Low-Light Image Enhancement for Autonomous Driving Systems using DriveRetinex-Net

Long Hoang Pham, Duong Nguyen-Ngoc Tran, and Jae Wook Jeon
Sungkyunkwan University, Department of Electrical and Computer Engineering
Suwon, South Korea
phlong@skku.edu, duongtran@skku.edu, jwjeon@yurim.skku.ar.kr

Abstract

Most autonomous driving algorithms are designed for normal-light images. Hence, insufficient lighting during image capture significantly degrades the visibility of images and hurts the performance of many computer vision systems. Retinex theory is an effective tool for enhancing the illumination and detail of images. In this paper, we collected a Low-Light Drive (LOL-Drive) dataset and applied a deep retinex neural network, named DriveRetinex, which was taught using this dataset. The deep Retinex-Net consists of two subnetworks: Decom-Net (decomposes a color image into a reflectance map and an illumination map) and Enhance-Net (enhances the light level in the illumination map). The whole architecture can be trained in an end-to-end fashion. Extensive experiments demonstrate that the proposed method not only achieves visually appealing low-light enhancement, but it also increases the accuracy of object detection in autonomous driving systems.

Keywords: low-light image enhancement, retinex theory, autonomous driving, image processing.

1. Introduction

Autonomous driving has been a trendy research topic in recent years. However, most existing algorithms developed for autonomous driving have been designed for normal-light images, i.e., daytime driving scenes. Hence, insufficient lighting in driving scenes at dusk or at night time significantly degrades the visibility of the images and hurts the performance of computer vision algorithms, including those used for object detection. The lack of sufficient lighting in the images can come from the limited capability of the capturing equipment or inappropriate configurations of the equipment. However, low-light environments are the main cause of the loss of details and low contrast in the images. Therefore, low-light image enhancement is important in autonomous driving research.

Another family of low-light enhancement methods is built based on the retinex theory [1]. The retinex theory states that a color image is composed of a light-independent reflectance map and a structure-aware smooth illumination map. In recent years, a handful of retinex-based methods have been proposed. Methods such as SSR [2] and MSRCR [3] were early attempts to smooth the illumination map using multi-scale Gaussian filters. SRIE [4] tried to estimate reflectance and illumination simultaneously using a weighted variational model. Meanwhile, LIME [5] estimated the illumination based on prior knowledge about the reflectance and used reflection as the final enhancement method. The authors in [6, 7] used the retinex theory to build joint low-light enhancement and noise removal. However, these methods still need hand-crafted constraints to decompose the reflectance and illumination. In reality, it is difficult to come up with robust constraints for various scenes. Hence, the performance of these methods relies on careful case-by-case parameter tuning.

In recent years, deep neural networks have been widely used in various low-level image processing tasks, including super-resolution, rain removal, and noise reduction. The authors in [8-11] proposed neural network retinex-based methods to decompose images into reflectance and illumination maps automatically, while also enhancing the illumination to obtain visually appealing, enhanced images. Hence, these methods have proven that deep neural networks are, by nature, good at learning the constraints for decomposition operations.

In this paper, a low-light image enhancement method, named DriveRetinex-Net, is presented by combining a deep neural network method and retinex theory. DriveRetinex-Net consists of two subnetworks: Decom-Net and Enhance-Net. The Decom-Net subnetwork is used to split the captured image into a reflectance map and an illumination map. Then, the Enhance-Net subnetwork is used to adjust the illumination map while maintaining consistency in large regions. DriveRetinex-Net learns both decomposition and enhancement tasks in a data-driven manner. To train the network, we used a dataset of low/normal-light image pairs from real driving videos. Extensive experiments demonstrate that the proposed method not only achieves visually appealing low-light enhancement, but it also increases the accuracy of object detection in autonomous driving systems.

The rest of the paper is organized as follows. Section 2 presents the dataset used to train the network. Section 3 discusses the architecture of the proposed DriveRetinex-Net method. Section 4 gives an exhaustive analysis of the experimental results, and Section 5 presents the conclusions of this study.

2. LOL-Drive Dataset

We collected a dataset, referred to as the Low-Light Drive (LoL-Drive) dataset, consisting of 10,000 low/normal-light image pairs. The images were captured from a variety of driving scenes, e.g., morning, afternoon, dusk, night, rain, and tunnel scenes. The distribution among the driving scenes can be seen in Fig. 2.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (2020R1A2C3011286).

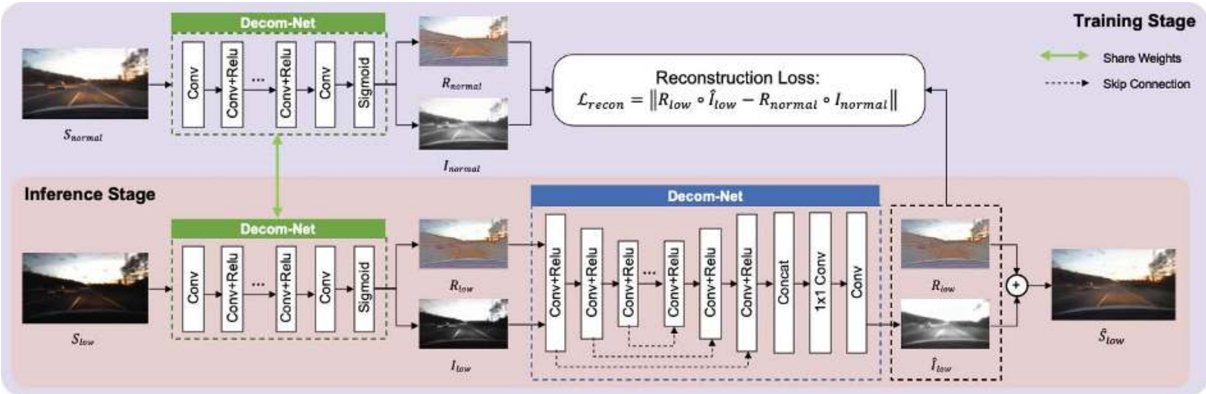


Figure 1: The architecture for the DriveRetinex-Net.

For morning and afternoon scenes, most low-light images are collected by decreasing the exposure time and ISO. On the other hand, normal-light images of darker scenes (dusk, night, rain, and tunnel scenes) are obtained by increasing the exposure time and ISO. Table 1 summarizes the values of the parameters used by Adobe Lightroom to get the low/normal-light image pairs. All images are then resized to a VGA resolution of 640×360 and converted into the PNG format. To the best of our knowledge, LoL-Drive is the first dataset containing large driving image pairs for low-light enhancement.

3. The Proposed Method

The retinex theory [1] assumes that observed color images can be decomposed into two components, i.e., reflectance and illumination maps, as denoted by

$$S = R \circ I \tag{1}$$

Here, S is the color image, R is the reflectance, and I is the illumination map. The reflectance map describes the details and features of the captured object, which and is considered to be consistent under any lightness conditions. The illumination map represents the various light levelsness on the objects. On With low-light images, it suffers from darkness and unbalances unbalanced illumination distribution. Motivated by the deep Retinex-Net method [4], which consists of two sub-networks: (i.e., Decom-Net and Enhance-Net), we propose an improved

Table 1: Parameters to generate low/normal-light image pairs.

Parameter	Value
Exposure	$-5 + 5F$
Highlights	$50\min\{Y, 0.5\} + 75$
Shadows	$-100\min\{Z, 0.5\}$
Vibrance	$-75 + 75F$
Whites	$16(5 - 5F)$

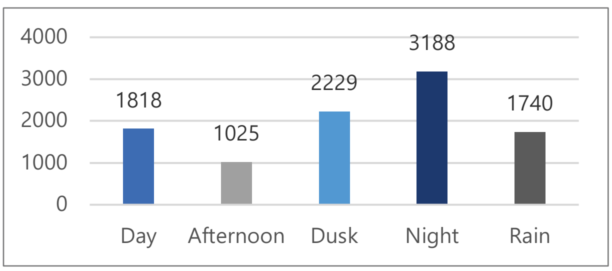
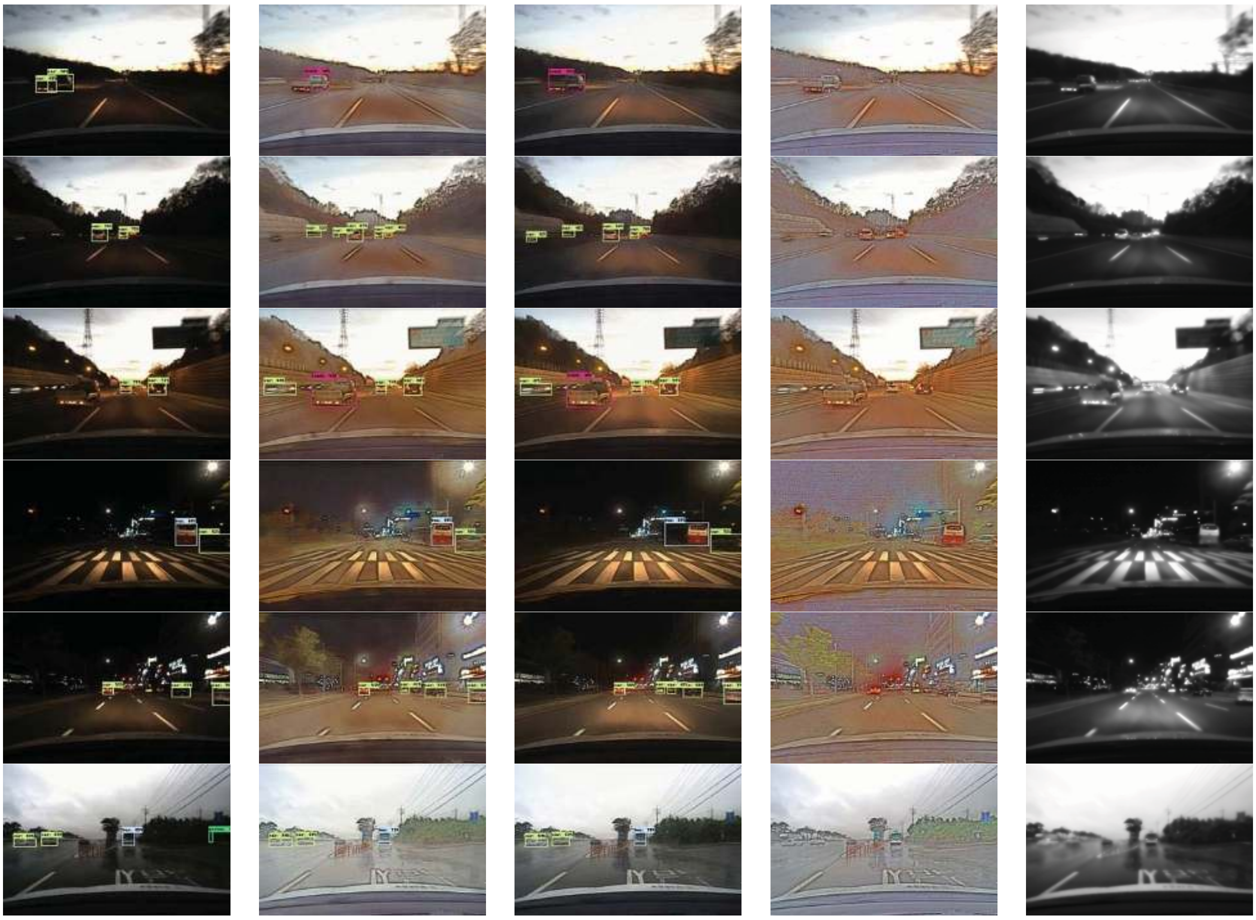


Figure 2: Image scene distribution in our dataset.

Retinex-Net to perform the reflectance/illumination decomposition and low-light enhancement.

The Decom-Net subnetwork learns to separate the color image into a consistent reflectance map between various illuminated images and the corresponding illumination map. Let denote S_{normal} as be the normal-light image and S_{low} as be the low-light image. In retinex theory, it assumes that the reflectance map is shared between S_{normal} and S_{low}. Hence, pairs of low/normal-light images are used as inputs at during the training stage, while only the low-light images are used in the testing stage. First, both S_{normal} and S_{low} are pushed through a 3×3 convolutional layer to extract features from the images. These features are then followed by several 3×3 convolutional layers with which have undergone a Rectified Linear Unit (ReLU) activation. In our experiments, a series of seven convolutional layers is used. A final 3×3 convolutional layer is used at the end; this with has a sigmoid function to project the reflectance maps (R_{normal}, R_{low}) and illumination maps (I_{normal}, I_{low}) into the range of $[0, 1]$.

The **Enhance-Net** subnetwork follows an encoder-decoder architecture. First, an encoder is used to successively down-sample I_{low} to a small scale so that the network can learn the illumination distribution in large regions. The down-sampling block consists of a convolutional layer with a stride of two and an ReLU. Second, a decoder is used to subsequently reconstruct the enhanced illumination map $\hat{I}_{low}$ with the large-scale illumination information. A multi-scale concatenation is applied to maintain the consistency of the global illumination and local illumination distribution in the up-sampling operation. The up-sampling block consists of the resized convolutional layer with ReLU activation. Skip connections are introduced from a down-sampling block to the corresponding up-sampling block by element-wise summation; this forces the network to learn residuals. The encoder-decoder architecture includes M pairs of down-sampling and up-sampling blocks. From our experiments, $M = 5$ is selected to provide good enhancement results while maintaining good processing speed. Finally, all the reconstructed channels are concatenated, and the result is pushed through a 1×1 convolutional layer and a 3×3 convolutional layer to reconstruct the enhanced illumination map $\hat{I}_{low}$. At the



(a) Input (b) Retinex-Net **(c) DriveRetinex-Net** (d) Reflectance map *R* (e) Illumination map *I*

Figure 3: The results of low-light image enhancement using the proposed DriveRetinex-Net (best view in color). (a) The low-light images as input. (b) Results from the Retinex-Net [8]. (c) Results from the proposed DriveRetinex-Net. EfficientDet object detection [13] results are also included in each image.

post-processing step, the $\hat{I}_{low}$ and R_{low} are combined to obtain the light-enhanced RGB image.

The loss function includes three terms: reconstruction loss $\mathcal{L}_{recon}$, invariable reflectance loss $\mathcal{L}_r$, and illumination smoothness loss $\mathcal{L}_s$:

$$\mathcal{L} = \mathcal{L}_{recon} + \lambda_r \mathcal{L}_r + \lambda_s \mathcal{L}_s \tag{2}$$

Here, λ_r and λ_s denote the coefficients used to balance the consistency of the reflectance and the smoothness of illumination maps. The reconstruction loss $\mathcal{L}_{recon}$ assumes that both R_{normal} and R_{low} can reconstruct the same image with the corresponding illumination map. Hence, for training Decom-Net, $\mathcal{L}_{recon}$ is defined as:

$$\mathcal{L}_{recon} = \sum_{i=low,normal} \sum_{j=low,normal} \lambda_{ij} \left\| R_i \circ I_j - S_j \right\|_1 \qquad \lambda_{ij} = \begin{cases} 0.001 & i \neq j \\ 1 & i = j \end{cases} \tag{3}$$

Alternatively, when training Enhance-Net, $\mathcal{L}_{recon}$ is expressed as:

$$\mathcal{L}_{recon} = \left\| R_{low} \circ \hat{I}_{low} - R_{normal} \circ I_{normal} \right\|_1 \tag{4}$$

The invariable reflectance loss $\mathcal{L}_r$ is used to constrain the consistency of reflectance maps:

$$\mathcal{L}_r = \| R_{low} - R_{normal} \|_1 \tag{5}$$

Additionally, the illumination smoothness loss $\mathcal{L}_r$ is defined as:

$$\mathcal{L}_s = \sum_{i=low,normal} \left\| \nabla I_i \circ \exp\left(-\lambda_g \nabla \mathrm{R_i}\right) \right\|_1 \tag{6}$$

Here, ∇ denotes the gradient including ∇_h (horizontal) and ∇_v (vertical), and λ_g denotes the coefficient balancing the strength of structure-awareness.

4. Experiments and Discussions

4.1 Implementation Details

Our LoL-Drive dataset and the LOL dataset [5] have a total of 11,500 image pairs. These are divided into 10,300 pairs for training, 600 pairs for validating, and 600 pairs for testing. The LOL dataset was used to introduce color variation to the training set. The DriveRetinex network is implemented using the TensorFlow 2.0 framework [9]. The Decom-Net subnetwork first takes a conv-layer without the activation layer, followed by seven convolutional layers with ReLU activation, and then ends with a conv-layer without the activation layer. The Enhance-Net subnetwork includes five down-sampling blocks and five up-sampling blocks. The Decom-Net and Enhance-Net subnetworks are first trained separately, and then the whole network is fine-tuned end-to-end using stochastic gradient descent (SGD) with back-propagation.

Table 2: Comparisons of PSNR and SSIM between Retinex-Net [8] and our proposed DriveRetinex-Net

	PSNR				SSIM			
	Afternoon	Dusk	Night	Rain	Afternoon	Dusk	Night	Rain
Retinex-Net [8]	57.2636	58.9709	59.1189	57.8193	0.4523	0.5212	0.2848	0.5753
DriveRetinex-Net	**64.2224**	**64.9454**	**67.9274**	**64.2802**	**0.7050**	**0.6887**	**0.5940**	**0.7562**

The batch size is set to be 32, and the Adam optimizer with a learning rate of 0.001 is used. The values for $\lambda_r, \lambda_s,$ and λ_g are set to 0.001, 0.1, and 10, respectively. The network was trained for 100 epochs on a computer with an NVIDIA Quadro RTX 8000. The whole network is light-weighted and can achieve, on average, 30 fps during the inference stage.

4.2 Evaluations

Fig. 3 presents the low-light image enhancement results from the test set of our LoL-Drive dataset. We also compare our results with the ones from Retinex-Net [8]. Our results clearly show better, visually pleasing, enhanced images. In the fourth and fifth columns of Fig. 3, the extracted reflectance and illumination maps are presented. The proposed method brightens the image while maintaining a consistent illumination distribution in local regions in the image. We can observe that the DriveRetinex-Net method can extract the objects' structure and detail well, even when they are hardly visible in the captured images.

On the other hand, the results of [8] are over-exposed. Hence, the network could not learn the illumination distribution well in outdoor scenes. This could be explained by the fact that the network only uses a few convolutional layers since it is designed to work with mostly indoor images.

The EfficientDet [13] object detection model is used to quickly evaluate the effectiveness of an enhanced image in vehicle detection operations. We can see that the bounding boxes in the enhanced images are superior to those of the original images and the results of Retinex-Net [8]. This demonstrates that the proposed method not only achieves visually appealing, low-light enhancement, but it also increases the accuracy of object detection in autonomous driving systems.

To quantitatively evaluate the effectiveness of the DriveRetinex-Net method, PSNR and SSIM metrics [14] are used. When the original image o is fixed, the higher the PSNR value the better. If $PSRN(o, d)$ is larger than $PSRN(o, n)$, it can be stated that the quality improvement of image d is better than that of image n. The resultant $SSIM$ index is a decimal value between -1 and 1, and a value of 1 is only reachable in the case of two identical sets of data. A value of 1 indicates perfect structural similarity, while a value of 0 indicates no structural similarity. The calculated results are shown in Table 2 and the visual comparisons can be found in Figs. 04. These results clearly show that the results from our proposed DriveRetinex-Net method outperform the state-of-the-art Retinex-Net method [8].

5. Conclusion

This paper demonstrated the effectiveness of a deep Retinex-Net model for enhancing the low-light driving images. The model can be trained in an end-to-end fashion and can learn to decompose the raw image into a reflectance and illumination map. Then the Enhance-Net subsequently increases the lightness in the illumination map. The final image is obtained by combining the newly enhanced illumination map and denoised reflectance map. Experimental results have shown that the proposed methods can produce visually good representation images for low-light scenes as well as improve the vehicle detection results. Future work can be performed to incorporate the Retinex-Net as a preprocessing branch to popular object detection deep learning models such as EfficientDet.

References

[1] E. H. Land, "The retinex theory of color vision," Scientific American, 237(6):108, 1997.

[2] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," IEEE Transactions on Image Processing, vol. 6, pp. 451-462, 1997.

[3] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multi-scale retinex for bridging the gap between color images and the human observation of scenes," IEEE Transactions on Image Processing, vol. 6, pp. 965-76, 1997.

[4] X. Fu, D. Zeng, Y. Huang, X. P. Xhang, X. Ding, "A fusion-based enhancing method for weakly illuminated images," Signal Processing, vol. 129, pp. 82-96, 2016.

[5] X. Guo, Y. Li, and H. Ling, "LIME: Low-light image enhancement via illumination map estimation," IEEE Transactions on Image Processing, vol. 26, no. 2, pp. 982–993, 2017.

[6] M. Li, J. Liu, W. Yang, and Z. Guo, "Joint denoising and enhancement for low-light images via retinex model," In International Forum on Digital TV and Wireless Multimedia Communications, pp. 91-99, 2017.

[7] M. Li. J. Liu. W. Yang, and Z. Guo, "Structure-revealing low-light image enhancement via robust retinex model," IEEE. Transactions on Image Processing, vol. 26, pp. 2828-2841, 2018.

[8] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," British Machine Vision Conference, 2018.

[9] W. Wang, C. Wei, W. Yang, and J. Liu, "GLADNet: Low-light enhancement network with global awareness," 13th IEEE International Conference on Automatic Face & Gesture Recognition, pp. 751-755, 2018.

[10] Y. Shi, X. Wu, and M. Zhu, "Low-light image enhancement algorithm based on retinex and generative adversarial network," arXiv, pp. 1–9, 2019.

[11] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo, "Structure-revealing low-light image enhancement via robust retinex

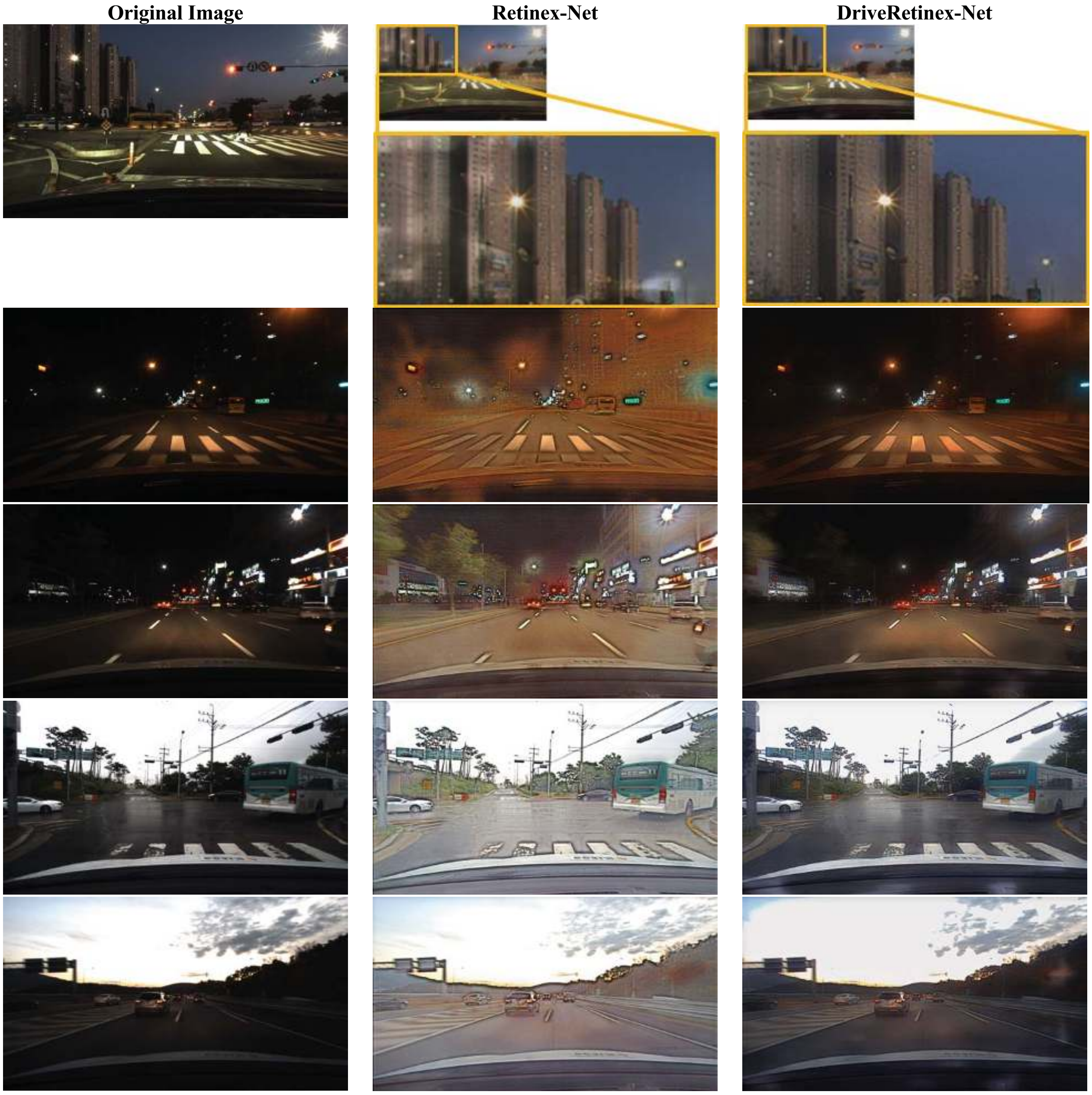


Figure 4: Comparison the results of the proposed DriveRetinex-Net and the Retinex-Net [8].
1st row, the proposed DriveRetinex-Net cam avoid the halo light errors since it learns the illumination distribution at multicale.
2nd – 5th rows, the RetinexNet wrongly enhance the illumination in the image, while the proposed DriveRetinex-Net can balance both local and global illumination enhancement.

model," IEEE Transactions on Image Processing, vol. 27, no. 6, pp. 2828–2841, 2018.

[12] TensorFlow 2.0. [Online]. Available: www.tensorflow.org

[13] M. Tan, R. Pang, Q. V. Le, "EfficientDet: Scalable and Efficient Object Detection," arXiv, 2019.

[14] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," Electronic Letter, vol. 44, no. 13, pp. 1–2, 2008.