ISSN 1213-6972 (print)
ISSN 1213-6964 (online)
Journal of WSCG
http://www.wscg.eu
Vol.31, No-1-2, 2023

ALIVE: Adaptive-Chromaticity for Interactive Low-light Image and Video Enhancement

Sumit Shekhar[1]
sumit.shekhar@hpi.uni-potsdam.de

Max Reimann[1]
max.reimann@hpi.uni-potsdam.de

Jobin Idiculla Wattaseril[1]
jobin.wattaseril@hpi.uni-potsdam.de

Amir Semmo[2]
amir.semmo@digitalmasterpieces.com

Jürgen Döllner[1]
doellner@uni-potsdam.de

Matthias Trapp[1]
matthias.trapp@hpi.uni-potsdam.de

[1]Hasso Plattner Institute, Digital Engineering Faculty, University of Potsdam,
Prof.-Dr.-Helmert-Str. 2-3, 14482 Potsdam, Germany.

[2]Digital Masterpieces GmbH,
August-Bebel-Str. 26, 14482 Potsdam, Germany.

ABSTRACT

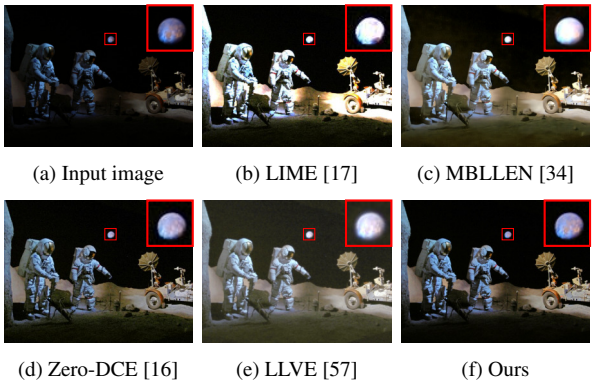
Image acquisition in low-light conditions suffers from poor quality and significant degradation in visual aesthetics. This affects the visual perception of the acquired image and the performance of computer vision and image processing algorithms applied after acquisition. Especially for videos, the additional temporal domain makes it more challenging, wherein quality is required to be preserved in a temporally coherent manner. We present a simple yet effective approach for low-light image and video enhancement. To this end, we introduce *Adaptive Chromaticity*, which refers to an adaptive computation of image chromaticity. The above adaptivity avoids the costly step of low-light image decomposition into illumination and reflectance, employed by many existing techniques. Subsequently, we achieve interactive performance, even for high resolution images. Moreover, all stages in our method consists of only point-based operations and high-pass or low-pass filtering, thereby ensuring that the amount of temporal incoherence is negligible when applied on a per-frame basis for videos. Our results on standard low-light image datasets show the efficacy of our method and its qualitative and quantitative superiority over several state-of-the-art approaches. We perform a user study to demonstrate the preference for our method in comparison to state-of-the-art approaches for videos captured in the wild.

Keywords

real-time, interactive, low-light, image, video, enhancement

1 INTRODUCTION

Due to unavoidable technical or environmental constraints, images and videos captured in poor lighting conditions suffer from severe degradation of visual quality. On most occasions, it is challenging for such visual media to be consumed for high-level tasks such as object detection or tracking due to deterioration or lack of information. Moreover, poor visual quality negatively impacts the overall aesthetics, and thus, the experience of end-users.



(a) Input image (b) LIME [17] (c) MBLLEN [34]

(d) Zero-DCE [16] (e) LLVE [57] (f) Ours

Figure 1: Comparison of LLIE results for three image-based (b to d) and one video-based (e) method. Our method (f) can brighten image while preserving details and avoiding artifacts in terms of over-exposedness, noise, and desaturation.

Numerous algorithms have been proposed for Low-light Image Enhancement (LLIE) (Fig. 1) and a few for video enhancement as well. A class of methods is based

Table 1: Comparing existing low-light image enhancement techniques in the context of interactivity. Here, the color green denotes the aspect which is favourable to interactive enhancement while the color red denotes otherwise.

	LIME [17]	SRIE [31]	MBLLEN [34]	RetinexNet [49]	Zero-DCE [16]	LLVE [57]	Ours
Provides enhancement editing at inference time?	Yes	Yes	No	No	No	No	Yes
Is the performance interactive?	No	No	Yes	Yes	Yes	Yes	Yes

on Retinex theory [25] that assumes the image to be a product of illumination and reflectance. Such Retinex-based approaches decompose the image into illumination and/or reflectance components, based on specific priors. However, finding effective priors is challenging and inaccuracies can result in artifacts and color deviations in the enhanced output. Further, the runtime for such a decomposition, employing a complex optimization process, is relatively long [32]. In comparison, deep-learning-based approaches are faster than conventional methods and learn the underlying prior using the given data distribution. However, they tend to suffer from limited generalization capability. The above could be due to limited/synthetic training data, ineffective network structures, or unrealistic assumptions [28]. Further, learning based approaches do not allow interactive editing of enhancement settings during test time and requires a complete re-training for this purpose. Therefore, we aim to develop a practical solution for LLIE, which adapts to different low-light conditions and also has low computational complexity for enabling interactive enhancement editing on commodity hardware (Tab. 1).

To achieve the above objective, we develop a method based on Retinex theory, the basis for various conventional and learning-based techniques. We avoid the compute-intensive decomposition step and propose an adaptive way to transition into baseline-reflectance (i.e., chromaticity) [4] via parameter tuning. We refer to it as *Adaptive Chromaticity (AC)*, which forms the basis for our approach. The adaptive transition into chromaticity can efficiently increase the output brightness while being robust against dark (or low-intensity) pixels. Moreover, it prevents amplification of sensor noises, which are common in low-light images. To further prevent noise amplification during enhancement, we decompose the input image into coarse and fine attributes, generally referred to as *base* and *detail* components respectively [2]. We generate multiple ACs for the *base* layer with varying levels of brightness followed by a multi-scale fusion step. Different levels of brightness prevents over/under-exposedness, while multi-scale fusion maintains spatial consistency. The *detail* layer is finally added to the result, thus preserving fine image details.

Unlike images, low-light video enhancement has received less attention. Application of image-based methods to videos on a per-frame basis is usually temporally incoherent and leads to flickering artifacts. Dark pixels, significantly contributing to noise amplification, is often the major source of temporal incoherence. Due to the ability of our method to robustly handle such pixels, the degree of incoherence is reduced significantly. Even the per-frame application of our image-based approach is superior to an existing video-specific approach. Our contributions are summarized as follows, we propose:

1. Adaptive Chromaticity (AC) to efficiently increase image brightness while preventing noise amplification.
2. An approach for interactive low-light image enhancement based on exposure fusion of multiple ACs.
3. A per-frame application of our image-based approach for videos that performs out-of-the-box without introducing significant temporal incoherence.

2 RELATED WORK

Low-Light Enhancement of Images:

One of the earliest algorithms for low-light image enhancement is based on Retinex theory. Jobson *et al.* [23, 22] propose center/surround Retinex at single-scale and multi-scale to achieve plausible results for dynamic range compression and color restoration. Various follow-up methods employ Retinex theory as their basis and propose complex optimization strategies to estimate reflectance and/or illumination for the purpose of low-light image enhancement [48, 14, 15, 17, 5, 31, 60, 13, 59, 41, 18]. Fu *et al.* [15] propose a weighted variational model for simultaneous reflectance and illumination estimation. Guo *et al.* [17] (LIME) perform refinement of an initial illumination map via a structure prior to obtain a well constructed illumination map thereby enabling enhancement. Li *et al.* [31] (SRIE) employ a fidelity term for gradients of the reflectance to reveal the structure details and also estimate a noise map out of their Retinex model. Ren *et al.* [41] propose a robust model to estimate reflectance and illumination maps simultaneously, with provision to suppress noise in the reflectance map. Most of the above techniques have long run-time involving CPU-based complex optimization solving for image decomposition. We also use the Retinex image-formation model as our premise. However, unlike existing techniques we do not perform the decomposition of image into reflectance and/or illumination layers, thus, achieving interactive performance on commodity hardware.

Another class of methods for low-light image enhancement is based on Histogram Equalization (HE), wherein

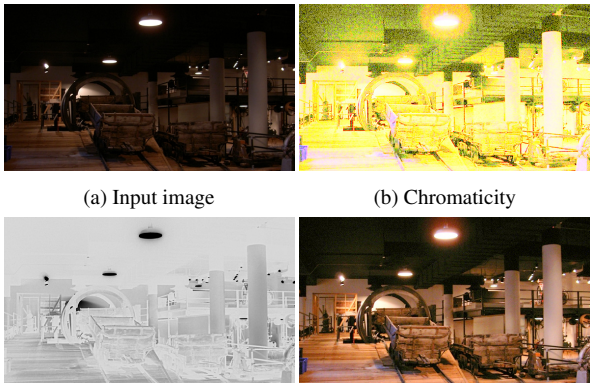
the histogram of the input image is stretched thereby improving its contrast [38]. Similar to Retinex-based approaches, various extension to the basic principle have been proposed [10, 1, 7, 26]. Celik and Tjahjadi [7] employ a variational approach for contrast enhancement using inter-pixel contextual information. Lee *et al*. [26] use a layered difference of 2D histograms and thus achieve better results than previous HE-based approaches. However, the primary focus of HE-based methods is contrast enhancement instead of physically-based illumination editing, thus having the potential risk of over- and/or under- exposed pixels.

Recently, deep learning has also been used substantially to address the problem of low-light image enhancement. Methods based on various learning strategies, such as supervised [33, 34, 49, 6, 40, 61, 51, 63], semi-supervised [54], unsupervised [16, 21, 27], and reinforcement learning [56] have been proposed. Lore *et al*. [33] present the first deep learning-based method in this context (LLNet) that employs stacked-sparse denoising autoencoder to lighten and denoise low-light images simultaneously. Lv *et al*. [34] (MBLLEN) propose an end-to-end multibranch network for simultaneous enhancement and denoising. Wei *et al*. [49] (RetinexNet) use Decom-Net for image decomposition followed by an Enhance-Net for illumination adjustment. The training is unsupervised for both the networks while the Enhance-Net also includes a joint denoising operation. Ren *et al*. [40] design an encoder-decoder network for global image enhancement and a separate recurrent neural network for further edge enhancement. Similar to Ren *et al*. [40], Zhu *et al*. [63] propose a method called EEMEFN, which consists of two stages: multi-exposure fusion and edge enhancement. Wang *et al*. [47] propose a network called DeepUPE to model image-to-image illumination and collect an expert-retouched dataset. Zhang *et al*. [61] propose a network called KinD based on Retinex theory and design a restoration module to counterbalance noise. Guo *et al*. [16] (Zero-DCE) estimates a set of best-fitting light-enhancement curves that iteratively enhances a given input image. The training is unsupervised and the method is efficient involving simple non-linear curve mapping. Chen *et al*. [8] collect a dataset named SID and train a U-Net [42] to estimate enhanced sRGB images from raw low-light images. Although learning-based methods can produce visually plausible results, they have limited generalization capability in comparison to conventional methods [28]. Moreover, unlike ours, most of the learning-based methods do not allow interactive editing of enhancement at inference time. For a new enhancement setting one has to retrain the network. Two methods which are closely related to our approach are that of Ying *et al*. [55] and Zheng *et al*. [62], both generate multiple images with different exposures followed by exposure fusion. Ying *et al*. employ a complex strategy with multiple steps to generate the exposure sequence followed by a computationally expensive optimization solving for fusion. The exposure sequence generation for Zheng *et al*. is relatively simpler than above, however, they make use of deep-learning to further enhance the sequence as an intermediate step. In comparison, our exposure sequence generation is straightforward and does not require any learning-based post-processing.

Apart from the above, existing techniques when applied on a per-frame basis, e.g., for videos, usually suffer from temporal incoherence. We prevent such inconsistency to a large degree by resorting to only point-based operations and high- or low- pass filtering.

Low-Light Enhancement of Videos:

In comparison to images, low-light video enhancement has received significantly less attention. One straightforward way to do so would be to stabilize a per-frame based application of low-light image enhancement technique using blind video consistent filtering approaches [3, 24, 43]. These techniques inherently make use of vision-based attributes such as optical flow [3, 24] or saliency masks [43] for temporal stabilization. However, computation of above vision-based attributes itself can be inaccurate/challenging for low light videos. Lv *et al*. [34] propose an extension for their learning based approach for images by replacing their 2D convolution layers with 3D ones and train it on synthetic video data. In order to collect real-world training data, Chen *et al*. [9] capture videos for static scenes with the corresponding long-exposure ground truths and ensure generalization for dynamic scenes by using a Siamese network. Jian and Zheng [20] develop a setup to capture bright and dark dynamic video pairs and subsequently train it using a modified 3D U-Net. However, with their sophisticated setup – consisting of two cameras, a relay lens and a beam splitter – the authors do not capture diverse scenes and objects as part of training data. Triantafyllidou *et al*. [45] propose a low-light video synthesis pipeline (SIDGAN) that maps "in the wild" videos into a corresponding low-light domain. The above approach employs a semi-supervised dual CycleGAN to produce dynamic video data (RAW-to-RGB) with intermediate domain mapping. In a recent work, Zhang *et al*. [57] (LLVE) enforce temporal stability for low-light video enhancement by predicting optical flow for a single image and synthesizing short range video sequences. However, their quality of enhancement is low in comparison to existing techniques (Sec. 4.4). We do not perform any temporal processing specific for videos, however our low-light image enhancement algorithm introduces only negligible temporal incoherence.



(a) Input image (b) Chromaticity

(c) Intensity Difference y (d) Adaptive Chromaticity

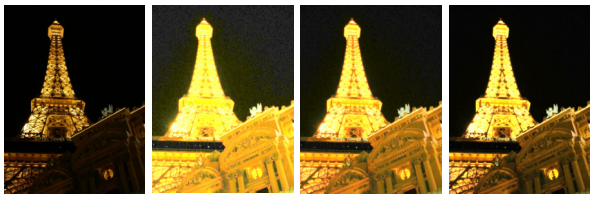
Figure 2: Given an input image (a), the noise in the chromaticity (b) is higher for low-intensity pixels with a larger intensity difference (c), which is significantly reduced for (d) adaptive chromaticity (with $\alpha = 0.3$ and $\gamma = 0.8$).

3 METHOD

According to the Retinex model, an image I can be expressed as the product of a *reflectance* layer $R \in \mathbb{R}^3$ and an *illumination* layer $L \in \mathbb{R}$ [25]: $I(x) = R(x) \times L(x)$, where the operator $\times$ denotes pixel-wise (x) multiplication. For the above equation to hold we assume only diffuse-reflection in the scene with monochromatic illumination. As a baseline, image "intensity" and "chromaticity" can be considered as the illumination and reflectance layer, respectively [4]. To compute image intensity one can employ different approaches, such as: norm or the maximum of the individual color channels. However, both does not yield desirable results for our purpose of perceptually plausible editing. To this end, we consider the *luma* (Y-channel in YCbCr color space) as our intensity operator $In(\cdot)$, since this satisfies the above objective. Chromaticity is correspondingly obtained by dividing the image with its intensity (Eqn. (1)). The above division operation is able to significantly reduce shading and shadows in the scene, which only affects the intensity, thus making the chromaticity relatively brighter than the input image. Moreover, it also acts as a normalizing factor for pixel color and saturates it, further making it appear perceptually bright. For an input image I with color channels r, g, and b in sRGB color space using 8-bit per channel (i.e., 24-bit color depth), we define intensity (following ITU-R BT.601) by the operator $In(\cdot)$ and chromaticity C as follows:

$$In(I) = 0.299r + 0.587g + 0.144b \text{ and } C = \frac{I}{In(I)}. \quad (1)$$

The brightening aspect of chromaticity is a preferable characteristic for low-light image enhancement. However, chromaticity suffers from undesirable artifacts in terms of *noise* and *color-shifts*, especially for low-intensity pixels (Fig. 2b).



(a) Input (b) y^2 (c) $\exp(y)$ (d) $\tan(y \cdot \frac{\pi}{2})$

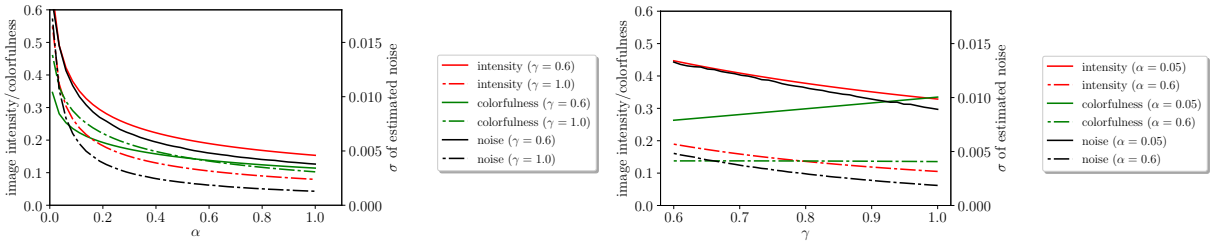
Figure 3: Our Single-Exposure (SE) output for $\alpha = 0.05$, and $\gamma = 0.7$ employing different adaptive functions $f(y)[= y^2, \exp(y), \tan(y \cdot \frac{\pi}{2})]$.

3.1 Adaptive Chromaticity

In order to preserve the brightening effect of chromaticity while avoiding artifacts, we introduce *Adaptive Chromaticity (AC)*. For identifying a low-intensity pixel, we compute the difference between pixel intensity, $In(\cdot)$, and the maximum intensity value $MaxIn$. For low-intensity pixels, this difference defined as $y = MaxIn - In(\cdot)$ would be comparatively larger. For example, for an intensity image encoded in the range of 0 to 1, $MaxIn = 1$ and for a low-intensity pixel p with $In(\cdot) = 0.05$ the difference $y(p) = 0.95$ is large. Similarly, for a high-intensity pixel q with $In(\cdot) = 0.8$ the difference $y(q) = 0.2$ is small (Fig. 2c). The above forms the basis for defining adaptive chromaticity (A_c), wherein we add an adaptive term, as a function of y, in the denominator while computing chromaticity (Eqn. (1)). To further increase the brightness and prevent color-shifts, we perform a non-linear scaling using *gamma correction*

$$A_c(I, \alpha, \gamma) = \left(\frac{I}{In(I) + \alpha f(y)} \right)^{\gamma}. \quad (2)$$

Here, $f(y)$ is a function in terms of y, α is a control parameter, and γ is a parameter for gamma correction. The adaptive function $f(y)$ should be chosen such that its value is close to zero when y is small and is substantially high for large values of y. Thus, by tuning the control parameter α, we can smoothly transition between the bright chromaticity (when $\alpha \to 0$) and a complete dark image (when $\alpha \to \infty$). The intuition behind the adaptive denominator in Eqn. (2) is that we divide by a larger value for low-intensity pixels as compared to high-intensity pixels, thereby, reducing undesirable artifacts. For adaptivity, a function f should be chosen that satisfies the above property and is efficient to compute. Among possible variants, y^2 and $\exp(y)$ produces desirable results. However, $f(y) = \tan(y \cdot \frac{\pi}{2})$ works significantly better in terms of noise reduction and also gives plausible results, see Fig. 3. The AC brightens an image while significantly reducing chromaticity-related artifacts (Fig. 2d) and forms the basis for our low-light image and video enhancement methodology.



(a) Impact of varying α on the resultant AC.

(b) Impact of varying γ on the resultant AC.

Figure 4: Changes in the characteristics of resultant Adaptive Chromaticity in terms of intensity, colorfulness and noise while varying α (Fig. 4a) and γ (Fig. 4b). Intensity is computed using Eqn. (1). Colorfulness represents the perceptual amount of saturation following [19]. Image noise is calculated using skimage estimate_sigma [11] based on a wavelet-based estimator [12] of the gaussian noise standard deviation σ. Metrics are computed and averaged over the LIME dataset [17].

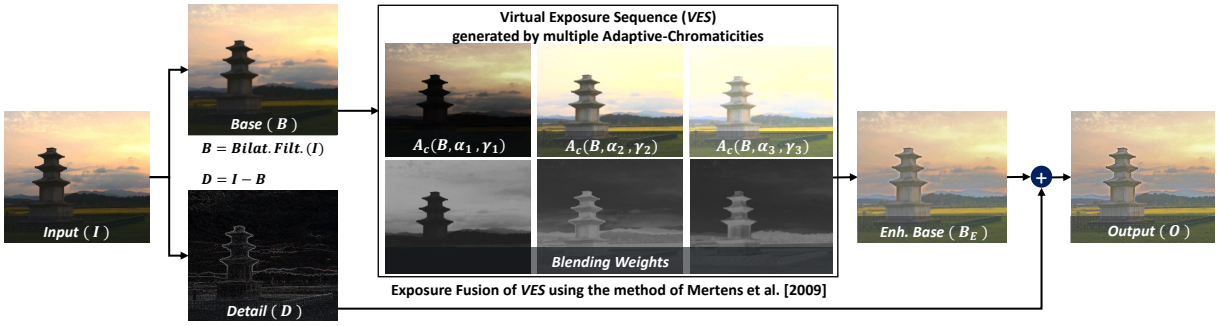


Figure 5: Flowchart of our low-light image enhancement algorithm. To prevent noise amplification we decompose the input image into *Base* and *Detail* layers. Subsequently, multiple Adaptive Chromaticities (ACs) are generated (Sec. 3.1) for the Base layer to create a Virtual Exposure Sequence (VES) (Sec. 3.2). Following to that, these images are blended guided by quality measures of contrast, saturation, and well-exposedness (Sec. 3.2). The above is performed in a multi-resolution fashion, as proposed by Mertens *et al.* [35]. Finally, the Detail layer is added to the enhanced Base layer to obtain the final output.

Parameter Analysis:

We analyse the changes in the characteristics of resultant AC in terms of image *intensity*, *colorfulness* and *noise* while varying the parameters α and γ in Fig. 4. Decreasing α leads to a quadratic increase in all three metrics (Fig. 4a). Moreover, note the significant decline in noise for higher values of alpha (> 0.8). On the other hand, decreasing γ linearly increases noise and intensity, while at the same time desaturates the image (Fig. 4b). The desaturating nature of γ plays a counterbalancing role to the effect of α in terms of *colorfulness* thereby preventing color-shifts. It is thus evident, that both α and γ needs to be adjusted to brighten the image while retaining the original saturation level, and also highlights that denoising is an essential requirement during low-light enhancement.

3.2 Our Approach for LLIE

To further reduce noise amplification during enhancement, we decompose the input image into *Base* (B) and *Detail* (D) components [2], and only enhance *Base*, as depicted in our full LLIE-algorithm flowchart in Fig. 5. We assume that most of the noise due to low-light conditions is captured in the high-frequency *Detail* layer. Thus, enhancing only *Base* layer will lead to negligible noise amplification. For base-detail decomposition we make use of Bilateral Filter [44], however, in principle, one can use any edge-preserving filter for this purpose. We use

$$B = BilatFilt(I,\ \sigma_s,\ \sigma_t) \quad D = I - B \tag{3}$$

where $\sigma_s = 1.0$ (spatial width) and $\sigma_t = 0.5$ (tonal range) works fine with most images (or video-frames). Following the above decomposition, the *Base* layer is enhanced via AC using a single-exposure (SE) or multiple-exposure (ME) setting. In either case, subsequently the *Detail* layer is added to the enhanced *Base* to obtain the final result (Fig. 6). For single-exposure, a single AC of base layer is assigned as its enhanced version (Fig. 6e). Multi-exposure enhancement involves computing multiple ACs of the base layer and is proposed as a two-step process consisting of Virtual Exposure Sequence (VES) *generation* and *fusion*.

VES Generation:

The overall exposedness of an image is increased by lowering α and/or γ values in Eqn. (2). However, the brightening effect of either of these parameters α or γ is slightly different. For lower values of α, increase in brightness comes at the cost of color-shifts (Fig. 7a, Fig. 7d). On the other hand, for lower γ values, an

(a) Input image (b) Chromaticity (c) Base (d) Detail (e) Single Exp. (SE) (f) Multi Exp. (ME)

Figure 6: For a low-light image (a) the corresponding chromaticity (b) has artifacts in terms of color-shifts and noise. To overcome noise amplification, we decompose the image into *Base* (c) and *Detail* (d) layers using a bilateral filter ($\sigma_s = 1.0, \sigma_t = 0.5$). Further, chromaticity-based artifacts are reduced by employing Adaptive Chromaticity (AC) and for the single-exposure approach, an enhanced image is obtained as the sum of AC ($\alpha = 0.1, \gamma = 0.8$) of Base layer and Detail layer (e). To further preserve details during enhancement, we use a multi-exposure fusion technique (3 exposure levels – $\alpha_1 = 0.03, \alpha_2 = 0.1, \alpha_3 = 2.0$ and $\gamma_1 = 0.7, \gamma_2 = 0.8, \gamma_3 = 0.5$ – and 4 pyramid levels) to obtain a high-quality output (f).

increase in brightness is accompanied with desaturation (Figs. 7d to 7f). For both α and γ, lower values leads to increase in noise (Fig. 7d) (see Sec. 3.1). Increasing the exposedness by tuning either α or γ is a point-based operation and does not respect the relative contrast within the image. The above leads to the problem, wherein already visible regions in the low-light image are over-exposed while increasing the brightness. It is similar to challenges in High Dynamic Range (HDR) photography, which aims to preserve all the details within an HDR scene.

We do not assume an HDR version of the image at our disposal, however we can generate different levels of brightness by varying the values of α and γ respectively. Thus, we generate a *virtual exposure sequence* for the given base layer by computing multiple ACs. For the base layer B, an exposure sequence $\{E_k | \, k = 1 \ldots N\}$ is obtained based on the parameter series $\{(\alpha_k, \gamma_k) \, | \, k = 1 \ldots N\}$, with

$$E_k = A_c(B, \alpha_k, \gamma_k). \tag{4}$$

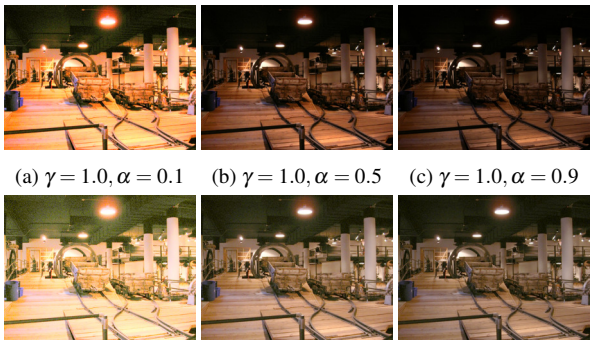
Subsequently, an HDR image can be generated using the above sequence of images and further tone-mapping can preserve details in both bright and dark regions while enhancing it [39].

VES Fusion:

For efficiency, we avoid the step of computing an HDR image, and directly fuse the multiple exposures into a high-quality, low dynamic range image using the exposure-fusion technique of Mertens *et al.* [35]. The well-exposedness of an image in the exposure sequence is determined based on quality measures of *contrast* (c_k), *saturation* (s_k), and *well-exposedness* (e_k) on a per-pixel basis.The three quality measures are combined into a joint weighting function

$$w_k = {c_k}^{\upsilon_c} \cdot {s_k}^{\upsilon_s} \cdot {e_k}^{\upsilon_e}, \tag{5}$$

where the above product can be seen as logical conjunction and the parameters υ_c, υ_s, and υ_e control the influence of individual quality measures. Finally, the obtained sequence of weight maps are normalized such that they sum up to one at each pixel location, thereby ensuring consistent results, as follows:



(a) $\gamma = 1.0, \alpha = 0.1$ (b) $\gamma = 1.0, \alpha = 0.5$ (c) $\gamma = 1.0, \alpha = 0.9$

(d) $\gamma = 0.5, \alpha = 0.1$ (e) $\gamma = 0.5, \alpha = 0.5$ (f) $\gamma = 0.5, \alpha = 0.9$

Figure 7: Virtual Exposure Sequence (VES) for the input image in Fig. 2: as a sequence of ACs generated by varying values of α and γ.

$$\widehat{w_k} = \frac{w_k}{\sum_{k=1}^{N} w_k}. \tag{6}$$

Once the weight maps are computed, a Laplacian pyramid $L(E_k)$ of each image and a Gaussian pyramid of each normalized weight map $G(\widehat{w_k})$ are generated. At each pyramid level l, the images are fused on per-pixel and per-color channel basis as

$$L(B_E)_l = \sum_{k=1}^{N} G(\widehat{w_k})_l L(E_k)_l. \tag{7}$$

The enhanced base layer, B_E, is obtained by collapsing the computed Laplacian pyramid $L(B_E)$. Following the above, we sum the detail layer (D) and the enhanced base layer (B_E) to obtain the final output O where,

$$O = B_E + \eta D, \tag{8}$$

and $\eta > 1$ is used to amplify the details in the final output [37]. However, large values of η (> 4.0) leads to halo-artifacts and unnatural looks. For most images

$\eta = 2.0$ gives visually plausible results. Note that the operations defined in Eqn. (1) till Eqn. (8) are all point-based where we have omitted the pixel-location x for the sake of clarity. All the steps in our method are efficiently summarized in Algo. 1.

4 RESULTS

4.1 Parameter Settings

The enhancement of the base layer for our Multi-Exposure (ME) version consists of two steps, for which the parameter settings are discussed as follows.

VES Generation:

Ideally, to capture fine details at different exposure levels, multiple images are required for the exposure sequence. However, the processing time will increase according to the number of images. Empirically, we determine three exposure levels ($N = 3$) as sufficient to preserve details at different levels of brightness. Further, we empirically determine $\gamma \in [0.6, 1.0]$ and $\alpha \in [0.01, 3.0]$ to result in well-exposed and less-noisy outputs. For most of the images, $\gamma_1 = 0.7, \alpha_1 = 0.03$ (high-exposure level), $\gamma_2 = 0.8, \alpha_2 = 0.1$ (mid-exposure level), and $\gamma_3 = 0.5, \alpha_3 = 2.0$ (low-exposure level) yield desirable results. For all the results in the paper, unless stated otherwise, we use the above parameter settings.

VES Fusion:

For exposure fusion, we set the weighting exponents for the quality measures to $\upsilon_c = \upsilon_s = \upsilon_e = 1$, as suggested by Mertens *et al*. [35]. During fusion, higher number of pyramid-levels facilitate the preservation of fine details. However, processing time increases with the number of levels, which is more pronounced for high-resolution images. Empirically, we determine four pyramid levels ($M = 4$) as a good trade-off between performance and quality.

4.2 Qualitative and Quantitative Results

We compare our results with state-of-the-art image-based methods: two conventional methods (SRIE [31] and LIME [17]), two supervised-learning based methods (MBBLEN [34] and RetinexNet [49]), a unsupervised-learning based method (Zero-DCE [16]), and a video-based method (LLVE [57]). The results are produced from publicly available source codes with respective parameter settings.

Images:

We test the above methods on images taken from the following datasets: LIME [17] (10 images), DICM [26] (44 images), NPE [48] (72 images), and VV [46] (24 images). For quantitative evaluation, we employ the Natural Image Quality Evaluator (NIQE) [36] metric to compare the performance of different methods on the above datasets. We choose this metric, as it is provides a completely blind quality measure for images and is based on only deviations from statistical regularities in natural images. Tab. 2 shows that overall we perform better than compared approaches except for Zero-DCE. We present qualitative comparison for enhanced image outputs in Fig. 8. The results of LIME(Fig. 8(b)) tends to be over-exposed, MBLLEN provides satisfactory brightening (Fig. 8(d)) however tends to over-smooth image details, the output of RetinexNet (Fig. 8(e)) seems to look unnatural, and for LLVE the results (Fig. 8(g)) appear to be hazy and desaturated. Our results are visually comparable to Zero-DCE and SRIE. However, in contrast to Zero-DCE, which requires a re-training of the complete network for a different degree of enhancement, our approach allows for interactive enhancement manipulation. Further, the slow optimization solving in SRIE makes it orders of magnitude slower than our approach Tab. 3. Moreover, the outcome of our user study, which includes a broad range of images (Fig. 9), indicates that overall our method is preferred over them.

For subjective evaluation of our method in the context of images, we perform a user study similar to Zhang *et al*. [57] comparing different techniques. We employ 9 different images (2 from LIME [17], 2 from DICM [26], 2 from NPE [48], and 3 from VV [46] datasets respectively) and compare 6 other techniques (5 image-based and 1 video-based) against our method. Thereby constituting 54 blind A/B tests which are presented in a random fashion to each participant. In total, 13 persons (7 female and 6 male) within the ages of 22 to 38 years participated in the study. We asked the participants to focus on the following aspects during comparison:

Exposure: As compared to the input, the enhanced image should be well-exposed, neither under- nor over-exposed.

Noise (and flickering): The enhanced image should have less noise (and flickering – in case of videos). However, the denoising should not be excessive as to remove details.

Color: The colors in the enhanced image should appear natural and it should not look over- or under-saturated.

For every low-light image, the participant is shown two enhanced versions of the image simultaneously (one of them is ours) and is asked to pick the version of their choice based on the above criteria. For the majority of

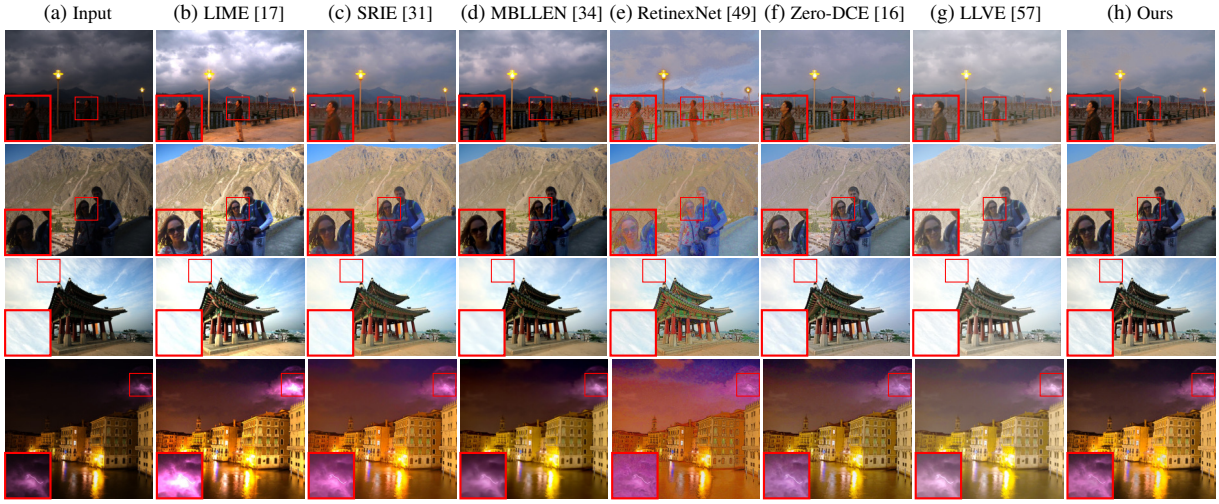


Figure 8: Low-light image enhancement results. Input images are taken from LIME [17], DICM [26], and VV [46] datasets.

Table 2: NIQE [36] (↓) values for images in LIME [17], DICM [26], NPE [48], and VV datasets. The best value is shown in red and the next best in blue.

Method	DICM	LIME	NPE	VV	Avg.
LIME	2.99	3.67	3.02	2.99	3.05
SRIE	3.27	4.29	3.45	3.25	3.42
MBLLEN	3.16	3.69	3.15	3.31	3.21
RetinexNet	3.59	3.63	3.62	**2.62**	3.45
LLVE	3.10	3.65	**2.98**	2.86	3.04
Zero-DCE	**2.48**	**3.10**	**2.92**	2.87	**2.79**
Ours	**2.84**	**3.22**	3.00	**2.66**	**2.92**

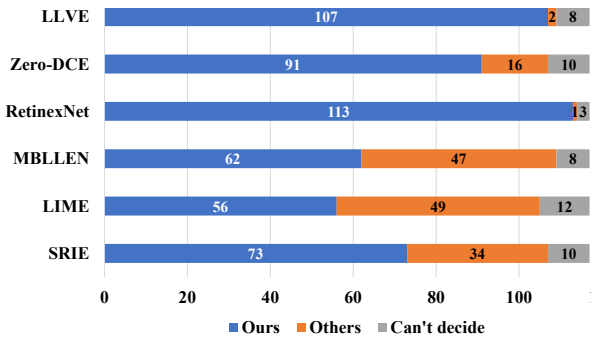


Figure 9: Statistics of user study results on low-light image enhancement. For 13 participants and 9 different images, we compare each existing method against ours through a total of 117 randomized A/B tests.

cases participants prefer our method against the existing approaches, see Fig. 9.

Videos:

To evaluate video-enhancement results, we perform a subjective user study similar to that of images explained in the previous paragraph. As test data, we make use of the challenging low-light videos provided by Li *et al.* in their survey LLIV [28]. In total, 13 persons (3 female, and 10 male) within the ages of 19 to 42 years participated in the study. Note that the above group of participants did not participate in the images-based user study to avoid any inherent bias between both the studies. The experiment consisted of 7 different low-light videos enhanced by ours and 6 other (5 image-based and 1 video-based) approaches. Two enhanced videos are shown to a participant simultaneously (one of them is ours), thereby constituting 42 blind A/B tests which are shown in a randomized order to each participant. Fig. 10 shows that our method surpasses all other methods including LLVE by a large margin.

4.3 Face Detection in the Dark

We investigate the performance of low-light enhancement methods for increasing the face-detection accuracy on low-light images. Specifically, following the settings presented in Li *et al.* [28], we use 500 randomly sampled images from the DARK FACE dataset [53] to measure performance of the state-of-the-art Dual Shot Face Detector (DSFD) [30] trained on the WIDER FACE dataset [52]. We use the author's DSFD implementation [29] with a non-maximum suppression threshold of 0.3 and evaluate using the dark face UG2 challenge evaluation tool [50]. Fig. 11 depicts the precision-recall curves as well as average precision (AP) under a 0.5 IoU threshold. The results show that all low-light enhancement methods achieve a substantial improvement in precision and recall over the unprocessed images (baseline result). For our method, the ME setting does not increase detection rates significantly. Moreover, we observe that shifting faces into a brightness and contrast range that the classifier

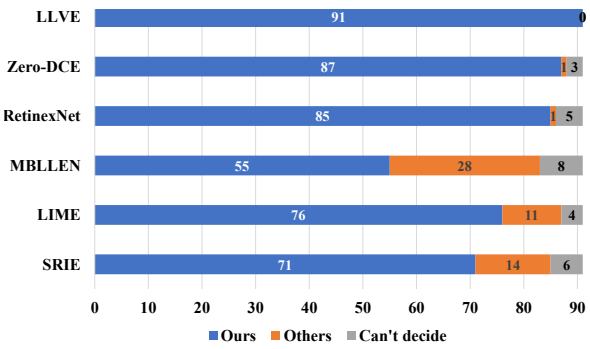


Figure 10: Statistics of the user study results on low-light video enhancement. For 13 participants and 7 different videos from LLIV [28] dataset, we compare each existing method against ours through a total of 91 randomized A/B tests.

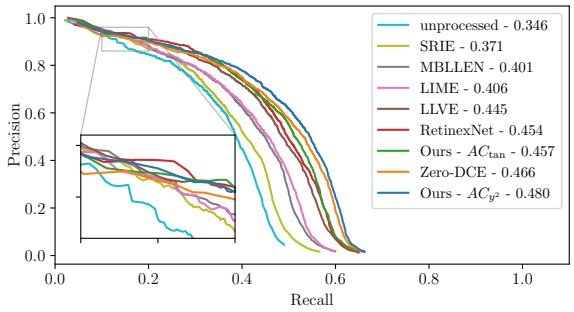


Figure 11: Precision-recall curves for face detection using DSFD [30] on dark-face images [53] enhanced using different LLIE methods. Average precision (AP) of each method is indicated in the legend, where "unprocessed" denotes the baseline AP on images in [53]. Our method uses adaptive chromaticity (AC) without exposure fusion, we compare two variants for $f(y)$, namely $f(y) = y^2$ with $\alpha = 0.25, \gamma = 0.6$ and $f(y) = \tan(y\frac{\pi}{2})$ with $\alpha = 0.25, \gamma = 0.3$.

has been trained on is crucial for accuracy improvement irrespective of overall image aesthetics. Thus we employ only Adaptive Chromaticity (AC) for this purpose. We investigate the performance of A_c (Eqn. (2)) for two different versions of $f(y)$, and find that while $f(y) = \tan(y\frac{\pi}{2})$ achieves visually more pleasing results, $f(y) = y^2$ outperforms all other LLIE methods on face detection. Overall our results show that AC adjustment is a simple and efficient pre-processing method for boosting detection accuracy on low-light images which outperforms more sophisticated techniques.

4.4 Run-time Performance Evaluation

All our experiments were performed on an consumer PC using Microsoft Windows 10 as operating system, with a 2.2 GHz (Intel i7) CPU, 16 GB of RAM, and a Nvidia GTX 1050 Ti graphics card with 4 GB VRAM. Our full algorithm, implemented with C++ and CUDA (v10.0), runs at real-time for VGA resolution images (Tab. 3) and at interactive frame rates on HD and FHD resolution images. Unlike ours, most of the existing

Table 3: Run-time performance of various methods in milliseconds. The top three run-time performance values for each resolution are shown in red, blue, and brown colors respectively. Note, that all learning-based methods except LIME and SRIE make use of the GPU. We were not able to run certain methods at higher resolutions due to Out-of-Memory (OOM) exceptions.

Method	VGA 640 × 480	HD 1280 × 720	FHD 1920 × 1080	QHD 2560 × 1440
LIME	580	1940	6450	**10180**
SRIE	11820	49830	OOM	OOM
MBLLEN	430	1300	3010	OOM
RetinexNet	1030	3710	7590	17540
LLVE	110	310	700	OOM
Zero-DCE	**4.69**	**11.77**	**25.75**	OOM
Ours SE	**4.87**	**12.91**	**28.59**	**49.49**
Ours ME	**59.58**	**180**	**410**	**740**

techniques are either not able to handle QHD resolution or are significantly slower for the given hardware configuration. Excluding the SE setting, our ME version performs better than all the other methods except Zero-DCE [16]. While AC forms the basis of our approach, more than 90% of the processing time is spent on multi-pyramid based exposure fusion. For the SE setting, the result obtained has artifacts in the form of over-exposedness and color-shifts, however, provides a reasonable approximation for the enhanced image. Thus, the SE version, our fast variant, can potentially serve as a preview of the enhanced output and allow for further interactive parameter editing.

5 DISCUSSION

Most of the existing methods, including ours, face three major challenges for LLIE. First is the trade-off between under- and over-exposedness. In order to expose the low-lit regions within an image, one might over-expose existing well-exposed parts. We approached the above to a large extent via adaptive computation of chromaticity and further by making use of an exposure sequence and multi-pyramid based blending. As a generic approach, one can compute the degree of exposure for different image regions, as an exposure mask, in a pre-processing step and use it for further processing. Second is the introduction and amplification of noise while enhancing images. To prevent the above, we first decompose the image into *base* and *detail* layers. However, a more sophisticated denoising scheme specifically tailored for low-light noise might perform better for this purpose. Thirdly, the enhancement process can result in changes in perceived color. In our approach, such changes are limited due to the counter-balancing effect of α and γ on the perceived colorfulness.

(a) Input (b) Ours (c) Ours + Denoising (d) MBLLEN (e) SRIE

Figure 12: Our result can further be improved by a post-processing denoising operation. Here, we compare our denoised-output (denoising done using FFDNET [58]) with that of MBLLEN [34] and SRIE [31].

Limitations: In order to tackle the issue of noise most of the existing techniques either employ denoising priors in their objective formulation [31], perform denoising as a post-processing operation [17], or introduce synthetic noise during training [34, 57]. We do not include any explicit denoising step in our methodology and still perform better both qualitatively and quantitatively. However, among the possible challenges in low-light image enhancement we are less effective in terms of noise-removal. The above is reflected to a certain degree during the user study where we observed that on certain occasions participants preferred the method of Lv *et al.* [34] and Li *et al.* [31] due to their less-noisy results. We conjecture that this preference can be shifted in our favor by performing a post-processing denoising operation. Note, that our denoised output in Fig. 12c has better quality and does not suffer from artifacts such as over-exposure (as in Fig. 12d) or color-shifts (as in Fig. 12e).

6 CONCLUSIONS & FUTURE WORK

This paper presents a simple yet effective technique to enhance low-light images and videos. The key to our approach is Adaptive Chromaticity that allows to efficiently increase the image brightness. Our SE version runs at real-time frame rates and can be used for a fast enhancement preview. To further improve results, we generate a virtual exposure sequence by computing multiple adaptive chromaticities for the *base* layer followed by a multi-pyramid based fusion. Our ME version runs at interactive frame rates, even for high resolution images. Experimental results validate the advancement of our approach in comparison to various state-of-the-art alternatives. For the above, we perform both quantitative and qualitative evaluation including subjective user studies. As part of future work we plan to include a denoising step in our algorithm and potentially use the multi-scale nature of exposure-fusion for this purpose. For videos, we plan to use the neighboring frames to improve the denoising as well as enhancement quality.

7 ACKNOWLEDGMENTS

We thank the participants who took part in the user study. This work was partially funded by the German Federal Ministry of Education and Research

Algorithm 1: Our Low-light Image Enhancement Algorithm

Input: Input image I, Bilateral Filter parametrs σ_s, σ_t, Adaptivity parameters $\alpha_1, \ldots, \alpha_N$, Gamma correction parameters $\gamma_1, \ldots, \gamma_N$, Exposure fusion parameters $\sigma, \upsilon_c, \upsilon_s, \upsilon_e$, Exposure levels – N, Pyramid levels – M, Additive parameter η

Output: Enhanced output image O

$B \leftarrow$ `BilateralFilter`(I, σ_s, σ_t) `// Base Layer`
$D \leftarrow I - B$ `// Detail Layer`
$wtSum \leftarrow 0$
for $k \leq 1$ *to* N **do**
  $E_k \leftarrow A_c(B, \alpha_k, \gamma_k)$ `// Generate exposure series`
  $w_k \leftarrow$ `ComputeWeights`$(E_k, \sigma, \upsilon_c, \upsilon_s, \upsilon_e)$ `// Eq. 5`
  $wtSum \leftarrow wtSum + w_k$
$outerSum \leftarrow 0$
for $k \leq 1$ *to* N **do**
  innerSum $\leftarrow 0$
  $\widehat{w_k} \leftarrow w_k / wtSum$
  $tmp1 \leftarrow E_k$
  $G(\widehat{w_k})_l \leftarrow \widehat{w_k}$
  for $l \leq 1$ *to* M **do**
    $tmp2 \leftarrow$ `GaussianFilter`$(tmp1, \sigma = l)$ `// "l" is the Gaussian Filter kernel width`
    $L(E_k)_l \leftarrow tmp1 - tmp2$ `// Laplacian pyramid of Base exposure levels`
    $G(\widehat{w_k})_l \leftarrow$ `GaussianFilter`$(G(\widehat{w_k})_l, \sigma = l)$
    $innerSum \leftarrow innerSum + G(\widehat{w_k})_l \cdot L(E_k)_l$
    $tmp1 \leftarrow tmp2$
  $innerSum \leftarrow innerSum + G(\widehat{w_k})_l \cdot tmp2$
  $outerSum \leftarrow outerSum + innerSum$
$B_E \leftarrow outerSum$ `// Enhanced Base Layer`
$O \leftarrow B_E + \eta D$ `// Enhanced Output Image`

(BMBF) through grants 01IS18092 ("mdViPro") and 01IS19006 ("KI-LAB-ITSE") and the Research School on "Service-Oriented Systems Engineering" of the Hasso Plattner Institute.

REFERENCES

[1] M. Abdullah-Al-Wadud et al. "A Dynamic Histogram Equalization for Image Contrast Enhancement". In: *IEEE Transactions on Consumer Electronics* 53.2 (2007), pp. 593–600. DOI: 10.1109/TCE.2007.381734.

[2] Soonmin Bae, Sylvain Paris, and Frédo Durand. "Two-Scale Tone Management for Photographic Look". In: *ACM SIGGRAPH 2006 Papers*. SIGGRAPH '06. 2006, pp. 637–645. DOI: 10.1145/1179352.1141935.

[3] Nicolas Bonneel et al. "Blind Video Temporal Consistency". In: *ACM Trans. Graph.* 34.6 (2015). ISSN: 0730-0301. DOI: 10.1145/2816795.2818107.

[4] Nicolas Bonneel et al. "Intrinsic Decompositions for Image Editing". In: *Computer Graphics Forum* 36.2 (May 2017), pp. 593–609. ISSN: 0167-7055. DOI: 10.1111/cgf.13149.

[5] Bolun Cai et al. "A Joint Intrinsic-Extrinsic Prior Model for Retinex". In: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017, pp. 4020–4029. DOI: 10.1109/ICCV.2017.431.

[6] Jianrui Cai, Shuhang Gu, and Lei Zhang. "Learning a Deep Single Image Contrast Enhancer from Multi-Exposure Images". In: *IEEE Transactions on Image Processing* 27.4 (2018), pp. 2049–2062. DOI: 10.1109/TIP.2018.2794218.

[7] Turgay Celik and Tardi Tjahjadi. "Contextual and Variational Contrast Enhancement". In: *IEEE Transactions on Image Processing* 20.12 (2011), pp. 3431–3441. DOI: 10.1109/TIP.2011.2157513.

[8] Chen Chen et al. "Learning to See in the Dark". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018, pp. 3291–3300. DOI: 10.1109/CVPR.2018.00347.

[9] Chen Chen et al. "Seeing Motion in the Dark". In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2019, pp. 3184–3193. DOI: 10.1109/ICCV.2019.00328.

[10] H.D. Cheng and X.J. Shi. "A simple and effective histogram equalization approach to image enhancement". In: *Digital Signal Processing* 14.2 (2004), pp. 158–170. ISSN: 1051-2004. DOI: https://doi.org/10.1016/j.dsp.2003.07.002.

[11] skimage v0.19.0 docs. *estimate_sigma*. 2022. URL: https://scikit-image.org/docs/stable/api/skimage.restoration.html#skimage.restoration.estimate_sigma (visited on 01/27/2022).

[12] David L Donoho and Iain M Johnstone. "Ideal spatial adaptation by wavelet shrinkage". In: *Biometrika* 81.3 (1994), pp. 425–455. ISSN: 0006-3444. DOI: 10.1093/biomet/81.3.425.

[13] Gang Fu, Lian Duan, and Chunxia Xiao. "A Hybrid L2 -Lp Variational Model For Single Low-Light Image Enhancement With Bright Channel Prior". In: *2019 IEEE International Conference on Image Processing (ICIP)*. 2019, pp. 1925–1929. DOI: 10.1109/ICIP.2019.8803197.

[14] Xueyang Fu et al. "A Probabilistic Method for Image Enhancement With Simultaneous Illumination and Reflectance Estimation". In: *IEEE Transactions on Image Processing* 24.12 (2015), pp. 4965–4977. DOI: 10.1109/TIP.2015.2474701.

[15] Xueyang Fu et al. "A Weighted Variational Model for Simultaneous Reflectance and Illumination Estimation". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 2782–2790. DOI: 10.1109/CVPR.2016.304.

[16] Chunle Guo et al. "Zero-DCE: Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020, pp. 1777–1786. DOI: 10.1109/CVPR42600.2020.00185.

[17] Xiaojie Guo, Yu Li, and Haibin Ling. "LIME: Low-Light Image Enhancement via Illumination Map Estimation". In: *IEEE Transactions on Image Processing* 26.2 (2017), pp. 982–993. DOI: 10.1109/TIP.2016.2639450.

[18] Shijie Hao et al. "Low-Light Image Enhancement With Semi-Decoupled Decomposition". In: *IEEE Transactions on Multimedia* 22.12 (2020), pp. 3025–3038. DOI: 10.1109/TMM.2020.2969790.

[19] David Hasler and Sabine E. Suesstrunk. "Measuring colorfulness in natural images". In: *Human Vision and Electronic Imaging VIII*. Vol. 5007. International Society for Optics and Photonics. 2003, pp. 87 –95. DOI: 10.1117/12.477378.

[20] Haiyang Jiang and Yinqiang Zheng. "Learning to See Moving Objects in the Dark". In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2019, pp. 7323–7332. DOI: 10.1109/ICCV.2019.00742.

[21] Yifan Jiang et al. "EnlightenGAN: Deep Light Enhancement Without Paired Supervision". In: *IEEE Trans. Image Process.* 30 (2021), pp. 2340–2349. DOI: 10.1109/TIP.2021.3051462.

[22] D.J. Jobson, Z. Rahman, and G.A. Woodell. "A multiscale retinex for bridging the gap between color images and the human observation of scenes". In: *IEEE Transactions on Image Processing* 6.7 (1997), pp. 965–976. DOI: 10.1109/83.597272.

[23] D.J. Jobson, Z. Rahman, and G.A. Woodell. "Properties and performance of a center/surround retinex". In: *IEEE Transactions on Image Processing* 6.3 (1997), pp. 451–462. DOI: 10.1109/83.557356.

[24] Wei-Sheng Lai et al. "Learning Blind Video Temporal Consistency". In: *Computer Vision – ECCV 2018*. Ed. by Vittorio Ferrari et al. 2018, pp. 179–195. DOI: 10.1007/978-3-030-01267-0_11.

[25] Edwin H Land and John J McCann. "Lightness and retinex theory". In: *Journal of the Optical Society of America* 61.1 (1971), pp. 1–11. DOI: 10.1364/JOSA.61.000001.

[26] Chulwoo Lee, Chul Lee, and Chang-Su Kim. "Contrast Enhancement Based on Layered Difference Representation of 2D Histograms". In: *IEEE Transactions on Image Processing* 22.12 (2013), pp. 5372–5384. DOI: 10.1109/TIP.2013.2284059.

[27] Hunsang Lee, Kwanghoon Sohn, and Dongbo Min. "Unsupervised Low-Light Image Enhancement Using Bright Channel Prior". In: *IEEE Signal Processing Letters* 27 (2020), pp. 251–255. DOI: 10.1109/LSP.2020.2965824.

[28] Chongyi Li et al. "Low-Light Image and Video Enhancement Using Deep Learning: A Survey". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021), pp. 1–1. DOI: 10.1109/TPAMI.2021.3126387.

[29] Jian Li and Yabiao Wang. *Tencent/FaceDetection-DSFD*. 2019. URL: https://github.com/Tencent/FaceDetection-DSFD.

[30] Jian Li et al. "DSFD: dual shot face detector". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 5060–5069. DOI: 10.1109/CVPR.2019.00520.

[31] Mading Li et al. "SRIE: Structure-Revealing Low-Light Image Enhancement Via Robust Retinex Model". In: *IEEE Transactions on Image Processing* 27.6 (2018), pp. 2828–2841. DOI: 10.1109/TIP.2018.2810539.

[32] Jiaying Liu et al. "Benchmarking Low-Light Image Enhancement and Beyond". In: *International Journal of Computer Vision* 129.4 (2021), pp. 1153–1184. ISSN: 1573-1405. DOI: 10.1007/s11263-020-01418-8.

[33] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. "LLNet: A deep autoencoder approach to natural low-light image enhancement". In: *Pattern Recognition* 61 (2017), pp. 650–662. ISSN: 0031-3203. DOI: https://doi.org/10.1016/j.patcog.2016.06.008.

[34] Feifan Lv et al. "MBLLEN: Low-Light Image/Video Enhancement Using CNNs". In: *British Machine Vision Conference 2018, BMVC 2018, Newcastle, UK, September 3-6, 2018*. BMVA Press, 2018, p. 220. URL: http://bmvc2018.org/contents/papers/0700.pdf.

[35] T. Mertens, J. Kautz, and F. Van Reeth. "Exposure Fusion: A Simple and Practical Alternative to High Dynamic Range Photography". In: *Computer Graphics Forum* 28.1 (2009), pp. 161–171. DOI: https://doi.org/10.1111/j.1467-8659.2008.01171.x.

[36] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. "Making a Completely Blind Image Quality Analyzer". In: *IEEE Signal Processing Letters* 20.3 (2013), pp. 209–212. DOI: 10.1109/LSP.2012.2227726.

[37] Franck Neycenssac. "Contrast Enhancement Using the Laplacian-of-a-Gaussian Filter". In: *CVGIP: Graph. Models Image Process.* 55.6 (1993), pp. 447–463. DOI: 10.1006/cgip.1993.1034.

[38] Stephen M. Pizer et al. "Adaptive histogram equalization and its variations". In: *Computer Vision, Graphics, and Image Processing* 39.3 (1987), pp. 355–368. ISSN: 0734-189X. DOI: https://doi.org/10.1016/S0734-189X(87)80186-X.

[39] Erik Reinhard et al. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.

[40] Wenqi Ren et al. "Low-Light Image Enhancement via a Deep Hybrid Network". In: *IEEE Transactions on Image Processing* 28.9 (2019), pp. 4364–4375. DOI: 10.1109/TIP.2019.2910412.

[41] Xutong Ren et al. "LR3M: Robust Low-Light Enhancement via Low-Rank Regularized Retinex Model". In: *IEEE Transactions on Image Processing* 29 (2020), pp. 5862–5876. DOI: 10.1109/TIP.2020.2984098.

[42] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Ed. by Nassir Navab et al. 2015, pp. 234–241. ISBN: 978-3-319-24574-4. DOI: 10.1007/978-3-319-24574-4_28.

[43] Sumit Shekhar et al. "Consistent Filtering of Videos and Dense Light-Fields Without Optic-Flow". In: *Vision, Modeling and Visualization*. 2019. DOI: 10.2312/vmv.20191326.

[44] C. Tomasi and R. Manduchi. "Bilateral filtering for gray and color images". In: *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*. 1998, pp. 839–846. DOI: 10.1109/ICCV.1998.710815.

[45] Danai Triantafyllidou et al. "Low Light Video Enhancement Using Synthetic Data Produced with an Intermediate Domain Mapping". In: *Computer Vision – ECCV 2020*. Ed. by Andrea Vedaldi et al. 2020, pp. 103–119. ISBN: 978-3-030-58601-0. DOI: 10.1007/978-3-030-58601-0_7.

[46] Vasileios Vonikakis. *Busting image enhancement and tone-mapping algorithms: A collection of the most challenging cases*. 2022. URL: https://sites.google.com/site/vonikakis/datasets (visited on 01/27/2022).

[47] Ruixing Wang et al. "Underexposed Photo Enhancement Using Deep Illumination Estimation". In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019, pp. 6842–6850. DOI: 10.1109/CVPR.2019.00701.

[48] Shuhang Wang et al. "Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images". In: *IEEE Transactions on Image Processing* 22.9 (2013), pp. 3538–3548. DOI: 10.1109/TIP.2013.2261309.

[49] Chen Wei et al. "RetinexNet: Deep Retinex Decomposition for Low-Light Enhancement". In: *British Machine Vision Conference 2018, BMVC 2018, Newcastle, UK, September 3-6, 2018*. BMVA Press, 2018, p. 155. URL: http://bmvc2018.org/contents/papers/0451.pdf.

[50] Dejia Xu. *Dark Face Eval Tool*. 2019. URL: https://github.com/Ir1d/DARKFACE_eval_tools.

[51] Ke Xu et al. "Learning to Restore Low-Light Images via Decomposition-and-Enhancement". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020, pp. 2278–2287. DOI: 10.1109/CVPR42600.2020.00235.

[52] Shuo Yang et al. "WIDER FACE: A Face Detection Benchmark". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 5525–5533. DOI: 10.1109/CVPR.2016.596.

[53] Wenhan Yang et al. "Advancing Image Understanding in Poor Visibility Environments: A Collective Benchmark Study". In: *IEEE Transactions on Image Processing* 29 (2020), pp. 5737–5752. DOI: 10.1109/TIP.2020.2981922.

[54] Wenhan Yang et al. "From Fidelity to Perceptual Quality: A Semi-Supervised Approach for Low-Light Image Enhancement". In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020, pp. 3060–3069. DOI: 10.1109/CVPR42600.2020.00313.

[55] Zhenqiang Ying, Ge Li, and Wen Gao. "A bio-inspired multi-exposure fusion framework for low-light image enhancement". In: *arXiv preprint arXiv:1711.00591* (2017). URL: https://arxiv.org/pdf/1711.00591.pdf.

[56] Runsheng Yu et al. "DeepExposure: Learning to Expose Photos with Asynchronously Reinforced Adversarial Learning". In: *Advances in Neural Information Processing Systems*. Ed. by S. Bengio et al. Vol. 31. Curran Associates, Inc., 2018. URL: https://proceedings.neurips.cc/paper/2018/file/a5e0ff62be0b08456fc7f1e88812af3d-Paper.pdf.

[57] Fan Zhang et al. "LLVE: Learning Temporal Consistency for Low Light Video Enhancement From Single Images". In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*. Computer Vision Foundation / IEEE, 2021, pp. 4967–

4976. DOI: 10.1109/CVPR46437.2021.00493.

[58] Kai Zhang, Wangmeng Zuo, and Lei Zhang. "FFDNet: Toward a Fast and Flexible Solution for CNN-Based Image Denoising". In: *IEEE Transactions on Image Processing* 27.9 (2018), pp. 4608–4622. DOI: 10.1109/TIP.2018.2839891.

[59] Qing Zhang, Yongwei Nie, and Wei-Shi Zheng. "Dual Illumination Estimation for Robust Exposure Correction". In: *Computer Graphics Forum* 38.7 (2019), pp. 243–252. DOI: 10.1111/cgf.13833.

[60] Qing Zhang et al. "High-Quality Exposure Correction of Underexposed Photos". In: *Proceedings of the 26th ACM International Conference on Multimedia*. MM '18. 2018, pp. 582–590. ISBN: 9781450356657. DOI: 10.1145/3240508.3240595.

[61] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. "Kindling the Darkness: A Practical Low-Light Image Enhancer". In: *Proceedings of the 27th ACM International Conference on Multimedia*. MM '19. Nice, France, 2019, pp. 1632–1640. DOI: 10.1145/3343031.3350926.

[62] Chaobing Zheng et al. "Single image brightening via multi-scale exposure fusion with hybrid learning". In: *IEEE Transactions on Circuits and Systems for Video Technology* 31.4 (2020), pp. 1425–1435. URL: https://arxiv.org/pdf/2007.02042.pdf.

[63] Minfeng Zhu et al. "EEMEFN: Low-Light Image Enhancement via Edge-Enhanced Multi-Exposure Fusion Network". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 34.07 (2020), pp. 13106–13113. DOI: 10.1609/aaai.v34i07.7013.