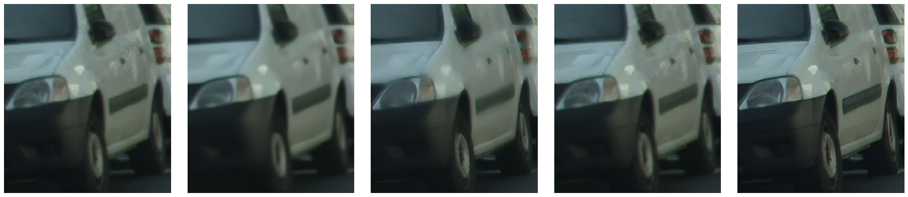
Semantic Segmentation Guided Real-World Super-Resolution: Supplementary Material

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 corresponds 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.

Method	Cityscapes MOR ↓	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	**1.21**	**1.45**

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



MZSR[9] DPSR[13] RealSR[2] DAN[5] Ours

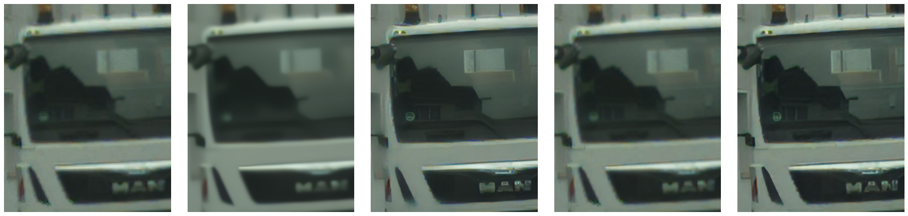
Figure 1: Question 1: Image patch from the CityScapes dataset.

MZSR[9] DPSR[13] RealSR[2] DAN[5] Ours

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

MZSR[9] DPSR[13] RealSR[2] DAN[5] Ours

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



MZSR[9] DPSR[13] RealSR[2] DAN[5] Ours

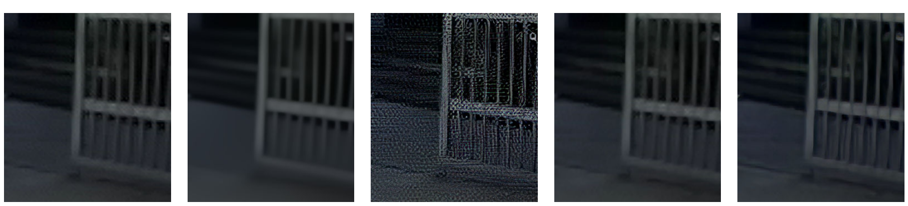
Figure 4: Question 4: Image patch from the CityScapes dataset.

MZSR[9] DPSR[13] RealSR[2] DAN[5] Ours

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

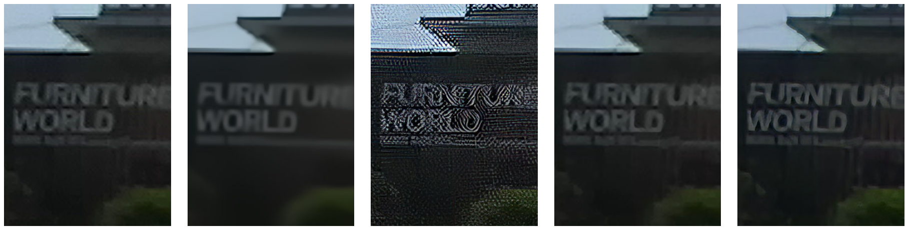
MZSR[9] DPSR[13] RealSR[2] DAN[5] Ours

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



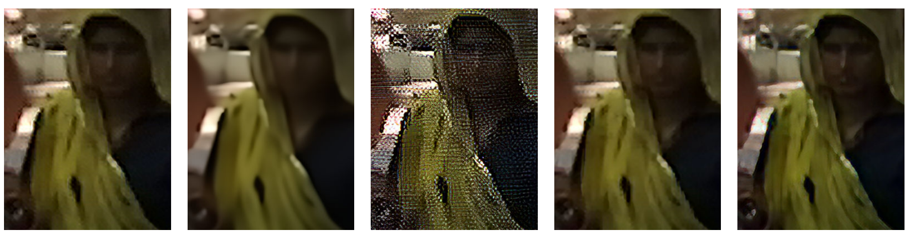
MZSR[9] DPSR[13] RealSR[2] DAN[5] Ours

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



MZSR[9] DPSR[13] RealSR[2] DAN[5] Ours

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



MZSR[9] DPSR[13] RealSR[2] DAN[5] Ours

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

MZSR[9] DPSR[13] RealSR[2] DAN[5] Ours

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 $\frac{1}{2}((10 - NQRM) + NIQE)$. 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)

Method	NIQE ↓	BRISQUE ↓	PIQE ↓	NRQM ↑	PI ↓	NIMA ↑	Meta-IQA ↑	MOR ↓
Bicubic [3]	5.62	45.14	75.75	6.60	4.51	4.62	0.245	-
ESRGAN [12]	**2.94**	**22.13**	**18.82**	**9.94**	**1.50**	4.95	0.247	-
MZSR [9]	6.09	56.17	72.74	7.37	4.36	4.88	0.231	3.33
DPSR [13]	5.97	53.33	87.00	9.32	3.33	4.83	0.240	4.41
RealSR [2]	2.95	29.67	24.50	9.13	1,91	4.87	0.236	2.75
DAN [5]	4.77	58.17	76.38	6.90	3,93	4.65	0.246	3.47
Ours	3.44	35.16	36.99	9.54	1.95	**5.04**	**0.254**	**1.21**

IDD (Real)

Method	NIQE ↓	BRISQUE ↓	PIQE ↓	NRQM ↑	PI ↓	NIMA ↑	Meta-IQA ↑	MOR ↓
Bicubic [3]	5.30	43.74	71.48	11.98	1.66	4.73	0.330	-
ESRGAN [12]	5.30	**33.44**	39.88	21.13	-2.92	4.94	0.325	-
MZSR [9]	5.12	54.36	81.15	11.76	1.68	5.00	0.330	2.96
DPSR [13]	5.42	58.45	88.70	8.81	3.31	4.92	0.330	3.16
RealSR [2]	4.63	39.52	31.00	**23.75**	**-4.56**	4.83	0.296	4.88
DAN [5]	4.61	57.40	86.32	10.75	2.68	4.77	**0.330**	2.48
Ours	**3.69**	37.48	**32.63**	14.88	-0.595	**5.03**	0.323	**1.45**

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

[1] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In *CVPR*, 2016.

[2] X. Ji, Y. Cao, Y. Tai, C. Wang, J. Li, and F. Huang. Real-world super-resolution via kernel estimation and noise injection. In *CVPR Workshops*, 2020.

[3] R. G. Keys. Cubic Convolution Interpolation for Digital Image Processing. *IEEE Trans Acoust. Speech Signal Process*, 1981.

[4] R. Lugmayr, M. Danelljan, and R. Timofte. Ntire 2020 challenge on real-world image super-resolution: Methods and results. *CVPR Workshops*, 2020.

[5] Zhengxiong Luo, Y. Huang, S. Li, L. Wang, and T. Tan. Unfolding the alternating optimization for blind super resolution. In *NeurIPS*, 2020.

[6] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang. Learning a no-reference quality metric for single-image super-resolution. *Comput. Vis. Image Underst. 158*, 2017.

[7] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-reference image quality assessment in the spatial domain. *TIP*, 2012.

[8] A. Mittal, R. Soundararajan, and A. C. Bovik. Making a "Completely Blind" Image Quality Analyzer. *IEEE Signal Processing Letters*, 2013.

[9] J. W. Soh, S. Cho, and N. I. Cho. Meta-transfer learning for zero-shot super-resolution. In *CVPR*, 2020.

[10] G. Varma, A. Subramanian, A. M. Namboodiri, M. Chandraker, and C. V. Jawahar. IDD: A dataset for exploring problems of autonomous navigation in unconstrained environments. In *WACV*, 2019.

[11] N. Venkatanath, D. Praneeth, C. B. Maruthi, S. C. Sumohana, and S. M. Swarup. Blind image quality evaluation using perception based features. In *NCC*, 2015.

[12] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *ECCV*, 2019.

[13] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Deep plug-and-play super-resolution for arbitrary blur kernels. In *CVPR*, 2019.