GAN architecture leveraging a Retinex model with colored illumination for low-light image restoration

Arthur Lecert, Aline Roumy, Renaud Fraisse, Christine Guillemot

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2023.0322000

GAN architecture leveraging a Retinex model with colored illumination for low-light image restoration

ARTHUR LECERT[1], Aline Roumy[1], Renaud Fraisse[2], Christine Guillemot[1] (Fellow, IEEE)
[1]Inria Rennes – Bretagne-Atlantique 263 Avenue Général Leclerc, 35042 Rennes Cedex, France (e-mail: firstname.lastname@inria.fr).
[2]Airbus Defence and Space, Toulouse, France

Corresponding author: Arthur Lecert (e-mail: arthur.lecert@inria.fr).

ABSTRACT In this work, we study the restoration of low-light images with outdoor scenes without ground truth. Until now, approaches in the literature have avoided using the Retinex decomposition model in an unsupervised way or have added constraining priors on the searched components. We propose here to relax the constraint of a grayscale illumination of the Retinex model. Indeed, according to the physics of light, it should include a colored illumination. Resulting from this new decomposition model, we formulate a new deep learning-based architecture inspired by the style transfer methods. Our method enables us to visualize the illumination (i.e. a complex style with the same dimensions as an image) and the reflectance (i.e. the content). It achieves more visually pleasing components compared to the state-of-the-art i.e. without artifact, without noise amplification and without hallucination with a simple restoration for each of the components.

INDEX TERMS Low light enhancement, Image decomposition, Image restoration, Inverse problems, Retinex model, Neural Networks

I. INTRODUCTION

MANY technological fields, such as self-driving vehicles, would benefit from more efficient algorithms for the restoration of outdoor nighttime images. To do so, three characteristics have to be considered. The first two are due to the image itself. Indeed, (i) the image is captured during the night, therefore the image is very dark with a low intensity signal, the noise level is high, but more importantly the light sources in the scene are artificial and thus mostly colored and not white. Second, (ii) the image contains an outdoor scene. Therefore, the image contains details of various types, with also a large range of depth of field. As for the last characteristic, it is related to the restoration task itself, which must (iii) preserve the integrity of the image. It is of great importance to avoid adding fake details for night vision tasks.

Due to the image characteristics (i) and (ii), the restoration method must be unsupervised. Indeed, obtaining night/day image pairs of the same outdoor scene is difficult and, existing datasets do not contain true day/night image pairs but rather estimated ones. For instance, in the database FiveK [8], the normal-light image is made by human experts. Other datasets consider pairs with modified exposure parameter (e.g. ISO) of the scene, which is different from a night/day illumination (e.g. NPE [48], LIME dataset which is called HDR by the original authors [45], VV [2], MEF [33], LOL [52], DICM which is a mix of images from USC-SIPI [51] and a True Color Kodak images database [1] or the multi exposure dataset from [3]). Therefore, supervised methods based on these datasets, such as KinD++ [64], learn to compensate the exposure parameter (e.g. ISO) change during the capture, but not the lighting change (e.g. colored streetlamps during the night versus white sun lighting at daytime). By contrast, we choose to work with the Waymo dataset [46]. This dataset does not contain any night/day image pairs of the same scene but provides true day and true night scenes with similar contents.

In a first set of approaches, the restoration of low-light images, either does not preserve the integrity of the image (requirement (iii)), or significantly increases the noise level. Indeed, Jiang *et al.* restore night images using adversarial generative networks in their EnlightenGAN approach [25]. The restored image is thus a pleasant and plausible image. However, elements, which were not present in the original image may be introduced. CLEGAN [53] use a similar method adding a regularization to maximize mutual information between low and normal-light images but still hallucinate details

in the output. This prior is even more powerful since satellite images have strong self-similarity. The authors train a neural network on a different type of dataset. The datasets they use contain mainly underexposed images changing exposure parameters of the camera sensor and not nighttimes images. Therefore, the required restoration is different from the one we seek with our method. For instance, they do not deal with the noise introduced in low-light conditions.

In another set of approaches, the Zero-DCE method [31] and the work of Wang *et al.* [49] consists of tone mapping functions. The former is based on a deep neural network while the latter is not. The latter doesn't need a dataset to train and has low complexity. In these two methods, no hallucination is added but the noise is increased. They do not deal with the noise in these images which is not ideal in our case.

To overcome the previous issues, Horn's interpretation of the Retinex theory [16] has been used to restore low-light images. This theory states that an image is a product of two components: illumination and reflectance, where the illumination contains the lighting dependent information and the reflectance the true color of the objects in the scene. Retinex-based low light image restoration methods therefore perform a decomposition of the low-light image into two components. Then, the restored image is either considered as the reflectance, or as the product between a gamma-corrected illumination and the estimated reflectance. These Retinex-based approaches have two drawbacks which come from the underdetermined nature of the separation problem, where the number of observations is smaller than the number of unknowns.

The first drawback observed in LIME [13], RetinexGAN [34], and RetinexDIP [30] is that these methods propose to explore solutions around an initialization of the illumination component as the maximum of the image over the color channels, and originally proposed in [13]. However, the method only explores the neighborhood of its solution. A second drawback, is that in all the Retinex-based methods, the prior of a grayscale illumination is used in order to reduce the number of degrees of freedom and facilitate the search for solutions to this inverse problem. In this paper, we propose to relax this constraint by defining a decomposition model based on the physics of light thanks to the Retinex model with a colored illumination. The extraction of the components then becomes an even more difficult problem and to solve it, we propose firstly the idea of extracting common information thanks to the physics of light, and define a GAN-based architecture allowing, *in fine*, a restoration of the images with low illumination in an unsupervised way.

Indeed, according to the physical definition of reflectance [23], this property of a material represents the fraction of the radiance reflected by a surface over the radiance received by this surface. However, a consequence of this definition, is that the value of the reflectance of a material varies with respect to the wavelength of the incident light. In other words, in nighttime images, where the lighting is colored, the estimated reflectance does not contain all the true colors of an object but only part of it, and the illumination is colored. Therefore, we propose a novel Retinex model that takes into account low-light characteristics, and decompose an image into the product of the reflectance and illumination, with two main differences. First, the reflectance is *the common information between daytime and nighttime image domains after a specific correction is applied*. Indeed, the low-light reflectance that we can estimate is only a portion of the whole estimated normal-light reflectance. Thus, the low-light reflectance is the common information but it is the best estimate of the reflectance we can possibly extract from the two domains because of this degradation. By contrast, previous contributions in low-light restoration assumed that the reflectance was equal under daytime and nighttime lighting. Second, the illumination is colored, whereas previous contributions considered grayscale illumination.

Another set of methods try to restore low-light images with additional metadata present in the RAW version of these images. For instance, in [19], the authors propose an approach to unprocess the images, correct them in the RAW image space and simulate back the image processing pipeline. In our case, we do not have access to the RAW metadata of the images. Thus we cannot use this information for the training phase of our algorithm. An interesting fact, however, is that the authors conclude that the linearity with respect to the irradiance on the sensor is critical to restore low-light images. We reach the same conclusion with the Retinex theory in Section II-B. To better estimate the Retinex components, reversing the nonlinear camera operations is important.

Our problem shares similarities with the source separation problem as it can be seen as an instantaneous mixture. However, our case differs from the methods referenced by the reviewer since the number of unknowns in the outputs is greater than the number of input variables. For instance, in the works of Yao *et al.* [56]–[58], the hyperspectral images have between a hundred to four hundred bands or channels for each sample. Moreover, the goal of unmixing is to decompose a spectrum into a collection of spectral signatures of pure materials (i.e. endmembers or classes) and their fractional abundances (i.e. abundance maps). This discrete set of classes can contain from three to twelve elements in the datasets shown in the papers which is far less than the dimension of the input samples. Therefore, this problem is overdetermined whereas in our case we seek to find the values of six variables out of a three-dimensional input and thus it is underdetermined.

In summary, the main contributions of our paper are as follows:

- We formulate different improvements to the original Retinex decomposition model thanks to a colored illumination and define new appropriate priors for the components. The first one is designed to avoid the scale ambiguity problem of the decomposition and the second one deals with the problem of a saturated sensor and its effect on the resulting components.

- We also propose a new architecture with deep neural networks inspired by state-of-the-art source separation and style transfer methods trained in an unsupervised fashion taking a single standard RGB image as input. This deep neural network has two branches, one for each of the component and outputs two colored images: the RGB illumination and the reflectance. It is trained with additional loss terms corresponding to physical priors such as the reflectance being the degraded common information between the night and daylight image distributions.
- We demonstrate the efficiency of our method compared to the competitors in the literature on a real world dataset without any ground-truth [46]. We then show the first visualization of the Retinex components following the physics of light as well as the original Horn's model [16] while only coarse approximations can be found in the literature.

II. IMPROVEMENTS TO THE RETINEX MODEL

A. BACKGROUND ON THE ORIGINAL RETINEX MODEL

The study of the human visual system which lead to the Retinex theory goes back to the fundamental work by Land *et al.* [28] quickly followed by Barrow *et al.* [4]. Through its history, this theory had diverse interpretations based on path, center/surround approaches or physics of light. We refer the reader to [41] for a more detailed review. However, the Retinex image decomposition model commonly found in the literature nowadays was first defined by Horn in [16] for one-channel grayscale images. In the context of low-light image restoration [13], [30], [52], [64], this model is extended to RGB images $I \in \mathbb{R}^{3n}$ as follows:

$$I = L. * R + \eta. \tag{1}$$

where $.*$ is the element-wise product, $L \in \mathbb{R}^{n}$ the light-dependent component known as the illumination map, $R \in \mathbb{R}^{3n}$ the complementary component named the reflectance, and η an additive Gaussian noise. Thus, the illumination is considered a grayscale image scaling the reflectance with a common factor for the different color channels. The reflectance is assumed to be Lambertian (i.e. the surface at every point in the scene is diffusely reflecting light rays). The incident angle of the irradiance can be ignored and the Bidirectional reflectance distribution function (BRDF) [37] is not used. Besides, any specular or ambient component is also neglected since these methods only take a single image as input which invalidate the use of models such as Phong's model of the reflection of the illumination on a surface [38].

B. A COLORED ILLUMINATION

Since the Retinex decomposition model is only valid if applied on the irradiance of the camera sensor, it cannot be used directly on a standard RGB image. The intensity of the image needs to be linear with respect to this irradiance. Thus, the non-linear camera operations (i.e. mainly gamma corrections or tone mappings) need to be reversed. Complex pipelines can be used to achieve this goal whether by estimating the camera response model [24], [26], [59] or by reversing each step of the image processing pipeline [7], [18], [32]. We assume in this paper that we can reverse the image processing pipeline by only inverting the gamma correction ($\gamma = 2.2$).

The reflectance is officially defined by [23] as the fraction of the radiance reflected by a surface over the radiance received by that surface. In the computer graphics community (e.g. [6], [9], [61]), the illumination is assumed to be a colored component. Since the spectral reflectance curves of the material present in the scene depends on the wavelength, this property is needed to accurately simulate the reflection of light rays. If the light sources in the scene are colored and not "white", the camera sensor only receives partial information about the whole spectral reflectance curve. Thus, the reflectance cannot be considered as the true color of the scene in this case. It is only a ratio map over the three bands, RGB, in the visible spectrum. Therefore, it can be counter-intuitive and not look like a realistic image. Different authors tried to reconstruct the spectral reflectance curve of the scene in a discrete fashion [5], [11], [55], [65] or in a continuous one [54]. However, identifying each material in low-light images is extremely challenging and ill-posed in practice without using strong priors on the diversity of the present elements because of the metamerism effect (i.e. one RGB color can be the result of different combinations of wavelength). Indeed, in a night image, all colors result from artificial lights (e.g. streetlamps, car headlights, ...) reflecting on the different objects in the scene and then going straight through the camera sensor. Since these artificial lights are colored and the reflectance spectra of the objects in the scene are highly non-linear, we only observe a tiny portion if not none of the "true" colors (i.e. the color under a white light) of the scene. We introduce a different definition of the Retinex decomposition to address these challenges. In the literature, indoor datasets such as LOL [52] don't fully represent the complexity of the degradation in outdoor images. Reducing the exposure to simulate a low-light image is too simplistic to capture the whole shift of the distribution. In this paper, we also do not consider the multiple scattering of light rays or the attenuation of the fog during the night in the outdoor scene to simplify the model as opposed to [36] for instance.

Instead, we define the reflectance as the corrected common information between two distributions of images (which are assumed to have similar scenes and objects but not paired images). Indeed, the reflectance extracted from low-light images is degraded and thus not equal to the one estimated from normal-light images. In that sense, the reflectance is really independent from the domain while the illumination is the light-dependent component and contains, for instance, the specific noise of low-light images.

With these definitions, we redefine eq. (1) as eq. (2).

$$I^{\gamma} = \left(\frac{L}{\alpha} + \eta\right). * \alpha R \tag{2}$$

where $I \in [0,1]^{3n}$ the RGB image, $L \in [0,1]^{3n}$ the il-

lumination, $\alpha \in \mathbb{R}^*$ a scaling factor as the decomposition leads to an infinity of solutions. $R \in [\varepsilon, 1-\varepsilon]^{3n}$ as the only object which absorbs all light is a black hole. On the contrary, perfect mirrors are still not widely commercialized and may not appear frequently in our everyday lives. $\varepsilon = 1e^{-8}$ in our experiments. Thus, we relax the original previous model of the grayscale constraint of the illumination. The component now has a local chrominance in addition to a local luminance value. In Section II-C, we quickly define an additional prior to reduce the solution set with the scaling factor α and extend a previously published prior in [30] to address a colored illumination. Since the illumination is supposed to contain the low-light noise and degradations, adding an illumination smoothness prior (e.g. the one in [52], [64] would be inefficient).

C. NEW PRIORS

1) Scale ambiguity (High reflectance prior)

The α factor introduced in the model (2) highlights the scale ambiguity problem in the Retinex decomposition. Any positive real value can lead to a plausible solution. To reduce even further the solution set we propose a new prior defined as follows:

$$\mathcal{L}_{HR} = \left\| \frac{1}{\alpha} \right\|_1 \tag{3}$$

$\alpha = 1$ as an initial value before the optimization process. As we are working with low-light images, we assume that the illumination should have the lowest possible value. On the other hand, minimizing this prior is equivalent to seeking for the highest reflectance. Intuitively, it can be seen as considering a high V-channel (HSV) for the reflectance, one with a low value for the illumination. This also means that the optimization process is biased against black bodies.

2) Exposure prior (RGB version)

In this section, we extend the exposure prior [30] to RGB images. As long as the camera sensor is not saturated by the light ray (i.e. $I_{c\in\{R,G,B\}} \neq 1$), the illumination can not be saturated as well (i.e. $L_{c\in\{R,G,B\}} \neq 1$). This prior is defined to prevent the trivial solution where $L = 1$ and thus $I = R$. Only light sources or overexposed regions in the input image should lead to such values in the components. This prior is defined as

$$\mathcal{L}_E = \sum_{c\in\{R,G,B\}} \left\| g(\tilde{L}) - g(L_c) \right\|_2^2 \tag{4}$$

where g is a threshold function $g(x) = \begin{cases} x, & x > 1-\epsilon \\ 0, & \text{otherwise} \end{cases}$ and $\tilde{L} = \max_{c\in\{R,G,B\}} I_c$, an approximation of the illumination following the work LIME by Guo *et al.* [13].

III. THE ARCHITECTURE

A. ARCHITECTURE CHOICES

In this section we describe the architecture shown in Fig. 1 that we propose to decompose an image. Since we do not have access to the ground-truth images and want a low execution time, we use a GAN-based architecture. To better separate the input image into the two components, we make use of two discriminators, one to generate each of the component respectively.

The network is composed of two branches to extract each component. We build on the YTMT source separation strategy [17] which consists of alternating positive ReLu on one branch and negative ReLu on the other to avoid losing information and to better connect the two networks together. We use two UNets [43]. The illumination branch receives as input an approximation of the illumination of the input image I.

In MUNIT [21], the authors managed to transfer the style of an image such that the resulting image belongs to another domain while preserving the content of the image. To perform that, they improve upon the work of [12] and the surprising result that instance normalization [47] followed a procedure to align the mean and variance of the content features with those of the style features [20]. By nature, this problem is similar to the Retinex decomposition problem if we consider the reflectance as the content we try to preserve and the illumination as a complex style (i.e. a whole RGB image instead of mean and variance parameters). Therefore, we add instance normalization modules to the reflectance extraction branch. The information of the style of the image flows through the illumination branch.

B. THE LOSS TERMS

1) The reconstruction loss

For ease of notation, we omit the α scaling factor in the following equations to compute the two estimated components $(\hat{L}, \hat{R})$.

$$\tilde{L} = \lambda * \max_{c\in\{R,G,B\}} I_c \tag{5}$$

$$G_R : I_d \in \mathbb{R}^{3n} \mapsto \hat{R}_d \in \mathbb{R}^{3n} \tag{6}$$

$$G_L : \tilde{L}_d \in \mathbb{R}^{3n} \mapsto \hat{L}_d \in \mathbb{R}^{3n} \tag{7}$$

where λ is the mean triplet RGB over the spatial dimensions of the input image, $d \in \{0, 1\}$ the label being equal to 1 for the normal-light domain and 0 for the low-light one.

To make sure that the two generated components we get can reconstruct the input image according to the Retinex model, we use the mean absolute error for the structure of the image and the angular error to ensure that the color is accurately recovered. This can be summed up as the following terms,

$$\mathcal{L}_{MAE} = \left\| I^\gamma - \hat{L}. * \hat{R} \right\|_1 \tag{8}$$

$$\mathcal{L}_{color} = \frac{\hat{L} \cdot \hat{R}}{\left\|\hat{L}\right\| \left\|\hat{R}\right\|} \tag{9}$$

We don't use the latent reconstruction terms like in [21] since applying the illumination of one image to the reflectance of another would result in an unrealistic and not plausible image and then would mislead the discriminators during the training process. The $\mathcal{L}^1$-norm guarantees that no information is lost

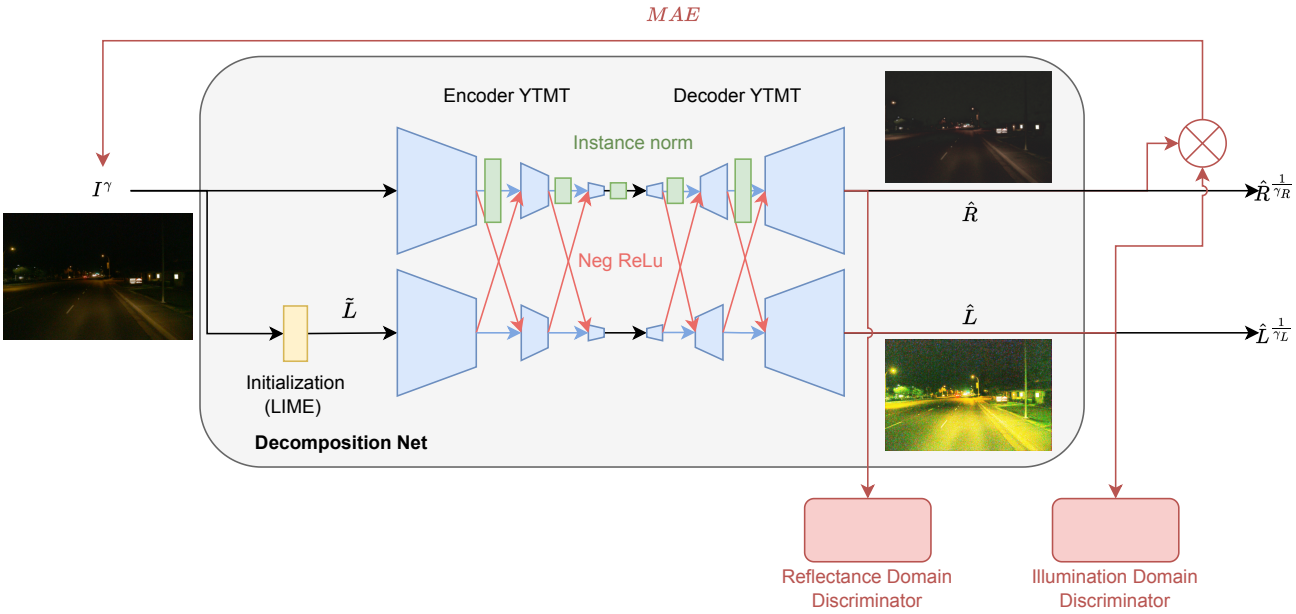

FIGURE 1. An illustration of architecture composed of two branches: the upper branch extracts the reflectance from the image normalizing the style of the features, the lower branch on the contrary keeps the style information to produce the illumination and takes as input an approximation of this component to facilitate the process. The two branches can swap information with the YTMT strategy [17] for source separation. Each component has its own discriminator to follow the definition of the Retinex components with common information. We reapply the gamma correction to display the images illustrating the variables.

during the decomposition process. However, we directly extract a noisy illumination instead of a noiseless version since it is easier to do so and then denoise the resulting component.

2) Domain discriminator adversarial functions

Our architecture relies on adversarial loss terms to find the components in an unsupervised fashion. We define the two discriminators function as follows,

$$D_R : \hat{R}_d \mapsto \hat{d} \tag{10}$$

$$D_L : \hat{L}_d \mapsto \hat{d} \tag{11}$$

$$\hat{d} \in \{0, 1\}.$$

where $\hat{d}$ is the estimated label resulting from each of the component. We use a multiscale discriminator architecture such as the one in [39]. The discriminators need to be able to identify the domain of the input component they get (i.e. separate each component according to their domain). We empirically find that the training of the generators is more stable with the Least Squares GAN [35] than the other versions. The parameters of (G_R, G_L) are fixed in this pass.

$$\mathcal{L}_{D_L} = \sum_{d\in\{0,1\}} \mathbb{E}_{L_d}\left[\left(D_L(G_L(L_d)) - d\right)^2\right] \tag{12}$$

$$\mathcal{L}_{D_R} = \sum_{d\in\{0,1\}} \mathbb{E}_{I_d}\left[\left(D_R(G_R(I_d)) - d\right)^2\right] \tag{13}$$

3) Generator adversarial functions

To train the generators (G_R, G_L), we fix the parameters of (D_R, D_L). The illumination generator should extract the component from the image and the domain should be accurately identified by the corresponding discriminator. On the contrary, we seek to extract information which cannot be classified by the reflectance discriminator between the low and normal-light domains. Therefore, we optimize it to align the low-light reflectance to the normal-light one. This leads to the following equations,

$$\mathcal{L}_{G_L} = \sum_{d\in\{0,1\}} \mathbb{E}_{L_d}\left[\left(D_L(G_L(L_d)) - d\right)^2\right] \tag{14}$$

$$\mathcal{L}_{G_R} = \sum_{d\in\{0,1\}} \mathbb{E}_{I_d}\left[\left(D_R(G_R(I_d)) - 1\right)^2\right] \tag{15}$$

4) The resulting optimization problem

As a result, we obtain the following problem to train the decomposition network,

$$\begin{aligned}\left(\hat{G}_R, \hat{G}_L\right) = \underset{G_R, G_L, \alpha}{\operatorname{argmin}}\ & \lambda_{MAE}\mathcal{L}_{MAE} + \lambda_{color}\mathcal{L}_{color} \\ & + \lambda_{HR}\mathcal{L}_{HR} + \lambda_E\mathcal{L}_E \\ & + \lambda_{adv}(\mathcal{L}_{G_L} + \mathcal{L}_{G_R})\end{aligned} \tag{16}$$

and for the Retinex components domain discriminators,

$$\left(\hat{D}_R, \hat{D}_L\right) = \underset{D_R, D_L}{\operatorname{argmin}}\ \mathcal{L}_{D_R} + \mathcal{L}_{D_L}. \tag{17}$$

C. VISUALIZATION & RESTORATION OF THE COMPONENTS

One of the key benefits of considering a complex style as the illumination (i.e. a style that has the same dimensions of

an image here) is that it can be visualized. Some examples of the obtained Retinex components are illustrated in Fig. 2. To the best of our knowledge, this is the first time that these components are linked to the style transfer literature and that the common and specific domain information are displayed. The reflectance is sharper than the input image. The dark areas present in it can be intuitively explained as loss or missing information about the scene. It shows that this component also needs a custom restoration. Besides, shadows and light rays coming from the car headlights and streetlamps still end up in the reflectance. However, the glare effect of the light sources are reduced and the colors are less saturated. On the other hand, the illumination contains the low-light noise and degradations. We seek to restore low-light images but not estimate a daylight version. Thus, we denoise the illumination component instead of the reflectance like in the previous works in the literature and avoid any tone mapping to avoid amplifying the noise. We use a weight map to denoise the component according to the structure of the scene in the reflectance in Fig. 3. Using the maximum of the reflectance over the color RGB channels gives more information than a simple gradient and guide the denoising network to strengthen the denoising process where the reflectance does not have a lot of information in the dark areas. We try to find the best compromise to keep the maximum of information in the image. Since the reflectance contains the textural details of the different objects of the scene, we seek to amplify this information to highlight and make it easier to distinguish the elements. The different works in the literature do it with a tone mapping function such as a gamma correction [13] or by the use of a neural network [31], [63]. Defining which type of function to apply here can be difficult without any ground-truth images or priors to control the exposure of the image (e.g. section 3.3 in [31]). Therefore, we decide to use the simple and efficient gamma correction with a low execution time. We empirically find the best gamma values by maximizing the LPC-SI metric [14]. We also found that we can get visually better results with a higher correction on the blue channel to balance the colored noise of the illumination. This effect is dataset-specific though and not mandatory in other cases. It may be due to the sensors that the authors used but we couldn't verify this hypothesis. We also tried a unique gamma for all the channels or in the HSV domain but the results were either too whitened by the process or the colors too saturated. An example of the restoration process of the illumination is shown in Fig. 3 and in Fig. 4 for the reflectance. Using a gamma correction on the reflectance does not reveal a hidden noise or another low-light degradation. This shows the high quality of the decomposition.

IV. RESULTS

A. METRICS & EVALUATION METHODOLOGY

As we are not seeking to approximate the distribution of daylight images with the restored images and we have no ground-truth, neither commonly used metrics like FID [15], IS [44] or CIS [22] nor classic reference-based metrics such as PNSR, SSIM [50] or LPIPS [62] can be used here. The LPC-SI metric [14] measures the sharpness of an image through local phase coherence of complex wavelet coefficients. Even though it cannot measure the whole low-light degradation, being sharp is one of the properties we desire for the result. We choose the best gamma values to apply for the gamma correction over the RGB color channels with a trade-off between the LPC-SI metric and the visual quality of the images.

As there are no ground-truth images available for our problem, we consider a two-steps process to evaluate the methods: First, we use a visual approach: we observe the level of noise and note if there are hallucinations in the outputs. Then, since there are no ground-truth in the datasets, we cannot use metrics with reference. Thus, we compare the methods using a non-visual test with the reference-free LPC-SI metric [14]. The results are illustrated in Table 2. The hallucinating methods can reach higher scores since they invent very sharp objects. On the contrary, ours gives the sharpest images among the methods which cannot hallucinate. The LPC-SI scores of the different methods are shown in Table 2.

B. IMPLEMENTATION DETAILS

We use the ADAM optimizer [27] with a fixed learning rate of $1e^{-4}$ optimized over 200 epochs, Pytorch [10] as framework and the Kornia library [42]. We empirically find the coefficients $\lambda_{\text{recon}} = 5e^1, \lambda_{color} = 1e^1, \lambda_{HR} = 1, \lambda_E = 5e^1, \gamma_R = 2, \gamma_G = 2, \gamma_B = 6, \sigma_R = 15, \sigma_G = 10, \sigma_B = 15$. We crop the images to the size 256x256 and group them by 3 to make a batch. To denoise the component, we use the plug-and-play denoising network trained on spatially varying noise in [29]. We found out that we get better results using the same noise level map for all color channels as illustrated in Fig. 6.

C. ABLATION STUDY

If we consider the original Retinex model with a grayscale smooth illumination, we get the results shown in Fig. 5. Then, even if we denoise the reflectance, we cannot obtain visually pleasing outputs here.

In Fig. 8, Retinex components of a daylight image are illustrated. Textural details such as the frontage of the buildings end up in the reflectance.

D. QUALITATIVE COMPARISON

1) On the Waymo dataset [46]

The state-of-the-art results are illustrated in Figs. 7 and 9. Fig. 7 contains the outputs of several style transfer methods. These works are mainly aiming at augmenting data to enhance datasets with the goal of training networks which will be robust to these modifications. CoMoGAN [39] simulates night images with daylight images and cannot do the reverse process as seen in Fig. 7b. In Figs. 7c and 7d, ManiFest [40] and MUNIT [22] completely modify parts of the image and do not preserve the integrity of the scene which is undesirable in our case.

In Fig. 9f, EnlightenGAN [25] hallucinates trees in the background as shown in the red squares which is obviously

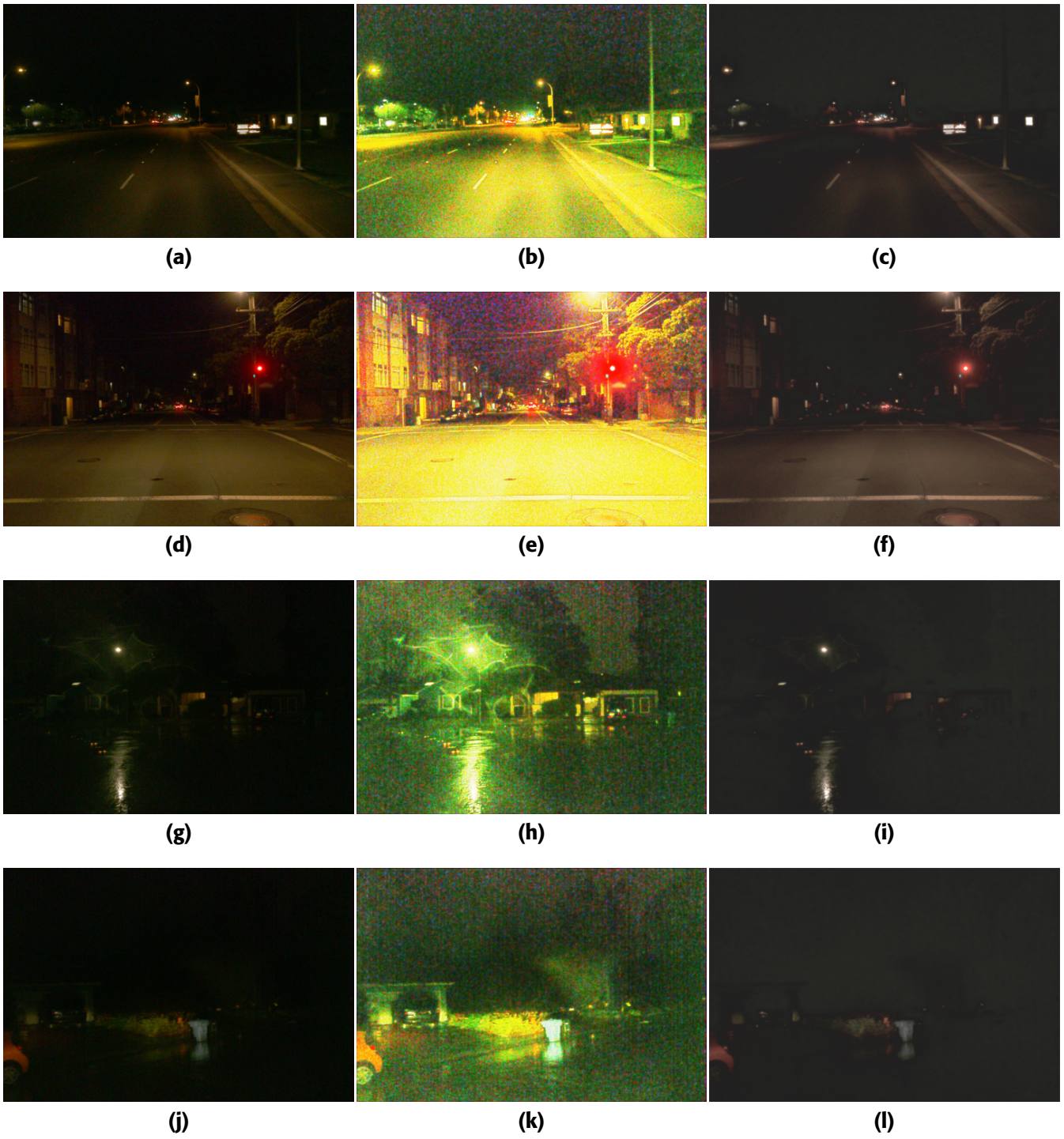

FIGURE 2. From left column to right column: the input image, the extracted raw illumination and reflectance. The illumination contains the specific low-light noise and degradations.

Dataset	Publicly available	Indoor/Outdoor	Contains paired images	Contains real-world scenes	Size
Waymo [46]	✓	Outdoor	✗	✓	128 093 normal-light images / 15419 low-light images
BDD [60]	✓	Outdoor	✗	✓	14772 low-light images
LOL [52]	✓	Indoor	✓	✓	500 pairs

TABLE 1. Properties of the different datasets we use throughout the paper.

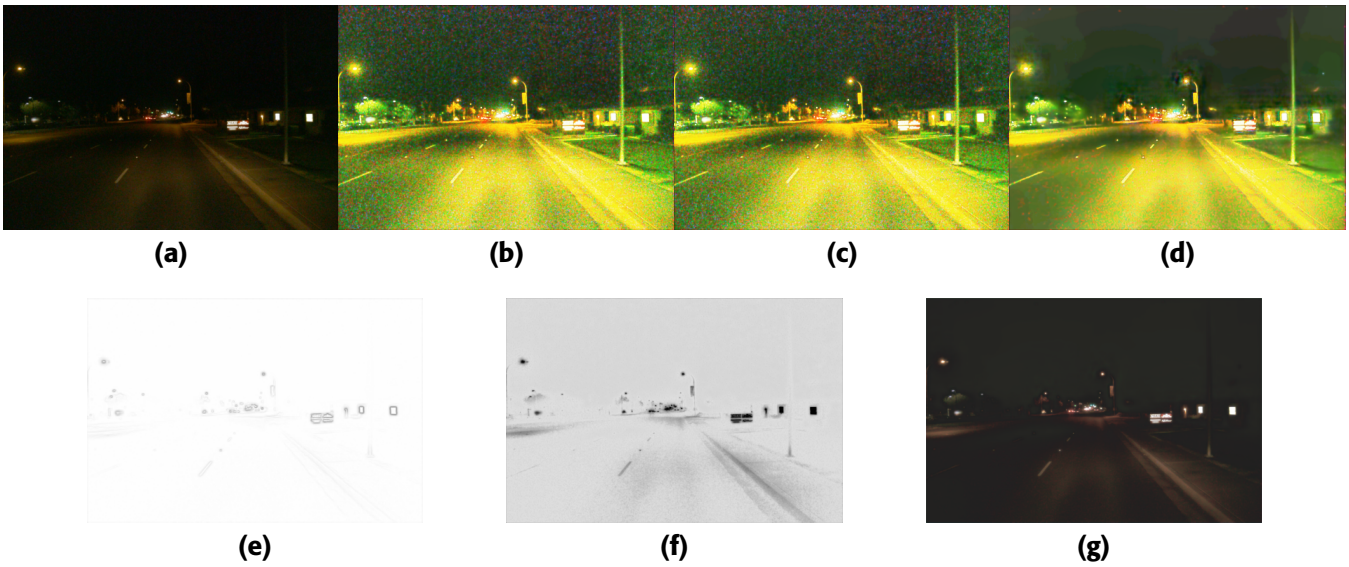

FIGURE 3. From left to right: the input image Fig. 3a, the extracted raw illumination Fig. 3b, the denoised component Fig. 3c, the component if we denoise before the decomposition Fig. 3d. To denoise the illumination, the noise map is weighted. Since the gradient of the reflectance Fig. 3e has less information about the structure of the scene than the weight map Fig. 3f as the inverse of the approximation of the illumination as defined in LIME [13], we use the latter here. We could not further denoise the image as it would lead to a loss of details. Fig. 3d shows that if we denoise the image before the decomposition, it only affects the illumination as the reflectance remains untouched Fig. 3g. We decide to denoise the component after the decomposition as the former would lead to some artifacts introduced by the denoising network and amplified by the restoration of the components.

FIGURE 4. From left to right: the input image, the extracted raw reflectance, the gamma corrected component. The low-light noise is not strengthened after the restoration which confirms the components are correctly separated according to the defined model. We restore with a different γ for each RGB color channel as we empirically find it leads to visually better output images. The higher gamma value for the blue color channel giving the component its non-natural hue is set to balance the green noise of the illumination. See Fig. 12 for an example of the components extracted from an image of the BDD100K dataset [60].

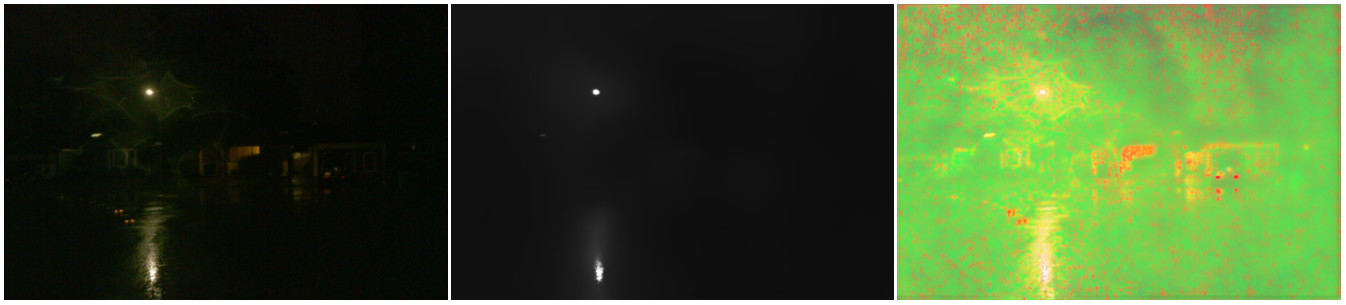

FIGURE 5. From left to right: the input image, the extracted raw illumination and reflectance if we consider a grayscale smooth illumination as previously used in the literature.

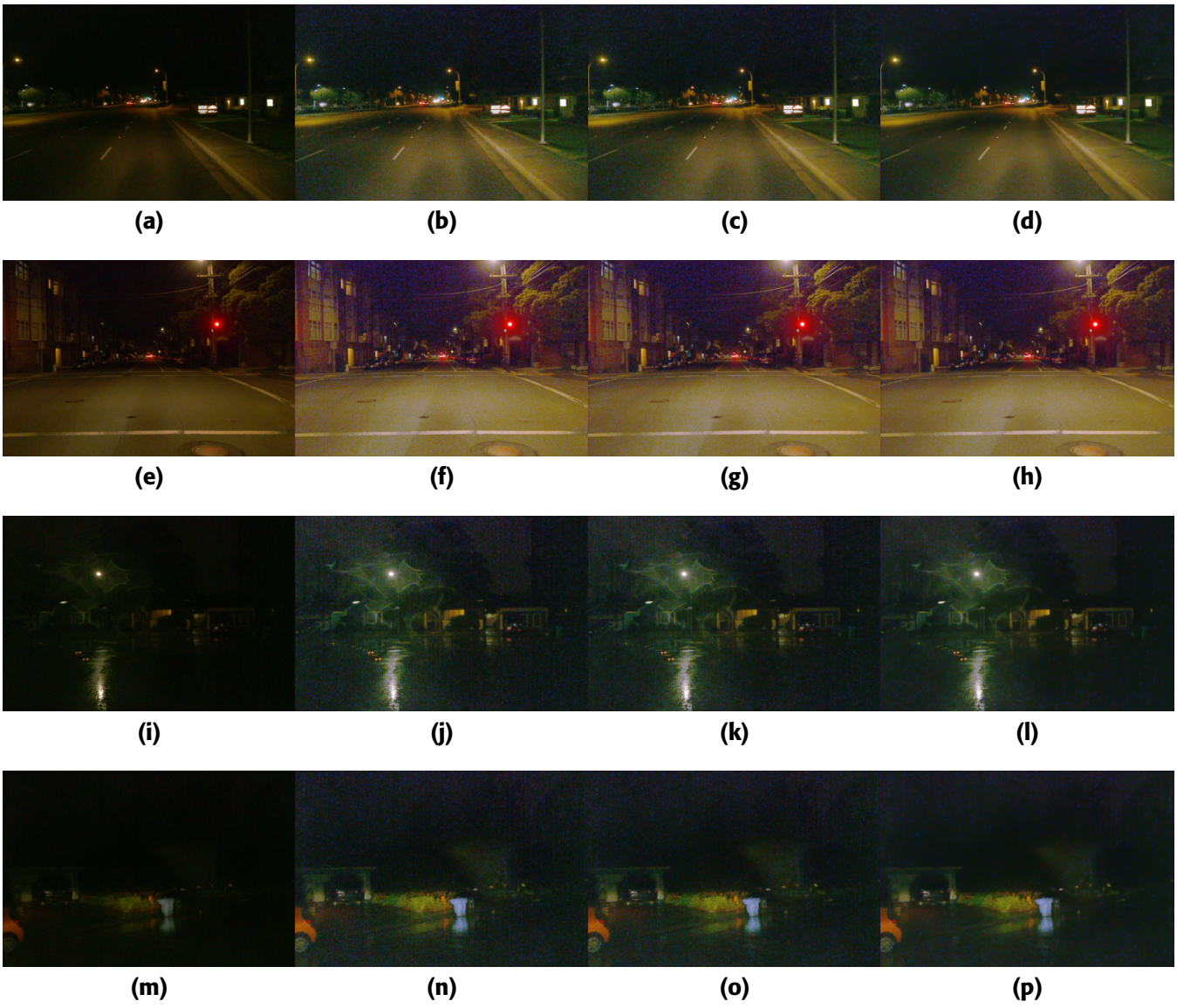

FIGURE 6. From left column to right column: the input image, the output of our method without denoising, with denoising and denoising with a different noise level for each RGB color channel.

FIGURE 7. Style transfer and data augmentation methods with from left to right: Fig. 7a The input image, Fig. 7b CoMoGAN [39], Fig. 7c ManiFest [40], Fig. 7d MUNIT [22] trained on the Waymo dataset. These style transfer methods do not preserve the integrity of the scene in the input image and add hallucinations.

not desirable in our case. However, the image is sharper. We fine-tune it on the same dataset to see if we could improve the results and still obtained hallucinations in Fig. 9g.

For the gamma correction, we apply the same gamma

(a) (b) (c)

FIGURE 8. From left column to right column: the input normal-light image, the extracted raw illumination and the reflectance. Textural details such as the frontage of the buildings end up in the reflectance which demonstrates the quality of the decomposition model.

Methods	LPC-SI(↑)
MUNIT [22]	0.95428
EnlightenGAN$_{\text{pretrained}}$ [25]	0.97610
EnlightenGAN$_{\text{waymo}}$ [25]	**0.97848**
Retinex DIP color [30]	0.94439
Retinex DIP gray [30]	0.94686
Zero-DCE [31]	0.96943
Gamma Correction$_{\text{HSV}}$	0.96746
Gamma Correction$_{\text{RGB}}$	0.96463
KinD++ [63]	0.96572
LIME [13]	0.96343
Ours	**0.97152**

TABLE 2. LPC-SI scores [14] on the Waymo dataset with respectively in blue methods that hallucinate and in black methods which don't. Scores in bold are the highest scores in each of the category. We obtain the best LPC-SI score among the non-hallucinating approaches.

values as with our method (i.e. with a higher gamma for the blue channel) in the RGB color space and the result is shown in Fig. 9h. The "fog" effect results from the gamma correction applied to all the color channels. The histogram of the blue channel is shifted to the right as shown in Fig. 11b. This effect is not present if we restore the V channel in the HSV color space instead, see Fig. 9i and its corresponding histogram Fig. 11c. With our method, it's not the case even if we restore the reflectance with a gamma correction on all RGB channels as shown in Figs. 11d and 11e. There is little to no difference applying a gamma correction on all RGB channels or V channel in HSV with respect to the histograms. Moreover, the histograms are flatter than the ones with gamma correction only which can be seen as histogram equalizations or contrast enhancement. We get visually more appealing results with our method than the gamma correction.

Looking at the histograms of the input images, the blue channel pixels have really low intensity and there are more information about the scene in the red and green channels.

Our method Fig. 9b leads to a visually better result than the competitors such as LIME, Zero-DCE or KinD++ Figs. 9c, 9j and 9k. Applying our network trained on the Waymo dataset [46] to the BDD100k dataset [60] leads to the results shown in Fig. 12. We only reduce the $\gamma_B = 2$ to the same value as the other color channels as this dataset does not suffer from the green hue. We obtain similar results with a network pretrained on another dataset which really highlights the generalization ability of the restoration. Moreover, the visual quality of the decomposition is on par with the one on the Waymo dataset. Nonetheless, we emphasize that this dataset is composed of images with similar scenes (same objects and backgrounds).

2) Failure cases

Fig. 10 illustrates the outputs of our method if the network is trained on the commonly known LOL dataset. Figs. 10a and 10b are respectively the input low-light image and its corresponding ground-truth normal-light image. Figs. 10c and 10d show the illumination and reflectance components we can extract if a colored illumination is considered. The dataset size being relatively small with around 500 paired images, a GAN-based architecture has trouble to learn in an unsupervised fashion. Specifically, the common information of widely diverse scenes is really challenging to estimate. If we consider a grayscale smooth illumination like in previous works, the constraints are strong enough to guide the decomposition but the low-light noise still ends up in the reflectance and makes it even more difficult to get rid of it. If we apply the network already trained on the Waymo dataset to the LOL dataset, we obtain results as shown in Figs. 10g and 10h. Since there is an enormous gap between the type of scene and degradation between the two datasets, the network experiences difficulty in extracting the correct information. We provide more information on the different datasets we use in Table 1.

V. CONCLUSION

In this paper, we proposed a new approach based on state-of-the-art source separation and style transfer methods to decompose in an unsupervised fashion outdoor nighttime images. We improved on the original Retinex model by extracting common information between the low and normal-light domain thanks to a colored illumination. Moreover, we also defined a new architecture with deep neural networks building on this physical model. To the best of our knowledge,

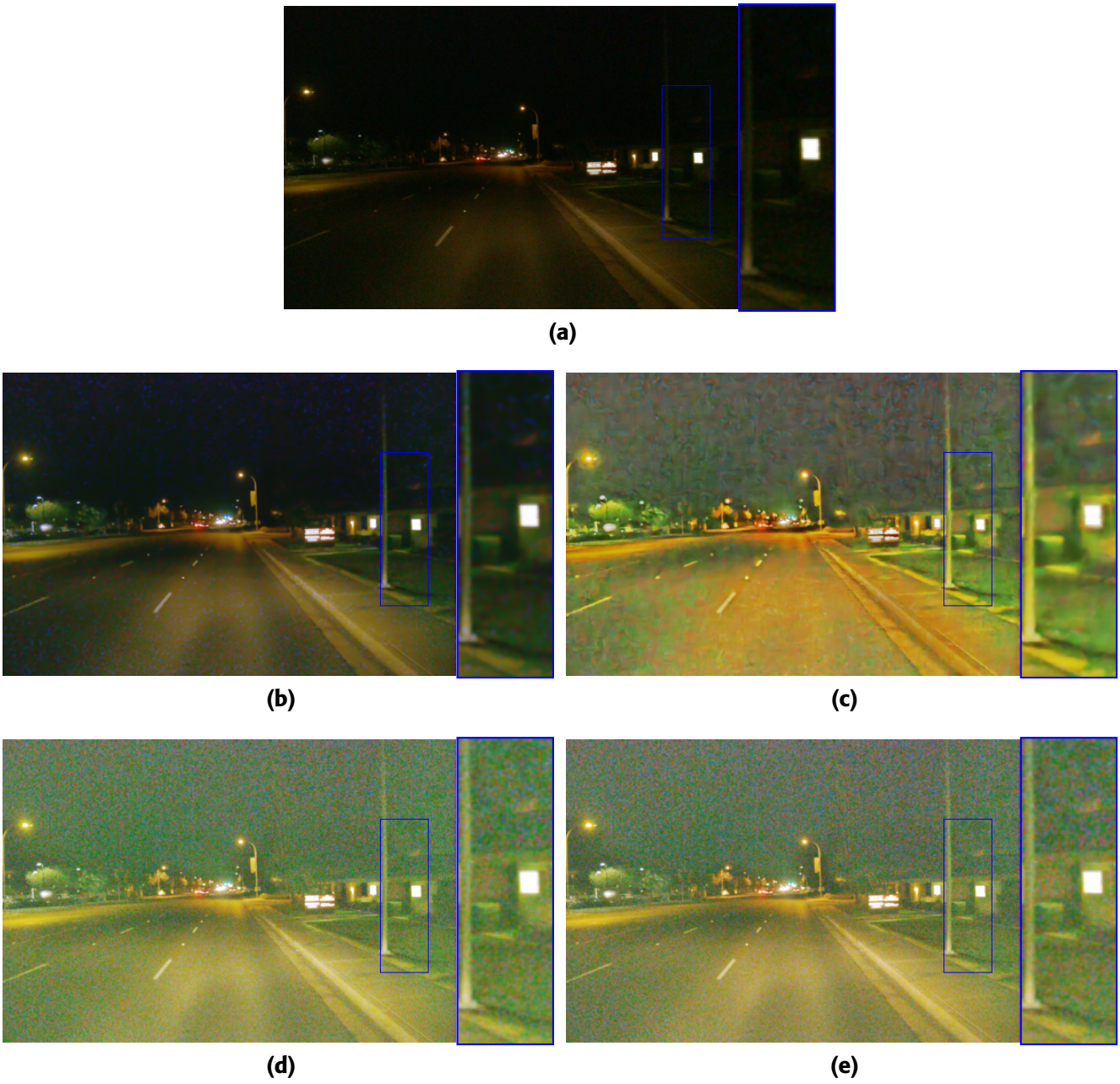

FIGURE 9. First part of the qualitative comparison between the restoration methods applied to the input image Fig. 9a and the state-of-the-art approaches. Our method Fig. 9b leads to a visually better result than the competitors with respectively Fig. 9c KinD++ [63], Fig. 9d Retinex DIP color [30], Fig. 9e Retinex DIP gray [30].

this is the first time this definition of the Retinex components is put into practice. It makes it feasible to visualize the complex style known as the illumination and the reflectance in an image. Applied to the Waymo dataset, our method is more stable and produces visually pleasing images without hallucinating parts of the image compared to the state-of-the-art methods. However, different aspects of the method could be improved in a future work. Indeed, each non-linear operation applied by the camera pipeline like a gamma correction or a specific tone mapping makes it more difficult to decompose the image. Reversing this pipeline from a single input image is already an active research field in the literature. See [32] for instance. Since it's another challenging inverse problem and our model is still valid in our context, reversing the camera pipeline is beyond the scope of this paper. It remains an interesting research direction to improve the accuracy of the estimation of the Retinex components. We also could find a better method to restore the components than a gamma correction and a denoising network.

REFERENCES

[1] True Color Kodak Images. https://www.r0k.us/graphics/kodak/.
[2] Vasileios Vonikakis - Datasets. https://sites.google.com/site/vonikakis/datasets.

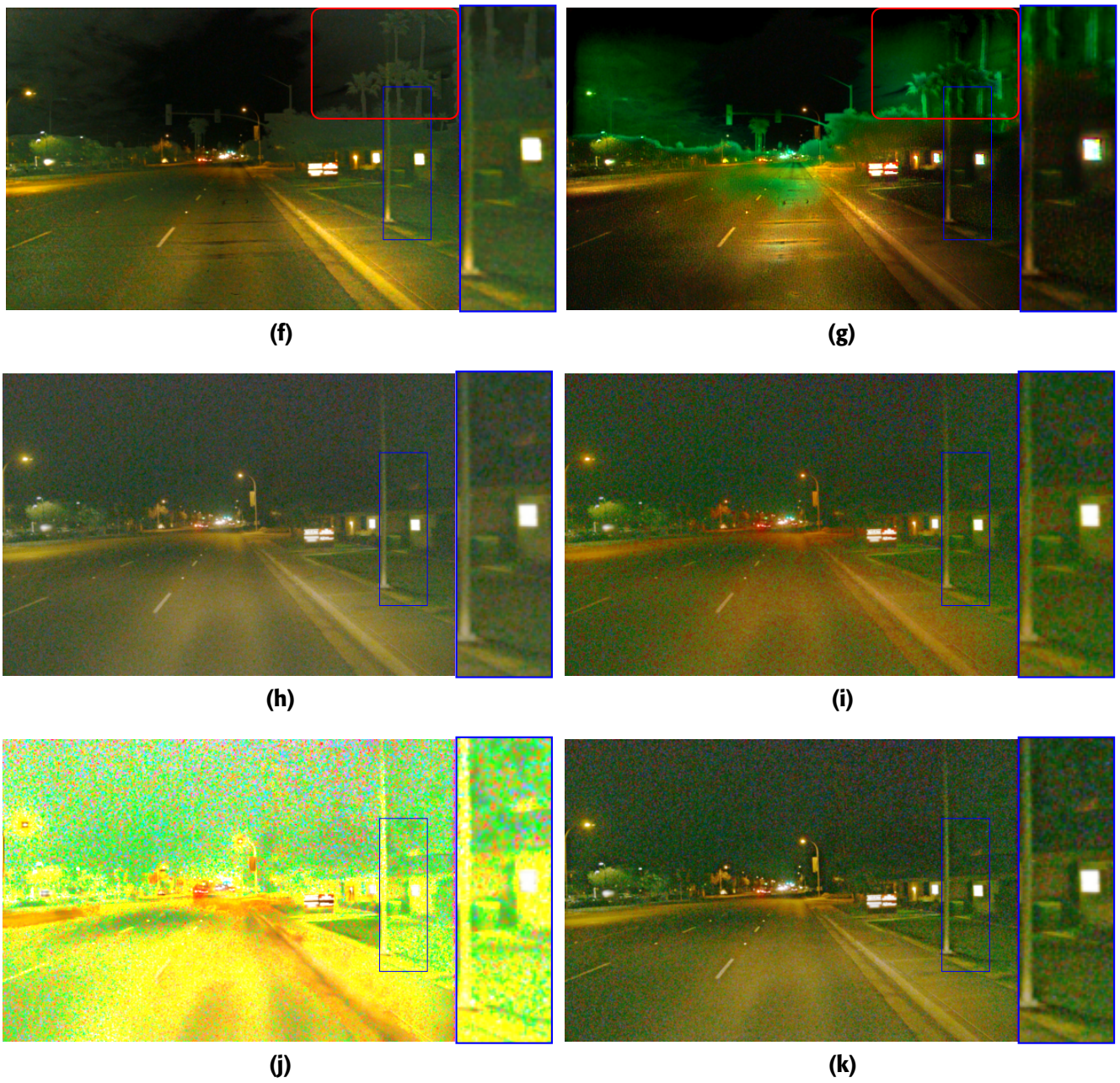

FIGURE 9. **Second part of the qualitative comparison between the restoration methods applied to the input image Fig. 9a and the state-of-the-art approaches. Fig. 9f EnlightenGAN [25] and Fig. 9g EnlightenGAN fine-tuned on the Waymo dataset [25] hallucinates trees at the top of the image as shown in the red squares. Applying gamma corrections in the RGB and HSV spaces leads to the outputs illustrated in Figs. 9h and 9i with an undesirable "fog" effect whitening the image. Fig. 9j LIME [13] and Fig. 9k Zero-DCE [31] amplify the noise in the darkest parts of the image.**

[3] Mahmoud Afifi, Konstantinos G Derpanis, Bjorn Ommer, and Michael S Brown. Learning multi-scale photo exposure correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9157–9167, 2021.
[4] Harry G. Barrow and J. Martin Tenenbaum. Recovering intrinsic scene characteristics from images. *Computer Vision Systems*, 1978.
[5] Ricardo Augusto Borsoi, Tales Imbiriba, José Carlos Moreira Bermudez, Cédric Richard, Jocelyn Chanussot, Lucas Drumetz, Jean-Yves Tourneret, Alina Zare, and Christian Jutten. Spectral Variability in Hyperspectral Data Unmixing: A Comprehensive Review. *IEEE Geoscience and Remote Sensing Magazine*, 9(4):223–270, December 2021.
[6] Mark Boss, Raphael Braun, Varun Jampani, Jonathan T. Barron, Ce Liu, and Hendrik P.A. Lensch. NeRD: Neural Reflectance Decomposition from Image Collections. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12664–12674, Montreal, QC, Canada, October 2021. IEEE.
[7] Tim Brooks, Ben Mildenhall, Tianfan Xue, Jiawen Chen, Dillon Sharlet, and Jonathan T Barron. Unprocessing images for learned raw denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11036–11045, 2019.
[8] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In *CVPR 2011*, pages 97–104. IEEE, 2011.
[9] Robert L Cook and Kenneth E. Torrance. A reflectance model for computer graphics. *ACM Transactions on Graphics (ToG)*, 1(1):7–24, 1982.
[10] Paszke et al. PyTorch: An Imperative Style, High-Performance Deep

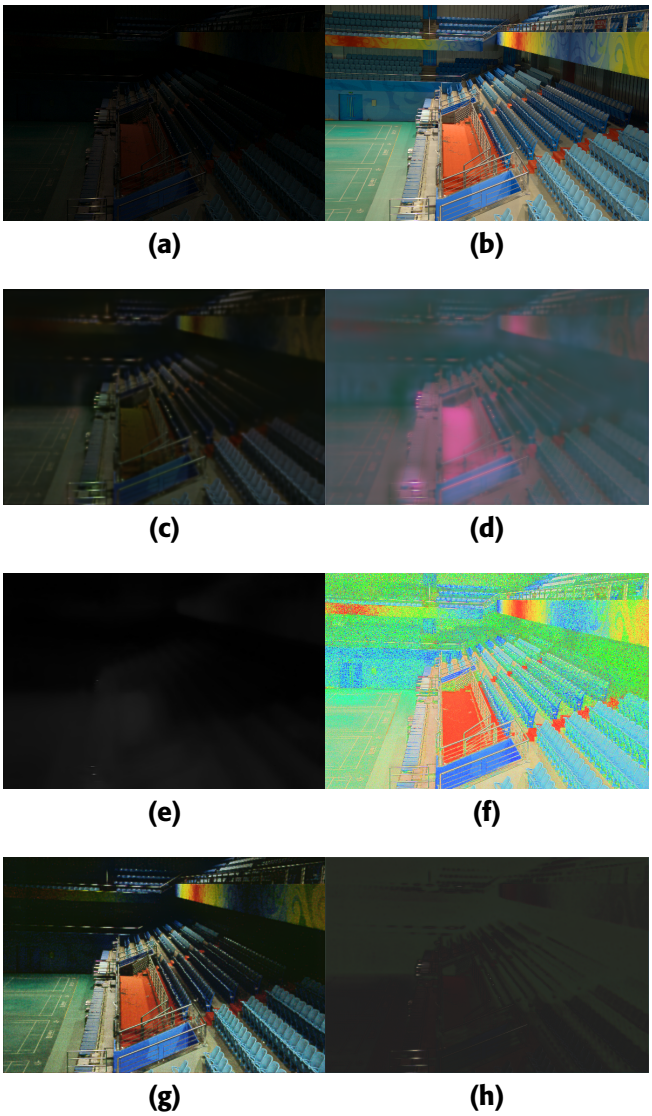

FIGURE 10. **Illustration of failure cases of the decomposition if applied to the LOL dataset on the input image Fig. 10a compared to the ground-truth image Fig. 10b. The model cannot extract the two components if trained with a colored illumination Figs. 10c and 10d because the dataset is too small and contains a high diversity of scenes. The same problem occurs if pretrained on the Waymo dataset [46] Figs. 10g and 10h. Using a grayscale illumination for training Figs. 10e and 10f, it recovers a smooth illumination which was also a previous prior in the literature.**

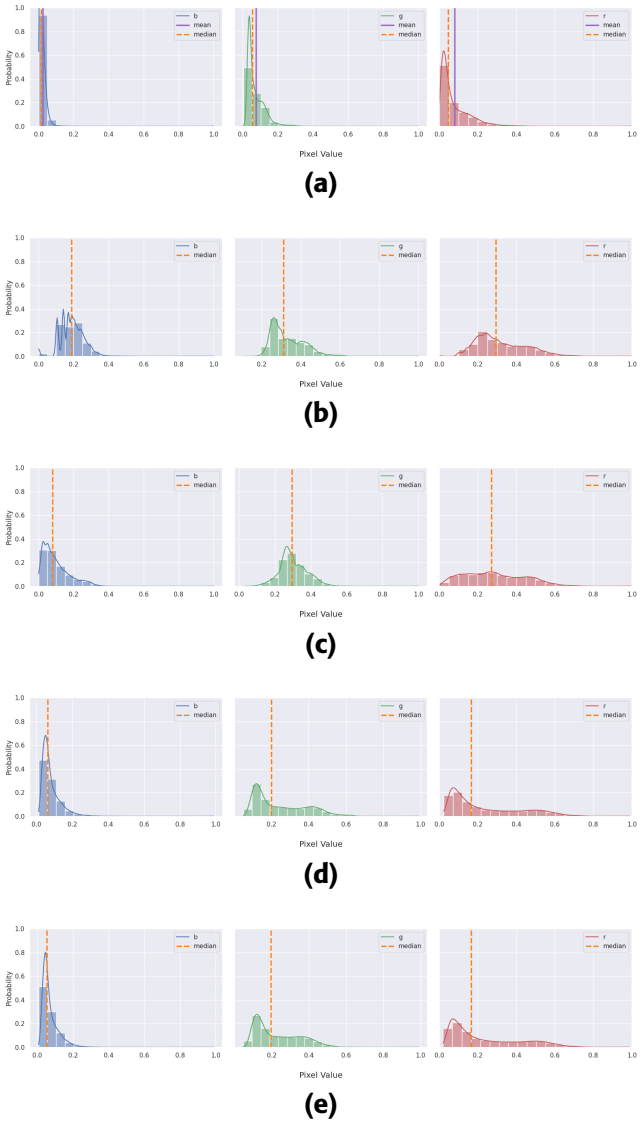

FIGURE 11. **Comparison of histograms between the input image Fig. 11a, the image after applying a simple gamma correction on all color channels in RGB Fig. 11b or the V channel in HSV Fig. 11c, the image restored with a gamma correction according to our method in RGB Fig. 11d or HSV Fig. 11e.**

Learning Library. *Advances in Neural Information Processing Systems*, December 2019.

[11] Biebele Joslyn Fubara, Mohamed Sedky, and Dave Dyke. RGB to Spectral Reconstruction via Learned Basis Functions and Weights. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1984–1993, Seattle, WA, USA, June 2020. IEEE.

[12] Leon Gatys, Alexander S Ecker, and Matthias Bethge. Texture synthesis using convolutional neural networks. *Advances in neural information processing systems*, 28, 2015.

[13] Xiaojie Guo, Yu Li, and Haibin Ling. LIME: Low-Light Image Enhancement via Illumination Map Estimation. *IEEE Trans. on Image Processing*, February 2017.

[14] R. Hassen, Zhou Wang, and M. M. A. Salama. Image Sharpness Assessment Based on Local Phase Coherence. *IEEE Transactions on Image Processing*, July 2013.

[15] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium, January 2018.

[16] Berthold K.P. Horn. Determining lightness from an image. *Computer Graphics and Image Processing*, 3(4):277–299, December 1974.

[17] Qiming Hu and Xiaojie Guo. Trash or treasure? an interactive dual-stream strategy for single image reflection separation. *Advances in Neural Information Processing Systems*, 34:24683–24694, 2021.

[18] Haofeng Huang, Wenhan Yang, Yueyu Hu, and Jiaying Liu. Raw-guided enhancing reprocess of low-light image via deep exposure adjustment. In *Proceedings of the Asian Conference on Computer Vision*, 2020.

[19] Haofeng Huang, Wenhan Yang, Yueyu Hu, Jiaying Liu, and Ling-Yu Duan. Towards low light enhancement with raw images. *IEEE Transactions on Image Processing*, 31:1391–1405, 2022.

[20] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1501–1510, 2017.

[21] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal Unsupervised Image-to-Image Translation. In Vittorio Ferrari, Martial Hebert,

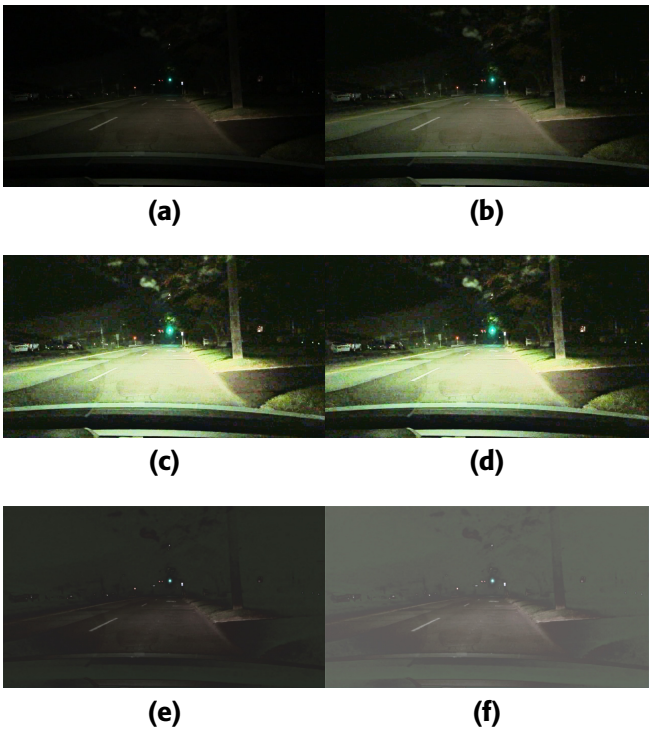

FIGURE 12. An example of the obtained Retinex components and the restoration output on an image coming from the BDD100k dataset [60]. We apply the network trained on the Waymo dataset [46]. From left column to right column: the input image Fig. 12a, the restored image Fig. 12b, the extracted raw illumination Fig. 12c and its restored version Fig. 12d, the extracted raw reflectance and its restored version Figs. 12e and 12f. The illumination still contains the low-light noise and degradations which strengthen the conclusion on the quality of the decomposition.

Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, volume 11207, pages 179–196. Springer International Publishing, Cham, 2018.

[22] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal Unsupervised Image-to-Image Translation. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, volume 11207, pages 179–196. Springer International Publishing, Cham, 2018.

[23] ISO-9288:2022. Thermal insulation — Heat transfer by radiation — Vocabulary. Technical report, ISO, Geneva, CH, August 2022.

[24] Jun Jiang, Dengyu Liu, Jinwei Gu, and Sabine Süsstrunk. What is the space of spectral sensitivity functions for digital color cameras? In *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, pages 168–179, January 2013.

[25] 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. *IEEE Transactions on Image Processing*, 2021.

[26] Seon Joo Kim, Hai Ting Lin, Zheng Lu, Sabine Süsstrunk, Stephen Lin, and Michael S. Brown. A New In-Camera Imaging Model for Color Computer Vision and Its Application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(12):2289–2302, December 2012.

[27] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*, 2015.

[28] Edwin H Land. The Retinex Theory of Color Vision. *Scientific American*, December 1977.

[29] Mikael Le Pendu and Christine Guillemot. Preconditioned plug-and-play ADMM with locally adjustable denoiser for image restoration mikael. *SIAM Journal on Imaging Sciences*, pages 1–30, 2022.

[30] Arthur Lecert, Renaud Fraisse, Aline Roumy, and Christine Guillemot. A new regularization for retinex decomposition of low-light images. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 906–910, October 2022.

[31] Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to Enhance Low-Light Image via Zero-Reference Deep Curve Estimation. *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2020.

[32] Yu-Lun Liu, Wei-Sheng Lai, Yu-Sheng Chen, Yi-Lung Kao, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Single-Image HDR Reconstruction by Learning to Reverse the Camera Pipeline. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1648–1657, Seattle, WA, USA, June 2020. IEEE.

[33] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual Quality Assessment for Multi-Exposure Image Fusion. *IEEE Transactions on Image Processing*, 24(11):3345–3356, November 2015.

[34] Tian Ma, Ming Guo, Zhenhua Yu, Yanping Chen, Xincheng Ren, Runtao Xi, Yuancheng Li, and Xinlei Zhou. RetinexGAN:Unsupervised Low-Light Enhancement With Two-Layer Convolutional Decomposition Networks. *IEEE Access*, 9:56539–56550, 2021.

[35] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2794–2802, 2017.

[36] S.G. Narasimhan and S.K. Nayar. Shedding light on the weather. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, pages I–665–I–672, Madison, WI, USA, 2003. IEEE Comput. Soc.

[37] Fred E. Nicodemus. Directional reflectance and emissivity of an opaque surface. *Applied Optics*, 4(7):767–775, July 1965.

[38] Bui Tuong Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, 1975.

[39] Fabio Pizzati, Pietro Cerri, and Raoul de Charette. CoMoGAN: Continuous model-guided image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14288–14298, 2021.

[40] Fabio Pizzati, Jean-François Lalonde, and Raoul de Charette. Manifest: Manifold deformation for few-shot image translation. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVII*, pages 440–456. Springer, 2022.

[41] Edoardo PROVENZI. Formalizations of the Retinex model and its variants with variational principles and partial differential equations. *Journal of Electronic Imaging*, December 2017.

[42] Edgar et al. Riba. Kornia: An Open Source Differentiable Computer Vision Library for PyTorch. *Winter Conference on Applications of Computer Vision*, October 2019.

[43] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[44] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, page 2234–2242, Red Hook, NY, USA, 2016. Curran Associates Inc.

[45] Pradeep Sen, Nima Khademi Kalantari, Maziar Yaesoubi, Soheil Darabi, Dan B. Goldman, and Eli Shechtman. Robust patch-based hdr reconstruction of dynamic scenes. *ACM Transactions on Graphics*, 31(6):1–11, November 2012.

[46] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2446–2454, 2020.

[47] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Improved Texture Networks: Maximizing Quality and Diversity in Feed-Forward Stylization and Texture Synthesis. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4105–4113, Honolulu, HI, July 2017. IEEE.

[48] 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, September 2013.

[49] Wencheng Wang, Zhenxue Chen, and Xiaohui Yuan. Simple low-light image enhancement based on Weber–Fechner law in logarithmic space. *Signal Processing: Image Communication*, 106:116742, 2022.
[50] Zhou et al. Wang. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. on Image Processing*, 2004.
[51] Allan G Weber. USC-SIPI Report #432 The USC-SIPI Image Database: Version 6.
[52] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep Retinex Decomposition for Low-Light Enhancement. *British Machine Vision Conference*, August 2018.
[53] Ling Xing, Hongyu Qu, Sheng Xu, and Yao Tian. CLEGAN: Toward Low-Light Image Enhancement for UAVs via Self-Similarity Exploitation. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–14, 2023.
[54] Ruikang Xu, Mingde Yao, Chang Chen, Lizhi Wang, and Zhiwei Xiong. Continuous spectral reconstruction from RGB images via implicit neural representation. In *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part V*, pages 78–94. Springer, 2023.
[55] Longbin Yan, Xiuheng Wang, Min Zhao, Maboud Kaloorazi, Jie Chen, and Susanto Rahardja. Reconstruction of Hyperspectral Data From RGB Images With Prior Category Information. *IEEE Transactions on Computational Imaging*, 6:1070–1081, 2020.
[56] Jing Yao, Danfeng Hong, Jocelyn Chanussot, Deyu Meng, Xiaoxiang Zhu, and Zongben Xu. Cross-attention in coupled unmixing nets for unsupervised hyperspectral super-resolution. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16*, pages 208–224. Springer, 2020.
[57] Jing Yao, Danfeng Hong, Lin Xu, Deyu Meng, Jocelyn Chanussot, and Zongben Xu. Sparsity-enhanced convolutional decomposition: A novel tensor-based paradigm for blind hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2021.
[58] Jing Yao, Deyu Meng, Qian Zhao, Wenfei Cao, and Zongben Xu. Nonconvex-sparsity and nonlocal-smoothness-based blind hyperspectral unmixing. *IEEE Transactions on Image Processing*, 28(6):2991–3006, 2019.
[59] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang. A New Low-Light Image Enhancement Algorithm Using Camera Response Model. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 3015–3022, Venice, October 2017. IEEE.
[60] Fisher Yu, Haofeng Chen, Xin Wang, Wenqi Xian, Yingying Chen, Fangchen Liu, Vashisht Madhavan, and Trevor Darrell. BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, June 2020.
[61] Ye Yu and William A. P. Smith. Outdoor inverse rendering from a single image using multiview self-supervision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7):3659–3675, 2022.
[62] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, April 2018.
[63] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond Brightening Low-light Images. *International Journal of Computer Vision*, 129(4):1013–1037, April 2021.
[64] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the Darkness: A Practical Low-light Image Enhancer. *ACM International Conference on Multimedia*, May 2019.
[65] Zhiyu Zhu, Hui Liu, Junhui Hou, Huanqiang Zeng, and Qingfu Zhang. Semantic-embedded Unsupervised Spectral Reconstruction from Single RGB Images in the Wild. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2259–2268, Montreal, QC, Canada, October 2021. IEEE.

ARTHUR LECERT received the Engineering degree from ESIEE Paris, France in 2020 and is currently working towards the Ph.D. degree from the University of Rennes 1, France.

His current research and study interests include deep learning and computer vision, signal processing as well as generative modeling.

ALINE ROUMY received the Engineering degree from Ecole Nationale Superieure de l'Electronique et de ses Applications (ENSEA), Cergy, France in 1996, the Master degree in 1997 and the Ph.D. degree in 2000 from the University of Cergy-Pontoise, France. During 2000-2001, she was a research associate at Princeton University, Princeton, NJ. In 2001, she joined INRIA, Rennes, France as a research scientist. She has held visiting positions at Eurecom and Berkeley University.

Her current research and study interests include the area of statistical signal and image processing, coding theory and information theory.

She has been a Technical Program Committee member and session chair at several international conferences, including ISIT, ICASSP, Eusipco. She is currently serving as a member of the French National University Council (CNU 61). She received the 2011 "Francesco Carassa" Best paper award.

RENAUD FRAISSE received the Engineering degree from Ecole Polytechnique, France in 1990, the Master degree in 1992 from Telecom Paris, France. He has been working for Airbus Defense and Space in Toulouse, France for more than 30 years first as a Satellite Image Quality team leader and Airbus expert and now as an Imaging Systems End to End Performances Senior Expert.

CHRISTINE GUILLEMOT (Fellow, IEEE) received the Ph.D. degree from the Ecole Nationale Superieure des Telecommunications (ENST) Paris, and the Habilitation for Research Direction degree from the University of Rennes. From 1985 to October 1997, she has been with FRANCE TELECOM, where she has been involved in various projects in the area of image and video coding and processing for TV, HDTV, and multimedia. From January 1990 to 1991, she worked at Bellcore, Murray Hill, NJ, USA, as a Visiting Scientist. She is currently the Director of Research at INRIA. Her research interests include signal and image processing, and computer vision. She has served as an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING (from 2000 to 2003 and from 2014 to 2016), the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (from 2004 to 2006), and the IEEE TRANSACTIONS ON SIGNAL PROCESSING (from 2007 to 2009). She has served as a Senior Member of the Editorial Board of the IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING (from 2013 to 2015). She is also a Senior Area Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING.