DSLR-Quality Photos on Mobile Devices with Deep Convolutional Networks –Supplemental material–

Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, Luc Van Gool
Computer Vision Laboratory, ETH Zurich, Switzerland
{ihnatova,nk,timofter,vanhoey,vangool}@vision.ee.ethz.ch

Abstract

Despite a rapid rise in the quality of built-in smartphone cameras, their physical limitations — small sensor size, compact lenses and the lack of specific hardware, — impede them to achieve the quality results of DSLR cameras. In this work we present an end-to-end deep learning approach that bridges this gap by translating ordinary photos into DSLR-quality images. We propose learning the translation function using a residual convolutional neural network that improves both color rendition and image sharpness. Since the standard mean squared loss is not well suited for measuring perceptual image quality, we introduce a composite perceptual error function that combines content, color and texture losses. The first two losses are defined analytically, while the texture loss is learned in an adversarial fashion. We also present DPED, a large-scale dataset that consists of real photos captured from three different phones and one high-end reflex camera. Our quantitative and qualitative assessments reveal that the enhanced image quality is comparable to that of DSLR-taken photos, while the methodology is generalized to any type of digital camera.

1. Introduction

These pages provide the following supplementary material[1]:

- *Results of the proposed method*: We show how our algorithm works on each of the 3 cameras proposed in the paper: iPhone (section 2), BlackBerry (section 3) and Sony (section 4). For every camera, we provide a set of image pairs: the left image is the original photo taken by the camera, and in the right image we show the results of applying our method.
- *Comparison with the baselines*: For pictures taken with iPhone (section 5), BlackBerry (section 6) and Sony (section 7) camera's, we show images of the same structure as Fig. 8 in the paper. We provide a set of 6 images arranged in the following way:

Original image	Result of Apple Photos automatic enhancer	Result of Dong *et al.* [1]
Result of Johnson *et al.* [3]	**Our result**	Corresponding DSLR image

- *User study*: For our user study that was carried out on 42 participants, we show the fraction of replies for every of 108 questions asked, as well as its mean value (in bold) for same questions carried out in different images. (section 8)
- *Loss analysis*: We evaluate the contribution of different loss terms used in our target error function. (section 9)
- *Results of [2]*: We analyze the results of another relevant method obtained on our dataset. (section 10)

[1] dped-photos.vision.ee.ethz.ch

2. Results of the proposed method: iPhone

iPhone original | Enhanced with our method

Figure 1: Image results of our method for iPhone DPED test images 1 to 4.

iPhone original | Enhanced with our method

Figure 2: Image results of our method for iPhone DPED test images 5 to 8.

iPhone original | Enhanced with our method

Figure 3: Image results of our method for iPhone DPED test images 9 to 12.

Figure 4: Image results of our method for iPhone DPED test images 13 to 16.

iPhone original | Enhanced with our method

Figure 5: Image results of our method for iPhone DPED test images 17 to 20.

3. Results of the proposed method: BlackBerry

BlackBerry original Enhanced with our method

Figure 6: Image results of our method for BlackBerry DPED test images 1 to 4.

BlackBerry original | Enhanced with our method

Figure 7: Image results of our method for BlackBerry DPED test images 5 to 8.

Figure 8: Image results of our method for BlackBerry DPED test images 9 to 12.

BlackBerry original | Enhanced with our method

Figure 9: Image results of our method for BlackBerry DPED test images 13 to 16.

BlackBerry original | Enhanced with our method

Figure 10: Image results of our method for BlackBerry DPED test images 17 to 20.

4. Results of the proposed method: Sony

Figure 11: Image results of our method for Sony DPED test images 1 to 4.

Sony original | Enhanced with our method

Figure 12: Image results of our method for Sony DPED test images 5 to 8.

Sony original | Enhanced with our method

Figure 13: Image results of our method for Sony DPED test images 9 to 12.

Sony original | Enhanced with our method

Figure 14: Image results of our method for Sony DPED test images 13 to 16.

Figure 15: Image results of our method for Sony DPED test images 17 to 20.

5. Comparison with the baselines: iPhone

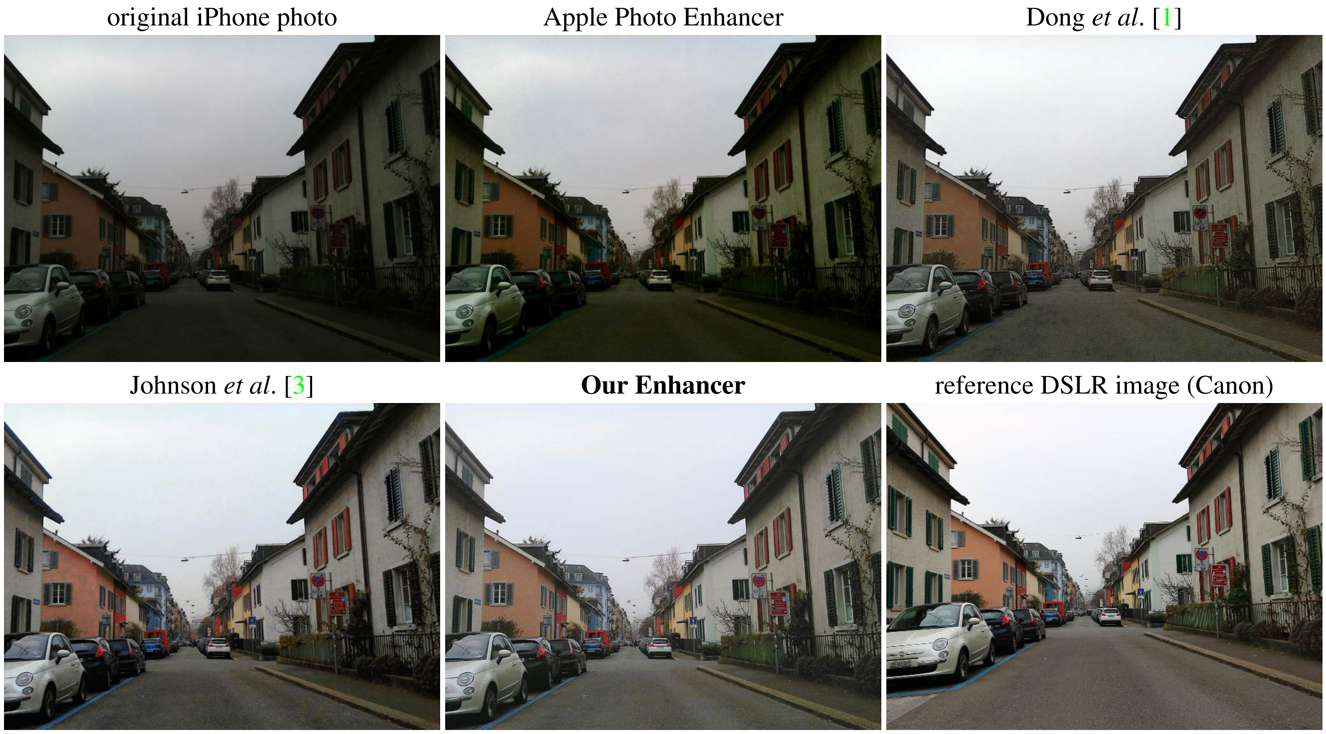

Figure 16: Comparison of 4 enhancers applied to an iPhone photo vs. the DSLR reference image (Canon).

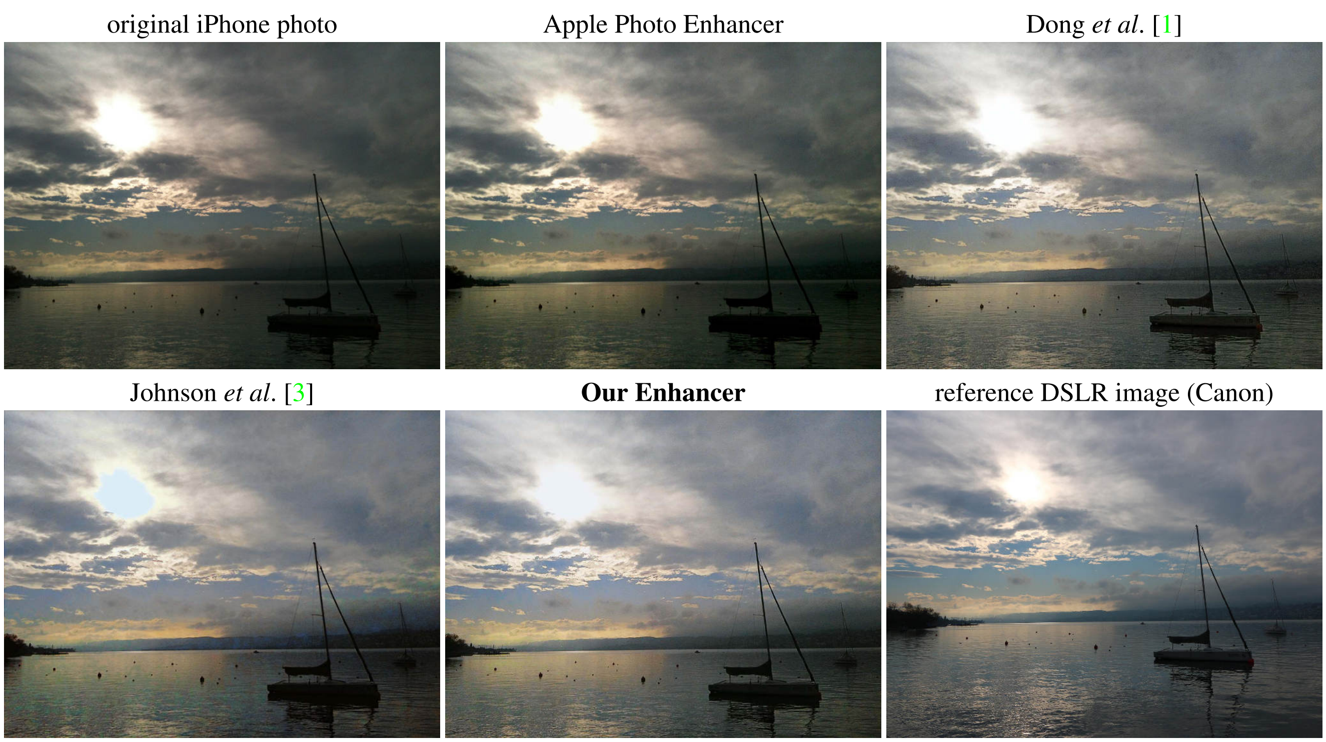

Figure 17: Comparison of 4 enhancers applied to an iPhone photo vs. the DSLR reference image (Canon).

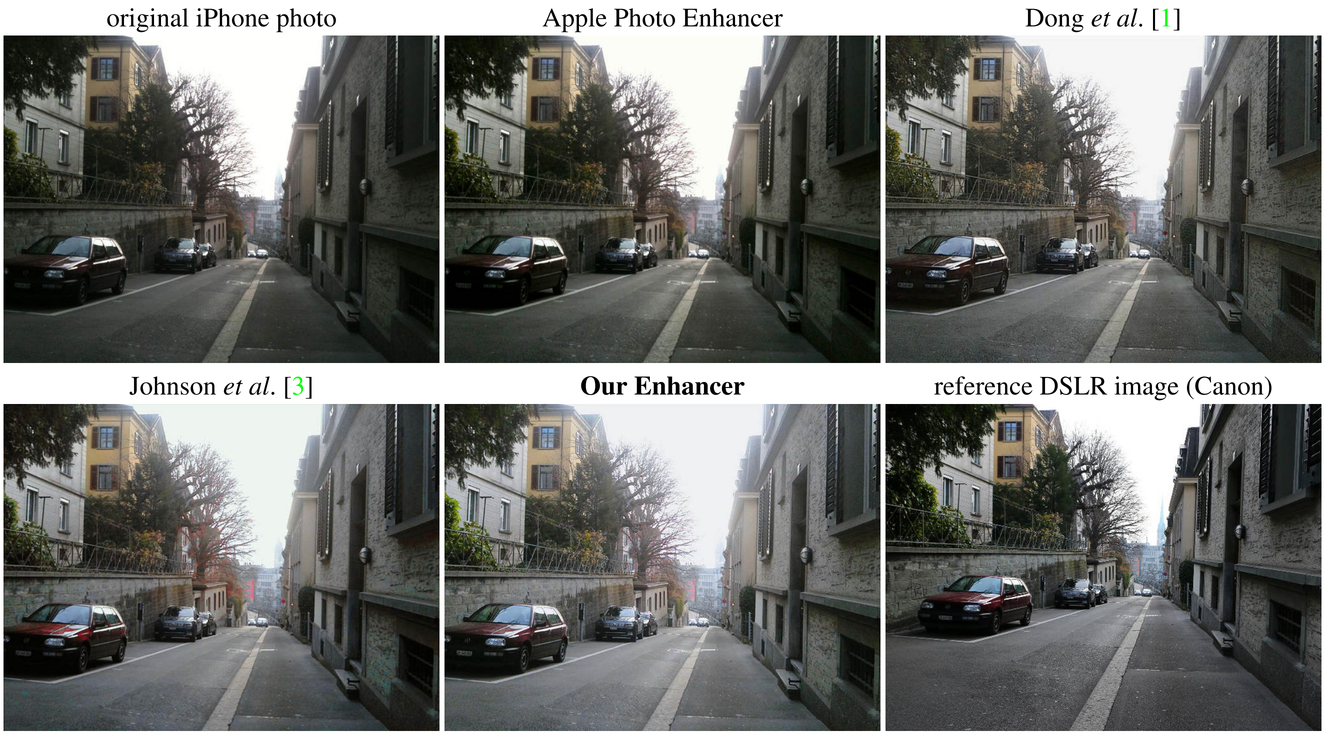

Figure 18: Comparison of 4 enhancers applied to an iPhone photo vs. the DSLR reference image (Canon).

Figure 19: Comparison of 4 enhancers applied to an iPhone photo vs. the DSLR reference image (Canon).

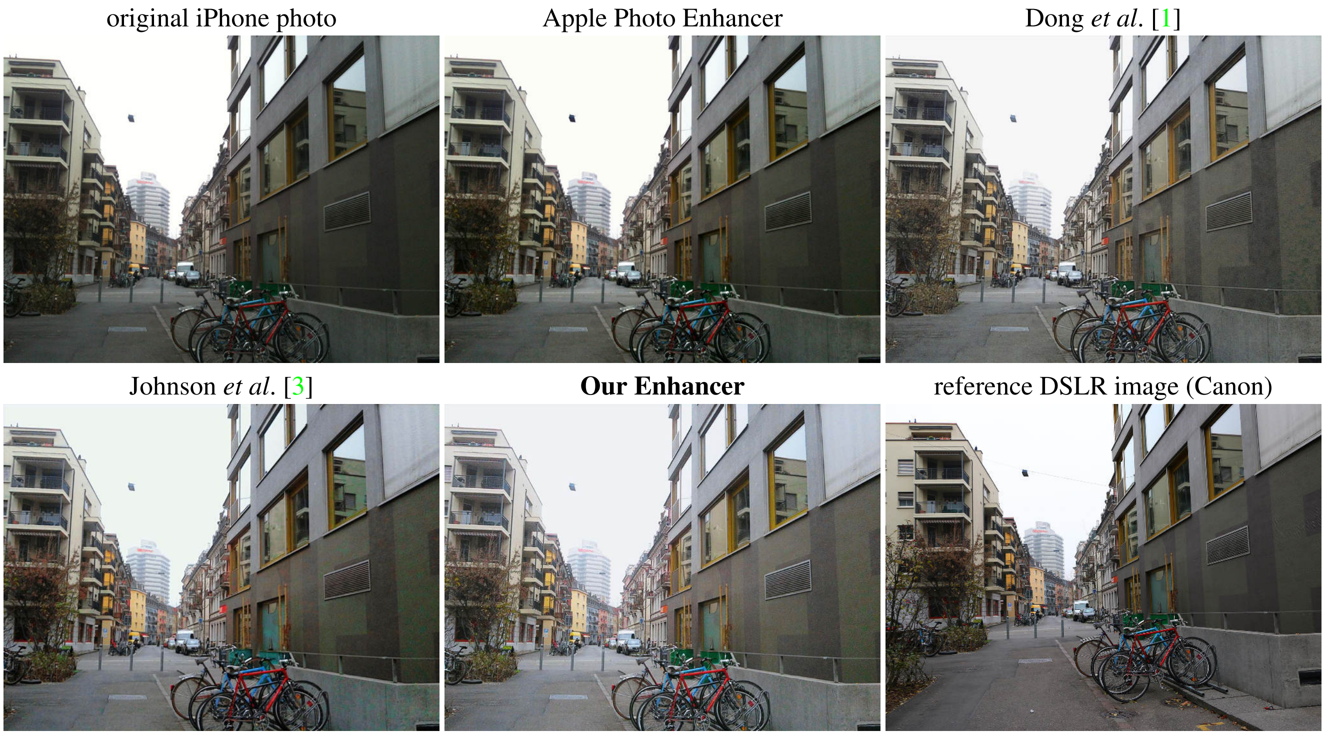

Figure 20: Comparison of 4 enhancers applied to an iPhone photo vs. the DSLR reference image (Canon).

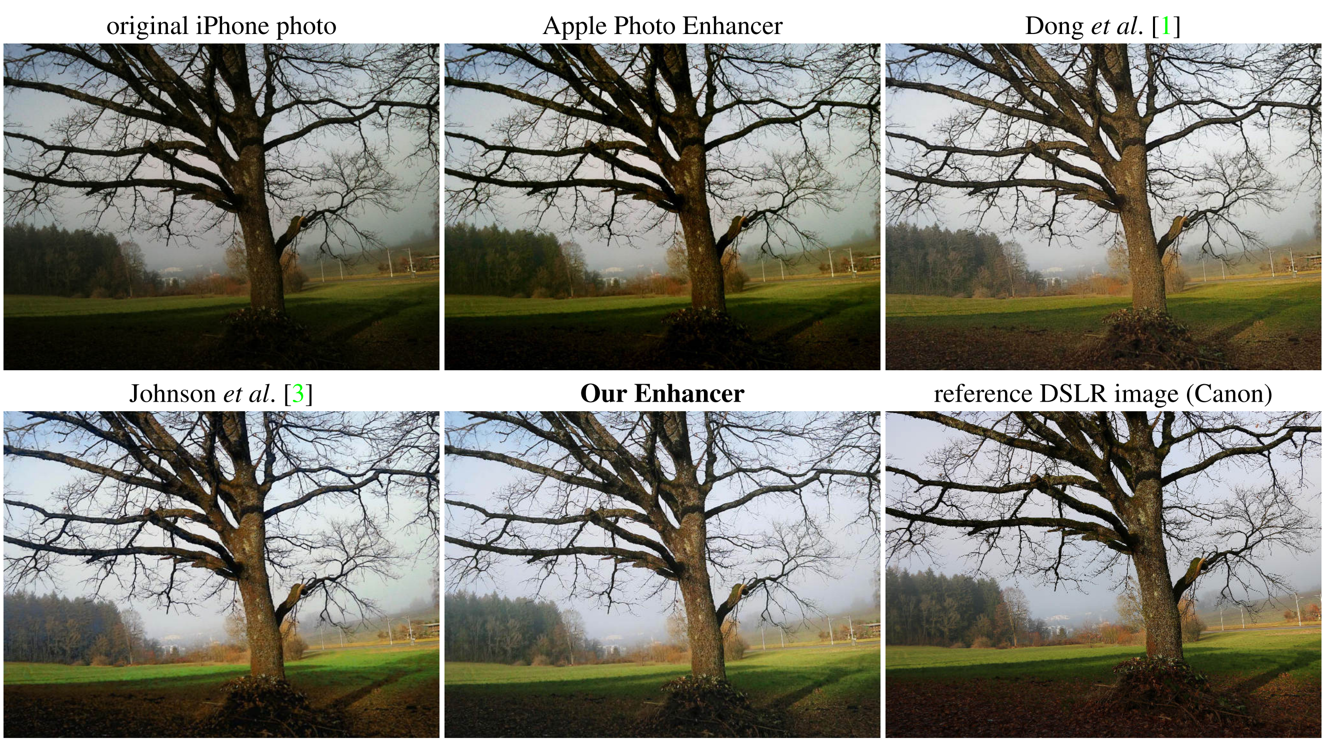

Figure 21: Comparison of 4 enhancers applied to an iPhone photo vs. the DSLR reference image (Canon).

6. Comparison with the baselines: BlackBerry

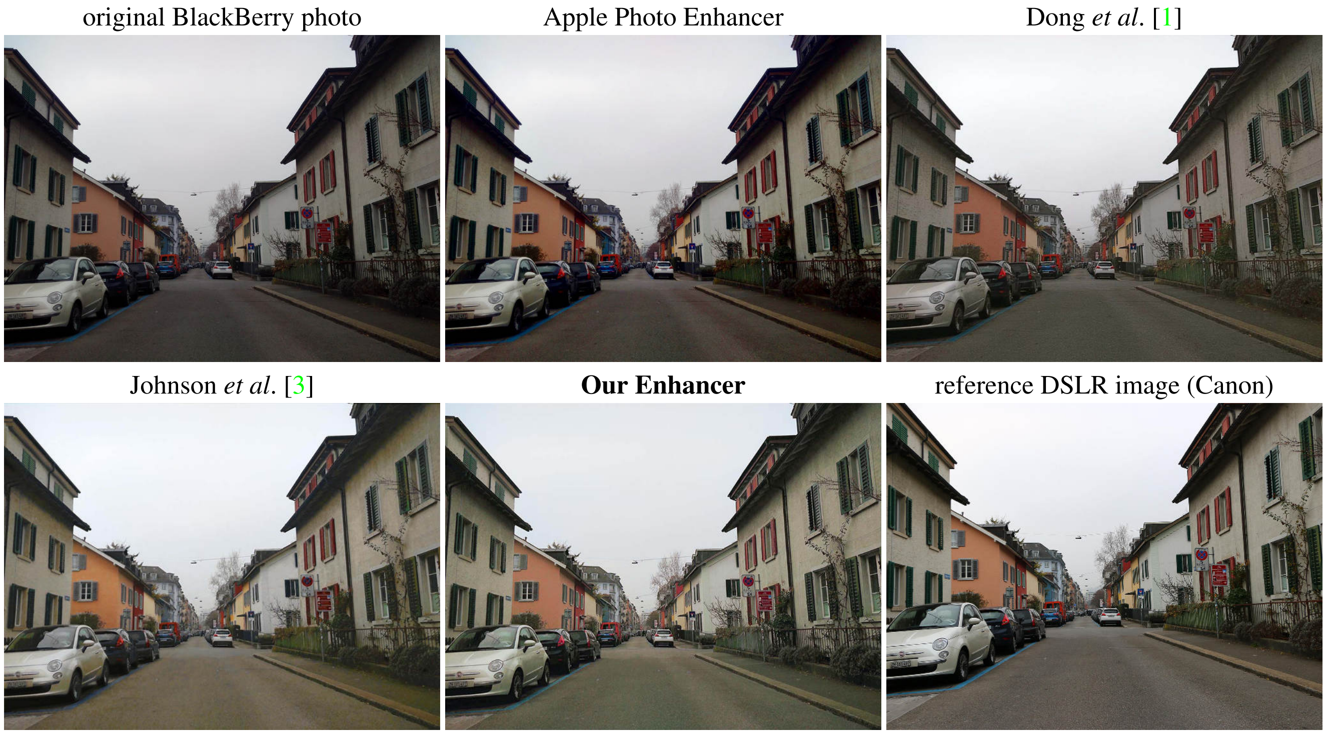

Figure 22: Comparison of 4 enhancers applied to a BlackBerry photo vs. the DSLR reference image (Canon).

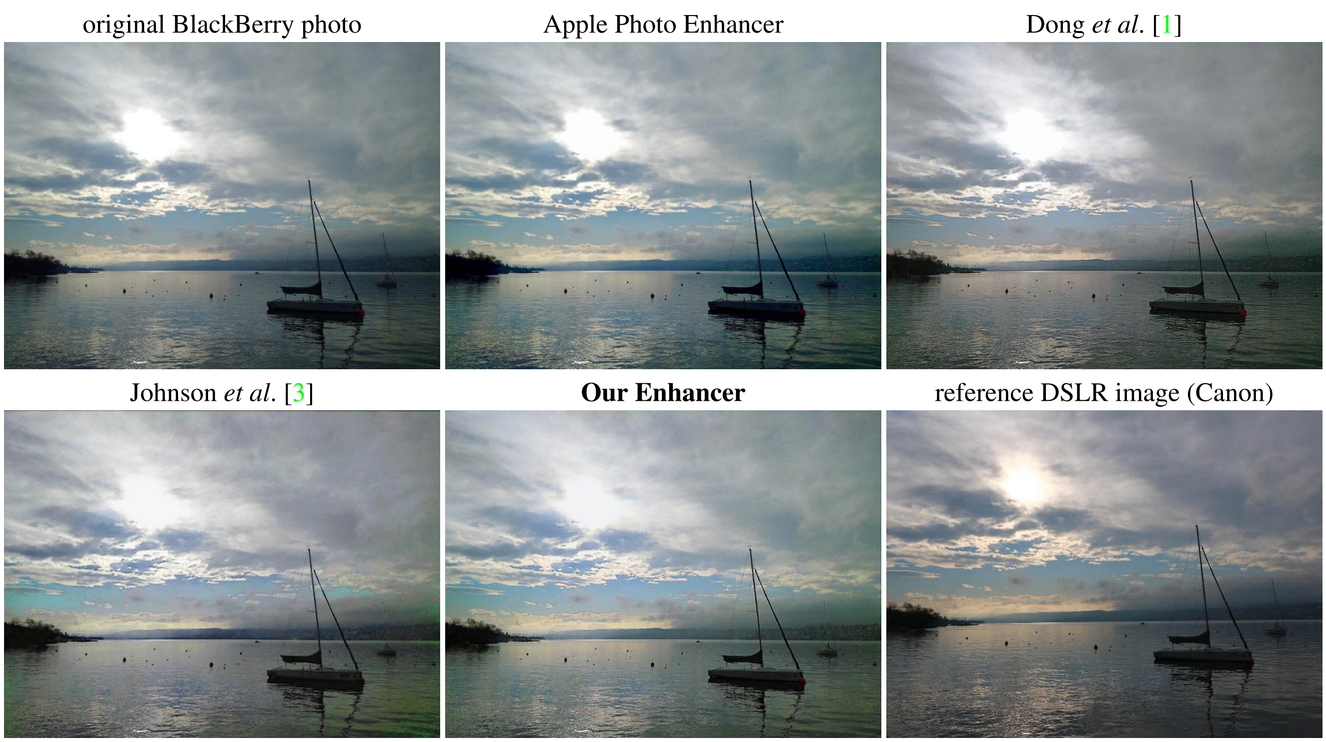

Figure 23: Comparison of 4 enhancers applied to a BlackBerry photo vs. the DSLR reference image (Canon).

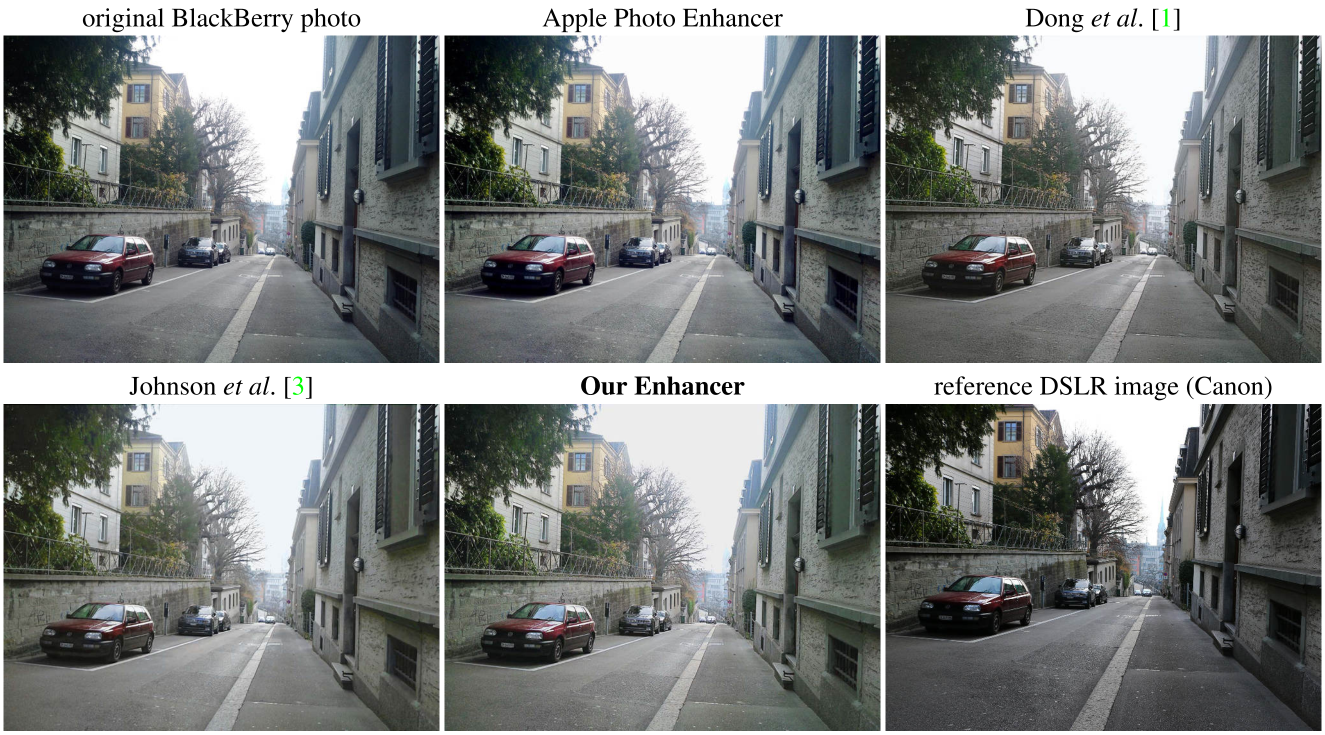

Figure 24: Comparison of 4 enhancers applied to a BlackBerry photo vs. the DSLR reference image (Canon).

Figure 25: Comparison of 4 enhancers applied to a BlackBerry photo vs. the DSLR reference image (Canon).

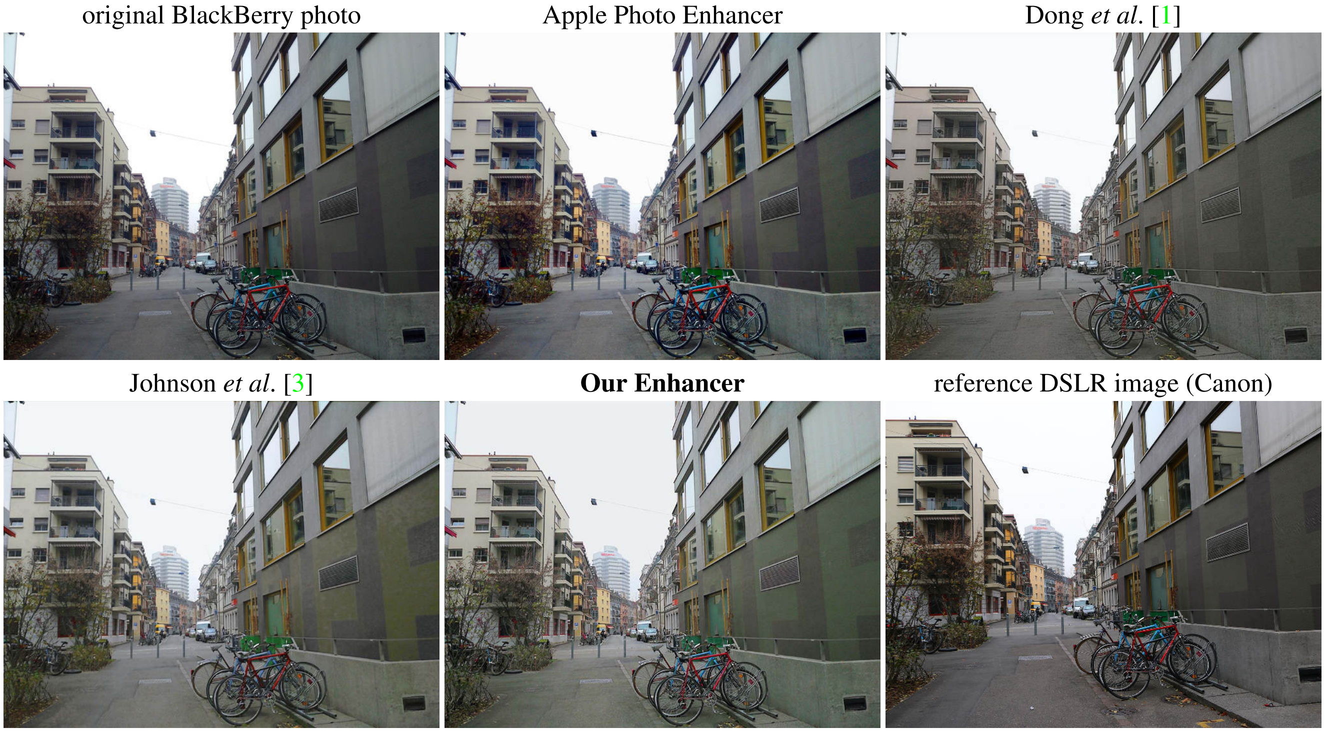

Figure 26: Comparison of 4 enhancers applied to a BlackBerry photo vs. the DSLR reference image (Canon).

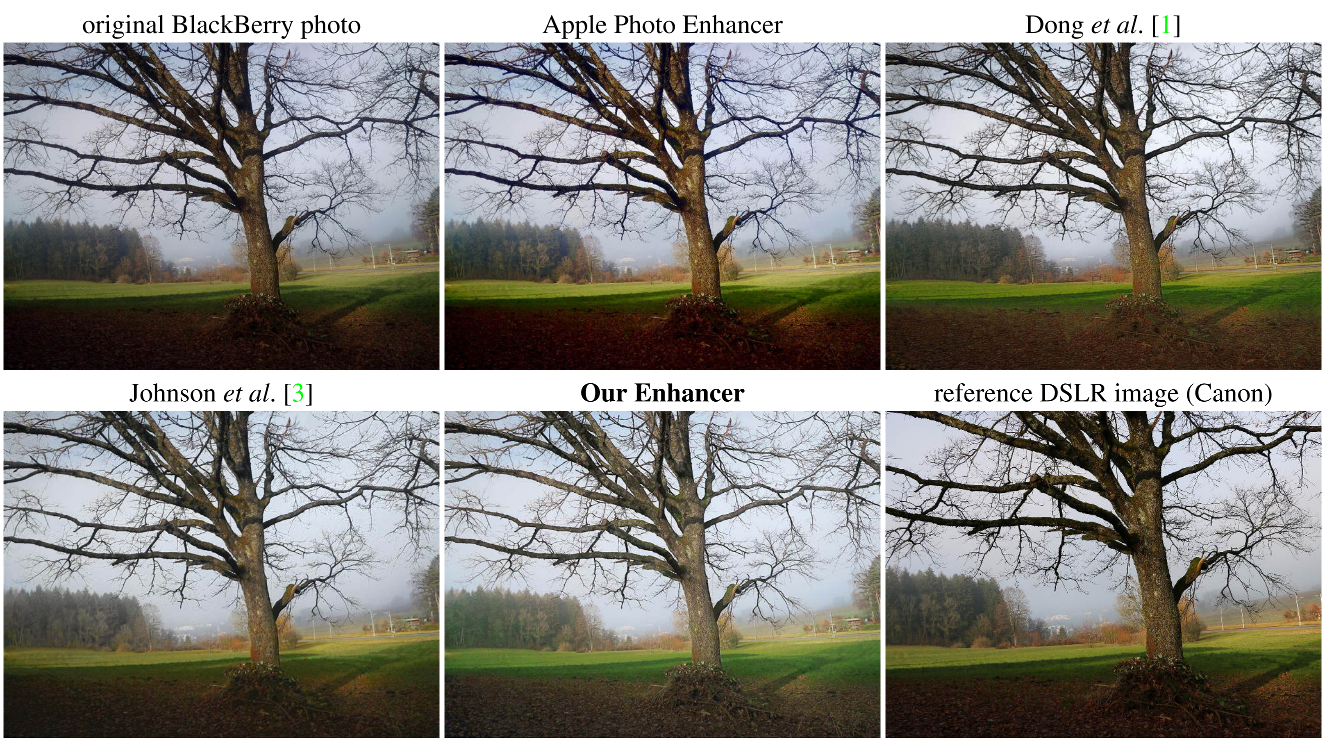

Figure 27: Comparison of 4 enhancers applied to a BlackBerry photo vs. the DSLR reference image (Canon).

7. Comparison with the baselines: Sony

original Sony photo | Apple Photo Enhancer | Dong *et al*. [1]

Johnson *et al*. [3] | **Our Enhancer** | reference DSLR image (Canon)

Figure 28: Comparison of 4 enhancers applied to a Sony photo vs. the DSLR reference image (Canon).

original Sony photo | Apple Photo Enhancer | Dong *et al*. [1]

Johnson *et al*. [3] | **Our Enhancer** | reference DSLR image (Canon)

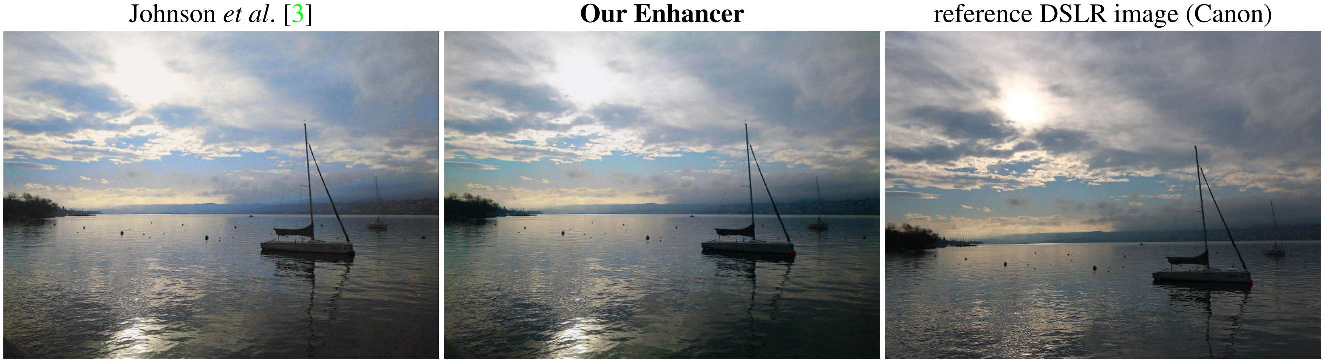

Figure 29: Comparison of 4 enhancers applied to a Sony photo vs. the DSLR reference image (Canon).

Figure 30: Comparison of 4 enhancers applied to a Sony photo vs. the DSLR reference image (Canon).

Figure 31: Comparison of 4 enhancers applied to a Sony photo vs. the DSLR reference image (Canon).

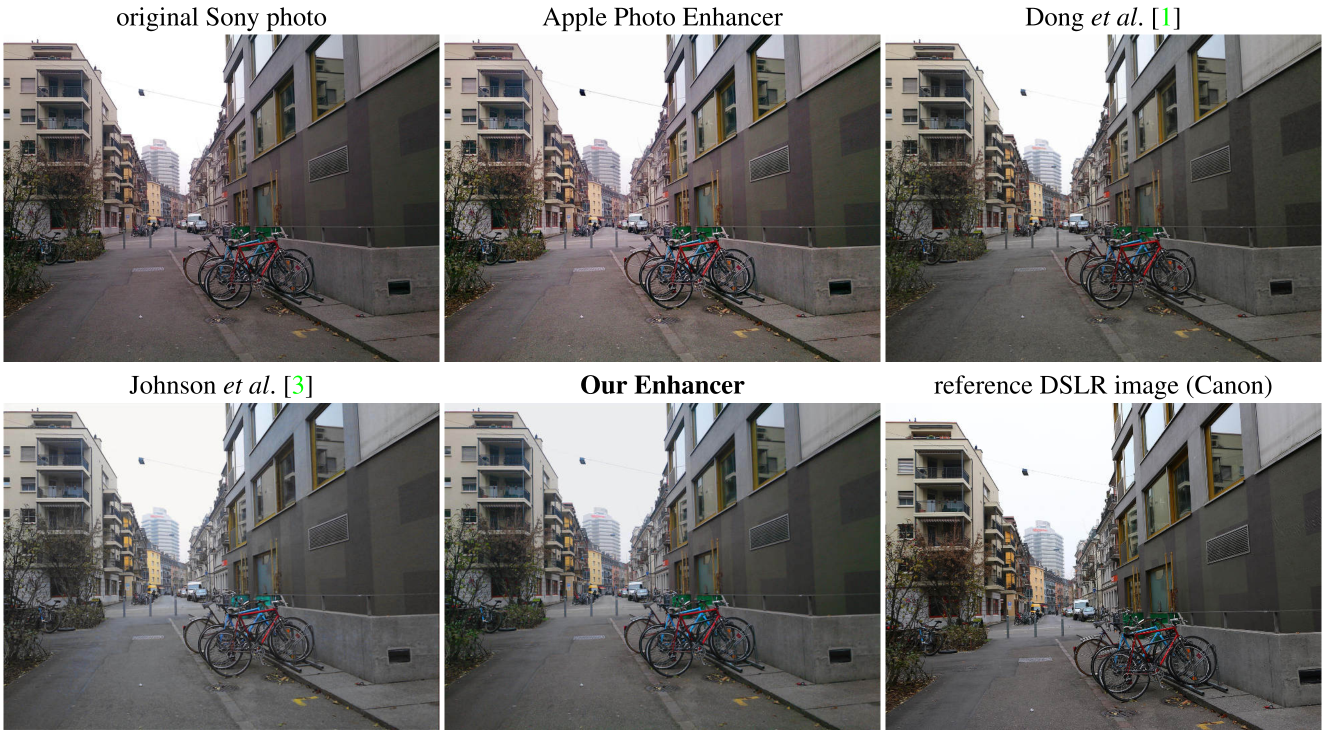

Figure 32: Comparison of 4 enhancers applied to a Sony photo vs. the DSLR reference image (Canon).

Figure 33: Comparison of 4 enhancers applied to a Sony photo vs. the DSLR reference image (Canon).

8. User Study

In this section, we show the detailed results of the user study. The subjects were shown a set of 9 images (see Fig. 11 in the paper) taken by a mobile device, a DSLR camera, and enhanced by the proposed method, Apple Photos enhancer and a human artist. On every question, the subjects were shown a pair of images and were asked to pick the best image.

Table 1: Photo by a DSLR camera preferred over a photo of the same scene taken by a mobile device. The DSLR camera is preferred for a large majority of cases.

	Img 1	Img 2	Img 3	Img 4	Img 5	Img 6	Img 7	Img 8	Img 9	Mean
BlackBerry	0.76	0.67	0.88	0.48	0.71	0.69	0.95	0.90	0.77	**0.76**
iPhone	0.78	0.98	0.90	0.93	0.95	0.90	0.93	0.88	1.00	**0.92**
Sony	0.83	0.77	0.74	0.74	0.93	0.26	0.83	0.83	0.90	**0.76**

Table 2: Photo improved by the proposed method preferred over the same photo without post-processing. The improved photos by the proposed method are preferred for a large majority of cases over the original smartphone photos.

	Img 1	Img 2	Img 3	Img 4	Img 5	Img 6	Img 7	Img 8	Img 9	Mean
BlackBerry	0.60	0.48	0.83	0.50	0.49	0.79	0.98	0.79	0.79	**0.69**
iPhone	0.77	0.93	0.71	0.86	0.74	0.85	0.86	0.79	0.90	**0.82**
Sony	0.62	0.86	0.77	0.62	0.95	0.64	0.69	0.79	0.77	**0.75**

Table 3: Photo by a DSLR camera preferred over the photo from a mobile device improved by the proposed method. Note that our method achieves the targeted quality level of the DSLR camera and that the preferences of the users are almost equally shared by our method and the DSLR camera.

	Img 1	Img 2	Img 3	Img 4	Img 5	Img 6	Img 7	Img 8	Img 9	Mean
BlackBerry	0.45	0.33	0.36	0.56	0.50	0.64	0.50	0.60	0.60	**0.50**
iPhone	0.33	0.21	0.50	0.64	0.64	0.38	0.51	0.31	0.62	**0.46**
Sony	0.33	0.43	0.24	0.38	0.67	0.69	0.34	0.62	0.57	**0.47**

Table 4: Different ways of improving the photos taken by the iPhone camera: ratios of preferences of the second option vs the first. Note the clear preference for the results of our automatic photo enhancement method when compared with the automatic Apple Enhancer and even the retouched photos by human artists.

	Img 1	Img 2	Img 3	Img 4	Img 5	Img 6	Img 7	Img 8	Img 9	Mean
Retouching vs proposed method	0.55	0.36	0.19	0.57	0.83	0.81	0.79	0.79	0.69	**0.62**
Apple Enhancer vs proposed method	0.62	0.95	0.72	0.83	0.83	0.83	0.74	0.71	0.81	**0.78**
Apple Enhancer vs retouching	0.67	0.95	0.76	0.83	0.21	0.21	0.21	0.36	0.71	**0.55**

9. Loss analysis

In this section, we study the contribution of different terms of the proposed perceptual loss function. For this purpose, we consider four different loss combinations: 1) the proposed one [color + content + texture], 2) [content + texture] loss, 3) [MSE + texture] loss and 4) [MSE] loss. For each of these target loss combinations, a CNN was trained on the DPED dataset and validated on its test subset. The results of this experiment are provided in Table 5 and visual results are shown in Fig. 34. As one can see, the adversarial network that stands behind the texture loss can cause significant color deviations, and the additional MSE term cannot effectively suppress them since it is not precise in this task (images are not perfectly aligned). Content loss shows better results in this case since it is less sensitive to image mismatches. Adding an extra color term further improves the resulting images, making the colors more saturated and closer to the target. Single MSE demonstrates high PSNR and SSIM values and natural color rendition while causing strong artifacts and slightly degrading image sharpness. Overall our proposed [color + content + texture] loss leads to the best visual results while at the same time achieves top SSIM scores.

Table 5: PSNR/SSIM scores for different loss functions.

Phone	**Color + Content + Texture**		Content + Texture		MSE + Texture		MSE	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
iPhone	20.08	**0.9201**	19.05	0.9166	20.11	0.9125	**20.56**	0.9198
BlackBerry	20.07	**0.9328**	19.64	0.9312	20.13	0.9241	**20.15**	0.9292
Sony	**21.81**	0.9437	21.59	0.9426	21.72	0.9416	21.35	**0.9453**

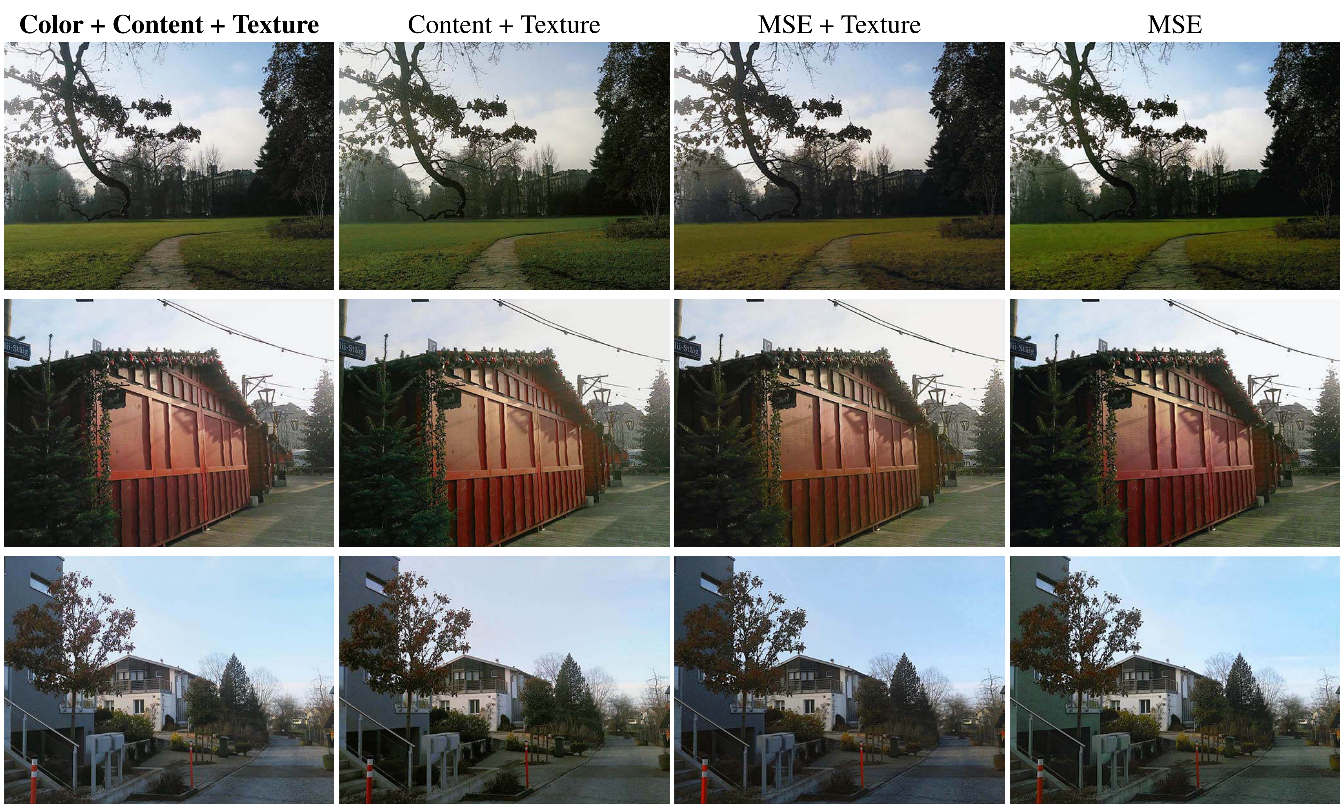

Figure 34: Result images for iPhone camera for 4 different target loss functions.

10. Results of Isola *et al*. [2]

Here, we demonstrate the results of the very recent image to image translation method proposed by Isola *et al*. [2] on our dataset (Figure 35). We trained the method of Isola *et al*. [2] under the same conditions as we did for our method and the other CNNs methods from our paper: Dong *et al*. [1] and Johnson *et al*. [3]. On one side, the network of Isola *et al*. delivers excellent results on adjusting image colors and brightness. However, due to the deconvolution mechanism of the suggested CNN architecture, the results have a strong checkerboard pattern. Note that the same artifacts are visible on most of the reported image results by the original authors [2]. Therefore, the considered network of Isola *et al*. is a suboptimal instrument for enhancing pictures taken by cameras of mobile devices, as for this task both high-resolution and high-quality output are required.

Figure 35: Example results by the encoder-decoder architecture [2].

References

[1] C. Dong, C. C. Loy, K. He, and X. Tang. *Learning a Deep Convolutional Network for Image Super-Resolution*, pages 184–199. Springer International Publishing, Cham, 2014. 1, 20, 21, 22, 23, 24, 25, 26, 27, 28, 31

[2] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *arxiv*, 2016. 1, 31

[3] J. Johnson, A. Alahi, and L. Fei-Fei. *Perceptual Losses for Real-Time Style Transfer and Super-Resolution*, pages 694–711. Springer International Publishing, Cham, 2016. 1, 20, 21, 22, 23, 24, 25, 26, 27, 28, 31

[2] https://phillipi.github.io/pix2pix/images/cityscapes_cGAN_AtoB/latest_net_G_val/index.html