Supplementary Material - Fourier Space Losses for Efficient Perceptual Image Super-Resolution

Dario Fuoli[1] Luc Van Gool[1,2] Radu Timofte[1]
[1]Computer Vision Lab, ETH Zurich, Switzerland [2]KU Leuven, Belgium
{dario.fuoli, vangool, radu.timofte}@vision.ee.ethz.ch

We provide additional evaluations and visual results for our main models, that support the conclusions in the paper. In order to show the generalization capabilities of our approach, we provide results on a third dataset (BSD100). Also, in addition to PSNR, SSIM, LPIPS and FID, we calculate popular no-reference metrics (Ma, NIQE, PI) on DIV2K and discuss their limitations for perceptual quality assessment. We also show an ablation study for our proposed Fourier space GAN architecture.

1. Quantitative Evaluations

The evaluation on 3 different datasets show the benefits and generalization capabilities of our proposed losses in comparison to previous approaches. The application of our losses directly in Fourier domain improves not only the perceptual quality, but also the restoration quality at the same time.

1.1. Urban100

In Tab. 1 we show the results of additional methods on Urban100 [5]. As already discussed, our losses show similar performance as on DIV2K (validation), which shows the generalizability of our proposed loss functions. Again, ESRGAN (Our losses) achieves the highest perceptual scores with a substantial improvement in FID of -1.43 over the version without our losses. Our losses in conjunction with IMDN [6] achieve comparable results with SRFlow despite the enormous difference in runtime (41ms vs. 1995ms). Ours (Full), our efficient implementation, outperforms all versions of RankSRGAN in every metric, and is also faster at inference time.

1.2. BSD100

In addition to DIV2K(val) and Urban100, we evaluate the performance also on BSD100 [9], another commonly used dataset, see Tab. 2. Again, ESRGAN (Our losses) performs best in terms of perceptual quality and also achieves high restoration quality. SRFlow has high restoration quality but can not compete in perceptual quality in comparison to all other methods. Ours (Full) outperforms all RankSRGAN models in all metrics, only RankSRGAN (NIQE) achieves a lower FID score.

1.3. DIV2K - No-reference Metrics

No-reference metrics are handcrafted quality assessment tools to quantify image quality without comparison to a ground truth. However, these metrics are limited for objective image quality quantification, because of the lacking reference. We would like to learn the true target distribution. We therefore chose FID [3] as a quality measure for distributional similarity, which together with LPIPS [12] quantifies perceptual quality.

Nevertheless, we list the results for no-reference quality metrics Ma [8], NIQE [10] and PI [2] for DIV2K in

Method	↑PSNR	↑SSIM	↓LPIPS	↓FID
SRFlow [7]	25.25	0.735	0.127	26.22
ESRGAN (Our losses) [11]	25.05	0.738	0.120	24.07
ESRGAN [11]	24.36	0.717	0.123	25.50
RankSRGAN (Ma) [13]	24.12	0.704	0.143	27.72
RankSRGAN (NIQE) [13]	24.52	0.715	0.143	27.47
RankSRGAN (PI) [13]	24.47	0.716	0.139	27.84
Ours (Full)	24.69	0.723	0.132	26.70

Table 1. Evaluation on Urban100 [5]. Red indicates best, blue second best.

Method	↑PSNR	↑SSIM	↓LPIPS	↓FID
SRFlow [7]	26.08	0.667	0.183	66.24
ESRGAN (Our losses) [11]	25.79	0.658	0.158	57.90
ESRGAN [11]	25.34	0.643	0.161	60.42
RankSRGAN (Ma) [13]	25.06	0.633	0.183	65.75
RankSRGAN (NIQE) [13]	25.52	0.642	0.178	61.52
RankSRGAN (PI) [13]	25.48	0.643	0.175	63.97
Ours (Full)	25.66	0.656	0.172	62.25

Table 2. Evaluation on BSD100 [9]. Red indicates best, blue second best.

Method	↑Ma	↓NIQE	↓PI
RankSRGAN (Ma) [13]	6.8142	2.6143	2.9000
RankSRGAN (NIQE) [13]	6.6923	2.7121	3.0099
RankSRGAN (PI) [13]	6.6794	2.6851	3.0029
SRFlow [7]	6.5230	3.5421	3.5095
ESRGAN (Our losses) [11]	6.5580	3.0388	3.2404
ESRGAN [11]	6.5937	3.0918	3.2491
Ours (WaveletSRNet losses) [4]	5.9682	4.9011	4.4664
Ours (Full)	6.6792	3.0836	3.2022

Table 3. Evaluation of no-reference metrics on DIV2K. Red indicates best, blue second best.

Tab. 3. As expected, the RankSRGAN [13] models perform the best, as they are explicitly trained with these metrics. Therefore, a direct comparison to all other methods is not fair. Interestingly, RankSRGAN (Ma) outperforms all other RankSRGAN models, even those that are trained for these specific metrics, which is unexpected. It is unclear why these inconsistencies arise, since RankSRGAN is trained for exactly these metrics.

Among the methods that are not explicitly trained for these metrics, our losses applied to ESRGAN and IMDN achieve the best results overall. Ours (Full) achieves the highest Ma and PI scores, ESRGAN (Our losses) achieves the best NIQE score.

2. Visual Results

We show a series of visual examples on all 3 datasets to asses the quality by visual inspection. In addition we provide PSNR and LPIPS for each method as quantitative metrics for restoration and perceptual quality respectively.

The metrics are in line with our quantitative evaluation overall. Note, we deliberately show some cases where the individual scores do not exactly match our overall quantitative evaluation. These differences on individual images arise due to different strengths and weaknesses of each method.

The application of our losses in general improves the restoration- and perceptual quality, as can be seen by the examples of ESRGAN, ESRGAN (Our losses) and Ours (Full). Our efficient setting with IMDN as generator achieves comparable performance to the larger model ESRGAN and especially the largest model SRFlow with highly improved runtimes. Ours (Full) in general also improves perceptual and restoration quality in comparison with RankSRGAN. Additionally, we observed that SRFlow tends to generate noisy output, even in areas of uniform color. ESRGAN tends to produce excess edges.

Method	↑PSNR	↑SSIM	↓LPIPS	↓FID
Ours (Config. 5, $\mathcal{L}_{GAN}^{\mathcal{F},3}$)	29.02	0.792	0.126	17.51
Ours (Config. 5, $\mathcal{L}_{GAN}^{\mathcal{F},5}$)	29.06	0.796	0.129	17.17
Ours (Config. 8, $\mathcal{L}_{GAN}^{\mathcal{F},3}$)	28.32	0.770	0.122	16.19
Ours (Config. 8, $\mathcal{L}_{GAN}^{\mathcal{F},5}$)	28.42	0.776	0.124	15.88

Table 4. Ablation of FFTGAN architecture on DIV2K [1]. Red indicates best.

3. Fourier GAN Architecture - Ablation

We provide additional analysis of our Fourier space GAN loss by training a smaller architecture with a reduced number of layers. We compare the full size GAN architecture ($\mathcal{L}_{GAN}^{\mathcal{F},5}$) with a reduced GAN architecture where the number of layers is set to 3 ($\mathcal{L}_{GAN}^{\mathcal{F},3}$). We test this setup in configuration 5 and 8 from our ablation study in Tab. 4. The higher complexity discriminator achieves consistently better scores in PSNR, SSIM and FID in both configuration 5 and 8. LPIPS is slightly improved when using $\mathcal{L}_{GAN}^{\mathcal{F},3}$. We suspect this could be in trade-off with FID due to an increased weight on the VGG-loss during training, when the discriminator is weaker.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017. 2

[2] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. The 2018 pirm challenge on perceptual image super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, September 2018. 1

[3] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30, pages 6626–6637. Curran Associates, Inc., 2017. 1

[4] Huaibo Huang, Ran He, Zhenan Sun, and Tieniu Tan. Wavelet-srnet: A wavelet-based cnn for multi-scale face super resolution. In *IEEE International Conference on Computer Vision*, pages 1689–1697, 2017. 2

[5] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5197–5206, 2015. 1

[6] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. Lightweight image super-resolution with information multi-distillation network. In *Proceedings of the 27th ACM International Conference on Multimedia (ACM MM)*, pages 2024–2032, 2019. 1

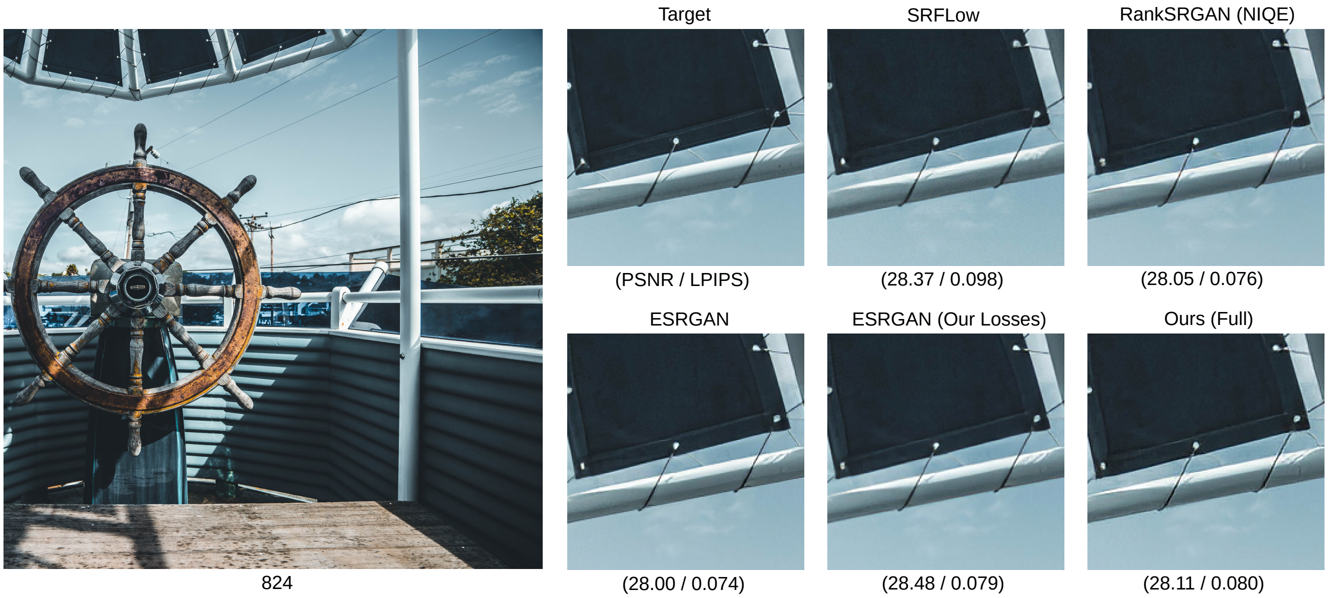

Figure 1. Visual examples on DIV2K, image 824.

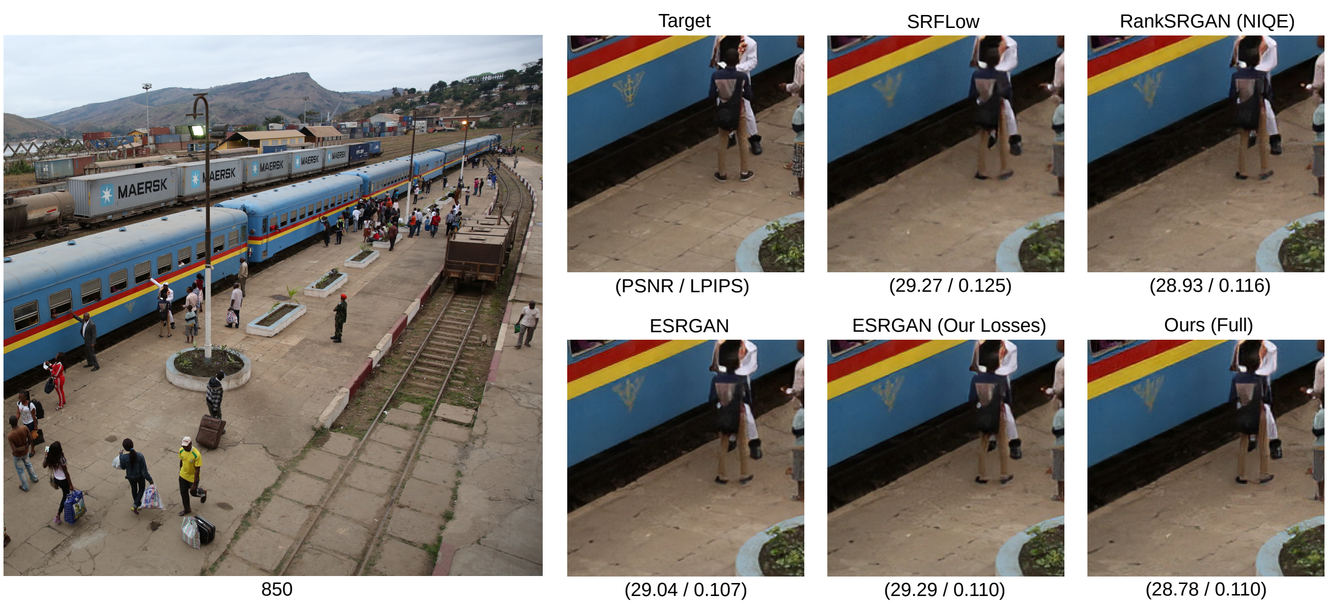

Figure 2. Visual examples on DIV2K, image 850.

[7] Andreas Lugmayr, Martin Danelljan, Luc Van Gool, and Radu Timofte. Srflow: Learning the super-resolution space with normalizing flow. In *ECCV*, 2020. 1, 2

[8] Chao Ma, Chih-Yuan Yang, Xiaokang Yang, and Ming-Hsuan Yang. Learning a no-reference quality metric for single-image super-rolution. *Computer Vision and Image Understanding*, pages 1–16, 2017. 1

[9] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 416–423 vol.2, 2001. 1

[10] A. Mittal, R. Soundararajan, and A. C. Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2013. 1

[11] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *The European Conference on Computer Vision Workshops*

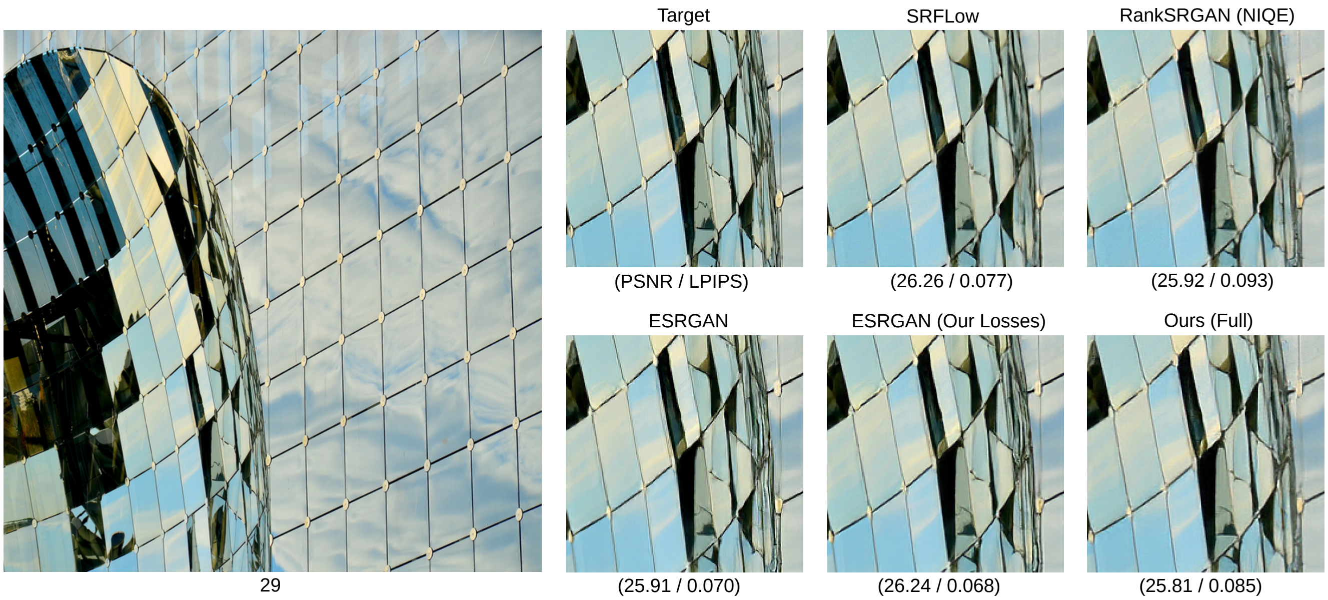

Figure 3. Visual examples on Urban100, image 29.

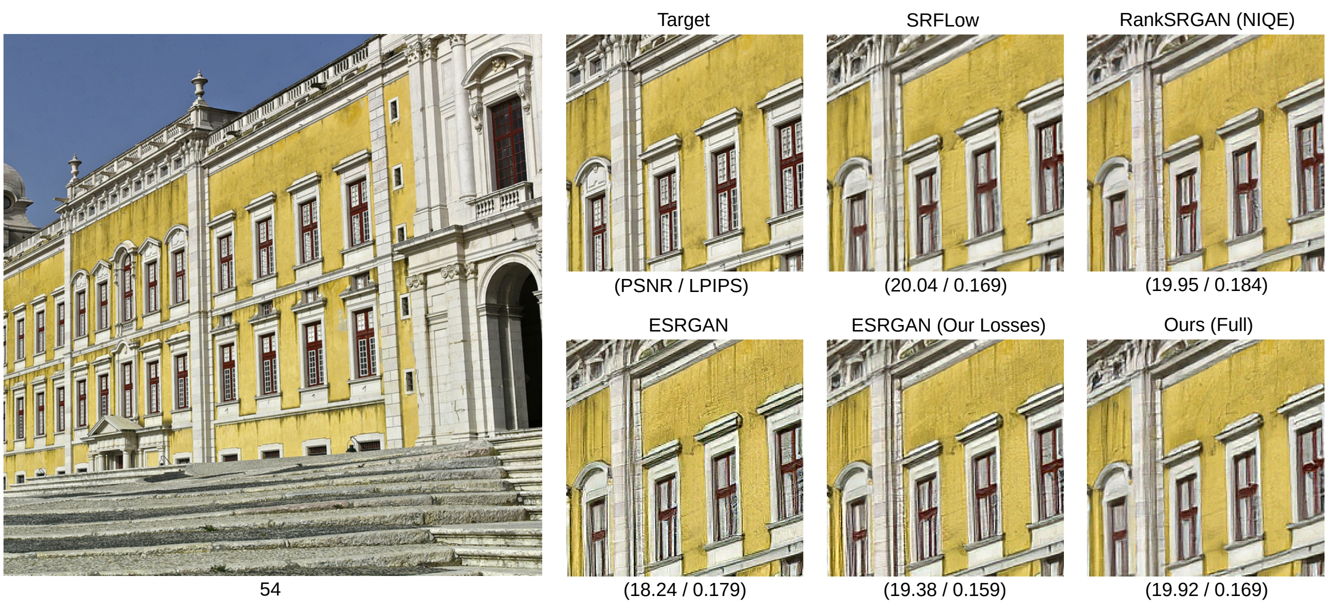

Figure 4. Visual examples on Urban100, image 54.

(ECCVW), September 2018. 1, 2

[12] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 1

[13] Wenlong Zhang, Yihao Liu, Chao Dong, and Yu Qiao. Ranksrgan: Generative adversarial networks with ranker for image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019. 1, 2

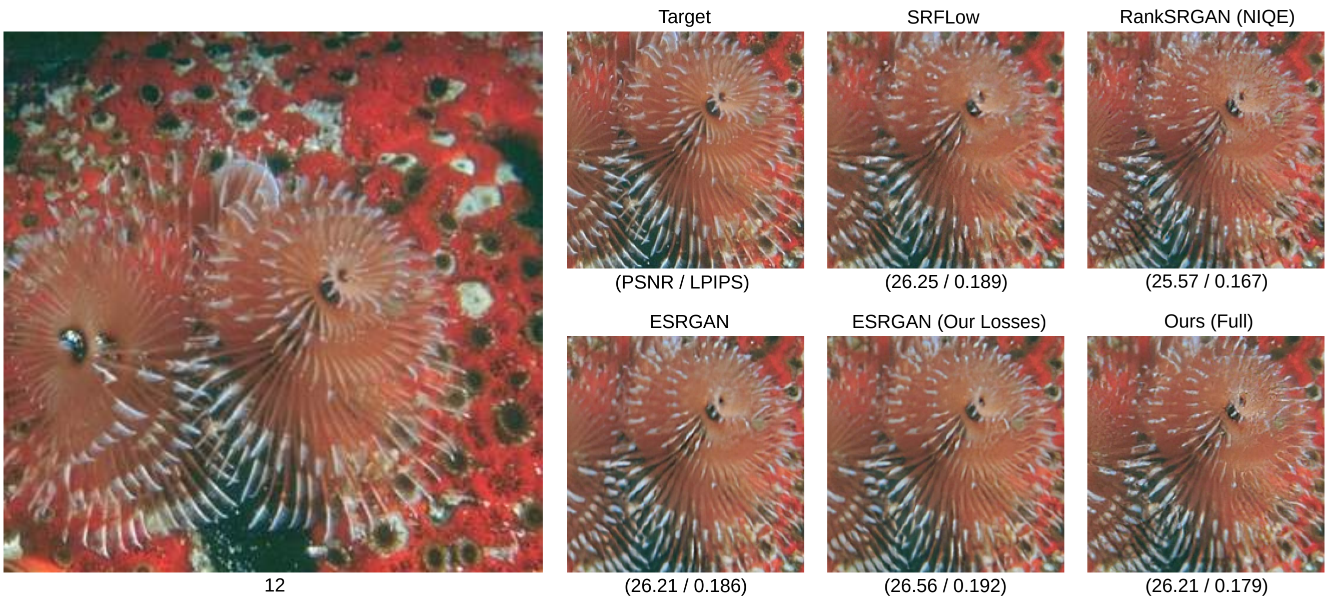

Figure 5. Visual examples on BSD100, image 12.

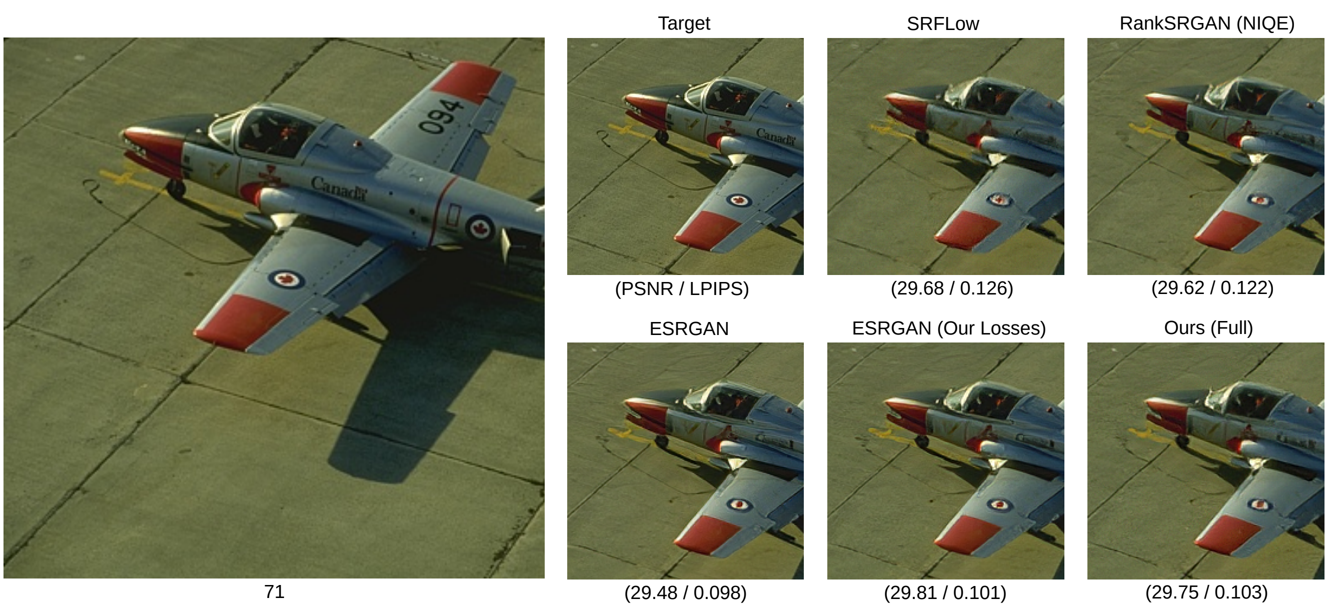

Figure 6. Visual examples on BSD100, image 71.