10.2352/EI.2022.34.8.IMAGE-288

Mix-loss trained bias-removed blind image denoising network

Yi Yang [a] [1] *, Chih-Hsien Chou* [b] *, Jan P. Allebach* [a]
[a]. *Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA.*
[b]. *Futurewei Technologies Inc., Santa Clara, CA 95050, USA.*

Abstract

Deep learning has been extensively studied in a variety of aspects of image denoising, including blind, unsupervised, universal, and perceptually oriented. This paper focuses on blind, bias-removed, mix loss optimized, and perceptually oriented image denoising task. We propose a network designed for AWGN image denoising. Our network removes the bias at each layer to achieve the benefits of scaling invariant network. Additionally, it implements a mix loss function to boost performance. We train and evaluate our denoising results using PSNR, SSIM, and the perceptual metric LPIPS, and demonstrate that our results achieve impressive performance evaluated with both objective and subjective IQA metrics.

1. Introduction

Due to the limitations of various recording devices, images are sensitive to random noise during acquisition. Noise is signal distortion that impedes image observation and information extraction. Thus, as a fundamental topic of image analysis and processing, image noise suppression aids our understanding of image statistics and processing.

There are numerous noise sources during the imaging process, including photon-shot noise, photon-to-charge conversion noise, analog-to-digital conversion noise, hot pixels, read noise, compression artifacts, and so on. The most frequently concerned noises in the literature are Additive White Gaussian Noise (AWGN), impulse (salt-and-pepper) noise, Poisson noise, and speckle noise. Impulse noise, speckle noise, and Poisson noise are caused primarily by defective manufacturing, bit errors, and an insufficient photon count. AWGN is most commonly seen in analog circuitry during image acquisition and transmission. In most image denoising tasks, the noise is assumed to be additive white Gaussian noise (AWGN).

Digital images can be viewed as a matrix or a two-dimensional signal containing gray level intensity or color channel values: $(x, y(x))$, where x is the index of a pixel; $y(x)$ is the value indicating the pixel's intensity at a given location in a grayscale image; or indicating a three-value array representing the red, green, and blue channels in a color image. An image AWGN noise model can be thought of in the following manner:

$$y(x) = m(x) + n(x) \qquad x \in Z^2 \tag{1}$$

where $m(x)$ represents the original image, $y(x)$ represents the noisy observation and $n(x)$ represents the Gaussian noise with zero mean and standard deviation σ. The denoising problem requires determining a function $f : R^N \rightarrow R^N$, that can accurately estimate the original image $m(x)$. This issue is typically resolved by minimizing the mean squared error : $f = argmin_g E||m(x) - g(y(x))||^2$. In deep learning, the denoising function g is parameterized by the weights of the network, so the optimization is performed over these parameters.

Image denoising algorithms have made significant strides with the development of imaging technology over the last few decades. Denoising algorithms are classified into the following categories according to their type [1]: Spatial domain filtering, which includes average filter, Weiner filter, bilateral filters, non-local mean and so on. Transform domain thresholding, which makes use of the Fourier transform (FT), fast Fourier transform (FFT), discrete cosine transform (DCT) and discrete wavelet transform (DWT). The most prevalent approach in this category is color block matching 3D collaborative filtering (CBM3D) [2]. Apart from these, additional well-established denoising approaches include random field based methods Markov random field (MRF) and hidden Markov models (HMM) [3]; as well as sparsity based methods such as K-singular value decomposition and convolution sparse representation (CSR) and the dictionary learning method [1].

Recently, as network architectures became more flexible, deep learning has been widely used to deal with image restoration problems, resulting in significant advances in image denoising. The convolutional neural networks (CNNs) were proposed and quickly gained popularity in image/video processing and low-level computer vision. As a result, CNNs were frequently used for AWGN denoising or feature extraction for removing real-world noise or hybrid noise. For addressing multiple low-level tasks by one model, a denoising CNN (DnCNN) [4] consisting of convolutions, batch normalization (BN), rectified linear unit (ReLU) and residual learning (RL) was proposed to deal with image denoising, super-resolution, and JPEG image deblocking. Taking into account the tradeoff between denoising performance and speed, a color non-local network (CNLNet) [5] efficiently reduced image noise by combining non-local self-similarity (NLSS) and CNN. In

[1] This work was done as a part of internship at Futurewei.

terms of blind denoising, a fast and flexible denoising CNN (FFDNet) [6] used different noise levels and the noisy image patch as the input of a denoising network, thus ensuring that the denoiser can adapt to all noise levels; and a convolutional blind denoising network (CBDNet) [7] removed the noise from the given real noisy image by two sub-networks for blind denoising. To address the lack of clean reference images, Generative Adversarial Network (GAN) based methods (e.g. GCBD) and CNN based unsupervised/weakly supervised methods (e.g. Noise2Noise, DIP, Noise2Self) are denoising methods that do not require clean images [8].

The idea of a bias-free CNN [9] also inspired us. It points out that CNN feedforward neural networks with ReLUs are piecewise affine, implying that the net network bias fluctuates wildly outside the training range, leading to the CNN overfitting to the noise levels within the training range and underfitting to the noise levels outside of the training range. The authors suggested that the issue could be ameliorated by removing additive (bias) terms at each stage of the network, resulting a bias-free CNN network (BF-CNN).

Although the loss layer is the primary driver of network learning, dealing with image restoration tasks is often accomplished by using L1 (MAE) or L2 (MSE). The cost function is typically set to the L2 for the majority of AWGN denoising tasks. However, L2 has a number of well-documented drawbacks, for instance, L2 has a weak correlation with perceived image quality by human observers [10]; L2 will significantly blur image details; and the use of L2 makes implicit assumptions about the noise, such as that the noise is additive and independent of the image, which is not suitable to real world noise.

Contributions

In this work, we focus on the CNN-based AWGN blind denoising task. We study the networks with and without bias, then train a bias-removed network, and compare its denoising performance to that of biased networks across various noise levels to demonstrate its superiority on the denoising task. We investigate the impact of loss functions on the performance of the CNN-based image restoration network and design a mix loss that combines MSE and SSIM for our bias-removed network. We train a denoiser and compare its performance to that of traditional MSE loss functions in the AWGN denoising task to demonstrate the advantage of mix loss. Additionally, we utilize not only the conventional image quality assessment metrics SSIM and PSNR, but also the cutting-edge perceptual image quality assessment (IQA) metric LPIPS to evaluate our denoising results and comparison results, showing that our results perform well on both objective and subjective IQA metrics. We have made the following contributions:

- First, we develop a bias-removed image denoising network, demonstrating that it is capable of handling the AWGN denoising over a wide range of noise levels while training over a very narrow range of noise levels. The bias-removed network achieves impressive denoising results on trained noise levels, and outperforms the state-of-the-art bias network on untrained noise levels.
- Second, we design a mix loss that combines MSE and SSIM, and use it as the loss function in our bias-removed network to further improve denoising performance. Our network is fully capable of denoising AWGN while retaining visually pleasing image details, so it outperforms the deep denoiser using a conventional L2 loss function in terms of perceptual image quality assessment.
- Finally, we study objective and subjective image quality assessment, and comprehensively evaluate our denoising results and comparison results using objective and subjective image quality assessment metrics.

2. Related Work

Deep learning-based image denoising has been extensively studied for denoising tasks with many aspects such as blind, unsupervised, universal and perceptually oriented. Our research focuses on a blind, bias-removed, mix loss optimized, and perceptually oriented image denoising task. Therefore, we will discuss related work and progress in these aspects.

Bias-removed network

Feedforward neural networks with rectified linear units (ReLUs) are piecewise affine since each of these is affine. For a noisy input image y, the function f_1 computed by a denoising neural network with bias may be written:

$$f_1 = W_L R(W_{L-1}...R(W_1 y + b_1) + ...b_{L-1}) + b_L = A_n y + b_n \quad (2)$$

where W_n is the weight of the convolutional layers, b_n is the bias (additive constants), R is the activation function ReLU, $A_n \in R^{N\times N}$ is the Jacobian of f_1 evaluated at input y, and b_n represents the net bias. Also, a bias-removed denoising neural network can be derived as follows:

$$f_2 = W_L R(W_{L-1}...R(W_1 y)) = A_n y \quad (3)$$

The work [9] demonstrates a viewpoint for several popular deep models, including DnCNN, Recurrent CNN, UNet and DenseNet, that these biased convolutional networks overfit to noise levels within the training range, but underfit to noise levels outside of the training range, because the net bias fluctuates wildly over the untrained noise level. This issue can be addressed by removing bias (additive terms) from each stage of the network, resulting in a bias-removed CNN. If bias-removed CNN has ReLU activations, the denoising map is locally homogeneous, and thus invariant to scaling. This property behaves as shown in the formula:

$$f_2(\alpha y) = \alpha f_2(y) \quad (4)$$

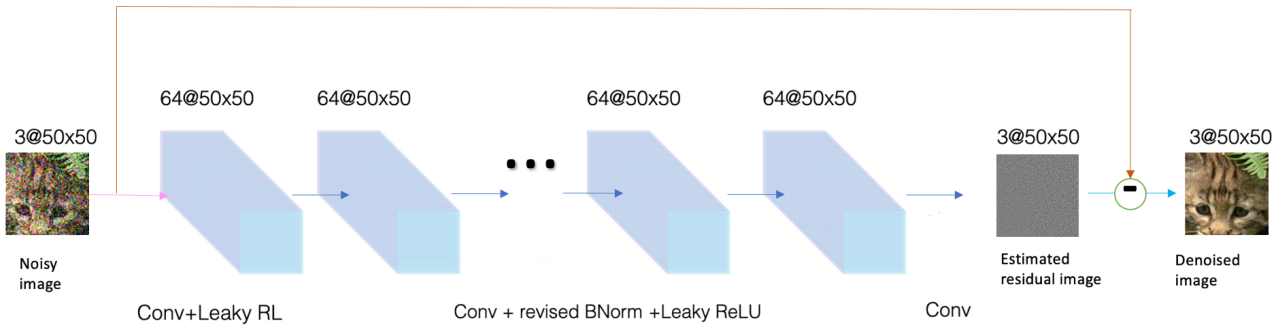

Figure 1. *The network architecture of our denoiser. We adopt 17 convolutional layers for the main stem of the network, each consisting of 3 × 3 filters and 64 channels, revised batch normalization according to bias-removal, and a leaky ReLU activation function. Also, the network includes residual blocks.*

The bias-removed neural network has the scaling invariant property, which means that rescaling the input by a constant value simply rescales the output by the same amount, and this property is intuitively desirable for a denoising method that operates on natural images.

Loss functions for image denoising

Different losses compute the similarity between the estimated image $g(y)$ and the clean image (ground truth) m in different ways. We list several important loss functions for image restoration tasks.

- Mean absolute error (MAE) or L_1 is defined as:

$$Loss_{L_1}(P) = \frac{1}{N}\sum |g(y(p)) - m(p)| \tag{5}$$

- Mean squared error (MSE) or L_2 is defined as:

$$Loss_{L_2}(P) = \frac{1}{N}(g(y(p)) - m(p))^2 \tag{6}$$

For Formulas (5) and (6), p is the index of the pixel, P is the patch, and N is the number of pixels in the patch; $g(y(p))$ and $m(p)$ are the values of the pixels in the processed patch and the ground truth, respectively.

- The Structural Similarity Index (SSIM) is a perceptual metric. SSIM is based on visible structures in the image. It is a perceptual metric used to quantify the image quality in terms of luminance, contrast, and structural similarity. SSIM is defined as:

$$SSIM(p) = \frac{2\mu_{g(y)}\mu_m + c_1}{\mu_{g(y)}^2 + \mu_m^2 + c_1}\frac{2\sigma_{g(y)m} + c_2}{\sigma_{g(y)}^2 + \sigma_m^2 + c_2} \tag{7}$$

  where $\mu_{g(y)}$ and μ_m are the averages of $g(y)$ and m, respectively, $\sigma_{g(y)}^2$ and σ_m^2 are the variance of $g(y)$ and m, respectively, and $\sigma_{g(y)m}$ is the covariance of $g(y)$ and m. Also, $c_1 = 1 \times 10^4$ and $c_2 = 9 \times 10^4$ are two constants, which are used to stabilize the division with a weak denominator. The SSIM loss is defined as:

$$Loss_{SSIM}(P) = 1 - SSIM(p) \tag{8}$$

In the image restoration task, each loss function has its advantages and disadvantages. The paper [10] focuses on the performance of various loss functions in multiple image restorations and proposes a mix loss function. Inspired by this work, we propose and develop a novel mix loss function that is suitable for our bias-removed denoising networks. We demonstrate that our denoised image quality outperforms those that were trained with conventional loss in terms of perceptual assessment.

Image quality assessment (IQA) metrics

While assessing the perceptual similarity of two images is straightforward for humans, the underlying processes are regarded to be highly complex. The widely used image quality assessment measures, such as PSNR and SSIM, are far too simplistic to account for subtleties in human perception. Therefore, many researchers have studied image quality assessment metrics based on human perceptual similarity.

The work [12] summarizes and compares a variety of models for evaluating the quality of full-reference images. They tested 11 IQA models, which include objective, subjective, conventional, and deep learning based models. They proved that the Learned Perceptual Image Patch Similarity model (LPIPS) [11] score, which is computed automatically by a pretrained network, is positively correlated with human subjective evaluation in the image denoising task. Thus, we use PSNR, SSIM, and LPIPS to comprehensively evaluate the denoised image quality in our work.

3. Main Body

We propose a network designed for AWGN image denoising. Our network removes bias at each layer to achieve the benefits of bias-removed networks. Additionally, it implements a mix loss

function to boost performance. AWGN noisy image and clean image pairs serve as inputs. The overall architecture is shown in Figure 1.

Network architectures

We developed our network architecture based on the benchmarking network, Denoising CNN (DnCNN) [4]. DnCNN is composed of 17 or 20 convolutional layers, depending on whether the task is non-blind or blind. Each layer of DnCNN comprises 3×3 filters and 64 channels, batch normalization, and a ReLU activation function. It has a skip connection from the initial layer to the final layer.

We set the depth of our AWGN denoising network to 17 layers. This depth is used by DnCNN to train a non-blind denoiser, which has a relatively shallower depth and thus requires a smaller dataset. We also use this depth to train a non-blind denoiser, but our non-blind denoiser will have the ability to perform blind denoising. Our network includes the following attributes:

- Bias removed. We removed all additive constants at each layer of the network and made corresponding changes to the batch normalization function.
- Mix loss optimized. We design and implement a mixed function of L2 and SSIM, as indicated by the following formula.

$$mixLoss = \alpha Loss_{L_2} + (1-\alpha) Loss_{SSIM} \tag{9}$$

where $Loss_{L_2}$ represents the MSE function of pixel-by-pixel comparison, $Loss_{SSIM}$ represents SSIM Loss function we mentioned before, and α is the loss weight set to 0.6. We chose the value for this parameter so that the contribution of the two losses would be roughly balanced. Also, based on a published reference [10], the results were not significantly sensitive to small variations of α.

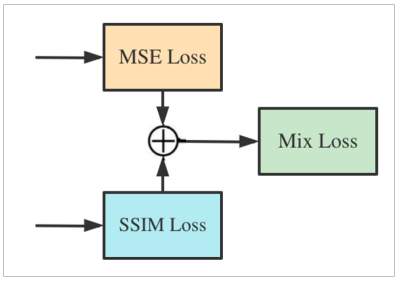

Figure 2. *Mix Loss.*

- Leaky ReLU used. We adapt the activation function to a leaky ReLU instead of the ReLU.

Thus, the network includes the first layer of convolutional layer and leaky ReLU, and then uses convolutional layer, revised batch normalization and leaky ReLU for layers $2 \sim (L-1)$. For the last layer, the convolutional layer is used to reconstruct the estimated residual image.

The denoiser trained from the above-mentioned network has a dual application: 1). It can be used to train blind denoisers directly, and the blind denoiser can produce cutting-edge denoising results for AWGN noisy images. 2). It can be used to train non-blind denoisers on limited noise ranges and smaller datasets, while the non-blind denoiser still maintains the ability of a blind denoiser to remove noise at any level, including levels beyond the training range. This attribute will be shown in detail in the subsequent results section.

Dataset

The training datasets are divided into two categories: grayscale images and color images. We used well-known datasets: The Berkeley Segmentation Dataset [13] includes the datasets BSD400, composed of 400 grayscale images; and BSD500, composed of 500 color images. The test datasets we used include grayscale image test datasets Set12 and Set68, and color image test datasets Set5 and Set14. All images are in png format.

We applied the following data augmentation and data preprocessing steps: 1). Using cubic interpolation, resize each image by different scales (1, 0.9, 0.8, and 0.7). 2). Flip or rotate each image at random. 3). Divide each image into small patches of 50×50 pixels. 4). Divide training images into a training set and a validation set at random. 5). Save training, validation, and test sets in an h5 (Hierarchical Data Format 5) file.

Training

After the data augmentation, we input a total of 655,900 small color patches with a size of 50×50 pixels. The loss function is mix loss combined by MSE and SSIM. We initialize the weights using the Kaiming weight initialization method [14] and use the Adam Optimizer [15]. For Adam, the learning rate = $1e^{-3}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 1e^{-8}$. We train on Nvidia Titan Xp graphics cards, check the Loss and PSNR per batch, and validate in the end of each epoch. We train a grayscale image denoiser on noise levels $\sigma \in [0,30]$ using the grayscale dataset and a color image denoiser on noise levels $\sigma \in [0,50]$ using the color dataset.

Results and Evaluation

In our work, we use PSNR, SSIM, and LPIPS to assess image quality. Numerous researchers have found that PSNR, which is derived from MSE, frequently contradicts human perceptual assessment. SSIM is a basic perceptually motivated metric, but it still represents only a limited perceptual assessment. However, PSNR and SSIM remain the most mainstream IQA metrics in the image processing area. For PSNR and SSIM, higher scores indicate greater similarity between the two images (denoised image and ground truth). Additionally, we use LPIPS to assess the image quality. LPIPS, as previously stated, is a cutting-edge deep-

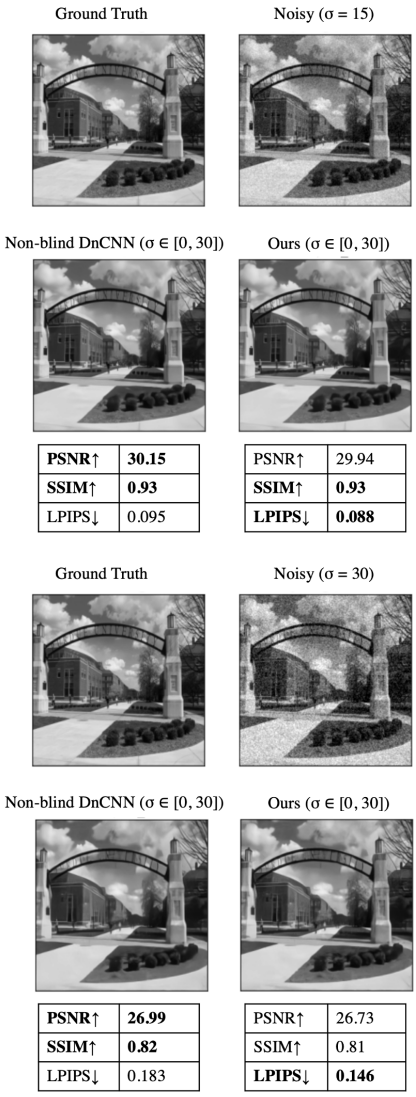

Figure 3. *Visualized results of denoising for σ = 15 and 30. The best values are highlighted in bold.*

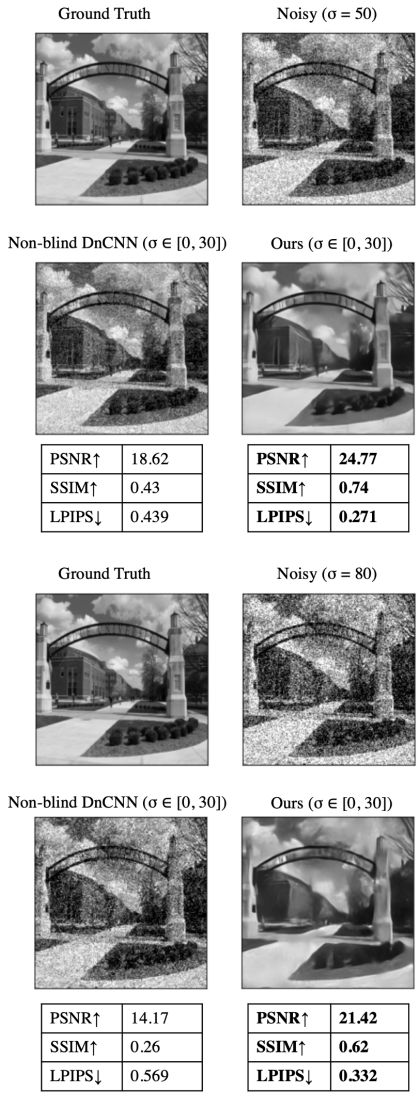

Figure 4. *Visualized results of denoising for σ = 50 and 80. The best values are highlighted in bold.*

learning-based perceptual assessment metric that quantifies image similarity in units of JND, and correlates positively with human perceptual assessment. For LPIPS, a lower score indicates a greater similarity.

The performance of the denoiser based on grayscale images, and the advantages of the bias-removed network is illustrated in Figures 3 and 4. The non-blind denoiser is trained at a given range of noise $\sigma \in [0,30]$, and compared to a non-blind DnCNN pretrained in the same range. The two models are compared at four different levels of noise: $\sigma = 15$, $\sigma = 30$, $\sigma = 50$ and $\sigma = 80$. $\sigma = 15$ and $\sigma = 30$ are two test noise levels that fall within the training range, while $\sigma = 50$ and $\sigma = 80$ are two significantly higher noise levels that exceed the training range. For $\sigma = 15$ and $\sigma = 30$, as illustrated in Figure 3, the denoising performance of the two models is comparable because both models are trained on this noise interval. Our results are slightly lower on PSNR but remain competitive on SSIM and better on LPIPS. $\sigma = 50$ and $\sigma = 80$, as illustrated in Figure 4, are noise levels higher than the training range for both models. While the non-blind DnCNN is nearly incapable of removing noise at these levels, our model is still capable of doing so effectively. While our denoising performance is insufficient at $\sigma = 80$ noise level, as a non-blind denoiser with narrow training range and shallow network depth, its blind denoising capability is still impressive.

The performance of the denoiser based on color images is illustrated in Figures 5, 6 and 7, and Table 1. Our color image denoiser is trained at a given range of noise $\sigma \in [0,50]$, and compared to a non-blind DnCNN are pretrained in the same range. We compared the two at $\sigma = 15$, $\sigma = 25$ and $\sigma = 50$ noise level, thus, they can all be regarded as blind denoisers. Table 1 shows

the quantitative results for the test dataset Set14.

Our results, even though they are a little lower on PSNR, are competitive on SSIM and impressive on LPIPS. At various noise levels, our denoiser achieves the best LPIPS scores. Considering that LPIPS metrics have a higher correlation with human perceptual assessment, our results outperform DnCNN in subjective assessment. Some qualitative results on images are shown in Figures 5, 6 and 7. We can see that our results preserve details while denoising. This is the benefit of using a relatively perceptual-aware SSIM in the loss function, which mitigates the over-smoothing property of MSE loss.

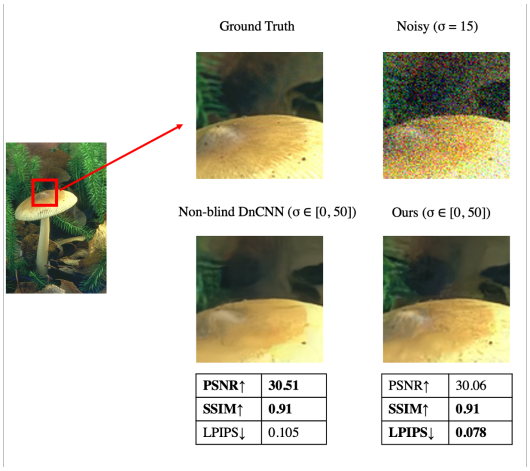

Figure 5. *Visualized results of denoising for σ = 15. The best values are highlighted in bold.*

Gaussian Denoising				
Dataset	Noise Level	IQA Scores	DnCNN	Ours
Set 14	σ = 15	PSNR ↑	**31.22**	31.05
		SSIM ↑	**0.9013**	0.9003
		LPIPS ↓	0.0217	**0.0187**
	σ = 25	PSNR ↑	**28.51**	28.32
		SSIM ↑	**0.8212**	0.8209
		LPIPS ↓	0.1312	**0.0956**
	σ = 50	PSNR ↑	**24.67**	24.33
		SSIM ↑	**0.7264**	0.7249
		LPIPS ↓	0.1763	**0.1378**

Table 1: Overall quantitative results compare with non-blind DnCNN on the same testset Set14. Our proposed method obtain better LPIPS values, and also perform competitive in PSNR and SSIM scores. The best values are highlighted in bold.

4. Conclusion

The modern CNN-based image denoising networks were investigated in this paper, and a CNN-based image denoising network that combines blind, bias-removed, and mix-loss was proposed and implemented. We train and evaluate our denoising results using PSNR, SSIM, and the perceptual metric LPIPS, and demonstrate that our results achieve impressive performance evaluated with both objective and subjective IQA metrics.

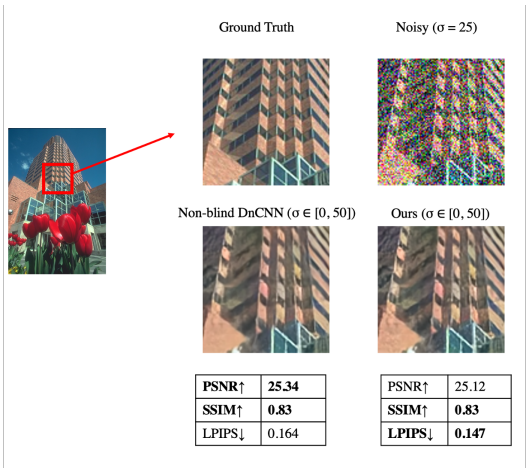

Figure 6. *Visualized results of denoising for σ = 25. The best values are highlighted in bold.*

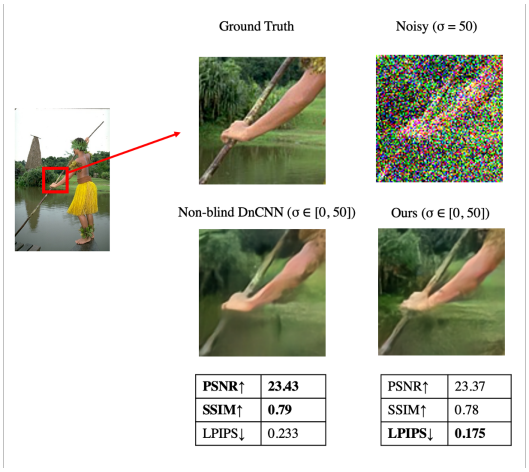

Figure 7. *Visualized results of denoising for σ = 50. The best values are highlighted in bold.*

Acknowledgment

Futurewei Technologies Inc. funded this research while the first author was an intern there. We are grateful to the company's IC Lab for the research assistance.

References

[1] Goyal B, Dogra A, Agrawal S, et al., "Image denoising review: From classical to state-of-the-art approaches", *Information Fusion*, 2020.

[2] Dabov K, Foi A, Katkovnik V, et al., "Image denoising by sparse 3-D transform-domain collaborative filtering", *IEEE Transactions on Image Processing*, 2007.

[3] Mardia K V, "Multi-dimensional multivariate Gaussian Markov random fields with application to image process-

ing", *Journal of Multivariate Analysis*, 1988.
[4] Zhang K, Zuo W, Chen Y, et al., "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising", *IEEE Transactions on Image Processing*, 2017.
[5] Lefkimmiatis S, "Non-local color image denoising with convolutional neural networks", *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
[6] Zhang K, Zuo W, Zhang L, "FFDnet: Toward a fast and flexible solution for cnn-based image denoising", *IEEE Transactions on Image Processing*, 2018.
[7] Guo S, Yan Z, Zhang K, et al., "Toward convolutional blind denoising of real photographs", *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.
[8] Lehtinen J, Munkberg J, Hasselgren J, et al., "Noise2Noise: Learning image restoration without clean data", *International Conference on Machine Learning*, 2018
[9] Mohan S, Kadkhodaie Z, Simoncelli E P, et al., "Robust And Interpretable Blind Image Denoising Via Bias-Free Convolutional Neural Networks", *International Conference on Learning Representations*, 2019.
[10] Zhao H, Gallo O, Frosio I, et al., "Loss functions for image restoration with neural networks", *IEEE Transactions on Computational Imaging*, 2016.
[11] Zhang R, Isola P, Efros A A, et al., "The unreasonable effectiveness of deep features as a perceptual metric", *IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
[12] Ding K, Ma K, Wang S, et al., "Comparison of full-reference image quality models for optimization of image processing systems", *International Journal of Computer Vision*, 2021.
[13] Arbelaez P, Fowlkes C, Martin D, *The Berkeley Segmentation Dataset and Benchmark*, Berkeley University, 2007.
[14] He K, Zhang X, Ren S, et al., "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification", *IEEE International Conference on Computer Vision*, 2015.
[15] Kingma D P, Ba J, "Adam: A method for stochastic optimization", *International Conference on Learning Representations*, 2014.

Author Biography

Yi Yang received her B.S. degree in Geomatics Engineering from Wuhan University and M.S. degree in Geomatics Engineering from Chinese Academy of Sciences. She is currently working on a Ph.D. in Electrical and Computer Engineering at Purdue University. Her primary area of research has been image processing, computer vision and machine learning.

Chih-Hsien Chou is currently a Principal Engineer in Futurewei Technologies, Inc. He has been working in research and development of real-time video / image processing algorithms for mass-produced IC products. He developed wide dynamic range, noise reduction, color correction / enhancement, video stabilization, and autofocusing algorithms for chip products since August 2013. He is the inventor or co-inventor of 15+ US and foreign patents. He has a B.S. degree from Tatung University, Taiwan, a M.S. degree and a Ph.D degree from University of Maryland, College Park, all in Electrical Engineering.

Jan P. Allebach is Hewlett-Packard Distinguished Professor of Electrical and Computer Engineering at Purdue University. Allebach is a Fellow of the IEEE, the National Academy of Inventors, the Society for Imaging Science and Technology (IS&T), and SPIE. He was named Electronic Imaging Scientist of the Year by IS&T and SPIE, and was named Honorary Member of IS&T, the highest award that IS&T bestows. He has received the IEEE Daniel E. Noble Award, the IS&T/OSA Edwin Land Medal, the IS&T Johann Gutenberg Prize, and is a member of the National Academy of Engineering.