Signal Processing: Image Communication 74 (2019) 175–190

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

Low-light image enhancement using Gaussian Process for features retrieval

Yuen Peng Loh [a,c], Xuefeng Liang [b,*], Chee Seng Chan [a,*]

[a] Centre of Image and Signal Processing, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, 50603, Malaysia
[b] School of Artificial Intelligence, Xidian University, Xi'an, Shaanxi 710126, China
[c] Faculty of Computing and Informatics, Multimedia University, Cyberjaya, 63100, Selangor, Malaysia

ARTICLE INFO

Keywords:
Low-light
Image enhancement
Gaussian Process
Convolutional neural network

ABSTRACT

Low-light is a challenging environment for image processing and computer vision tasks, either in contrast enhancement for better visibility and quality, or application oriented tasks such as detection. We found that the current trend of low-light enhancement research is heavily on quality improvement. In this work, we aim to shift the focus towards a more functional direction, that is enhancement that prioritizes feature retrieval. For this reason, we first propose to model low-light enhancement as a set of localized functions using Gaussian Process ($\mathcal{GP}$) that is trained at runtime using data from a simple Convolutional Neural Network (CNN) to provide the necessary feature information as reference. The CNN is in turn trained using large amount of synthetic data, based upon the luminance distribution of real world low-light images to learn the relationship between features and pixels. Secondly, we also proposed two new evaluation metrics to better assess enhancement algorithms to support high level computer vision tasks, namely, local features matching and intensity histogram similarity. In our experiments, our proposed low-light enhancement framework outperforms the state-of-the-arts with significant improvement in Recall, F_1, and F_2-score of SIFT features matching, and achieve comparable results for l_1-norm distance of histograms as well as the PSNR. Moreover, our analysis of the performance showed that the PSNR quality metric is not only unable to assess the practicality of the results, but also inappropriately gives high assessment to low visual quality images.

1. Introduction

Low-light is a challenging environment for human vision as the lack of visibility affects a person's ability to perform tasks. Therefore, computer vision algorithms that can provide assistance in such conditions are highly valuable. However, current research works related to the low-light domain are mostly on image aesthetic enhancement instead of applications, like object detection, that can be developed into practical intelligent vision systems, such as visual surveillance and autonomous car driving [1].

The motivation of our work is to shift the focus of low-light image[1] enhancement research from the aesthetics driven studies towards a functional practice, i.e. enhancement in support of computer vision applications. For this reason, a change is necessary not only in the formulation of enhancement frameworks but also the evaluation schemes. Hence, in this paper, we set out to achieve two objectives, (1) to propose a low-light image contrast enhancement framework that primarily retrieves features that were degraded by low illumination and contrast, while visual quality is secondary, and (2) to propose new evaluation metrics that would assess the ability of enhancement algorithms to retrieve features.

Low-light image enhancement is a non-trivial task because of the non-uniformity of scene luminance, as shown in Fig. 1. It can be seen that Fig. 1(a) is a uniformly dark low-light image, however, there are others with weak light sources in the dark environments, such as Fig. 1(b), or reflected from surfaces, as shown in Fig. 1(c). Researchers mostly seek to counter the over and under enhancement problems caused by such phenomenon through increasingly sophisticated methods of manipulating the illumination of the image. However, to the best of our knowledge, none has explicitly considered the prospects of restoring features that could be used as a result of the enhancement. As we will demonstrate, state-of-the-art low-light image enhancement algorithms produce results with good visual quality but do not necessarily retrieve the most features.

In this paper, we make two major proposals. First, is a framework that addresses the challenge of low-light enhancement from the perspective of features retrieval. Based on our analysis of low-light image characteristics, we noted that the enhancement functions has to be localized within an image for optimal results, i.e. individual functions to brighten, maintain, or darken, specific regions or pixels. To this end, we employed *Gaussian Process regression* to construct a distribution of such functions with the support of a *Convolutional Neural Network* to introduce feature enhancement functions into the distribution. Secondly, we propose two

* Corresponding authors.
E-mail addresses: lohyuenpeng@siswa.um.edu.my, yploh@mmu.edu.my (Y.P. Loh), xliang@xidian.edu.cn (X. Liang), cs.chan@um.edu.my (C.S. Chan).

[1] Low-light images are images captured in environment with low illumination such as nighttime, twilight, etc.

https://doi.org/10.1016/j.image.2019.02.001
Received 6 November 2018; Received in revised form 25 January 2019; Accepted 1 February 2019
Available online 14 February 2019

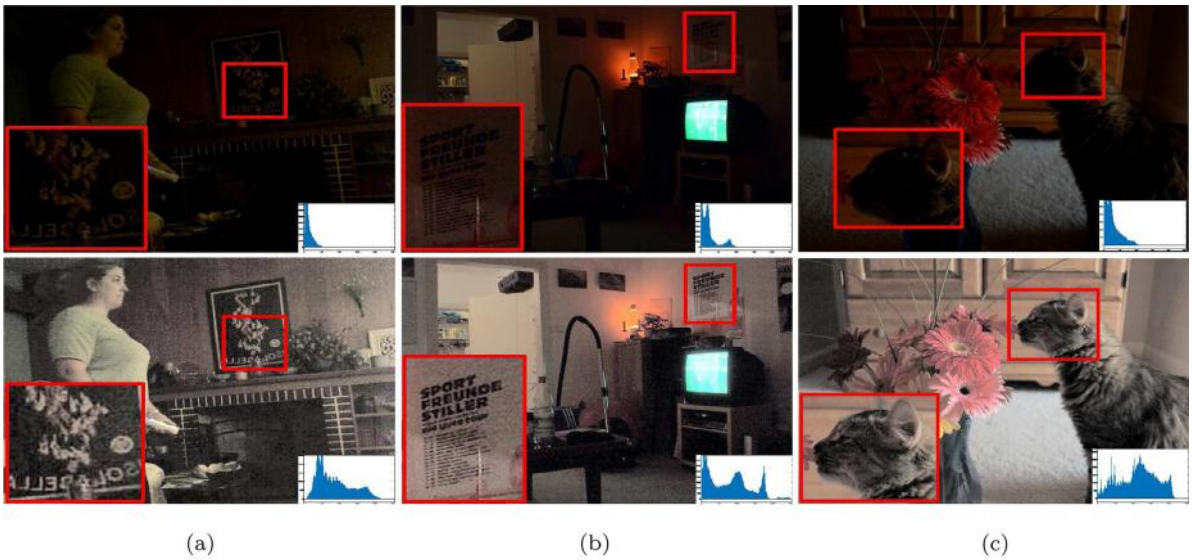

Fig. 1. (Top) Low-light images, and (Bottom) our contrast enhancement results, with their respective intensity histograms. The low-light images are not only brightened, but the details of the image content are also visibly improved, such as the (a) poster design, (b) text, and (c) cat. More details are provided in Sections 5.2 and 5.3. These sample images are from the MS COCO dataset. [Best viewed in color].

new evaluation metrics that focus on features retrieval to demonstrate the usability of image enhancement algorithms for high level computer vision tasks. Namely, the *precision, recall, and F-score* of local features matching, and l_1*-norm distance* measure of intensity histogram.

Experimental results showed that our proposed framework outperformed the state-of-the-art [2–7] in our newly proposed feature matching evaluation while comparable in the l_1-norm distance of histogram as well as the standard image quality measure, the peak signal-to-noise ratio (PSNR). Moreover, qualitative evaluation of enhancement on real low-light images from the ExDark dataset [1] showed that our proposal is able to overcome problems such as color distortion, saturation, over-enhancement, and noise.

2. Related works

The research works in computer vision tackle low-light problems from two perspectives, (1) through the use of hardware supports such as cameras equipped with infrared sensors or thermal imagers [8–11]; or (2) using enhancement algorithms on low-light images or videos [2–5,12–19]. The former are commonly used to address surveillance and application based problems but the required cameras are costly and do not show realistic images, e.g. only silhouettes of pedestrians and vehicles can be observed. Whereas the latter, namely low-light image enhancement is generally applied for image quality improvement. Our interest lies in developing low-light image enhancement that not only improves the quality but supports intelligent computer vision applications.

We further discuss a few researches that are closely related to this area as follows:

Low-light imaging. Low-light imaging enhancement is a computational photography pipeline that processes raw data from camera sensor to construct the enhanced RGB output. [20] is a notable work that proposed a deep neural network approach to learn the whole pipeline from end-to-end to avoid noise amplification and error accumulation found in the conventional pipeline. However, the function of this work is limited by the camera sensor where the model has to be trained specifically for a given sensor. In our work, we focus on the image enhancement component which is not restricted by the camera sensors or hardware.

HDR imaging. This field of work addresses the problem of low quality images due to improper exposure that could be caused by low-light conditions, inappropriate camera settings and/or its limited dynamic range. One of the approaches is the stack-based HDR [21–23], where multiple images of varying exposures/low dynamic range are captured and fused to improve the image quality. However, a drawback of such methods is the multiple image requirement where the movement of dynamic objects will result in subtle differences within each image and subsequently cause HDR algorithms to produce results with "ghosting" artifacts. The other approach is the single image contrast enhancement where the proposed methods have similarity with the low-light image enhancement works as well. Though, we note that the difference between these two domains is that, HDR imaging addresses over or under exposed images regardless if the image is captured in low-light or bright conditions, to produce the best quality image, whereas low-light image enhancement specifically focus on low-light environments such as nighttime. The latter is the main target area of our work.

Low-light enhancement. Low-light image enhancement specifically addresses images captured in low-light conditions such as nighttime, where the common goal is to brighten and improve the contrast of the image for better visual quality and show details that were hidden in darkness. The various proposed works can be categorized into three categories. The first is statistically model and manipulate [12–14] the distributions of low-light images, either intensities or high frequency coefficients, to improve the image contrast and brightness, such as histogram equalization and its variants [12,28,29]. The second category is the transformation model approach [6,15,16,19,24] that uses parameterized functions or trained models to perform transformation mapping from low-light image space to bright image space, meanwhile preserve the contextual information. The third category uses the Retinex model [30] that takes into consideration both the contextual information and light intensity, whereby the main assumption is that a color image can be decomposed to reflectance and illumination components to represent the aforementioned elements respectively. By manipulating the illumination component and merging with the reflectance, these methods [3–5,25] had shown impressive results and recently, deep learning approaches have been proposed based on the Retinex theory as well [7,26]. Additionally, [2,17,18] have proposed methods that resemble the Retinex theory by implementing the dark channel prior algorithm made for image dehazing [27]. This approach is mainly sparked by the observation where inverted low-light images exhibit similar characteristics to images captured in hazy weather.

Table 1
Existing research works on low-light image enhancement. Category (A) Statistical methods, (B) Transformational methods.

Cat.	Literature	Color space	Method
A	Huang et al. (2013)[12]	Value (V) component of HSV	Adaptive gamma correction with weighted distribution modification on the intensity histogram.
	Lim et al. (2015) [13]	Grayscale	Histogram equalization on Gaussian approximated distribution of noise-free image pixels.
	Łoza et al. (2013) [14]	Value (V) component of HSV	Dual-Tree Complex Wavelet Transform (DT-CWT) on V, exponential contrast enhancement on high-pass wavelet coefficients and CLAHE on low-pass coefficients.
B	Wu (2011) [15]	RGB	Optimal contrast tone mapping by linear programming.
	Fu et al. (2012) [16]	RGB	Color estimation model modulated by a single parameter determined by statistical observation on large data and then post processed by sparse coding.
	Loh and Chan (2015) [19]	Grayscale	Invert local features such as SIFT and HOG using paired dictionary learning.
	Lore et al. (2017) [24]	Grayscale	Stacked sparse denoising autoencoder model trained on synthesized low-light images.
	Wang et al. (2016) [6]	RGB	Global illumination estimation and details reconstruction using deep learning model

Table 2
Existing research works on low-light image enhancement. Category (C) Retinex methods, (D) Other methods.

Cat.	Literature	Color space	Method
C	Fu et al. (2016) [3]	RGB	Weighted multi-scale fusion of luminance and contrast improved illumination component of the Retinex model.
	Fu et al. (2016) [4]	RGB and HSV	Retinex model decomposition by weighted variational model and illumination enhancement by gamma correction.
	Guo et al. (2017) [5]	RGB	Illumination map estimation and refinement by structure-aware smoothing model.
	Li et al. (2018) [25]	Value (V) component of HSV	Addition of noise term into Retinex model for simultaneous decomposition and noise reduction.
	Shen et al. (2017) [26]	RGB	Multi-scale Retinex using convolutional neural network.
	Chen et al. (2018) [7]	RGB	Convolutional neural network for joint Retinex decomposition and low-light enhancement.
D	Dong et al. (2011) [17]	RGB	Invert low-light image and apply dark channel prior dehazing algorithm [27].
	Zhang et al. (2012) [18]	RGB	Enhancement using dehazing algorithm where light transmission is estimated from the image luminance instead of the scene depth as used in the standard dehazing.
	Li et al. (2015) [2]	RGB	Dehazing with adaptive weight coefficient for light transmission estimation.

Tables 1 and 2 contains the summary of aforementioned literature related to low-light image enhancement.

Instead, our proposal using $\mathcal{GP}$ is a statistical model that performs pixel mapping, which we integrate with features related enhancement functions for supporting computer vision applications. Thus, the main difference between our work and these existing studies is that our objective aims at feature retrieval, contrary to visual quality improvement only.

3. Problem formulation

The typical aim of low-light image contrast enhancement is to improve brightness and contrast of a low-light image so that the content is visible and aesthetically pleasing to observers. This is supported by the choice of evaluation metrics used in current works [20,24–26], such as PSNR, and structural similarity index (SSIM) that assess image quality. Our objective is to perform low-light image enhancement with the goal of feature retrieval in support of automated vision systems where illumination and lighting variation is one of the key problems [31]. For this reason, we reexamine the challenges that are present in low-light images.

Studies on object detection commonly work on bright images where the contrast is relatively uniform throughout the image with clear contrast and details [1], as shown in the bottom row of Fig. 2. Bright images have high scene luminance provided by strong sources of light, such as the sun, that is able to encompass large areas. On the contrary, low-light images have low illumination where the objects' appearance lack details and look invisible. This is because the lighting in low-light environment is provided by limited sources that are comparatively weaker, such as twilight, street lights, or car lights to name a few. As a result, objects are only apparent when near to the light source but become increasingly obscure as they move further away and brings about considerable illumination variations within one image.

Fig. 3 shows some examples of our analysis on bright (Fig. 3(a)) and low-light (Fig. 3(b) & 3(d)) images. In the bright images, the average luminance, Y from the YC_bC_r color space, of the patches are relatively high, above 90, and consistent, where the maximum difference is only 80 between the blue and red boxes of Fig. 3(a). However, in the low-light images, not only is there lower illumination, the average luminance reduces when the patches are further away from the light source and have very large differences (the maximum difference is 159 between patches of blue and yellow boxes in Fig. 3(b)). Furthermore, when there are more than one light source available, the luminance variation becomes even more severe, as shown in Fig. 3(d).

Therefore, in order to formulate a solution for this problem, we look back at the definition of light behavior in the real world, stated by the

Fig. 2. A single low-light image comprises varying luminosity (i.e. shadows, many light sources), while the lighting in a bright image is relatively consistent. Top row: Samples of the real low-light images used in the analysis; Bottom row: Samples of the bright images used in the synthesis. [Best viewed in color].

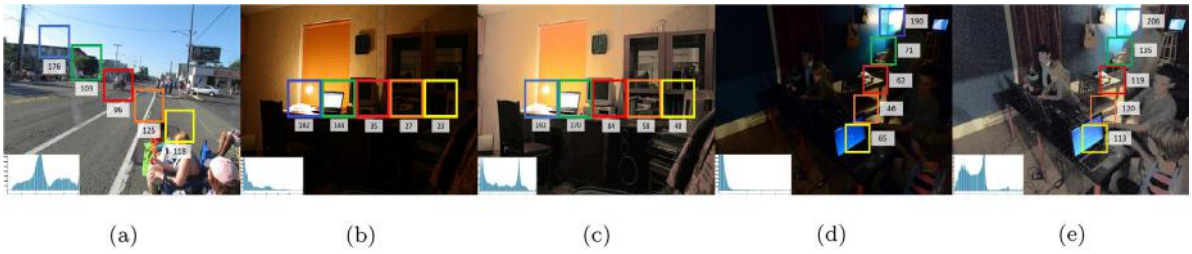

Fig. 3. Examples of bright 3(a) image, and low-light 3(b) & 3(d) images with their respective enhanced images 3(c) & 3(e), along with their local illumination values and global intensity histograms. The average luminance (Y of YC_bC_r color space) of the patches (colored boxes) in the bright image are relatively high and consistent, while the patches in the low-light images have decreasing values as the distance from the light source increases. [Best viewed in color].

inverse-square law of distance:

$$E(p) = \frac{L(p)\cos^2\theta dA}{l^2} \tag{1}$$

where E is the irradiance or intensity per unit area on the object, L is the radiance from light source, cos θ is the foreshortening, dA is the unit area, and l is the distance between the light source and object. Based on this, the light intensity of an object depends on radiance of the light source (e.g. sun, street lights, etc.), the foreshortening,[2] and distance. Intuitively, the enhancement operation can be understood as increasing the radiance as the distance increase, however, we note that such straightforward operation does not fully benefit object features. For example, the edge of an object as a result of foreshortening is a notable feature for object detection. Simply reversing the foreshortening effect or brightening the pixels would diminish its discriminative ability, thus, in such cases maintaining or darkening the pixels may be more beneficial for the features. Hence, the low-light image enhancement can be modeled as:

$$I_B(x) = I_D(x)\mathcal{F}(L, l_x, \phi_x), \tag{2}$$

$$\text{s.t.} \quad L = \{L_1, L_2, \ldots, L_i\};$$
$$l_x = \{l_{x_1}, l_{x_2}, \ldots, l_{x_j}\}; \tag{3}$$
$$\phi_x = \{\phi_{x_1}, \phi_{x_2}, \ldots, \phi_{x_j}\};$$

where $I_B(x)$ is the enhanced image that has an appropriate contrast and relative uniform intensity distribution, $I_D(x)$ is the captured low-light image that has relatively low illuminance, low contrast and intensity variation, x is a pixel or small patch in the image, and $\mathcal{F}(L, l_x, \phi_x)$ is the mapping operator defined by the light source L, distance l and features ϕ. However, estimating the mapping operator is not a simple task due to the following reasons:

- Diverse types of light sources L have individual light strength providing distinct levels of intensity;
- Multiple sources of light $\{L_1, L_2, \ldots, L_i\}$ increases the non-uniformity of the scene luminance;
- Any point of a scene can locate at a varying distance $\{l_{x_1}, l_{x_2}, \ldots, l_{x_j}\}$ between the object and source of light.
- Image pixels may form features of a single or different objects $\{\phi_{x_1}, \phi_{x_2}, \ldots, \phi_{x_j}\}$

Hence, the best course of action is to perform localized enhancement on each pixel where not only will that address the varying L and l, but also the appropriate enhancement can be performed on pixels that comprise features. Therefore, we transform Eq. (2) into $I_B(x) = f(I_D(x))$ where $f(\cdot)$ represents the set of enhancement functions that models the relationship of pixels in the low-light image $I_D(x)$ with the target bright pixels of $I_B(x)$.

4. Proposed method

We have established that the relationship of pixels between $I_D(x)$ and $I_B(x)$ are localized, hence $f(\cdot)$ is not a single function acting on all pixels, but a collection of different functions that acts on the respective pixels. Therefore, we propose the use of the Gaussian Process ($\mathcal{GP}$) to constrain the enhancement functions into a *distribution of functions*.

4.1. Gaussian Process revisited

A Gaussian Process defines a distribution over function f that estimates an output y from the marginal distribution of functions $\mathbb{P}(f(x_1), f(x_2), \ldots, f(x_k))$ of finite inputs $x = \{x_1, x_2, \ldots, x_k\}$. It is parameterized by a mean function $m(x)$ and a covariance function $k(x_{tr}, x_{ts})$ such that $f(x) \sim \mathcal{GP}(m(x), k(x_{tr}, x_{ts}))$, where the joint distribution of training and test outputs is:

$$\begin{bmatrix} y_{tr} \\ y_{ts} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(x_{tr}) \\ m(x_{ts}) \end{bmatrix}, \begin{bmatrix} K(x_{tr}, x_{tr}) & K(x_{tr}, x_{ts}) \\ K(x_{ts}, x_{tr}) & K(x_{ts}, x_{ts}) \end{bmatrix}\right). \tag{4}$$

[2] the reduction of surface area as seen from a particular point of observation.

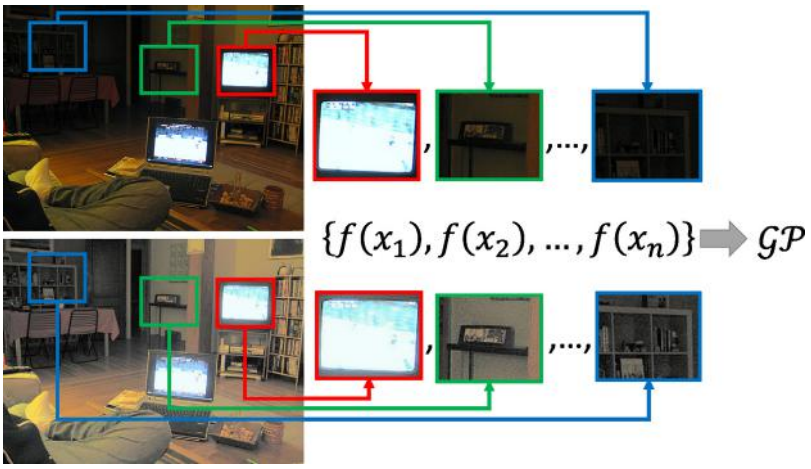

Fig. 4. Our intuition is to have localized enhancement functions for each region/pixel, hence we utilized $\mathcal{GP}$ to govern them into a distribution of functions. (Top: Low-light image, Bottom: Contrast enhanced image). [Best viewed in color].

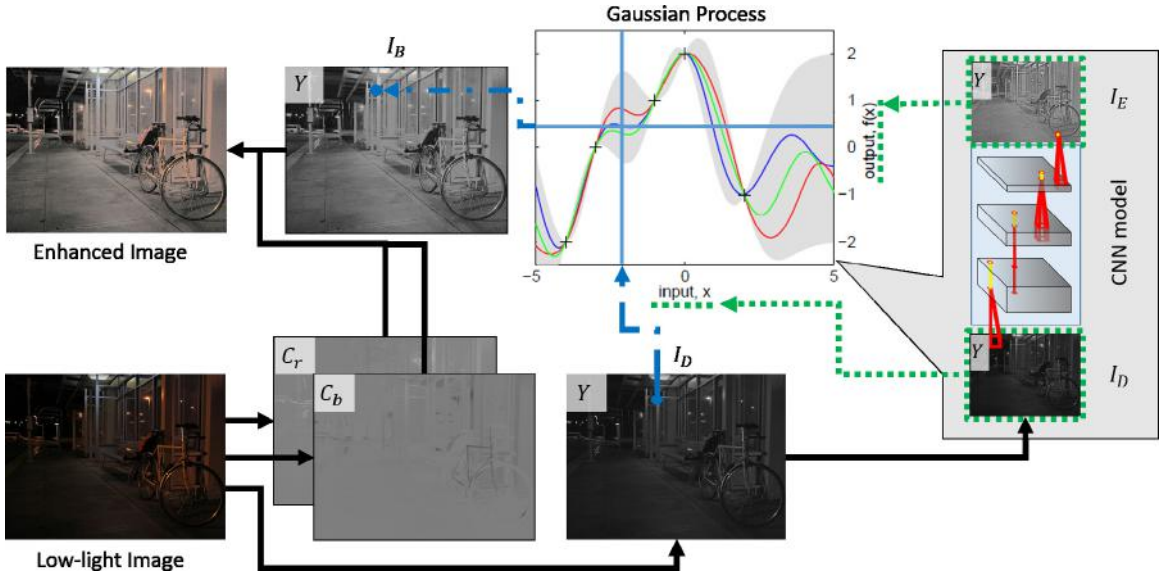

Fig. 5. Overall framework of the low-light image enhancement using $\mathcal{GP}$ regression. As the construction of the $\mathcal{GP}$ requires both input and output training data, a CNN is engaged as the intermediate model for output data generation. [Best viewed in color].

y_{tr} refers to the training outputs and y_{ts} is the testing outputs in which $y = [y_{tr}, y_{ts}]$. Similarly, $x = [x_{tr}, x_{ts}]$ are the training and testing inputs. $K(x_{tr}, x_{ts})$ denotes the covariance matrix between the training and testing inputs, along with $K(x_{tr}, x_{tr})$, $K(x_{ts}, x_{tr})$, and $K(x_{ts}, x_{ts})$ as the covariance of their respective pairings. Given the observation y_{tr} from x_{tr}, the output y_{ts} can be estimated with x_{ts} from the conditional distribution $\mathbb{P}(y_{ts}|x_{tr}, y_{tr}, x_{ts}) \sim \mathcal{N}(\mu, \Sigma)$ where $\mu = m(x_{ts}) + K(x_{ts}, x_{tr})[K(x_{tr}, x_{tr})]^{-1}(y_{tr} - m(x_{tr}))$ and $\Sigma = K(x_{ts}, x_{ts}) - K(x_{ts}, x_{tr})[K(x_{tr}, x_{tr})]^{-1}K(x_{tr}, x_{ts})$.

4.2. Modeling contrast enhancement with $\mathcal{GP}$

Our objective is to estimate a corresponding enhanced image I_B given a single low-light image I_D. As we have determined that the enhancement functions for low-light images are localized, the distribution of functions $\mathbb{P}(f(x_1), f(x_2), \ldots, f(x_k))$ are therefore the varied local luminance mapping functions as shown in Fig. 4, where the inputs x refer to local patches or pixels in the low-light image.

Specifically, in our implementation the testing inputs are the pixels from the low-light image, i.e. $x_{ts} = \{p_D | p_D \in I_D\}$, while the testing outputs are the corresponding reference pixels, i.e. $y_{ts} = \{p_B | p_B \in I_B\}$.

Fig. 5 shows the overall flow of the proposed enhancement framework where the reference pixels for training the $\mathcal{GP}$ is provided by a trained CNN model that extract features and project them into pixel space.

The construction of the $\mathcal{GP}$ is specified by mean and covariance functions given the data. Hence, for these priors, we use the zero mean function $m(x) = 0$ to simplify the modeling process and allow the relationship between x_{tr} and x_{ts} to be fully defined by the covariance function $k(x_{tr}, x_{ts})$. We chose squared exponential as the covariance function:

$$k(x_{tr}, x_{ts}) = \sigma_f^2 \exp\left(-\frac{(x_{tr} - x_{ts})^2}{2d^2}\right), \tag{5}$$

where hyperparameters, σ_f^2 is the data variance, and d is the length scale that defines the smoothness of the $\mathcal{GP}$. These hyperparameters $\theta_{\mathcal{GP}} = \{\sigma_f, d\}$ determines the form of the distribution of function, and are inferred from the low-light data using conjugate gradients to optimize the log marginal likelihood: $\mathbb{L} = \log\mathbb{P}(y_{tr}|x_{tr}, \theta_{\mathcal{GP}})$. As the posterior distribution is data dependent, each image is therefore enhanced by image exclusive optimal hyperparameters in our framework. For the training data, we define the patches of $m \times n$ pixels of the given low-light image I_D as the training inputs x_{tr} whereas the training outputs y_{tr} are patches of the similar pixels dimension and spatial location

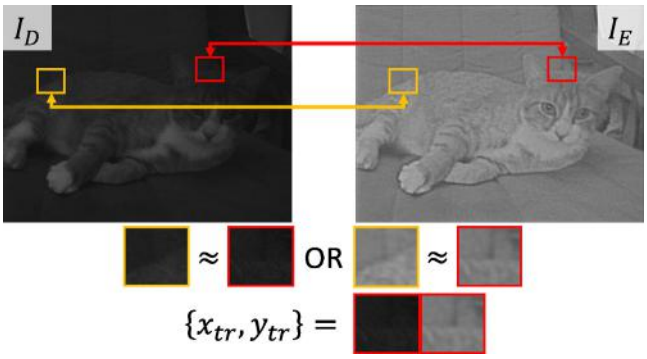

Fig. 6. $\mathcal{GP}$ training input I_D (in luminance channel, Y) and output I_E, and training data optimization. Note that the training data x_{tr} and y_{tr} are corresponding patches from I_D and I_E respectively, enabling edge/texture relationships to be preserved in the $\mathcal{GP}$. If there are multiple patches pairs that are similar, only one pair is used for training to minimize computational cost. [Best viewed in color].

from a corresponding reference I_E, as illustrated in Fig. 5. To this end, the training inputs $\bar{P}_D = \{\bar{P}_{D,1}, \bar{P}_{D,2}, \ldots, \bar{P}_{D,k}\}$ and outputs $\bar{P}_E = \{\bar{P}_{E,1}, \bar{P}_{E,2}, \ldots, \bar{P}_{E,k}\}$ are the average intensities of local patches, $P_D = \{P_{D,1}, P_{D,2}, \ldots, P_{D,k} | P_D \subseteq I_D\}$ and $P_E = \{P_{E,1}, P_{E,2}, \ldots, P_{E,k} | P_E \subseteq I_E\}$.

From our study, we found that the sizes of P_D and P_E heavily influence the posterior distribution and quality of the enhancement, where smaller sizes give better results. This is because the region size constrains the precision of the mapping distribution. As the size decreases, more constraints are introduced and brings more precise distribution. However, there is a trade-off where more constraints will drastically increase the computational cost. Due to this reason, we searched for an inflection point, and fixed P_D and P_E as patches of 4×4 pixels. In addition, many patches in an image bear a superficial resemblance, we therefore further optimize the computation by using only one region-pair instead of multiple similar ones as shown in Fig. 6, i.e. $\{x_{tr}, y_{tr}\} = \{P_{D,i}, P_{E,i}\}$ if $\{\bar{P}_{D,i} \approx \bar{P}_{D,j} | \{\bar{P}_{D,i}, \bar{P}_{D,j}\} \subset \bar{P}_D\} \vee \{\bar{P}_{E,i} \approx \bar{P}_{E,j} | \{\bar{P}_{E,i}, \bar{P}_{E,j}\} \subset \bar{P}_E\}$ where $i \neq j$. The average intensity of the patches are used as the similarity measure.

4.3. $\mathcal{GP}$ Training data

The $\mathcal{GP}$ is able to model low-light enhancement as localized functions, hence, able to differentiate the functions to brighten, maintain, or even darken the pixels for best results. By itself, the $\mathcal{GP}$, akin to other statistical models, will build a distribution based on the given input and output training pixels irregardless of the spatial structure or features. Thus, the challenge is to provide the training data where the input and output training pixels reflect the features of the image.

For this reason, we adopt a CNN model to learn the feature to pixel relationship from a large external data which will then be provided to the $\mathcal{GP}$ to construct the distribution. The choice of CNN for this learning is due to its renowned ability to learn features in an unsupervised manner from a supervised task training and also its successful performance in pixel-wise transformation works such as image denoising [32] and super-resolution [33]. We also note that there are works that adopt CNNs to directly perform the enhancement [6,7], however, we observed that such models are still lacking with respect to the localized problem of low-light enhancement. As known, CNN models consist of a set design of layers and parameters to be trained from a large collection of data. The multi-layered approach can provide multi-function enhancement, however, once the model is trained, the enhancement functions are fixed for one optimal enhancement across different images, i.e. globally optimized. On the contrary, our intuition for solving this problem comes from the study of low-light image and the necessity for localized enhancement, which lead us to propose the $\mathcal{GP}$ as a locally optimized solution, and thus we merely used the CNN to provide the specified data for training the $\mathcal{GP}$.

We modified [33]'s architecture by incorporating an additional layer to achieve our objective as shown in Fig. 7. The CNN model is trained to perform synthetic low-light image I_{D_s} to bright image I_B mapping operation $g : I_{D_s} \mapsto I_B$ where weights of the trained model capture representative features of a low-light image and projected back to pixel values in the bright spectrum as a result. Hence, whenever given a real low-light image I_D, the model will be able to extract object features and produce pixel values, as seen in Fig. 6 as example, to be the output training data I_E for the $\mathcal{GP}$.

We used the Mean Squared Error (MSE) as the loss function to train the model:

$$\mathcal{L}(I_E, I_B) = \frac{1}{n}\sum_{i=1}^{n} \|I_E - I_B\|^2 \tag{6}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \|g(I_D|\theta_{\text{CNN}}) - I_B\|^2, \tag{7}$$

where n refers to the number of training data. This function is minimized by stochastic gradient descent and backpropagation with the following weight and bias updates for each layer:

$$w_i^{\text{new}} = w_i^{\text{old}} - \eta_i \frac{\partial g}{\partial w}(w_i^{\text{old}}),\ b_i^{\text{new}} = b_i^{\text{old}} - \eta_i \frac{\partial g}{\partial b}(b_i^{\text{old}}) \tag{8}$$

where $i = \{1, 2, 3, 4\}$, w are the convolution weights, b are the biases, and η_i is the learning rate for the convolution layers. We set $\eta_1 = \eta_2 = 10^{-4}$, and $\eta_3 = \eta_4 = 10^{-5}$ to promote network convergence. It should be noted that the aim is not to train the model to replicate the exact bright image, but to reconstruct the extracted features into pixel space. Therefore, minimizing MSE updates the weight to produce the reference pixels I_E to be in the range of pixel values from real images as it is a crucial element for building an appropriate $\mathcal{GP}$ distribution.

4.3.1. CNN training

To the best of our knowledge, currently available data make use of exposure settings to simulate bright and low-light data [7,20]. However, the scope of our study is targeted on real low-light images with dynamic objects, similar to those provided by [1], which is impractical to be simulated with varying camera exposure. Therefore, we synthetically darken real bright images I_b containing objects to generate the artificial low-light counterpart I_d. This will enable us to generate a significantly large amount of data to train the CNN model for feature extraction and pixel mapping.

The darkening operation we used is a combination of contrast scaling and gamma correction:

$$I_d = C_{lim} I_b^{\gamma}, \tag{9}$$

where C_{lim} is the upper intensity limit of I_d and γ is the gamma value. Different combinations of C_{lim} and γ are applied to generate different levels of low-light images to be learned.

A similar approach is used by [24] to generate their training data as well, however we justify our scheme by doing an analysis on real low-light images in comparison to our synthesis outcome. We extracted 150 low-light images from the ExDark dataset, inclusive of both indoor and outdoor environments, for analysis. Examples of these images are shown in Fig. 2. These images were divided into non-overlapping patches and the average luminance of each patch was obtained. These values were then binned and the distribution trend was observed for patch sizes of 9×9, 17×17 and 32×32. As shown in Fig. 8(a), the number of patches extracted in the real low-light images logarithmically decreases as the intensity level increases, irrespective of the patch sizes.

Based on this observation, we expected that by synthetically darkening bright images using Eq. (9), the low-light image patches could exhibit a similar trend. To this end, 150 bright images were randomly sampled from the Microsoft COCO dataset (MS COCO) [34] (examples shown in bottom row of Fig. 2), where Fig. 8(b) shows that the distri-

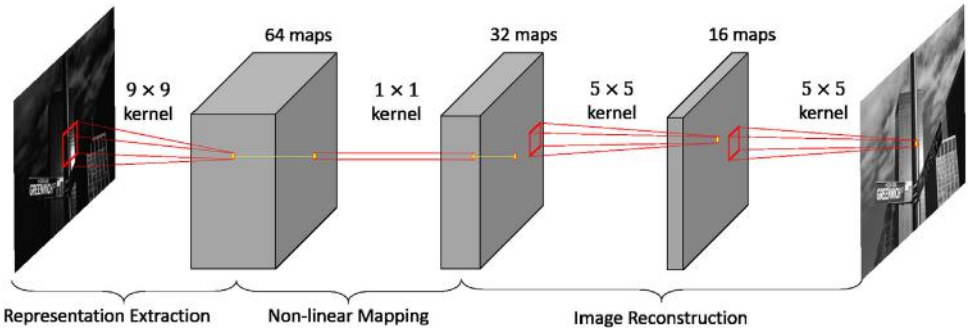

Fig. 7. Convolutional Neural Network architecture used for generating $\mathcal{GP}$ training output.

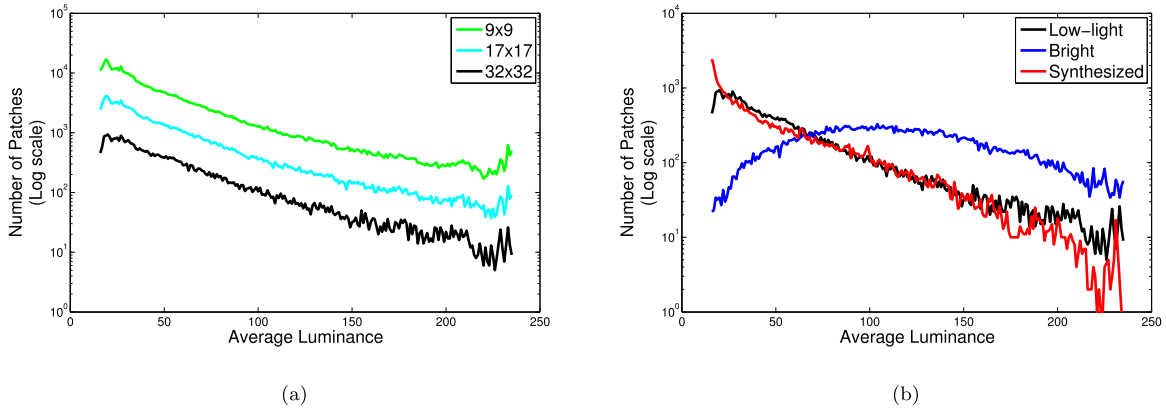

Fig. 8. Distribution comparison of average intensities of image patches. (a) Between different patch sizes from the real low-light images only. (b) Between patches from both the real and synthesized low-light images with patch size of 32 × 32. [Best viewed in color].

Fig. 9. Examples of low-light images synthesized from one bright image using different configurations. (a) $C_{lim} = 250, \gamma = 1$; (b) $C_{lim} = 200, \gamma = 2$; (c) $C_{lim} = 150, \gamma = 3$; (d) $C_{lim} = 100, \gamma = 4$; (e) $C_{lim} = 50, \gamma = 5$.

bution of average luminance in their patches (blue) greatly differs from the low-light images (black). We then darken these bright images with the combination of $C_{lim} = \{250, 200, 150, 100, 50\}$ and $\gamma = \{1, 2, 3, 4, 5\}$ to produce 25 levels of darkening for a single bright patch. Fig. 9 shows five examples of different darkening levels of the same image. The intensity distribution of the synthesized patches of size 32 × 32 shows a similar trend (red) as that of real low-light image patches in Fig. 8(b), particularly for lower intensity levels. Though synthetic, this approximation towards real data is sufficient to train a reliable transformation model and more importantly enable the generation of large enough data for the CNN model to be optimized across many variations of low-light conditions. Subsequently, the model can produce reliable training data for the $\mathcal{GP}$.

5. Experiments

This section describes the implementation details of our proposed method, including training data, and evaluation results in comparison to the latest work in low-light image contrast enhancement. We compare our proposal to the state-of-the-arts including CLAHE [28], Li et al. (LACE) [2], Fu et al. (FBE) [3], Fu et al. (WVM) [4], Guo et al. (LIME) [5], Li et al. (SRRM) [25], Tanaka et al. (GBLE) [35], Wang et al. (GladNet) [6], and Chen et al. (RetinexNet) [7]. LACE [2] is the luminance adaptive contrast enhancement method built up from the image haze removal algorithm [27], without the denoising module stated in the paper. To establish a fair comparison with the other methods that do not explicitly deal with noise, the code of [2] is reimplemented from [27][3] based on the details given by the paper whereas the codes of [3],[4] [4],[5] [5],[6] [35],[7] [25],[8] [6],[9] and [7][10] were used as provided by the authors.

[3] https://github.com/sjtrny/Dark-Channel-Haze-Removal.
[4] http://smartdsp.xmu.edu.cn/weak-illumination.html.
[5] http://smartdsp.xmu.edu.cn/cvpr2016.html.
[6] https://sites.google.com/view/xjguo/lime.
[7] http://www.ok.sc.e.titech.ac.jp/res/IC/LowLight/.
[8] http://www.icst.pku.edu.cn/struct/Projects/RRM.html.
[9] https://github.com/weichen582/GLADNet.
[10] https://daooshee.github.io/BMVC2018website/.

Fig. 10. Example of the contrast enhancement on a real low-light image, and the intensity of each pixel before and after enhancement of the respective methods (arranged in ascending order of pixel values from the original low-light image). [Best viewed in color].

We use the conventional quality metric as well as propose new information retrieval metrics to validate our method. These quantitative evaluations were performed on a sub-dataset we sampled from the validation data of MS COCO that consists of 300 bright images which we darken according to the scheme stated in Section 4.3.1. This gives a total of 7,500 testing images for quantitative evaluation. We also make use of the real low-light images from the ExDark dataset [1] for qualitative assessment.

5.1. Implementation details

Generally, our proposed method can be applied to images in RGB color space, grayscale, or even luminance (Y) channel. Our experimental implementation uses the luminance channel Y for lower computation complexity, whereas the chrominance components (C_bC_r) are unaltered and only used for producing the final colored image. Similarly, the reference pixels estimated by the CNN model is in the Y channel.

For the CNN training data, we used the entire MS COCO training set which contains 82,783 images for the synthesis. Each image provides 26 training pairs (including the original bright patch paired with itself), hence providing a total of 2,152,358 images for CNN training and validation. The CNN model was trained using the full images resized to 256×256 pixels and normalized to the range of $[0, 1]$. We chose to train the model using full images to retain the contextual information and high level features of objects within the images, which would otherwise be "broken" if patches were extracted from the training images instead. Therefore, the CNN trained with full images would ensure the model is able to capture effective object features, whereas the local intensity variation will be dealt with by the $\mathcal{GP}$.

5.2. Qualitative evaluation

We first compare the results of our proposed method qualitatively with the state-of-the-arts. Fig. 10 shows an example of real low-light image with irregular lighting and the results produced by each method. In the regions bounded by the blue boxes, we can see that all methods result in better contrast after the enhancement. Our method, and LIME show the best brightness, contrast and sharpness, for example, the bicycle wheel bounded by the red boxes. However, LIME's result has more noise and color distortion as shown in the yellow bounding

Fig. 11. Example of the contrast enhancement on a synthesized low-light image, and the intensity of each pixel before and after synthesis and enhancement (arranged in ascending order of pixel values from the original bright image). [Best viewed in color].

boxes. It can be observed that the area next to the light source appears noisy with unnatural pinkish hue whereas our result look more natural. Moreover, the result by GBLE has the "halo" effect and the RetinexNet appears very artificial.

This observation is similar to the results of the synthesized low-light image in Fig. 11. The cow in the red bounding box and edges of the tree in the yellow bounding box of our result have better contrast than most of the methods whereas LACE, LIME and SRRM suffer from color distortions, such as the clouds in the blue bounding box showing a purplish hue. LIME does have the most aesthetically pleasing result contributed by its vibrant colors, nonetheless the details retrieved by our proposed method are still comparable even though not as apparent to us because human perceptions are sensitive to colors.

We also show that our proposal indeed enhances the darkened sample towards the original image in the pixel intensity distributions in Figs. 10 and 11. Our results' are very much less scattered than CLAHE, LACE, and GBLE while the distributions of FBE, WVM, and SRRM are still lower than the pixel value of the original bright image (blue curve) indicating under enhancement. As for LIME, the distribution is able to somehow match the original bright image but is relatively more scattered. Moreover, we found that LIME particularly fails in some scenarios as shown in Fig. 13. GladNet also shows a very good result with a matching distribution shape, however, the values are still slightly low while contrarily, the distribution of the result by RetinexNet shows a distinct difference from the rest, corresponding to its artificial output. Especially notable in Fig. 11 is the result of our proposed $\mathcal{GP}$ model where it produced a distribution that closely matches the original bright image. We attribute this to the ability of our model to perform localized enhancement that allows structural awareness contributed by the features information.

More examples of results are shown in Fig. 12–13 where various common problems arise for the current state-of-the-arts while our proposed method is still able to perform well. One of which is the noise problem, as shown in Fig. 12, that is particularly obvious for LIME, GBLE, and RetinexNet, whereas methods like LACE, FBE, and SRRM over enhance when a light source such as a street light is captured in the image. GladNet's result even has a slight "ring" of light on the surroundings. The examples in Fig. 13 shows obvious color distortions such as an unnaturally blue sea, and a reddish toilet seat. However, we do note that our results' lack color vibrancy because we only address the luminance channel in our model, but we show that it is not a hindrance to our target for higher level applications in the quantitative evaluations.

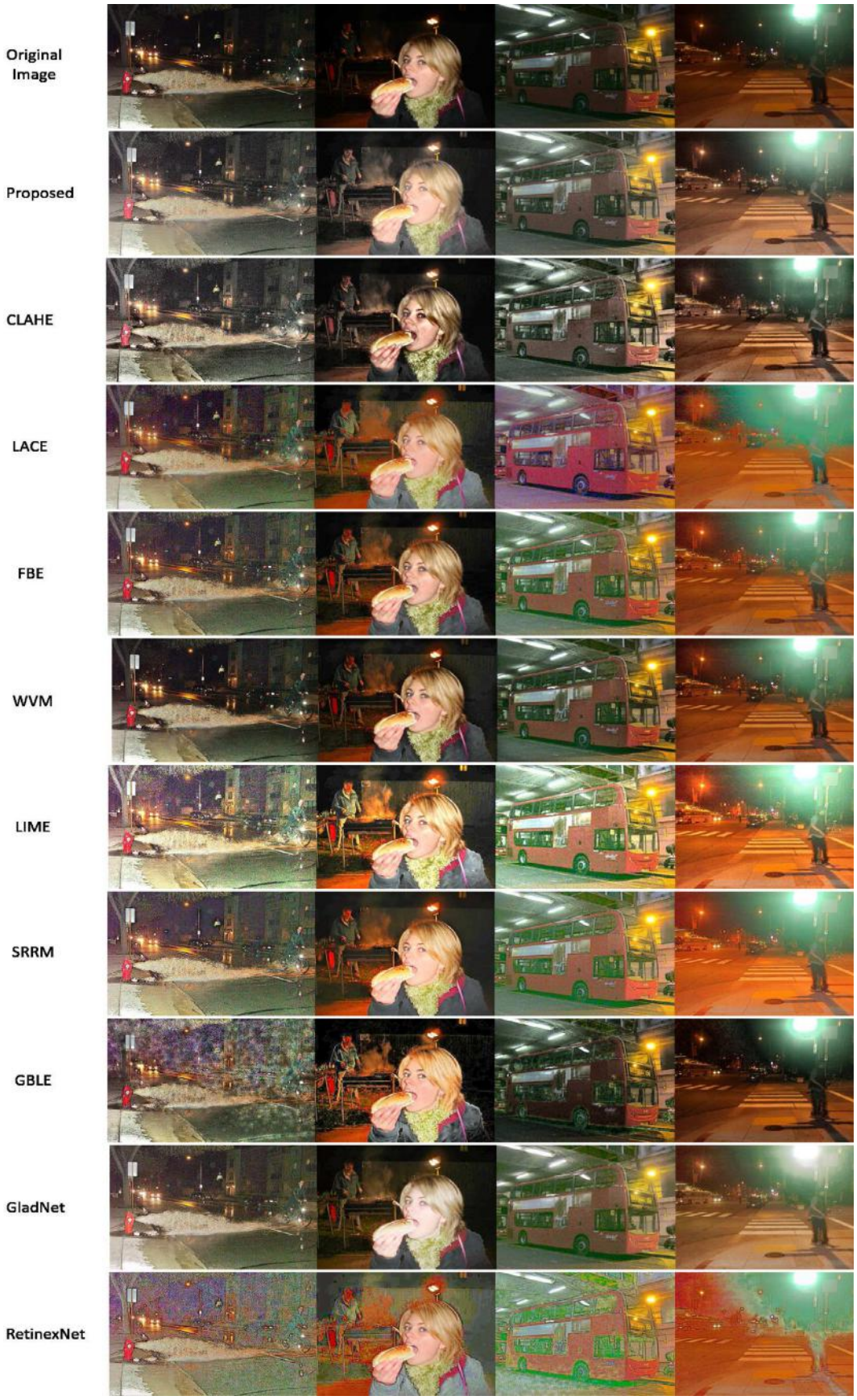

Fig. 12. Contrast enhancement results of real low-light images.

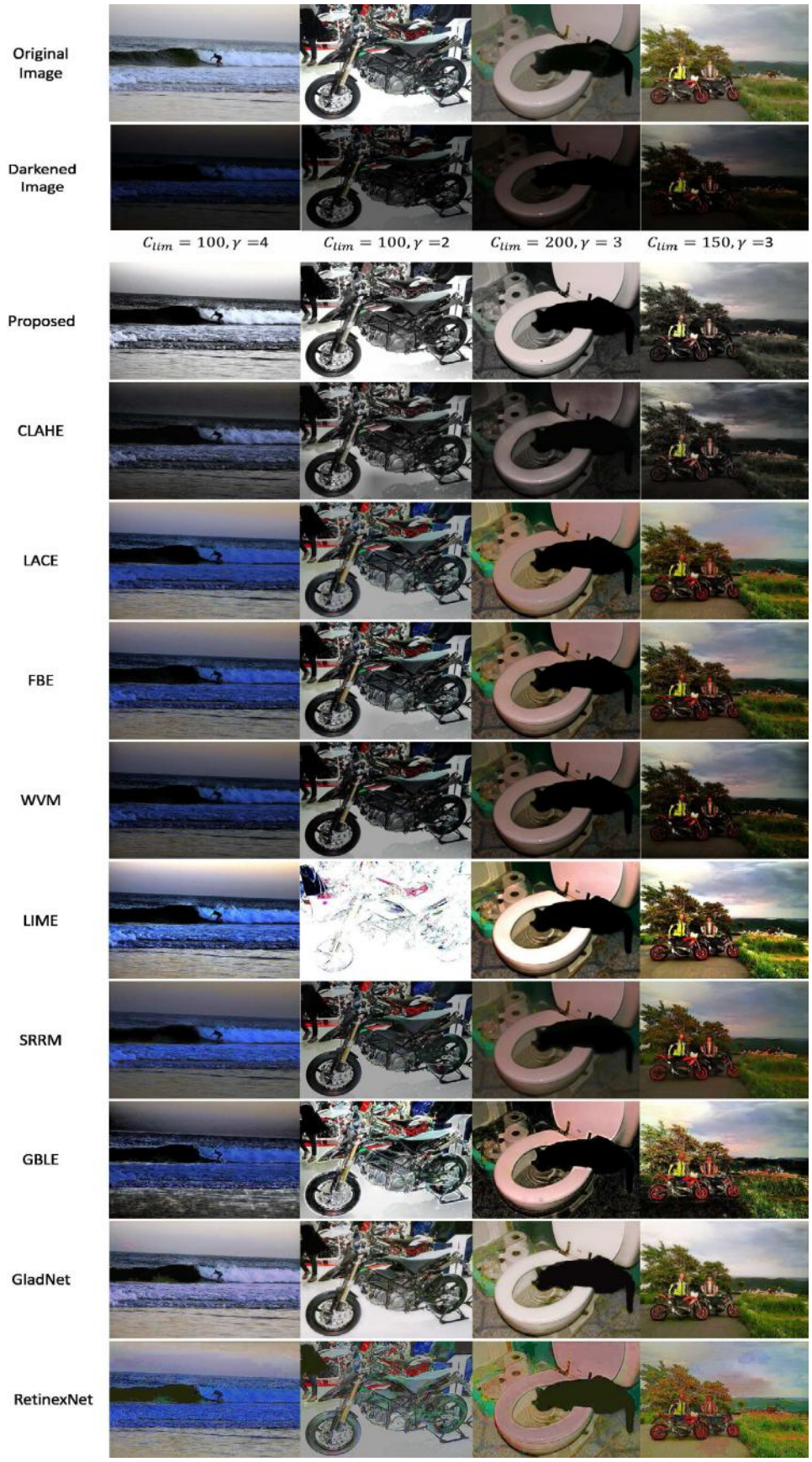

Fig. 13. Contrast enhancement results of synthesized low-light images.

Table 3
Average PSNR, computation time, precision, recall, and F-scores of feature matching.

Approach	PSNR	Time (s)	Features matching			
			Precision	Recall	F_1-score	F_2-score
Darkened	10.44 (±0.120)	–	0.4514	0.1711	0.2090	0.1824
CLAHE [28]	12.03 (±0.084)	**0.02**	0.4632	0.3928	0.3404	0.3549
LACE [2]	14.95 (±0.077)	15.97	**0.6358**	0.4305	0.4820	0.4445
FBE [3]	14.68 (±0.087)	0.17	0.5959	0.4659	0.4831	0.4639
WVM [4]	12.88 (±0.102)	3.03	0.5794	0.3458	0.3722	0.3496
LIME [5]	15.08 (±0.075)	0.07	0.3205	0.6463	0.4076	0.5062
SRRM [25]	14.10 (±0.083)	4.71	0.5779	0.3395	0.3714	0.3466
GBLE [35]	13.26 (±0.073)	0.24	0.2284	0.3303	0.2513	0.2850
GladNet [6]	**18.32** (±0.077)	0.04	0.5388	0.5704	0.5200	0.5403
RetinexNet [7]	14.63 (±0.052)	0.05	0.4477	0.2073	0.2594	0.2230
Proposed	16.25 (±0.078)	1.25	0.4745	**0.6563**	**0.5292**	**0.5871**

Table 4
Average l_1-norm distance of global color intensity histogram, and local 32 × 32 pixels patch intensity histograms.

Approach	Global	Local
Darkened	$3.96 \times 10^5 (\pm 2.85 \times 10^3)$	587.0 (±3.56)
CLAHE [28]	$3.14 \times 10^5 (\pm 1.62 \times 10^3)$	536.4 (±2.29)
LACE [2]	$2.95 \times 10^5 (\pm 2.54 \times 10^3)$	518.2 (±2.14)
FBE [3]	$2.89 \times 10^5 (\pm 2.70 \times 10^3)$	505.8 (±2.49)
WVM [4]	$3.28 \times 10^5 (\pm 2.99 \times 10^3)$	540.9 (±2.68)
LIME [5]	$2.73 \times 10^5 (\pm 2.55 \times 10^3)$	492.8 (±1.41)
SRRM [25]	$3.00 \times 10^5 (\pm 1.33 \times 10^3)$	524.2 (±2.16)
GBLE [35]	$2.89 \times 10^5 (\pm 1.69 \times 10^3)$	505.5 (±2.25)
GladNet [6]	$\mathbf{2.26 \times 10^5}$ $(\pm 1.15 \times 10^3)$	**447.6** (±2.30)
RetinexNet [7]	$2.70 \times 10^5 (\pm 1.10 \times 10^3)$	490.0 (±1.85)
Proposed	2.54×10^5 $(\pm 0.98 \times 10^3)$	*477.0* (±1.21)

5.3. Quantitative evaluation

Our quantitative assessments were carried out on 3 evaluation metrics: the PSNR, local features matching, and l_1-norm color histogram distance. We show the measures of the synthetically darkened test images enhanced by LACE, FBE, WVM, LIME, SRRM, GBLE, GladNet, RetinexNet, and our proposed framework, the $\mathcal{GP}$.

5.3.1. PSNR

Table 3 shows the average PSNR results (with their 95% confidence intervals) calculated for all RGB channels of the tested images, and the average processing time for images of size 256 × 256 pixels. Our proposed method shows comparable performance to the state-of-the-art solutions with satisfactory computation time[11]. However, we note a conflict with the qualitative assessment in Figs. 10 and 11 where LACE, FBE, and WVM produced images with better visual quality than the RetinexNet model however, the average PSNR of RetinexNet is comparable or better than them. Fig. 14 additionally shows three examples enhanced by all the methods where the PSNR displays similar inconsistency. The results of RetinexNet appear quite unnatural and have notable noise, particularly in the top and bottom examples, but still show better PSNR values.

Based on this finding, we believe that PSNR is not the ideal measure for low-light image enhancement. Moreover, such quality measurements do not bring out the significance of low-light image contrast enhancement for feature retrieval in support of higher level application such as detection. We therefore introduce local features matching and histogram l_1-norm distance as new metrics to evaluate the ability of enhancement algorithms to improve valuable details of low-light images.

[11] We used Intel Xeon Processor E5-2623 v3 at 3.00 GHz without GPU acceleration for testing.

5.3.2. Local features matching

Before the breakthrough of learned features, local features were the forerunners for detection and recognition tasks, and are still used in part to this day [36–38]. Hence, we make use of detected local features as a gage for useful information content retrieved by enhancing low-light images. Furthermore, we heighten the reliability of this measure by matching features detected from the enhancement results to the original bright image to ensure the retrieved details are not "false" features from noise and artifacts created by the enhancements. We then calculate the precision Pr, recall Rc, and F-score based on the information retrieval context as follows:

$$Pr = \frac{|q_{rlv} \cap q_{rtv}|}{q_{rtv}}; Rc = \frac{|q_{rlv} \cap q_{rtv}|}{q_{rlv}}; F_\beta\text{-score} = (1+\beta^2)\frac{Pr \cdot Rc}{(\beta^2 \cdot Pr) + Rc}, \tag{10}$$

where q_{rlv} refers to feature points extracted from the original bright image and q_{rtv} are feature points from the enhanced image, while $|q_{rlv} \cap q_{rtv}|$ indicates the correctly matched points. β is the weight variable for the precision and recall in computing the F-score. A higher β puts more weight on recall than precision.

In our experiments, we use SIFT [39] as the local features for evaluation as it can serve as a fast, lightweight and impartial measurement, instead of trained features that are very data reliant. We set the peak threshold to be 10 so as to only remain points detected from regions with strong contrast, and define correctly matched points as features with matching descriptors, location, scale and orientation. We emphasize on the useful information that enhancement can retrieve, hence the results were recall is more weighted with $\beta = 2$ (F_2-score) is included in our evaluation. Table 3 shows that our proposed method outperforms all methods in recall, as well as both F_1 and F_2 scores. We analyzed and found that the LACE method has the best precision because it retrieves less local features in total. Even so, for the average scores where at equal weights for precision and recall in the F_1-score, their performance fall behind while our proposed method is best performing. We also note that the GladNet shows the most balanced performance among all the methods with a comparable F_1-score to ours.

Fig. 14 shows some examples of the features detected from the enhancement results and matched to the features detected in the original bright images. The darkening of the images significantly impacts the features extractable from the objects, particularly in the bottom row example of Fig. 14 where the synthesized low-light image only have two features matched. Nevertheless, each enhancement method is able to retrieve some features, particularly, our proposed $\mathcal{GP}$ with the most retrievals. Moreover, the examples also showed that, high PSNR does not translate to better feature retrieval. Both LIME and GladNet are admittedly the best performing methods visually, as seen in Fig. 14, but from the standpoint of practicality, the feature retrieval our method achieved is more useful.

5.3.3. l_1-norm

The second evaluation method we propose is the histogram distance measure using l_1-norm. We calculated the l_1-norm between color histograms of the original bright image and the enhanced images. We performed the comparison on both global image intensity histogram and the histograms of local patches in the image. We obtain the local histograms by dividing the image into non-overlapping patches of 32 × 32 pixels and then calculate the average distance of all the patches. We set the intensity histograms to have 32 bins for our assessment.

Table 4 shows the average distance of the color histogram distance measure and their 95% confidence intervals. Similar to features retrieval, the enhancement methods were able to shorten the distance of both types of histograms. While GladNet is the best performing method in this evaluation, our proposed method comes in at second place for both global and local histogram measures as marked in italics, with comparable results even though we did not explicitly enhance the color content like the others. We deduced that, in the efforts of enhancing the

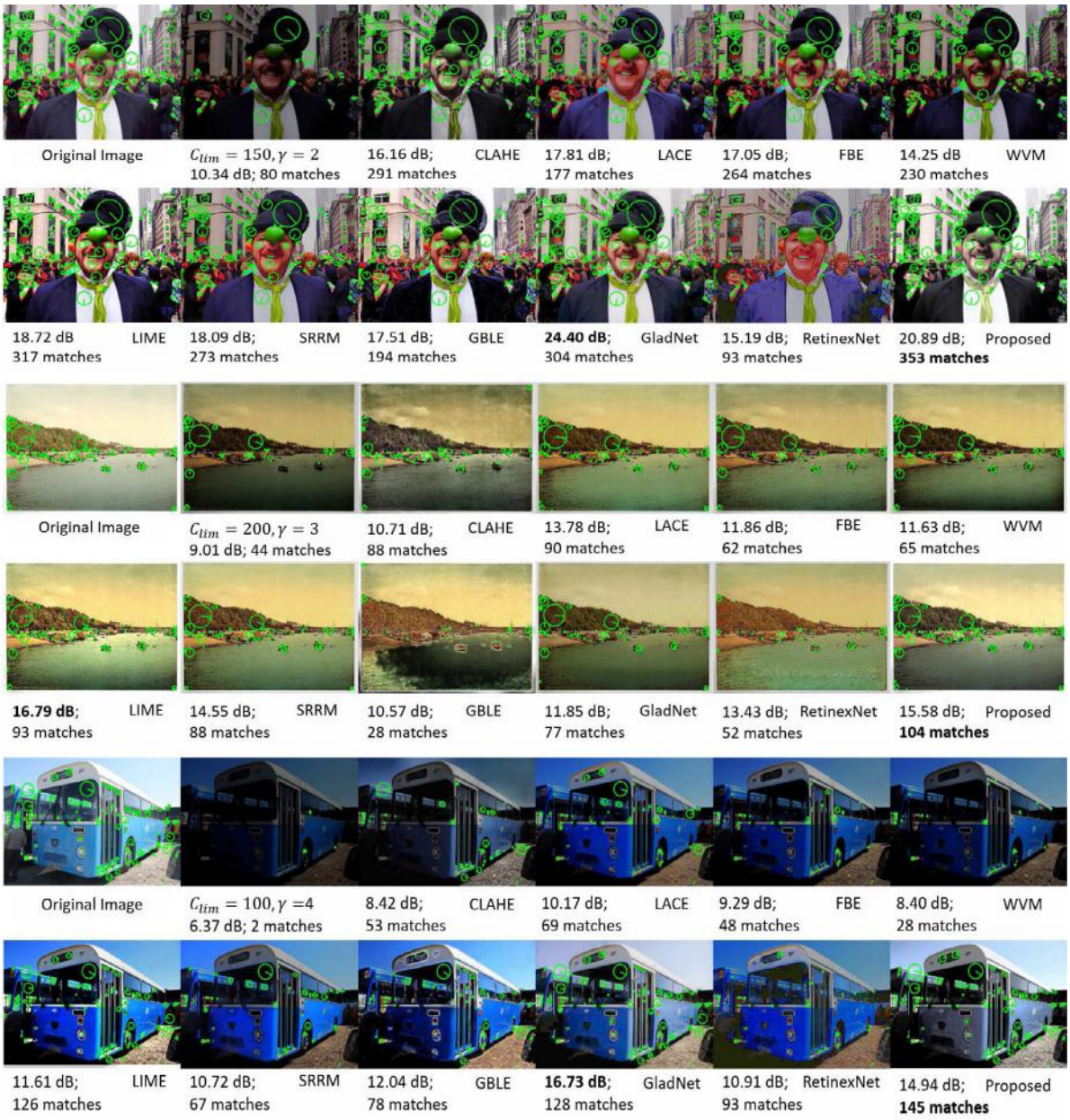

Fig. 14. PSNR values and SIFT features matched in synthetic low-light images using different enhancements methods. [Best viewed in color].

low-light images, most of the methods skewed the color spectrum, hence the mismatch with the reference image, as notable in the top example of Fig. 14 where the black suit jacket of the man has been altered to be blue, particularly by LACE and RetinexNet. Contrarily, our method is able to bring the image closer to the original state, although less pleasing to human observation.

5.4. Study of $\mathcal{GP}$ training data

In our proposed framework, we stressed on the aim to perform localized enhancement for features retrieval. The training of the $\mathcal{GP}$ depends on the data that would provide the necessary enhancement relationship of features, which we do so by a CNN model. The CNN model learns image features and projects them into pixel values to act as the training output for the runtime training of the $\mathcal{GP}$.

We therefore conducted a study by using various state-of-the-art enhancement methods to provide the reference pixels for the $\mathcal{GP}$ training. We selected the CLAHE, LIME, and GladNet for this study as CLAHE is a well established and fast contrast enhancement method, while LIME is due to its visual vibrancy, and GladNet for its overall balanced performance in our evaluations.

Table 5
Performance of $\mathcal{GP}$ using training output generated by our CNN, CLAHE, LIME, and GladNet.

Approach	PSNR	Features matching				Histogram l_1-norm	
		Precision	Recall	F_1-score	F_2-score	Global	Local
CNN	*16.25*	*0.4745*	**0.6563**	*0.5292*	*0.5871*	2.54×10^5	*477.0*
CLAHE [28]	14.03	0.4478	0.4726	0.4250	0.4407	2.93×10^5	521.8
LIME [5]	15.58	0.4287	0.5850	0.4679	0.5187	2.62×10^5	484.7
GladNet [6]	**17.04**	**0.5054**	*0.6511*	**0.5433**	**0.5918**	$\mathbf{2.44 \times 10^5}$	**465.7**

Table 5 shows the average performance of different models where using our original model still achieve the best recall, though the training by GladNet shows improved results for all the other evaluations. On the other hand, the performance of the $\mathcal{GP}$s trained by the pixels from

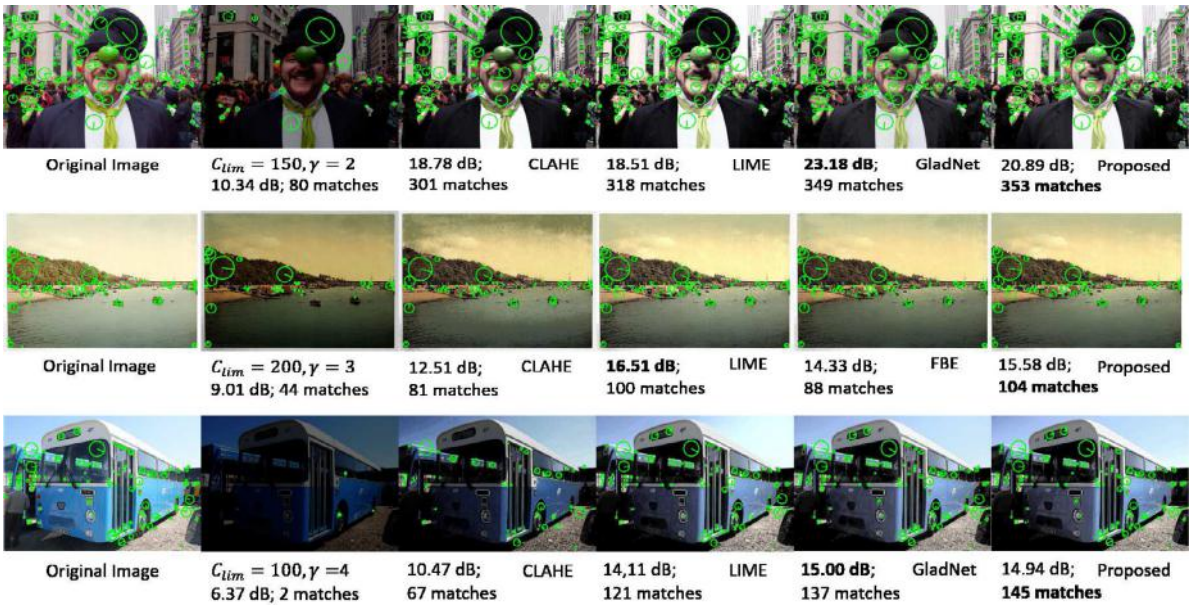

Fig. 15. PSNR values and SIFT features matched in synthetic low-light images using $\mathcal{GP}$ trained by different reference values. [Best viewed in color].

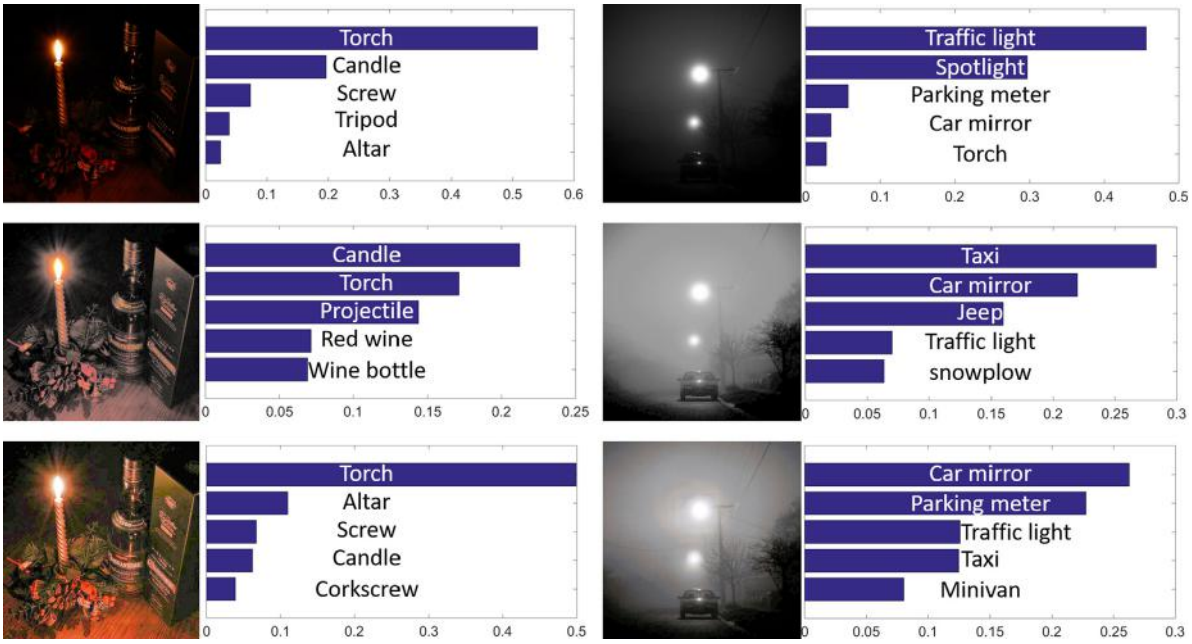

Fig. 16. Top 5 object classification results of low-light image (top), enhanced by our proposed method (middle), and enhanced by GladNet (bottom), using ResNet-101 trained on ImageNet. [Best viewed in color].

CLAHE and LIME are unable to compete with the deep models. We make two deductions from these results. First, the CLAHE and LIME are unable to provide the $\mathcal{GP}$ with features aware reference data for the training. As we note, these methods enhances the visual aspects of the low-light images, where they do not explicitly consider the structures of the content for the enhancement. Therefore, the $\mathcal{GP}$ is not a simple pre-processing algorithm.

The second deduction is based on the results using our CNN model and GladNet that significantly outperforms both the CLAHE and LIME training. These results clearly support our intuition that the ability of CNNs to capture the necessary features to pixels relationship provides the $\mathcal{GP}$ with the appropriate reference values for training the model. We note that the GladNet architecture is very much more complex than our 3-layer CNN, hence we believe the slight performance boost is in its ability to learn better features. More importantly, we attribute this increment to the localized enhancement brought by the distribution modeling of the $\mathcal{GP}$ that is lacking in CNN-based models, as made apparent by the results of our framework surpassing that of CNN only models, i.e. GladNet and RetinexNet. Fig. 15 shows example results from this study. Except for the CLAHE trained model that suffers from under enhancement, the results from models trained by LIME and GladNet shows similar quality to our proposed model.

5.5. Object classification test

In this section, we take a step further to look into the potential of low-light image enhancement in assisting the popular computer vision task, object classification. To do so, we performed tests using the 101-layer Residual Network [40] (ResNet-101) on real low-light images, and their enhanced versions by our proposed framework and also GladNet. We directly used the ResNet-101 model that is trained on ImageNet[12] to do the classification without any re-training or fine-tuning involved as we would like to observe the outcome of the originally optimized model when given low-light and enhanced images.

Fig. 16 shows example of the results we obtained where we can see notable changes in the classification results before and after the enhancements have been done. For the left example, the model is unable to classify the bottle of wine in the low-light image, likely due to the low illumination as compared to the candle. The image enhanced by our model enabled the model to not only detect the candle, but also the initially missed wine bottle. On the other hand, the GladNet result seem to have deteriorated the image for the classification, even though for human vision, it is easier for us to see. Nevertheless, the right example shows that both enhancements are able to somewhat improve the performance, classifying the car as a taxi that was missed by the model in the original low-light image.

This test illustrates the feasibility of incorporating image enhancement as a support for practical applications. Additionally, this demonstration also shows that our proposed method can indeed serve as a backing for tasks such as object classification and our proposed evaluation metrics is consistent with the purpose we have set them out to assess.

6. Conclusion

In this paper, we look into the enhancement problem with the objective of information retrieval instead of aesthetic restoration. We modeled low-light image enhancement as a distribution of localized enhancement functions using Gaussian Process trained at runtime with reference data generated from a CNN. We trained the CNN using large data synthesized based on luminance statistics of real images so that it learns the relationship of feature to pixel values. Thus, the reference generated trains the $\mathcal{GP}$ to perform features aware enhancement. Furthermore, we introduced two information retrieval assessment that points out the practicability of enhancement results for high level applications instead of the traditional image quality assessment that has been widely used in the field. Our proposed framework is a new approach that is unlike the conventional low-light image enhancement methods and shows competitive performance, both qualitatively and quantitatively, in the new evaluation metrics. Additionally, our analysis found that the PSNR quality metric miss assesses image quality where evidently artificial images are still given high scores, hence, our shift of focus with the new evaluation metrics would bring a new direction to the field.

In the future work, we intend to further improve low-light object detection to meet the underlying motivation of this research work, meanwhile, to also address the noise issue that commonly affects camera sensors when capturing low-light images, as well as improve the color reproduction.

Acknowledgments

This research is supported in part by the Fundamental Research Grant Scheme (FRGS) MoHE Grant FP004-2016, from the Ministry of Education Malaysia; and in part by the Postgraduate Research Grant (PPP) PG002-2016A, from University of Malaya. Also, we gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Z GPU used for this research.

[12] Trained model obtained from http://www.vlfeat.org/matconvnet/pretrained/.

Conflict of interest

No conflict of interest.

References

[1] Y.P. Loh, C.S. Chan, Getting to know low-light images with the exclusively dark dataset, Comput. Vis. Image Understanding 178 (2018) 30–42.
[2] L. Li, R. Wang, W. Wang, W. Gao, A low-light image enhancement method for both denoising and contrast enlarging, in: ICIP, 2015.
[3] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, J. Paisley, A fusion-based enhancing method for weakly illuminated images, Signal Process. 129 (2016) 82–96.
[4] X. Fu, D. Zeng, Y. Huang, X.-P. Zhang, X. Ding, A weighted variational model for simultaneous reflectance and illumination estimation, in: CVPR, 2016, pp. 2782–2790.
[5] X. Guo, Y. Li, H. Ling, Lime: Low-light image enhancement via illumination map estimation, IEEE Trans. Image Process. 26 (2) (2017) 982–993.
[6] W. Wang, C. Wei, W. Yang, J. Liu, Gladnet: Low-light enhancement network with global awareness, in: FG, IEEE, 2018, pp. 751–755.
[7] C. Wei, W. Wang, W. Yang, J. Liu, Deep retinex decomposition for low-light enhancement, in: BMVC, 2018.
[8] D. Kang, H. Han, A.K. Jain, S.-W. Lee, Nighttime face recognition at large standoff: Cross-distance and cross-spectral matching, Pattern Recognit. 47 (12) (2014) 3750–3766.
[9] B. Qi, V. John, Z. Liu, S. Mita, Use of sparse representation for pedestrian detection in thermal images, in: CVPR-W, 2014.
[10] X. Zhao, Z. He, S. Zhang, D. Liang, Robust pedestrian detection in thermal infrared imagery using a shape distribution histogram feature and modified sparse representation classification, Pattern Recognit. 48 (6) (2015) 1947–1960.
[11] L.T. Yuan, S.K. Swee, T.C. Ping, Infrared image enhancement using adaptive trilateral contrast enhancement, Pattern Recognit. Lett. 54 (2015) 103–108.
[12] S.-C. Huang, F.-C. Cheng, Y.-S. Chiu, Efficient contrast enhancement using adaptive gamma correction with weighting distribution, IEEE Trans. Image Process. 22 (3) (2013) 1032–1041.
[13] J. Lim, J.-H. Kim, J.-Y. Sim, C.-S. Kim, Robust contrast enhancement of noisy low-light images: Denoising-enhancement-completion, in: ICIP, 2015.
[14] A. Łoza, D.R. Bull, P.R. Hill, A.M. Achim, Automatic contrast enhancement of low-light images based on local statistics of wavelet coefficients, Digit. Signal Process. 23 (6) (2013) 1856–1866.
[15] X. Wu, A linear programming approach for optimal contrast-tone mapping, IEEE Trans. Image Process. 20 (5) (2011) 1262–1272.
[16] H. Fu, H. Ma, S. Wu, Night removal by color estimation and sparse representation, in: ICPR, 2012.
[17] X. Dong, G. Wang, Y.A. Pang, W. Li, J.G. Wen, W. Meng, Y. Lu, Fast efficient algorithm for enhancement of low lighting video, in: ICME, 2011.
[18] X. Zhang, P. Shen, L. Luo, L. Zhang, J. Song, Enhancement and noise reduction of very low light level images, in: ICPR, 2012.
[19] Y.P. Loh, C.S. Chan, Unveiling contrast in darkness, in: ACPR, IEEE Computer Society, 2015.
[20] C. Chen, Q. Chen, J. Xu, V. Koltun, Learning to see in the dark, in: CVPR, 2018.
[21] P. Sen, N.K. Kalantari, M. Yaesoubi, S. Darabi, D.B. Goldman, E. Shechtman, Robust patch-based hdr reconstruction of dynamic scenes, ACM Trans. Graph. 31 (6) (2012) 203–1.
[22] T.-H. Oh, J.-Y. Lee, Y.-W. Tai, I.S. Kweon, Robust high dynamic range imaging by rank minimization, IEEE Trans. Pattern Anal. Mach. Intell. 37 (6) (2015) 1219–1232.
[23] J. Cai, S. Gu, L. Zhang, Learning a deep single image contrast enhancer from multi-exposure images, IEEE Trans. Image Process. 27 (4) (2018) 2049–2062.
[24] K.G. Lore, A. Akintayo, S. Sarkar, Llnet: A deep autoencoder approach to natural low-light image enhancement, Pattern Recognit. 61 (2017) 650–662.
[25] M. Li, J. Liu, W. Yang, X. Sun, Z. Guo, Structure-revealing low-light image enhancement via robust retinex model, IEEE Trans. Image Process. 27 (6) (2018) 2828–2841.
[26] L. Shen, Z. Yue, F. Feng, Q. Chen, S. Liu, J. Ma, Msr-net: Low-light image enhancement using deep convolutional network, arXiv preprint arXiv:1711.02488.
[27] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. 33 (12) (2011) 2341–2353.
[28] K. Zuiderveld, Contrast limited adaptive histogram equalization, Graph. Gems (1994) 474–485.
[29] M. Kaur, J. Kaur, J. Kaur, Survey of contrast enhancement techniques based on histogram equalization, Int. J. Adv. Comput. Sci. Appl. 2 (7) (2011) 137–141.
[30] E.H. Land, et al., The Retinex Theory of Color Vision, Citeseer, 1977.
[31] H.-C. Wang, Y.-C. Lai, W.-H. Cheng, C.-Y. Cheng, K.-L. Hua, Background extraction based on joint gaussian conditional random fields, IEEE Trans. Circuits Syst. Video Technol. 28 (11) (2018) 3127–3140.
[32] V. Jain, S. Seung, Natural image denoising with convolutional networks, in: NIPS, 2009.
[33] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2) (2015) 295–307.
[34] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft coco: Common objects in context, in: ECCV, Springer, 2014.

[35] M. Tanaka, T. Shibata, M. Okutomi, Gradient-based low-light image enhancement, arXiv preprint arXiv:1809.09297.
[36] F.S. Khan, R.M. Anwer, J. van de Weijer, A.D. Bagdanov, M. Vanrell, A.M. Lopez, Color attributes for object detection, in: CVPR, 2012.
[37] S. Zhang, R. Benenson, B. Schiele, Filtered channel features for pedestrian detection, in: CVPR, 2015.
[38] R. Mottaghi, Y. Xiang, S. Savarese, A coarse-to-fine model for 3d pose estimation and sub-category recognition, in: CVPR, 2015.
[39] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[40] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: CVPR, 2016, pp. 770–778.