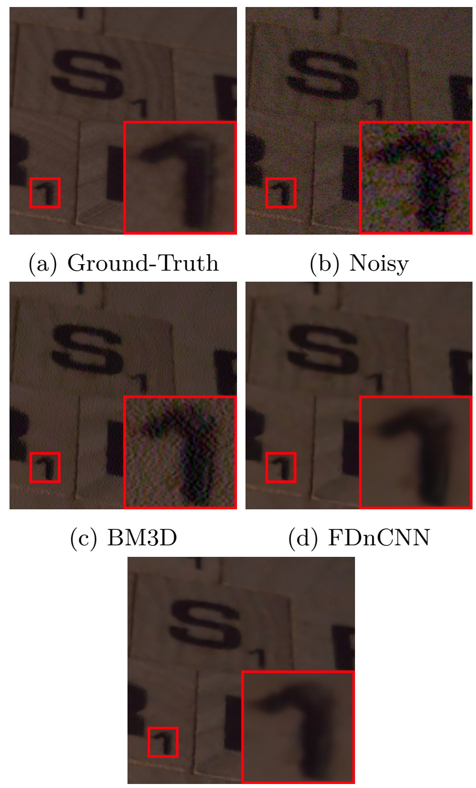
Deep Residual Autoencoder for Real Image Denoising

Doğa Yılmaz
Özyeğin University
Istanbul, Turkey
doga.yilmaz.11481@ozu.edu.tr

M. Furkan Kıraç
Özyeğin University
Istanbul, Turkey
furkan.kirac@ozyeğin.edu.tr

Abstract

Image denoising is one of the fundamental problems in low-level computer vision since it has found more and more real-world applications every day. Various approaches have been used for image denoising throughout the years such as block-matching and 3D filtering (BM3D). In the recent years learning-based approaches have outperformed the traditional methods such as BM3D. However, most of these learning-based methods makes the assumption that the real-world noise is fully modeled with various noise model such as additive white Gaussian noise (AWGN). These methods struggle to achieve outstanding performance when it comes to real-world noide. With the recent release of real-world noise datasets such as Smartphone Image Denoising Dataset (SIDD) and Darmstadt Noise Dataset (DND), the limitation caused by lack of real world noise data has eliminated. In this paper, we propose a deep convolutional autoencoder network combined with symmetric residual connections for real image denoising. We used the real-world images provided by SIDD for the training of the proposed model. Also, we have experimented with L1, L2, SSIM, MS-SSIM and sum of L1 and MS-SSIM loss functions in order to optimize the performance of our proposed model both qualitatively and quantitative. Our experimental results show that our proposed model outperforms the traditional methods and offers similar performance with state-of-the-art methods in blind real image denoising.



(a) Ground-Truth (b) Noisy
(c) BM3D (d) FDnCNN
(e) Ours

Figure 1: A and B are ground-truth and noisy image pair from SIDD dataset [1]. C through E are sample denoised images using [5], [19] and Our method respectively. (Best viewed on high-resolution display.)

1. Introduction

Image denoising aims at removing the noise of a given noisy image which is an essential task in low-level computer vision. Nowadays, with the increase of digital imaging, image denoising has found many real-world use cases such as medical image denoising. In the literature there can be found considerable amount of research done for the removal of various noise models (e.g., salt and pepper noise, additive white Gaussian noise (AWGN)). Although those noise models are used to represent real-world noise, there is a considerable difference between current noise models and real-world noise [1].

In the literature, learning based methods are proven their performance. Recent state-of-the-art image denoising methods such as FDnCNN[19] and REDNet [10] which use deep Convolutional Neural Networks (CNNs) have remarkable denoising performance on various noise models. However, since real-world noise is not fully represented by any noise model, performance of these methods are limited by the noise model used in the training of the models [1]. Even in some cases, depending on the noise model and noise level, traditional methods such as BM3D [5] can outperform the learning-based methods.

Another factor which affects the performance of learning-based methods is the loss function used while training the model. In the literature mean square error (L1 loss) and mean absolute error (L2 loss) are widely used as loss function. However, these functions do not consider the nature of Human Visual System (HVS). As an example, L2 loss assumes that the impact of noise is independent from the characteristics of the image [23]. Due to the sensitivity of the HVS to local luminance, contrast and structure, this assumption causes poor result for human observers [15].

In this paper, we propose a deep convolutional autoencoder with skip connections in order to pass valuable information to the deeper layers of the model. Our model also utilizes several loss functions in order to boost its denoising performance. Also, we trained our model with real-world noisy and ground truth image pairs provided by Smartphone Image Denoising Dataset (SIDD) [1]. Our model can achieve competitive results with state-of-the-art methods such as BM3D [5], FDnCNN [19] and REDNet [10] in blind image denoising in color images.

The rest of the paper is organized as follows. Section 2 provides a brief summary of existing image denoising methods, noise types, datasets and image quality metrics used in the literature. Section 3 presents the proposed method and discusses about its features. In Section 4 we talk about our experimental setup and extensive evaluations. We present our results in Section 5 and conclude the paper by discussing our findings in Section 6.

2. Related Work

In this section, we will present existing traditional and state-of-the-art image denoising methods. Then, we will discuss about available image denoising datasets. Also, we will present image quality assessments which is necessary when measuring the quality of an image in a quantitative manner.

2.1. Traditional Methods

BM3D [5] is widely known as a traditional image denoising method in the literature. It uses effective filtering in 3D transform domain by combining sliding-window transform processing with block matching. Burger et al. [3] showed that when the noisy image does not contain any regular structure, denoising performance BM3D is decreased. Also, it is shown that a plain multi-layer perceptron (MLP) can achieve similar denoising performance with BM3D by Burger et al. [3].

2.2. State-Of-The-Art Methods

One of the recent additions to image denoising literature is denoising autoencoders. Autoencoders aim to learn an approximation to identity function. By taking this property of autoencoders into account, denoising autoencoders forces the model to learn reconstruction of the input given its noisy version [7].

In the literature there are many proposed methods for image denoising which uses denoising autoencoders [7, 16, 4, 17]. Gondara [7] proposed an autoencoder based denoiser for medical imaging domain. Xie [16] et al. and Ye [17] et al. proposed stacking multiple denoising autoencoders in order to better model the noise.

2.3. Datasets

As it is with many machine learning related research, data is one of the major factors which limits the performance. Image denoising is no exception. However, recently with the release of new datasets such as SIDD [1] and Darmstadt Noise Dataset (DND) [13], the bottleneck caused by the lack of high-quality data has decreased. In the literature SIDD and DND are also used as a benchmarking tool for image denoising methods.

2.4. Image Quality Assessment

Image Quality Assessment (IQA) is another related important research topic. It aims to measure the quality of the images in a quantitative way. For image denoising domain, having a quantitative metric to measure the quality of the images precisely is vital for optimizing learning based denoising models. Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) [15] are widely used in the literature for measuring the quality of images. Feature Similarity Index (FSIM) [21] is another quality metric which measures the dissimilarity between two images based on local information. Furthermore, Zhao [22] et al. proposed a novel method which is a combined version of Multi-Scale Structural Similarity (MS-SSIM) and mean square error (MSE) in order to eliminate each methods drawback.

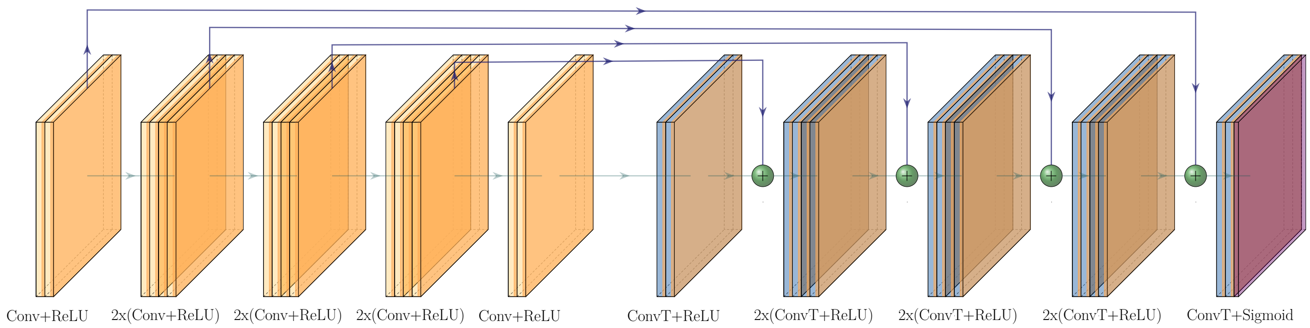


Figure 2: Proposed network architecture where "+" denotes element wise sum of feature maps.

3. Proposed Method

We propose a deep fully convolutional autoencoder with residual connections in order to reduce the degradation problem which occurs in deep networks. We benefit from Convolutional (Conv.), Transposed Convolutional (ConvT.), Rectified Linear Unit (ReLU) [11] and Sigmoid layers in the architecture.

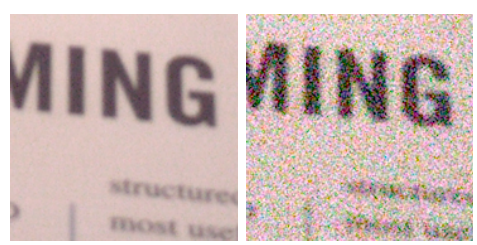
3.1. Network Architecture

The network takes a color image as input with size $3xMxN$ and passes it through eight times convolution and ReLU (Conv+ReLU) layers with one padding applied to keep the input size constant. We refer to these layers as encoder layers. Output of the encoder layers are then passed to the decoder layers which are seven transposed convolution and ReLU (ConvT+ReLU) layers followed by one transposed convolution and sigmoid layer. We use Sigmoid function in output layer to clip the output values of our network between 0 and 1. Like encoder layers, decoder layers also have one padding applied to keep the input size constant. Throughout the network we placed symmetric residual connections which transfers high-level information to deeper layers in order to reduce the effects of degradation problem. Visualization of our architecture can be seen in the figure 2.

We used 3 channel in, 64 channel out convolutional layer as input layer and 64 channel in, 3 channel out Transposed Convolutional layer as output layer. All other layers have 64 channel input and 64 layer output. We also used 3x3 kernel size for convolution and transposed convolution layers throughout the network. We used ADAM optimizer [9] with weight decay rate of 0.05 and learning rate 10^{-4}. At 60% and 90% percent of the training learning rate is multiplied by 10^{-1}. As loss function we used many variations of L1, L2, PSNR, SSIM, MS-SSIM and FSIM [21] functions.

3.2. Training Dataset

We preferred Smartphone Image Denoising Dataset (SIDD) [1] due to its quality and amount of data it provides. SIDD dataset provides high-quality real-world noisy images and their noise free ground truths which are taken with smartphones and DSLR cameras respectively. SIDD dataset has 3 versions which are small, medium and full version. Small version of the dataset contains 160 image pairs and it is approximately 6 GB. Medium version of the dataset contains 320 image pairs and approximately 12 GB. Full version of the dataset contains 12000 image pairs and approximately 450 GB. We used medium version of the SIDD dataset due to its manageable size. As training dataset, we randomly selected 90% of medium version of the dataset. Remaining 10% of the data is used as validation set. Before the training phase we randomly cropped 128×128 patches from images batch size x number of epochs times and saved them for fast training.



(a) Ground-Truth (b) Noisy

Figure 3: Sample cropped patches from SIDD dataset.

3.3. Validation Dataset

As we mentioned in the training dataset section, we used 10% of the SIDD medium version dataset as validation set. We recorded validation loss by using validation dataset at each epoch in order to better understand and optimize the training process of the proposed model. We calculated validation loss as mean loss throughout the validation set for each epoch.

4. Experiments

We focused on removing real world noise from color images. To implement, train and evaluate our proposed method we used PyTorch [12]. All the experiments are conducted in Python 3.8.2 environment using PyTorch 1.6.0 running on a PC with Intel(R) Core(TM) i7-3770K CPU and Nvidia GTX 1080 GPU with 8 GB video memory. The training of a single model can be done in about 3 hours. Some demonstrations of the proposed method are available at: https://yilmazdoga.com/deep_residual_autoencoder_for_real_image_denoising also, the implementation is available at: https://github.com/yilmazdoga/Deep_Residual_Autoencoder_for_Real_Image_Denoising.

4.1. Performance Evaluation Criteria

In order to evaluate the performance of our proposed model, we needed a reliable metric to quantitatively measure the quality of the resulting denoised image. In the literature there are many IQA metrics which can be used for evaluating the quality of denoised images produced by our proposed method. Neural Image Assessment (NIMA) [14] and Deep Image Structure and Texture Similarity (DISTS) [6] can be given as examples of state-of-the-art image quality evaluation metrics. However, these state-of-the-art metrics are not fully adopted by the literature. For the sake of comparability, we are evaluating the performance of our proposed method with traditional metrics which are SSIM and PSNR.

4.2. Test Datasets

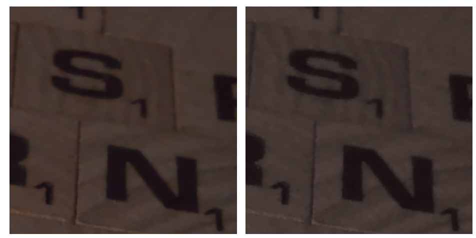
Two test sets are used for evaluating color image denoising performance which are Darmstadt Noise Dataset (DND) and SIDD. DND dataset consists of 50 high-resolution images with realistic image noise. We used noisy images provided by DND as our first test dataset without applying any manipulation. From SIDD dataset we used 256 by 256 image patches from images which are not used in our training or validation sets as our second test dataset.

4.3. Effects of Residual Learning

To test the effects of residual learning, we trained two networks with and without residual connections. Training of these two networks was done using the same dataset which is the medium sized version of SIDD dataset. We observed that that network with residual connections perform better than the network without residual connections. In the table 1 average PSNR and SSIM scores of both models can be found. Also, a sample image denoised with both networks can be found in the figure 4.

Table 1: Average PSNR and SSIM scores of proposed network with and without residual connections.

Method Name	Average PSNR	Average SSIM
With Residual Connections	37.75	0.898
Without Residual Connections	32.68	0.872



(a) (b)

Figure 4: Sample denoised image using proposed model with residual connections (a), without residual connections (b).

4.4. Effects of Different Loss Functions

Table 2: Average PSNR and SSIM scores of proposed network with L1, L2, SSIM, MS-SSIM and L1 + MS-SSIM loss functions.

Loss Function	Average PSNR	Average SSIM
L1	37.93	0.895
L2	37.71	0.891
SSIM	37.25	0.900
MS-SSIM	35.91	0.895
L1 + MS-SSIM	37.75	0.898

Beside optimizing the performance of our network, we also aimed to optimize the loss function we used with our network. With the objective of using a better loss function we trained our model with 5 different loss functions which are L1, L2, SSIM, MS-SSIM, and L1 + MS-SSIM which is proposed by Zhao [22] et al. Performance comparison of all loss functions can be found in table 2. Also, sample images denoised with all loss functions can be found in figure 5.

(a) L2 (b) L1 (c) SSIM (d) MS-SSIM (e) L1 + MS-SSIM

Figure 5: Sample images denoised with proposed network using L1, L2, SSIM, MS-SSIM and L1 + MS-SSIM loss functions.

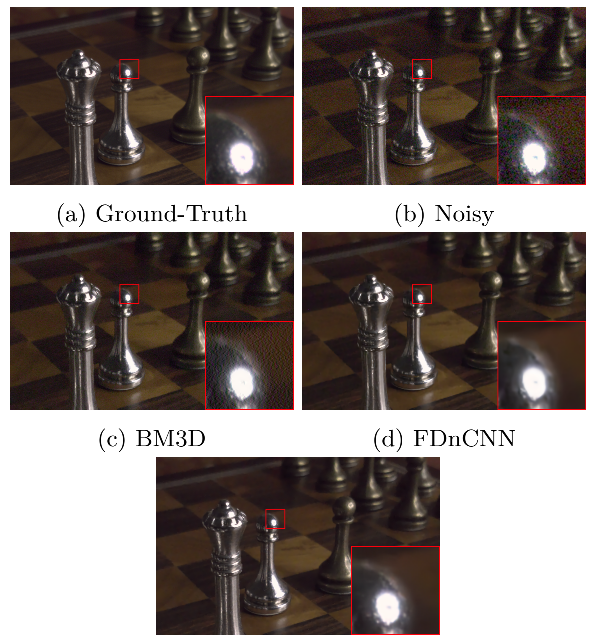
5. Results

In this section we present our results in blind real-noise denoising in color images both quantitatively and qualitatively. We picked our proposed model with L1 + MS-SSIM loss function as the best performer and all of the results shown in this section are obtained using that model. The results are compared with state-of-the-art methods which DND dataset benchmark page provides. We used DND datasets for conducting comparison with state-of-the-art methods.

Table 3: Benchmark results provided by DND dataset webpage.

Method Name	Average PSNR	Average SSIM
GMSNet-B [8]	40.237	0.9616
MIRNet [18]	39.88	0.9563
RIDNet [2]	39.2555	0.9528
Ours	37.9069	0.9391
FFDNet+ [20]	37.6107	0.9415
BM3D [5]	34.51	0.8507
DnCNN [19]	32.4296	0.79
Original Noisy	29.836	0.7018

At table 3 benchmark results of best performer model with the benchmark results of state-of-the-art and traditional methods are given. More benchmark data of other denoising methods are available at DND dataset webpage. From the benchmark data we can say that our proposed method outperforms traditional methods. However, although there are some state-of-the-art methods which our method is similar in terms of performance there are some state-of-the-art methods which outperform our method. Also, when we look at qualitative results are shown in the figure 6 it can be seen that our method outperforms traditional methods and offers similar performance with state-of-the-art methods. All in all, the performance of our proposed method outperforms traditional methods however, it falls behind when it comes to cutting edge of state-of-the-art-methods.



(a) Ground-Truth (b) Noisy (c) BM3D (d) FDnCNN (e) Ours

Figure 6: A and B are high-definition (1920x1200) ground-truth and noisy image pair from SIDD dataset [1]. C through E are sample denoised images using [5], [19] and Our method respectively. (Best viewed on high-resolution display.)

6. Conclusions

In this paper, we propose a deep autoencoder combined with residual connections as a solution to the blind real image denoising problem. We used SIDD [1] to obtain the real-world noisy image data in order for our model to learn real world noise. Results show that our network can outperform traditional methods and also can achieve similar performance with state-of-the-art methods in blind real image denoising. We believe that our proposed model has potential to achieve improved denoising performance by adding feature attention module.

References

[1] Abdelrahman Abdelhamed, Stephen Lin, and Michael S. Brown. A high-quality denoising dataset for smartphone cameras. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018.

[2] Saeed Anwar and Nick Barnes. Real image denoising with feature attention. CoRR, abs/1904.07396, 2019.

[3] 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, pages 2392–2399, 2012.

[4] Kyunghyun Cho. Boltzmann machines and denoising autoencoders for image denoising, 2013.

[5] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space. In 2007 IEEE International Conference on Image Processing, volume 1, pages I – 313–I – 316, 2007.

[6] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P. Simoncelli. Image quality assessment: Unifying structure and texture similarity, 2020.

[7] L. Gondara. Medical image denoising using convolutional denoising autoencoders. In 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), pages 241–246, 2016.

[8] Dong-Wook Kim, Jae Ryun Chung, and Seung-Won Jung. Grdn:grouped residual dense network for real image denoising and gan-based real-world noise modeling, 2019.

[9] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.

[10] Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang. Image restoration using very deep fully convolutional encoder-decoder networks with symmetric skip connections. CoRR, abs/1603.09056, 2016.

[11] Vinod Nair and Geoffrey Hinton. Rectified linear units improve restricted boltzmann machines vinod nair. volume 27, pages 807–814, 06 2010.

[12] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. In NIPS 2017 Workshop on Autodiff, 2017.

[13] Tobias Plötz and Stefan Roth. Benchmarking denoising algorithms with real photographs, 2017.

[14] Hossein Talebi and Peyman Milanfar. Nima: Neural image assessment. IEEE Transactions on Image Processing, 27(8):3998–4011, Aug 2018.

[15] Zhou Wang, Alan Bovik, Hamid Sheikh, and Eero Simoncelli. Image quality assessment: From error visibility to structural similarity. Image Processing, IEEE Transactions on, 13:600 – 612, 05 2004.

[16] Junyuan Xie, Linli Xu, and Enhong Chen. Image denoising and inpainting with deep neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, Advances in Neural Information Processing Systems 25, pages 341–349. Curran Associates, Inc., 2012.

[17] X. Ye, L. Wang, H. Xing, and L. Huang. Denoising hybrid noises in image with stacked autoencoder. In 2015 IEEE International Conference on Information and Automation, pages 2720–2724, 2015.

[18] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement, 2020.

[19] 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, 26(7):3142–3155, 2017.

[20] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for CNN based image denoising. CoRR, abs/1710.04026, 2017.

[21] L. Zhang, L. Zhang, X. Mou, and D. Zhang. Fsim: A feature similarity index for image quality assessment. IEEE Transactions on Image Processing, 20(8):2378–2386, 2011.

[22] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. IEEE Transactions on Computational Imaging, PP:1–1, 12 2016.

[23] H. Zhao, O. Gallo, I. Frosio, and J. Kautz. Loss functions for image restoration with neural networks. IEEE Transactions on Computational Imaging, 3(1):47–57, 2017.