Abstract

During the past years, deep convolutional neural networks have achieved impressive success in image denoising. In this paper, we propose a densely self-guided wavelet network (DSWN) for real world image denoising. The basic structure of DSWN is a top-down self_guidance architecture which is able to efficiently incorporate multi-scale information and extract good local features to recover clean images. Moreover, such a structure requires a smaller number of parameters and enables us to achieve better effectiveness than Unet structure. To avoid information loss and achieve a better receptive field size, we embed wavelet transform into DSWN. In addition, we apply densely residual learning to convolution blocks to enhance the feature extraction capability of the proposed network. At the full resolution level of DSWN, we adopt a double branch structure to generate the final output. One branch of them tends to pay attention to dark areas and the other performs better on bright areas. Such a double branch strategy is able to handle the noise at different exposures. The proposed network is validated by BSD68, Kodak24 and SIDD+ benchmark. Additional experimental results show that the proposed network outperforms most state-of-the-art image denoising solutions.

1. Introduction

Image denoising is a fundamental task in low-level vision and an important pre-processing step in many other vision tasks. Traditional methods [1] usually address image denoising by domain transform [2], non-local algorithm [3], Markov Random Fields (MRFs) [4], etc. However, these methods need to manually set parameters and refer a complex optimization problem for the testing stage. With the rapid development of deep learning technology, numerous advanced approaches [5-7] have been developed and achieve impressive success. By involving the strategy of residual learning and adding batch normalization (BN) and ReLU activate function into deep architectures, DnCNN [8] is proposed to handle Gaussian blind denoising and achieves a much higher peak signal to noise ratio (PSNR) than conventional state-of-the-art approaches [9]. For the pursuit of highly accurate denoising results, some follow-up works have been proposed to remove additive white Gaussian noise (AWGN) [10, 11]. To involve multi-scale information, some most advanced end-to-end methods [12-14] apply Unet [15] as their basic structure and add some dense residual block in each level. Although these methods obtain competitive performance on benchmark datasets, their heavy computation and memory footprint hinder their application.

To seek a better trade-off between denoising performance and the consumption of computational resources, self-guided neural network (SGN) [16] is proposed for image denoising task by a top-down guidance strategy. SGN generates multi-resolution inputs with the PixelUnShuffle [17] before any convolutional operation. DSWN (as shown in Figure 2) which is able to improve performance of SGN (Figure 1) and require...
Incorporation of multi-scale information

To extract multi-scale information for image denoising and single image super resolution (SISR) tasks,PixelShuffle and wavelet transform have been proposed to replace pooling and interpolation to avoid information loss. With self-guidance strategy and PixelShuffle, SGN [16] greatly improved the memory and runtime efficiency. Bae et al. proposed a wavelet residual network (WavResNet) [22] for image denoising and SISR and find wavelet subbands benefits learning convolutional neural network (CNN). Similarly, a deep wavelet super-resolution (DWSR) method [23] is propose to recover missing details on subbands. Both WavResNet and DWSR only consider single level wavelet decomposition. Multi-level wavelet transform is considered by MWCNN [24] to achieve better receptive field size and avoid down-sampling information loss by embedding wavelet transform into CNN architecture. MWCNN owns more power to model both spatial context and inter-subband dependency by embedding DWT and IDWT to CNN. In this paper, our proposed network adopts the same method as MWCNN to incorporate multi-scale information with a totally different architecture from MWCNN.

3. Densely Self-guided Wavelet Network (DSWN)

In this section, we firstly introduce the overall network structure and then introduce the details of DSWN.

3.1. Overall Structure of DSWN

Our proposed denoising network is shown in Figure 2. A top down self-guidance architecture is used to better exploit image multi-scale information. Information extracted at low resolution is gradually propagated into the higher resolution sub-networks to guide the feature extraction processes. Instead of PixelShuffle and PixelUnShuffle, DWT and IDWT are used to generate multi-scale inputs. Before any convolution operation, DSWN uses wavelet transform to transform the input image to three smaller...
scales. At the full resolution layer, we adopt a double branch structure consists of a residual learning branch and an end-to-end learning branch. In the rest of this paper, we simply refer to these two branches as residual branch and end2end branch. For our network, we observed that the residual branch focuses on bright areas and the end2end branch focuses on dark areas. Therefore, we use both of these two branches at the full resolution level to further improve the performance, especially when the network needs to work on noisy images with different ISO at the same time. In addition, we find batch normalization is harmful for the denoising performance and do not use any normalization layer in this network. For each level, we add densely connected residual (DCR) [13] one or two blocks as shown in Figure 3.

3.2. Detail Structure of DSWN

The top level of DSWN works on the smallest spatial resolution to extract large scale information. The top sub-network contains two Conv+PReLU layers (the orange box in Figure 2) and a DCR block (the red box in Figure 2). The DCR block simultaneously applies dense connectivity and residual learning to remove the noise of input images accurately and solve the vanishing-gradient problem.

At the middle two levels, 1×1 Convolutional kernel layers are used to merge information extracted from different resolution. The network structure of the middle sub-networks is similar to the structure of the top sub-network. As for the full resolution level, we add more DCR blocks with skip connections to enhance the feature extraction capability of DSWN after merging information from all the scales. For the residual branch, DSWN has a global residual connection between the input image and the final estimation. We add a Tanh activation function at the end of the end2end branch. The final output is the simple average result of the residual branch and the end2end branch. By adding gradient loss, our network is able to achieve better retention of details without reducing PSNR. In order to ensure the fairness of the comparative experiment, in the experimental part of this paper, our network only uses L1 loss for training.

4. Experimental Results

In this section, we first introduce the training details and then provide experimental results on different datasets. We compare DSWN with several state-of-the-art denoising approaches.

4.1. Experimental Setting

DIV2K [25] training and validation datasets provide an adequate amount of high quality images. Most state-of-the-art image denoising solutions select the DIV2K as their training dataset [12, 16]. To train our DSWN model, we
Figure 4: Denoising results of the conventional methods and the proposed method on BSD68 dataset ($\sigma = 50$).

Figure 5: Denoising results of the conventional methods and the proposed method on BSD68 dataset ($\sigma = 30$).

Table 1: RGB image denoising results on BSD68 dataset.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Noise Level</td>
<td></td>
<td>PSNR (dB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma = 10$</td>
<td>28.30</td>
<td>35.89</td>
<td>36.12</td>
<td>36.14</td>
<td>36.06</td>
<td>35.97</td>
<td>36.45</td>
<td>36.91</td>
</tr>
<tr>
<td>$\sigma = 30$</td>
<td>19.03</td>
<td>29.72</td>
<td>30.32</td>
<td>30.31</td>
<td>30.22</td>
<td>30.36</td>
<td>30.41</td>
<td>30.72</td>
</tr>
<tr>
<td>$\sigma = 50$</td>
<td>14.91</td>
<td>27.36</td>
<td>27.92</td>
<td>26.96</td>
<td>27.86</td>
<td>28.01</td>
<td>28.02</td>
<td>28.29</td>
</tr>
</tbody>
</table>

Table 1: RGB image denoising results on BSD68 dataset.

Also use DIV2K dataset as training and validation dataset for AWGN task. The training dataset consists of 800 high quality images. The resolution of each of these images is $1920 \times 1080$. The DIV2K validation dataset consists of 100 images; the quality of each image is similar to that of the training dataset. For the testing datasets, we use the BSD68 [28] and dataset Kodak dataset [29] which are used by some recent denoising networks [30]. The Kodak dataset consists of 24 images, each of which has a resolution of $768 \times 512$. The BSD68 dataset consists of 68 images, each of which has a resolution of $321 \times 481$. As for real noise removal task, we conduct experiments on Smartphone Image Denoising Dataset (SIDD) benchmark [31]. The SIDD training dataset consists of 320 images from 10 scenes under different lighting conditions using five representative smartphone cameras. The test dataset is SIDD+ dataset [32] which is generated by NTIRE 2020 Real Image Denoising Challenge with a similar procedure as the SIDD benchmark. Compared with the SIDD validation dataset, the SIDD+ dataset contains more details.
Table 2: RGB image denoising results on Kodak24 dataset.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Noise Level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ = 10</td>
<td>28.24</td>
<td>36.57</td>
<td>36.58</td>
<td>36.80</td>
<td>36.70</td>
<td>36.64</td>
<td>37.33</td>
<td>37.35</td>
</tr>
<tr>
<td>σ = 30</td>
<td>18.93</td>
<td>30.89</td>
<td>31.28</td>
<td>31.39</td>
<td>31.24</td>
<td>31.70</td>
<td>31.95</td>
<td>33.08</td>
</tr>
<tr>
<td>σ = 50</td>
<td>14.87</td>
<td>28.62</td>
<td>28.94</td>
<td>29.10</td>
<td>29.92</td>
<td>29.42</td>
<td>29.67</td>
<td>29.97</td>
</tr>
</tbody>
</table>

| Noise Level | PSNR (dB) | SSIM           |            |             |             |           |           |       |
| σ = 10   | 0.6607    | 0.9432         | 0.9446     | 0.9462      | 0.9448      | 0.9455    | 0.9508    | 0.9488 |
| σ = 30   | 0.2755    | 0.8459         | 0.8579     | 0.8596      | 0.8581      | 0.8599    | 0.8736    | 0.8794 |
| σ = 50   | 0.1557    | 0.7772         | 0.7915     | 0.7949      | 0.7939      | 0.7949    | 0.8160    | 0.8241 |

When training our model, we randomly crop 256 × 256 patches from the training images. The input patches of the proposed network are randomly flipped and rotated for data augmentation. The parameters of network are Xavier initialized [33]. We train the whole network for 300 epochs overall. The learning rate is initialized as 1e−4 at the first 200 epochs and reduce to 5e−5 in the next 50 epochs. We finetune our model at the last 50 epochs with a 1e−5 learning rate. For optimization, we use Adam optimizer [34] with β1 = 0.5, β2 = 0.999 and batch size equals to 1. We use L1 loss which is a PSNR-oriented optimization in the training process [35]. The experiments are implemented using a NVIDIA RTX 2080Ti GPU.

4.2. Performance comparison

We compare our proposed network with several state-of-the-art image denoising solutions: CBM3D [9], DnCNN [8], FFDNet [19], IRCNN [27], SGN [16] and DHDN [13], where DHDN is more complex than DSWN. To compare the performance, we determined the peak-signal-noise-ratio (PSNR) [36] and structural similarity (SSIM) [37] as the objective measurements. Table 1 lists the average PSNR of the compared methods and the proposed method for sRGB images in BSD68 dataset. Figure 4 and Figure 5 show some example results in BSD68 dataset with different noise levels. From Table 1, we can see that our DSWN shows the best PSNR and SSIM in all the noise levels. DHDN is the second best method. It adopts twice as many DCR blocks as DSWN and is harder to train. Although DHDN adopts more DCR block in its network architecture, DSWN is still able to achieve a PSNR which is about 1.37 higher than DHDN on average. Figure 4 and Figure 5 show some detail results of DnCNN, SGN and DSWN, where DnCNN is a classical denoising network and SGN has a similar structure to DSWN. We can see all the three denoising networks are able to achieve a obvious improvement compared with noisy images. DSWN is better than DnCNN and SGN in some details such as the texture details of the statue in Figure 4 and the beard of the tiger in Figure 5. Our proposed method is able to handle different noise levels and reserve more details at the same time.
Table 2 shows DSWN outperforms the compared methods in most cases. From Figure 6 and Figure 7, we can conclude a similar conclusion to the BSD68 dataset. At a higher noise level, denoising networks tend to smooth the noise image too much, because the network is difficult to distinguish true details from noise. DSWN can better preserve details at high noise levels, such as eyelashes and window textures.

4.3. NTIRE 2020 real image denoising challenge

The proposed method is initially proposed to participate in NTIRE 2020 real image denoising challenge [32]. The purpose of the challenge is to remove unspecified noise from images. The training dataset is SIDD benchmark which includes images with various noise levels, dynamic ranges, and brightness. Such a noise model is more complex than the AWGN task. During the challenge, we tried several different solutions on SIDD benchmark dataset and finally fix DSWN as our final solution. Our network achieves a competitive PSNR on the testing dataset and is very efficient among all the solutions for this challenge.

In this section, we first present denoising results generated by DSWN and some compared denoising networks. Then, we conduct some ablation studies on different network structures. In the section, we evaluate the proposed method on SIDD+ validation dataset which provided by the organizers of NTIRE 2020 real image denoising challenge.

4.3.1 Performance comparison

We compare our proposed network with several state-of-the-art image denoising solutions in our SIDD+ dataset which includes images with various noise levels, dynamic ranges, and brightness. Such a noise model is more complex than the AWGN task. During the challenge, we tried several different solutions on SIDD benchmark dataset and finally fix DSWN as our final solution. Our network achieves a competitive PSNR on the testing dataset and is very efficient among all the solutions for this challenge.

Table 3: Denoising results on SIDD+ validation dataset.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR (dB)</td>
<td>26.63</td>
<td>34.03</td>
<td>34.01</td>
<td>27.35</td>
<td>35.87</td>
<td>32.28</td>
<td>35.51</td>
<td><strong>36.94</strong></td>
</tr>
<tr>
<td>SSIM</td>
<td>0.6622</td>
<td>0.9139</td>
<td>0.9192</td>
<td>0.7331</td>
<td>0.9407</td>
<td>0.8669</td>
<td>0.9411</td>
<td><strong>0.9574</strong></td>
</tr>
</tbody>
</table>
Figure 9: Denoising results of the conventional methods and DSWN on SIDD+ validation dataset.

Figure 9 shows example results of the compared methods and DSWN. The denoised result of DSWN is better than other methods, especially at the edges of the hat, hair and eyebrows. Owing to the real images denoising task is more difficult than AWGN task, some of the compared methods only slightly improved the performance.

In this challenge, we average eight output images of eight input images; these are generated by a combination of a flip and rotation of an input image. Using such an ensemble strategy, the PSNR of our method is able to be further improved by 0.1.

4.3.2 Ablation Study

Table 4: Ablation study on two branches of DSWN on SIDD+ validation dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>Residual branch only</th>
<th>End2End branch only</th>
<th>Two Residual branches</th>
<th>DSWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR (dB)</td>
<td>36.67</td>
<td>36.44</td>
<td>36.86</td>
<td>36.94</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.9515</td>
<td>0.9466</td>
<td>0.9533</td>
<td>0.9574</td>
</tr>
</tbody>
</table>

At the full resolution level of DSWN, we adopt a double branch structure: a residual branch and an end2end branch. At first, we tried such a scheme as an ensemble strategy. However, we observe an interesting phenomenon: the residual branch tends to ignore the dark areas and the end2end branch tends to highlight the dark details. As a result, these two branches collaborate to generate output images with appropriate colors as shown in Figure 10. To prove the effectiveness of the double branch structure, we conduct additional experiments as shown in Table 4. We can see DSWN is able to achieve better PSNR and SSIM than using any branch only. If we replace the end2end branch by another residual branch, our DSWN is still better. This demonstrates that the improvement is not only from the increase of network parameters. In addition, the results from the same two residual branches are similar (Figure 11).
In addition, our proposed the main structure of DSWN is inspired by SGN. The most obvious structural difference between the two is the incorporation of multi-scale information. DSWN adopt DWT and IDWT to replace PixelShuffle and PixelUnShuffle in SGN. Both of them are no loss of information. In Table 5, we show the ablation study of these two methods. The results suggest that wavelet transform is better than PixelShuffle strategy. This might be because the frequency domain is more suitable for denoising and wavelet transform is able to enlarge receptive field.

Table 5: Ablation study on down / up sampling method of DSWN on SIDD+ validation dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>PixelShuffle / PixelUnShuffle</th>
<th>DWT/IDWT</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR (dB)</td>
<td>36.59</td>
<td>36.94</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.9548</td>
<td>0.9571</td>
</tr>
</tbody>
</table>

5. Conclusion

In this paper, we proposed a densely self-guided wavelet network (DSWN) for image denoising. DSWN adopts a top-down manner to denoise images. Wavelet transform is adopted to generate input variations with different spatial resolutions before any convolutional operation. Then, we embed a DCR block into three low spatial resolution levels, respectively. At the full resolution level, we employ more DCR blocks and a double branch structure to further improve the quality of output images. The proposed DSWN was validated on AWGN task and real world image denoising benchmark and demonstrated excellent efficiency in NTIRE 2020 real image denoising challenge. DSWN is able to generate higher quality denoising results than the compared state-of-the-art methods.

References


