Progressive Image Enhancement under Aesthetic Guidance

Xiaoyu Du
University of Electronic Science and Technology of China
Chengdu, China
duxy.me@gmail.com

Xun Yang
National University of Singapore
Singapore, Singapore
xunyang@nus.edu.sg

Zhiguang Qin
University of Electronic Science and Technology of China
Chengdu, China
qinzg@uestc.edu.cn

Jinhui Tang*
Nanjing University of Science and Technology
Nanjing, China
jinhuitang@njust.edu.cn

ABSTRACT

Most existing image enhancement methods function like a black box, which cannot clearly reveal the procedure behind each image enhancement operation. To overcome this limitation, in this paper, we design a progressive image enhancement framework, which generates an expected "good" retouched image with a group of self-interpretable image filters under the guidance of an aesthetic assessment model. The introduced aesthetic network effectively alleviates the shortage of paired training samples by providing extra supervision, and eliminate the bias caused by human subjective preferences. The self-interpretable image filters designed in our image enhancement framework, make the overall image enhancing procedure easy-to-understand. Extensive experiments demonstrate the effectiveness of our proposed framework.

CCS CONCEPTS

• **Computing methodologies** → **Computer vision**; **Image processing**.

KEYWORDS

Image Manipulation, Image Enhancement, Aesthetic Assessment

ACM Reference Format:
Xiaoyu Du, Xun Yang, Zhiguang Qin, and Jinhui Tang. 2019. Progressive Image Enhancement under Aesthetic Guidance. In *International Conference on Multimedia Retrieval (ICMR '19), June 10–13, 2019, Ottawa, ON, Canada.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3323873.3325055

1 INTRODUCTION

Image enhancement encompasses the processes of altering images, aiming to increase the image qualities in some aspects especially in aesthetic metrics. This task could be manually conducted with some professional tools, such as Photoshop, Lightroom, etc. However, due to the complexity of these tools, manually enhancing images is a time-consuming task and also rely on strong domain knowledge. Therefore, automatic image enhancement techniques are highly desired to assist or even replace time-consuming manual processing.

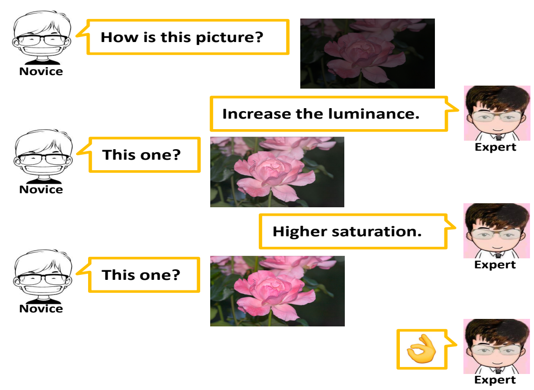

Figure 1: An example that a novice enhances an image under the guidance of an expert.

As a key technique in image processing, neural networks are widely used in image enhancement models [2, 4, 10]. However, they usually process the images in an end-to-end fashion and function like a "black box". It is difficult to understand the overall enhancing procedure. In contrast to existing "black box" models, we aim to develop an easy-to-interpret image enhancement framework that can reveal the image enhancement steps adopted for an image. Figure 1 illustrates an example of the procedure that a novice enhances images under the guidance of a human expert. As shown in Figure 1, the novice adjusts the lotus image in a heuristic pattern with several steps. At each step, the novice shows his retouched image and the expert gives an improvement suggestion to make the image better. The novice would obtain an optimal adjustment strategy with a series of expert suggestions. By summarizing the suggestions, the adjustment strategy becomes interpretable.

Motivated by such a novice-expert processing procedure in Figure 1, we propose a **P**rogressive **I**mage **E**nhancement framework under **A**esthetic guidance (PIEA), which consists of two key modules: Image Enhancer (IE) and Image Guider (IG). In PIEA, we make Image Enhancer (IE) and Image Guider (IG) play the *conversation* between the novice and the expert in a heuristic process. Specifically, IE is implemented by a group of resolution-independent differentiable image filters in [7] that represent conventional image post-processing operations, which is self-interpretable and understandable to users, unlike the black-box solutions of most CNNs

*Corresponding Author.

ICMR '19, June 10–13, 2019, Ottawa, ON, Canada

ACM ISBN 978-1-4503-6765-3/19/06...$15.00
https://doi.org/10.1145/3323873.3325055

based models. IG is implemented by a well-trained aesthetic assessment model on a separate aesthetic dataset, which contains abundant aesthetic domain knowledge. In our experiments, IG is practically specified to NIMA (Inception-v2) [18]. The parameters in NIMA are well trained on the AVA dataset [14] and they are fixed during the application of PIEA. Given an image, IG yields an aesthetic score that represents the public aesthetic awareness. In our PIEA framework, we assume that a well-enhanced image could have a high aesthetic score from IG, or otherwise. Then IG provides the supervision information to IE for adjusting the parameters of image filters. Except for the aesthetic guidance, our proposed PIEA does not require any supervision from human experts due to two reasons: (1) Image enhancement is a subjective task. Different human experts usually have different preferences in the retouching style. Such a human bias may confuse the model. (2) Collecting high-quality paired training data from human experts is also costly. Besides, PIEA also does not require any training process for learning parameters. It is a *plug and play* model that processes each input by several rounds of parameter searching of image filters under the supervision of aesthetic assessment model. Specifically, PIEA progressively enhances the images via several conventional image editing steps. In each step, there are three phases, 1) IE retouches the raw image with current filter settings and shows the retouched image to IG, 2) IG evaluates the current retouched image and gives IE a feedback to increase image quality, and 3) IE updates its filter settings by refreshing the filter parameters. After several steps, IE obtains an optimal adjustment strategy for the raw image.

There are two key advantages of PIEA:

(1) **Easy-to-interpret:** For each input, PIEA can reveal the adopted image enhancement steps dominated by a group of resolution-independent differentiable image filters that represent conventional image editing operations, which is self-interpretable and understandable to users.
(2) **Plug and Play:** PIEA is a *Plug and Play* model that does not require any training process. For each input, it performs several rounds of parameter searching of image filters under the supervision of aesthetic assessment model. It is a light-weight model which can run on a mobile platform efficiently.

2 RELATED WORK

Recent image enhancement methods have two folds, 1) enhancing the images in a specific property, such as exposure [11] and contrast [6], and 2) directly constructing the retouched image with overall adjustments [3, 4, 8–10]. In order to make the overall adjustments interpretable, many methods redefine the enhancing model by assembling a sequence of interpretable image filters [5, 16, 17]. Training these methods requires paired data. However, due to the complex human cognition, the expected retouched images should not be unique in realities. Moreover, collecting high-quality paired data (not just distort the original image) is costly. Therefore, EXPOSURE [7] proposes a GAN-based unsupervised method to decouple the pairs.

We here introduce aesthetic assessment models, which quantifies images in aesthetic metrics, i.e., the human cognition-based metrics [12, 13]. We believe the aesthetic assessment models (e.g. NIMA [18]) somehow have obtained some inherent rules due to two observations: 1) in dataset AVA [14], the accuracy NIMA [18] predicting the image qualities achieves 81.51%, and 2) in dataset FiveK [1], NIMA [18] gives similar ratings (averagely around 4.7) to the image sets retouched by experts, while it gives lower ratings (averagely around 4.3) to the raw image set.

The plug & play pattern is inspired by PPGN [15]. Since the progressive method benefits the comprehensions on enhancing strategy, this iterative approach seems much suitable for image enhancement methods.

3 METHODOLOGY

Figure 2 exhibits the overview of our model. Separated by the dashed lines, the left part is the Image Enhancer (IE), the middle part is the Image Guider (IG), and the right part is the adjusting target. We then elaborate IE and IG separately, and illustrate the enhancing procedure.

3.1 Image Enhancer

Image Enhancer (IE) takes in the raw image denoted as $\mathbf{R}$ and generates the retouched image $\mathbf{A}$ via N pixel-independent and differentiable image filters. Let $\mathcal{F}_1, \mathcal{F}_2, ..., \mathcal{F}_N$ indicate the N filters, and π_i indicate the necessary parameters for filter $\mathcal{F}_i$, the retouched image $\mathbf{A}$ is obtained via Equation 1,

$$\begin{aligned} \mathbf{X}_1 = \mathcal{F}_1(\mathbf{R}|\pi_1), \mathbf{X}_i = \mathcal{F}_i(\mathbf{X}_{i-1}|\pi_i), \\ \mathbf{A} = \mathbf{X}_N = \mathcal{F}_N(\mathbf{X}_{N-1}|\pi_N), \end{aligned} \tag{1}$$

where $\mathbf{X}_i$ indicate the output image of the i-th filter. Through the joint N filters, the raw image $\mathbf{R}$ is retouched to $\mathbf{A}$. In this work, following EXPOSURE [7], we use the filters on exposure, gamma, white balance, saturation, tone, contrast, black and white, and color.

3.2 Image Guider

Image Guider (IG) takes in the retouched image $\mathbf{A}$ and scores the image in aesthetic metric. Let $\mathbf{X}$ be an image, IG (*i.e.,* an aesthetic assessment model) scores the image with Equation 2. The score S indicates the aesthetic quality of $\mathbf{X}$, the higher the better.

$$S = IG(\mathbf{X}). \tag{2}$$

In order to guide the adjustment of IE, IG would give an improvement suggestion on the current retouched image to increase the score S. Let $\bar{S}$ denote the max score IG would give, the suggestion from IG aims to reduce the gap between current score S and the theoretical max score $\bar{S}$ as,

$$L = (\bar{S} - S)^2. \tag{3}$$

In this work, IG is specified to a well-trained NIMA (Inception-v2) model [18], which achieves the accuracy of 81.51% on AVA dataset [14]. The weights of NIMA are fixed in our framework. For each input image, NIMA predicts a vector $\mathbf{s} \in [0, 1]^{10}$ indicating the probability of the integer scores in $\{1, ..., 10\}$. We score an image with the score expectation computed by

$$S = \sum_{i=1}^{10} i \cdot s_i. \tag{4}$$

Thus S is a float number in $[1, 10]$ and the theoretical max score $\bar{S}$ is 10.

3.3 Progressive Enhancing Procedure

The process of improving the raw image R is equivalent to the process of reducing L. Since the weights of IG are fixed, $\mathbf{A}$ is the

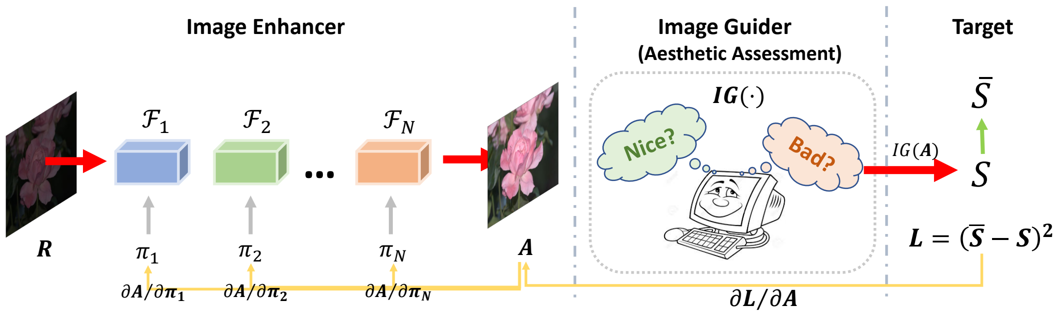

Figure 2: Overview of PIEA.

only choice to affect the score S. In order to maximize the reduction of L, $\mathbf{A}$ should be updated along the direction of $\frac{\partial L}{\partial \mathbf{A}}$. Subsequently, the retouched image $\mathbf{A}$ relies on the filters which are controlled by the filter parameters π_*. Since all the filters are differentiable, the π_* should be updated along the direction of $\frac{\partial L}{\partial \pi_*}$. To summarize, in order to reduce L, the π_* should be updated as,

$$\pi_i = \pi_i - \lambda \cdot \frac{\partial L}{\partial A}\frac{\partial A}{\partial \pi_i}, \forall i = 1, 2, ..., N. \tag{5}$$

where λ is the learning rate of π_i.

Algorithm 1 illustrates the entire enhancing procedure. The only input for this system is the raw image $\mathbf{R}$. At the beginning, all the filter parameters π_* are initialized, followed by M iterations. In each iteration, there are three key phases:

(1) Generating the retouched image $\mathbf{A}$ with the certain π_*,
(2) Scoring the retouched image $\mathbf{A}$ with S, and computing the target variable L,
(3) Updating the parameters π_* according to Equation 5.

The enhancing process has several iterations. In each iteration, there would be an updating suggestion for the set of π_* to generate a better-retouched image. After all the iterations have done, the set of π_* indicates the expected enhancing strategy.

Algorithm 1: Enhancing Procedure

```
Input: R
Output: {A, π_1, π_2, ··· , π_N}
Initialize {π_1, π_2, ··· , π_N};
for iteration ← 1 to M do
    /* Phase 1: Enhance the raw image R (Eq. 1)                  */
    X_0 = R;
    for i ← 1 to N do  X_i = F_i(X_{i-1}) ;
    A = X_N;
    /* Phase 2: Scoring A (Eq. 2) and computing L (Eq. 3)       */
    S = IG(A) ;
    L = (S̄ − S)^2 ;
    /* Phase 3: Update the parameters π_* (Eq. 5)               */
    for i ← 1 to N do  π_i = π_i − λ · (∂L/∂A)(∂A/∂π_i) ;
end
return {A, π_1, π_2, ··· , π_N};
```

4 EXPERIMENTS

4.1 Settings

Dataset. We select FiveK [1] as our dataset. FiveK is established above 5,000 digital images taken with SLR cameras. In FiveK, each image has six copies, including the raw image and five copies retouched by five experts (noted as A, B, C, D, and E) respectively. Usually, the raw images are treated as low-quality images while the retouched images are treated as high-quality images. Following the previous work[1], we take the images retouched by expert C as the ground truth if there is no specific explanation. Following EXPOSURE [7], we split the 5,000 images to training set (2,000 images), validation set (2,000 images) and testing set (1,000 images). It is notable that there is no training phase for PIEA, thus only the validation set and testing set are used to obtain the results of PIEA.

Parameter Settings. Following the previous work[1], the resolutions of the raw images are 500×333. Since the implementations of IE and IG are independent of image resolution, we do not resize the raw images. If there is no special explanation, the hyper-parameters λ and M are set to 10^{-3} and 20, respectively. To ensure that the enhancing process starts from the raw images, the filters with initial parameters would not change their input image, except the exposure filter. In our experiment, we roughly initialize the parameter of exposure filter with 1.0 to get out of the too dark condition, because the aesthetic model would achieve extraordinary weak performance on the very dark and very bright images. One possible reason is that the aesthetic dataset AVA [14] lacks these two kinds of images. In the future, we may explore a proper way to initialize the luminance and improve the aesthetic model.

4.2 Quantative Comparison

We compare our model PIEA with the state-of-the-art models in quantitative metrics. We select three baselines. EXPOSURE [7] is the strong baseline also enhancing the images with filters. CycleGAN [19] can learn image transforming strategy with GAN automatically, and Pix2pix [10] is a supervised method on paired image data. Following the previous work [7], we measure the patch-based histogram intersection of luminance, contrast, and saturation between the model outputs and the ground truth. The results in Table 1 demonstrates that PIEA outperforms the baselines, except the Luminance of Pix2pix, which is the performance of a supervised method trained with paired images. PIEA also outperforms the baselines on other experts in MIT-Adobe FiveK. Due to space limitation, the results are not listed here.

4.3 Performances

Figure 3 exhibits the retouched images. The images of EXPOSURE are from the GitHub of EXPOSURE[2]. The images retouched by EXPOSURE present a bit overexposure or underexposure in a high

[1] https://github.com/yuanming-hu/exposure/wiki/Preparing-data-for-the-MIT-Adobe-FiveK-Dataset-with-Lightroom
[2] https://github.com/yuanming-hu/exposure/issues/20

Table 1: Quantitative results on expert C in MIT-Adobe FiveK

Approach	Luminance	Contrast	Saturation
PIEA	86.4%	87.8%	87.8%
EXPOSURE [7]	71.3%	83.7%	69.7%
CycleGAN [19]	61.4%	71.1%	82.6%
Pix2pix [10]	92.4%	83.3%	86.5%

probability. In contrast, PIEA performs well in most cases. The bad images, such as the woman image, go underexposure due to the very darkness of the raw images. Roughly increasing the luminance is insufficient to lighten these images. According to the weak performance of NIMA on the dark images, it is hard to guide the image to a nice performance. Besides, PIEA seems to prefer bright colors. The slightly exaggerated colors make the images seem like photographic artworks.

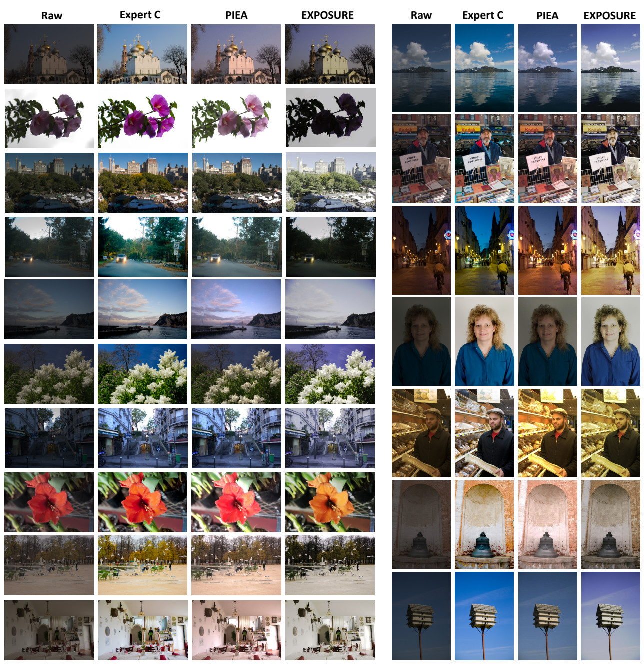

Figure 3: A display of the raw images from FiveK and their corresponding retouched images enhanced by Expert C, PIEA and EXPOSURE.

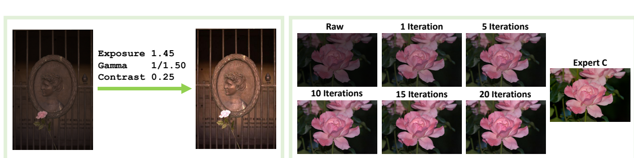

Figure 4: The interpretable adjusting strategy (left) and the progressive enhancing process (right).

Figure 4 demonstrates the interpretable adjusting strategy and the progressive process of PIEA. The left part of Figure 4 presents the key filters to enhance the left image. With the name and the parameters of the filters, anyone who knows a bit about the post-processing concepts could fully reproduce the enhancing process.

The right part of Figure 4 shows the phases in the enhancing process. PIEA enhances the raw image progressively. The later iteration would generate a better image. After 20 iterations, the retouched image has been close to the image retouched by expert C.

4.4 Impact of Learning Rate

We here explore the impact of learning rate λ. By enhancing the images with different learning rate λ, we compute the histogram intersections between the retouched images and the ground truth. The top subfigure in Figure 5 shows that the learning rate 10^{-3} outperforms the smaller and the larger learning rates. The possible reason is, the small learning rate is insufficient to obtain the nice parameters in 20 iterations, while the large learning rate makes the parameters hard to converge to an optimal set. The bottom subfigure in Figure 5 displays the retouched images with different learning rate. It is obvious that the large learning rate causes an over adjustment.

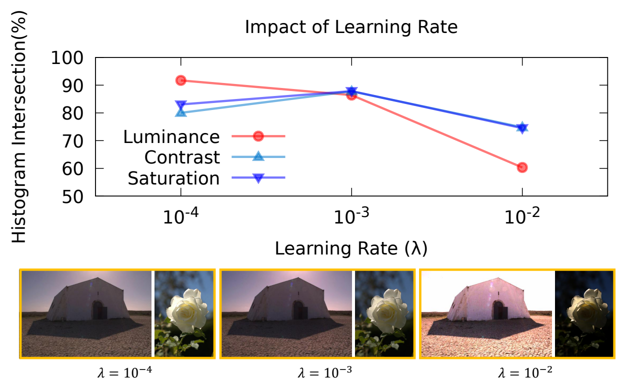

Figure 5: Impact of learning rate.

5 CONCLUSION

We presented a novel progressive image enhancement framework PIEA, which enhances the raw images in a progressive fashion under aesthetic guidance. The introducing of the aesthetic model promotes the improvement of image enhancement via abundant aesthetic dataset and reduces the bias caused by human personalized preferences. The progressive enhancing strategy makes the enhancing process easy-to-understand. With the plug & play setting, PIEA is free of any training process for learning parameters, which mitigates the requirements of costly paired data. The experiments on FiveK datasets show that PIEA outperforms the state-of-the-art methods. In the future, we will explore a better way to update filter parameters more efficiently and improve the aesthetic assessment models to adapt to the weak images.

ACKNOWLEDGMENTS

This work was partially supported by the National Key Research and Development Program of China under Grant 2016YFB1001001, the National Natural Science Foundation of China (Grant No. 61732007 and U1611461), and the National Science Foundation of China - Guangdong Joint Foundation (Grant No. U1401257). This research is also part of NExT++, supported by the National Research Foundation, Prime Ministers Office, Singapore under its IRC@Singapore Funding Initiative.

REFERENCES

[1] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. 2011. Learning Photographic Global Tonal Adjustment with a Database of Input / Output Image Pairs. In *CVPR*.

[2] Qifeng Chen, Jia Xu, and Vladlen Koltun. 2017. Fast image processing with fully-convolutional networks. In *ICCV*, Vol. 9. 2516–2525.

[3] Yu-Sheng Chen, Yu-Ching Wang, Man-Hsin Kao, and Yung-Yu Chuang. 2018. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In *CVPR*. 6306–6314.

[4] Yubin Deng, Chen Change Loy, and Xiaoou Tang. 2018. Aesthetic-driven image enhancement by adversarial learning. In *MM*. ACM, 870–878.

[5] Ryosuke Furuta, Naoto Inoue, and Toshihiko Yamasaki. 2019. Fully Convolutional Network with Multi-Step Reinforcement Learning for Image Processing. In *AAAI*.

[6] KGGZ Gu, Guangtao Zhai, Weisi Lin, and Min Liu. 2016. The Analysis of Image Contrast: From Quality Assessment to Automatic Enhancement. *IEEE Transaction on Cybernetics* 46, 1 (2016), 284–297.

[7] Yuanming Hu, Hao He, Chenxi Xu, Baoyuan Wang, and Stephen Lin. 2018. Exposure: A White-Box Photo Post-Processing Framework. *ACM Transactions on Graphics* 37, 2 (2018), 26.

[8] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. 2017. DSLR-quality photos on mobile devices with deep convolutional networks. In *ICCV*. 3277–3285.

[9] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. 2018. WESPE: weakly supervised photo enhancer for digital cameras. In *CVPR*. 691–700.

[10] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *CVPR*. IEEE, 5967–5976.

[11] Yuma Kinoshita, Sayaka Shiota, and Hitoshi Kiya. 2018. Automatic Exposure Compensation for Multi-Exposure Image Fusion. In *ICIP*. IEEE.

[12] Shu Kong, Xiaohui Shen, Zhe Lin, Radomir Mech, and Charless Fowlkes. 2016. Photo aesthetics ranking network with attributes and content adaptation. In *ECCV*. Springer, 662–679.

[13] Shuang Ma, Jing Liu, and Chang Wen Chen. 2017. A-lamp: Adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment. In *CVPR*. 4535–4544.

[14] Naila Murray, Luca Marchesotti, and Florent Perronnin. 2012. AVA: A large-scale database for aesthetic visual analysis. In *CVPR*. IEEE, 2408–2415.

[15] Anh Nguyen, Jeff Clune, Yoshua Bengio, Alexey Dosovitskiy, and Jason Yosinski. 2017. Plug & play generative networks: Conditional iterative generation of images in latent space. In *CVPR*. IEEE, 3510–3520.

[16] Mayu Omiya, Edgar Simo-Serra, Satoshi Iizuka, Hiroshi Ishikawa, Edgar Simo-Serra, Satoshi Iizuka, and Hiroshi Ishikawa. 2018. Learning Photo Enhancement by Black-Box Model Optimization Data Generation. *SIGGRAPH Asia Technical Briefs* 37 (2018).

[17] Jongchan Park, Joon-Young Lee, Donggeun Yoo, and In So Kweon. 2018. Distort-and-recover: Color enhancement using deep reinforcement learning. In *CVPR*. 5928–5936.

[18] Hossein Talebi and Peyman Milanfar. 2018. Nima: Neural image assessment. *IEEE Transactions on Image Processing* 27, 8 (2018), 3998–4011.

[19] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2242–2251.