Parallel Feature Pyramid Network for Image Denoising

School of Electrical Engineering
Korea University
Seoul, Korea
Email: {sjcho, khuhm, swkim, swji}@dali.korea.ac.kr, sjko@korea.ac.kr

Abstract—Image denoising is a classical and essential task in consumer electronics equipped with cameras. Recently, the convolutional neural network (CNN)-based denoising methods have been widely studied. These methods adopt single-scale features to separate image structures from the noisy observation. Single-scale features, however, have limitation in covering the full characteristics of image structures at different scales. In this paper, we propose a novel denoising network that makes use of the multi-scale feature pyramid where each feature map represents the characteristics of image structure at different scales. We then combine these multi-scale features to obtain the contextual information and utilize it to effectively generate clear denoised results. Experimental results show that our network achieves superior performance to other conventional methods.

I. INTRODUCTION

Image denoising is an indispensable topic in image processing. Since noise corruption is inevitable during image acquisition, image denoising has been employed to the image signal processing module of cameras. Over the past decades, various model-based image denoising methods have been proposed, such as block-matching and 3D filtering (BM3D) [1] and weighted nuclear norm minimization (WNNM) [2].

With the recent development of the convolutional neural networks (CNNs), the denoising performance has been significantly advanced. Based on CNN, Zhang et al. proposed a denoising CNN (DnCNN) [3] which adopts batch normalization (BN) [4] and a residual learning strategy and achieves state-of-the-art performance. However, DnCNN uses single-scale features that is insufficient to cover full characteristics of image structures at different scales. As a result, DnCNN often produces visual artifacts in the resultant image.

In this paper, we present a multi-scale denoising network adopting the feature pyramid generated by gradually reducing the size of the feature maps. We then aggregate the contextual information in these multi-scale feature maps to effectively obtain clear denoised results. Experimental results demonstrate that the proposed network achieves outstanding performance as compared to the conventional methods.

II. PROPOSED METHOD

Fig. 1 shows an overall architecture of the proposed network. Following DnCNN, we adopt residual learning using global skip connection to predict the noise. In this figure, the “Convs” block is composed of two 3 × 3 convolution layers except for the first one, which is composed of three consecutive 3 × 3 convolution layers. Each convolution layer is followed by the BN [4] and rectified linear units (ReLU). To extract multi-scale contextual information from the noisy input, we first construct a feature pyramid which consists of feature maps with different sizes. Let the first Convs block produce a feature map with a size of \( W \times H \). Then, we employ the spatial pyramid pooling (SPP) [5] to form a feature pyramid, \( F = \{ f_0, f_1, f_2 \} \), where \( f_n \), the \( n \)th level feature pyramid of \( F \), is the feature map with a spatial size of \( W/2^n \times H/2^n \). Each level of \( F \) contains the information of its corresponding scale of the input.

To utilize contextual information at different scales, we use the multi-scale context aggregation (MSCA) [6] module. As shown in Fig. 1, the MSCA module produces the levels of an output feature pyramid, \( G = \{ g_0, g_1, g_2 \} \). For example, consider that we generate the \( n \)th level of \( G \), \( g_n \), and have the feature maps of \( F \) with \( C \) output channels. We apply the Convs block to each pyramid level of \( F \) in parallel to extract the appropriate feature at each scale. For the \( n \)th level feature map, the Convs block generate a feature map with \( C/2 \) channels; for the other levels, that is the Convs block produce a feature map with \( C/4 \) channels. We then match the size of the feature maps to \( W/2^n \times H/2^n \) by using up-sampling and down-sampling operations. Finally, the feature maps are concatenated and then fed into another Convs block to obtain \( g_n \).

We assume that \( f_0 \) having the same spatial resolution as the input contains primary information for separating the noise from the noisy input. Therefore, we adjust the filter number of the convolution layer in the last Convs block of MSCA module so that the number of output channels of \( g_0 \) is \( C/2 \) and that of \( g_1 \) and \( g_2 \) is \( C/4 \). We up-sample \( g_1 \) and \( g_2 \) to match the size of \( g_0 \) and concatenate the feature maps to form a single feature map having \( C \) channels. The last Convs block is then applied to produce the residual image which only contains the estimated noise.

III. EXPERIMENTAL RESULT

We evaluated the proposed network on Berkeley segmentation dataset (BSD) 68 and color BSD (CBSD) 68. The peak signal-to-noise ratio (PSNR) was used for evaluating the denoising performance.
Fig. 1. The architecture of the proposed network

For fair comparison, we trained the DnCNN and our network on the same data set. We used 4K and 200 images from the training set of the Waterloo Exploration dataset (WEDS) and BSD, respectively. We added Gaussian noise of a certain sigma to the images and train the networks at that noise level. From each image, we randomly selected 20 different patches with a size of $64 \times 64$. With the batch size of 64, our network was trained for 150 epochs via the ADAM optimizer with the weight decay of $10^{-3}$. The learning rate started from $10^{-3}$ and decreased by a factor of 0.98 at each epoch.

Fig. 2 shows the denoising results of a sample image from CBSD68 with the noise level of 50. The artifacts seen in the results of DnCNN are not observed in the proposed method. In this example, our model achieves 0.5 dB higher PSNR than DnCNN.

Table I lists the average PSNR results on BSD68 and CBSD68, respectively. As can be seen in the table, our network outperforms other conventional methods on the datasets. Although we used fewer learning parameters than DnCNN (485K vs. 558K), the performance of our network is superior to that of DnCNN. In particular, at the noise level of 50, the proposed method boosts the PSNR result by about 0.39dB.

IV. CONCLUSION

In this paper, we proposed a novel denoising network which exploits multi-scale feature pyramid. We adopt the spatial pyramid pooling operator and generate the feature pyramid where each feature of the pyramid contains the information of different scales. We then combine these multi-scale features to obtain the contextual information and utilize it to effectively generate clear denoised results. Experimental results show that our network achieves superior performance to other conventional methods.

ACKNOWLEDGMENT

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2014-3-00077, Development of global multi-target tracking and event prediction techniques based on real-time large-scale analysis).

REFERENCES