1 Mean Opinion Ranking

This section elaborates on the mean opinion ranking (MOR) evaluation conducted on the super-resolved images from five different methods included in our comparison. To simplify the ranking process, we have excluded images upsampled with bicubic interpolation and the pre-trained ESRGAN. The MOR provides a direct measure of the image quality as perceived by humans. We compute the MOR by asking the participants in the study to rank the super-resolved images in terms of overall image quality, where 1 and 5 correspond to best and worst image quality, respectively. For easier judgement of the image quality, we use cropped patches from each image, as the fine details of the images would otherwise be lost due to re-scaling in the survey framework. The survey consists of 10 questions, where the participants are shown the results of all methods for a given image at a time. To avoid bias, we randomly shuffle the order of the presented images. We include 5 real low-resolution images from both the Cityscapes [1] and IDD [10] datasets in our survey. All images must be given a unique rank by the participant. We compute the final MOR by averaging the assigned ranks for each method over all images and participants for the two datasets. In total 20 persons participated in the survey. The MOR results can be seen in Table 1. The images used for each question in the survey can be seen below.

\[
\begin{array}{lcc}
\text{Method} & \text{Cityscapes MOR} & \text{IDD MOR} \\
MZSR \ [9] & 3.33 & 2.96 \\
DPSR \ [13] & 4.41 & 3.16 \\
RealSR \ [2] & 2.75 & 4.88 \\
DAN \ [5] & 3.47 & 2.48 \\
Ours & \textbf{1.21} & \textbf{1.45} \\
\end{array}
\]

Table 1: Results of the MOR of super-resolved real low-resolution images from the Cityscapes and IDD datasets. Lower values are better.
Figure 1: Question 1: Image patch from the CityScapes dataset.

Figure 2: Question 2: Image patch from the CityScapes dataset.

Figure 3: Question 3: Image patch from the CityScapes dataset.
Figure 4: Question 4: Image patch from the CityScapes dataset.

Figure 5: Question 5: Image patch from the CityScapes dataset.

Figure 6: Question 6: Image patch from the IDD dataset.

Figure 7: Question 7: Image patch from the IDD dataset.
Figure 8: Question 8: Image patch from the IDD dataset.

Figure 9: Question 9: Image patch from the IDD dataset.

Figure 10: Question 10: Image patch from the IDD dataset.
2 Additional quantitative results

For completeness, we also evaluate our method on additional no-reference image-quality assessment (NR-IQA) metrics including: NIQE [8], BRISQUE [7], PIQE [11], NRQM [6], and PI. PI is a weighted score computed as $\frac{1}{2}((10 - NRQM) + NIQE)$. However, as NR-IQA is a challenging problem, the aforementioned methods are known to correlate poorly with human ratings [4]. Hence, in our work rely on the mean opinion ranking results as an accurate measure of the quality of the reconstructed images.

Cityscapes (Real)

<table>
<thead>
<tr>
<th>Method</th>
<th>NIQE ↓</th>
<th>BRISQUE ↓</th>
<th>PIQE ↓</th>
<th>NRQM ↑</th>
<th>PI ↓</th>
<th>NIMA ↑</th>
<th>Meta-IQA ↑</th>
<th>MOR ↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bicubic [3]</td>
<td>5.62</td>
<td>45.14</td>
<td>75.75</td>
<td>6.60</td>
<td>4.51</td>
<td>4.62</td>
<td>0.245</td>
<td>-</td>
</tr>
<tr>
<td>ESRGAN [12]</td>
<td>2.94</td>
<td>22.13</td>
<td>18.82</td>
<td>9.94</td>
<td>1.50</td>
<td>4.95</td>
<td>0.247</td>
<td>-</td>
</tr>
<tr>
<td>M2SR [9]</td>
<td>6.09</td>
<td>56.17</td>
<td>72.74</td>
<td>7.37</td>
<td>4.36</td>
<td>4.88</td>
<td>0.231</td>
<td>3.33</td>
</tr>
<tr>
<td>DPSR [13]</td>
<td>5.97</td>
<td>53.33</td>
<td>87.00</td>
<td>9.32</td>
<td>3.33</td>
<td>4.83</td>
<td>0.240</td>
<td>4.44</td>
</tr>
<tr>
<td>RealSR [2]</td>
<td>2.95</td>
<td>29.67</td>
<td>24.50</td>
<td>9.13</td>
<td>1.91</td>
<td>4.87</td>
<td>0.236</td>
<td>2.75</td>
</tr>
<tr>
<td>DAN [5]</td>
<td>4.77</td>
<td>58.17</td>
<td>76.38</td>
<td>6.90</td>
<td>3.93</td>
<td>4.65</td>
<td>0.246</td>
<td>3.47</td>
</tr>
<tr>
<td>Ours</td>
<td>3.44</td>
<td>35.16</td>
<td>36.99</td>
<td>9.54</td>
<td>1.95</td>
<td>5.04</td>
<td>0.254</td>
<td>1.21</td>
</tr>
</tbody>
</table>

Table 2: Quantitative results on the Cityscapes and IDD validation sets. ↑ and ↓ indicate whether higher or lower values are desired, respectively. As seen, the traditional metrics (NIQE-PI) correlates poorly with MOR. However, the more recent NIMA metric shows good correlation with MOR. Our method achieves superior NIMA and MOR results on both datasets.
References


