# Zero-Reference Low-Light Image Enhancement


Arun Talkad
arun.talkad@sjsu.edu
*San Jose State University*

Kumuda BG Murthy
kumuda.benakanahalliguru
prasadamurt@sjsu.edu
*San Jose State University*

Raghava D Urs
raghavadevaraje.urs@sjsu.edu
*San Jose State University*

Shiv Kumar Ganesh
shivkumar.ganesh@sjsu.edu
*San Jose State University*



**Abstract— The research introduces a new method called Zero-Reference Low Light Enhancement, which treats light enhancement as a deep network task of image-specific curve estimation. Our method uses DCE-Net, a lightweight deep network, to predict pixel-wise and high-order curves for image dynamic range modification. Zero-DCE is intriguing because it makes no assumptions about reference images during training, i.e. it doesn't require any paired or unpaired data. This is accomplished via a series of properly constructed non-reference loss functions that implicitly quantify the network's enhancement quality and drive its learning. Extensive tests on a variety of benchmarks show that our technology outperforms state-of-the-art methods both qualitatively and numerically.**



***Keywords—Zero-DCE, Low Light image enhancement***


## I. INTRODUCTION

Many images are taken in less-than-ideal lighting settings due to unforeseen environmental and/or technical factors constraints. Inadequate and imbalanced are two of them. wrong positioning, lighting conditions in the environment of things in front of a bright light source, as well as under-exposure while the image is being captured Low-light photography suffers from several issues. Aesthetic quality is degraded, and information is transmitted in an inadequate manner. The former has an impact on the viewer's experience, while the latter sends out the wrong message, such as inaccurate object/face identification.

To address this problem, a deep learning-based image enhancement method is introduced in this paper. The Zero-Reference based approach can work in a variety of lighting situations, including non-uniform and low lighting. We reformulate the issue as an image-specific curve estimation problem rather than conducting image-to-image mapping. The suggested approach takes a low-light image as input and outputs high-order curves. These curves are then used for pixel-wise adjustment on the dynamic range of the input to obtain an enhanced image.

The curve estimation is carefully crafted to preserve the improved image's range as well as the contrast of surrounding pixels. The suggested network is lightweight, and it may be used to repeatedly approximate higher-order curves for more reliable and precise dynamic range correction.

The zero-reference nature of our deep learning-based method, as opposed to existing CNN-based [1,2] and GAN-based methods [3,4], means it does not require any paired or even unpaired data in the training phase. This is made feasible by a collection of specially constructed nonreference loss functions that consider multi-factor light enhancement, including spatial consistency loss, exposure control loss, color constancy loss, and illumination smoothness loss. We show that Zero-DCE can compete with existing approaches that need paired or unpaired data for training even with zero-reference training.

In both qualitative and quantitative indicators, our Zero-DCE technique outperforms the competition. More crucially, it can improve high-level visual tasks like face detection without putting a lot of strain on the computer.

## II. RELATED WORK

Light enhancement is achieved using HE-based technologies by increasing the dynamic range of an image. The global [5, 6] and local [7, 8] histogram distribution of images are both modified. There are also several ways based on the Retinex theory [9], which divides an image into reflectance and illumination.

Light enhancement is posed as an illumination estimation issue since the reflectance component is generally considered to be constant under all lighting situations. Several techniques have been presented based on the Retinex idea. When dealing with photos of nonuniform lighting, Wang et al. [10] devised an approach that preserves naturalness and information. A weighted variation model was presented by Fu et al. [11] to estimate the reflectance and illumination of an input image concurrently. Guo et al. [12] initially calculated a coarse illumination map by looking for the greatest intensity of each pixel in RGB channels, then refined the coarse illumination map using a structure prior. A novel Retinex model that takes noise into account was proposed by Li et al. [13]. An optimization problem was used to estimate the illumination map.

Unlike traditional approaches that rely on possibly erroneous physical models or change the distribution of the picture histogram by chance, the suggested Zero-DCE method achieves an improved result by image specific curve mapping. This method allows for image brightness augmentation while avoiding unrealistic artefacts.

Yuan and Sun [14] proposed an automatic exposure correction method in which a global optimization algorithm estimates the S-shaped curve for a given image and curve mapping pushes each segmented region to its optimal zone.

Unlike [14], our Zero-DCE is a totally data-driven method that considers several light enhancement parameters in the construction of the non-reference loss functions, resulting in improved resilience, a wider picture dynamic range adjustment, and reduced computing overhead.

CNN-based and GAN-based methods are the two main branches of data-driven methods. Because most CNN-based solutions use paired data for supervised training, they use a lot of resources. The paired data is frequently acquired exhaustively through automatic light degradation, altering camera settings during data capture, or synthesizing data via image retouching. The LLNet [15] was trained on data simulated with random Gamma correction; the LOL dataset [16] of paired low/normal light images was collected by varying the exposure time and ISO during image acquisition; and the MIT-Adobe FiveK dataset [3] contains 5,000 raw images, each of which has five retouched images produced by trained experts.

Wang et al. [1] recently suggested an underexposed photo improvement network based on illumination map estimation. This network was trained using paired data that had three experts retouch it. Light enhancement techniques based on paired data are understandably unfeasible in many aspects, given the significant cost of gathering adequate paired data as well as the inclusion of factitious and unrealistic data in deep model training. The limited generalization power of CNN-based algorithms reflects these limits. When these technologies are applied to real-world photos of varying light intensities, artefacts and color casts are typical.

The advantage of unsupervised GAN-based approaches is that they do not require paired data for training. EnlightenGAN [3], a pioneering unsupervised GAN-based system for learning to enhance low-light photos from unpaired low/normal light data. The network was trained using discriminators and loss functions that were carefully developed. Unsupervised GAN-based solutions, on the other hand, frequently necessitate the careful selection of unpaired training data. In three ways, the proposed Zero-DCE outperforms existing data-driven techniques. First, it investigates a novel learning technique based on zero reference, which eliminates the need for both paired and unpaired data. Second, the network is trained using non-reference loss functions that have been properly constructed. This technique allows for an implicit evaluation of output image quality, with the findings being repeated for network learning. Finally, our approach is both efficient and cost-effective. Our zero-reference learning methodology, lightweight network structure, and effective non-reference loss functions all contribute to these benefits.

## III. The Data

### A. Dataset

The most important task of a machine learning project is to have a good dataset and apply the right methodologies to it. Throughout the paper, we follow the various steps in KDD to enhance the low light image.

The LOL dataset is composed of 500 low-light and normal-light image pairs and divided into 485 training pairs and 15 testing pairs. The low-light images contain noise produced during the photo capture process. Most of the images are indoor scenes. All the images have a resolution of 400×600.

The data under consideration can be found at the link - https://drive.google.com/file/d/157bjO1_cFuSd0HWDUuAmcHRJDVyWpOxB/view

### B. Data Pre-processing

The images from the LOL dataset are preprocessed before using them in the training and validation phases. The data preprocessing involves three steps: decomposition, downscaling and normalization.

*1) Decoding:* The PNG image data is in the form of a compressed string. The PNG data is decoded into 8-bit integer tensors to make computation easier.

*2) Downscaling:* Downscaling is the reduction of the image to a specific resolution. Neural networks require inputs of the same size and the training process is made faster by downscaling the images to a pre-determined size. The images have a resolution of 400 by 600 pixels, it is reduced to 256 by 256 in this study.

*3) Normalization:* Neural Nets reach convergence more quickly if the data is normalized. In the study, all the image data is divided by 255 to ensure that they are in the range [0,1].

The operations are carried out using the TensorFlow data pipeline which allows for batch operations to be performed.

## IV. Methodology

Fig.1. depicts the Zero-DCE framework. Given an input image, a Deep Curve Estimation Network (DCE-Net) is used to estimate a series of best-fitting Light-Enhancement curves (LE-curves). The framework then successively applies the curves to all pixels in the input's RGB channels, resulting in the final enhanced image. In the next sections, we go over the important components of Zero-DCE, such as the LE-curve, DCE-Net, and non-reference loss functions.

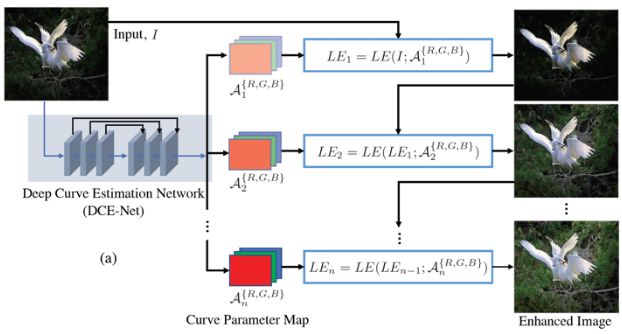


*Fig. 1. Zero-DCE Architecture*

### A. Light Enhancement Curve (LE curve)

We seek to build a type of curve that can automatically translate a low-light image to its enhanced version, where the self-adaptive curve parameters are purely reliant on the input image, based on the curve’s adjustment used in photo editing software. In the creation of such a curve, there are three goals:

- To avoid information loss due to overflow truncation, each pixel value in the enhanced image should be in the normalised range of [0,1].
- This curve should be monotonous to preserve the differences (contrast) of neighboring pixels.

- The form of this curve should be as simple as possible and differentiable in the gradient backpropagation process.

To satisfy these three goals, we create a quadratic curve that looks like this:

$$LE(I(\mathbf{x});\alpha) = I(\mathbf{x}) + \alpha I(\mathbf{x})(1 - I(\mathbf{x})),$$

*Equation 1*

where x denotes pixel coordinates, $LE(I(\mathbf{x});\alpha)$ is the enhanced version of the given input $I(\mathbf{x}), \alpha \in [-1,1]$ is the trainable curve parameter, which adjusts the magnitude of LE-curve and also controls the exposure level. Each pixel is normalized to [0; 1] and all operations are pixel-wise.

### B. Higher-Order Curve

The LE-curve defined in Eq. (1) can be applied iteratively to enable more versatile adjustment to cope with challenging low-light conditions. Specifically,

$$LE_n(\mathbf{x}) = LE_{n-1}(\mathbf{x}) + \alpha_n LE_{n-1}(\mathbf{x})(1 - LE_{n-1}(\mathbf{x})),$$

*Equation 2*

where $n$ is the number of iterations, which controls the curvature. In this paper, we set the value of $n$ to 8, which can deal with most cases satisfactory. Eq. (2) can be degraded to Eq. (1) when $n$ is equal to 1.

### C. Pixel-Wise Curve

A higher-order curve can adjust an image within a wider dynamic range. Nonetheless, it is still a global adjustment since is used for all pixels. Global mapping tends to over-under- enhance local regions. To address this problem, we formulate as a pixel-wise parameter, i.e., each pixel of the given input image has a corresponding curve with the best fitting to adjust its dynamic range. Hence, Eq. (2) can be reformulated as:

$$LE_n(\mathbf{x}) = LE_{n-1}(\mathbf{x}) + \mathcal{A}_n(\mathbf{x}) LE_{n-1}(\mathbf{x})(1 - LE_{n-1}(\mathbf{x})),$$

*Equation 3*

where $A$ is a parameter map with the same size as the given image. Here, we assume that pixels in a local region have the same intensity (also the same adjustment curves), and thus the neighboring pixels in the output result still preserve the monotonous relations. In this way, the pixel-wise higher-order curves also comply with three objectives.

### D. DCE-Net

We present a Deep Curve Estimation Network to learn the mapping between an input image and its best-fitting curve parameter maps (DCE-Net). The DCE-Net takes a low-light image as input and produces a series of pixel-wise curve parameter mappings for higher order curves as output.

A simple CNN with seven convolutional layers and symmetrical concatenation is used. The ReLU activation function is followed by 32 convolutional kernels of size 3*3 and stride 1 in each layer. The down sampling and batch normalization layers, which destroy the relationships between neighboring pixels, are removed. The Tanh activation function follows the last convolutional layer, producing 24 parameter maps for 8 iterations ($n$ = 8), with three curve parameter maps for each of the three channels required for each iteration. The supplementary material contains the DCE-Net architecture in depth.

We present a set of differentiable non-reference losses that allow us to evaluate the quality of improved images in order to enable zero-reference learning in DCE-Net. To train our DCE-Net, we use the four categories of losses listed below.

### E. Loss of Spatial Consistency

The spatial consistency loss $L_{spa}$ promotes enhanced image spatial coherence by retaining the difference between contiguous regions in the input image and its enhanced version:

$$L_{spa} = \frac{1}{K}\sum_{i=1}^{K}\sum_{j\in\Omega(i)}(|(Y_i - Y_j)| - |(I_i - I_j)|)^2,$$

*Equation 4*

where $K$ is the number of local regions and, $\Omega(i)$ is the four neighboring regions (top, down, left, right) centered at the region $i$. We denote $Y$ and $I$ as the average intensity value of the local region in the enhanced version and input image, respectively.

### F. Exposure Control Loss

We build an exposure control loss $L_{exp}$ to manage the exposure level to restrain under-/over-exposed zones. The exposure control loss is the difference between a local region's average intensity value and the well-exposedness level $E$. The loss $L_{exp}$ can be expressed as:

$$L_{exp} = \frac{1}{M}\sum\nolimits_{k=1}^{M}|Y_k - E|,$$

*Equation 5*

Where $M$ represents the number of nonoverlapping local regions of size 16 by 16, $Y$ is the average intensity value of a local region in the enhanced image.

### G. Color Constancy Loss

We design a color constancy loss to correct the potential

$$L_{col} = \sum\nolimits_{\forall(p,q)\in\varepsilon}(J^p - J^q)^2, \varepsilon = \{(R,G),(R,B),(G,B)\}.$$

*Equation 6*

color deviations in the enhanced image and build the relations among the three adjusted channels. The color constancy loss $L_{col}$ can be expressed as:

where $Jp$ denotes the average intensity value of $p$ channel in the enhanced image,$(p,q)$ represents a pair of channels.

### H. Illumination Smoothness Loss

To preserve the monotonicity relations between neighboring pixels, we add an illumination smoothness loss to each curve parameter map $A$. The illumination smoothness loss $L_{tvA}$ is defined as:

$$L_{tv_{\mathcal{A}}} = \frac{1}{N}\sum_{n=1}^{N}\sum_{c\in\xi}(|\nabla_x\mathcal{A}_n^c| + \nabla_y\mathcal{A}_n^c|)^2, \xi = \{R,G,B\},$$

*Equation 7*

where $N$ is the number of iterations, $\nabla_x$ and $\nabla_y$ represent the horizontal and vertical gradient operations, respectively.

*I. Total Loss*

The total loss can be expressed as:

$$L_{total} = L_{spa} + L_{exp} + W_{col}L_{col} + W_{tv_A}L_{tv_A},$$

*Equation 8*

where $W_{col}$ and $W_{tvA}$ are the weights of the losses.

## V. Technology Setup

The initial experiment was conducted in Google Colab with a smaller number of epochs, while the finalized trial, with many epochs, storing the model and creating endpoints for prediction were handled in the Google Cloud Platform. A training and a prediction pipeline were created for this study. The overall technical architecture of the system is as shown in Fig . 2.

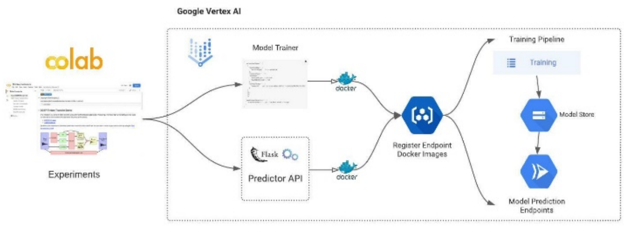


*Fig. 2 MLOps Training and Prediction*

### A. Training pipeline

- The training algorithm for the model and any preprocessing steps are transformed into reusable python modules and the different processes are linked into a single python module.
- The python module was embedded in a Docker image using a pre-existing GCP compliant template. The required library installations, codes etc. were all coded into the docker image.
- The docker image is thus created, it is registered in Google Container Registry, a custom service that allows us to store and spin docker images.
- The docker image thus created is used to schedule a model training workflow in Vertex AI. If the workflow is successful, the generated model is stored in the specified storage bucket.

### B. Prediction Pipeline

- The Endpoints service in Vertex AI can be directly linked to the training pipeline to serve the models stored in the storage bucket directly. But the model in the study required customized data pre-preprocessing before the model could consume it. To alleviate this problem, a custom API layer, compliant with Google Endpoint Specs was created using Flask, a python-based web framework.

- The Flask application along with the pre-requisite data was containerized and registered in Google Container Registry. The containerized application was then used to create a custom prediction endpoint.

## VI. Experimentation

The goal of DCE-Net is to estimate a set of best-fitting light-enhancement curves (LE-curves) given an input image. The framework then maps all pixels of the input's RGB channels by applying the curves iteratively to obtain the final enhanced image. To bring the capability of wide dynamic range adjustment into full play, we incorporate both low-light and over-exposed images into our training set.

The experiment involves the training and evaluation of pre-processed LOL dataset using DCE network, convolutional neural network.

- After hyper parameter tuning, neural network of 8 convolutional layers was considered for prediction.
- The first five convolutional layers have ReLU activation function, and the last convolutional layer has tanh.
- To enable zero-reference learning in DCE-Net, we use a set of differentiable zero-reference losses that allow us to evaluate the quality of enhanced images.
- The color constancy loss is used to correct the potential color deviations in the enhanced image.
- Exposure Loss measures the distance between the average intensity value of a local region and a preset well-exposedness level (set to `0.6`).
- To preserve the monotonicity relations between neighboring pixels, the illumination smoothness loss is added to each curve parameter map.
- The spatial consistency loss encourages spatial coherence of the enhanced image by preserving the contrast between neighboring regions across the input image and its enhanced version.

The dataset was divided into training and validation datasets each undergoing the requisite preprocessing.

The training set is fed to the model built above for around 100 epochs. Later the model was evaluated using validation dataset. The model generated through training can be seen in the Fig.3.

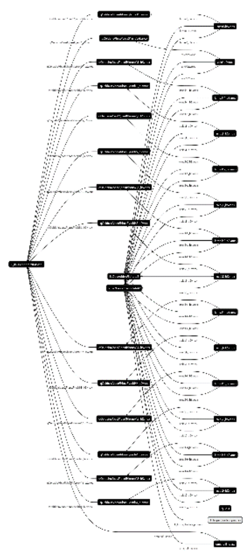

*Fig. 3. Zero-DCE Model*

We can see the evaluation results in next section.

## VII. Evaluation and Results

### A. Quantitative Perspective

The final observed losses after 100 epochs of training in the validation dataset are as shown in the Table 1.

| Loss | Value |
| --- | --- |
| Total | **1.3587** |
| Illumination Smoothness | **0.0239** |
| Spatial Consistency | **0.2618** |
| Color Constancy | **0.0429** |
| Exposure | **1.0301** |

*Table 1 Validation Loss*

The exposure, illumination smoothness and total loss were observed to decrease significantly with training. The color constancy loss plateaus after a while because the optimal contrast has been reached. But the contributing factor for the reduction of overall loss is the illumination smoothness and exposure loss which indicates that the model was able to address the blurring caused by low-light the most in the images.

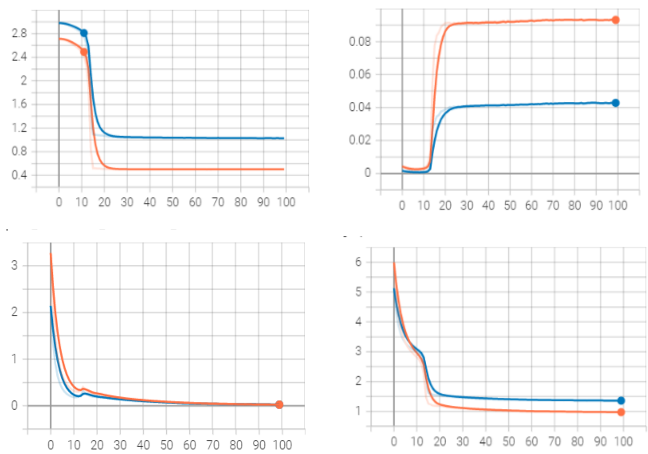


*Fig. 4. Exposure, Color Constancy, Illumination Smoothness, and total loss (Orange: Training; Blue: Validation)*

The validation loss curve indicates that the training weights capture the complexity of the data and can replicate the results on future data.

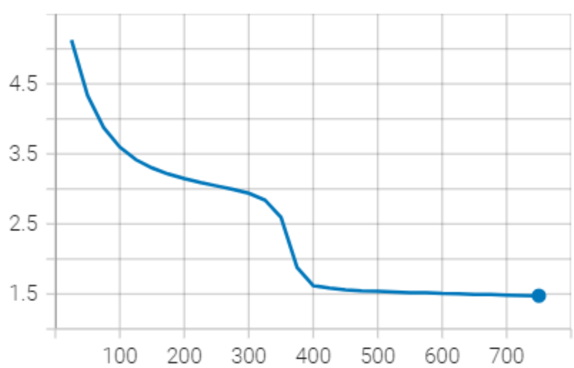


*Fig. 5. Validation Loss Curve*

### B. Visual Perspective

In the study, it was observed that Zero-DCE was able to reduce the losses significantly over hundred epochs providing very crisp images with a richer visual palette.

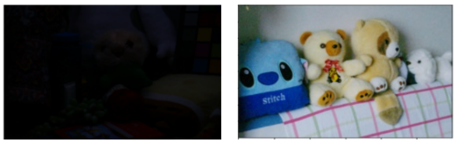

*Fig. 6. Original Vs Model Derived*

The auto contrast feature in Pillow which uses a histogram based contrasting algorithm was used as the baseline to showcase how the model performs. The most noticeable feature is the saturation of colors and improvement in brightness.


*Fig. 7. Original Vs Auto Contrast Vs Model Derived*

## VIII. Conclusion

For low-light image improvement, we proposed a deep network. It can be trained from start to finish with only zero reference photos. This is accomplished by recasting the low-light picture enhancement problem as an image-specific curve estimation problem and constructing a series of differentiable nonreference losses. Experiments show that our technology outperforms existing light enhancement techniques. In future work, we'll strive to incorporate semantic information to solve difficult scenarios and account for noise effects.


## Acknowledgment

The authors wish to thank Professor Vijay Eranti for the continued guidance and support. This project was part of academic project submitted to Department of Computer engineering, San Jose State University.


## References


[1] Ruixing Wang, Qing Zhang, Chi-Wing Fu, Xiaoyong Shen, Wei-Shi Zheng, and Jiaya Jia. Underexposed photo enhancement using deep illumination estimation.

[2] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In BMVC, 2018.

[3] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang

Wang. EnlightenGAN: Deep light enhancement without paired supervision.

[4] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycleconsistent adversarial networks.

[5] Dinu Coltuc, Philippe Bolon, and Jean-Marc Chassery. Exact histogram specification. IEEE Transactions on Image Processing, 15(5):1143–1152, 2006.

[6] Haidi Ibrahim and Nicholas Sia Pik Kong. Brightness preserving dynamic histogram equalization for image contrast enhancement. IEEE Transactions on Consumer Electronics, 53(4):1752–1758, 2007.

[7] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. IEEE Transactions on Image Processing, 22(12):5372–5384, 2013.

[8] J Alex Stark. Adaptive image contrast enhancement using generalizations of histogram equalization. IEEE Transactions on Image Processing, 9(5):889–896, 2000.

[9] Edwin H Land. The retinex theory of color vision. Scientific American, 237(6):108–128, 1977.

[10] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Transactions on Image Processing, 22(9):3538–3548, 2013.

[11] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In CVPR, 2016.

[12] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing, 26(2):982–993, 2017.

[13] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust retinex model. IEEE Transactions on Image Processing, 27(6):2828–2841, 2018.

[14] Lu Yuan and Jian Sun. Automatic exposure correction of consumer photographs. In ECCV, 2012.

[15] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognition, 61:650–662, 2017.

[16] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In BMVC, 20