Learning to Jointly Generate and Separate Reflections
Supplementary Material

Daiqian Ma$^{1,2}$, Renjie Wan$^3$, Boxin Shi$^{2,4}$, Alex C. Kot$^3$, Ling-Yu Duan$^{1,2,4,*}$

The SECE of Shenzhen Graduate School, Peking University, Shenzhen, China$^1$
The National Engineering Lab for Video Technology, Peking University, Beijing, China$^2$
School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore$^3$
The Peng Cheng Laboratory, Shenzhen, China$^4$

{madaiqian, shiboxin, lingyu}@pku.edu.cn, {rjwan,eackot}@ntu.edu.sg

Contents:

1 - Network Structure
2 - Dataset Gap
3 - Additional Results
4 - Reference
The comparisons between residual-blocks structure and U-Net structure

<table>
<thead>
<tr>
<th>Architecture</th>
<th>$\text{SSIM}_r$</th>
<th>SSIM</th>
<th>PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>U-Net</td>
<td>0.851</td>
<td>0.886</td>
<td>24.35</td>
</tr>
<tr>
<td>Res-blocks</td>
<td>0.858</td>
<td>0.892</td>
<td>24.32</td>
</tr>
</tbody>
</table>

As shown above, the performance of these two structures are very close. We apply residual-blocks structure as the network architecture because it is more flexible to incorporate the entanglement and disentanglement mechanism.
1 - Network Structure

The details of the separator and generator
The setup of data collection in our experiment.
We evaluate the generalization ability of each method to cross the domain gap.

The domain gap indicates the differences between datasets. It may cause noticeable performance drop when training and testing on datasets from different domains. For example, as shown in this page, the special glass coating in the dataset of Zhang18 [8] makes the behavior of their dataset fairly different from our proposed dataset and SIR² dataset [5]. A network model trained on the dataset suitable for Zhang18 may not generalize well on our proposed dataset.
The performances of pre-trained models released by Zhang18 and Wan18 on SIR$^2$ and Zhang18 datasets.

<table>
<thead>
<tr>
<th>Model</th>
<th>SIR$^2$ dataset</th>
<th>Zhang18 dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhang18</td>
<td>0.860/23.13</td>
<td>0.821/21.30</td>
</tr>
<tr>
<td>Wan18</td>
<td>0.895/24.03</td>
<td>0.741/19.22</td>
</tr>
</tbody>
</table>

It is observed that both two models only perform well on one dataset but worse on the other dataset. The domain gap between these two datasets is obvious and we have observed similar phenomenon in the following experiment.
Due to the lack of reflection images in the dataset of Zhang18 [8], it is difficult to train the methods like Wan18 [8] and our method on this dataset. For fair comparison, we train all the learning-based methods on our training dataset and evaluate their performances on the dataset of Zhang18 [8] to see whether they can generalize well on different datasets.

It is observed that our method can achieve consistent performances on the dataset of Zhang18 [8], but the performances of Zhang18 [8] trained on our training dataset are lower than the performances reported in their paper (SSIM: 0.821, PSNR: 21.30). It shows that our method has better generalization ability to cross the domain gap. However, as we mentioned in the previous page, the dataset gap is a challenging yet important problem and we will continue working on this part to improve the generalization ability in our future work.

<table>
<thead>
<tr>
<th>Method</th>
<th>SSIM</th>
<th>PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>LB14</td>
<td>0.752</td>
<td>19.05</td>
</tr>
<tr>
<td>WS16</td>
<td>0.640</td>
<td>17.92</td>
</tr>
<tr>
<td>NR17</td>
<td>0.650</td>
<td>18.41</td>
</tr>
<tr>
<td>FY17</td>
<td>0.757</td>
<td>19.10</td>
</tr>
<tr>
<td>CycleGAN</td>
<td>0.683</td>
<td>18.22</td>
</tr>
<tr>
<td>Zhang18</td>
<td>0.780</td>
<td>19.69</td>
</tr>
<tr>
<td>Wan18</td>
<td>0.784</td>
<td>19.54</td>
</tr>
<tr>
<td>Eq. (1)</td>
<td>0.765</td>
<td>19.48</td>
</tr>
<tr>
<td>Ours</td>
<td>0.791</td>
<td>19.92</td>
</tr>
<tr>
<td>Ours + Eq. (1)</td>
<td><strong>0.798</strong></td>
<td><strong>20.03</strong></td>
</tr>
</tbody>
</table>
3 - Additional Results from the Generator
Examples of mixture images compared with [10]
3 - Additional Results on SIR$^2$ dataset
3 - Additional Results on SIR² dataset
3 - Additional Results on SIR$^2$ dataset
3 - Additional Results on $\text{SIR}^2$ dataset
3 - Additional Results on SIR$^2$ dataset

[Images showing different results for input images, ground truth, and various methods such as FY17, NR17, Zhang18, Wan18, and Ours.]
3 - Additional Results on SIR$^2$ dataset
3 - Additional Results on FY17 dataset
3 - Additional Results on FY17 dataset
3 - Additional Results on Internet Images
3 - Additional Results on Internet Images

- Input
- Zhang18
- Wan18
- Ours
3 - Additional Results on the High-resolution Images

Input
3 - Additional Results on the High-resolution Images

Ours (R)
Comparing the above result with the low-resolution result of Figure 9 in the main paper, we can find that though our method recovers the reflection part, the background is harder to recover clearly, which leaves more space for further exploration.
4 - Reference