2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM)

Single Low-light Image Enhancement by Fusing Multiple Sources

Zhuang Feng
Hefei University of Technology
Hefei, China
fengzhuang1994@gmail.com

Zhiyuan Zhou
Hefei University of Technology
Hefei, China
zhouzhiyuan.hfut@gmail.com

Shijie Hao
Hefei University of Technology
Hefei, China
hfut.hsj@gmail.com

Yong Ge
University of Arizona
Tucson, USA
yongge@email.arizona.edu

Richang Hong
Hefei University of Technology
Hefei, China
hongrc.hfut@gmail.com

Meng Wang
Hefei University of Technology
Hefei, China
eric.mengwang@gmail.com

***Abstract*—The visual quality of photographs can be seriously affected by various low-light conditions. The image enhancement methods based on fusion achieve satisfying results. However, this poses challenges to the situation that there are no fusion sources and only a single low-light image is at hand. In this paper, we propose a single low-light enhancement method by fusing multiple sources, including the original image and its intermediate enhancements generated by a simplified Retinex model. We combine these sources at a mid-level image representation based on a patch-based decomposition model. Compared with other state-of-the-art methods, visual and quantitative results show that our method effectively improves visual effects in terms of lightness, color harmony and vividness.**

***Index Terms*—Image Enhancement, Low-light Image, Fusion, Patch Decomposition**

I. INTRODUCTION

Taking photographs has become a common activity of our daily life. However, due to poor lightness or casual shooting, we often produce unsatisfied photographs such as images with low contrast or unclear details shown in Fig.1. Automatic low-light image enhancement techniques are thus highly desirable in many real-world applications.

By acquiring High Dynamic Range (HDR) image data, the low-light issue can be solved by a tone mapping technique that compresses the range pixel intensities [1]. Alternatively, the family of Multiple Exposure Fusion (MEF) methods [2] [3] [4] also tackles the low image contrast issue by fusing multiple input images with different exposures. These methods achieve good performance on the task of enhancing image contrast. These methods need advanced imaging mechanisms, such as generating a much larger range of pixel value or generating different value for a pixel. Nevertheless, in many situations, we only have a single image composed of 8-bit pixels at hand, which poses challenges for the above methods.

Research on single-image enhancement methods has also been conducted for many years. The histogram-based [5] and Retinex-based [6] methods perform well in enhancing the lightness of an image. Nevertheless, as there is usually no hint on how to tune the enhancement strengths, they often suffer from unnatural effects, such as over-exaggerated details and color distortion. To solve this issue, it is useful to produce homologous sources from the input with camera response function estimation [7] or intermediate enhancements [8].

Following this technical roadmap, we propose a single low-light image enhancement method, of which the flowchart is shown in Fig.2. Our method is highlighted in its capacity of obtaining enhanced images with good lightness and color. Two factors contribute to this performance. First, we generate multiple enhancements with a simplified Retinex model directly parameterized by the enhancing strength. Second, based on these enhancing sources, a model of image decomposition and fusion at the patch level is adopted, which simultaneously relieves artifacts and improves the color of the enhanced result. Both the qualitative and quantitative comparison with several state-of-the-art methods validate the effectiveness of our method.

In the following, we briefly introduce the related research on low-light image enhancement. As for the single-input methods, they are usually built upon the physical assumption that an image scene can be seen as the multiplication of a reflectance layer and an illumination layer [9] [10]. A typical framework of is to enhance the illumination layer separated from an input image. The key lies in the intrinsic image decomposition, which can be time-consuming. Differently, the LIME method [6] utilizes a simplified Retinex model, in which the illumination layer is initially estimated by the MaxRGB technique. The LIME method is very computationally efficient. Dong et al. [11] use the haze-removal technique by assuming that the inverse of a low-light image resembles a hazy one. The dehazing technique is applied on the inverse image to produce the low-light enhancement result. A common issue of these single-image enhancement methods is that over- or under-enhanced results often occur, as the enhancement parameters are usually set in an ad-hoc way.

Fig.1. 1st row: low-light images; 2ed row: our enhanced images

The research was supported by the National Nature Science Foundation of China (Grant No. 61772171)

978-1-5386-5321-0/18/$31.00 ©2018 IEEE

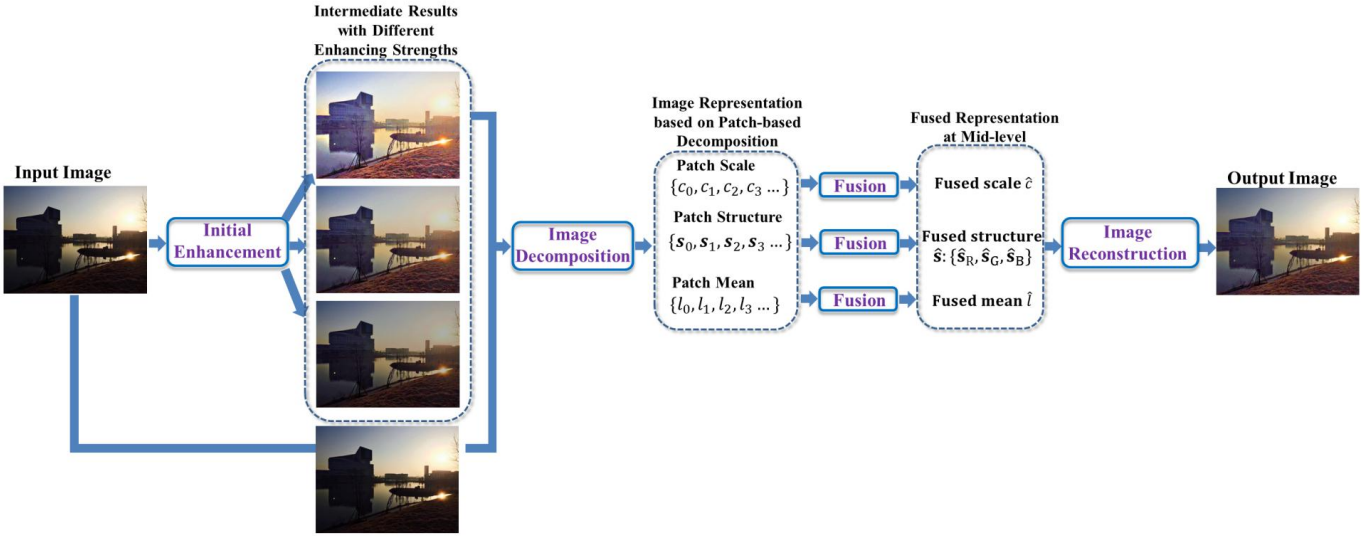

Fig.2. The framework of our method.

As for the methods based on multi-source fusion, the above issue can be well addressed by using multiple inputs with different exposures [2] [3]. The key factor is to design the fusion model. In [2], each source image is decomposed into three parts. The fusion is conducted independently on each part, and the output is obtained by composing these fused parts again. We adopt this technical roadmap in our research. Similarly, by decomposing an image into the detail layer and structure layer, Kou et al. [4] propose an exposure fusion method based on multi-scale fusion process.

The fusion-based methods mentioned above meet fusion source insufficiency when only a single input image is available. One feasible roadmap is to generate several intermediate enhanced results as the fusion sources. Ying et al. propose a camera response model [7] to simulate the multi-exposure process, and linearly combine them [12] [13]. Fu et al. [8] generate multiple enhancements based on different enhancement models, and fuse them through a multi-resolution image pyramid. As the fusion process of these methods is pixel-wise, artifacts are often introduced. Of note, the color channels are often left unchanged in these methods.

II. OUR METHOD

In this section, we introduce our method for low-light image enhancement. As shown in **Fig. 2**, the technical roadmap of our method is simple and straightforward. In general, the input image $\mathbf{I}_0 \in \Re^{3N^2}$ is firstly enhanced with a same enhancing model with different filtering strengths. Along with $\mathbf{I}_0$, the intermediate enhanced images $\{\mathbf{I}_1,...,\mathbf{I}_k\}$ are represented by using a patch-based decomposition model. The patches from the sources $\{\mathbf{I}_0,\mathbf{I}_1,...,\mathbf{I}_k\}$ are fused respectively for each image component, i.e. c,$\mathbf{s}$,and l. The final result can be obtained by reconstructing the fused $\hat{c}$,$\hat{\mathbf{s}}$,and $\hat{l}$. The details are presented in the following.

To generate multiple initial enhancements, we propose a fast simplified Retinex model:

$$\mathbf{I}_\epsilon = \mathbf{I}_0/(\mathbf{T} + \epsilon) \tag{1}$$

In the model, $\mathbf{I}_0$ is the original image, $\mathbf{T}$ is the refined illumination map, and ϵ is the enhancing strength that parameterizes the initially enhancements. We use a computationally efficient technique to produce $\mathbf{T}$. For each pixel $p \in \mathbf{I}_0$, the max value of its RGB channels is firstly extracted:

$$\mathbf{T}_0(p) = \max_{c\in\{R,G,B\}} \mathbf{I}^c(p) \tag{2}$$

Different from [6], we just apply the off-the-shelf fast guided filter [15] to refine $\mathbf{T}_0$, and obtain a piece-wise constant map $\mathbf{T}$. The spatial and range parameters of the guided filter are constantly set as $0.1 \cdot \min\{\text{ImWidth,ImHeight}\}$ and 0.01 throughout our experiments. By imposing different enhancing strengths $\{\epsilon_1,...,\epsilon_{K-1}\}$ on $\mathbf{I}_0$, the produced sources $\{\mathbf{I}_1,...,\mathbf{I}_K\}$ are able to provide scalable appearances of a same scene, as exemplified in the left part of **Fig.2**.

We totally have $K+1$ sources $\{\mathbf{I}_0,\mathbf{I}_1,...,\mathbf{I}_K\}$ for the fusion process. For an image patch $\mathbf{P}$ of each source, we adopt the patch-based image decomposition [2]:

$$\begin{aligned}\mathbf{P} &= \|\mathbf{P} - \mu_\mathbf{P}\| \cdot \frac{\mathbf{P} - \mu_\mathbf{P}}{\|\mathbf{P} - \mu_\mathbf{P}\|} + \mu_\mathbf{P} \\ &= \left\|\tilde{\mathbf{P}}\right\| \cdot \frac{\tilde{\mathbf{P}}}{\|\tilde{\mathbf{P}}\|} + \mu_\mathbf{P} = c \cdot \mathbf{s} + l\end{aligned} \tag{3}$$

In Eq. 3, for each squared patch $\mathbf{P}$, we stack its RGB channels together into a $3M^2$-length column vector (M is the patch width). Here $\|\cdot\|$ is the L2 norm, c is the patch scale, $\mathbf{s}$ is the patch structure, and l is the patch mean intensity. The decomposed c, $\mathbf{s}$, and l can be seen as a mid-level representation of $\mathbf{I}_k$. In the following, different fusion models are separately applied on the three components.

First, an nonlinear max-fusion is applied to the patch scale c:

$$\hat{c} = \max_{1\le k\le K} c_k \tag{4}$$

Second, a linear weight fusion is constructed for the patch structure $\mathbf{s}$:

$$\bar{\mathbf{s}} = \frac{\sum_{k=1}^{K} c^\rho \mathbf{s}_k}{\sum_{k=1}^{K} \mathbf{s}_k} \tag{5}$$

From Eq.5, we observe that the fused $\bar{\mathbf{s}}$ is jointly determined by $\{\mathbf{s}_k\}$ of the multiple sources. The fusion weights are determined by the exponential of patch scales $\{c_k\}$, where $\rho \ge 0$ is a hyper-parameter. The obtained $\bar{\mathbf{s}}$ is further

normalized as $\hat{\mathbf{s}} = \bar{\mathbf{s}}/\|\bar{\mathbf{s}}\|$. With Eq.4 and Eq.5, the fusion models tend to weigh more on strong patches, and still consider the impact of weak patches. Third, we also use a weighted linear fusion for the patch mean:

$$\hat{l} = \frac{\sum_{k=1}^{K} L(\mu_k^0, l_k) l_k}{\sum_{k=1}^{K} L(\mu_k^0, l_k)} \quad (6)$$

In Eq.6, $L(\cdot,\cdot)$ describes how well the lightness of l_k is in $\mathbf{I}_k$:

$$L(\mu_k^0, l_k) = \exp\left(-\frac{(\mu_k^0 - 0.5)^2}{2\sigma_g^2} - \frac{(l_k - 0.5)^2}{2\sigma_l^2}\right) \quad (7)$$

where μ_k^0 is the global mean of $\mathbf{I}_k$, σ_g and σ_l control the spreads of the Gaussian distribution tails.

Finally, we disconnect the stacked $\hat{\mathbf{s}}$ back into the RGB channels, and reconstruct them with the obtained $\hat{c}$ and $\hat{l}$ according to Eq.3:

$$\hat{\mathbf{P}}_\phi = \hat{c} \cdot \hat{\mathbf{s}}_\phi + \hat{l} \quad (8)$$

where $\phi \in \{R,G,B\}$ enumerates the three color channels. We use a sliding window with the stride length of $B = \lfloor M/2 \rfloor$ to reconstruct each patch of the result image, and the pixels in overlapping regions are averaged. In this way, the reconstructed $\mathbf{I}_f$ is taken as the final result. In our research, the fusion parameters $M, \rho, \sigma_g, \sigma_l$ are typically set as in [2].

The complexity of our method is analyzed as follows. The complexity of multiple source generation is $O(KN^2)$. All the later stages of patch decomposition, fusion and reconstruction are achieved with a sliding patch window going through the whole pixel grid. So the complexity of these part is $O(CM^2N^2)$, where C is a constant.

III. EXPERIMENTS

In experiments, we validate our method by comparing it with several state-of-the-art methods, including LIME [6], DEHAZE [11], MSRCR [14], MF [8], and BIMEF [13]. All the codes were run on a laptop with 8G ROM and a 2.6G Hz CPU. For our method, we empirically set $\{\epsilon_1,...,\epsilon_{K-1}\}$ as $\{0.12,0.2,0.4\}$, which indicate strong, medium and weak enhancing strengths. The parameters of all the methods for comparison were set by default according to the codes provided by their authors. As shown in Fig. 3, the images for experimental validation are collected from the Internet.

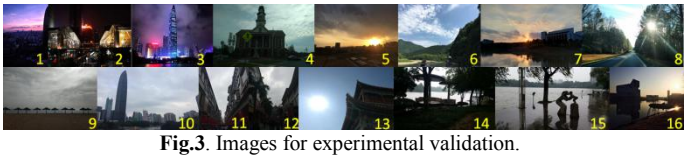

Fig.3. Images for experimental validation.

First, we conduct qualitative visual comparison. In Fig. 4, we show several enhancing results of our method and their competitors, as well as the zoomed-in regions. In general, the non-fusion methods (LIME, DEHAZE, MSRCR) generate unnatural effects by over-exaggerating the texture, edge or global mean intensity. In contrast, the fusion-based methods (MF, BIMEF, and our method), largely avoid this issue, as the fusion process is able to automatically harmonize different sources. Among the fusion-based group, our method introduces much less artifacts (e.g. the layered effect of the zoomed-in night sky region in the 3rd row of **Fig. 4**) and produces a vivid color (e.g. the zoomed-in region of the 1st and 2nd row of **Fig. 4**). The reasons are two-fold. First, the patch-based computation makes our method robust to artifacts to some extent. Second, as the RGB channels of $\{\mathbf{I}_0,\mathbf{I}_1,...,\mathbf{I}_k\}$ are jointly considered in the decomposition and fusion, our method is able to improve the color distribution.

Then, we adopt several quantitative measures to evaluate the effectiveness of our method. We first use a full-reference DRIM metric [16], of which the results are shown in **Fig. 5**. Two low-light images are randomly chosen from the public BID dataset [17], in which their reference images with normal image contrast are available as the ground truth. The DRIM result is visualized in color at a pixel-level, where fewer colors mean a better result. Generally, we can also observe the similar trend that fusion-based group outperforms the group with no fusion. Specifically, among the fusion-based ones, our method achieves slightly better DRIM result than the other two competitors.

To evaluate the performance in terms of visual aesthetics, we use an aesthetic ranking network [18] to grade all the images shown in Fig.3 and their enhancements. First, the well-trained network facilitates the performance evaluation of our task, as the ground truth of low-light enhancement is usually unavailable in real-world applications. Second, the network is able to grade the fine-grained attributes of an image, including color harmony, color vividness, and lightness. As for the score value, a larger one indicates a better aesthetic attribute, and a negative one indicates the visual appearance is below the average level. We have the following observations from Table 1. First, for most images (except the 3rd image in Fig.3), the enhancing models significantly improve the visual aesthetics in terms of the three attributes. Second, compared with MF and BIMEF, our method achieves the best performance in the mass.

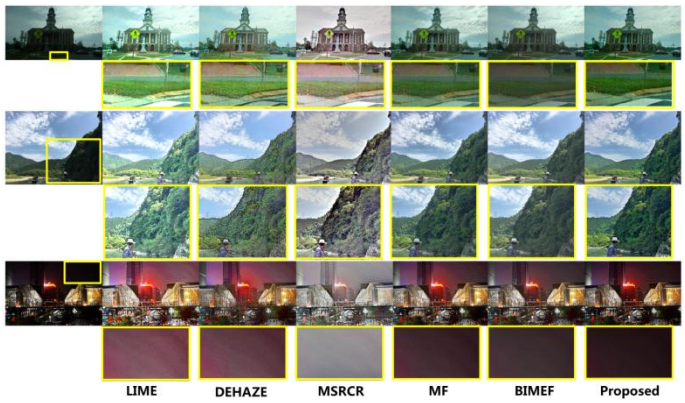

Fig.4. Visual comparison between the state-of-the-art methods and ours (Better with a good screen brightness and a large view).

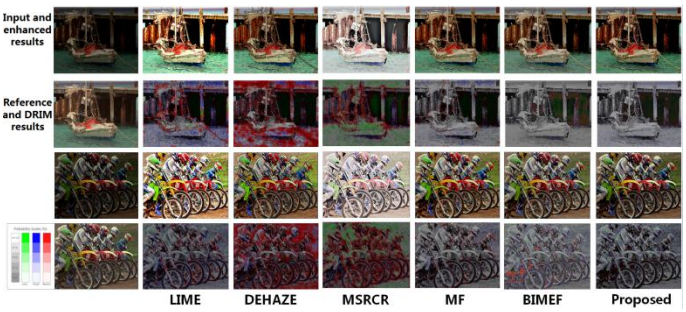

Fig.5. Comparison of contrast distortion based on DRIM [15].

TABLE I. Quantitative aesthetic scores (color harmony, color vividness, lightness)

	Input	MF	BIMEF	Proposed
1	(0.3886,0.1425,0.6873)	(0.3719,0.3051,0.7277)	(0.3746,0.2445,0.7013)	(**0.3969**,**0.3640**,**0.7329**)
2	(0.1973,0.3276,0.6655)	(**0.2953**,0.5674,0.9002)	(0.2674,**0.6069**,**0.9554**)	(0.2187,0.5597,0.8971)
3	(**0.2135,0.3685,0.3798**)	(0.1476,0.3290,0.2698)	(0.1294,0.2722,0.1745)	(0.0950,0.2632,0.1872)
4	(-0.0627，-0.3697,-0.5051)	(**0.4264**,-0.0039,-0.0396)	(0.2680,-0.0847,-0.1342)	(0.3797,**0.0499**,**-0.0252**)
5	(0.2157,-0.1165,0.4714)	(0.3511,0.1280,0.6728)	(0.2701,-0.0291,0.5352)	(**0.3627**,**0.1510**,**0.7119**)
6	(0.1442,-0.3217,-0.1937)	(0.3828,-0.2192,-0.1299)	(0.3070,-0.2381,-0.1221)	(**0.6332**,**-0.0261**,**-0.0933**)
7	(0.1096,-0.1401,0.1201)	(0.2212,0.1237,0.4292)	(0.2523,0.0113,0.3365)	(**0.2676**,**0.2842**,**0.5439**)
8	(0.0286,-0.3357,-0.3311)	(0.0847,-0.2953,-0.2191)	(0.0647,-0.3693,-0.3016)	(**0.2822**,**-0.2396**,**-0.1202**)
9	(-0.0241,-0.6724,-0.6598)	(0.0612,-0.6044,-0.2784)	(0.0406,-0.6377,-0.3742)	(**0.1276**,**-0.5040**,**-0.1423**)
10	(-0.0212,-0.5118,-0.7583)	(0.1012,-0.3971,**-0.5247**)	(0.0238,-0.4562,-0.6681)	(**0.1215**,**-0.3342**,-0.5312)
11	(0.1755,-0.3061,-0.1679)	(**0.1875**,0.2001,-0.0909)	(0.1541,-0.1144,-0.1849)	(0.1468,**0.2707**,**-0.0788**)
12	(-0.0116,-0.5269,-0.3208)	(-0.0162,-0.1052,-0.4020)	(-0.0502,-0.3605,-0.4643)	(**0.0687**,**0.1356**,**-0.2481**)
13	(0.0651,-0.3304,-0.3834)	(0.1644,-0.0260,**-0.1392**)	(0.1529,-0.2575,-0.1646)	(**0.2345**,**0.1743**,-0.1797)
14	(-0.1921,-0.4883,-0.7497)	(0.6686,-0.0294,-0.1053)	(0.6414,-0.0206,-0.0953)	(**0.6895**,**0.0565**,**-0.0586**)
15	(0.2272,-0.4494,-0.2115)	(0.4312,**-0.2382**,-0.1168)	(0.3570,-0.3319,-0.1805)	(**0.4620**,-0.2471,**-0.0872**)
16	(-0.1655,-0.5515,-1.0002)	(-0.0226,-0.2759,-0.2123)	(-0.0349,-0.4217,-0.4998)	(**-0.0013**,**-0.1091**,**-0.1547**)

IV. Conclusion and discussions

In this paper, we propose a low-light image enhancement method based on the mid-level fusion. The qualitative and quantitative experiments validate the effectiveness of our method. Of note, the framework of our method is quite open. Currently, we generate all the sources by the same model with different enhancing strengths. This might limit the final performance, as the homogenous enhancements from the simplified Retinex model might have common pros and cons. We plan to introduce different filtering models that better mutually benefit each other.

References

[1] K. Ma, H. Yeganeh, K. Zeng, Z. Wang, "High Dynamic Range Image Compression by Optimizing Tone Mapped Image Quality Index," IEEE TIP, 24(10):3086–3097, 2015

[2] K. Ma, H. Li, H. Yong, Z. Wang, D. Meng, L. Zhang, Robust Multi-Exposure Image Fusion: A Structural Patch Decomposition Approach, IEEE TIP, 26(5): 2519-2532, 2017

[3] M. Bertalmio, S. Levine, Variational Approach for the Fusion of Exposure Bracketed Pairs, IEEE TIP, 22(2): 712–723, 2013

[4] F. Kou, Z. Wei, W. Chen, X. Wu, C. Wen, Z. Li, Intelligent Detail Enhancement for Exposure Fusion, IEEE TMM, 20(2): 484-495, 2018

[5] C. Lee, C. Lee, C. Kim, Contrast Enhancement based on Layered Difference Representation of 2D Histograms, IEEE TIP, 22(12): 5372–5384, 2013

[6] X. Guo, Y. Li, H. Ling, LIME: Low-light Image Enhancement via Illumination Map Estimation. IEEE TIP, 26(2): 982-993, 2017

[7] Z. Ying, G. Li, Y. Ren, R. Wang, W. Wang, A New Low-Light Image Enhancement Algorithm Using Camera Response Model. ICCVW, 2017

[8] X. Fu, D. Zeng, Y. Huang, Y. Liao, X. Ding, J. Paisley, A Fusion-based Enhancing Method for Weakly Illuminated Images. Signal Processing, 129: 82-96, 2016

[9] X. Fu, D. Zeng, Y. Huang, X. Zhang, X. Ding, A Weighted Variational Model for Simultaneous Reflectance and Illumination Estimation. CVPR, 2016

[10]H. Yue, J. Yang, X. Sun, F. Wu, C. Hou, Contrast Enhancement Based on Intrinsic Image Decomposition, IEEE TIP, 26(8): 3981-3994, 2017

[11] X. Dong, G. Wang, Y. Pang, Fast Efficient Algorithm for Enhancement of Low Lighting Video. ICME, 2011

[12] Z. Ying, G. Li, Y. Ren, R. Wang, W. Wang, A New Image Contrast Enhancement Algorithm Using Exposure Fusion Framework. CAIP, 2017

[13] Z. Ying, G. Li, W. Gao, A Bio-Inspired Multi-Exposure Fusion Framework for Low-light Image Enhancement. ArXiv, abs/1711.00591, 2017

[14] D. Jobson, Z. Rahman, G. Woodell, A Multi-scale Retinex for Bridging the Gap between Color Images and the Human Observation of Scenes, IEEE TIP, 6(7): 965–976, 1997

[15] K. He, J. Sun, Fast Guided Filter. ArXiv, abs/1505.00996, 2015

[16] T. Aydin, R. Mantiuk, K. Myszkowski, H. Seidel, Dynamic range independent image quality assessment, ACM TOG, 27(3): Article 69, 2008

[17] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, TID2008: A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics, Advances of Modern Radioelectronics, 10: 30-45, 2009

[18] S. Kong, X. Shen, Z. Lin, R. Mech, C. Fowlkes, Photo Aesthetics Ranking Network with Attributes and Content Adaptation. ECCV, 2016