Learning-based Noise Component Map Estimation for Image Denoising

Sheyda Ghanbaralizadeh Bahnemiri, Mykola Ponomarenko,
and Karen Egiazarian, *Fellow, IEEE*

***Abstract*—A problem of image denoising, when images are corrupted by a non-stationary noise, is considered in this paper. Since, in practice, no a priori information on noise is available, noise statistics should be pre-estimated prior to image denoising. In this paper, deep convolutional neural network (CNN) based method for estimation of a map of local, patch-wise, standard deviations of noise (so-called *sigma-map*) is proposed. It achieves the state-of-the-art performance in accuracy of estimation of sigma-map for the case of non-stationary noise, as well as estimation of a noise variance for the case of an additive white Gaussian noise. Extensive experiments on image denoising using estimated sigma-maps demonstrate that our method outperforms recent CNN-based blind image denoising methods by up to 6 dB in PSNR, as well as other state-of-the-art methods based on sigma-map estimation by up to 0.5 dB, providing, at the same time, better usage flexibility. A comparison with the ideal case, when denoising is applied using ground-truth sigma-map, shows that a difference of corresponding PSNR values for the most of noise levels is within 0.1-0.2 dB, and does not exceed 0.6 dB.**

***Index Terms*—Image denoising, non i.i.d. noise, blind noise parameters estimation, deep convolutional neural networks**

I. Introduction

Image denoising is one of the most studied problems of image processing. Acquired images are often exposed to noise due to low light intensity, camera quality, transmission errors, etc. [1]. During the last decades, a large number of denoising methods have been proposed, such as methods based on local transform domain (e.g. sliding Discrete Cosine Transform (DCT) filtering) [2]–[4] and wavelet-based [5]–[7]) methods, nonlocal collaborative denoisers (e.g. BM3D [8]) [9], [10], and learning (e.g. Convolutional Neural Networks (CNN)) based methods [11]–[17]. Most of these denoisers assume that a noise is Additive White Gaussian Noise (AWGN), and its standard deviation σ is either known or pre-estimated (σ is a constant for the whole image) [18]–[21].

In this paper, we assume that the observed image is corrupted by a noise with non-i.i.d. pixel-wise Gaussian distribution (similar to the noise model considered in [22]):

$$y_{i,j} \sim \mathcal{N}(y_{i,j}|x_{i,j}, \sigma_{i,j}^2), i = 1, 2, ..., L, j = 1, 2, ..., K, \quad (1)$$

where $\mathcal{N}(\cdot|\mu, \sigma^2)$, here and later in this paper, denotes the Gaussian distribution with mean μ and variance σ^2; $x_{i,j}$ and $y_{i,j}$ are $(i, j)^{th}$ pixel values of unknown clean image $\mathbf{X}$ and noisy image $\mathbf{Y}$, respectively; $L \times K$ is image size, and a matrix $\mathbf{\Sigma} = [\sigma_{i,j}]$ of size $L \times K$ defines the so-called *sigma-map* [22]. Note, that in the case of i.i.d. AWGN, sigma-map becomes a matrix with all elements having value σ.

Early methods of sigma-map estimation perform robust analysis of transform coefficients of image patches [23], [24]. However, these methods may produce large estimation errors for fine details, edges, and texture; thus, they will not work well when input image is corrupted by a small level of noise.

During the last decade, CNN becomes a widespread tool in image processing and analysis. One popular trend in image denoising is to develop fully blind methods which do not require any input noise parameters [12], [14]. However, these methods are unable to provide a quality of denoising comparable to those of the state-of-the-art methods, which use sigma-map as an auxiliary input. The recent state-of-the-art method for joint sigma-map estimation and noise suppression, called VDNet, is proposed in [22].

The main goal of the paper is to estimate a sigma-map from a given noisy image, with a further use of it as an auxiliary input in image denoising methods. A special attention is given to the case when a noise level is small. It is the most appealing case from the practical point of view, as well as a very difficult case for existing sigma-map estimation methods, as it will be demonstrated later.

In this paper, we propose a novel method of sigma-map estimation, based on the SDNet network, designed by combining U-Net [25] and ResNet [26] architectures, similar to the DRUNet network used for denoising and super-resolution [27], but taking only noisy images as an input. SDNet is trained to predict only a noise level map, unlike VDNet, which also produces a denoised image. Our previously proposed deep convolutional autoencoder-based method, DCAE [28], has been designed to preserve an informative image component and trained on noise-free patches. In contrary, SDNet is trained on noisy patches and directly predicts a sigma-map.

We also propose an innovative strategy to prepare and collect images for the training set, and an improved method of generation of patches in the custom training loop. SDNet achieves a superior accuracy of sigma-map estimation, outperforming state-of-the art. A combination of SDNet with image denoisers, which take a predicted sigma-map as an auxiliary input, achieves superior denoising capability, significantly outperforming state-of-the-art blind image denoising methods.

A structure of the paper is as follows. The proposed network and the training process are described in Section II. Sections III and IV analyse the results of sigma-map estimation and its use in image denoising. Conclusions are given in Section 5.

This manuscript was submitted on September 20, 2021 and resubmitted on February 14, 2022. Authors are with Tampere University, Finland, (e-mail: sheyda.ghanbaralizadehbahnemiri@tuni.fi; mykola.ponomarenko@tuni.fi; karen.eguiazarian@tuni.fi).

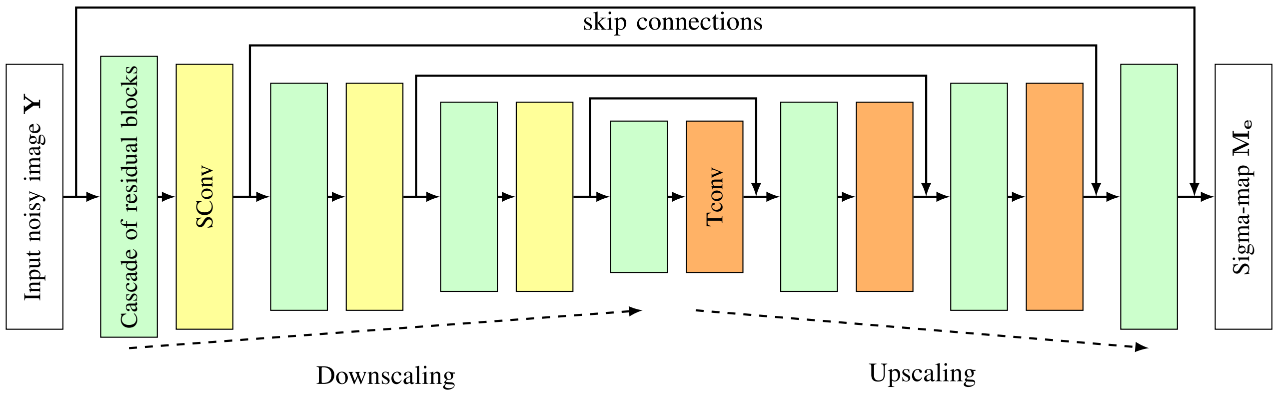


Fig. 1: Structural scheme of proposed SDNet (SConv - stride convolution, TConv - transpose convolution)

II. Proposed network for blind sigma-map estimation

A. Network Design

A structural scheme of the proposed neural network, SDNet, is presented in Fig.1. Inspired by the DRUNet [27] network architecture, SDNet combines U-Net [25] and ResNet [26] in a single design. SDNet suits well to our task of sigma-map estimation since analysis is performed at several image scales. In the middle part of SDNet, each pixel of activations corresponds to 8x8 pixels of the input image. This part of SDNet provides good estimation accuracy for large homogeneous areas. At the same time, a pixel-wise analysis in the last quarter of SDNet provides good robustness to fine details and textures.

A detailed structure of SDNet, demo scripts, pre-trained models, real-life and full-size illustrations of sigma map estimation and image denoising are available in https://github.com/SheydaGhb/SDNet.

B. Image Set for Training

Images used for neural network training are often distorted by noise and compression artifacts, which decreases efficiency of training and introduces a bias in noise level estimation by a pre-trained network. Our strategy is different: we use only near distortions-free images. In the combination with other important details of custom training loop proposed in this paper, this strategy significantly increases a quality of training.

To train SDNet, we have generated and accurately selected 4238 images from different datasets as it is described below.

We designed a new Tampere21 image database [29], which contains 1000 near noise-free color images, obtained by Canon EOS 250D camera with controlled shooting conditions using a methodology described in [21].

The second set of 1000 images is selected from four datasets: KonIQ10k [30], FLIVE [31], NRTID [32] and SPAQ [33]. The merged mean opinion scores (MOS) [34] of these four databases has been used to select high quality images having the best MOS values.

The remaining images are collected from the following sources: 217 images from Flickr2K database [35], 123 images from Waterloo Exploration Database [36], 103 images from DIV2k database [37], and 1795 images from different photo hostings. We have used metric KonCept512 [30] pre-trained on six databases (KonIQ10k [30], Live-in-the-Wild [38], FLIVE [31], NRTID [32], HTID [39] and SPAQ [33]) to collect high quality images from the above mentioned image sources.

C. Custom Training Loop

SDNet is trained to a predict a sigma map on noisy patches which are generated from a clean image by adding a noise to it with a given ground truth sigma map (M_t) according to formula (1). Mean square error between predicted and ground truth sigma maps is used to correct weights of SDNet during the training.In this paper, a relative error ε_m is used as a qualitative criterion of estimated sigma-map:

$$\varepsilon_m = \frac{||\boldsymbol{M_e} - \boldsymbol{M_t}||_2}{n||\boldsymbol{M_t}||_2} \quad (2)$$

where $\boldsymbol{M_e}$ is the estimated maps, n is a number of images in the test set. For testing purposes, three non-stationary sigma-map models introduced in [22] are selected as ground truth map to create a noisy test set.

For efficient image denoising a relative error ε_m has to be smaller than 0.1 [40]. This condition cannot be guaranteed if one will use the conventional custom loss functions, since the pre-trained network may produce too large relative errors for small noise standard deviations. To overcome this, we set the mean variance of sigma-map, σ^2_{av} for generated patches to follow a half-normal distribution:

$$\sigma^2_{av} = |\mathcal{N}(0, R^2)|, \quad (3)$$

where we set $R = 40$ for the training.

This provides a larger number of small σ_{av} values, and, therefore, their larger weights in the training. As a result, SDNet will better minimize a prediction error for small σ_{av} values, while providing a comparable ε_m for large σ_{av}. At the same time, the pre-trained SDNet should give good prediction results also for very large σ_{av} values (over 100).

Authors of [22] have used smooth surfaces to generate $\boldsymbol{M_t}$ values. However, in practice, this may limit an applicability of the pre-trained network. In this paper, we have generated the $\boldsymbol{M_t}$ for an input patch by:

$$\sigma_{i,j} = (\sigma^2_{av} B_{i,j}/\bar{B})^{0.5}, i = 1, 2, ..., L, j = 1, 2, ..., K; \quad (4)$$

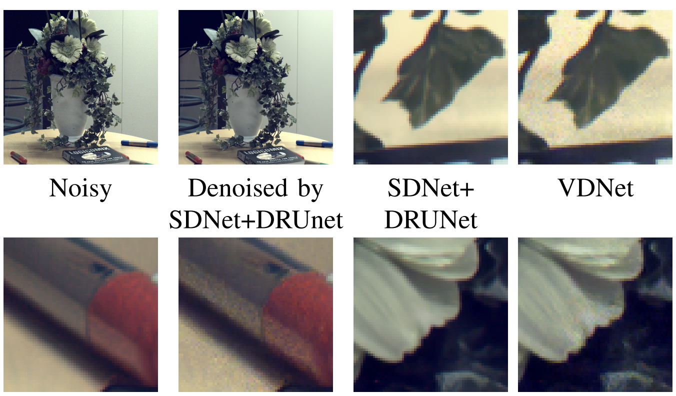
where $\boldsymbol{B} = [B_{i,j}]$ is a brightness of the fragment of a randomly selected image from the training set, $\bar{B}$ is the mean of $\boldsymbol{B}$. The corresponding input patch $\boldsymbol{P}$ is generated for training by:

$$P_{i,j} = \mathcal{N}(I_{i,j}, \sigma_{i,j}^2), i = 1, 2, ..., L, j = 1, 2, ..., K; \quad (5)$$

where $\boldsymbol{I} = [I_{i,j}]$ is an image fragment of size 128x128.

Ground-truth map, σ_{av}=5 | Noisy image | SDNet ε_m=0.12 | VDNet ε_m=0.27

Fig. 2: Examples of sigma-map estimation from a noisy image by VDNet and color SDNet. Ground truth sigma-map is generated following a setup in [22]



Noisy | Denoised by SDNet+DRUnet | SDNet+ DRUNet | VDNet

SDNet+ DRUNet | VDNet | SDNet+ DRUNet | VDNet

Fig. 3: Visual comparison of denoising by DRUNet+SDNet and VDNet for a real life image

Image fragments $\boldsymbol{I}$ for generation of patches are selected in the following way. First, a random image from the training set is extracted. Next, a fragment of size 128x128 is cropped from this image to increase the presence of patches with fine details and textures in the training set. This is done by the algorithm described in [41] (selecting three image fragments). To decrease over-learning, random rotations (on 90°, 180°, 270°) and mirroring of selected fragments are used. The half of input patches $\boldsymbol{P}$ are clipped by restricting pixel values to be in the [0...255] interval. This is done to make a pre-trained SDNet effective for both clipped and non-clipped noise cases.

The design and training of SDNet has been carried out in Matlab R2021a, using the custom training loop. Overall, 150000 iterations with minibatch size 32 were performed. MSE has been used as a loss function to provide smaller number of outliers in sigma-map estimates. Adam optimizer was used with the learning rate for the first 100000 iterations and for the last 50000 iterations.

D. Modification to color images

SDNet has been separately trained for color and grayscale images. To provide a compatibility with the setup used in [22], for color images, noise was independently added to each image color channel according to (5) using the same sigma-map.

III. Numerical Analysis:Sigma-map Estimation

In this section, we analyze a quality of sigma-maps estimation by the pre-trained SDNet. A set of 30 color images (28 images from Tampere17 dataset [21] in addition to Barbara and Baboon images) are selected for testing due to their rich content, fine details and texture, which make them challenging for denoising task. Noise is added to these images according to (1), using three ground truth sigma-maps introduced in [22]. We have multiplied these sigma-maps by a factor to provide a specified mean variance σ_{av}^2 of a sigma-map. In total, 630 noisy images (30 x 3 x 7 noise levels) are included in the test set. We have used 5-folds cross-validation for the training-testing both for color and for grayscale versions of SDNet. Tables I, II and III contain averaged values over these 5 test sets.

A. Quality of estimation of sigma-map for non-stationary noise

We have used ε_m defined in (2) as a qualitative criterion ($n = 72$). The relative errors are demonstrated in Table II, to compare quality of sigma-map estimation by the proposed method SDNet, methods RHDCT [24], LADCT [23], DCAE [28], and the state-of-the-art VDNet [22]. Sigma-maps for grayscale SDNet are separately estimated for image color channels and then averaged.

SDNet outperforms all methods providing 2..5 times smaller ε_m than VDNet. Fig 2 shows an example of ground-truth map and sigma-maps estimated by VDNet [22] and color SDNet.

B. Quality of estimation of standard deviation of AWGN

The proposed SDNet can be applied for a special case when all $\sigma_{i,j} = \sigma$, which corresponds to the case of AWGN. For a comparative amalysis we have selected three methods: IEDD [21], PCA [20] and WTP [42], which are state-of-the-art for estimation of σ of AWGN. We have also included in this analysis VDNet [22]. Estimated σ for a given image by SDNet and VDNet are calculated as medians of estimated sigma-maps. As a quality criterion we used a relative error of estimation of STD, defined by:

$$\varepsilon = \frac{||\boldsymbol{\sigma_e} - \sigma_t||_2}{n\sigma_t} \quad (6)$$

where σ_t is a true value of AWGN STD, n is a number of estimates (here $n = 24$), $\boldsymbol{\sigma_e}$ is a vector of estimated standard deviations for test images. Both clipped and non-clipped noise are considered in this analysis (see Table III).

SDNet attains the best accuracy, outperforming other methods almost for all cases. Moreover, for σ_t equal to 3 and 5, SDNet estimation error is twice smaller than that of the nearest competitor. It is interesting to observe that all four methods IEDD, PCA, WTP and VDNet fail for large σ_t for

TABLE I: Image denoising in the case of non-stationary noise, PSNR[dB] and SSIM

σ_{av}	Noisy images		CBDNet (blind)		DnCNN (blind)		VDNet		FFDNet+ color SDNet		FFDNet + true sigma-map		DRUNet+ VDNet		DRUNet+ grayscale SDNet		DRUNet+ color SDNet		DRUNet+ true sigma map	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
5	34.2	0.953	31.7	0.950	34.1	0.973	37.0	0.981	36.5	0.980	36.8	0.981	37.1	0.981	**37.6**	0.983	**37.6**	**0.983**	38.1	0.984
7	31.3	0.919	31.5	0.947	33.3	0.966	35.3	0.973	35.0	0.971	35.2	0.973	35.5	0.974	35.7	0.974	**35.8**	**0.975**	36.1	0.977
10	28.3	0.864	31.0	0.938	32.1	0.954	33.5	0.961	33.2	0.959	33.3	0.959	33.7	0.962	33.7	0.962	**33.9**	**0.963**	34.1	0.964
15	24.8	0.774	29.8	0.917	30.5	0.933	31.4	0.941	31.1	0.936	31.2	0.936	31.6	0.940	31.6	0.940	**31.7**	**0.942**	31.8	0.943
20	22.4	0.693	28.6	0.893	29.3	0.912	29.9	**0.920**	29.6	0.912	29.7	0.912	30.0	0.918	**30.1**	0.918	**30.1**	**0.920**	30.2	0.920
30	19.1	0.562	26.8	0.847	27.4	0.870	27.8	**0.880**	27.5	0.863	27.5	0.864	**27.9**	0.876	27.8	0.874	**27.9**	0.876	27.9	0.876
45	16.0	0.424	25.0	0.783	25.3	0.813	25.6	**0.826**	25.1	0.793	25.1	0.794	**25.7**	0.820	25.4	0.809	25.5	0.812	25.5	0.813

TABLE II: Relative estimation error ε_m of sigma-maps

σ_{av}	LADCT	RHDCT	DCAE	VDNet	SDNet grayscale	SDNet color
5	1.87	1.69	0.41	0.31	0.26	**0.19**
7	1.24	1.11	0.32	0.21	0.18	**0.12**
10	0.80	0.69	0.28	0.15	0.13	**0.07**
15	0.49	0.40	0.22	0.11	0.09	**0.04**
20	0.35	0.27	0.21	0.10	0.06	**0.03**
30	0.24	0.19	0.25	0.11	0.04	**0.02**
45	0.20	0.19	0.31	0.13	0.03	**0.02**

clipped noise case (values are marked in Table III in italic and underlined), while SDNet demonstrates very good accuracy both for clipped and non-clipped noise cases.

IV. Numerical Analysis: Denoising Using Estimated Sigma-map

Here we use estimated sigma-maps in image denoising. Comparing denoised images obtained by the different sigma-map estimators, we show how much one can boost the performance of denoising by improving accuracy of sigma-map estimation. As denoisers in this study we use DRUNet and FFDNet [13], since their network architecture accepts noise sigma-map as an auxiliary input. Three state-of-the-art CNN-based blind denoising methods, DnCNN [12], CBDNet [14] and VDNet [22], have been selected for the comparison. The peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) are used as quality criteria.

Note that DnCNN was originally designed to address the problem of Gaussian denoising with unknown noise level, while CBDNet is taking one step further and includes real-world noisy-clean image pairs as training input.

Table I shows average PSNR and SSIM values for 72 denoised test images. Furthermore, for denoisers equipped with the sigma-map estimation by the proposed SDNet, we provide also the results for ideal case of the denoising when the true sigma-map are used. Small differences between PSNR and SSIM values of the results produced by the denoisers accepting true sigma-map with the same denoisers accepting SDNet estimated sigma-map, demonstrate the high accuracy of sigma-map estimation and its influence on denoising performance. As one can see from Table I, a combination of SDNet with DRUNet yields the best denoising performance, and the gap between best results and the results from ideal case (using true sigma-maps) is less than 0.6 db in PSNR and none in SSIM.

Although the same comparison is performed for the case of FFDNet denoiser and the gap is smaller, quality of denoising using FFDNet is about 1 db lower than that of DRUNet. In the case of blind denoisers, quality of denoised images is much lower, within the range of $2 \sim 6$ dB for low noise levels.

SDNet outperforms VDNet since its accuracy in estimating sigma map is higher than that of VDNet as well as other state-of-the-art methods. Thus, it directly affects a denoising quality. It is also interesting that for small noise levels DRUNet with sigma map estimated by VDNet provides better results than VDNet denoiser itself. Image denoising based on the proposed sigma-map estimation in the combination with DRUNet denoiser significantly outperforms the blind denoisers. While blind denoisers can distort texture and fine details, the denoising methods based on a sigma-map estimated by the proposed SDNet overcome this problem and preserve fine image details especially in the case of low noise levels.

Fig 3 shows an example of a real life image denoising. Enlarged fragments of the denoised images show that the proposed sigma map estimation by SDNet, combined with the DRUNet, produces better visual quality images than those by VDNet [22].

TABLE III: ε for non-clipped/clipped AWGN

σ	IEDD	PCA	WTP	VDNet	SDNet
3	0.407/0.407	0.573/0.573	0.508/0.507	1.783/1.782	**0.144/0.150**
5	0.204/0.204	0.272/0.274	0.271/0.269	0.906/0.905	**0.090/0.094**
7	0.110/0.114	0.167/0.166	0.176/0.174	0.560/0.559	**0.058/0.062**
10	0.054/0.073	0.098/0.097	0.126/0.122	0.325/0.323	**0.031/0.034**
15	0.033/0.071	0.057/0.053	0.079/0.075	0.171/0.168	**0.014/0.013**
20	0.021/*0.077*	0.044/0.038	0.057/0.057	0.110/0.105	**0.017/0.016**
30	0.014/*0.114*	0.031/0.036	0.029/0.055	0.063/0.066	**0.011/0.011**
50	0.013/*0.157*	0.022/*0.069*	0.012/*0.103*	0.040/*0.089*	**0.009/0.009**
75	0.013/*0.207*	0.023/*0.172*	**0.007**/*0.176*	0.032/*0.150*	**0.007/0.007**

V. Conclusion

We have proposed an effective CNN-based method, SDNet, for sigma-map estimation. The network architecture, training set generation and selection of training patches were described. A comparison of SDNet with the state-of-the-art methods shows that SDNet outperforms them in estimation accuracy of sigma-maps, for both clipped and non-clipped noise cases. For small levels of noise estimation errors of SDNet in average are twice smaller than those of nearest methods.

A comparative analysis of denoising efficiency using estimated sigma maps demonstrates that DRUNET with SDNet sigma-map estimator provides the state-of-the-art performance, very close to the ideal case, when the true sigma-maps are used in DRUNet.

REFERENCES

[1] J. Astola and P. Kuosmanen, *Fundamentals of nonlinear digital filtering*. CRC press, 2020.

[2] V. Lukin, R. Oktem, N. Ponomarenko, and K. Egiazarian, "Image filtering based on discrete cosine transform," *Telecommunications and Radio Engineering*, vol. 66, no. 18, pp. 1685–1701, 2007.

[3] O. B. Pogrebnyak and V. V. Lukin, "Wiener discrete cosine transform-based image filtering," *Journal of Electronic Imaging*, vol. 21, no. 4, p. 043020, 2012.

[4] A. Foi, V. Katkovnik, and K. Egiazarian, "Pointwise shape-adaptive dct for high-quality denoising and deblocking of grayscale and color images," *IEEE transactions on image processing*, vol. 16, no. 5, pp. 1395–1411, 2007.

[5] R. Oktem, L. Yarovslavsky, and K. Egiazarian, "Signal and image denoising in transform domain and wavelet shrinkage: A comparative study," in *9th European Signal Processing Conference (EUSIPCO 1998)*. IEEE, 1998, pp. 1–4.

[6] D. L. Donoho and I. M. Johnstone, "Minimax estimation via wavelet shrinkage," *The annals of Statistics*, vol. 26, no. 3, pp. 879–921, 1998.

[7] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Transactions on Image processing*, vol. 12, no. 11, pp. 1338–1351, 2003.

[8] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[9] A. Buades, B. Coll, and J. M. Morel, "On image denoising methods," *CMLA Preprint*, vol. 5, 2004.

[10] M. Mahmoudi and G. Sapiro, "Fast image and video denoising via nonlocal means of similar neighborhoods," *IEEE signal processing letters*, vol. 12, no. 12, pp. 839–842, 2005.

[11] V. Jain and S. Seung, "Natural image denoising with convolutional networks," *Advances in neural information processing systems*, vol. 21, pp. 769–776, 2008.

[12] 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.

[13] K. Zhang, W. Zuo, and L. Zhang, "Ffdnet: Toward a fast and flexible solution for cnn-based image denoising," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4608–4622, 2018.

[14] 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*, 2019, pp. 1712–1722.

[15] J. Chen, J. Chen, H. Chao, and M. Yang, "Image blind denoising with generative adversarial network based noise modeling," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3155–3164.

[16] K. Isogawa, T. Ida, T. Shiodera, and T. Takeguchi, "Deep shrinkage convolutional neural network for adaptive noise reduction," *IEEE Signal Processing Letters*, vol. 25, no. 2, pp. 224–228, 2017.

[17] Y. Wang, X. Song, and K. Chen, "Channel and space attention neural network for image denoising," *IEEE Signal Processing Letters*, vol. 28, pp. 424–428, 2021.

[18] A. Danielyan and A. Foi, "Noise variance estimation in nonlocal transform domain," in *2009 International Workshop on Local and Non-Local Approximation in Image Processing*. IEEE, 2009, pp. 41–45.

[19] X. Liu, M. Tanaka, and M. Okutomi, "Noise level estimation using weak textured patches of a single noisy image," in *2012 19th IEEE International Conference on Image Processing*. IEEE, 2012, pp. 665–668.

[20] S. Pyatykh, J. Hesser, and L. Zheng, "Image noise level estimation by principal component analysis," *IEEE transactions on image processing*, vol. 22, no. 2, pp. 687–699, 2012.

[21] M. Ponomarenko, N. Gapon, V. Voronin, and K. Egiazarian, "Blind estimation of white gaussian noise variance in highly textured images," *Electronic Imaging*, vol. 2018, no. 13, pp. 382–1, 2018.

[22] 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.

[23] V. V. Lukin, D. V. Fevralev, N. N. Ponomarenko, S. K. Abramov, O. B. Pogrebnyak, K. O. Egiazarian, and J. T. Astola, "Discrete cosine transform- based local adaptive filtering of images corrupted by nonstationary noise," *Journal of Electronic Imaging*, vol. 19, no. 2, p. 023007, 2010.

[24] A. A. Shulev, A. Gotchev, A. Foi, and I. R. Roussev, "Threshold selection in transform-domain denoising of speckle pattern fringes," in *Holography 2005: International Conference on Holography, Optical Recording, and Processing of Information*, vol. 6252. International Society for Optics and Photonics, 2006, p. 625220.

[25] 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*. Springer, 2015, pp. 234–241.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[27] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *arXiv preprint arXiv:2008.13751*, 2020.

[28] S. G. Bahncmiri, M. Ponomarenko, and K. Egiazarian, "Deep convolutional autoencoder for estimation of nonstationary noise in images," in *2019 8th European Workshop on Visual Information Processing (EUVIP)*. IEEE, 2019, pp. 238–243.

[29] [Online]. Available: https://webpages.tuni.fi/imaging/tampere21/

[30] V. Hosu, H. Lin, T. Sziranyi, and D. Saupe, "Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment," *IEEE Transactions on Image Processing*, vol. 29, pp. 4041–4056, 2020.

[31] Z. Ying, H. Niu, P. Gupta, D. Mahajan, D. Ghadiyaram, and 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*, 2020, pp. 3575–3585.

[32] N. Ponomarenko, O. Eremeev, V. Lukin, and K. Egiazarian, "Statistical evaluation of no-reference image visual quality metrics," in *2010 2nd European Workshop on Visual Information Processing (EUVIP)*. IEEE, 2010, pp. 50–54.

[33] 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.

[34] A. Kaipio, M. Ponomarenko, and K. Egiazarian, "Merging of mos of large image databases for no-reference image visual quality assessment," in *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*. IEEE, 2020, pp. 1–6.

[35] B. Lim, S. Son, H. Kim, S. Nah, and 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*, 2017, pp. 136–144.

[36] K. Ma, Z. Duanmu, Q. Wu, Z. Wang, H. Yong, H. Li, and L. Zhang, "Waterloo Exploration Database: New challenges for image quality assessment models," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 1004–1016, Feb. 2017.

[37] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 126–135.

[38] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 372–387, 2015.

[39] M. Ponomarenko, S. G. Bahnemiri, K. Egiazarian, O. Ieremeiev, V. Lukin, V.-T. Peltoketo, and J. Hakala, "Color image database htid for verification of no-reference metrics: Peculiarities and preliminary results," in *2021 9th European Workshop on Visual Information Processing (EUVIP)*. IEEE, 2021, pp. 1–6.

[40] V. V. Lukin, S. K. Abramov, N. N. Ponomarenko, M. L. Uss, M. S. Zriakhov, B. Vozel, K. Chehdi, and J. T. Astola, "Methods and automatic procedures for processing images based on blind evaluation of noise type and characteristics," *Journal of applied remote sensing*, vol. 5, no. 1, p. 053502, 2011.

[41] M. Ponomarenko, S. G. Bahnemiri, and K. Egiazarian, "Deep convolutional network for spatially correlated rayleigh noise suppression on terrasar-x images," in *2020 IEEE Ukrainian Microwave Week (UkrMW)*. IEEE, 2020, pp. 458–463.

[42] X. Liu, M. Tanaka, and M. Okutomi, "Noise level estimation using weak textured patches of a single noisy image," in *2012 19th IEEE International Conference on Image Processing*. IEEE, 2012, pp. 665–668.