Image denoising using convolutional neural network

Behnam Latifi
Department of Electrical Engineering
Amirkabir University of Technology
Tehran, Iran
behnamlatifi@aut.ac.ir

Abolghasem Raie
Department of Electrical Engineering
Amirkabir University of Technology
Tehran, Iran
raie@aut.ac.ir

***Abstract*—Noise removal is one of the classic problems in low-level machine vision and image processing. After camera invention, image noise removal became necessary. Its purpose is to eliminate noise while preserving the edges and other precise and vital details as much as possible. Due to the outstanding results produced by deep learning in a variety of domains, this study introduces a novel model for image denoising that uses a mix of dilated and regular convolution to remove noise from degraded photos with a specified noise level. The proposed model is assessed on the BSD68 dataset, and the findings reveal that it outperforms models such as DnCNN and FFDNet at all noise levels.**

Keywords—Image denoising, convolutional neural network, dilated convolutional neural network

I. INTRODUCTION

Denoising, often known as noise removal, is the process of recovering the original picture from a damaged image by eliminating undesired noise. Noise reduction is a critical field of study because it underpins a wide variety of applications, including object recognition, digital entertainment, and remote sensing photography. Denoising is a necessary preprocessing step for segmentation, feature extraction, and texture analysis, among other operations. During capture, transmission, storage, and retrieval, noise degrades the picture. As the number of image sensors per unit area rises, cameras become increasingly susceptible to noise, making noise reduction a critical step in enhancing picture quality. Shot noise is low in many technical applications due to the high-quality sensors in cameras. However, in certain applications, such as remote sensing and medical instruments, the shot noise can be quite loud.

Images may be damaged with various noises. In this paper, the removal of Additive White Gaussian noise (AWGN) is considered. The reasons for choosing this noise are:

1) Thermal noise occurs in analog circuits of cameras. This kind of noise has a Gaussian distribution with zero mean and is the most often seen in real-world applications.
2) AWGN provides an ideal testbed for evaluating noise removal methods.

Denoising may be accomplished in three ways: spatial domain, transform domain, and learning-based techniques. Spatial approaches make use of the correlations observed in the majority of natural pictures. For each pixel, a sequence of candidates will be employed in the filtering process. Spatial filters may be classified as local or non-local filters, depending on how the candidates are chosen. Local approaches are time efficient. However, if the noise level (standard deviation) is high, these approaches may not perform effectively, since noise weakens the correlation between neighboring pixels. Non-local filters outperform local filters at high noise levels. However, a significant disadvantage of non-local filters is that they may also result in inhomogeneity, such as over smoothing of picture edges.

The second category is transformation domain methods. Decomposition in the wavelet domain is a popular technique. However, strategies for transformation domains have downsides. Because most transform domain approaches employ fixed basis functions, they have difficulty representing realistic images using diverse patterns. Additionally, the number of coefficients used to represent an image patch is equal to the number of pixels included inside the image patch, resulting in artifacts. Thus, in order to address wavelet's weaknesses, it is important to have a redundant dictionary. BM3D [1] achieved extraordinary results in noise cancellation with redundancy in the transform domain and non-local classification in

2022 30th International Conference on Electrical Engineering (ICEE) | 978-1-6654-8087-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICEE55646.2022.9827021

the spatial domain. Wavelet-based methods perform better than spatial domain methods because they have extraordinary properties such as sparsity.

Sparse representations became popular as a field of study and were used to retrieval problems with the emergence of machine learning. The main concept behind dictionary-based learning approaches is that denoising is accomplished by the learning of a large number of image data patches. Each pixel in the predicted picture may be described in this dictionary as a linear combination of patches. Clustering with K-SVD [2], LSSC [3], and CSR [4] are the best dictionary learning techniques that have been introduced. While the majority of dictionary-based learning approaches have shown acceptable performance, the primary limitation of sparse models is their linearity and inability to solve non-convex optimization issues.

In the last decade, deep learning has attracted the attention of many researchers in machine vision. The reasons for using deep learning are: The deep learning approach can be used in almost any field of application and does not require a precisely designed feature. Instead, optimal parameters are learned automatically to get the job done. The deep learning approach is highly scalable. Moreover, as the number of data increases, the performance of traditional machine learning approaches remains constant. Still, the performance of deep learning approaches increases. Denoising has previously been accomplished using convolutional neural networks (CNN), auto-encoder networks (AE), and generative adversarial networks (GAN). CNN was utilized as the basis for our model in this paper.

The remainder of this paper is organized as follows. Sec. II reviews existing deep learning-based denoising methods. Sec. III presents the proposed image denoising model. Sec. IV reports the experimental results. Sec. V concludes the paper.

II. RELATED WORK

Many efforts have been made to denoise the images with the help of deep learning. For example, in [5], the authors combined two main ideas for noise elimination: using a CNN as an image processing architecture and an unsupervised learning process. This study shows that convolutional networks have comparable performance and, in some cases, are better than Markov random field methods. In [6], an effort was made to train a multilayer perceptron to map from noisy to clean pictures. The authors have shown that by training the neural network on a massive database, their model can obtain good performance compared to state-of-the-art methods. In addition, the proposed method is compatible with noise that has been less studied in various sources. A new approach has been used in [7] by combining sparse coding and deep auto-encoder networks to eliminate noise. A unique training procedure for unsupervised learning of features is designed for noise removal and blind inpainting operations. The performance of this method in noise removal is comparable to K-SVD, which is mainly used in the sparse coding method. The proposed process also introduces ways to solve the problem of blind inpainting that has not been presented before. The authors collected data from the Internet and trained the network, and tested the trained network for only four images. In [8], a flexible learning framework based on nonlinear reaction-diffusion models is described. In contrast to older nonlinear diffusion models, all parameters, including filters and influence functions, are trained simultaneously. The results of this model on public datasets demonstrate that this strategy outperforms other methods. The proposed model is compelling due to the simplicity of the structure and small emission steps and it is very suitable for implementation on the GPU. BMCNN uses non-local self-similarities with CNN to solve the noise reduction problem [9]. GAN has also been used to eliminate noise [10]. The proposed DnCNN in [11] uses batch normalization and ResNet to perform noise removal operations. This network conducts denoising operations with the unknown noise level and performs SISR and JPEG image deblocking. FFDNet [12] uses the noise level map and the noisy image to deal with different noise levels. This network on CPU and GPU is faster than BM3D. IRCNN [13] combines model-based optimization with CNN to solve the noise reduction problem, which can perform powerful noise reduction operations with a single model against different noise levels. Also, adding dilated convolution to the network improves the noise reduction performance.

In addition, many other methods have performed well in image denoising. For example, integration of dilated convolution and ResNet has been used to denoise the image [14]. Combining different expert systems is also a good choice for image noise removal [15]. General denoiser networks [16] and deep CNN trained by prior knowledge of image pixels also effectively eliminate impulse noise [17]. In general, deep learning approaches outperform all traditional methods, and the most successful outcomes to date have been acquired using this sort of noise reduction methods.

III. THE PROPOSED DENOISING MOSEL

To remove noise, several models based on deep learning have been utilized, and our proposed model is based on convolutional networks. Figure 1 illustrates the proposed network design. The network input is a noisy image $y = x + n$, in

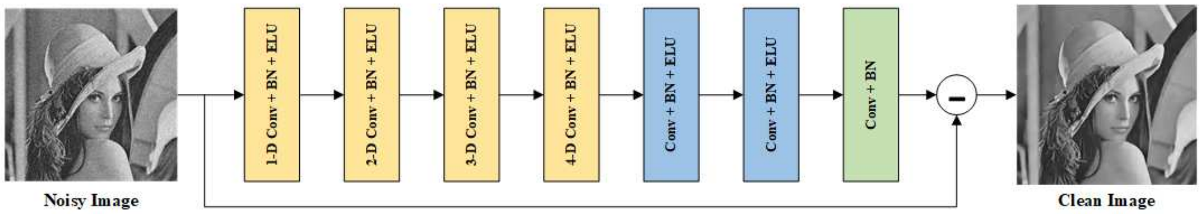

Fig. 1. The architecture of the proposed denoiser network

which x is a clean image and n will be Gaussian noise. The proposed model is trained to learn a residual mapping R(y) ≈ n, and the denoised image is estimated as follows: x = y - R(y). According to [18], optimizing the residual mapping is much simpler when the original mapping is identical. Note that the noisy image is very similar to the clean image (especially with low noise level). Therefore, F(y) ≈ x is identical mapping, and the residual learning formulation is more appropriate for image denoising. According to the findings in [11], integrating residual learning with batch normalization leads in a quicker, more stable convergence and improved noise reduction performance regardless of the optimization approach chosen. We employed batch normalization and residual learning in our model architecture for the reasons stated above.

In general, the model's performance in denoising images (particularly those with low-level noise) is proportional to its complexity (number of parameters), hence the model must be sufficiently complicated to perform effectively. High-level noise significantly increases the damage of photos compared to low-level noise. In order to rebuild highly damaged images, we need more background information which is directly related to the network receptive field. The receptive field of the network is equal to the picture size utilized to reconstruct one noisy pixel. Naturally, a bigger receptive field will recover more background information. Today's machine vision applications are real-time, and the existence of a noise-canceling block in the system is undeniable. As a result, the model's speed must be very fast, and the denoiser unit must quickly give a clean picture to the following blocks. In total, three critical factors come into play while building the suggested network architecture: the number of parameters, the network receptive field, and the model execution time.

Now, we are confronted with a multiobjective optimization challenge while constructing network architecture. There are two primary methods for expanding the receptive field. The first option is to raise the network's depth, while the second approach is to increase the kernels' size. Both of these strategies, as will be seen, are utilized to increase the number of parameters. Both of these ways increase the run time, but there is another way to expand the receptive field without increasing the run time, and that is to use dilated convolution instead of regular convolution. Dilated convolution expands the receptive field while maintaining the kernel parameters. As a result, we employed dilated convolution in the suggested model's first four layers. Naturally, regular convolution was used in the final layers since they are responsible for noise reconstruction and in order to rebuild each pixel, neighbor features are required, but with dilated convolution, spatial neighborhood information is not properly used. There are many approaches for increasing the number of network parameters: increasing the network's depth, increasing the number of filters in the layers, or increasing the kernel size. Because we employed dilated convolution, we did not need to increase the size of the kernels. We discovered that increasing the number of filters had a smaller influence on execution time than increasing the network depth when testing with various models. Without expanding the network depth, we increased the number of filters. Finally, the suggested model's properties are such that the total number of filters in all layers (excluding the last layer, which has one filter) is 64. The size of the kernels is 3×3 in the bottom layer, 7×7 in the fifth layer, and 5×5 in the other layers. The number of trainable parameters of the model is 613250 and receptive field of network is 65×65. In addition, we have used the ELU activation function to speed up learning in our work.

IV. EXPERIMENTAL RESULTS

To train the proposed model for Gaussian denoising with known noise level, we consider five noise levels, i.e., σ = 15, 25, 35, 50, and 75. We set the patch size as 40×40 and crop 128×8000 patches from the training set of BSD500 to train the model. The evaluation criterion of the proposed model is considered PSNR like other sources and articles. MSE is used as cost function to learn the residual mapping, and we use the Adam optimization algorithm and the mini-batch of size 128.

TABLE I. The average PSNR (dB) results of different methods on the BSD68 dataset

Methods	BM3D [1]	WNNM [19]	TNRD [8]	IRCNN [13]	UNLNet [16]	DnCNN [11]	FFDNet [12]	Proposed
σ = 15	31.07	31.37	31.42	31.63	31.47	31.72	31.63	31.76
σ = 25	28.57	28.83	28.92	29.15	28.96	29.23	29.19	29.27
σ = 35	27.08	27.30	-	-	27.50	27.69	27.73	27.78
σ = 50	26.52	25.87	25.97	26.19	26.04	26.23	26.29	26.32
σ = 75	24.21	24.40	-	-	-	24.64	24.79	24.80

TABLE II. Run time (in milliseconds) of the proposed method on images of size 256×256, 512×512 and 1024×1024

Input size	256×256	512×512	1024×1024
Run time	44.77	70.51	147.98

To prevent overfitting and to choose the best model, the validation set of BSD500 was utilized, which has 100 photos. On the BSD68 dataset, the suggested network was tested. To get an accurate average PSNR, we corrupted the pictures of the experimental dataset 1000 times with different noise levels and reported the average noise reduction results as the suggested model's final result. Table 1 summarizes the mean PSNR values obtained using various approaches on the BSD68. In Figure 2, five noisy photos with varying noise levels were picked from the BSD68 dataset to demonstrate the visible outcomes of the noise reduction procedure.

The Google Colab system was utilized to train the suggested model, and the GPU employed in this study is a Tesla V100-SXM2-16GB. Each training epoch lasts around ten minutes. Comparing the execution times of various models without taking into account the system utilized by the researchers is an incorrect technique, and an accurate comparison of the execution times of the models should be done on the same software and hardware platform. As a result, we avoid comparing our model's execution time to that of others and instead represent our model's execution time in Table 2. However, as is customary in publications, the trained model is evaluated using three photos of varying size. As can be observed, the suggested model has a very low execution time, which is quite desired.

V. CONCLUSION

According to the observed findings, the proposed model outperforms models such as DnCNN and FFDNet at all noise levels. At the present, researchers are concentrating their efforts on non-local approaches in deep learning and developing a model for other image restoration tasks such as SISR, JPEG deblocking, and blind denoising. In the future, we'll aim to build a network that uses non-local approaches to reduce noise in pictures with unknown noise levels, and we'll tune hyperparameters and use booster to enhance the existing model.

References

[1] 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.

[2] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image processing,* vol. 15, no. 12, pp. 3736-3745, 2006.

[3] J. Mairal, F. R. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *ICCV*, 2009, vol. 29 :Citeseer, pp. 54-62.

[4] W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in CVPR 2011, 2011: IEEE, pp. 457-464.

[5] V. Jain and S. Seung, "Natural image denoising with convolutional networks", in *Advances in neural information processing systems*, 2009, pp. 769-776.

[6] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?", in *2012 IEEE conference on computer vision and pattern recognition*, 2012: IEEE, pp. 2392-2399.

[7] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks", in *Advances in neural information processing systems*, 2012, pp. 341-349.

[8] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration", *IEEE transactions on pattern analysis and machine intelligence,* vol. 39, no. 6, pp. 1256-1272, 2016.

Fig. 2. Denoising results of five images from BSD68 dataset

[9] B. Ahn and N. I. Cho, "Block-matching convolutional neural network for image denoising", *arXiv preprint arXiv:1704.00524,* 2017.

[10] S. Tripathi, Z. C. Lipton, and T. Q. Nguyen, "Correction by projection: Denoising images with generative adversarial networks", *arXiv preprint arXiv:1803.04477,* 2018.

[11] 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.

[12] 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.

[13] K. Zhang,W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration", in *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3929–3938, July 2017.

[14] T. Wang, M. Sun, and K. Hu, "Dilated deep residual network for image denoising", in *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, 2017: IEEE, pp. 1272-1279.

[15] J. H. Choi, O. A. Elgendy, and S. H. Chan, "Optimal combination of image denoisers", *IEEE Transactions on Image Processing,* 2019.

[16] S. Lefkimmiatis, "Universal denoising networks: a novel CNN architecture for image denoising", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3204-3213.

[17] G. Wang, G. Wang, Z. Pan, and Z. Zhang, "Multiplicative noise removal using deep CNN denoiser prior", in *2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, 2017: IEEE, pp. 1-6.

[18] 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.

[19] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 2862-2869.