SROBB: Targeted Perceptual Loss for Single Image Super-Resolution

A. Supplementary Material

In this supplementary material, first, we provide additional qualitative and quantitative results on super-resolution benchmarks such as Set5 [1], Set14 [7] and BSD100 [6]. Then, we present more details of our extensive user study, as well as the time span taken by the users for decision making.

A.1. Results on standard benchmarks

A.1.1 Quantitative results

In this subsection, we conduct an evaluation study based on the quantitative metrics. Table 1 summarizes the average of SSIM, PSNR and LPIPS values of the Set5 and Set14 images, respectively. Because of the fact that the human eye is most sensitive to luma information, we compute the PSNR and SSIM values only for the intensity (luma) channel in YCbCr space.

As also emphasized in the paper, these metrics would not reflect the reconstruction quality; the reconstructed images using both our method and the SRGAN are not ranked first in terms of mentioned metrics, however, they generate more realistic and appealing super-resolved images comparing to the other methods. Therefore, here, we only present the qualitative results on the BSD100 test set.

Testset	Metric	Bicubic	SRCNN	SelfExSR	LapSRN	SRGAN	SROBB	HR image
Set5	SSIM	0.811	0.863	0.862	0.884	0.848	0.817	1.0
	PSNR	28.43	30.51	30.34	31.54	29.41	28.93	∞
	LPIPS	0.340	0.214	0.171	0.121	0.083	0.087	0.0
Set14	SSIM	0.704	0.756	0.757	0.772	0.739	0.678	1.0
	PSNR	26.01	27.52	27.41	28.19	26.04	25.43	∞
	LPIPS	0.440	0.332	0.301	0.312	0.148	0.162	0.0

Table 1. Comparison of bicubic interpolation, SRCNN [2], SelfExSR [3], LapSRN [4], SRGAN [5] and SROBB (ours) on the Set5 and Set14 test sets. Red color indicates the best measures (SSIM, PSNR [dB], LPIPS) and blue color indicates the second bests. The visual comparison of the images from these test sets are shown in Figures 2 and 3.

A.1.2 Qualitative results

In this part, we evaluate and compare the visual results of our method with SRGAN and bicubic interpolation methods on random images from the BSD100 test set, as well as the images from the Set5 and Set14 datasets, respectively. Figure 1 corresponds to the reconstructed images from the BSD100 dataset. Results on Set5 are shown in Figure 2 while Figure 3 shows some reconstructed images from the Set14. The upscaling factor of all images is set to four (Best viewed in zoom in).

A.2. Details of the user study

Figure 4 shows a screenshot of the survey that we used to evaluate our proposed method. The subjects were shown five reconstructed images and were asked to choose the image that looks more appealing to them. We also added the real high-resolution image in the same page as the reference. We cropped each image vertically to be able to fit all versions of the same image side by side within a single page. The height of the images are remained the same as the original size.

Options	Cannot decide	Only pixel -wise loss	With perc- eptual loss	With targeted perceptual loss
Average time	22.60	23.85	24.09	17.81

Table 2. The average of decision making duration [seconds] for users to choose the reconstructed images of each method.

Time span analysis for the user decision making In total, 51 persons participated in our ablation study. Among them, eight persons have been subject to a new experimental setting, under an additional controlled situation: we recorded the time span that each user spent to respond each question. As each user has different speed to complete the survey, we normalized all times to the average time by all users, 10:52 minutes (in average, 18.62 seconds per question). Table 2 shows the time that users spend to choose each of the following options: 1- the reconstructed image only by pixel-wise loss, 2- pixel-wise loss and standard perceptual loss, 3- pixel-wise loss and targeted perceptual loss (this work), and finally, 4- the "Cannot decide" option (The adversarial loss term is used for both 2 and 3). For images, where our method was the preferred choice, the average time span taken by the users to make decision was relatively shorter than other methods. We can conclude that, in cases where SROBB was not the winning choice, the difference between the super-resolved images using different loss terms was less significant, therefore, users had more difficulties to choose the best option. Meanwhile, users seem to be more sure when they are voting in favor of reconstructed images by the SROBB method. As a future work, to be able to validate this conclusion and to be sure that time span for decision making is not biased by the type of the image, this experiment needs to be extended with significantly more number of images.

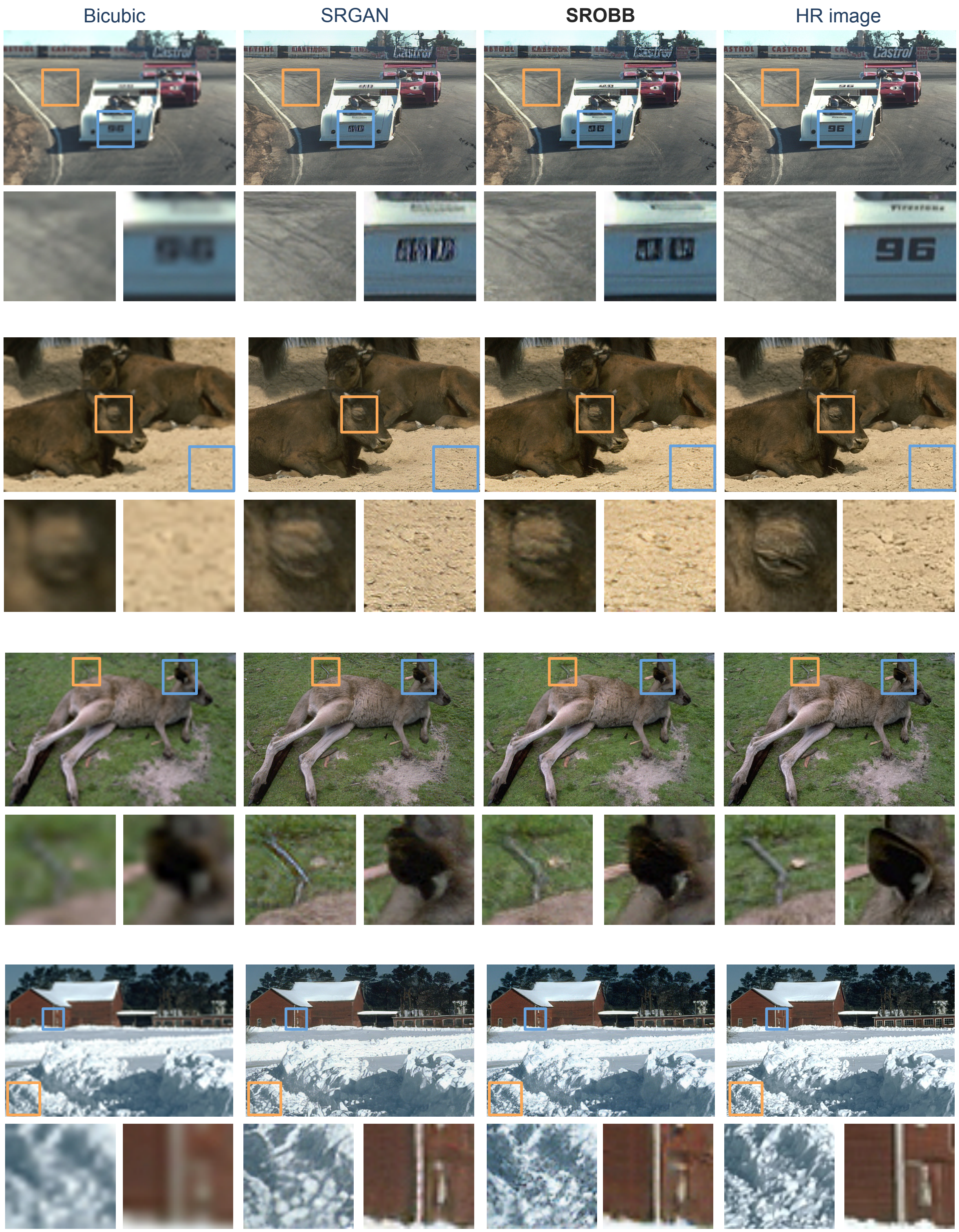

Figure 1. Qualitative results on random images from BSD100 [6] using bicubic interpolation, SRGAN[5], SROBB (ours), respectively. Zoom in for the best view. [4× upscaling]

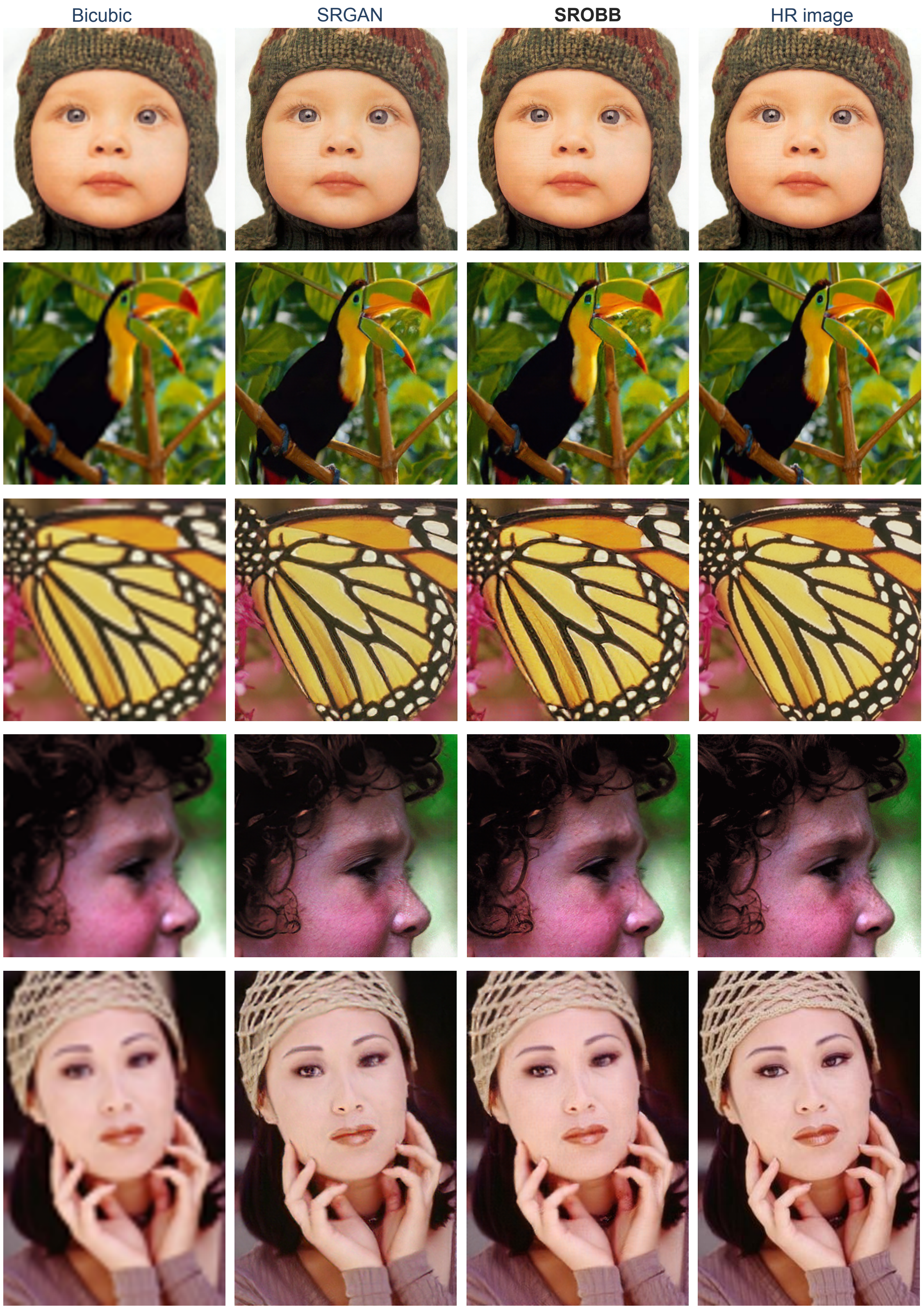

Figure 2. Qualitative results on the images from Set5 [1] using bicubic interpolation, SRGAN[5], SROBB (ours), respectively. Zoom in for the best view. [4× upscaling]

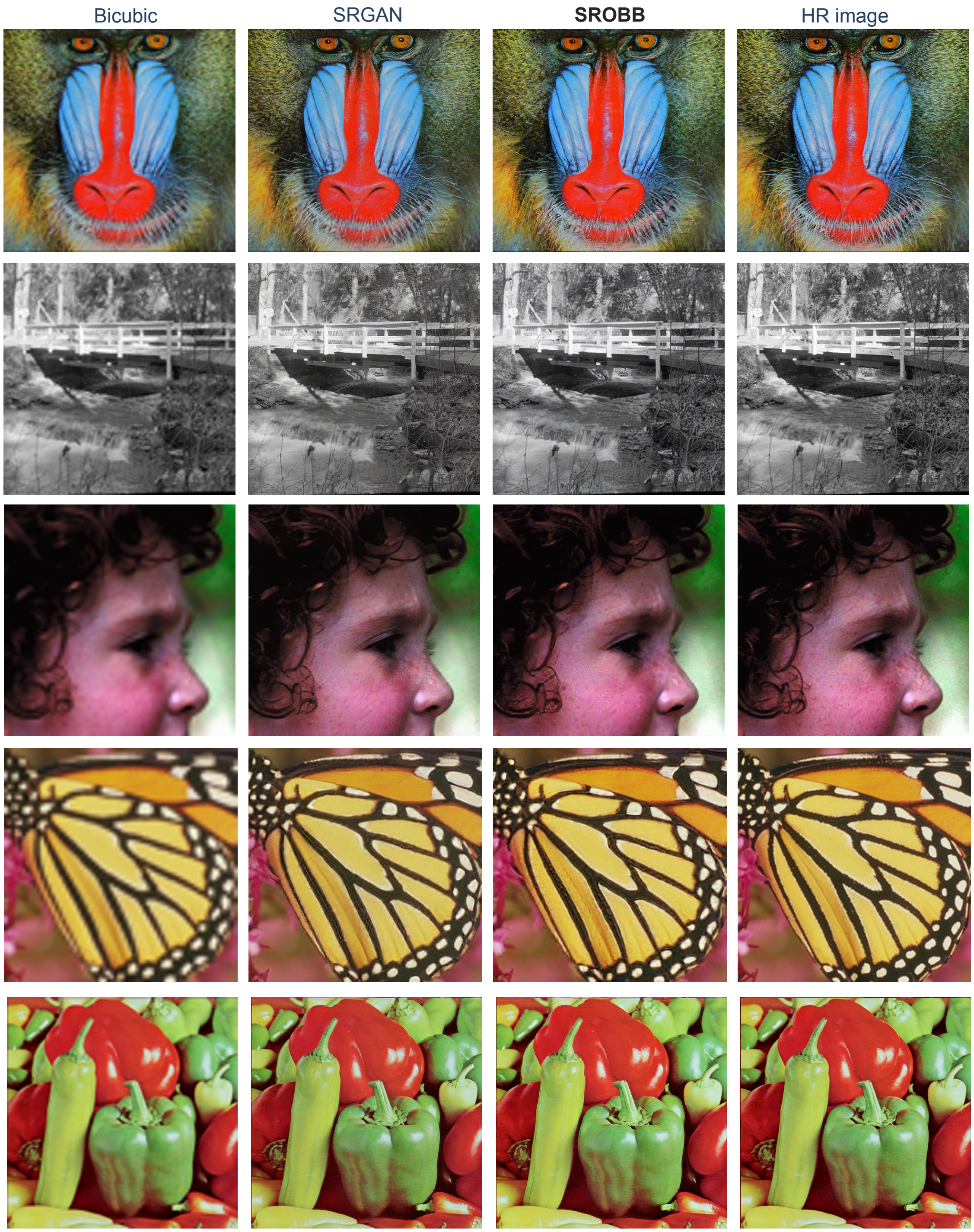

Figure 3. Qualitative results on the images from Set14 [7], using bicubic interpolation, SRGAN[5], SROBB (ours), respectively. Zoom in for the best view. [4× upscaling]

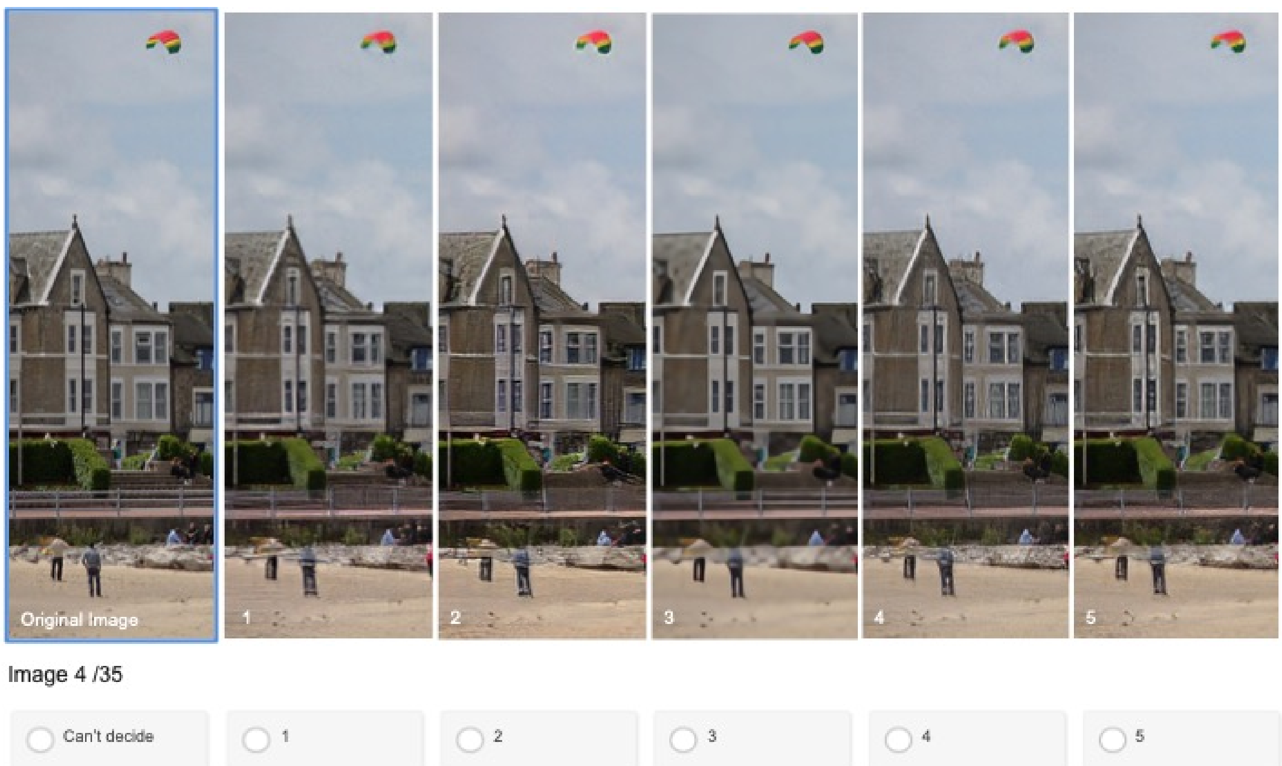

Figure 4. Example screenshot of our online survey, to perform a user study and compare our method to state-of-the-art PSNR and GAN-based approaches. In total, 46 persons participated in this survey and 1610 votes were obtained. Users selected the images produced by SROBB (ours) 38.3% while ESRGAN, SFT-GAN, SRGAN, RCAN, and "Cannot decide" had 27.1%, 13.9%, 12.5%, 4.7%, and 8.3% of the votes, respectively. In total, in 42.9% of images we were the winning choice by the majority of votes for SROBB.

References

[1] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie-Line Alberi Morel. Low-Complexity Single-Image Super-Resolution based on Nonnegative Neighbor Embedding. In *British Machine Vision Conference (BMVC)*, Guildford, Surrey, United Kingdom, Sept. 2012.

[2] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *CoRR*, abs/1501.00092, 2015.

[3] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *IEEE Conference on Computer Vision and Pattern Recognition)*, 2015.

[4] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. *CoRR*, abs/1704.03915, 2017.

[5] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi. Photo-realistic single image super-resolution using a generative adversarial network. *CoRR*, abs/1609.04802, 2016.

[6] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, 02 2001.

[7] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Proceedings of the 7th International Conference on Curves and Surfaces*, pages 711–730, Berlin, Heidelberg, 2012. Springer-Verlag.