Low-light Image and Video Enhancement via Selective Manipulation of Chromaticity

SUMIT SHEKHAR, Hasso Plattner Institute, University of Potsdam, Germany
MAX REIMANN, Hasso Plattner Institute, University of Potsdam, Germany
AMIR SEMMO, Digital Masterpieces GmbH, Germany
SEBASTIAN PASEWALDT, Digital Masterpieces GmbH, Germany
JÜRGEN DÖLLNER, Hasso Plattner Institute, University of Potsdam, Germany
MATTHIAS TRAPP, Hasso Plattner Institute, University of Potsdam, Germany

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 various computer vision and image processing algorithms applied after acquisition. Especially for videos, the additional temporal domain makes it more challenging, wherein we need to preserve quality 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 allows us to avoid the costly step of low-light image decomposition into illumination and reflectance, employed by many existing techniques. All stages in our method consist 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 algorithm and its qualitative and quantitative superiority over several state-of-the-art techniques. For videos captured in the wild, we perform a user study to demonstrate the preference of our method in comparison to state-of-the-art approaches.

CCS Concepts: • Computing methodologies → Image-based rendering; Image processing; Computational photography.

Additional Key Words and Phrases: 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 information and aesthetic quality. Furthermore, it is challenging for such visual media to be used for high-level tasks such as object detection or tracking due to a lack of visual information. Further, poor visual quality negatively affects the visual experience of end-users.

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 on Retinex theory, which assumes the image to be a product of illumination and reflectance. Most of the existing Retinex-based approaches decompose the image into illumination and/or reflectance components, based on specific prior(s). However, finding an effective prior 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 [Liu et al. 2021]. In comparison, deep learning-based solutions 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 [Li et al. 2021]. Therefore, we aim to develop a practical solution for LLIE which adapts to different low-light conditions and also has low computational complexity for interactive performance on commodity hardware.

To achieve the above objective, we adopt a simple strategy based on Retinex theory, the basis for various conventional and learning-based methods. We avoid the computationally costly decomposition step and propose an adaptive way to slowly transition into baseline-reflection (i.e., chromaticity) [Bonneel et al. 2017]. 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. Further, it prevents amplification of sensor noises, to a large degree, common in low-light images. With respect to dark pixels, our approach consistently produces better results for both low and very-low lighting conditions. We generate multiple such ACs with varying level of brightness followed by a multi-scale fusion step. Different levels of brightness prevents over/under-exposedness while multi-scale fusion preserves 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 temporally incoherent and often leads to flickering artifacts. Dark pixels significantly contribute in noise amplification.
leading to temporal incoherence. Due to our ability to robustly handle such pixels the amount of temporal incoherence is reduced significantly. Even the per-frame application of our image-based solution is superior to an existing video-specific approach. Our contributions are summarized as follows, we propose:

1. Adaptive Chromaticity to efficiently increase image brightness while preventing amplification of noise.
2. An approach for low-light image enhancement based on exposure fusion of various ACs of the given image.
3. An per-frame application of our image-based solution for videos, which works out-of-the-box without introducing significant temporal incoherence.

2 BACKGROUND AND 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. [1997a, 1997b] 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 [Cai et al. 2017; Fu et al. 2019, 2015, 2016; Guo et al. 2017; Li et al. 2018; Ren et al. 2020; Wang et al. 2013; Zhang et al. 2019a]. Fu et al. [2016] propose a weighted variational model for simultaneous reflectance and illumination estimation. Guo et al. [2017] perform refinement of an initial illumination map via a structure prior to obtain a well constructed illumination map thereby enabling enhancement. Ren et al. [2020] 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 [Pizer et al. 1987]. Similar to Retinex-based approaches, various extension to the basic principle have been proposed [Abdullah-Al-Wadud et al. 2007; Celik and Tjahjadi 2011; Cheng and Shi 2004; Lee et al. 2013]. Celik and Tjahjadi [2011] employ a variational approach for contrast enhancement using inter-pixel contextual information. Lee et al. [2013] 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 tackle the problem of low-light image enhancement. Methods based on various learning strategies, such as supervised [Cai et al. 2018; Lore et al. 2017; Lv et al. 2018; Ren et al. 2019; Wei et al. 2018; Xu et al. 2020; Zhang et al. 2019b; Zhu et al. 2020], semi-supervised [Yang et al. 2020a], unsupervised [Guo et al. 2020; Jiang et al. 2021; Lee et al. 2020], and reinforcement learning [Yu et al. 2018] have been proposed. Lore et al. [2017] 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. [2018] propose an end-to-end multibranch network for simultaneous enhancement and denoising. Ren et al. [2019] design an encoder-decoder network for global image enhancement and a separate recurrent neural network for further edge enhancement. Similar to Ren et al., Zhu et al. [2020] propose a method called EEMEFN, which consists of two stages: multi-exposure fusion and edge enhancement. Wang et al. [Wang et al. 2019] propose a network called DeepUPE to model image-to-image illumination and collect an expert-retouched dataset. Zhang et al. [Zhang et al. 2019b] propose a network called KinD based on Retinex theory and design a restoration module to counterbalance noise. Chen et al. [Chen et al. 2018] collect a dataset named SID and train a U-Net [2015] 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 [Li et al. 2021]. Two methods which are closely related to our approach are that of Ying et al. [2017] and Zheng et al. [2020], 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 quite 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 [Bonneel et al. 2015; Lai et al. 2018; Shekhar et al. 2019]. These techniques inherently make use of vision-based attributes such as optical flow [Bonneel et al. 2015; Lai et al. 2018] or saliency masks [Shekhar et al. 2019] for temporal stabilization. However, computation of above vision-based attributes itself will be potentially inaccurate/challenging for low-light videos. Lv et al. [2018] 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. [2019] 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 [2019] develop a setup to capture bright and dark dynamic video pairs and subsequently train it using a modified 3D U-Net. However, their sophisticated setup – consisting of two cameras, a relay lens and a beam splitter – is difficult for general usage in the wild. Triantafyllidou et al. [2020] 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. [2021] 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,
Exposure Fusion of Virtual Exposure Sequence using the method of Mertens et al. [2009]

Denoising (NLM)

Input Image

VES

\( A_c(I, \alpha_i, \gamma_i) \)

VES - Fused

Enhanced Output (O)

Adaptive Chromaticity

3 METHOD

According to the Retinex model, an image \( I \) can be expressed as the product of a reflectance layer \( R \) and an illumination layer \( L \) [Land and McCann 1971]: \( I = R \times L \), where the operator \( \times \) denotes pixel-wise multiplication. As a baseline, image "intensity" and "chromaticity" can be considered as the illumination and reflectance layer, respectively [Bonneel et al. 2017]. One can employ different approaches to compute image intensity, such as: norm or the maximum of the individual color channels. However, it does not yield desirable results for our purpose of perceptually plausible editing (see supplementary material). We consider the luminance (Y-channel in YCbCr color space) as our intensity operator \( I_n(\cdot) \) 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 \( I(\cdot) \) and chromaticity \( C(\cdot) \) as follows:

\[
I_n(I) = 0.299 \cdot r + 0.587 \cdot g + 0.114 \cdot b \quad \text{and} \quad C = \frac{I}{I_n(I)} \quad (1)
\]

The brightening effect 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. 3b).

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, \( I_n(\cdot) \), and the maximum intensity value \( MaxI_n \). For low-intensity pixels, this difference defined as \( y = MaxI_n - I_n(\cdot) \) would be comparatively larger. For example, for an intensity image encoded in the range of 0 to 1, \( MaxI_n = 1 \) and for a low-intensity pixel \( p \) with \( I_n(\cdot) = 0.05 \) the difference \( y(p) = 0.95 \) is large. Similarly, for a high-intensity pixel \( q \) with \( I_n(\cdot) = 0.8 \) the difference \( y(q) = 0.2 \) is small (Fig. 3). The above forms the basis for defining adaptive chromaticity (\( A_c(\cdot) \)), wherein we add an adaptive term in the denominator while computing chromaticity (Eqn. (1)). To further increase the brightness, we perform a non-linear scaling using gamma correction

\[
A_c(I, \alpha, \gamma) = \left( \frac{I}{I_n(I) + \alpha(\gamma f(y)+h)} \right)^\gamma \quad (2)
\]

Here, \( f(y) \) is a function in terms of \( y \), \( \alpha \) is a control parameter, \( h \) is a small constant, and \( \gamma \) 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 significantly large value of \( y \). Thus, by tuning the control parameter \( \alpha \) we can smoothly translate 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, we can choose a function \( f(y) \) which satisfies the above property, we use \( f(y) = y^2 \) which is efficient to compute and gives plausible results. The AC brightens up an image while significantly reducing these
Already visible regions in the low-light image get 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 a HDR scene.

We do not have an HDR version of the image at our disposal, however, we can generate an exposure sequence, with varying values of $\alpha$ and $\gamma$. One can generate an HDR image using the above sequence of images and further tone-map it to preserve details in both bright and dark regions while enhancing it [Reinhard et al. 2010]. Thus, we generate a virtual exposure sequence for the given input image by computing ACs with varying brightness by setting the parameters $\alpha$ and $\gamma$. For an image $I$, 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

$$E_k = A_c(I, \alpha_k, \gamma_k).$$

**VES Fusion.** For efficiency, we skip 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. [2009]. 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 $(x)$ basis (see supplementary material). The three quality measures are combined into a joint weighting function

$$w_k(x) = c_k(x) \cdot s_k(x) \cdot e_k(x),$$

where the above product can be seen as logical conjunction and the parameters $v_c$, $v_s$, and $v_e$ 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 $x$, thereby ensuring consistent results, as follows:

$$\bar{w}_k(x) = \frac{w_k(x)}{\sum_{k=1}^{N} w_k(x)}.$$  

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

$$L(O)_l = \sum_{k=1}^{N} G(\bar{w}_k)_l L(E_k)_l.$$  

The final output is obtained by collapsing the computed Laplacian pyramid $L(O)$. Following the above, we employ a denoising operation (similar to DAC) to remove any remaining noise. All the steps in our method are efficiently summarized in an algorithm in the supplementary material.

### 4 RESULTS

#### 4.1 Parameter Settings

Our method has three major steps, for which the parameter settings are discussed in the following.

**VES Generation.** Ideally, to capture fine details at different exposure levels, multiple images are required for the exposure sequence. However, with increase in number of images processing time will increase accordingly. Empirically, we determine three exposure levels ($N = 3$) as sufficient to obtain visually plausible results. For any given scene we keep $\gamma$ as constant, thus $\gamma_1 = \gamma_2 = \gamma_3 = \gamma$. Empirically, we determine $\gamma \in [0.6, 1.0]$ to give well-exposed and less-noisy results. For most of our results, we set $\gamma = 0.6$ (for low-noise images) or $\gamma = 0.9$ (for high-noise images). Empirically, we determine $\alpha \in [0.1, 3.5]$ to yield plausible output. Unlike $\gamma$, we set...
three different values of $\alpha$ for a given scene to obtain three different exposure levels respectively. For most of the results in the paper, we set these as $\alpha_1 = 0.15$ (high-level of brightness), $\alpha_2 = 0.6$ (mid-level of brightness), and $\alpha_3 = 0.85$ (low-level of brightness). Otherwise we mention the used parameters in the caption or in the supplementary.

**VES Fusion and Denoising.** For exposure fusion, we set the weighting exponents for the quality measures to $\nu_c = \nu_b = \nu_c = 1$, as suggested by Mertens et al. [2009]. During fusion, higher number of pyramid-levels helps in preserving fine details. However, with increase in number of levels processing time increases accordingly which is more pronounced for high-resolution images. Empirically, we determine four pyramid levels ($M = 4$) as sufficient to obtain visually plausible results.

For denoising, the NLM approach requires two parameters threshold ($th$) and level ($lv$). For us, $th = 0.7$ and $lv = 1.5$ works best for most of the cases. Otherwise we mention the used parameters in the caption or in the supplementary. On lowering the threshold value significantly, severe denoising leads to loss in details.

### 4.2 Qualitative and Quantitative Evaluation

We compare our results with state-of-the-art image-based methods: two conventional methods (SRIE [2018] and LIME [2017]), two supervised-learning based methods (MBLLEN [2018] and RetinexNet [2018]), an unsupervised-learning based method (Zero-DCE [2020]), and a video-based method (LLVE [2021]). The results are produced from publicly available source codes with given parameter settings.

**Images.** We test the above methods on images taken from the following datasets: LIME [2017], DICM [2013], NPE [2013], and VV [2022]. For quantitative evaluation, we employ the Lightness Order Error (LOE) metric to compare the performance of different methods on the above datasets. Tab. 1 shows that we perform better than compared approaches except for MBLLEN. However, visually we are able to better preserve the details in comparison to MBLLEN. We provide such comparison for enhanced image outputs in Fig. 9.

The results of LIME (Fig. 9(b)) tends to be over-exposed, MBLLEN provides satisfactory brightening (Fig. 9(d)) however tends to over-smooth image details, the output of RetinexNet (Fig. 9(e)) do not look natural, and for LLVE the results (Fig. 9(g)) appear to be hazy and desaturated. Our results look visually comparable to Zero-DCE and SRIE, however we are able to better preserve details (e.g., the sky in Row-1) and brighten image details in a large dynamic range scenario (e.g., human faces in Row-2).

**Videos.** To evaluate video-enhancement results, we make use of the challenging low-light videos provided by Li et al. in their survey LLIV [2021]. We perform a subjective user study with participants to evaluate the performance of different techniques. In total, 22 people (3 female, 18 male, and 1 non-binary) within the ages of 10 to 50 years participated in the study. The experiment consists 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. We asked the participants to focus on the following aspects during comparison:

- **Exposure:** As compared to the input, the output video should be well-exposed, neither under- nor over-exposed.
- **Noise and flickering:** The output video should have less noise and flickering. However, the denoising should not be excessive as to remove details.
- **Color:** The color in the output video should appear natural and it should not look over- or under-saturated.

Fig. 7 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. [2021], we use 500 randomly sampled images from the DARK FACE dataset [2020b] to measure performance of the state-of-the-art Dual Shot Face Detector (DSFD) [2019] trained on the WIDER FACE dataset [2016]. We use the author’s DSFD implementation [2019] with a non-maximum suppression threshold of 0.3 and evaluate using the dark face UG2 challenge evaluation tool [2019]. Fig. 8 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 significant improvement in precision and recall over the unprocessed images. Overall, both our method variants outperform all other methods, with the exception of a precision threshold above 0.85, where RetinexNet [2018] has marginally better precision-recall rates. Our best performing variant uses a simple AC adjustment without denoising or exposure fusion, indicating that more sophisticated methods may smooth or otherwise discard high-frequency information important for face detection.

4.4 Run-time Performance Evaluation

All our experiments were performed on an average 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. 2) and at interactive frame rates on HD and FHD resolution images. Unlike ours, most of the existing techniques are either not able to handle QHD resolution or are very slow for the given hardware configuration. Excluding DAC, our full version performs better than all the other methods except Zero-DCE [2020]. While AC forms the basis of our approach, more than 90% of the processing time is spent on multi-pyramid based exposure fusion. If we simply denoise the AC, the result thus obtained has artifacts in the form of over-exposedness and lack of details however is already comparable to existing approaches (Fig. 5). The DAC, our fast variant, can thus potentially serve as a preview of the enhanced output and 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 degree 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 remove this noise, we use NLM denoising that provides plausible results. However, improved and efficient denoising technique specially tailored for noises in low-lit images will give better results. Thirdly, the enhancement process can result in changes in perceived color. For us, such change is limited due to counter-balancing effect of $\alpha$ and $\gamma$ on the perceived colorfulness (see supplementary material).

**Limitation**: Among the above challenges we are least effective in terms of noise-removal as we employ a moderate denoising scheme for the sake of better run-time performance and handling of high-resolution images. Further, for certain images we might require careful fine tuning of parameters for a better trade-off.

6 CONCLUSIONS AND FUTURE WORK

This paper presents a simple yet efficient technique to enhance low-light images and videos. The key to our approach is Adaptive Chromaticity that allows to increase the image brightness in a straightforward manner. The DAC is already comparable to state-of-the-art methods and can be potentially used for a fast enhancement preview. To further improve results, we generate a virtual exposure sequence by computing multiple adaptive chromaticities for the given low-light image followed by a multi-pyramid based fusion. 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 a subjective user study. We believe that our approach can be used to improve the visual quality of low-light images for further processing. As part of future work we would like to improve the denoising step of our algorithm and potentially use the multi-scale nature of exposure-fusion for this purpose. For videos we would like to...
use the neighboring frames to improve the denoising as well as enhancement quality.

REFERENCES

M. Abdollahi-Al-Wadud, Md. Hasanul Kabir, M. Ali Akber Dewan, and Oksams Chae. 2007. A Dynamic Histogram Equalization for Image Contrast Enhancement. IEEE Transactions on Consumer Electronics 53, 2 (2007), 593–600. https://doi.org/10.1109/TCE.2007.381734

Nicolas Bonneel, Balazs Kovacs, Sylvain Paris, and Kavita Bala. 2016. Intrinsic Decompositions for Image Editing. Computer Graphics Forum 36, 2 (May 2017), 593–609. https://doi.org/10.1111/cgf.13149

Nicolas Bonneel, James Tompkin, Kalyan Sunkavalli, Deqing Sun, Sylvain Paris, and Hanspeter Pfister. 2013. Blind Video Temporal Consistency. ACM Trans. Graph. 34, 6, Article 196 (oct 2015), 9 pages. https://doi.org/10.1145/2816795.2818107

A. Buades, B. Coll, and J.-M. Morel. 2005. A non-local algorithm for image denoising. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05). Vol. 2. 60–65 vol. 2. https://doi.org/10.1109/CVPR.2005.38

Bolun Cai, Xianming Xu, Kading Guo, Kui Jia, Bin Hu, and Dacheng Tao. 2017. A Joint Intrinsic-Extrinsic Prior Model for Retinex. In 2017 IEEE International Conference on Computer Vision (ICCV). 4020–4029. https://doi.org/10.1109/ICCV.2017.431

Jianrui Cai, Shuhang Gu, and Lei Zhang. 2018. Learning a Deep Single Image Contrast Enhancer from Multi-Exposure Images. IEEE Transactions on Image Processing 27, 4 (2018), 2049–2062. https://doi.org/10.1109/TIP.2018.2794218

Turgay Celik and Tardi Tjahjadi. 2011. Contextual and Variational Contrast Enhancement. IEEE Transactions on Image Processing 20, 12 (2011), 3431–3441. https://doi.org/10.1109/TIP.2011.2157513

Chen Chen, QiFeng Chen, Jia Xu, and Vladlen Koltun. 2018. Learning to See in the Dark. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 3291–3300. https://doi.org/10.1109/CVPR.2018.00347

H.D. Cheng and X.J. Shi. 1971. Lightness and retinex theory. Journal of the Optical Society of America 61, 1 (1971), 1–11. https://doi.org/10.1364/JOSA.61.000001

Chunlue Guo, Ling Huo, Chuanli Li, and Runmin Cong. 2020. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 1777–1786. https://doi.org/10.1109/CVPR42600.2020.00185

Hsuan Yang. 2018. Learning Blind Video Temporal Consistency. In Computer Vision – ECCV 2018; Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (Eds.). 179–195. https://doi.org/10.1007/978-3-030-01234-2_10

Edwin H Land and John J McCann. 1971. Lightness and retinex theory. Journal of the Optical Society of America 61, 1 (1971), 1–11. https://doi.org/10.1364/JOSA.61.000001

Chunlue Guo, Ling-Hao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. 2021. EnlightenGAN: Deep Light Enhancement Without Paired Supervision. IEEE Trans. Image Proc. 30 (2021), 2340–2349. https://doi.org/10.1109/TIP.2020.3051462

D.J. Jobson, Z. Rahman, and G.A. Woodell. 1997a. A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE Transactions on Image Processing 6, 7 (1997), 965–976. https://doi.org/10.1109/83.637722

D.J. Jobson, Z. Rahman, and G.A. Woodell. 1997b. Properties and performance of a center/surround retinex. IEEE Transactions on Image Processing 6, 3 (1997), 451–462. https://doi.org/10.1109/83.57536

Wei-Sheng Lai, Jia-Bin Huang, Oliver Wang, Eli Shechtman, Ersin Yumer, and Ming-Hsuan Yang. 2018. Learning Blind Video Temporal Consistency: In Computer Vision – ECCV 2018; Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (Eds.). 179–195. https://doi.org/10.1007/978-3-030-01234-2_10

Jilin Li, and Feiyue Huang. 2019. DSFD: dual shot face detector. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 5060–5069. https://doi.org/10.1109/CVPR.2019.00742

Pan Zhou, and Zhangyang Wang. 2021. EnlightenGAN: Deep Light Enhancement Without Paired Supervision. IEEE Trans. Image Proc. 30 (2021), 2340–2349. https://doi.org/10.1109/TIP.2020.3051462

Chunlue Guo, Gao, and Runmin Cong. 2020. Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 1777–1786. https://doi.org/10.1109/CVPR42600.2020.00185

Xiaojie Guo, Yu Li and Haibin Ling. 2017. LIME: Low-Light Image Enhancement via Illumination Map Estimation. IEEE Transactions on Image Processing 26, 2 (2017), 982–993. https://doi.org/10.1109/TIP.2016.2639450

Fig. 9. Low-light image enhancement results. Input images are taken from LIME [2017], DICM [2013], VV [2022], and LOL [2018] datasets.
Kim Gwon Lore, Adelotun Akintayo, and Soumik Sarkar. 2017. LLNet: A deep auto-encoder approach to natural low-light image enhancement. *Pattern Recognition* 61 (2017), 656–662. https://doi.org/10.1016/j.patcog.2016.06.008

Fei Fei Lu, Feng Lu, Jianhua Wu, and Chongsoon Lim. 2018. MBLLEN: Low-Light Image/Video Enhancement Using CNNs. In *British Machine Vision Conference 2018, BMVC 2018, Newcastle, UK, September 3-6, 2018. BMVA Press*, 220. http://bmvc2018.org/ contents/papers/090.pdf

T. Mertens, J. Kautz, and F. Van Reeth. 2009. Exposure Fusion: A Simple and Practical Alternative to High Dynamic Range Photography. *Computer Graphics Forum* 28, 1 (2009), 161–171. https://doi.org/10.1111/j.1467-8659.2008.01171.x

Stephen M. Fizer, E. Philip Amburn, John D. Austin, Robert Cromptt, Ari Geselowitz, Troy Greer, Bart ter Haar Romeny, John B. Zimmerman, and Karel Zuiderveld. 1987. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing* 39, 3 (1987), 355–368. https://doi.org/10.1016/0734-189X(87)90018-X

Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. 2010. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann.

Wenqi Ren, Sifei Liu, Lin Ma, Qianqian Ju, Xiangyu Xu, Xiaochun Cao, Jumping Du, and Ming-Hsuan Yang. 2019. Low-Light Image Enhancement via a Deep Hybrid Network. *IEEE Transactions on Image Processing* 28, 9 (2019), 4364–4375. https://doi.org/10.1109/TIP.2019.2910412

Xutong Ren, Wenhan Yang, Wen-Huang Cheng, and Jaying Liu. 2020. LR3M: Robust Low-Light Enhancement via Low-Rank Regularized Retinex Model. *IEEE Transactions on Image Processing* 29 (2020), 5862–5876. https://doi.org/10.1109/TIP.2020.2984908

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi (Eds.) 234–241.

Sumit Shekhar, Amir Sammo, Matthias Trapp, Oakan Tursun, Sebastian Pasewaldt, Karol Myszkowski, and Jürgen Döllner. 2019. Consistent Filtering of Videos and Dense Light-Fields Without Optic-Flow. In *Vision, Modeling and Visualization*, Hans-Jörg Schulz, Matthias Teschner, and Michael Wimmer (Eds.). https://doi.org/10.2312/vmv.20191326

Danai Triantafyllidou, Sean Moran, Steven McDonagh, Sarah Parisot, and Gregory Slabaugh. 2020. Low Light Video Enhancement Using Synthetic Data Produced with an Intermediate Domain Mapping. In *Computer Vision – ECCV 2020*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). 103–119.

Vasileios Vonikakis. 2022. Busting image enhancement and tone-mapping algorithms: A collection of the most challenging cases. https://sites.google.com/site/vonikakis/datasets

Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jaya Jia. 2019. Underexposed Photo Enhancement Using Deep Illumination Estimation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6842–6850. https://doi.org/10.1109/CVPR.2019.00701

Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. 2013. Naturalness Preserved Enhancement Algorithm for Non-Uniform Illumination Images. *IEEE Transactions on Image Processing* 22, 9 (2013), 3538–3548. https://doi.org/10.1109/TIP.2013.2261309

Chen Wei, Wenjing Wang, Wenhan Yang, and Jaying Liu. 2018. Deep Retinex Decomposition for Low-Light Enhancement. In *British Machine Vision Conference 2018, BMVC 2018, Newcastle, UK, September 3-6, 2018. BMVA Press*, 155. http://bmvc2018.org/contents/papers/0451.pdf

Deja XU. 2019. Dark Face Eval Tool. https://github.com/bridd/DARKFACE_eval_tools

Ke Xu, Xin Yang, Baocai Yin, and Runyun W.H. Lau. 2020. Learning to Restore Low-Light Images via Decomposition-and-Enhancement. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2278–2287. https://doi.org/10.1109/CVPR42600.2020.00235

Shuo Yang, Ping Luo, Chen Change Loy, and Xiaosou Tang. 2016. WIDER FACE: A Face Detection Benchmark. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*

Wenhan Yang, Shiq Wang, Yunfeng Yue, Yue Wang, and Jaying Liu. 2020b. 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)*. 3060–3069. https://doi.org/10.1109/CVPR42600.2020.00313

Wenhan Yang, Ye Yuan, Wenqi Ren, Jaying Liu, Walter J. Scheirer, Zhangyang Wang, and Zhang. 2020b. Advancing Image Understanding in Poor Visibility Environments: A Collective Benchmark Study. *IEEE Transactions on Image Processing* 29 (2020), 5757–5752. https://doi.org/10.1109/TIP.2020.2981922

Zhiqiang Ying, Ge Li, and Wen Gao. 2017. A bio-inspired multi-exposure fusion framework for low-light image enhancement. arXiv preprint arXiv:1711.00591 (2017).

Runsheng Yu, Wenyu Liu, Yasen Zhang, Zhi Qu, Deli Zhao, and Bo Zhang. 2018. DeepExposure: Learning to Expose Photos with Asynchronously Reinforced Adversarial Learning. In *Advances in Neural Information Processing Systems*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Eds.), Vol. 31. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2018/file/a5e0ff62be0b08456fc7f1e88812af3d-Paper.pdf

Fan Zhang, Yu Li, Shaodou You, and Ying Fu. 2021. 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. 4967–4976

Kai Zhang, Wangmeng Zuo, and Lei Zhang. 2018. FTI-Net: Toward a Fast and Flexible Solution for CNN-Based Image Denoising. *IEEE Transactions on Image Processing*. 27, 9 (2018), 4608–4622. https://doi.org/10.1109/TIP.2018.2839891

Qing Zhang, Yongwei Nie, and Wei-Shi Zheng. 2019a. Dual Illumination Estimation for Robust Exposure Correction. *Computer Graphics Forum* 38, 7 (2019), 243–252. https://doi.org/10.1111/cgf.13833

Yonghua Zhang, Jiaoyin Zhang, and Xiaojie Guo. 2019b. Kinding the Darkness: A Practical Low-Light Image Enhancer. In *Proceedings of the 27th ACM International Conference on Multimedia (Nice, France) (MM’19)*. 1632–1640. https://doi.org/10.1145/3343801.3350926

Chaoxing Zheng, Zhengguo Li, Yi Yang, and Shiqian Wu. 2020. Single image brightening via multi-scale exposure fusion with hybrid learning. *IEEE Transactions on Circuits and Systems for Video Technology* 31, 4 (2020), 1425–1435.

Minfeng Zhu, Pingho Pan, Wei Chen, and Yi Yang. 2020. EEMFN: Low-Light Image Enhancement via Edge-Enhanced Multi-Exposure Fusion Network. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 07 (April 2020), 13106–13113. https://doi.org/10.1609/aaai.v34i07.10713