https://doi.org/10.2352/EI.2022.34.14.COIMG-218

FiveNet: joint image demosaicing, denoising, deblurring, super-resolution and clarity enhancement

Mykola Ponomarenko[1], Vladimir Marchuk[2] and Karen Egiazarian[1]
[1]Tampere University, Tampere, Finland
[2] Don State Technical University, Rostov-on-Don, Russia

Abstract

In this paper, a convolutional neural network for joint image demosaicing, denoising, deblurring, super-resolution and clarity enhancement is proposed. The network inputs are four-channel Bayer CFA image (R, G, G, B) and three channels of the same size containing distortions maps, namely, noise level map, blur level map, and clarity degradation map. It is shown that the designed network FiveNet can effectively process images with the mix of five different distortions. It is also demonstrated that adding clarity enhancement into the processing chain can additionally increase image quality (by up to 3-4 dB in PSNR). A small dataset ClarityDegr120 of color images with different clarity degradations and enhancements is designed using images processed by FiveNet. Mean opinion scores (MOS) for the test set are collected. The MOS prove that clarity enhancement can significantly increase image visual quality. A comparative analysis using the MOS demonstrates a low correspondence between image quality metrics and human perception for the clarity enhancement task.

Introduction

During last decade convolutional neural networks (CNN) show great progress in solving various image processing and analysis tasks. CNNs provide state-of-the-art performance quality in image denoising, demosaicing, deblocking, restoration, and super-resolution [1-3]. Now it is the area of intensive research, which goes in many directions.

One of them is automation of image denoising and image restoration. Several fully blind methods, including DnCNN [4], CBDNet [5], VDNet [6] were proposed. At the same time, better results and universality are demonstrated by the methods which take noise levels map as an additional input of CNN (for example, DRUNet [1]). This noise level map should be preliminary estimated.

Another research direction is a design of end-to-end solutions, which combine several image enhancement routines in one CNN. For example, JDnDmSR+ CNN can simultaneously perform image demosaicing, denoising and super-resolution, as well as any combination of these tasks, providing better quality of processed images than a sequential step by step processing [3]. Such an end-to-end solution potentially allows to replace whole processing chains of digital cameras by a single CNN.

This paper is dedicated to CNN design for the problem of joint demosaicing, denoising, deblurring, super-resolution and clarity enhancement, and addresses the following two main challenges.

The first challenge is a simultaneous use of three distortions maps as inputs of the same CNN. To provide an effective CNN's training, these maps should have similar dynamic range of values, and all combinations of distortions should be covered by the training set.

The second challenge is a CNN-based implementation of clarity enhancement [7-9]. In contrast with denoising, clarity enhancement can be used to increase visual quality of noise free and sharp images. However, to the best of our knowledge, there are no image databases with MOS containing clarity changes. Because of this, full-reference and no-reference metrics are neither trained nor verified for clarity distortions and enhancements. This complicates a usage of the metrics in the CNN's training.

Clarity is applied to a relatively large image areas, and selected CNN's architecture should be suitable for that. At the same time, the architecture should be also suitable for inverse imaging tasks, such as demosaicing, denoising, deblurring and super-resolution. This adds an extra difficulty to the task.

In the paper, we propose a novel neural network architecture able to perform end-to-end image processing including clarity enhancement. An intensive numerical analysis of the network's efficiency including collecting MOS is provided.

Proposed network

A structural scheme of the proposed CNN is shown in Fig. 1. We called this network FiveNet, because it simultaneously performs five image enhancement routines: demosaicing, deblurring, denoising, super-resolution and clarity enhancement, at five image scales, with five skip connections between the scales.

FiveNet has an input with 7-channels (4 channels of RGGB of Bayer CFA, clarity degradation map, blur level map and noise level map) and 3-channels of RGB output. Due to demosaicing and super-resolution by the factor of 2, output dimensions are four times larger than the dimensions of the input. For training we have used 64x64x7 input patches and 256x256x3 output patches. A pretrained network can work with 7-channels input of any dimensions.

As a base, we have utilized DRUNet denoising network [1] which combines U-Net [10] and ResNet [11]. We added a transposed convolution and residual blocks in the network input to interpolate Bayer CFA image. After this, we added another transposed convolution to prepare a preliminary enlarged image for super-resolution. Also, we added one more scaling level to the network, i.e., there are four downscaling and upscaling cascades in the network in comparison to three in the DRUNet.

Added new scaling level is important for clarity enhancement. To enhance image clarity, one should be able to process relatively large local areas of size 32x32 pixels or even 64x64 pixels. Due to added downscaling level, this network in the largest scale works with pixels of size 16x16. For 3x3 convolutions this means regions of size 48x48. Several residual blocks in a row increase the size of the processed area even more.

Generation of input patches

The training set should contain as many as possible images with excellent visual quality. It can provide in some extent the presence in the training set of large number of image regions with good image clarity.

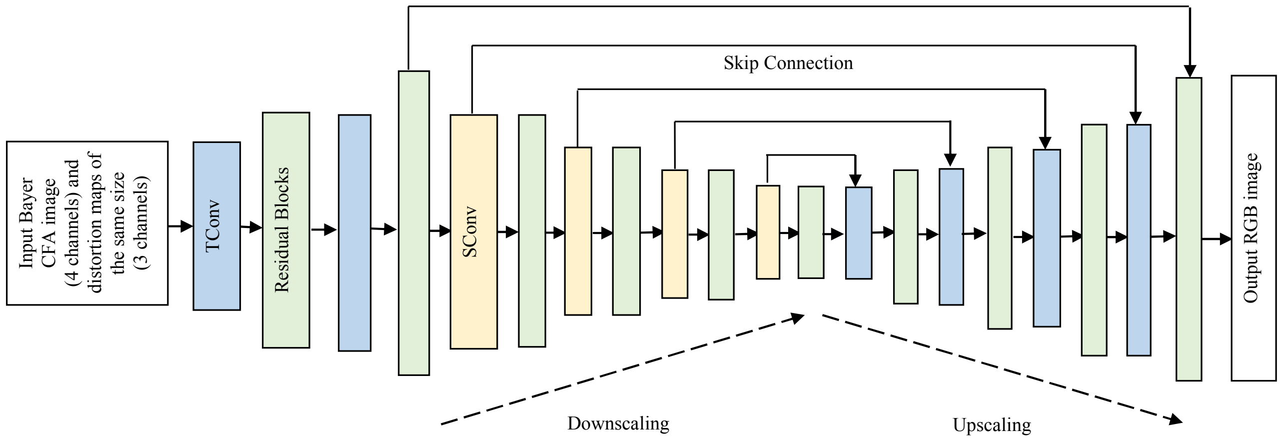


Fig. 1. Structure scheme of the proposed network for end-to-end image enhancement (TConv – Transpose convolution, SConv - Stride convolution)

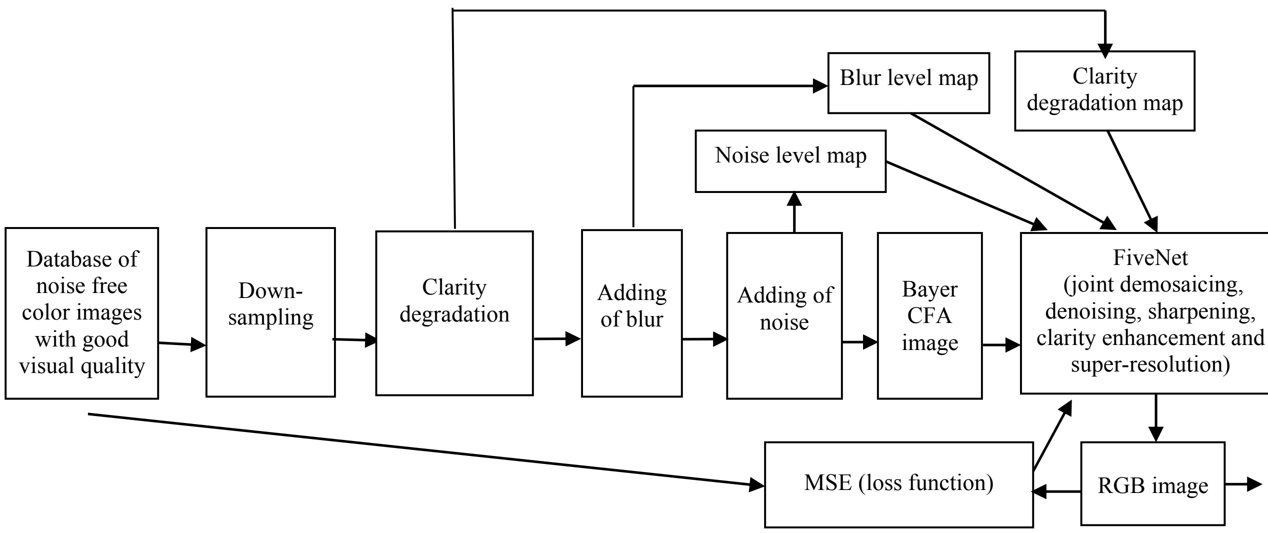


Fig. 2. Structural scheme of FiveNet training

For the training set, we used only images with low level of noise and good visual quality. A part of images with the largest MOS values were selected from image databases with MOS (KonIQ10k [12], FLIVE [13], NRTID [14] and SPAQ [15]).

A part of images was collected from different publicly available image databases (such as DIV2K [16], Flickr2K [17]) using quality predictions of KonCept512 metric [12]. In total, 3360 images were selected.

Fig. 2 shows a scheme of FiveNet training and order of steps of generation of training patches and distortion maps.

Adding noise and blur to an image are well formalized and simple routines. At the same time there is no formalized definition of image clarity in a theory as well as of process of clarity degradation. Moreover, known practical implementations of clarity degradation such as Photoshop and Matlab's local Laplacian filtering [18] often changes brightness of large homogeneous areas (see Fig. 3). It is clearly seen from that Figure that brightness of sky region on both images Fig.3,b and Fig.3,c is changed. As a result, according to distortion map Fig.3,d, it is the most distorted image region. But it should not be for clarity degradation. For the training we need a clarity degradation routine in the meaning "decreasing of local contrasts", but not in the meaning "changing brightness of large flat gradients".

Let us describe in detail the clarity degradation algorithm used in the paper.

For FiveNet training we have used input patches of size 256x256 pixels. We cannot use smaller patches, because for the largest downscaling a patch of size 256x256 transforms into 16x16 pixels. It is barely enough for network training without any serious influence of edge effects.

To decrease image clarity, we have used Matlab's local Laplacian filtering.

As a first step we have applied Matlab's *locallapfilt(I, 0.5, 2)* to a given ground truth image ***I***. A resulting image ***A*** is shown in Fig. 7, b.

a b c d

Fig. 3. Illustration of a typical problem with practical algorithms of clarity degradation: a) source image, b) result of clarity degradation by Photoshop, c) result of clarity degradation by Matlab's local Laplacian filtering, d) distortion map for the image (c)

a b c d e

Fig. 4. Example of clarity decreasing in a training patch generation: a) ground truth image I, b) image **A**, c) image **B**, d) image **C** for K=0.8, e) Image **C** after adding a blur

Then we calculated image $\boldsymbol{B}$ as $\boldsymbol{I} + (\boldsymbol{A}-\boldsymbol{I})F(\boldsymbol{I})$, where $F(\boldsymbol{I})$ is a weighting function of pixel intensity values.

Proposed $F(\boldsymbol{I})$ for the range 0…255 is given in Fig. 5 and provides decreasing of local contrasts only for the middle tones.

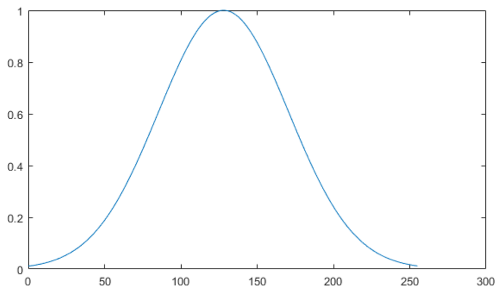


Fig. 5. Weighting function provides larger distortions for middle tones

We stretch this function to actual dynamic range of a given image. A resulting image $\boldsymbol{B}$ is shown in Fig. 4, c.

Finally, image $\boldsymbol{C}$ is created as $\boldsymbol{I} + (\boldsymbol{B}-\boldsymbol{I})K$, where K (0..1) is a selected strength of clarity decreasing. Image $\boldsymbol{C}$ for K=0.8 is shown in Fig. 4, d.

After clarity decreasing, we added a blur to the processed path using Matlab's *imgaussfilt(C, blurl)*, where *blurl* is a selected blur level. Image $\boldsymbol{C}$ from the Fig. 4, d after adding blur with *blurl*=0.55 is shown in Fig. 4, e.

Finally, after clarity degradation and blur, we add to the patch additive white Gaussian noise (AWGN) independently for each RGB color channel.

In our training patch generation, we used random K from 0 to 1, *blurl* selected as *0.35+abs(randn)* and standard deviation of AWGN selected as *abs(randn)*25*.

To provide robustness of trained FiveNet to situations where only one or two distortions are presented, we trained FiveNet as it is shown in the Table 1.

Table 1. Distortions in generated patches

Presented distortions (C - clarity degradations, B - blur, N - noise)	Percentage
C and B and N	65%
None	5%
B and N	5%
C and B	5%
C and N	5%
N	5%
B	5%
C	5%

All distortions maps are calculated as smoothed versions of absolute differences between input and distorted patches.

Training environment and parameters

FiveNet training was performed in Matlab R2021a environment using a computer with 24 GB GPU. In our experiments, we have used miniBatchSize = 8, "Adam" optimizer, LearningRate 0.00001 with decay 0.000001 for 130000 iterations.

Training patches were cropped in a custom training loop from images of the training set. To increase the presence of patches with textures and fine details, we have used a technique described in [19]. Additional mirroring and rotations were applied to reduce overlearning.

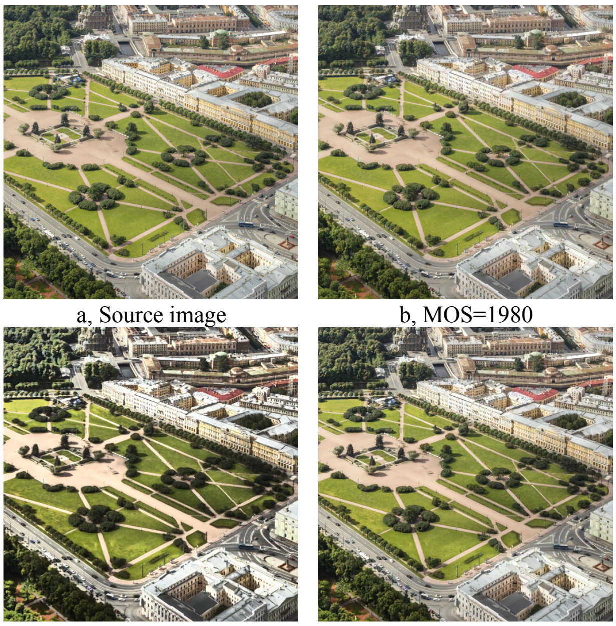
We pretrained FiveNet using MSE as a loss function for 100000 iterations. After this, we continued training and trained three separate versions of FiveNet for MSE, MAE and wMSE (with W=5) [20] loss functions adding another 30000 iterations.

FiveNet contains many layers with 64, 128, 256 and 512 filters. Because of this, it takes 115 Mb in the memory. To optimize the network size, we trained a lite version of FiveNet. We divided all filter numbers by 2, except layers with 64 filters. As a result, the network called FiveNetLite takes only 34 Mb in memory. We trained FiveNetLite using MSE loss function for 130 000 iterations.

Experiments with collecting MOS

It is unclear which metric corresponds well to human perception for the case of clarity degradation. To clarify this question at least to a certain degree, we created a small image dataset and collected mean opinion scores (MOS) for the dataset. It allows to carry out of a comparative analysis of different full-reference and no-reference metrics. Also, we can compare results of FiveNet training for different loss functions.

For test set generation, we used ground truth image presented in Fig. 6, a.



a, Source image b, MOS=1980

c, MOS=1640 d, MOS=2275

Fig. 6. Images without noise and blur for different levels of clarity enhancement

The test set contains 120 images including both distorted and processed by FiveNet images. Distorted images were obtained by applying to distorted Bayer CFA images Matlab's "demosaic" function. To eliminate an influence of strong residual errors of super-resolution, we downscaled by 2 (without interpolation) all images in the output of FiveNet. Due to this all images in the test set are 512x512 pixels and do not have residual errors of a super-resolution process.

Table 2 includes list of images in the test set, which is called ClarityDegr120. Here M=0.75 means that clarity degradation map is multiplied by 0.75.

MOS for the ClarityDegr120 test set were collected using Glicko rating system and pairwise comparisons. Four experienced observers collected 1800 judgments (30 for each image). On each step of experiment image with better visual quality from a proposed pair was selected.

Table 2. List of images in the ClarityDegr120 test set

Position in the test set	Description	Processing
001	Downscaled source image	-
002	Distorted, only mosaicing (no other distortions)	-
003	Distorted, only mosaicing	FiveNet+MSE
004-008	Distorted, only mosaicing	FiveNet+MSE, increased clarity with different levels
009-015	Source image with AWGN with different levels	-
016-050	Random distortions parameters	FiveNet+MSE
051-060	10 distorted images with selected mix of distortions	-
061-070	The same 10 images	FiveNet+MSE,M=0.75
071-080	The same 10 images	FiveNet+MSE
081-090	The same 10 images	FiveNet+MSE,M=1.5
091-100	The same 10 images	FiveNet+MSE,M=2.5
101-110	The same 10 images	FiveNet+MAE
111-120	The same 10 images	FiveNet+wMSE

We calculated for images of ClarityDegr120 test set several no-reference image quality metrics. Also, for images 2-120 using 1-st image as a reference image we calculated several full-reference image visual quality metrics. Table 3 contain Spearman rank order correlation coefficient (SROCC) between the MOS and considered metrics.

Table 3. SROCC between MOS of ClarityDegr120 and considered metrics values

	Metric	SROCC
No-reference metrics	KonCept512 [12]	0.9
	FISH [21]	0.61
	SMetric [22]	0.62
Full-reference metrics	PSNR	0.84
	MDSI [23]	0.95
	PSNR-HVS-M [24]	0.83
	SSIM [25]	0.92
	CSSIM [26]	0.87
	FSIMc [27]	0.91

ClarityDegr120 set contains many "simple" for the metrics distorted images. Due to this, SROCCs in the table are relatively large and not too informative. However, the Table 4 helps to select a metric which potentially can be used in clarity enhancement tasks. However, more statistics should be collected to make reliable conclusions.

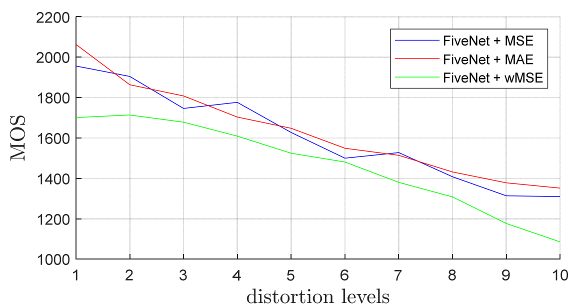


Fig. 7. Comparison of FiveNet pretrained with different loss functions

Fig. 7 shows curves of MOS for FiveNet trained with different loss functions. One can see, that FiveNet with MSE and MAE provides comparable results while combination of FiveNet with wMSE definitely fails for the task.

Fig. 8 shows curves for different multiplication factor for clarity degradation map.

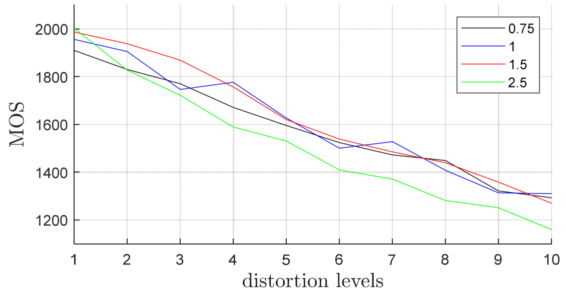


Fig. 8. Comparison of FiveNet with different multiplication factor for input clarity degradation map

It looks like the factor 2.5 is not good for large distortions (noise and blur). However, there is not enough information to make confident conclusions, especially for small distortion levels. More precise experiments are needed.

Fig. 6 shows images for "only demosaicing case" with artificially added clarity enhancement. There is no clarity enhancement for the image Fig 6,b (MOS=1980) and different levels of enhancement for other images. It is clearly seen, that a moderate clarity enhancement can significantly increase image visual quality (Fig. 6,d, MOS=2275). At the same time too much clarity increase can significantly downgrade image quality (Fig. 6,c, MOS=1640).

Thereby, a good method of estimation of an optimal clarity enhancement level for image regions is needed. Design of corresponding algorithms is actual.

Fig. 10 shows two curves. The first is MOS versus PSNR for images 004-008 + 003 (clarity increase). The second is MOS versus PSNR for images 009-0015 (AWGN).

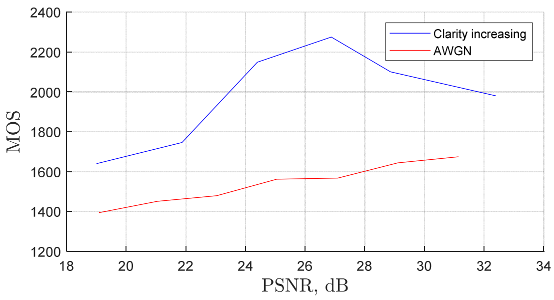


Fig. 10. PSNR vs MOS for clarity increase and AWGN

It is clearly seen that PSNR significantly overestimates clarity distortions. It is also interesting, that there is an optimum on the clarity increase curve.

Comparative analysis of designed networks for end-to-end image enhancement

In this section we will compare effectiveness of FiveNet+MSE, FiveNet+MAE, FiveNet+wMSE and FiveNetLite. For the purpose we selected set of 24 color 512x512 images, which were not used in FiveNet training.

Demosaicing and super-resolution

Super-resolution is the most difficult and non-effective operation in the list of processing routines (super-resolution, demosaicing, denoising, clarity enhancement). Let us show it.

Table 4 contains results for the super-resolution and demosaicing case (there is no noise, no blur and no clarity degradation on distorted images). Here and bellow distorted images are Matlab's "demosaic" images after bicubic upsampling.

Table 4. Demosaicing and super-resolution, PSNR, dB

Noisy	FiveNet+MSE	FiveNet+MAE	FiveNet+wMSE	FiveNetLite
23.2	24.3	24.3	23.8	24.2

It is clearly seen that PSNR for processed images is increased only on 1 dB in comparison to distorted images. The average PSNR value 24.3 dB is the maximal PSNR value which can be reached for this task for the test set even without presence of noise, blur, and clarity degradation. The main factor which limits quality of processed images is the difficulty of super-resolution task.

Clarity enhancement

Hereinafter we will assume that we have estimated map of distortions. Table 5 contains results for super-resolution, demosaicing and clarity enhancement (there is no noise and no blur on distorted images).

Table 5. Demosaicing, super-resolution and clarity enhancement, PSNR, dB

Noisy	FiveNet+MSE	FiveNet+MAE	FiveNet+wMSE	FiveNetLite
22.9	24.2	24.1	23.7	24.0
21.6	23.6	23.5	23.1	23.5
20.2	22.8	22.7	22.3	22.7
19.3	22.1	22.2	21.8	22.0

It is clearly seen that by a clarity enhancement it is possible to increase image quality up to 3 dB. Note that FiveNetLite provides almost the same results as FiveNet.

Processing of images with presence of blur, noise and clarity degradations

Table 6 contains results for simultaneous processing of all distortions (demosaicing, deblurring, denoising, clarity enhancement and super-resolution).

Table 6. Simultaneous processing of all distortions, PSNR, dB

Noisy	FiveNet+MSE	FiveNet+MAE	FiveNet+wMSE	FiveNetLite
22.5	24.1	24.0	23.4	23.9
21.1	23.3	23.3	22.6	23.2
19.2	21.6	21.5	21.0	21.5
17.7	20.5	20.5	19.9	20.5

Results in Table 6 confirm ability of FiveNet to simultaneously perform several enhancement routines, significantly increasing image visual quality.

The difference between FiveNet and FiveNetLite is still small (0.1 dB PSNR or 0.5 dB MDSIPSNR) as well as difference between FiveNET+MSE and FiveNet+wMSE. Results in the Table 6 show that wMSE metric is not suitable for FiveNet training.

Conclusions

The proposed FiveNet is the first CNN, which can perform simultaneously five image enhancement routines effectively taking into accounting three maps of distortions levels.

It is shown that FiveNet can significantly (up to 3-4 dB in PSNR) increase quality of processed image by a clarity enhancement.

Pretrained FiveNet and ClarityDegr120 image set with MOS are available in http://ponomarenko.info/fivenet.

Acknowledgments

Vladimir Marchuk would like to acknowledge the financial support of the Russian Federation represented by the Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-15-2021-997 of 09/28/2021), and Mykola Ponomarenko - the financial support of Huawei-Tampere University project 3114100158, FlexISP.

References

[1] K. Zhang, Y. Li, W. Zuo, L. Zhang, L.Van Gool, R. & Timofte, “Plug-and-play image restoration with deep denoiser prior”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 17 p, 2021.

[2] A. Dhanalakshmi, G. Nagarajan, “Convolutional Neural Network-based deblocking filter for SHVC in H. 265”, Signal, Image and Video Processing, 14:1635-1645, 2020.

[3] W. Xing, K. Egiazarian, “End-to-End Learning for Joint Image Demosaicing, Denoising and Super-Resolution”, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3507-3516, 2021.

[4] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising,” IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3142–3155, 2017.

[5] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, “Toward convolutional blind denoising of real photographs”, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1712–1722, 2019.

[6] Z. Yue, H. Yong, Q. Zhao, L. Zhang, and D. Meng, “Variational denoising network: Toward blind noise modeling and removal”, arXiv preprint arXiv:1908.11314, 2019.

[7] X. Ji, J. Cheng, J. Bai, T. Zhang, M. Wang, “Real-time enhancement of the image clarity for traffic video monitoring systems in haze”, In 2014 7th International Congress on Image and Signal Processing, pages 11-15, 2014.

[8] Y.Q. Zhang, Y. Ding, J.S. Xiao, J. Liu, Z. Guo, “Visibility enhancement using an image filtering approach”, EURASIP Journal on Advances in Signal Processing, (1):1-6, SpringerOpen, 2012.

[9] X. Bai, Y. Li, F. Zhou, “Measure of image clarity using image features extracted by the multiscale top-hat transform”, Journal of Optics, 14(4):045402, IOP Publishing, 2012.

[10] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241, 2015.

[11] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, 2016.

[12] V. Hosu, H., Lin, T., Sziranyi, D. Saupe, “KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment”, IEEE Transactions on Image Processing, 29, pp. 4041-4056, 2020.

[13] Z. Ying, H. Niu, P. Gupta, D. Mahajan, D. Ghadiyaram, A. Bovik, “From patches to pictures (PaQ-2-PiQ): Mapping the perceptual space of picture quality”, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3575-3585, 2020.

[14] N. Ponomarenko, O. Eremeev, K. Egiazarian, V. Lukin, “Statistical evaluation of no-reference image visual quality metrics”, in Proceedings of EUVIP, Paris, France, 5p, 2010.

[15] Y. Fang, H. Zhu, Y. Zeng, K. Ma, and Z. Wang, “Perceptual quality assessment of smartphone photography”, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3677-3686.

[16] E. Agustsson, R. Timofte, “NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study”, In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2017.

[17] B. Lim, S. Son, H. Kim, S. Nah, K. Mu Lee, “Enhanced deep residual networks for single image super-resolution”, In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 136-144, 2017.

[18] S. Paris, S.W. Hasinoff, and J. Kautz, "Local Laplacian filters: edge-aware image processing with a Laplacian pyramid", ACM Trans. Graph. 30.4: 68, 2017.

[19] M. Ponomarenko, S.G. Bahnemiri, K. Egiazarian, "Deep Convolutional Network for Spatially Correlated Rayleigh Noise Suppression on TerraSAR-X Images", in Proceeding of IEEE Ukrainian Microwave Week (UkrMW), pp. 458-463, 2020.

[20] N. Ponomarenko, S. Krivenko, K. Egiazarian, V. Lukin, J. Astola, “Weighted mean square error for estimation of visual quality of image denoising methods”, in Proceedings of VPQM, 4 p, 2010.

[21] P. Vu, D. Chandler, "A fast wavelet-based algorithm for global and local image sharpness estimation", IEEE Signal Processing Letters, pp. 423-426, 2012.

[22] N. Ponomarenko, V. Lukin, O. Eremeev, K. Egiazarian, J. Astola, "Sharpness metric for no-reference image visual quality assessment", Image Processing: Algorithms and Systems X and Parallel Processing for Imaging Applications II. International Society for Optics and Photonics, vol. 8295, 11 p, 2012.

[23] H.Z. Nafchi, A. Shahkolaei, R. Hedjam, & M. Cheriet, “Mean deviation similarity index: Efficient and reliable full-reference image quality evaluator”, IEEE Access, 4:5579-5590, 2016.

[24] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, V. Lukin, “On between-coefficient contrast masking of DCT basis functions”, in: Third International Workshop on Video Processing and Quality Metrics, pp. 1-4, 2007.

[25] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, “Image quality assessment: from error visibility to structural similarity”, IEEE Transactions on Image Processing, 13, pp. 600-612, 2004.

[26] M. Ponomarenko, K. Egiazarian, V. Lukin, V. Abramova, "Structural Similarity Index with Predictability of Image Blocks", in Proceedings of International Conference MMET 2018, July 2-5, 4p, 2018.

[27] L. Zhang, L. Zhang, X. Mou, D. Zhang, “FSIM: A Feature Similarity Index for Image Quality Assessment”, IEEE Transactions on Image Processing, 20, pp. 2378-2386, 2011.