LQAID: LOCALIZED QUALITY AWARE IMAGE DENOISING USING DEEP CONVOLUTIONAL NEURAL NETWORKS

Sathya Veera Reddy Dendi, Chander Dev, Narayan Kothari, and Sumohana S. Channappayya

Indian Institute of Technology Hyderabad

ABSTRACT

In this paper we propose the Localized Quality Aware Image Denoising (LQAID) technique for image denoising using deep convolutional neural networks (CNNs). LQAID relies on local quality estimates over global cues like noise standard deviation since the perceptual quality of a noisy image is typically spatially varying. Specifically, we use localized quality maps generated using DistNet, a spatial quality map estimation method. These quality maps are used to augment the noisy image and guide the denoising process. The augmented noisy image is denoised using a deep fully convolutional network (FCN) trained using mean square error (MSE) as the loss function. The proposed approach shows state-of-the-art performance both qualitatively and quantitatively on two vision datasets: TID 2008 and BSD500. We also show that the proposed approach possesses excellent generalization ability. Lastly, the proposed approach is completely blind since it neither requires information about the strength of the additive noise nor does it try to explicitly estimate it.

***Index Terms*—** Distortion map, denoising, fully convolutional network and perceptual quality.

1. INTRODUCTION

Image denoising is perhaps one of the oldest and most widely studied problems in the computer vision community. We believe that one of the primary reasons for image denoising to be a challenge is the spatially varying perception of noise. It is well-known that the perception of noise is influenced by the local signal strength (or local signal variance). For example, if we apply additive white Gaussian noise (AWGN) noise uniformly to a pristine natural image, the human visual system (HVS) will not perceive distortions equally across the image. High texture regions of an image mask distortions due to noise to a greater extent compared to low texture regions. This perceptual property of the HVS provides us the motivation for our work. We hypothesize that image denoising that is guided by local quality (or distortion) estimates is much more effective than using global cues such as noise standard deviation. A challenge however is to be able to objectively localize image distortion. Importantly, localized noise strength in the image cannot be obtained from a single noise parameter like its standard deviation.

The Structural SIMilarity (SSIM) index [1], a full reference image quality assessment (FRIQA) algorithm allows us to objectively localize image distortion, but it *cannot* be used in our blind setting because it is computed using both the reference and noisy images, thereby posing a further challenge to implementing our hypothesis. An alternative approach is to estimate the distortion map in the blind or no reference image quality assessment (NRIQA) setting. Several NRIQA algorithms provide distortion maps at varying levels of spatial resolutions [2] [3] [4]. Our prior work DistNet [2] has the ability to localize distortions at a higher resolution compared to other NRIQA techniques, and is used in this work.

Given the long, vast and rich history of the image denoising problem, the literature is replete with several excellent solutions. However, we will only briefly review recent and relevant methods due to space constraints. One of the early deep learning based contributions is image denoising using stacked auto-encoder [5]. Xie *et al.* [6] proposed using a combination of sparse representation and DNN with pre-trained denoising auto-encoders. Lefkimmiatis *et al.* [7] proposed image denoising using CNN techniques based on non-local image modeling. Zhang *et al.* [8] proposed an image denoising technique by combining discriminative learning based methods and model-based optimization methods. In general, model-based optimization methods are slow but accurate while discriminative learning methods are fast but they are task specific. A combination of these methods has the advantages of both. DnCNN [9] is a feed-forward neural network based image denoising technique with residual batch normalization. RED Net [10] is an image restoration technique using deep fully convolutional encoder-decoder like neural network with symmetric skip connections. Skip connections pass the information that may be lost due to the depth of the network. Chen *et al.* [11] proposed a blind denoising technique using a generative adversarial network (GAN). These deep learning based methods have moved the state-of-the-art significantly forward.

The proposed approach is presented next, followed by performance evaluation, discussion, and conclusions.

We thank Visvesvaraya PhD scheme, Media Asia Lab, MeitY, Government of India for the financial support.

2. PROPOSED APPROACH

Our proposed approach is motivated from the fact that local and global perceptual quality of a natural image need not be same. This can be observed in Figure 1, where a pristine image is distorted with AWGN (with mean 0 and standard deviation σ of 50) resulting in the noisy image and corresponding localized distortion map generated using DistNet [2] are shown. We perceive more noise on the face compared to the hair. This observation is incorporated into the proposed denoising approach using localized distortion maps, where we used a FCN for denoising with symmetric skip connections. The DistNet and the proposed denoising approach using FCN are described next.

2.1. DistNet [2]

DistNet is a distortion map generation technique which accepts as input a natural image and generates a distortion map as the output. The network architecture of DistNet is similar to that of SegNet [12] which has a convolutional autoencoder like structure. It is trained using distorted images as inputs and their corresponding SSIM maps [1] as target labels. The distortion maps generated using DistNet have been shown to have excellent perceptual agreement with SSIM maps. In the interest of space, we refer the readers to DistNet [2] for architecture and implementation details. In this work, we have not trained the DistNet further but rather have used the weights of the DistNet network as-is.

2.2. LQAID using FCN

The network architecture of the LQAID using FCN is motivated from [10] and incorporates an extra branch for augmenting the main denoising branch with the distortion maps as shown in Figure 1. The network architecture of LQAID using FCN has two parts: an encoder and a decoder. The encoder has two branches; one of which accepts the noisy image as input and the other accepts the distortion map of the noisy image as input. Each branch has 10 convolutional layers and the convolutional layer output of the distortion map branch is concatenated to the convolutional layer output of the noisy image branch. This way we augment the noisy image with the localized quality map to achieve quality guided denoising. In the decoder, we use another set of 10 convolutional layers and each convolutional layer output is concatenated with the corresponding symmetric convolutional layer output from the encoder. In total, the proposed LQAID using a FCN has 30 convolutional and 10 concatenation layers. Each convolution layer has 128 filters with size 3×3. We describe the dataset preparation and training procedure next.

To train LQAID, we synthetically created a dataset by taking 4797 pristine images from the Waterloo Challenge Dataset [13] and generating their distorted versions by adding white Gaussian noise (AWGN). We chose 4 values for the noise standard deviation ($\sigma = 10, 30, 50$ and 70) that result in 4 different quality levels and a total of 19188 noisy images. The pristine and noisy images are further divided into patches of size $224 \times 224 \times 3$ to match the input size requirement of the DNNs. The distortion maps of all the noisy image patches are generated using DistNet [2].

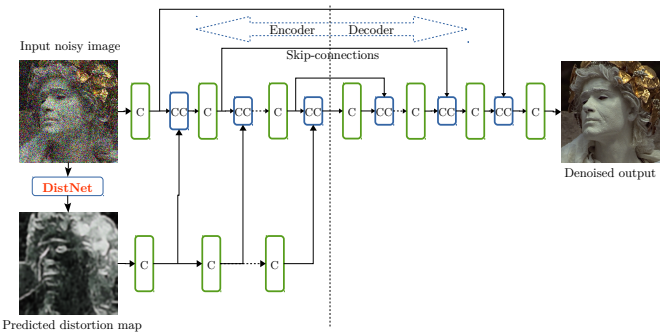


Fig. 1: The network architecture of proposed LQAID using FCN: C indicates convolution layer and CC indicates concatenation layer. Each convolution layer has 128 filters and the size of each filter is 3×3.

Let us assume that a pristine image x is corrupted with additive noise $n \sim \mathcal{N}(0, \sigma^2)$ so that the noisy image can be defined as $y = x + n$. The distortion map of the noisy image is defined as $d = \mathcal{G}(y)$, where $\mathcal{G}(.)$ represents DistNet which maps a noisy image to its corresponding localized distortion map. The goal of this work is to learn a function $\mathcal{F}(.)$ that can "best" map a noisy input image to a denoised output image. The function $\mathcal{F}(.)$ accepts as input the noisy image y, the distortion map d, and is parameterized by Θ. The parameters Θ have to be learnt from the training data such that the function $\mathcal{F}(.)$ can "best" denoise the input image. In this work, we define "best" to be the function that minimizes the mean square error (MSE) between the denoised image and the pristine image.

Let $\mathcal{X} = \{x_1, x_2 ... x_N\}$, $\mathcal{Y} = \{y_1, y_2, ... y_N\}$ and $\mathcal{D} = \{d_1, d_2, ... d_N\}$ be the set of corresponding pristine images, noisy images, and localized distortion maps respectively. We optimize the LQAID using mean square error (MSE) as the loss function and RMSProp as the optimization algorithm. The loss function $\mathcal{L}(\Theta)$ is defined as:

$$\mathcal{L}(\Theta) = \frac{1}{N} \sum_{i=1}^{N} \| \mathcal{F}(y_i, d_i; \Theta) - x_i \|^2 \qquad (1)$$

The training dataset is split in the ratio 80:20 for training and validation respectively. The network is trained over several epochs and the early stopping criterion is employed to avoid over-fitting. The performance of the proposed LQAID using FCN is evaluated on two popular vision datasets: TID2008 [14] and BSD500 [15] using qualitative and quantitative analyses as well as an ablation test. Performance

evaluation is described in Section 3.

3. PERFORMANCE EVALUATION

We demonstrate the effectiveness of the proposed approach using qualitative and quantitative evaluation as well as ablation testing.

3.1. Qualitative evaluation

Figure 2 shows the qualitative comparison of the proposed approach with a few state-of-the-art techniques using a popular (*monarch*) image that is corrupted with AWGN with standard deviation (σ) set to 50 and having zero mean. We can clearly observe that the proposed approach outperforms the state-of-the-art methods qualitatively. Specifically, we would draw the reader's attention to the head and wings of the butterfly as well as the flowers in the background (along the left image edge). We observe that in Figure 2g, the higher texture region is denoised better compared to the state-of-the-art methods. We attribute this improvement to the perceptual guidance provided by the distortion maps as compared to using global cues.

We also evaluated the proposed approach by adding signal-dependent noise. By signal-dependent noise we mean noise strength that varies with signal strength. The high variance regions of the image are distorted with higher strength noise while low variance regions are distortion with lower strength noise, and vice-versa. Figure 3 shows the results of the proposed approach on signal-dependent noise. In Figure 3a, the image is of size $224 \times 224 \times 3$. It is divided into non-overlapping patches of size $56 \times 56 \times 3$ and corrupted with AWGN noise whose σ depends on the patch variance. To find the σ value for a particular patch, we normalized the patch variances such that they sum up to one. We then chose σ to be $100\times$ normalized variance of that patch. The denoised result of the proposed approach is shown in Figure 3b. Similarly, in Figure 3c we applied higher noise to low variance patches and the denoised image is shown in Figure 3d. It is to be noted that the proposed approach works well even in this scenario despite it being trained to denoise spatially uniform noise.

3.2. Quantitative Evaluation

Table 1 shows the competitiveness of the proposed approach in quantitative terms by comparing the image quality of the proposed method with state-of-the-art methods as measured using the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index [1] over the TID2008 and BSD500 datasets. The numbers reported in the tables are average scores over the entire dataset. We again observe that the proposed approach outperforms the state-of-the-art methods. We would like to point to the reader that columns in italic correspond to noise levels not used for training. The proposed approach shows competitive performance even on previously unseen noise strengths and thereby demonstrates the generalization ability of our method.

3.3. Ablation Test

In this section, we demonstrate the efficacy of the distortion map in the proposed approach by performing an ablation test. Specifically, we train and test the proposed network (that uses a FCN) with and without augmenting the distortion map and report its performance quantitatively in Table 2. We can clearly observe that the distortion map contributes to improved performance. Through this ablation test we demonstrate the importance and the utility of the distortion map in providing perceptual guidance for image denoising.

4. DISCUSSION

As summarized in the previous section, Tables 1, 2 and Figures 2, 3 show the quantitative and qualitative performance of the proposed approach respectively. From these tables and figures it is clear that the proposed approach outperforms the state-of-the-art methods both quantitatively and qualitatively. Further, the generalization ability of our approach is also demonstrated in Table 1 through the noise levels shown in italics, and through the signal-dependent noise experiment shown in Figure 3. We attribute the high performance of our proposed algorithm to the ability of the distortion maps to guide the denoising function with perceptual cues. This is especially evident in both the high frequency and the low frequency regions of the images in our qualitative examples. It is also worth noting that our method provides these perceptual cues in a completely blind setting where no prior information about the pristine image or the noise is available. We believe that this approach could be applied to other tasks such as image restoration and image super resolution.

5. CONCLUSIONS

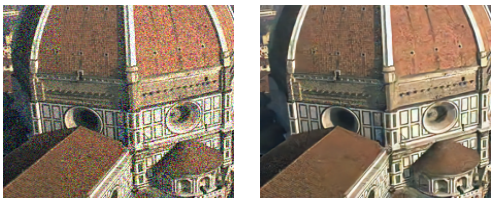
We presented a deep neural network based image denoising approach called LQAID that is aided by local quality (distortion) information. We demonstrated that our denoising approach outperforms the state-of-the-art denoising methods both qualitatively and quantitatively on two image datasets. Further, we showed that the proposed approach has excellent generalization ability by testing it with noise levels not used for training and with signal-dependent noise. Our work also shows the importance of using local perceptual distortion cues for denoising as opposed to using standard global noise cues. To the best of our knowledge, this is the first completely blind denoising approach that makes of local distortion information. As future work, we would like explore this augmentation approach in other restoration tasks such as image in painting, super resolution and deblurring.

(a) Pristine Image; PSNR/SSIM | (b) Noisy Image, 18.48 dB/0.4245 | (c) Predicted DistNet map | (d) FFDNet [16], 26.56 dB/0.866 | (e) DnCNN [9], 26.59 dB/0.8478 | (f) IRCNN [8], 26.34 dB/0.8534 | (g) LQAID using FCN, **28.21 dB/0.8923**

Fig. 2: Qualitative evaluation of the LQAID and comparison with state-of-the-art denoising techniques: Figure (a), (b) and (c) are the reference, noisy images and predicted distortion map of noisy image using DistNet respectively, Figure (d)-(f) are denoised results using state-of-the-art techniques. Figure (g) is generated using proposed LQAID.

Table 1: Quantitative demonstration of the proposed LQAID on TID2008 [14] / BSD500[15] with AWGN using average PSNR (in dB) and SSIM. The noise levels in italic were not used for training.

PSNR							
σ	10	*20*	30	*40*	50	*60*	70
BM3D [17]	35.17/34.72	*31.46/31.19*	29.28/29.11	*28.02/27.51*	26.66/26.56	*25.39/25.68*	25.14/24.92
EPLL [18]	34.56/34.11	*31.07/30.93*	29.16/28.86	*27.79/27.46*	26.47/26.13	*24.92/25.12*	24.71/24.59
DnCNN [9]	33.37/32.96	*30.07/29.61*	28.25/27.76	*26.10/26.50*	25.96/25.51	*24.94/24.55*	23.73/23.42
IRCNN [8]	34.40/34.00	*30.95/30.39*	29.06/28.45	*27.86/27.25*	26.92/26.32	*21.39/21.24*	16.84/16.83
FDDNet [16]	34.44/34.00	*31.02/30.46*	29.20/28.58	*27.98/27.35*	27.07/26.43	*26.36/25.72*	25.77/25.14
LQAID	**35.50/35.02**	***31.65/31.22***	**30.55/29.94**	***29.12/28.58***	**28.17/27.58**	***27.30/26.74***	**26.51/25.98**
SSIM							
σ	10	*20*	30	*40*	50	*60*	70
BM3D [17]	0.968/0.961	*0.938/0.931*	0.912/0.907	*0.874/0.867*	0.856/0.848	*0.828/0.835*	0.816/0.806
EPLL [18]	0.941/0.939	*0.931/0.924*	0.899/0.897	*0.871/0.855*	0.843/0.829	*0.814/0.826*	0.811/0.802
DnCNN [9]	0.903/0.906	*0.825/0.826*	0.762/0.760	*0.709/0.706*	0.659/0.656	*0.598/0.597*	0.515/0.518
IRCNN [8]	0.923/0.929	*0.859/0.861*	0.805/0.805	*0.766/0.761*	0.730/0.723	*0.384/0.396*	0.205/0.218
FDDNet [16]	0.924/0.929	*0.861/0.863*	0.811/0.810	*0.771/0.767*	0.738/0.731	*0.711/0.702*	0.688/0.677
LQAID	**0.972/0.974**	***0.945/0.944***	**0.929/0.926**	***0.908/0.904***	**0.887/0.882**	***0.867/0.861***	**0.847/0.840**



(a) High variance - High noise, SSIM = 0.7170 | (b) LQAID using FCN, SSIM = 0.9313

(c) High variance - Low noise, SSIM = 0.6531 | (d) LQAID using FCN, SSIM = 0.9138

Fig. 3: Performance evaluation of the LQAID on signal-dependent noise.

Table 2: Performance comparison of the LQAID without and with augmenting the distortion map on BSD200 [15] dataset.

PSNR in dB				
σ	10	30	50	70
Without	33.65	29.03	26.87	25.41
With	**34.92**	**29.35**	**26.92**	**25.53**
SSIM				
σ	10	30	50	70
Without	0.9678	0.9159	0.8618	0.8316
With	**0.9736**	**0.9189**	**0.8683**	**0.8321**

6. REFERENCES

[1] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[2] Sathya Veera Reddy Dendi, Chander Dev, Narayan Kothari, and Sumohana S Channappayya, "Generating image distortion maps using convolutional autoencoders with application to no reference image quality assessment," *IEEE Signal Processing Letters*, vol. 26, no. 1, pp. 89–93, 2018.

[3] Le Kang, Peng Ye, Yi Li, and David Doermann, "Convolutional neural networks for no-reference image quality assessment," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1733–1740.

[4] KVSNL Manasa Priya, Balasubramanyam Appina, and Sumohana Channappayya, "No-reference image quality assessment using statistics of sparse representations," in *2016 International Conference on Signal Processing and Communications (SPCOM)*. IEEE, 2016, pp. 1–5.

[5] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 1096–1103.

[6] Junyuan Xie, Linli Xu, and Enhong Chen, "Image denoising and inpainting with deep neural networks," in *Advances in neural information processing systems*, 2012, pp. 341–349.

[7] Stamatios Lefkimmiatis, "Non-local color image denoising with convolutional neural networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017, pp. 5882–5891.

[8] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang, "Learning deep cnn denoiser prior for image restoration," in *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*. IEEE, 2017, pp. 2808–2817.

[9] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei 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.

[10] Xiaojiao Mao, Chunhua Shen, and Yu-Bin Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Advances in neural information processing systems*, 2016, pp. 2802–2810.

[11] Jingwen Chen, Jiawei Chen, Hongyang Chao, and Ming Yang, "Image blind denoising with generative adversarial network based noise modeling," in *Computer Vision and Pattern Recognition (CVPR), 2018 IEEE Conference on*. IEEE, 2018.

[12] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *arXiv preprint arXiv:1511.00561*, 2015.

[13] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang, "Waterloo exploration database: New challenges for image quality assessment models," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 1004–1016, 2017.

[14] N Ponomarenko, "Tid2008-a database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, pp. 30–45, 2009.

[15] DoronTal Jitendra Malik DavidMartin, Charless-Fowlkes, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int'l Conf. Computer Vision*, July 2001, vol. 2, pp. 416–423.

[16] Kai Zhang, Wangmeng Zuo, and Lei 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.

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

[18] Daniel Zoran and Yair Weiss, "From learning models of natural image patches to whole image restoration," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 479–486.