Supplementary Material of NTGAN: Learning Blind Image Denoising without Clean Reference

Rui Zhao
rick10.zhao@connect.polyu.hk
Daniel P.K. Lun
enpklun@polyu.edu.hk
Kin-Man Lam
enkmlam@polyu.edu.hk

Department of Electronic and Information Engineering,
The Hong Kong Polytechnic University,
Hong Kong, China

1 Statistics consistency analysis on the noise

In the paper, we adopt the noise generation model in [8] to synthesize the target observation $z$ from a real-world image $y$ as follows:

$$z = y + n$$

$$n(y, M_z) = f_{BPD}(f_{crf}(L + n_s(L) + n_c)) - f_{BPD}(f_{crf}(L)),$$

(1)

where $f_{crf}$ and $f_{icrf}$ represent the camera response function and the inverse camera response function, respectively. $n_s$ and $n_c$ account for the noise components that are dependent and independent of the signal $y$, respectively. We assume that the noise residing in the real image $y$ is heterogeneous Gaussian distributed, with the dependency on the clean signal $x$. Therefore, in the irradiance plane, $L_y$ can be represented as follows:

$$L_y = n_s(L_x) + n_c,$$

(2)

where $L_y = f_{icrf}(y)$ and $L_x = f_{icrf}(x)$.

In addition, the synthetic observation $z$ in the irradiance plane can be approximated as:

$$L_z = L_y + n_s(L_y) + n_c \approx L_x + n_s'(L_x) + n_c',$$

(3)

where $n_s'$ denotes the signal-dependent component in $z$ with respect to the clean signal, and $n_c'$ denotes the signal-independent component, combining the signal-independent noise from Eqns.(1) and (2). Eqn. (3) indicates that the resultant noise in $z$ also follows a heterogeneous Gaussian distribution. Therefore, the adopted noise generation model guarantees the consistency of the noise distribution in the source and target observations.

© 2020. The copyright of this document resides with its authors. It may be distributed unchanged freely in print or electronic forms.
2 Evaluation on synthetic AWGN

To further evaluate the denoising performance of NTGAN, we apply it to the synthetic additive white Gaussian noise (AWGN), and compare its performance with the other state-of-the-art denoisers, including BM3D [3], WNNM [4], DnCNN [15], FFDNet [14], RIDNet [1], N2N [7], DIP [13], N2S [2], and N2V [6]. To make a fair comparison, we change the original noise generation model in the paper, and make it to generate AWGN for producing paired samples in training. Specifically, the training pairs are generated as:

\[ y = x + n_y, \quad z = x + n_z, \]

where both \( \sigma_y \) and \( \sigma_z \) are uniformly sampled from the range \([0, 75/255]\), and \( x \) represents the clean signal. In other words, the training pairs are obtained by independently adding two small AWGNs to the clean image. We collect all the 4,744 images from the Waterloo Exploration database [9] to form the training set, and use Eqn. (4) to generate the training pairs. We follow the same settings mentioned in the paper to train up NTGAN and evaluate it on BSD68 [11] with the noise level set to 15, 25, and 50, respectively. The results are summarized in Table 1. It can be observed that NTGAN achieves comparable or even better results, compared to the supervised non-blind denoisers, i.e., DnCNN and FFDNet. However, NTGAN performs slightly worse than RIDNet when facing AWGN. The reason is that the distribution of AWGN is much easier to be learned, and thus the proposed noise transfer strategy loses its advantage by serving as an augmentation approach for learning complicated noise distributions. NTGAN achieves about 31.72dB, 29.28dB, and 26.37dB with the noise level set to 15, 25, and 50, respectively, which basically outperforms the other unsupervised deep denoisers by a large margin.

<table>
<thead>
<tr>
<th>Type</th>
<th>Traditional methods</th>
<th>Supervised CNNs</th>
<th>Unsupervised CNNs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>BM3D</td>
<td>WNNM</td>
<td>DnCNN</td>
</tr>
<tr>
<td>( \sigma = 15 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSNR</td>
<td>31.08</td>
<td>31.32</td>
<td>31.73</td>
</tr>
<tr>
<td>( \sigma = 25 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSNR</td>
<td>28.57</td>
<td>28.83</td>
<td>29.23</td>
</tr>
<tr>
<td>( \sigma = 50 )</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: The quantitative results on the grayscale BSD68 images corrupted with AWGN. The best results are highlighted in **bold.**

3 Qualitative comparison on DND

Here, we present the qualitative results on DND [10] in Figs. 1 and 2 to show the perceptual quality of the restored images based on NTGAN. It is clear from the figures that NTGAN acquires a better ability in recovering the texture and edge details. The traditional methods, e.g., BM3D [3], WNNM [4], and NI [11], still retain some noise in the restored images. The supervised denoisers, e.g., DnCNN+ [15], RIDNet [1], and VDNet [14], oversmooth the images. CBDNet [5] produces the images with visible artefacts on the edges. The proposed NTGAN can recover the image with fewer artefacts and more details.
Figure 1: Evaluation on the perceptual qualities of a real-world noisy image from DND, restored by different methods. BM3D: 23.95dB, WNNM: 25.63dB, NI: 27.28dB, DnCNN+: 32.26dB, CBDNet: 31.40dB, RIDNet: 34.30dB, VDNet: 34.08dB, NTGAN: 35.10dB.

Figure 2: Evaluation on the perceptual qualities of a real-world noisy image from DND, restored by different methods. BM3D: 30.91dB, WNNM: 30.94dB, NI: 32.23dB, DnCNN+: 33.29dB, CBDNet: 33.62dB, RIDNet: 34.09dB, VDNet: 33.89dB, NTGAN: 34.37dB.
References


