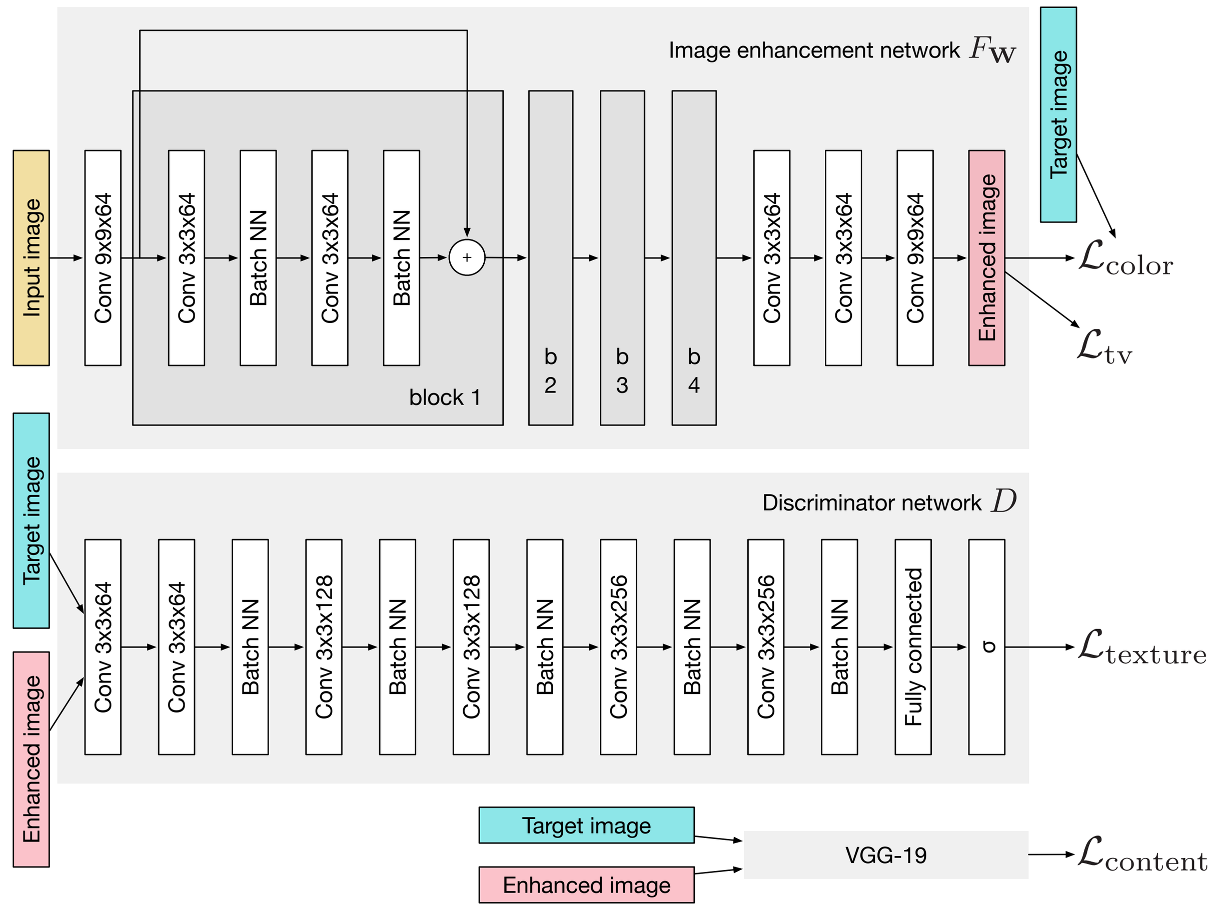
DSLR-Quality Photos on Mobile Devices with Deep Convolutional Networks

Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, Luc Van Gool **dped-photos.vision.ee.ethz.ch** Computer Vision Laboratory, ETH Zurich, Switzerland

Contribution

- A novel approach for the **photo enhancement task** based on learning a mapping function between photos from mobile devices and a DSLR camera.
- A new large-scale **DPED dataset** consisting of over 22K photos taken synchronously by a DSLR camera and 3 low-end cameras of smartphones.
- Experiments measuring objective and subjective quality **demonstrating** the advantage of the enhanced photos over the originals and, at the same time, their **comparable quality with the DSLR** counterparts.

Proposed Photo Enhancer



Given a low-quality photo I_s (source image), the goal is to reproduce the image I_t (target image) taken by a DSLR camera. A deep residual CNN $F_{\mathbf{W}}$ is used to learn this translation function, and is trained to minimize a loss function consisting of the following terms:

• **Color loss:** To measure the color difference between the enhanced and target images, we propose applying a Gaussian blur and computing Euclidean distance between the obtained representations:

$$\mathcal{L}_{\text{color}}(X,Y) = \|X_b - Y_b\|_2^2,$$

where X_b and Y_b are the blurred images X and Y.

• **Texture loss:** We build upon GANs to learn a suitable metric for measuring texture quality. The discriminator CNN observes fake (improved) and real (target) grayscale images, and its goal is to predict whether the input image is real or not. It is trained to minimize the cross-entropy loss function, and the texture loss is defined as a standard generator objective ($F_{\mathbf{W}}$ and D – generator and discriminator):

$$\mathcal{L}_{\text{texture}} = -\sum_i \log D(F_{\mathbf{W}}(I_s), I_t).$$

where $F_{\mathbf{W}}$ and D – generator and discriminator nets.

• **Content loss:** We define our content loss based on the activation maps $\psi_j()$ produced by the ReLU layers of the pre-trained VGG-19 network:

$$\mathcal{L}_{\text{content}} = \alpha \|\psi_j(F_{\mathbf{W}}(I_s)) - \psi_j(I_t)\|.$$

• **Final loss:** The final loss is defined as a weighted sum of previous losses with the following coefficients:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{content}} + 0.4 \cdot \mathcal{L}_{\text{texture}} + 0.1 \cdot \mathcal{L}_{\text{color}} + 400 \cdot \mathcal{L}_{\text{tv}}$$

Proposed DPED Dataset

DPED dataset consists of photos taken in the wild synchronously by three smartphones and one DSLR camera. The devices were mounted on a tripod and activated remotely by a wireless control system. The photos were captured in automatic mode during the daytime in a wide variety of places and in various illumination and weather conditions.

Camera	Sensor	Photo quality
iPhone 3GS	3 MP	Poor
BlackBerry Passport	13 MP	Mediocre
Sony Xperia Z	13 MP	Average
Canon 70D DSLR	20 MP	Excellent

Experiments

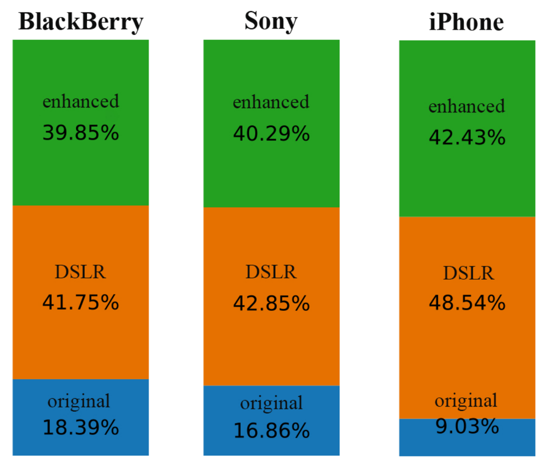
We quantitatively compare Apple Photo Enhancer (APE), Dong et al., Johnson et al., and our method on the considered task. First, one can note that our method is the best in terms of SSIM, at the same time producing images that are cleaner and sharper, thus perceptually performs the best. On PSNR terms, our method competes with the state of the art: it slightly improves or worsens depending on the dataset, i.e., on the actual phone used: alignment issues could be responsible for these minor variations.

Phone	APE		Dong et al.		Johnson et al.		**Ours**	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
iPhone	17.28	0.8631	19.27	0.8992	**20.32**	0.9161	20.08	**0.9201**
BlackBerry	18.91	0.8922	18.89	0.9134	**20.11**	0.9298	20.07	**0.9328**
Sony	19.45	0.9168	21.21	0.9382	21.33	0.9434	**21.81**	**0.9437**

From left to right, top to bottom: original iPhone photo and the same image after applying, respectively: APE, Dong et al., Johnson et al., our generator network, and the corresponding DSLR image.

User Study

To measure overall quality we designed a no-reference user study where subjects are repeatedly asked to choose the better looking picture out of a displayed pair. The results indicate that in all cases both pictures taken with a DSLR as well as pictures enhanced by the proposed CNN are picked much more often than the original ones taken with the mobile devices. When subjects are asked to select the better picture among the DSLR-picture and our enhanced picture, the choice is almost random. This means that the quality difference is inexistent or indistinguishable, and users resort to chance.



Acknowledgements

The work is supported by the ETH Zurich General Fund (OK), Toyota via the project TRACE-Zurich, the ERC grant VarCity, and an NVidia GPU grant.

Visual Results

iPhone 3GS BlackBerry Passport Sony Xperia Z

Limitations

Two typical artifacts that can appear on the processed images are color deviations and high contrast levels. Although they often cause rather plausible visual effects, in some situations this can lead to content changes that may look artificial, i.e. greenish asphalt in the second image. Another notable problem is noise amplification – due to the nature of GANs, they can effectively restore high frequency-components. However, high-frequency noise is emphasized too. Note that this noise issue occurs mostly on the lowest-quality photos (i.e., iPhone), not on the better phone cameras.

Try it out at phancer.com: