 shows some examples of the original and preprocessed images (using D2BGAN) along with the detections obtained. Additional examples are shown in Figs. 17 and 18. It can be seen that the number of detected faces increased after preprocessing. We used S3FD face detectors and YOLOV3 object detectors to show the visual results. Training and fine-tuning were performed using Tinaface detector. Such experiments primarily aim to show how preprocessing the images using *D2BGAN* model improves the results of state-of-the-art detection algorithms. Hence, all experiments report the detection results using a standard algorithm post enhancement application. For example, Fig. 16 shows two false detections on the original image. In the processed image one more face was detected, with no false detections. Although it would be interesting to train the detection network jointly with the enhancement network, it is currently not covered in our scope.

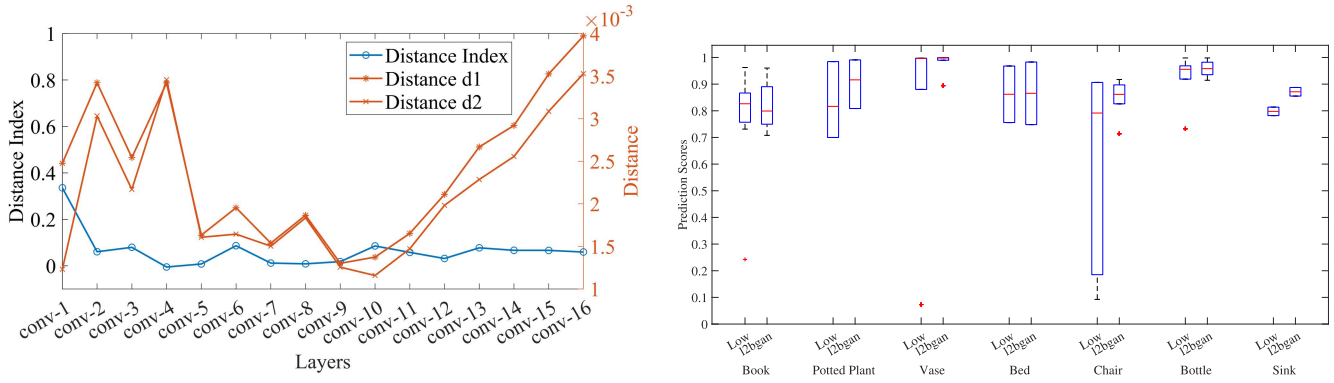
(a) (b)

FIGURE 18. Face and object detection performances on original and pre-processed image from ExDark dataset. (a) is original while (b) is the processed version using D2BGAN.

We also performed two simple experiments to analyze the quality of the classification features and object detection scores obtained from the D2BGAN images compared with the low-light images. We take low-light and normal (ground-truth) images from the LOL dataset, and obtained D2BGAN images by processing the low-light images. We further computed the distances $d2$ and $d1$ for each layer of VGG-19 trained using ImageNet. $d2$ and $d1$ were obtained using the feature differences of normal-D2BGAN images and normal-low image pairs respectively. We then computed a distance index $(d1 - d2)/(d1 + d2)$ as shown in Fig. 19 (left). We observe that: (a) $d1$ is greater than $d2$, hence, D2BGAN features are closer to normal for all layers; (b) although the distance index decreases as we move towards higher layers, it is still greater than 0; (c) interestingly, both low and D2BGAN features are closest to the normal between layer conv9-10, after which they increase again. Hence extracting features from this layer may be beneficial. In Fig. 19 (right) we compare the object detection scores of low and D2BGAN images. 40 images were selected from the LOL dataset. It is observed that D2BGAN generally provides a greater mean score and lower deviation.

F. EVALUATION ON RAW IMAGES

In some studies, the image enhancement task was carried out on RAW input data, which possess a large amount of information, not present in the JPEG or PNG versions. Our aim is to enhance images whose RAW versions are not available. This is indeed the case for many applications where post-processing needs to be performed on displayable formats. However,we performed a simple experiment in which we have processed a RAW image using our network,



(a) (b)

FIGURE 19. (a) depicts the distance index obtained from classification features of low, normal and processed image. (b) depicts the object detection scores using low and processed images.

(a) (b)

(c) (d)

FIGURE 20. (a) Original low light RAW; (b) ground truth; and RAW image processed with (c) MBLLEN and (d) D2BGAN.

without retraining. The results obtained can be found in Fig. 20. The low-exposure RAW image and normal exposure ground truth images were taken from the SID dataset (See-in-the-dark) [51] It would be interesting to assess the performance of the network after fine-tuning it with RAW images.

V. CONCLUSION

We presented an unpaired GAN-based image enhancement operation using cycle consistency, geometric consistency and illumination consistency. Visual and objective results on the standard benchmark datasets show that our D2BGAN provides competitive results. It is observed that D2BGAN can

enhance real images suffering from typical artifacts, without considerably amplifying blocking artifacts. In most cases, the images exhibited show smooth enhancements without color saturation. One of the main advantages of the D2BGAN is its better generalization across different datasets without providing separate dataset-based models. Its performance particularly on LOL and BDD make it suitable for the image enhancement operations required for detection tasks. We have particularly focused on JPEG images even though a simple experiment on RAW data is discussed. The application of D2BGAN for processing images to be used for face detection yielded significant improvements.

It would be interesting to see how joint training and domain adaptation can influence D2BGAN to provide superior results on datasets such as DarkFace. In general image-enhancement and detection tasks are treated independently often resulting in poor end results. The idea is to use a joint loss function that simultaneously learns to enhance images and detect objects/faces. Another approach could be to use object priors during enhancement training to ensure that proper features are learned. This should prove to be far more effective than content/perceptual loss of the entire image. Future research directions include exploring the scope of improving image quality and detection ability with backlit images. Global attention maps such as blur maps may be able to differentiate in-focus areas from background regions.

what are attention maps?

REFERENCES

[1] C. Wang, J. Peng, and Z. Ye, "Flattest histogram specification with accurate brightness preservation," *IET Image Process.*, vol. 2, no. 5, pp. 249–262, Oct. 2008.

[2] T. Arici, S. Dikbas, and Y. Altunbasak, "A histogram modification framework and its application for image contrast enhancement," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 1921–1935, Sep. 2009.

[3] S. Hashemi, S. Kiani, N. Noroozi, and M. E. Moghaddam, "An image enhancement method based on genetic algorithm," in *Proc. Int. Conf. Digit. Image Process.*, Mar. 2009, pp. 167–171.

[4] K. Nakai, Y. Hoshi, and A. Taguchi, "Color image contrast enhacement method based on differential intensity/saturation gray-levels histograms," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, Nov. 2013, pp. 445–449.

[5] T. Celik and T. Tjahjadi, "Contextual and variational contrast enhancement," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3431–3441, Dec. 2011.

[6] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation of 2D histograms," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5372–5384, Dec. 2013.

[7] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, and J. Paisley, "A fusion-based enhancing method for weakly illuminated images," *Signal Process.*, vol. 129, pp. 82–96, Dec. 2016.

[8] Q. Wang, X. Fu, X.-P. Zhang, and X. Ding, "A fusion-based method for single backlit image enhancement," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 4077–4081.

[9] 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.

[10] Y. Guo, Y. Lu, R. W. Liu, M. Yang, and K. T. Chui, "Low-light image enhancement with regularized illumination optimization and deep noise suppression," *IEEE Access*, vol. 8, pp. 145297–145315, 2020.

[11] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A weighted variational model for simultaneous reflectance and illumination estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2782–2790.

[12] B. Cai, X. Xu, K. Guo, K. Jia, B. Hu, and D. Tao, "A joint intrinsic-extrinsic prior model for retinex," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4020–4029.

[13] K. G. Lore, A. Akintayo, and S. Sarkar, "LLNet: A deep autoencoder approach to natural low-light image enhancement," *Pattern Recognit.*, vol. 61, pp. 650–662, Jan. 2017.

[14] W. Wang, C. Wei, W. Yang, and J. Liu, "GLADNet: Low-light enhancement network with global awareness," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 751–755.

[15] Q. Li, H. Wu, L. Xu, L. Wang, Y. Lv, and X. Kang, "Low-light image enhancement based on deep symmetric encoder–decoder convolutional networks," *Symmetry*, vol. 12, no. 3, p. 446, Mar. 2020.

[16] Y. Zhang, X. Guo, J. Ma, W. Liu, and J. Zhang, "Beyond brightening low-light images," *Int. J. Comput. Vis.*, vol. 129, no. 4, pp. 1013–1037, Apr. 2021.

[17] Y. Wang, R. Wan, W. Yang, H. Li, L.-P. Chau, and A. C. Kot, "Low-light image enhancement with normalizing flow," in *Proc. AAAI Conf. Artif. Intell.*, 2022, pp. 1–9.

[18] 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 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 1780–1789.

[19] Z. Ni, W. Yang, S. Wang, L. Ma, and S. Kwong, "Towards unsupervised deep image enhancement with generative adversarial network," *IEEE Trans. Image Process.*, vol. 29, pp. 9140–9151, 2020.

[20] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input/output image pairs," in *Proc. CVPR*, Jun. 2011, pp. 97–104.

[21] 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 Trans. Image Process.*, vol. 30, pp. 2340–2349, 2021. Is D2BGAN better than EnlightenGAN?

[22] A. Anoosheh, T. Sattler, R. Timofte, M. Pollefeys, and L. V. Gool, "Night-to-day image translation for retrieval-based localization," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 5958–5964.

[23] H. Fu, M. Gong, C. Wang, K. Batmanghelich, K. Zhang, and D. Tao, "Geometry-consistent generative adversarial networks for one-sided unsupervised domain mapping," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2427–2436.

[24] J.-Y. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman, "Toward multimodal image-to-image translation," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst. (NIPS)*. Red Hook, NY, USA: Curran Associates, 2017, pp. 465–476.

[25] H. Ibrahim and N. S. P. Kong, "Brightness preserving dynamic histogram equalization for image contrast enhancement," *IEEE Trans. Consum. Electron.*, vol. 53, no. 4, pp. 1752–1758, Nov. 2007.

[26] B. Li, S. Wang, and Y. Geng, "Image enhancement based on retinex and lightness decomposition," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 3417–3420.

[27] H. Liu, X. Sun, H. Han, and W. Cao, "Low-light video image enhancement based on multiscale retinex-like algorithm," in *Proc. Chin. Control Decis. Conf. (CCDC)*, May 2016, pp. 3712–3715.

[28] W. Ren, S. Liu, L. Ma, Q. Xu, X. Xu, X. Cao, J. Du, and M.-H. Yang, "Low-light image enhancement via a deep hybrid network," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4364–4375, Sep. 2019.

[29] R. Wang, Q. Zhang, C.-W. Fu, X. Shen, W.-S. Zheng, and J. Jia, "Underexposed photo enhancement using deep illumination estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 6842–6850.

[30] Y. Guo, X. Ke, J. Ma, and J. Zhang, "A pipeline neural network for low-light image enhancement," *IEEE Access*, vol. 7, pp. 13737–13744, 2019.

[31] H. Jiang and Y. Zheng, "Learning to see moving objects in the dark," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7323–7332.

[32] F. Lv, F. Lu, J. Wu, and C. Lim, "MBLLEN: Low-light image/video enhancement using CNN," *Brit. Mach. Vis. Conf. (BMVC)*, 2018, p. 4.

[33] L. Tao, C. Zhu, G. Xiang, Y. Li, H. Jia, and X. Xie, "LLCNN: A convolutional neural network for low-light image enhancement," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.

[34] A. Ignatov, N. Kobyshev, R. Timofte, and K. Vanhoey, "DSLR-quality photos on mobile devices with deep convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3297–3305.
[35] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *Proc. Brit. Mach. Vis. Conf.*, 2018, Art. no. 61772043.
[36] Y. Zhang, J. Zhang, and X. Guo, "Kindling the darkness: A practical low-light image enhancer," in *Proc. 27th ACM Int. Conf. Multimedia*, New York, NY, USA, Oct. 2019, pp. 1632–1640.
[37] 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 *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10561–10570.
[38] G. Kim, D. Kwon, and J. Kwon, "Low-LightGAN: Low-light enhancement via advanced generative adversarial network with task-driven training," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 2811–2815.
[39] H. Lee, K. Sohn, and D. Min, "Unsupervised low-light image enhancement using bright channel prior," *IEEE Signal Process. Lett.*, vol. 27, pp. 251–255, 2020.
[40] A. Sharma and R. T. Tan, "Nighttime visibility enhancement by increasing the dynamic range and suppression of light effects," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 11972–11981.
[41] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.
[42] R. Mechrez, I. Talmi, and L. Zelnik-Manor, "The contextual loss for image transformation with non-aligned data," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 768–783.
[43] C. Lee, C. Lee, and C.-S. Kim, "Contrast enhancement based on layered difference representation," in *Proc. 19th IEEE Int. Conf. Image Process.*, Sep. 2012, pp. 965–968.
[44] K. Ma, K. Zeng, and Z. Wang, "Perceptual quality assessment for multi-exposure image fusion," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3345–3356, Nov. 2015.
[45] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3538–3548, Sep. 2013.
[46] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2012.
[47] N. Venkatanath, D. Praneeth, B. M. Chandrasekhar, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features," in *Proc. 21st Nat. Conf. Commun. (NCC)*, Feb. 2015, pp. 1–6.
[48] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.
[49] W. Zhang, K. Ma, G. Zhai, and X. Yang, "Uncertainty-aware blind image quality assessment in the laboratory and wild," *IEEE Trans. Image Process.*, vol. 30, pp. 3474–3486, 2021.
[50] K. Ma, H. Fu, T. Liu, Z. Wang, and D. Tao, "Deep blur mapping: Exploiting high-level semantics by deep neural networks," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 5155–5166, Oct. 2018.
[51] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3291–3300.