ESTIMATION OF SIGNAL-DEPENDENT NOISE LEVEL FUNCTION USING MULTI-COLUMN CONVOLUTIONAL NEURAL NETWORK

Jingyu Yang[†] , *Xin Liu*[†] , *Xiaolin Song*[*†] , *Kun Li*[‡]

† School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China
‡ School of Computer Science and Technology, Tianjin University, Tianjin 300072, China

ABSTRACT

To estimate the levels of signal-dependent noise (SDN) from a single image is challenging. This paper proposes a novel method to estimate the noise level function (NLF) from a single image using a Multi-column Convolutional Neural Network (MC-Net) with an end-to-end architecture. The MC-Net is trained on a synthesized dataset containing noisy images with known NLFs, and it allows to learn rich hierarchical features using three sub-networks. Moreover, this method performs end-to-end training to retain more details for pixel-wise noise level estimation. Experimental results indicate that our method is accurate and robust to estimate NLFs of SDN for various types of images.

Index Terms— Noise estimation, signal-dependent noise, convolutional neural network, deep learning

1. INTRODUCTION

Several types of noise, e.g. dark-current noise, shot noise, and quantization noise [1], are coupled together in digital imaging systems, which make noise characteristics complex and difficult to model. Thus, it is challenging to estimate the noise levels accurately, which is vital important to achieve high performance denoising and adjust the parameters in image processing. Many methods have been proposed to estimate the noise level for additive white Gaussian noise (AWGN) [2, 3, 4], which is the simplest yet the most popular signal-independent noise (SIN) model. However, image sensors often generates signal-dependent noise (SDN) mainly due to the highly nonlinear camera response.

Image sensor noise is often modeled as signal-dependent noise (SDN), whose levels vary with intensity levels referred to as noise level function (NLF) with respect to intensity levels. NLF estimation from a single image is challenging to distinguish whether the signal variations are due to noise, or due to texture and color variations. Uss et al. [5] proposed an NLF estimation method for hyperspectral imaging systems from multiple images; Liu et al. [6] addressed the SDN estimation from a single image based on Bayesian inference which assumes a piece-wise smooth prior model and operates on the image segmentation. Yang et al. [7] proposed a SDN estimation method based on sparse representation of NLF. Nam et al. [8] proposed a data-driven approach for determining the parameters of the noise model. However, if the assumed model is violated for highly-textured images, the estimation of noise levels would be inaccurate, especially for low-occurrence intensities.

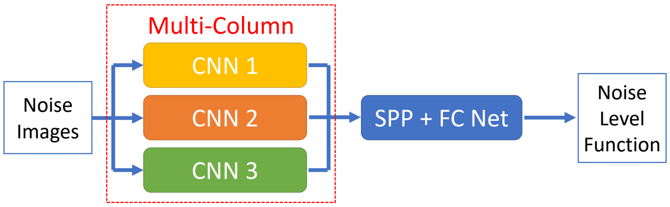


Fig. 1. The framework of our method. The input is the observed noisy images and the output is the estimation of NLFs.

With the capability of deep learning in solving both discrimination and regression problems, almost all computer vision problems have been investigated by deep neural networks with different structures, achieving satisfactory performance in many research fields. Jain et al. [9] and Xu et al. [10] used convolutional neural network for denoising. Dong et al. [11] used Super-Resolution Convolutional Neural Network (SRCNN) for image super-resolution, and proposed a Faster Super-Resolution Convolutional Neural Network (FSRCNN) in Ref. [12]. However, to the best of our knowledge, no existing work has investigated NLF estimation using convolutional neural network, which could be a very strong candidate in solving this problem with proper structure and training.

Without explicitly imposing an image model, we propose a novel SDN level estimation method using Multi-column Convolutional Neural Network (MC-Net) shown in Fig.1, which learns a noise level function from a single noisy image. This network is trained end-to-end on noisy images synthesized from BSD-500 dataset with known NLFs as labels, and is tested with images to validate the fitting accuracy. Experimental results show that the proposed MC-Net achieves

This work was supported by the National Natural Science Foundation of China under Grant 61372084, Grant 61571322 and Grant 61520106002.
*Corresponding Author: songxl@tju.edu.cn

accurate NLF estimation for images of various characteristics, even for images with skew intensity distributions.

2. METHOD

This section first introduces the SDN model, and then presents the proposed MC-Net for NLF estimation.

2.1. Noise Model for Image Sensors

SDN models describe the noise levels as a function of intensity levels [6, 7, 13]. Note that our proposed method is a data-driven approach that is not limited to particular noise models. In order to compare with the existing methods, without loss generality, we use the following model refered in Ref. [7] to synthesis noise images.

$$\begin{aligned} I &= f(L_I), \\ I_N &= f(L_I + n_s + n_c) + n_q, \end{aligned} \tag{1}$$

where I is the ideal noise-free image, I_N is the observed noisy image, $f(\cdot)$ is the camera response function (CRF), $n_s \sim N(0, L_I\sigma_s^2)$ denotes all the noise components (assumed zero mean Gaussian noise with variance $L_I\sigma_s^2$) that are dependent on irradiance L_I, $n_c \sim N(0, \sigma_c^2)$ is the independent noise before gamma correction, and n_q is the additional quantization noise, which is ignored in this model since it is usually small compared with other noise.

To evaluate the performance of proposed method, we use noise synthesis according to the SDN model in Eq. (1), with the original image I and noisy image I_N, the noise level function (NLF) with respect to the intensity level I is estimated as $\sigma(I) = \sqrt{E[(I_N - I)^2]}$. It can be further rewritten as

$$\sigma(I; f, \sigma_s, \sigma_c) = \sqrt{E[(I_N(f^{-1}(I), f, \sigma_s, \sigma_c) - I)^2]}, \tag{2}$$

where $I_N(\cdot)$ is the noise synthesis process.

2.2. Multi-column CNN for NLF estimation

The rationale of noise estimation is to infer the noise level from noisy pixels with similar intensities. Pixels with the same intensity level can distribute locally in particular areas and/or non-locally over the whole image. For reliable estimation, both local and non-local correlation should be fully utilized. We found that a standard one column convolutional neural network owns this capability. As shown in Fig. 2, the baseline CNN consists of three convolutional layers followed by max-pooling layers and two fully-connected hidden layers. All layers use Rectified Linear Unit (ReLU) [14] as activation functions. The input of the network is a noisy color image and the output is the estimated NLF. The local kernels in convolutional and max-pooling layers exploit local correlation while the one-to-one connections in fully-connected hidden layers is capable of exploring non-local correlation. Therefore, the structure of CNN is quite suitable for the task of NLF estimation.

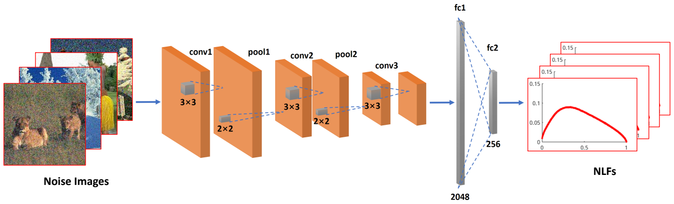


Fig. 2. Architecture of one column CNN.

However, the baseline structure is not adequate to provide multi-scale processing, which could be beneficial for NLF estimation. According to the theory of statistical estimation, the estimation variance could be significantly reduced by increasing the number of independent samples, which suggests using lager kernels in the convolutional layer. However, natural images contains both smooth regions and textured regions. The main challenge in NLF estimation is to suppress the influence of image variations (mainly due to edges and textures) in noise level estimation. Large kernels are advantageous for large smooth regions, but would overestimate the noise level due to the inclusion of image variations. For reliable estimation, we should introduce a strategy to adapt the kernel size to image content.

Inspired by Multi-column Deep Neural Networks (MDNNs) [15] for image classification, we use the multi-column structure to provide different kernel sizes in different columns, whose overall architecture is illustrated in Fig. 3. There are three parallel CNNs whose structures are the same but filters of convolutional layers are with different sizes to capture spatially-varying smoothness over natural images. Strides of the first filter in each column are set to two and average pooling is applied for each 2 × 2 region to reduce the size of feature maps. ReLU is adopted as the activation function. The feature maps of three paths are concatenated and fed into a spatial pyramid pooling (SPP) layer [16] which produces a fixed dimension output, followed by two fully-connected hidden layers that output the NLF estimation.

SPP for NLF estimation. A standard CNN requires the input image to have a fixed size since the output of the last convolutional layer needs to have a predefined dimensionality. For this reason, we have to first resize the input image to fixed spatial dimension. Such a resolution adaptation does not affect the overall performance in many applications such as classification and detection, but would certainly change the noise characteristics to be estimated. To solve this problem, motivated by the success of SPP-net [16], we add a spatial pyramid pooling layer between the convolutional layers and the fully-connected layers of the network. The SPP layer extracts and aggregates the features of the last convolutional layer through spatial pooling, where the size of the pooling regions depends on the size of the input. Thus, we can maintain

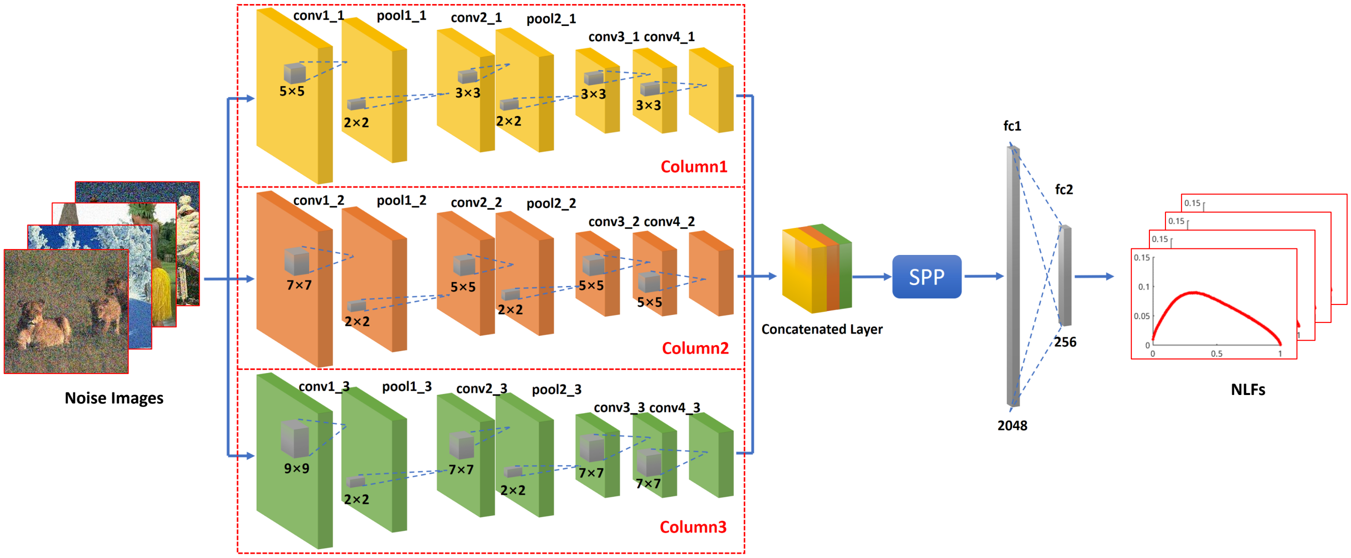


Fig. 3. Architecture of the proposed MC-Net. For each convolutional layer, the number of filters are 64 while the filter size for each layer is presented in the figure. Average pooling is applied for each 2 × 2 region. The feature maps of three columns are concatenated and fed into a SPP layer and followed by two fully-connected layers that output the NLF estimation. Batch normalization is added between convolution layer and ReLU.

the fixed output dimensionality for the SPP layer and make the network to adapt to different sizes of input images. Further details of SPP is refered to [16].

Batch Normalization. Batch normalization [17] is proposed to alleviate the internal covariate shift by incorporating a normalization step and a scale-and-shift step before the nonlinearity in each layer. Two free parameters updated by back-propagation are added per activation. The incorporation of bach normalization brings not only faster and stabler convergence in training, but also better performance.

2.3. Training

Denote by $D = \{(I_j, p_j), j \in [1, N]\}$ the training set, where (I_j, p_j) denotes the j^{th} training pair, I_j denotes the observed noisy image, and p_j is the corresponding NLF. The number of training pairs is N.

We learn the parameters $\widehat{\Theta}$ of the proposed MC-Net by minimizing the following loss function:

$$\widehat{\Theta} = \arg\min_{\Theta} \sum_{j \in [1,N]} \frac{1}{2N} \|\hat{p_j}(I_j, \Theta) - p_j\|_2^2 \qquad (3)$$

where $\hat{p_j}(I_j, \Theta)$ is the estimated NLF for the input image I_j by MC-Net parametrized with Θ.

Just like typical training of CNN, we optimize parameters by error back-propagation with the mini-batch stochastic gradient descent (SGD) algorithm. However, it is difficult to train the three-column CNN simultaneously because of computational complexity and the problem of gradient vanishing. Inspired by the pre-training of restricted Boltzmann machine (RBM) [18], we pre-train each single column CNN to estimate NLFs, and then use these pre-trained models to initialize the three-column MC-Net and fine-tune the parameters simultaneously.

3. EXPERIMENTS AND RESULTS

3.1. Setting

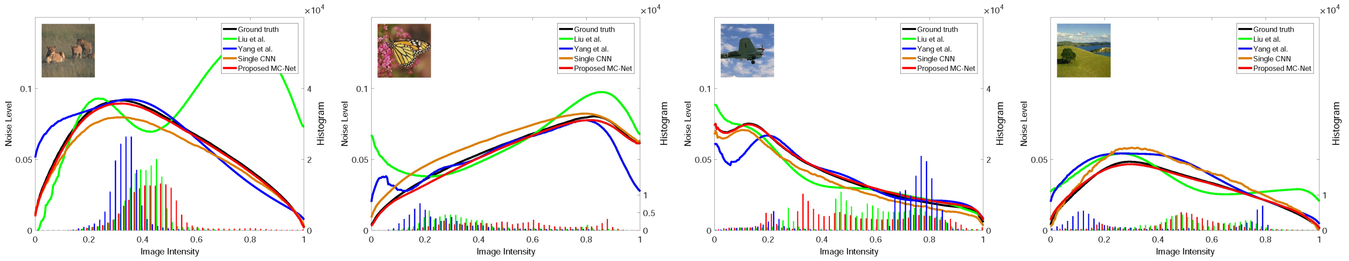
Datasets. Following Ref. [19, 20], we use the same 400 training images, and crop a 256×256 region from each image. Then, noise with different noise levels is added to each region, resulting in over 60000 training samples of size 256×256.

Implementation Details. In the training of MC-Net, we set the weight decay to 0.0001, the momentum to 0.9 and the mini-batch size to 64. The learning rate is 0.001 for pre-training and 0.0001 for fine-tuning. In convolutional layers and fully connected layers, we use Xavier to initialize the weights and the biases are initialized to one.

Our method is implemented in Caffe [21], and we run experiments on a computer equipped with an Intel Core i7 CPU, 32GB RAM, and NVIDIA GeForce GTX TITAN X GPU.

3.2. Experimental Results

We use two datasets to evaluate the performance of our method: one is the dataset containing 100 natural images from Berkeley segmentation dataset (BSD500) and the other



(a)

(b)

(c)

(d)

Fig. 4. NLFs recovered from synthetic SDN different parameters (a) $\{CRF(60), 0.08, 0.04\}$, (b) $\{CRF(170), 0.16, 0.04\}$, (c) $\{CRF(30), 0, 0.03\}$, (d) $\{CRF(60), 0.04, 0.02\}$, comparing with Liu et al. [6], Yang et al. [7] and single CNN. (Best viewed on screen)

Table 1. NLF estimation results in MSE ($\times 10^{-4}$)

	Lena	Starfish	Baboon	Barbara	Boat	Girl	Goldhill	House	Pen	Pepper	Avg.
Liu et al.	2.70	4.07	3.73	2.34	2.78	3.32	3.02	12.4	9.36	1.51	4.53
Yang et al.	0.20	1.09	1.54	0.49	0.58	0.25	**0.12**	6.20	0.17	0.13	1.08
Single CNN	0.31	0.98	2.30	0.87	0.76	0.37	0.31	4.50	0.51	0.24	1.12
Proposed MC-Net	**0.12**	**0.76**	**1.31**	**0.22**	**0.45**	**0.12**	0.13	**2.80**	**0.10**	**0.09**	**0.61**

is the dataset containing some popular test images (e.g., lena, Peppers, Starfish and Barbara).

We compare the proposed MC-Net method with three noise estimation methods on images with synthetic SDN, including Bayesian-based method [6], sparse presentation based method [7], and a standard single convolutional neural network. Following [7], for the Bayesian-based method, the SLIC superpixel method instead of K-means in [6] is used to over-segment the image into more coherent regions.

As show in Fig. 4, four test images are polluted by synthetic SDN with four sets of parameters $\{CRF(60), 0.08, 0.04\}$, $\{CRF(170), 0.16, 0.04\}$, $\{CRF(30), 0, 0.03\}$ and $\{CRF(60), 0.04, 0.02\}$. The four test images are diverse in characteristics such as background colors, textural features, and intensity distributions. The histograms of intensity levels are shown to suggest narrow or flat distribution of the range of intensities. As show in Fig. 4(a), the image *lions* has a quite narrow range of intensity and mainly contains textures, no method is effective enough for the low end $(0.0 \sim 0.2)$ and high end $(0.8 \sim 1.0)$ except our MC-Net. The reason is that our MC-Net is able to learn efficient features with the multi-scale setting from image content with spatially-varying smoothness and supports. Fig. 4(b)~Fig. 4(d) show NLFs recovery of images with various noise levels. Bayesian-based method [6] does not provide satisfactory results and sparse representation based method [7] is suitable to flat and abundant intensity distribution but ineffective for low-occurrence intensities. The standard single column CNN is able to learn the shape of NLFs, but suffers from over-or under-estimation. Our method can recover stable and accurate NLFs from images with different characteristics and various noise levels.

Table 1 shows results in terms of mean squared error (MSE) between ground truth and estimated NLFs of ten natural images polluted by synthetic SDN with parameters $\{CRF(60), 0.08, 0.04\}$. These images are not included in the training dataset. The results by our method have the lowest MSEs for most cases. Besides, we evaluate the performance of proposed method in a test set containing over 10000 synthetic noise images, and the average MSE is 0.62×10^{-4}. Under the same desktop computer in the CPU mode, the running time of our method for testing an image of size 256×256 is $0.30s$ while the methods in [7] and [6] takes $1.19s$ and $3.16s$, respectively. These results demonstrate that proposed method significantly outperform the state-of-the-art in SDN estimation task in terms of both accuracy and computation.

4. CONCLUSION

This paper proposes a multi-column convolutional neural network (MC-Net) to estimate the signal-dependent noise level from a single image. The proposed MC-Net is trained with an end-to-end architecture, and each column has the same architecture with different convolutional kernel size in order to retain details from each pixel and acquire different scale feature maps of each layer. Experimental results show that our method is accurate and robust for diverse images with various NLFs. In future work, we will apply accurate NLFs to denoising tasks of natural images and seek more effective data-driven methods for signal-dependent noise (SDN) processing.

5. REFERENCES

[1] Yanghai Tsin, Visvanathan Ramesh, and Takeo Kanade, "Statistical calibration of ccd imaging process," in *IEEE International Conference on Computer Vision*, 2001, pp. 480–480.

[2] Mohammed Ghazal and Aishy Amer, "Homogeneity localization using particle filters with application to noise estimation," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 1788–96, 2011.

[3] R. C. Bilcu and M. Vehvilainen, "New method for noise estimation in images," in *IEEE-Eurasip Nonlinear Signal and Image Processing*, 2005, p. 25.

[4] Xiaohui Yuan and Bill P. Buckles, "Subband noise estimation for adaptive wavelet shrinkage," in *International Conference on Pattern Recognition*, 2004, pp. 885–888.

[5] M. L Uss, B Vozel, Vladimir V Lukin, and K Chehdi, "Local signal-dependent noise variance estimation from hyperspectral textural images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 469–486, 2011.

[6] Ce Liu, Richard Szeliski, Sing Bing Kang, C. Lawrence Zitnick, and William T. Freeman, "Automatic estimation and removal of noise from a single image," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 30, no. 2, pp. 299–314, 2007.

[7] J. Yang, Z. Gan, Z. Wu, and C. Hou, "Estimation of signal-dependent noise level function in transform domain via a sparse recovery model.," *IEEE Transactions on Image Processing*, vol. 24, no. 5, pp. 1561–72, 2015.

[8] Seonghyeon Nam, Youngbae Hwang, Yasuyuki Matsushita, and Seon Joo Kim, "A holistic approach to cross-channel image noise modeling and its application to image denoising," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1683–1691.

[9] Viren Jain and H. Sebastian Seung, "Natural image denoising with convolutional networks.," in *Conference on Neural Information Processing Systems*, 2008, pp. 769–776.

[10] Qingyang Xu, Chengjin Zhang, and Li Zhang, "Denoising convolutional neural network," in *IEEE International Conference on Information and Automation*, 2015, p-p. 1184–1187.

[11] Chao Dong, Change Loy Chen, Kaiming He, and Xiaoou Tang, *Learning a Deep Convolutional Network for Image Super-Resolution*, Springer International Publishing, 2014.

[12] Chao Dong, Chen Change Loy, and Xiaoou Tang, "Accelerating the super-resolution convolutional neural network," in *European Conference on Computer Vision*. Springer, 2016, pp. 391–407.

[13] Xinhao Liu, Masayuki Tanaka, and Masatoshi Okutomi, "Estimation of signal dependent noise parameters from a single image," in *IEEE International Conference on Image Processing*, 2013, pp. 79–82.

[14] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, no. 2, pp. 2012, 2012.

[15] Ciregan Dan, Ueli Meier, and Jrgen Schmidhuber, "Multi-column deep neural networks for image classification," in *IEEE Conference on Computer Vision & Pattern Recognition*, 2012, pp. 3642–3649.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 37, no. 9, pp. 1904–16, 2015.

[17] Sergey Ioffe and Christian Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *Computer Science*, 2015.

[18] G. E. Hinton, S Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets.," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[19] Yunjin Chen and Thomas Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *arXiv preprint arXiv:1508.02848*, 2015.

[20] 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. PP, no. 99, pp. 1–1, 2016.

[21] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 675–678.