A Fast and Practical CNN Method for Artful Image Regeneration

Xiaolin Wu
McMaster Univ
xwu@ece.mcmaster.ca

Qifan Gao
Shanghai Jiao Tong Univ
gaoqifan@sjtu.edu.cn

Zhenhao Li
Shanghai Jiao Tong Univ
mr.richor@gmail.com

Shenglei Li
Shanghai Jiao Tong Univ
lishenglei@sjtu.edu.cn

ABSTRACT

Although artists' actions in photo retouching appear to be highly nonlinear in nature and very difficult to characterize analytically, we find that the net effects of interactively editing a mundane image to a desired appearance can be modeled, in most cases, by a parametric monotonically non-decreasing global tone mapping function in the luminance axis and by a global affine transform in the chrominance plane. This allows us to greatly simplify the existing CNN methods for mimicking the artists in photo retouching, and design a new artful image regeneration network (AIRNet). The objective of AIRNet is to learn the image-dependent parameters of the luminance tone mapping function and the affine chrominance transform, rather than learning the end-to-end pixel level mapping as in the standard practice of current CNN methods for image restoration and enhancement. The proposed new approach reduces the complexity of the neural network by two orders of magnitude, and as a side benefit, it also improves the robustness and the generation capability at the inference stage.

ACM Reference Format:
Xiaolin Wu, Qifan Gao, Zhenhao Li, and Shenglei Li. 2020. A Fast and Practical CNN Method for Artful Image Regeneration. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Posters (SIGGRAPH '20 Posters), August 17, 2020.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3388770.3407428

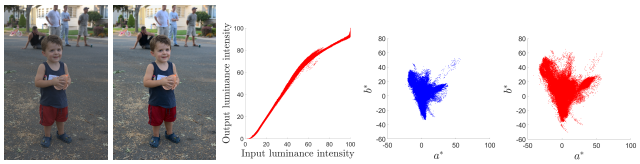
1 INTRODUCTION AND MOTIVATION

A big and unique challenge of automating photo retouching, which we call automatic artful image regeneration (AIR), is the great difficulty in analytically quantifying or modeling what makes an image more appealing. Indeed, even experienced photographers and postprocessing artists struggle to precisely explain visual esthetics; for most people their answers to the question are something alone the line "I know a beautiful image when I see it". In essence, AIR is a cognitive task not a signal processing one like image restoration. Complex, multifaceted, highly intuitive cognitive tasks like AIR are suited for a deep learning solution.

We propose a CNN framework for the AIR task, called AIRNet. AIRNet is motivated by our newly gained insight into the following properties of photo retouching results. The net effects from a mundane image to its final beautified version can be satisfactorily modeled by two independent parametric models in CIELab color space, one for luminance and one for chrominance. Fig. 1(a) and (b) are an input image $I(L, a, b)$ and a retouched output image $I'(L', a', b')$ in MIT-Adobe FiveK dataset[Bychkovsky et al. 2011]. Fig. 1(c) is the scatter plot of input-output pixel pairs (L, L') in luminance. The scatter plot reveals a monotonically increasing and smooth transfer function $L' = T_L(L)$ from the input luminance to the output luminance. Fig. 1 (d) and (e) are the distributions of chrominance vectors before and after photo retouching. One can observe that the point cloud of (a', b') in Fig. 1(e) is approximately the result of translating, rotating (if necessary) and scaling the chrominance vectors (a,b) in Fig. 1(d), i.e., $(a', b') = T_C(a, b)$ where T_C is approximately an affine transform. These behavior hold for the majority of image pairs of the MIT-Adobe FiveK dataset.

SIGGRAPH '20 Posters, August 17, 2020, Virtual Event, USA

ACM ISBN 978-1-4503-7973-1/20/08.
https://doi.org/10.1145/3388770.3407428



(a) (b) (c) (d) (e)

Figure 1: (a) original image ©MIT-Adobe FiveK ; (b) retouched image I'; (c) scatter plot of input-output pixel pairs (L, L') in luminance, and the chrominance distribution in a-b plane of (d) I and (e) I'.

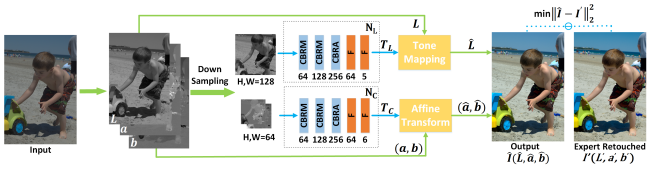


Figure 2: Architecture of AIRNet. Symbols C, B, R, M, A, F represent convolution layer, batch normalization, ReLU, max pooling layer (stride=2), average pooling layer and fully connected layer respectively.

2 AIRNET

The proposed parametric transform models T_L and T_C for photo retouching reduce the AIR problem to that of estimating the models' parameters. As these parameters are image dependent and the dependencies are difficult to analytically characterize, machine learning naturally becomes a method of choice. The proposed AIRNet automates AIR by generating the image dependent parameters of T_L and T_C. As shown in Fig.2, AIRNet comprises two independent subnetworks N_L and N_C that take the luminance and chrominance components of an image as input respectively, and construct the transforms T_L and T_C. T_L is applied to the luminance component $\hat{L} = T_L(L)$ and T_C to the chrominance component $(\hat{a}, \hat{b}) = T_C(a, b)$ of the input image $I(L, a, b)$ to produce the output image $\hat{I}(\hat{L}, \hat{a}, \hat{b})$. AIRNet optimizes the parameters of T_L and T_C for minimal ℓ_2 loss of $\hat{I}$ with respect to the expert retouched image I' (ground truth).

We conduct a comparison study of the proposed AIRNet method versus five state-of-the-art deep learning-based AIR methods:

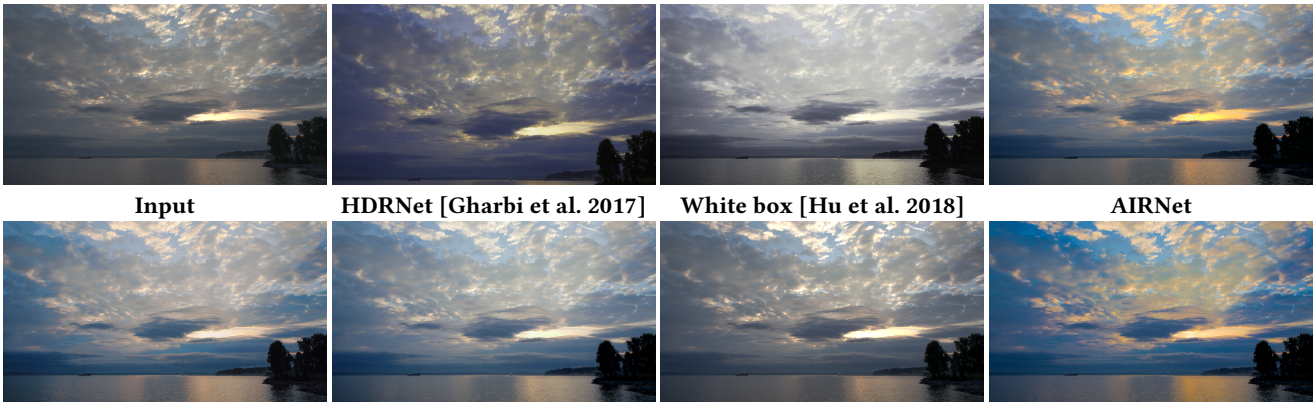


Figure 3: Comparison of different methods on an MIT-Adobe FiveK image ©.

Method	Parameters ($\times 10^6$)	GFLOPs
DPE	3.338	1614
D&R	70.89	481.5
DPED	0.401	1604
HDRNet	1.446	0.670
White box	25.68	19.70
AIRNet	0.865	0.107

Table 1: Comparison of model sizes and FLOPs for processing an input image of size 1920×1080 with existing methods.

	PSNR↑	SSIM↑	NIQE↓	PAM↑	LOD↓
Input	-	-	3.8158	3.0326	-
D&R	22.19	0.8173	3.5376	3.3010	1.3282
DPE	**23.45**	0.8233	3.6604	3.3255	1.8912
White box	18.75	0.6897	3.7000	3.2110	2.8573
HDRNet	22.76	**0.8556**	3.5598	3.2420	1.1850
DPED	21.97	0.7510	4.6457	3.3290	4.4942
AIRNet	22.98	0.8246	**3.4902**	**3.3375**	**1.0090**
Expert C	-	-	3.3710	3.4054	-

Table 2: Performance comparison on 1500 test images in MIT-Adobe FiveK dataset. for metrics marked by "↑", the larger the better; for those with "↓", the smaller the better.

DPE[Chen et al. 2018], HDRNet[Gharbi et al. 2017], DPED[Ignatov et al. 2017], White Box[Hu et al. 2018] and D&R[Park et al. 2018]. Table 1 lists the model size (measured by the number of model parameters), and the number of floating point operations (FLOPs) of the above methods given a 1920×1080 input image as input. AIRNet has significantly lower computation cost than its opponents thanks to its much simplified parameterized learning task backed by strong priors. It is worth mentioning that the reason for DPED's high FLOP number despite having the fewest parameters is the absence of downsampling operations during inference. Table 2 compares the six methods by two common similarity metrics, PSNR and SSIM. AIRNet is the runner-up on both, narrowly behind DPE and HDRNet. But the latter two methods employ much larger CNN models and hence have much higher complexity. As criticized in [Blau and Michaeli 2018], signal level distortion metrics such as PSNR or SSIM often have poor correlation to perceptual image quality. Therefore we evaluate the methods by the following metrics as well: 1) naturalness image quality evaluator (NIQE) [Mittal et al. 2012], a widely used image naturalness metric; 2) personalized aesthetics model (PAM) [Ren et al. 2017], a state-of-the-art approach for objective image aesthetics rating; and 3) luminance ordinary distortion (LOD) [Li and Wu 2017], a measure of the distortion of artifacts due to over-enhancement (such as halos). These metrics put more emphasis on observers' visual experience. As reported in Table 2, AIRNet outperforms its rivals on all the aesthetic ratings. Note that DPE and HDRNet fall behind here in spite of the low distortion relative to the ground truth. Fig. 3 exhibits the output images of the competing methods (see supplementary materials and the representative image for more examples). As shown in the figure, AIRNet produces desirably high contrast and vivid color appearance, resulting in the most faithful reproduction of the expert's retouched images. Other methods are somewhat lacking in visual appeal.

REFERENCES

Yochai Blau and Tomer Michaeli. 2018. The perception-distortion tradeoff. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 6228–6237.

Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. 2011. Learning Photographic Global Tonal Adjustment with a Database of Input / Output Image Pairs. In *The Twenty-Fourth IEEE Conference on Computer Vision and Pattern Recognition.*

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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 6306–6314.

Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédo Durand. 2017. Deep bilateral learning for real-time image enhancement. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 118.

Yuanming Hu, Hao He, Chenxi Xu, Baoyuan Wang, and Stephen Lin. 2018. Exposure: A white-box photo post-processing framework. *ACM Transactions on Graphics (TOG)* 37, 2 (2018), 26.

Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. 2017. DSLR-quality photos on mobile devices with deep convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision.* 3277–3285.

Zhenhao Li and Xiaolin Wu. 2017. Learning-Based Restoration of Backlit Images. *IEEE Transactions on Image Processing* 27, 2 (2017), 976–986.

Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. 2012. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters* 20, 3 (2012), 209–212.

Jongchan Park, Joon-Young Lee, Donggeun Yoo, and In So Kweon. 2018. Distort-and-recover: Color enhancement using deep reinforcement learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 5928–5936.

Jian Ren, Xiaohui Shen, Zhe Lin, Radomir Mech, and David J Foran. 2017. Personalized image aesthetics. In *Proceedings of the IEEE International Conference on Computer Vision.* 638–647.