Blind DCT-based prediction of image denoising efficiency using neural networks

Oleksii Rubel
Department of Information and Communication Technologies
National Aerospace University "KhAI"
Kharkiv, Ukraine
o.rubel@khai.edu

Andrii Rubel
Department of Information and Communication Technologies
National Aerospace University "KhAI"
Kharkiv, Ukraine
rubel.andrew@gmail.com

Vladimir Lukin
Department of Information and Communication Technologies
National Aerospace University "KhAI"
Kharkiv, Ukraine
lukin@ai.kharkov.com

Karen Egiazarian
Signal Processing Laboratory
Tampere University of Technology
Tampere, Finland
karen.eguiazarian@tut.fi

***Abstract*— Visual quality of digital images acquired by modern mobile cameras is crucial for consumers. Noise is one of the factors that can significantly reduce visual quality of acquired data. There are many image denoising methods able to efficiently suppress noise. However, often in practice denoising does not provide sufficient enhancement of images or even demonstrates visual quality reduction compared to observed noisy data. This paper considers the problem of prediction of denoising efficiency of images in a blind manner under additive white Gaussian noise condition. The proposed technique does not require a priori knowledge of a noise variance and uses a moderate amount of image data for analysis. The denoising efficiency prediction employs neural networks (all-to-all connected multi-layer perceptron) to create a regression model. Image statistics obtained in the spectral domain are used as input data and the state-of-the-art visual quality metrics are considered as outputs of the network. As a target denoising method, block matching and 3D filtering (BM3D) technique is used. It is demonstrated that the obtained neural networks are compact and overall prediction procedure is fast and has an appropriate accuracy to confidently answer to the question: "Do we need to denoise an image?" The full dataset, executable code and demo Android application is available at https://github.com/asrubel/EUVIP2018.**

Keywords—image quality, image denoising, prediction algorithms, multi-layer neural network, mobile applications.

I. INTRODUCTION

A continuous progress in development of mobile devices, especially digital cameras, has offered many opportunities to obtain images or video with high resolution and visual quality [1]. These characteristics usually meet the requirements from consumers regardless which conditions digital images have been captured. However, in practice, visual quality of images can be unsatisfactory due to many reasons. The dominant negative factor is a noise that can influence visual quality dramatically and reduce it significantly [2]. The most popular and commonly accepted noise model in image processing is an additive white Gaussian noise (AWGN) [1, 3]. For cases when a noise presence is obvious in the sense of low visual quality of original data, denoising is usually applied [3].

Currently, there are many denoising methods that are able to restore images corrupted by noise, majority of them are designed to suppress AWGN [3]. To choose a proper method for a practical use, one has to take into account the major requirements to filters that depend upon application. In particular, in mobile imaging, the requirement of computational complexity arises in addition to denoising efficiency. Ideally, a filter applied to a given image should be fast and provide an improvement of visual quality of corrupted images in a degree obvious to a consumer. However, not all denoising techniques are able to satisfy both abovementioned requirements. First, a denoising filter could exhibit insignificant efficiency of noise reduction and, as a result, it can be unable to improve visual quality of a given image. In this situation, denoising occurs to be inefficient (often useless) and it wastes computational resources of a mobile device. Second, denoising could bring additional distortions and a loss of image details due to inadequate image processing. Then, again, it is not reasonable to perform denoising in such a case as well.

Therefore, it could be beneficial to design a fast and accurate predictor to assess a denoising efficiency for a given filter and/or a given image and to make a corresponding decision concerning filter's use for image enhancement. The decision to apply or not a denoising method must be rational and based on some quantitative analysis. In other words, before any filtering applied to an image, its visual quality and noise characteristics have to be assessed [4]. Note, that these two procedures can be rather sophisticated.

Recall that a visual quality assessment can be performed using three types of metrics: full-reference (with a comparison to a corresponding undistorted image – a reference), reduced-reference (using some information or extracted features to help in an assessment) and no-reference [5]. The first type is impossible in the considered case due to the absence of reference image. The second type still exploits some extra data that can vary from an image to image. The third type might have high computational burden and use large amount of exploited image data. Due to these reasons, a denoising efficiency predictor should be performed in a "silent" mode without use of large amount of data and keeping in mind the computational restrictions. Besides, a predictor has to assess filtering efficiency adequately.

Noise intensity, if known in advance or pre-estimated with an appropriate accuracy, can partly clarify the situation. Estimates of noise characteristics (noise variance or standard deviation in AWGN case) [4] can be useful in prediction of filtering efficiency if these estimates are accurate enough. Besides, AWGN standard deviation is a necessary parameter for most denoising methods [3, 4, 6]. However, methods of automatic estimation of noise standard deviation may require large amount of computations (comparable with those of filtering) to detect regions of interest in order to extract noise components. In addition, errors in estimation of noise standard deviation can lead to errors in filtering efficiency prediction [4]; thus, it is better to avoid their estimation in filtering efficiency prediction.

A goal of this paper is to propose an accurate and fast method of denoising efficiency prediction for BM3D filter [6]. BM3D is one of the state-of-the-art methods of AWGN removal according to different visual quality metrics. The novelty of our approach in comparison to the previous studies [7-10] is in the following: a) the proposed method does not use estimation of noise characteristics and performs image processing in a fully blind manner; b) special neural networks (NN) are used for filtering efficiency prediction. Note that predictions for the BM3D filter are close to predictions for some other advanced filters [10].

The paper is structured as follows. Section II describes our previous work devoted to the problem of denoising efficiency prediction. It considers metrics that can be used in visual quality assessment. Section III introduces the proposed method with detailed specification of the NN-based predictor, input parameters, used datasets and training process. Section IV contains the obtained results for denoising efficiency prediction using various metrics, used separately and in groups. The implementation of the proposed method on Android platforms is discussed. Finally, the conclusions are given.

II. Denoising Efficiency Prediction Background

Let us briefly recall the main principles of the denoising efficiency prediction. The following hypotheses are assumed [7-10]:

- there is at least one or many output parameters (quality metrics) able to adequately characterize a denoising efficiency;
- there is at least one or many input parameters that can be calculated easily and quickly for a considered image and that are connected with a quality of observed noisy image and its properties;
- there exists a connection between output and input parameters allowing easy and fast calculation of output parameters using the estimated input parameters; this connection is available to the moment the prediction is carried out since this connection (equation, approximation, neural network) has been obtained in advance (off-line);
- denoising efficiency prediction presumes fast estimation of input parameters and, then, fast calculation of output parameters that can be used for various purposes: analyze quality of original images, study efficiency of denoising, undertake a decision on denoising expedience, etc.

The results of the previous research in the problem of interest are presented in papers [7-10]. In [7], the NN-based predictor of denoising efficiency has been proposed. It exploits block-wised statistical parameters in the Discrete Cosine Transform (DCT) domain to obtain several integral parameters describing the entire noisy image. For each image block transformed by DCT, and hard-thresholding applied to the transform coefficients, a probability of assigning zero values to transformed block coefficients is established. As a threshold parameter, 2σ, has been used, where σ denotes AWGN noise standard deviation (thus, the assumption is that σ has to be a priori known). After such a probability $P_{2\sigma}$ is calculated for all image blocks (usually 1000 blocks covering different image regions), four parameters of the obtained distribution, namely, mean, variance, skewness and kurtosis, are used as inputs to a neural network. As a NN, a feed-forward neural network with a hidden layer, that has 10 neurons, was used. For training, 70% of initial dataset (containing 1280 images) have been used, other 30 % were used for validation.

Note that an improvement of visual quality after filtering can be described by the difference between two values of a visual quality metric – for filtered and noisy images. This difference can indicate expedience to apply a denoising procedure. As the outputs of NN-based predictor, the corresponding improvements of PSNR and PSNR-HVS-M [11] were used. As the denoising techniques, the BM3D and DCT-based filter (which is a simplified version of BM3D) have been used. These two filters deal with images corrupted by noise in a block-wised manner in DCT domain. Unlike DCT-based filter, BM3D employs search of similar blocks and joint collaborative filtering that are computationally costly procedures. The proposed predictor has demonstrated high accuracy for BM3D in terms of R^2 (coefficient of determination): 0.957 for I-PSNR and 0.908 for I-PSNR-HVS-M, respectively (the prefix I denotes an improvement of the corresponding metric value).

The paper [8] considers the problem of texture image denoising. It has been shown that efficiency of the state-of-the-art filters for these cases are close to the potential limits. Strictly speaking, it is often reasonable to avoid denoising textures after their detection. The additional reason for this deals with a low sensitivity of human visual system to noise at high frequencies for small-grain textures. Therefore, it is highly desirable to predict denoising efficiency for images which are fully textured and have complex structure. Later, in the paper [9], these results have been extended to various state-of-the-art filters and metrics. Different from the paper [7], a simple exponential function has been used for prediction to reduce the complexity of the method. Additionally, the thresholding local procedure has been optimized using parameter 0.5σ instead of 2σ. Finally, in the paper [10], multiple input parameters have been exploited to improve the prediction performance.

Let us briefly recall the requirements to a predictor. As it has been stated above, denoising can use computational resources without any sufficient improvement of visual quality (even visual quality drop can be observed). Then, a predictor should be accurate to undertake a correct decision. Computational costs of a prediction method must be significantly lower than those of denoising, i.e. use a minimum amount of input data and have a simplicity of applied operations. The predicting model must be implementable on different hardware and software platforms.

Denoising efficiency can be expressed by various metrics of visual quality. Taking into account the considered types of distortions (noise and distortions due to denoising) it is reasonable to use the most suitable metrics. These metrics are desired to have the highest values of Spearman rank correlation coefficient (SROCC) for databases that have these kinds of distortions, e.g. from TID2013 database. Keeping this in mind, the following full-reference metrics have been exploited for further analysis and training of NN-based predictor. The traditional metric for denoising benchmarking is the peak signal-to-noise ratio (PSNR). The weighted version of this metric, WSNR [12], employs contrast sensitivity functions for visual quality assessment. Information content weighted peak signal-to-noise ratio (IW-PSNR) [13] is another modification of PSNR. For adapting PSNR to human visual system and considering the masking effect at high frequencies and contrast changing, the family of metrics has been proposed, including PSNR-HVS [14], PSNR-HVS-M [11], PSNR-HA [15] and PSNR-HMA [15]. Structure similarity index (SSIM) [16] is another widely used quality metric, based on luminance, contrast and structural similarity. Multi-scale SSIM (MS-SSIM) [17] employs different scales for analysis and provides more reliable results than SSIM. Gradient magnitude similarity deviation (GMSD) [18] relies on the fact that image gradient is sensitive to different degrees of distortions. Recently, the high-accuracy information content weighted SSIM, similarly to PSNR (IW-SSIM) [13], was proposed. New metrics based on new pooling model, ADD-SSIM and ADD-GSIM [19], analyze distortion distribution by image content and distortion. Gradient similarity measure (GSM) [20] is based on the idea that low-level features are important for HVS. Most apparent distortion index (MAD) [21] exploits detection and appearance model to assess visual quality. The advanced algorithm internal generative model (IGM) [22] is strongly connected with neuroscience and modelling perceiving process of visual data. Feature similarity index (FSIM) [23] describes denoising performance in terms of fine details preservation. Haar wavelet-based perceptual similarity index (Haar-PSI) [24] achieves correlation with human opinion scores according to benchmarks. Visual saliency-induced index (VSI) [25] is not PSNR- or SSIM-based metric and it uses another mechanism to assess visual quality of distorted image. Sparse feature fidelity (SFF) [26] is created via transformation of images into sparse representations in the primary visual cortex. Spectral residual based similarity (SR-SIM) [27] is based on a specific spectral residual visual saliency model that relies on the idea that visual saliency map is related to a visual quality. Another metric working in transform domain is RF-SIM [28] – Riesz-transform based feature similarity index that effectively represents low-level features. DCT subbands similarity metric (DSS) [29] measures changes in structural information in DCT domain over subbands. Contrast and visual saliency similarity induced index (CVSSI) [30] utilizes visual saliency attraction in attention of HVS and contrast importance in images perceiving by humans.

III. NN-based Blind Denoising Efficiency Predictor

From metrics listed above and their brief descriptions, it is evident that peculiarities of HVS can play a great role in visual quality assessment for denoising efficiency prediction. The main hypothesis exploited by these metrics is an importance heterogeneity of different level information for HVS. Roughly speaking, the low-level details of an image, i.e. that can be represented mostly at low frequencies, are more important to human eye than high frequency content. Such regions as textures or regular structures in transform domain are placed at high frequencies. Therefore, if we can assess a character of information in image distribution in terms of frequency, the potential "complexity" of denoising can be estimated properly.

Let us consider two extreme scenarios of denoising of different image content. Suppose that denoising is performed in a block-wise manner. That means dealing with image content in some transform domain. For example, the BM3D filter operates in DCT domain with 8x8 pixels blocks and performs search of similar blocks in DCT domain to apply collaborative filtering for them. The first scenario is a block with only low-level content placed at low frequencies. It might have sparse representation of signal in image block that makes the local denoising procedure much easier in both senses of hard thresholding (noise components removal) and similar blocks search. Contrary, if an image block contains high frequency content in DCT domain, noise removal from mixed signal components becomes difficult. For the first scenario, signal allocation promises benefits in filtering. For the second scenario, it is irrational to use denoising conversely. The situations between these two extreme scenarios are worth considering.

We propose the following set of local parameters to describe the structural complexity of signals in image blocks in the sense of frequencies allocation. The transformed image block in DCT domain is divided into 4 stripe areas that contain different coefficients (Fig. 1). All corresponding coefficients are colored and enumerated by an index of this stripe. Thus, the first stripe contains dominantly low frequency content, the fourth consists of only high frequencies. Other stripes contain intermediate frequencies, respectively. Note that after applying a DCT the

0	1	1	1	1	2	2	2
1	1	1	1	2	2	2	3
1	1	1	2	2	2	3	3
1	1	2	2	2	3	3	3
1	2	2	2	3	3	3	4
2	2	2	3	3	3	4	4
2	2	3	3	3	4	4	4
2	3	3	3	4	4	4	4

Fig. 1. Selection of DCT coefficients of image block taking into account energy allocation in the striped manner.

following energy allocation parameter evaluation is performed:

$$W_{k,l\in m,n} = \frac{\sum_{k,l\in m,n} B_{kl}^2}{\sum_{k,l} B_{kl}^2 - B_{00}^2} \quad (1)$$

where B is a transformed block, k and l are indices of block coefficients, m and n are the indices related to a considered stripe, W is a local allocation of the corresponding frequency content. The first coefficient with indices (0,0), corresponding to the mean intensity level of the image block, is not used in analysis.

When the map of these local estimations is obtained for the entire image, the corresponding distributions have to be described in the following way. Similar to [6], we use first four moments such as mean, variance, skewness and kurtosis to describe the shape of distribution. As a result, we have 16 statistical parameters that characterize intently energy allocation in analyzed noisy images. We have used maximum 1000 analyzed image blocks to estimate these parameters.

As a prediction model, we have used feedforward neural network with Bayesian regularization as a training function. As targeted outputs, 24 metrics listed in Section II are used. For training, 300 original grayscale noise-free images from Tampere17 database [31] have been taken and distorted by AWGN with 28 noise levels (σ varied from 3 to 30 with a step 1). We chose this database due to presence of highly textural images. Totally, there are 8400 data samples of metrics' values for filtered and noisy images as well as calculated prediction parameters. The validation is performed in two ways. The first is a self-validation approach where full dataset is permuted and divided into two parts: 80% for training and 20% for validation. The second way is a cross-dataset validation where the initial dataset is exploited fully for training and the images from another dataset are used for validating the trained NNs. As a validation dataset, the mix of reference images from database TID2013 [32] and textural images from USC-SIPI dataset [33] are employed. The last dataset contains only highly textured images with different types of textures. Similarly, to the training dataset, corresponding 28 noise levels are applied for distortions modelling.

IV. Results and Implementation

The exploited architecture of the used NNs for all single-metric predictors is a multi-layer all-to-all connected perceptron with three hidden layers of 16, 8 and 4 neurons respectively. To provide more reliable results, each NN has been trained for each metric 1000 times. The obtained results have been averaged for both self-cross-dataset validations. To characterize the prediction performance of the proposed method, root mean square error (RMSE) and adjusted coefficient of determination R^2 (that describes the goodness of fit) have been used. Recall that adjusted R^2 have to be studied in situations where the number of inputs are high. Note, that R^2 equal to 0.9 and higher means strong connection between two variables or group of variables. In addition, to show the stability of obtaining the prediction results, standard deviations of RMSE and adjusted R^2 have been calculated for each metric predictor. The corresponding data characterizing prediction performance of visual quality metrics are given in Table 1.

One can see from the presented results that all variants of single-metric improvement by NN-based predictor have good performance in terms of adjusted R^2 that is close to 0.9, or higher for self-dataset validation. However, there are two groups of metrics improvement predictors: the first has an appropriate performance – adjusted R^2 is close to 0.9 (I-SSIM, I-ADD-SSIM, I-ADD-GSIM, I-MAD, I-IGM, I-FSIM, I-VSI, I-SFF, I-SR-SIM, I-RFSIM, I-DSS, I-CVSSI) and the second has high performance expressed by R^2 that is higher than 0.92 (I-PSNR, I-WSNR, I-IW-PSNR, I-PSNR-HVS, I-PSNR-HVS-M, I-PSNR-HA, I-PSNR-HMA, I-MS-SSIM, I-GMSD, I-GSM, I-Haar-PSI). It should be noted that prediction performance in terms of RMSE are also appropriate or high for SNR-based metrics (I-PSNR, I-WSNR, I-IW-PSNR, I-PSNR-HVS, I-PSNR-HVS-M, I-PSNR-HA and I-PSNR-HMA) – they do not exceed 1 dB and are even close to 0.5 dB except I-PSNR. Similar remark can be made for most of SSIM-based metrics (I-SSIM, I-MS-SSIM, I-GMSD, I-IW-SSIM, I-ADD-SSIM, I-ADD-GSIM, I-FSIM and I-SR-SIM) where RMSE values are about 0.01 (the range of possible values for these metrics is from 0 to 1). Standard deviations of RMSE and adjusted R^2 values for 1000 realizations of predictors training are also given to demonstrate reliability of the proposed predictors.

The results of prediction performance using cross-dataset validation shows that for all predictors the performance decreases when training dataset and validating datasets consist from different images. Besides, this seems to be close to real-life situations. Summarizing the obtained results from the right part of Table I, one can conclude the following. The NN-based predictors for SNR-based metrics improvement as I-PSNR, I-WSNR, I-PSNR-HVS, I-PSNR-HVS-M, I-PSNR-HA and I-PSNR-HMA can perform well under AWGN conditions for various images. Moreover, such SSIM-based metrics as I-MS-SSIM, I-GMSD, I-IW-SSIM, I-ADD-SSIM including also I-GSM and I-Haar-PSI demonstrate good performance as well. All mentioned metrics' predictors (marked bold in Table I) have appropriate accuracy: the adjusted R^2 is close to 0.9 and higher. This gives us a confidence that prediction using these metrics can be made in real-life applications.

Note that such SNR- or SSIM-based metrics used in predictors are related or can be correlated between each other. After analyzing SROCC values between sets of metrics improvements values, it has been established that there are two groups of metrics for which multi-target prediction is possible. Therefore, these two groups (I-PSNR, I-WSNR, I-PSNR-HVS, I-PSNR-HVS-M, I-PSNR-HA and I-PSNR-HMA) and (I-MS-SSIM, I-GMSD, I-IW-SSIM, I-ADD-SSIM, I-GSM and I-Haar-PSI) are used as outputs for multi-metrics' improvement predictors that could be useful [34]. For such predictors, NN architectures similar to single-metrics cases are exploited. These NNs also have three hidden layers with 16, 2 * (number of metrics) and (number of metrics) neurons. The prediction results using cross-dataset are shown in Table II. Prediction accuracy is still good and comparable to single-metrics versions. Recall that the use of two or more metric in decision undertaking can be expedient.

TABLE I. PREDICTION PERFORMANCE OF VISUAL QUALITY METRICS IMPROVEMENT BY DENOISING

Metric Improvement	Self-Dataset Validation				Cross-Dataset Validation			
	RMSE	*σ_{RMSE}*	*Adjusted R^2*	*σ_{R^2}*	*RMSE*	*σ_{RMSE}*	*Adjusted R^2*	*σ_{R^2}*
I-PSNR	0.6625	0.0421	0.9644	0.0053	**0.8775**	**0.0380**	**0.9229**	**0.0069**
I-WSNR [12]	0.4072	0.0813	0.9477	0.0301	**0.3785**	**0.0509**	**0.9221**	**0.0252**
I-IW-PSNR [13]	0.3884	0.0409	0.9209	0.0212	0.3703	0.0270	0.8706	0.0207
I-PSNR-HVS [14]	0.5241	0.0575	0.9633	0.0103	**0.5905**	**0.0362**	**0.9407**	**0.0078**
I-PSNR-HVS-M [11]	0.5458	0.0760	0.9604	0.0156	**0.6142**	**0.0550**	**0.9366**	**0.0148**
I-PSNR-HA [15]	0.5333	0.0523	0.9619	0.0102	**0.6060**	**0.0400**	**0.9358**	**0.0093**
I-PSNR-HMA [15]	0.5437	0.0609	0.9610	0.0115	**0.6128**	**0.0424**	**0.9352**	**0.0097**
I-SSIM [16]	0.0291	0.0016	0.8946	0.0129	0.0390	0.0022	0.7264	0.0328
I-MS-SSIM [17]	0.0144	0.0021	0.9606	0.0174	**0.0145**	**3.8905e-04**	**0.9597**	**0.0022**
I-GMSD [18]	0.0129	8.1860e-04	0.9285	0.0103	**0.0133**	**5.0005e-04**	**0.9100**	**0.0070**
I-IW-SSIM [13]	0.0132	0.0028	0.9564	0.0286	**0.0136**	**4.8868e-04**	**0.9489**	**0.0037**
I-ADD-SSIM [19]	0.0013	6.7543e-05	0.8946	0.0117	**0.0012**	**2.6626e-05**	**0.8997**	**0.0044**
I-ADD-GSIM [19]	0.0013	8.1163e-05	0.8900	0.0158	0.0012	4.5547e-05	0.8659	0.0103
I-GSM [20]	0.0021	1.8122e-04	0.9284	0.0153	**0.0022**	**6.3648e-05**	**0.9095**	**0.0053**
I-MAD [21]	4.5527	0.2943	0.8900	0.0174	5.0494	0.2487	0.8577	0.0150
I-IGM [22]	0.0097	5.1904e-04	0.9030	0.0113	0.0104	4.5313e-04	0.8471	0.0142
I-FSIM [23]	0.0176	0.0016	0.9163	0.0189	0.0197	6.1032e-04	0.8604	0.0086
I-Haar-PSI [24]	0.0290	0.0021	0.9389	0.0096	**0.0332**	**0.0016**	**0.8999**	**0.0101**
I-VSI [25]	0.0075	3.6615e-04	0.8912	0.0106	0.0091	2.0429e-04	0.8124	0.0086
I-SFF [26]	0.0039	2.7051e-04	0.9024	0.0151	0.0048	4.0025e-04	0.7724	0.0462
I-SR-SIM [27]	0.0103	7.3635e-04	0.8805	0.0191	0.0108	3.7720e-04	0.8395	0.0112
I-RFSIM [28]	0.0416	0.0025	0.8804	0.0148	0.0447	0.0043	0.8203	0.0377
I-DSS [29]	0.0513	0.0030	0.8926	0.0147	0.0485	0.0018	0.8875	0.0083
I-CVSSI [30]	0.0132	6.9545e-04	0.9065	0.0102	0.0131	9.4356e-04	0.8750	0.0199

TABLE II. JOINT PREDICTION PERFORMANCE OF VISUAL QUALITY METRICS IMPROVEMENT

Joint prediction for improvement of PSNR, WSNR, PSNR-HVS, PSNR-HVS-M, PSNR-HA, PSNR-HMA			Joint prediction for improvement of MS-SSIM, GMSD, IW-SSIM, GSM, Haar-PSI		
Metric	*RMSE*	*Adjusted R^2*	*Metric*	*RMSE*	*Adjusted R^2*
I-PSNR	0.8971	0.9194	I-MS-SSIM [17]	0.0146	0.9586
I-WSNR [12]	0.3997	0.9140	I-GMSD [18]	0.0141	0.8982
I-PSNR-HVS [14]	0.5993	0.9388	I-IW-SSIM [19]	0.0131	0.9529
I-PSNR-HVS-M [11]	0.6241	0.9347	I-GSM [20]	0.0328	0.9020
I-PSNR-HA [15]	0.6047	0.9360	I-Haar-PSI [24]	0.0024	0.8885
I-PSNR-HMA [15]	0.6128	0.9353			

The proposed method is implemented both in MATLAB and Python using TensorFlow framework. Demo application is made on Android to demonstrate the real work of proposed method. It uses neural networks API available since Android 8.1 (API level 27) or higher to run the trained NNs on mobile devices. Evaluation of input prediction parameters is implemented using C++ and runs on Android Native Development Kit (NDK). The models of trained neural networks are saved and ported to Android platform. Please note that for correct application a device must fully support the Camera2 API (full or level_3 support categories). All listed resources and supplementary materials are available at https://github.com/asrubel/EUVIP2018.

V. CONCLUSIONS

Noise existing in a digital image significantly decreases its visual quality. In practice, there are cases when denoising does not provide sufficient efficiency. Handling these cases is crucial for mobile imaging in the sense of computational costs. This paper presents the NN-based predictor of denoising efficiency characterized by improvements of visual quality metrics. The proposed predictors perform fully blindly without any necessity of a priori knowledge of AWGN variance. They use simple input statistics and compact neural networks with accuracy appropriate for real-life applications. In addition, the multi-metric predictors are designed using single NN that is able to forecast simultaneously various metrics' improvements. The obtained results demonstrate that it is possible to give a strong rationale for applying or skipping a denoising step. Finally, the implementation of NN-based predictors on Android platform is provided.

REFERENCES

[1] I. W. Pratt, Digital image processing. Fourth Edition. N.Y, USA.: Wiley-Interscience, 2007, 812 p.

[2] Z. Wang and A. C. Bovik, Modern Image Quality Assessment. San Mateo, CA, USA: Morgan Claypool, 2006.

[3] A. Pižurica, "Image Denoising Algorithms: From Wavelet Shrinkage to Nonlocal Collaborative Filtering," Wiley Encyclopedia of Electrical and Electronics Engineering, pp. 1-17, 2017.

[4] V. Abramova, V. Lukin, S. Abramov, O. Rubel, B. Vozel, K. Chehdi, J. Astola, and K. Egiazarian, "On requirements to accuracy of noise estimation in prediction of DCT-based filter efficiency," Telecommunications and Radio Engineering, vol. 75, no. 2, pp. 139-154, 2016.

[5] Z. Wang, "Applications of Objective Image Quality Assessment Methods [Applications Corner]," in IEEE Signal Processing Magazine, vol. 28, no. 6, pp. 137-142, Nov. 2011.

[6] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," IEEE Trans. on Image Processing, vol. 16, no. 8, pp. 2080-2095, 2007.

[7] A. Rubel, A. Naumenko, and V. Lukin, "A neural network based predictor of filtering efficiency for image enhancement," in Proc. Microwaves Radar and Remote Sensing Symposium (MRRS), Kiev, Ukraine, 2014, pp. 14-17.

[8] A. Rubel, V. Lukin, M. Uss, B. Vozel, O. Pogrebnyak, and K. Egiazarian, "Efficiency of texture image enhancement by DCT-based filtering," Neurocomputing, vol. 175, pp. 948-965, 2016.

[9] O. Rubel, V. Lukin, S. Abramov, B. Vozel, K. Egiazarian, and O. Pogrebnyak, "Efficiency of texture image filtering and its prediction," Signal, Image and Video Processing, vol. 10, no. 8, pp. 1543-1550, 2016.

[10] O. Rubel, V. Lukin, S. Abramov, B. Vozel, O. Pogrebnyak, and K. Egiazarian, "Is Texture Denoising Efficiency Predictable?," International Journal of Pattern Recognition and Artificial Intelligence, vol. 32, no. 01, p. 32, 2018.

[11] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin, "On between-coefficient contrast masking of DCT basis functions," in Proc. 3rd Int. Workshop Video Process. Qual. Metrics Consum. Electron, Scottsdale, USA, Jan. 2007, 4 p.

[12] T. Mitsa and K. Varkur, "Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms," in Proc. IEEE International Conference on Acoustics Speech and Signal Processing, vol. 5, Apr. 1993, pp. 301-304.

[13] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," IEEE Trans. Image Process., vol. 20, no. 5, pp. 1185-1198, May 2011.

[14] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli, "New full-reference quality metrics based on HVS," in Proc. Second International Workshop on Video Processing and Quality Metrics, Scottsdale, USA, Jan. 2006, 4 p.

[15] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, and M. Carli, "Modified image visual quality metrics for contrast change and mean shift accounting," in Proc. 11th Int. Conference The Experience of Designin and Application of CAD Systems in Microelectronics, Ukraine, 2011, pp. 305-311.

[16] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process., vol. 13, no. 4, pp. 600-612, Apr. 2004.

[17] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structucal similarity for image quality assessment," in Proc. IEEE 37th Asilomar Conf. Signals, Syst., Comput., Pacific Grove, CA, Nov. 2003, pp. 1398-1402.

[18] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," IEEE Trans. Image Process., vol. 23, no. 2, pp. 684–695, Feb. 2014.

[19] K. Gu, S. Wang, G. Zhai, W. Lin, X. Yang, and W. Zhang, "Analysis of Distortion Distribution for Pooling in Image Quality Prediction," IEEE Transactions on Broadcasting, vol. 62, no. 2, pp. 446-456, 2016.

[20] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," IEEE Trans. Image Process., vol. 21, no. 4, pp. 1500–1512, Apr. 2012.

[21] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," J. Electron. Imagining, vol. 19, no. 1, pp. 011006:1-011006:21, 2010.

[22] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism," IEEE Trans. Image Process., vol. 22, no. 1, pp. 43-54, Jan. 2013.

[23] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," IEEE Trans. Image Process., vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[24] R. Reisenhofer, S. Bosse, G. Kutyniok, and T. Wiegand, "A Haar Wavelet-Based Perceptual Similarity Index for Image Quality Assessment," Signal Processing: Image Communication, vol. 61, pp. 33-43, Feb. 2018.

[25] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment," IEEE Transactions on Image Processing, vol. 23, no. 10, pp. 4270-4281, Oct. 2014.

[26] H-W. Chang, H. Yang, Y. Gan, and M-H. Wang, "Sparse feature fidelity for perceptual image quality assessment," IEEE Trans. on Image Processing, vol. 22, no. 10, pp. 4007-4018, Oct. 2013.

[27] L. Zhang and H. Li, "SR-SIM: A fast and high performance IQA index based on spectral residual," in Proc. IEEE Int. Conf. Image Process., Orlando, FL, USA, Sep. 2012, pp. 1473-1476.

[28] L. Zhang, L. Zhang, and X. Mou, "RFSIM: a feature based image quality assessment metric using Riesz transforms," in Proc. IEEE Int. Conf. Image Process., 2010, pp. 321-324.

[29] A. Balanov, A. Schwartz, Y. Moshe, and N. Peleg, "Image quality assessment based on DCT subband similarity," in Proc. IEEE Int. Conf. Image Process. (ICIP), 2015, pp. 2105-2109.

[30] H. Jia and T. Wang, "Contrast and visual saliency similarity induced index for image quality assessment," arXiv preprint arXiv:1708.06616, 2017.

[31] M. Ponomarenko, N.Gapon, V.Voronin, and K. Egiazarian, "Blind estimation of white Gaussian noise variance in highly textured images," mage Processing: Algorithms and Systems XVI, 2018, 5 p.

[32] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. Jay Kuo, "Image database TID2013: Peculiarities, results and perspectives," Signal Processing: Image Communication, vol. 30, Jan. 2015, pp. 57-77.

[33] USC-SIPI Image database. [Online]. Available: (http://sipi.usc.edu/database/database.php?volume=textures).

[34] F. Boudjenouia, K. Abed-Meraim, A. Chetouani, R. Jennane, "On the use of image quality measures for image restoration," Image Processing Theory Tools and Applications (IPTA), 6th International Conference on. IEEE, 2016, pp. 1-6.