Learning Photo Enhancement by Black-Box Model Optimization Data Generation

Mayu Omiya
Waseda University
Tokyo, Japan
omiya_mayu@akane.waseda.jp

Edgar Simo-Serra
Waseda University
Tokyo, Japan
ess@waseda.jp

Satoshi Iizuka
University of Tsukuba
Ibaraki, Japan
iizuka@cs.tsukuba.ac.jp

Hiroshi Ishikawa
Waseda University
Tokyo, Japan
hfs@waseda.jp

ABSTRACT

We address the problem of automatic photo enhancement, in which the challenge is to determine the optimal enhancement for a given photo according to its content. For this purpose, we train a convolutional neural network to predict the best enhancement for given picture. While such machine learning techniques have shown great promise in photo enhancement, there are some limitations. One is the problem of *interpretability*, i.e., that it is not easy for the user to discern what has been done by a machine. In this work, we leverage existing manual photo enhancement tools as a black-box model, and predict the enhancement parameters of that model. Because the tools are designed for human use, the resulting parameters can be interpreted by their users. Another problem is the difficulty of obtaining training data. We propose generating supervised training data from high-quality professional images by randomly sampling realistic *de*-enhancement parameters. We show that this approach allows automatic enhancement of photographs without the need for large manually labelled supervised training datasets.

CCS CONCEPTS

• **Computing methodologies** → **Image manipulation**; *Image processing*;

KEYWORDS

photo enhancement, black-box optimization, machine learning

ACM Reference Format:
Mayu Omiya, Edgar Simo-Serra, Satoshi Iizuka, and Hiroshi Ishikawa. 2018. Learning Photo Enhancement by Black-Box Model Optimization Data Generation. In *SIGGRAPH Asia 2018 Technical Briefs (SA '18 Technical Briefs), December 4–7, 2018, Tokyo, Japan.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3283254.3283286

(a) input (b) output

Figure 1: Automatic photo enhancement with our approach. (a) Input image and (b) enhanced image. The image is "Ojito Sunset" by John Fowler licensed under CC BY 2.0.

1 INTRODUCTION

Although taking a photo is nearly instantaneous, post-processing for obtaining the desired visual characteristics, such as brightness and color properties, can take many minutes or even hours. Given that professional photographers may take thousands of images in a single photo-shoot, the burden of post-processing them all is intractable. In this work, we focus on automatically enhancing photographs using machine learning techniques.

The challenge is that optimal enhancing parameters for each image is different, and not only because of simple features such as the overall brightness and contrast. Rather, the content of the image must be taken into account for best result. This defines the problem as a content-based mapping of images to enhancements. Convolutional neural network is best suited for such image interpretation problems.

An enhancement is a mapping of images to images. In our approach, it is defined by the parameters for an off-the-shelf professional photo enhancement software, which is treated as a black box that turns the parameters into the enhancement. By predicting enhancement parameters instead of directly modifying the pixel values, we are able to easily interpret the results, as well as further improve the image manually.

We train a convolutional neural network to predict the best enhancement parameters for a given picture. For such training, we need a large amout of data, in this case many pairs of a non-optimal input photo and a set of parameters that enhances it as well as possible. Making such pairs manually is a daunting task, so we generate them from high-quality professional photographs that represent the desired results of enhancement. For a given professional photograph, we randomly sample realistic photo *de*-enhancement parameters, which allows creating 'amateur' pre-enhancement versions of each professional photograph. We then use optimization to obtain the parameters necessary to convert the de-enhanced image to the professional image. Finally, the de-enhanced image and the obtained parameters are used as the supervised training data to train our automatic photo enhancement model.

We evaluate our model on challenging real-world photographs, and compare with existing methods using a perceptual user study.

SA '18 Technical Briefs , December 4–7, 2018, Tokyo, Japan
© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.
This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *SIGGRAPH Asia 2018 Technical Briefs (SA '18 Technical Briefs), December 4–7, 2018, Tokyo, Japan*, https://doi.org/10.1145/3283254.3283286.

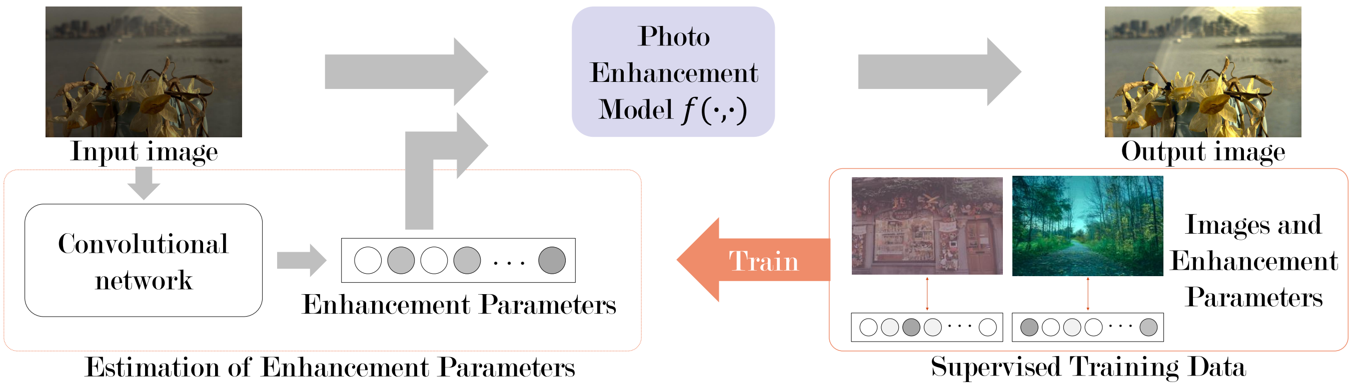

Figure 2: Overview of our approach. We train a CNN to predict the photo enhancement parameters for a given input image, which can then be used with a photo enhancement model. The supervised training data consists of pairs of input images and photo enhancement parameters.

2 RELATED WORK

Koyama et al. [Koyama et al. 2016] proposed a system that aids enhancement of photographs by learning a preference model as a photo enhancement model for each user. However, since learning is done from the edit history, it is necessary for the user to manually correct various kinds of photos in order to improve the enhancement accuracy. Kaufman's method [Kaufman et al. 2012] automatically corrects the contrast and saturation by a fixed amount in the regions other than face and sky, which it recognizes. Therefore, the enhancement is rather uniform and it is difficult to make a correction that matches the contents of the photo. Yan et al. [Yan et al. 2016] proposed a method to automatically enhance photos by feeding feature descriptors calculated from an image into deep neural network using the Adobe-FiveK datasets [Bychkovsky et al. 2011]. The calculation of the feature descriptors depends on scene detection [Tighe and Lazebnik 2010] and object recognition [Viola and Jones 2001] for a limited number of classes such as sky and road, and it does not work on images where class estimation and accurate segmentations are difficult to obtain. In Deep Photo Enhancer [Chen et al. 2018], Chen et al. use GANs to learn to enhance photos. Though the purpose is slightly different, Lee et al. [Lee et al. 2016] proposed a method to convert the style of photos. In this method, pairs of photos before and after enhancement are not necessary as teacher data.

The commercial software Adobe Photoshop[1] and Adobe Lightroom[2] have automatic photo enhancement functions. Adobe Lightroom features an automatic enhancement assisted by the latest artificial intelligence called Adobe Sensei. It uses thousands of photos with high-quality enhancement, machine learning platform and neural networks.

3 PROPOSED APPROACH

An overview of our approach can be seen in Fig. 2. We create training data by *de*-enhancing professional photographs and use optimization to obtain photo enhancement parameters. Then, the pairs of de-enhanced images and photo enhancement parameters can be used to train a convolutional neural network which, in combination with the photo enhancement model, can be used for the automatic enhancement of photographs, as well as allowing for further manual editing.

For our approach, we assume we have a photo enhancement model, but make no assumptions on the model itself, other than that it has a finite number of parameters, representable as a vector, that can be used to enhance a particular image.

3.1 Automatic Photo Enhancement Model

Our photo enhancement model is based on a convolutional neural network that predicts the photo enhancement parameters of a given photograph. Given that each photo enhancement parameter can have different ranges, we normalize each photo-enhancement parameter to be predicted by subtracting its mean and dividing by its standard deviation over the training dataset. The optimization of the model can be then formulated as:

$$\arg\min_{\phi} \frac{1}{N} \sum_{i=0}^{N} \left| c(x_i;\phi) - \hat{\theta}_i^* \right|^2 \tag{1}$$

where $c(x;\phi)$ is the convolutional neural network with input image x and parameters ϕ, and $\hat{\theta}^*$ is the normalized photo enhancement parameter vector of length N.

Instead of training from scratch, we fine-tune a VGG19 [Simonyan and Zisserman 2014] model pre-trained on ImageNet [Krizhevsky et al. 2017]. In particular, we replace the last layer with one with N outputs, and train the model using stochastic gradient descent.

3.2 Generating Training Data

Pairs of non-enhanced images and enhancement parameters are necessary to train our automatic photo enhancement software. An overview of our data creation approach can be seen in Fig. 3. To create the training data, we randomly sample parameters to de-enhance high-quality images into simulated amateur photos. As the random sampling from the parameter space will inevitably lead to images unnatural even for an amateur, we propose using a separate classification network to only keep natural-looking de-enhanced photos. We manually annotate randomly de-enhanced photos to train the classification network and use 6,000 images for training. This allows quickly creating de-enhanced images for training data with a reduced bias towards unnatural images. As a classification network, we use a VGG16 [Simonyan and Zisserman 2014] model

[1]https://www.adobe.com/uk/products/photoshop.html
[2]https://www.adobe.com/uk/products/photoshop-lightroom.html

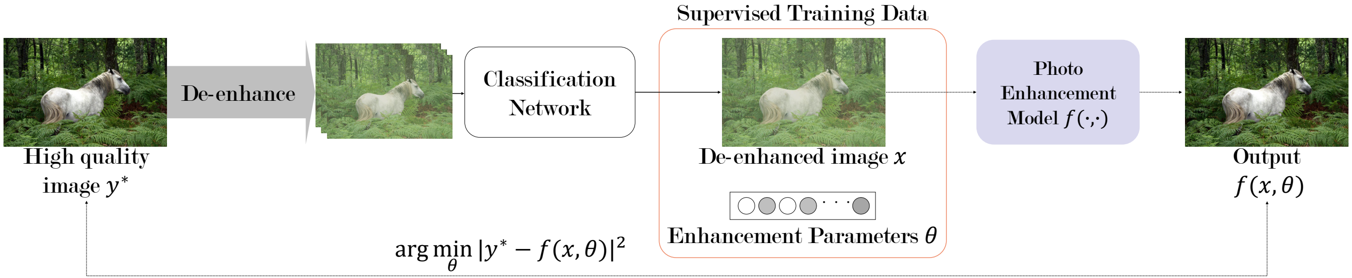

Figure 3: Overview of the data generation process. High-quality professional photographs are de-enhanced to produce lower quality ones. Photos that are too unrealistic are removed by using a classification network. Afterwards, the parameters needed to enhance the de-enhanced image back to the original are estimated by optimization. The pairs of de-enhanced photos and enhancement parameters are used as the supervised training data for the automatic enhancement model. Horse image is "Horse" by lago licensed under CC BY 2.0.

fine-tuned to the task of classifying de-enhanced photos as realistic or not. To check if a photo is suitable for training or not, we set a threshold τ_s on the classification score.

Since we do not assume the photo enhancement model to be invertible, in order to obtain the enhancement parameters to convert the de-enhanced photo back to the original photo we resort to black-box optimization. In particular, for a given de-enhanced image x and a original photo y^*, we optimize

$$\arg\min_{\theta} \left| y^* - f(x,\theta) \right|^2 \quad , \qquad (2)$$

where $f(\cdot,\cdot)$ is the photo enhancement model and θ is the enhancement parameters. In optimizing, there are two points to be aware of. One is that there may be multiple combinations of enhancement parameters and operations that lead to the same result, which might confound the enhancement-learning network. This problem can be solved by using a sufficient number of data. Secondly, the order of functions performed within the enhancement model affects the results. In consideration of this, we use a fixed function order. For optimization, we use the Nelder-Mead (NM) method [Nelder and Mead 1965], which does not require the explicit computation of gradients. In the NM method, an n-dimensional domain is searched by using a simplex which is a $(n+1)$-vertex polytope in a n-dimensional domain. Optimization is done by repeating the reflection, expansion, and contraction of the simplex within the domain.

The full approach to generate the supervised data can thus be described as follows:

(1) Randomly sample from enhancement parameters and create a de-enhanced photo.
(2) Classify the de-enhanced photo as suitable for training or not. If not, go to (1).
(3) Use optimization to find the enhancement parameters that converts the de-enhanced photo back to the original.
(4) Check if optimization loss is within a threshold τ_o. If not, go to (1).
(5) Add the de-enhanced photo and the enhancement parameters to the training dataset.

4 EXPERIMENTS

As the high-quality pictures, it would be best to use professional ones. For cost reasons, however, for this experiments we instead obtained 13, 750 images from Flickr's Interestingness photos, which are pictures in Flickr chosen based on the number of: photo tags, groups they belong to, people who saw the images, favorite people, etc., and which seem to us to be of very high-quality. For the model, we set τ_s to 2.5 and τ_o to 0.066.

Table 1: Overview of the different functions we use in our black-box photo enhancement model and the number of associated parameters.

Function Name	Number of Parameters
exposure, brightness	3
shadow, highlight	5
contrast	1
saturation, vibrance	2
white balance	3
color correction	5
color contrast	2
total	21

4.1 Photo Enhancement Model

As the black-box photo enhancement model, we used the Darktable [3] open-source software package, which contains a large number of different photo enhancement functions. The functions we used and their number of parameters are shown in Table 1. In particular, we focus on a diverse set of functions that allow for correcting exposure, brightness, shadows, color, ... We note that we treated the set of functions as a black-box model and did not make any hypothesis on their characteristics nor functionality. This allowed incorporating stronger non-linear non-local functionality for the photo enhancement.

4.2 Comparison with Existing Approaches

We compare the results by our approach with those by existing approaches on images taken from Flickr in Fig. 4. In particular, we compare against Adobe Photoshop automatic tone enhancement, contrast and color enhancement, and Deep Photo Enhancer (DPE) [Chen et al. 2018]. We can see how our approach was able to significantly improve the brightness of the input image. In the bottom row, our approach was able to improve the white balance in contrast with the other two approaches.

[3] https://www.darktable.org/

(a) Input Image (b) Adobe Photoshop (c) DPE [Chen et al. 2018] (d) Ours

Figure 4: Comparison of automatic enhancement results with existing approaches. More results are in the supplementary materials. The top image is "Sheffield Park and Garden 22-10-2010" by Karen Roe licensed under CC BY 2.0. The bottom image is "Strasbourg" by francois schnell licensed under CC BY 2.0.

Table 2: Results of the perceptual user study.

	Ours	Input	Photoshop	DPE
vs. Ours	-	24.4%	41.2%	33.1%
vs. Input	**75.6%**	-	85.3%	75.0%
vs. Photoshop	**58.8%**	14.7%	-	35.7%
vs. DPE	**66.9%**	25.0%	64.3%	-

4.3 Perceptual User Study

We conducted a user study by showing the result of automatic enhancement by our method, Adobe Photoshop, and DPE [Chen et al. 2018] to 20 participants. The scores are shown in Table 2. In the user study, we used a total of 50 input images and 200 output images by each method. The participants were shown two images randomly selected from all the images and selected the one that looks more natural and generally better. For each pair of dueling methods, the percentage of its output selected by the user is the score. The proposed method was evaluated as more natural and easier to see than Adobe Photoshop and DPE [Chen et al. 2018]. Adobe Photoshop came the closest, but note that this commercial program was presumably trained using thousands of image pairs before and after hand-enhancement.

5 CONCLUSIONS

We have presented a fully automatic approach for photo enhancement that is also amenable to further manual editing. Our approach is based on predicting the enhancement parameters of off-the-shelf black-box photo enhancement software, and allows for easily interpretability of the results. We generate training data by de-enhancing professional photographs and estimate enhancement parameters through optimization, which allows us to circumvent the need for expensive annotation data.

ACKNOWLEDGMENTS

This work was partially supported by JST ACT-I (Iizuka, Grant Number: JPMJPR16U3), JST PRESTO (Simo-Serra, Grant Number: JPMJPR1756), and JST CREST (Ishikawa, Iizuka, and Simo-Serra, Grant Number: JPMJCR14D1).

REFERENCES

Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. 2011. Learning photographic global tonal adjustment with a database of input/output image pairs. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, Colorado Springs, CO, USA, 97–104.

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.

Liad Kaufman, Dani Lischinski, and Michael Werman. 2012. Content-Aware Automatic Photo Enhancement. *Comput. Graph. Forum* 31, 8 (Dec. 2012), 2528–2540. https://doi.org/10.1111/j.1467-8659.2012.03225.x

Yuki Koyama, Daisuke Sakamoto, and Takeo Igarashi. 2016. SelPh: Progressive Learning and Support of Manual Photo Color Enhancement. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2520–2532. https://doi.org/10.1145/2858036.2858111

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2017. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* 60, 6 (May 2017), 84–90. https://doi.org/10.1145/3065386

Joon-Young Lee, Kalyan Sunkavalli, Zhe Lin, Xiaohui Shen, and In So Kweon. 2016. Automatic content-aware color and tone stylization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2470–2478.

John A Nelder and Roger Mead. 1965. A simplex method for function minimization. *The computer journal* 7, 4 (1965), 308–313.

Karen Simonyan and Andrew Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. (09 2014).

Joseph Tighe and Svetlana Lazebnik. 2010. Superparsing: scalable nonparametric image parsing with superpixels. In *European conference on computer vision*. Hersonissos, Heraklion, Crete, Greece, 352–365.

Paul Viola and Michael Jones. 2001. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, Vol. 1. IEEE, Kauai, HI, USA, I–I.

Zhicheng Yan, Hao Zhang, Baoyuan Wang, Sylvain Paris, and Yizhou Yu. 2016. Automatic photo adjustment using deep neural networks. *ACM Transactions on Graphics (TOG)* 35, 2 (2016), 11.