A NEW REGULARIZATION FOR RETINEX DECOMPOSITION OF LOW-LIGHT IMAGES

Arthur Lecert[1], Renaud Fraisse[2], Aline Roumy[1], Christine Guillemot[1]

[1] INRIA, Rennes, France, [2] Airbus Defence and Space, Toulouse, France

ABSTRACT

We study unsupervised Retinex decomposition for low light image enhancement. Being an underdetermined problem with infinite solutions, well-suited priors are required to reduce the solution space. In this paper, we analyze the characteristics of low-light images and their illumination component and identify a trivial solution not taken into consideration by the previous unsupervised state-of-the-art methods. The challenge comes from the fact that the trivial solution cannot be completely eliminated from the feasible set as it corresponds to the true solution when the low-light image contains a light source or an overexposed area. To address this issue, we propose a new regularization term which only remove absurd solutions and keep plausible ones in the set. To demonstrate the efficiency of the proposed prior, we conduct our experiments using deep image priors in a framework similar to the recent work RetinexDIP and an in-depth ablation study. Finally, we observe no more halo artefacts in the restored image. For all-but-one metrics, our unsupervised approach gives results as good as the supervised state-of-the-art indicating the potential of this framework for low-light image enhancement.

Index Terms— Low light enhancement, Image decomposition, Image restoration, Inverse problems, Retinex model, Neural Networks

1. INTRODUCTION

Many image processing algorithms (e.g., for object detection, classification, segmentation, recognition, scene understanding and 3D reconstruction) have been designed for normal lighting conditions, and do not perform well in low light environment. Indeed, in this environment, image processing tasks fail because the camera sensor captures a limited number of photons. This leads to very low signal-to-noise ratio, but also to lightness and color deviations.

The Retinex model [1, 2] has been shown to be an effective a priori to perform low-light image restoration. According to this theory, it is assumed that any image is a product of two components: reflectance and illumination. The reflectance is a color image and defines the natural color of an object irrespective of the illumination condition. Whereas, the illumination models the lighting condition as a one-channel image. Once the image is decomposed, one can apply component specific processing: color deviation correction for the reflectance, and lighting enhancement for the illumination. This leads to a very effective normal-light image reconstruction; and shows the effectiveness of the Retinex decomposition.

The state-of-the-art methods for Retinex decomposition are based on deep neural networks trained on datasets in an end-to-end manner [3]. While they can efficiently decompose an image with powerful supervised priors knowing the ground-truth, relying on a dataset remains a problem. Indeed, it is extremely difficult to obtain ideal pairs of low-light/normal-light images to isolate the degradation. Thus, when these networks are applied on natural out-of-distribution images, the restored images may contain artifacts or color deviations. Different implementations of the core Retinex prior (i.e. structure-aware smooth illumination component) have been proposed in the signal-prior based approach [4] and in the supervised frameworks of Wei et al. [5] and Zhang et al. [3]. Recent works depart from supervised learning and focus on unsupervised learning instead to address the difficulty of collecting ground-truth data. For the sake of fairness, we will also compare the performance of our approach against EnlightenGAN [6] and Zero-DCE [7]. These solutions require a careful selection of the training data such as a dataset containing multi-exposure unpaired images or a manual inspection to remove images of medium brightness. Still, few methods try to decompose an image following the Retinex model in an unsupervised fashion.

A recent work by Zhao et al. [8] combined the Retinex decomposition with deep image priors [9] to generate the two intrinsic components of the image. Then, they enhance the low-light image by restoring the illumination with a gamma correction. The authors designed an illumination consistency prior to constrain the generated illumination to be close to an approximation of the illumination, a first plausible guess. This is a hard constraint on the set of reachable solutions and therefore the deep image priors can't explore as much solutions as they should.

The contributions of this paper are manifold. First, we rewrite the retinex decomposition as a scale mixture which is a widespread and well-studied model in the image processing literature especially in the wavelet domain with Gaussian priors [10, 11, 12]. In the context of low light images, we identify a trivial solution (i.e. when the scaling factor is equal to one), and analyze its properties. Then, we propose a new prior that address this problem while still letting the deep im-

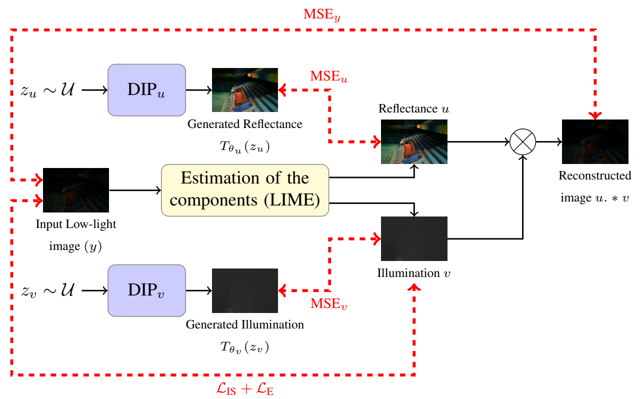


Fig. 1. Retinex decomposition scheme: The DIPs $(T_{\theta_u}, T_{\theta_v})$, initialized with random noises (z_u, z_v), produce two initial estimate $(T_{\theta_u}(z_u), T_{\theta_v}(z_v))$. They are defined in Eq. (4)-(6). These are further improved with the general loss (8) to produce the final components $\hat{u}$ and $\hat{v}$.

age priors explore as much plausible solutions as before. Finally, we propose to restore the image with two gamma corrections, one for each component. The fact that component-specific corrections are used, shows the necessity to decompose the low-light image. We demonstrate the effectiveness of our decomposition achieving good performance using simple gamma corrections. Finally, we observe no more halo artefacts on the restored image. For all-but-one metrics, our approach gives results as good as the supervised state-of-the-art indicating the potential of a generative framework for low-light image enhancement.

2. INITIAL PROBLEM STATEMENT AND BACKGROUND

2.1. The Retinex Model and estimating its components

According to the Retinex theory [1, 2], an image $y \in \mathbb{R}^{3n}$ is a noisy observation of the product of two components

$$y = v. * u + \eta. \tag{1}$$

where $.*$ is the element-wise product, $v \in \mathbb{R}^n$ the illumination map, $u \in \mathbb{R}^{3n}$ the reflectance component of this image, and η the additive Gaussian noise. This model can be seen as a scale mixture since one pixel is the product of a scaling factor (i.e the illumination) applied to all three RGB components of the reflectance. The illumination component contains the lightness information of the scene including shadows or light sources. The reflectance consists of the intrinsic color of the elements of the scene regardless of the exposure conditions.

Computing the elements of a scale mixture is an underdetermined problem and inherently ill-posed. Indeed, we need to estimate $4n$ variables from $3n$ observations in the context of the Retinex decomposition. There is a need of suited priors to regularize the loss function and therefore to reduce the solution space. Thus, we seek to estimate $(\hat{v}, \hat{u})$ such that

$$(\hat{v}, \hat{u}) = \underset{v,u}{\operatorname{argmin}} \|y - v. * u\|_2^2. \tag{2}$$

2.2. Background on the common priors

In a ideal decomposition, the illumination map should not contain any texture details but still keep the structure of the scene. Thus, we first use the structure-aware illumination smoothness prior in [5],

$$\mathcal{L}_{IS} = \left\| \frac{\nabla v}{\max(\nabla y, \epsilon)} \right\|_1. \tag{3}$$

If ∇y has high values, the penalty will be small to smooth the surface whereas low values will give a high penalty to force the illumination component v to be close.

Using deep image priors to generate the components, we get the following optimization problem,

$$(\hat{\theta}_v, \hat{\theta}_u) = \underset{\theta_v \in \mathbb{R}^p, \theta_u \in \mathbb{R}^{p'}}{\operatorname{argmin}} \|y - T_{\theta_v}(z_v). * T_{\theta_u}(z_u)\|_2^2. \tag{4}$$

$$T_{\theta_u} : z_u \sim \mathcal{U}(0,1)^{3n} \to T_{\theta_u}(z_u) \in \mathbb{R}^{3n} \tag{5}$$

$$T_{\theta_v} : z_v \sim \mathcal{U}(0,1)^{n} \to T_{\theta_v}(z_v) \in \mathbb{R}^{n} \tag{6}$$

where (z_u, z_v) are the input noises, and $(T_{\theta_u}(z_u), T_{\theta_v}(z_v))$ their respective outputs.

Because the illumination has to contain only low frequencies, high frequencies of the image including noise end up in the reflectance. A core property of the deep image prior is to be robust to noise and to converge faster on naturally looking images [9]. To further reduce the noise, we add a TV penalty $\rho_{TV}(u)$ to the problem following the work in [13].

3. THE PROPOSED PRIOR

3.1. The trivial solution problem and the exposure prior

$v_i = 1$, and thus $u_i = y_i$, should only be admissible when there is a light source or an overexposed area in y at pixel i. Indeed, the illumination can be smoothed out by the prior, the reconstruction of the low-light image still correct and yet the problem could occur. In RetinexDIP [8], they use an Illumination consistency prior which ties the component to the maximum of the low-light image over the color channels. This constrains the component to be close to a first plausible guess. The authors then try to find the best decomposition in the direct neighborhood of the approximation. This reduces the possibility to improve the process further than the initial guess. We propose a new regularization in order to only accept the trivial solution when it is feasible. We define the exposure prior as follows,

$$\mathcal{L}_{\mathrm{E}} = \left\| g\Big(\max_{c \in \{R,G,B\}} y_c\Big) - g\Big(T_{\theta_v}(z_v)\Big) \right\|_2^2 \tag{7}$$

where g is a threshold function $g(x) = \begin{cases} x, & x > t \\ 0, & \text{otherwise} \end{cases}$. We choose $t = 0.9$ here so that this constraint only affects the

Methods	PSNR(↑)	SSIM(↑)	LPIPS(↓)	Runtime (in *s*)
Supervised				
KinD [3]	17.26	**0.77**	**0.187**	1.47
KinD [3] DecompNet → γCorr	17.91	0.64	0.321	1.06
Unsupervised				
LIME [4]	10.10	0.38	0.383	0.26
EnlightenGAN [6]	17.48	0.65	0.322	0.12
Zero-DCE [7]	14.86	0.56	0.335	**0.0012**
RetinexDIP [8]	11.69	0.48	0.351	20.89
Ours → γCorr	**18.11**	0.68	0.306	2220

Table 1. Best and second-best results are highlighted in bold, and blue respectively. KinD [3] relies on paired degraded/ground-truth images to extract its priors. LIME [4] is a signal-prior based approach. Zero-DCE [7] and EnlightenGAN [6] are unsupervised methods but still needs to be trained on a dataset of unpaired multi-exposure images. On the contrary, RetinexDIP [8] and ours are *fully unsupervised*. The latter outperforms the unsupervised competitors while being close to the supervised one. Since we build on the deep image prior [9], it has the same drawback being the high computation time. The difference between RetinexDIP [8] and ours is due to the different number of epochs as discussed in 4.5.

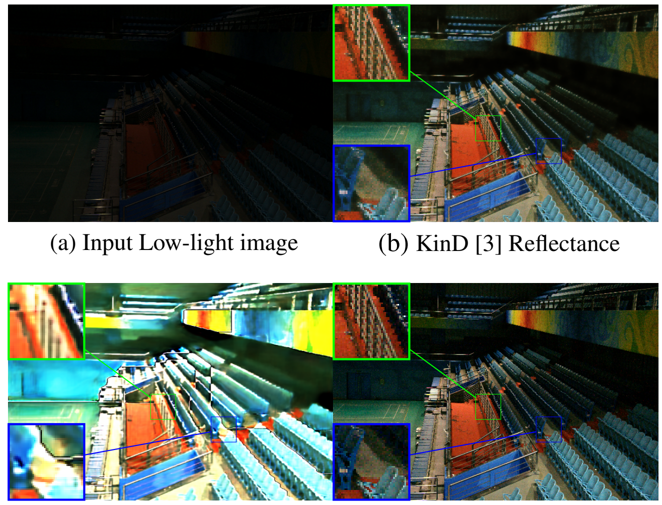
high values of the component. Therefore, the solution set includes the ones with $v_i = 1$ when $y_i = 1$ but forbids the illumination to be equal to one if there is no light source or overexposed regions in the input low-light image.

3.2. New problem formulation

Instead of minimizing over the parameter space of a DIP, the authors in [14] proposed to relax this constraint to improve the performance expanding the solution space. We seek the best compromise between the data fidelity term and the DIPs outputs reaching potentially better solutions. Therefore, we define the additional terms $\|T_{\theta_v}(z_v) - v\|_2^2$ and $\|T_{\theta_u}(z_u) - u\|_2^2$ to keep the estimated components close to the outputs of the DIPs. Each generated component is able to drift from its respective DIP solution. Consequently, in the SUB-DIP formulation [14], the structure of the CNN is really considered as a prior and not as a hard constraint unlike DIP [9], DoubleDIP [15] or RetinexDIP [8]. This is an additional difference between our work and RetinexDIP [8]. We now seek to estimate the best parameters of the DIPs $(\hat{\theta}_v, \hat{\theta}_u)$ as well. Hence, the optimization problem becomes

$$\begin{aligned}(\hat{v}, \hat{u}, \hat{\theta}_v, \hat{\theta}_u) = \underset{v,u,\theta_v,\theta_u}{\operatorname{argmin}} \; & \|y - v.*u\|_2^2 + \lambda_{IS}\mathcal{L}_{IS} \\ & + \lambda_{DIP_u}\|T_{\theta_u}(z_u) - u\|_2^2 + \lambda_{DIP_v}\|T_{\theta_v}(z_v) - v\|_2^2 \\ & + \lambda_{TV}\rho_{TV}(u) + \lambda_E\mathcal{L}_E. \end{aligned} \tag{8}$$

We initialize the components u and v as proposed in LIME [4], i.e., as $v = \max_{c \in \{R,G,B\}} y_c$, $u = y./v$. The complete decomposition process is illustrated in Figure 1.



(a) Input Low-light image (b) KinD [3] Reflectance

(c) RetinexDIP [8] (d) Ours

Fig. 2. Our *fully unsupervised* approach achieves on par intrinsic components with the KinD [3] network trained in an end-to-end fashion on the same dataset. Even with 12000 iterations, RetinexDIP [8] reflectance is cartoon-like.

4. EXPERIMENTS

4.1. Restoration of the components

To restore the components, we take a random subset of 30 paired images from the training set, decompose them with our method and estimate the gamma values. We found that we get better results using a different value for each component. Thus, we use two unique values for all images. This demonstrates the necessity of the Retinex decomposition.

4.2. Implementation details

We use the ADAM optimizer [16] with a fixed learning rate of $1e^{-4}$ and 12000 optimization steps, Pytorch [17] as framework and the Kornia library [18]. We empirically find the coefficients $\lambda_{IS} = 1e^{-4}, \lambda_E = 1e^2, \lambda_{TV} = 1e^{-10}, \lambda_{DIP_u} = 1e^{-2}, \lambda_{DIP_v} = 1e^{-1}, \gamma_u = 0.4, \gamma_v = 0.2$.

4.3. Quantitative comparison

We evaluate our method on the test set of the LOL dataset [5] composed of 500 low/normal-light image pairs taken from real scenes by changing exposure time and ISO.

Methods	PSNR(↑)	SSIM(↑)	LPIPS(↓)	NIQMC(↑)	CPCQI(↑)	NIQE(↓)
RetinexDIP 300 iter [8]	11.69	0.48	0.351	3.718755	1.209930	8.035102
RetinexDIP 12000 iter [8]	11.95	0.49	0.354	3.648840	**1.216181**	8.132518
RetinexDIP 12000 iter [8] → γCorr	17.46	**0.69**	0.380	3.951571	0.461452	**4.536125**
Ours 300 iter → γCorr	16.14	0.59	0.397	3.332199	0.624579	7.168487
Ours 12000 iter → γCorr	**18.11**	0.68	**0.306**	4.207515	0.915857	5.882059
Ours 12000 iter → γ_vCorr only	14.89	0.55	0.331	**4.331838**	1.147143	5.890018

Table 2. Even with the same restoration process and number of iterations, our approach gives the best scores for most of the metrics.

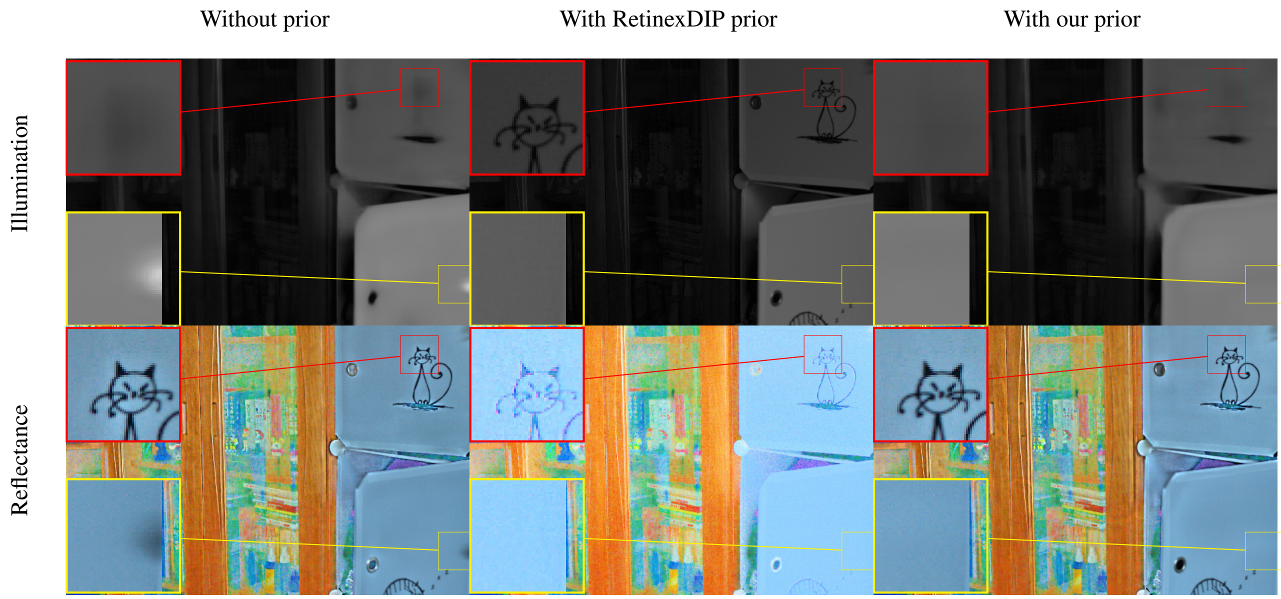


Fig. 3. Without any prior, the components can contain artefacts (yellow square). Although RetinexDIP prior solves this issue, the textural details are still in the illumination after the decomposition (red square). Our prior gets the best of both worlds.

We adopt the following metrics to evaluate the performance of our approach: PSNR, SSIM [19], LPIPS [20], NIQE [21], CPCQI [22] and NIQMC [23]. Therefore, we hope to measure the whole phenomenon of the low light degradation thanks to pixel-wise, classic and learned perceptual metrics.

The chosen state-of-the-art competitors are KinD [3], LIME [4], EnlightenGAN [6], Zero-DCE [7] and RetinexDIP [8]. The first one is a well-known completely supervised network trained on the LOL dataset. The followings methods are unsupervised, either traditional or trained with unpaired data and the latter uses deep image priors in a similar framework. Table 1 summarizes the results. Our method outperforms the unsupervised approaches on most of the metrics while being close to KinD [3].

Since the only available dataset of ground-truth reflectance and illumination [24] is only composed of 16 images, we compare our components against those of the Decomposition Net of KinD [3] and RetinexDIP [8] in Table 1 and in Figure 2.

4.4. Qualitative results

We visually compare the different components generated by the solutions in Figure 2. We obtain visually pleasing components with our approach close to the supervised competitor, KinD [3]. On the contrary, although we use 12000 iterations to get better components out of RetinexDIP [8], the reflectance is still cartoon-like and contains artefacts.

Methods	PSNR(↑)	SSIM(↑)	LPIPS(↓)	NIQMC(↑)	CPCQI(↑)	NIQE(↓)
Without prior	17.36	**0.68**	0.311	3.781431	0.798175	5.931700
RetinexDIP prior	**17.72**	0.67	0.325	**3.932083**	0.793341	6.173116
Our prior	17.44	**0.68**	**0.306**	3.824624	**0.813172**	**5.836579**

Table 3. Our prior achieves the best scores for most of the metrics.

4.5. Ablation and hyperparameter study

To compare the efficiency of the priors, we implement the illumination consistency prior of RetinexDIP [8] in our framework for a fairer ablation study. We reduce the weight of the illumination smoothness as it can alleviate the problem without solving it. Consequently, the visual quality of the components are subpar. The results are shown in Figure 3 and Table 3. With our prior, our approach achieves better scores and leads to a better decomposition without artefacts.

Since RetinexDIP [8] uses only 300 iterations as default compared to the 12000 iterations of our method, we analyze both methods in Table 2 by changing this parameter. We also include the results when applying gamma corrections on the components of RetinexDIP [8]. As shown in Table 2, our approach achieves good performance even if we reduce the number of iterations to 300.

5. CONCLUSION

In this work, we identified a trivial solution problem and proposed a new regularization term to fix it. We have demonstrated its efficiency in an in-depth ablation study. Our framework achieves visually pleasing intrinsic components on par with state-of-the-art supervised methods and outperforms the unsupervised competitors.

6. REFERENCES

[1] Edwin H Land, "The Retinex Theory of Color Vision," *Scientific American*, Dec. 1977.

[2] Harry G. Barrow and J. Martin Tenenbaum, "Recovering intrinsic scene characteristics from images," *Computer Vision Systems*, 1978.

[3] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo, "Kindling the Darkness: A Practical Low-light Image Enhancer," *ACM International Conference on Multimedia*, May 2019.

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

[5] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu, "Deep Retinex Decomposition for Low-Light Enhancement," *British Machine Vision Conference*, Aug. 2018.

[6] 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 Trans. on Image Processing*, 2021.

[7] Chongyi Li, Chunle Guo, and Chen Change Loy, "Learning to Enhance Low-Light Image via Zero-Reference Deep Curve Estimation," *IEEE Conference on Computer Vision and Pattern Recognition*, 2020.

[8] Zunjin Zhao, Bangshu Xiong, Lei Wang, Qiaofeng Ou, Lei Yu, and Fa Kuang, "RetinexDIP: A Unified Deep Framework for Low-light Image Enhancement," *IEEE Trans. on Circuits and Systems for Video Technology*, 2021.

[9] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky, "Deep Image Prior," *IEEE Conference on Computer Vision and Pattern Recognition*, Apr. 2018.

[10] Martin J Wainwright and Eero Simoncelli, "Scale Mixtures of Gaussians and the Statistics of Natural Images," in *Advances in Neural Information Processing Systems*, 2000.

[11] Martin J. Wainwright, Eero P. Simoncelli, and Alan S. Willsky, "Random Cascades on Wavelet Trees and Their Use in Analyzing and Modeling Natural Images," *Applied and Computational Harmonic Analysis*, July 2001.

[12] Odelia Schwartz, Terrence J Sejnowski, and Peter Dayan, "Assignment of Multiplicative Mixtures in Natural Images," *Advances in Neural Information Processing Systems*, 2004.

[13] Jiaming Liu, Yu Sun, Xiaojian Xu, and Ulugbek S. Kamilov, "Image Restoration Using Total Variation Regularized Deep Image Prior," *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2019.

[14] Alexander Sagel, Aline Roumy, and Christine Guillemot, "SUB-DIP: Optimization on a subspace with deep image prior regularization and application to superresolution," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2020.

[15] Yossi Gandelsman, Assaf Shocher, and Michal Irani, "Double-DIP: Unsupervised Image Decomposition via Coupled Deep-Image-Priors," *IEEE Conference on Computer Vision and Pattern Recognition*, June 2019.

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

[17] Paszke et al., "PyTorch: An Imperative Style, High-Performance Deep Learning Library," *Advances in Neural Information Processing Systems*, Dec. 2019.

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

[19] Zhou et al. Wang, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Trans. on Image Processing*, 2004.

[20] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang, "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric," *IEEE Conference on Computer Vision and Pattern Recognition*, Apr. 2018.

[21] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "Completely Blind" Image Quality Analyzer," *IEEE Signal Processing Letters*, Mar. 2013.

[22] Ke Gu, Dacheng Tao, Jun-Fei Qiao, and Weisi Lin, "Learning a No-Reference Quality Assessment Model of Enhanced Images With Big Data," *IEEE Trans. on Neural Networks and Learning Systems*, Apr. 2018.

[23] Ke Gu, Weisi Lin, Guangtao Zhai, Xiaokang Yang, Wenjun Zhang, and Chang Wen Chen, "No-Reference Quality Metric of Contrast-Distorted Images Based on Information Maximization," *IEEE Trans. on Cybernetics*, Dec. 2017.

[24] Roger Grosse, Micah K Johnson, Edward H Adelson, and William T Freeman, "Ground truth dataset and baseline evaluations for intrinsic image algorithms," *International Conference on Computer Vision*, 2009.