

LFIEM: Lightweight Filter-based Image Enhancement Model

Oktai Tatanov¹ {oktai.tatanov@vk.com}, Aleksei Samarin^{1,2} {aleksei.samarin@vk.com} ${}^{1}_{VK}$ ${}^{2}Saint-Petersburg State University$



Abstract

Photo retouching features are being integrated into a growing number of mobile applications. Current learning-based approaches enhance images using large convolutional neural network-based models, where the result is received directly from the neural network outputs. This method can lead to artifacts in the resulting images, models that are complicated to interpret, and can be computationally expensive. In this paper, we explore the application of a filter-based approach in order to overcome the problems outlined above. We focus on creating a lightweight

solution suitable for use on mobile devices when designing our model. A significant performance increase was achieved through implementing consistency regularization used in semi-supervised learning. The proposed model can be used on mobile devices and achieves competitive results compared to known models.

Proposed pipeline

Illustration of our pipeline is shown in the figure. It consists of several independent blocks, the

Filters description

Below we provide description of image transfor-

number of which depends on the number of used filters. In every *i*-th block, a small version of the original image I_{so} passes through the parameter generator h_i that produces parameters p_i for the corresponding filter f_i . After separately applying the filters to the original image I_o , the final enhanced image I_e is the sum of the original image and filter outputs. Overall, our model can be written as

$I_e = I_o + \sum_{i=1}^{n} f_i(I_o, h_i(I_{so})),$

where n is number of selected filters.



mations (filters) from the best configuration of our model.

We used a unified notation for filters: I_{in} is the input image, I_{out} is the output image, (x, y)stands for image pixel coordinates, c stands for a color channel (red, green or blue) and t, u are trainable parameter unless otherwise specified. To achieve good time performance, we used only the RGB color space for image representation. For uniform description, we used channel values in the [0, 1] range.

The following image transformation performs automatic *exposure* correction:

$$I_{out}[x,y] = I_{in}[x,y] \cdot 2^t.$$

The trainable *linear* image *transformation*, is an additional important mapping. It can be described with the following expression:

 $I_{out}[x,y] = P \cdot I_{in}[x,y] + b,$

where $P \in \mathbb{R}^{3 \times 3}$ stands for the trainable affine

Example of image enhancement by the proposed model on a photo from the MIT Adobe FiveK dataset. (Left) Original image. (Right) Enhanced image by our best model. This model does not lead to artifacts and produces good results even on detailed images.



Overview of the proposed model on the example of two filters (contrast and exposure). The resized original image passes through parameter generators that produce filter parameters. The final enhanced image is the

mapping matrix, $b \in \mathbb{R}^3$ – trainable vector in RGB color space.

Finally, the *channel-wise* image *color transformation* was also used. The transformation is composed of a triplet of functions that are applied to the red, green, and blue color channels respectively. Each function is a linear combination of the elements $f_1, f_2, ..., f_n$ of a n-dimensional basis, the coefficients for which are calculated from the output of the neural network. Therefore, a channel value for each pixel of the input image is evaluated by the following formula:

$$I_{out}^{c}[x,y] = I_{in}^{c}[x,y] + \sum_{i=1}^{n} u_{ic} \cdot f_i(I_{in}^{c}[x,y]),$$

where $f_1, f_2, ..., f_n$ – functional basis mentioned above, and u_c – trainable parameters (one parameter for each channel).

Because of its proven effectiveness, we considered only the piece-wise basis and used the fol-

sum of the original image and filter outputs.

Conclusion

We have proposed a lightweight image enhancement model. The lack of complex architecture and the small number of weights (101,547 trainable parameters) allows deploying our solution on mobile devices. In addition, our model outperforms the vast majority of known image enhancement models and contains at least ten times less parameters than most of them. We used a contemporary consistency regularization approach that allowed our model to achieve the best results in comparison to other models. In the process of creating our model, we also performed a comprehensive comparative analysis of various transformations used to improve image quality.

Acknowledgements

Authors are thankful to Pavel Kalaidin for useful discussions, Viktoriya Loginova for proofreading.

lowing set of functions:

 $f_i(x) = \max(0, 1 - |(n - 1) \cdot x - i + 1|), i \in 1..n,$

where x is a value of the current pixel of the input image.

Quantitative results

Pix2Pix 54M - 0.857	
Dist. & Rec. 153M - 0.905	
DPE 2.2M 23.89 0.906	
CE+PRNL $>30M$ 24.19 0.915	!
Ours 101K 24.77 0.911	