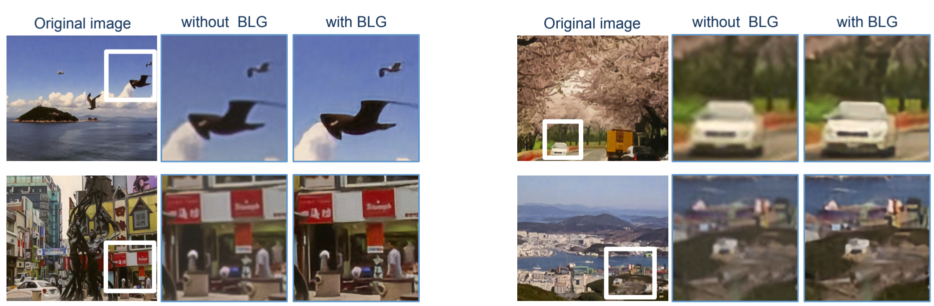
A. Supplementary Material

In this supplementary material, first, we use ESRGAN [8] and RCAN [10] methods for the SR decoder as the second step of our proposed method (RBSR) to demonstrate the generalization capability of the bicubic look-alike generator. Then, we provide a quantitative analysis of the proposed approach, by using the RealSR [2] testset images. In addition, we present more details concerning the computational cost of the proposed method, as well as the extensive user study we conducted in order to compare RBSR with other SOTA methods.

A.1. Generalization capabilities of the bicubic look-alike generator

Our proposed approach (RBSR) is a two step procedure. The first step transforms the real LR image using the bicubic look-alike generator. The second step uses any generic SR decoder trained on bicubically downsampled images, taking the transformed LR image as input. In the paper, for the qualitative comparison and the user study, we used a pre-trained EDSR network for this second step. Here, we show the robustness and generalizability of our two step approach by replacing the EDSR network with pretrained ESRGAN and RCAN models. To do so, we compare the results of these models on real LR images and our transformed LR images obtained from the bicubic look-alike generator. Experimental results demonstrate that these SR methods generate more plausible results with greater perceptual quality when fed with transformed LR images instead of real LR images (see Figure 1).

(a) RCAN [10]



(b) ESRGAN [8]

Figure 1: Comparison results of RCAN (a) and ESRGAN (2) methods on original images from the RealSR dataset and our transformed LR images, generated by our bicubic look-alike generator (BLG). Experimental results demonstrate that these SR methods generate more plausible results with greater perceptual quality when fed with transformed LR images instead of real LR images.

A.2. Quantitative results

In this work we tackle the real-world SR problem, where the downsampling operator is not known and therefore no ground-truth is available. Hence, calculating distortion metrics such as PSNR and SSIM is not possible for test images that truly reflect this problem (original images from smartphones, TV streams, etc.). Although, as mentioned previously, RealSR [2] is the only dataset with physically produced high and low-resolution image pairs and is the closest existing dataset to real low and high resolution pairs.

Table 1 shows the SSIM and PSNR values estimated between super-resolved images of RealSR LR test images and their HR counterparts, using bicubic upsampling, EDSR-real [4], the RealSR network [2], DPSR [9] and our proposed method. The training details of each method is presented in Section 4.3 of the main manuscript. We also add the perception index (PI) metric to our evaluation; this index combines two no-reference image quality measures of Ma et al. [6] and NIQE [7] and was shown to have a higher correlation with human opinion than other commonly used metrics [1]. As PI is a no-reference metric, it can be also used for test images that have no ground-truth.

Dataset	Method	bicubic	SRResNet	RCAN	EDSR-real	DPSR	RealSR	**RBSR**
RealSR	SSIM	0.77	0.79	0.80	0.81	0.79	0.81	**0.82**
	PSNR	26.63	26.98	27.11	26.51	27.02	**28.05**	26.54
	PI	9.28	9.06	9.19	7.94	9.12	8.97	**7.76**
DIV2K HR	SSIM/PSNR	-	-	-	no ground-truth	-	-	-
	PI	10.02	9.62	9.81	9.01	9.36	9.19	**8.48**
DPED (cellphones)	SSIM/PSNR	-	-	-	no ground-truth	-	-	-
	PI	10.24	9.91	10.02	9.62	9.73	9.55	**7.92**
TV Streams	SSIM/PSNR	-	-	-	no ground-truth	-	-	-
	PI	11.52	10.71	10.64	**10.04**	11.19	10.32	10.15

Table 1: Comparison of bicubic interpolation, SRResNet [3], RCAN [10], EDSR [4], DPSR [9], RealSR [2] and RBSR (ours) on different presented test sets. Best measures (SSIM ↑, PSNR [dB] ↑, PI ↓) are highlighted in bold.

A.3. Ablation study

In this section, we perform another study to investigate the effectiveness of each proposed component of the bicubic look-alike generator. We compare the performance of our network trained with the combinations of different settings such as different loss functions, and trainings with and without copying mechanism. These setting are listed in Table 2. We calculate PSNR and SSIM for each setting on RealSR [2] test set, the only available dataset with ground-truth for real-world SR task. For each setting, SSIM and PSNR values are calculated after upsampling the picture by a fixed $\times 4$ SR decoder and comparing it to the RealSR ground-truth.

Name	Description	SSIM	PSNR
$RBSR_{MSE}$	only $\mathcal{L}_{MSE}$ loss	0.788	27.69
$RBSR_{E}$	only $\mathcal{L}_1$ loss	0.792	**27.95**
$RBSR_{EP}$	$\mathcal{L}_1 + \mathcal{L}_{perceptual}$	0.811	26.98
$RBSR_{EPA}$	$\mathcal{L}_1 + \mathcal{L}_{perceptual} + \mathcal{L}_{adversarial}$	0.798	26.60
$RBSR_{EBA}$	$\mathcal{L}_1 + \mathcal{L}_{bicubic\,perceptual} + \mathcal{L}_{adversarial}$	**0.835**	26.73
$RBSR$	$\mathcal{L}_1 + \mathcal{L}_{bicubic\,perceptual} + \mathcal{L}_{adversarial} +$ Copying mechanism	0.820	26.54

Table 2: Comparing the effect of each proposed component of the bicubic look-alike generator on LR and HR images of [2] test set. Best measures (SSIM ↑, PSNR [dB] ↑) are highlighted in bold. As mentioned earlier, **these metrics are not directly correlated to the perceptual quality, therefore, we chose our best baseline based on qualitative comparison shown in Figure 5 and Figure 6 of the manuscript, comparing $RBSR_{EPA}$ to $RBSR_{EBA}$ and $RBSR_{EBA}$ to $RBSR$, respectively.**

As it is already emphasized in the Section 4 of the manuscript, the distortion metrics are not directly correlated to the perceptual quality as judged by human raters, therefore, we chose our best baseline based on qualitative comparisons such as Figure 5 and Figure 6 of the mains manuscript. Our best baseline is then compared to state-of-the-art works on real-world SR by an extensive user study, following the standard procedure of the ICCV AIM 2019 challenge [5] on Real-world Super-Resolution.

A.4. Computational cost

In our paper, we compared our two step approach (RBSR), our end-to-end comparison (EDSR-real), RealSR [2], and DPSR [9]. In terms of computational cost, both RealSR and DPSR have different disadvantages. RealSR's network calculations take place in the high-resolution space, incurring a heavy memory overhead cost. For example, running the model on CPU requires 19 GB of RAM for an image of size 1200×1200, which is the maximum possible. DPSR is an iterative algorithm, requiring multiple forward passes and multiple deblurring steps in order to converge to an acceptable solution; DPSR uses an iterative approach by default for real LR images. Hence, these two algorithms have either high memory overhead or high computation time overhead. In contrast, RBSR requires two forward passes per input image. The first network is relatively lightweight, as it operates exclusively in the LR space. The second network can be any generic SR decoder for bicubically downsampled images. The complete pipeline (using EDSR as the SR decoder) reconstructs 1024×768 pixel images at 26.9 FPS, using a GeForce GTX 1080 Ti. Our end-to-end setting (EDSR-real) reconstructs the same size images at 33.7 FPS using the same GPU.

A.5. Details of the user study

Figure 2 shows a screenshot of the survey that we used to evaluate our proposed method. The subjects were shown four reconstructed images (one from each algorithm) and were asked to choose the image that looked most appealing to them.

Figure 2: Example screenshot of our online survey to perform a user study and compare our method to state-of-the-art real-world SR approaches. In total, 41 people participated in this survey.

We note that no reference image was shown, since the vast majority of the images had no ground truth. In sum, 41 people participated in user study. We refer the reader to the **supplementary material B** for the full version of the survey and additional statistical results for each image. An interactive version of this survey will be available online, after the acceptance of the paper.

References

[1] Blau, Y., Mechrez, R., Timofte, R., Michaeli, T., Zelnik-Manor, L.: 2018 PIRM challenge on perceptual image super-resolution. CoRR **abs/1809.07517** (2018), `http://arxiv.org/abs/1809.07517`

[2] Cai, J., Zeng, H., Yong, H., Cao, Z., Zhang, L.: Toward real-world single image super-resolution: A new benchmark and a new model. In: Proceedings of the IEEE International Conference on Computer Vision (2019)

[3] Ledig, C., Theis, L., Huszar, F., Caballero, J., Aitken, A.P., Tejani, A., Totz, J., Wang, Z., Shi, W.: Photo-realistic single image super-resolution using a generative adversarial network. CoRR **abs/1609.04802** (2016), `http://arxiv.org/abs/1609.04802`

[4] Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) pp. 1132–1140 (2017)

[5] Lugmayr, A., Danelljan, M., Timofte, R., Fritsche, M., Gu, S., Purohit, K., Kandula, P., Suin, M., Rajagopalan, A.N., Joon, N.H., Won, Y.S., Kim, G., Kwon, D., Hsu, C.C., Lin, C.H., Huang, Y., Sun, X., Lu, W., Li, J., Gao, X., Bell-Kligler, S.: Aim 2019 challenge on real-world image super-resolution: Methods and results (2019)

[6] Ma, C., Yang, C., Yang, X., Yang, M.: Learning a no-reference quality metric for single-image super-resolution. CoRR **abs/1612.05890** (2016)

[7] Mittal, A., Soundararajan, R., Bovik, A.: Making a completely blind image quality analyzer. Signal Processing Letters, IEEE **20**, 209–212 (03 2013). https://doi.org/10.1109/LSP.2012.2227726

[8] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Loy, C.C.: Esrgan: Enhanced super-resolution generative adversarial networks. In: ECCV Workshops (2018)

[9] Zhang, K., Zuo, W., Zhang, L.: Deep plug-and-play super-resolution for arbitrary blur kernels. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1671–1681 (2019)

[10] Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: ECCV (2018)