Supplementary Material for
SGSST: Scaling Gaussian Splatting Style Transfer

Bruno Galerne[1,2], Jianling Wang[1], Lara Raad[3], and Jean-Michel Morel[4]
[1]Université d'Orléans, Université de Tours, CNRS, IDP, UMR 7013, Orléans, France
[2]Institut Universitaire de France (IUF)
[3]Instituto de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de la República [4]City University of Hong Kong

Supplementary material description Our supplementary material consists of the following elements:
- The present document with additional details and figures.
- The project website: https://www.idpoisson.fr/galerne/sgsst/ with rendered videos, including one video showing the stylized scenes of the main paper's teaser and videos for the 40 comparison experiments (see Figures 18 to 26).
- The source code used for all experiments availble at https://github.com/JianlingWANG2021/SGSST based on the public source codes[1] for 3DGS [4] training and SPST [3].

Note that due to space constraints all the images of this document have been compressed.

1. Ablation on the number of scales

As explained in the main paper, the number of scales n_s is set automatically to use all available scales, the coarsest resolution having sides larger than 256 for VGG19 statistics to be reliable. Figure 9 presents an ablation on the number of scales n_s showing the results for different values of n_s and the corresponding close-ups of these results (after an initial color transfer for first 10k iterations using coarsest scale $n_s = 4$ for all examples). One can observe that when using only the large resolution images ($n_s = 1$) the pattern of the style transfer are limited to HR details. High-quality style transfer is only achieved when using all scales.

2. UHR style transfers of the teaser figure

Due to space limitation, style images of the main paper's teaser figure have been displayed as tiny images regardless of their resolution. Figures 10 to 17 show the eight pairs of images of this figure in full size to better appreciate the multiscale details of the style images and their corresponding stylized results. Each style image is displayed at the same resolution as the rendered view so that one can observe that the style features are reproduced with the same size (see e.g. the stone wall of Figure 15).

3. Comparison experiments

As said in the main paper, we performed a thorough comparative study using 40 3D style transfer experiments using 9 different scenes from previous works [1, 2, 4] and various style images. We compare our results with the NeRF-based ARF [6] and the 3DGS-based StyleGaussian [5] algorithms using their public implementations[2].

Figures 18 to 26 display a rendered view for each of these 40 experiments. Let us recall that for the HR scenes (Figures 18 to 22) our approach is the only one working at high-resolution. While SGSST produces outputs having the content size, StyleGaussian outputs are limited in resolution to a maximal width of 1600 or maximal height of 1200, and for ARF the content images have been downscaled by a factor 4 to obtain a low-resolution input suitable for ARF (see Section 4 below). Video versions of these figures are available at: https://www.idpoisson.fr/galerne/sgsst/comparison_web.html.

Moreover, we provide with Figure 27 a second version of the comparison figure (Figure 4) with close views to highlight the texture consistency of each method.

4. ARF and high resolution inputs

ARF [6] uses Nearest Neighbor Feature Matching (NNFM) of a single layer of VGG features for fine tuning a plenoxel radiance field [2]. It produces high-quality results at moderate resolution. While comparing our results with ARF, we observed that this algorithm does not produce visually satisfying results for high-resolution scenes. This is illustrated by Figure 28 where one can observe that the style transfer quality decreases as the input size increases. To allow a fair

[1]https://github.com/graphdeco-inria/gaussian-splatting; https://github.com/bgalerne/scaling_painting_style_transfer

[2]https://github.com/Kai-46/ARF-svox2; https://github.com/Kunhao-Liu/StyleGaussian

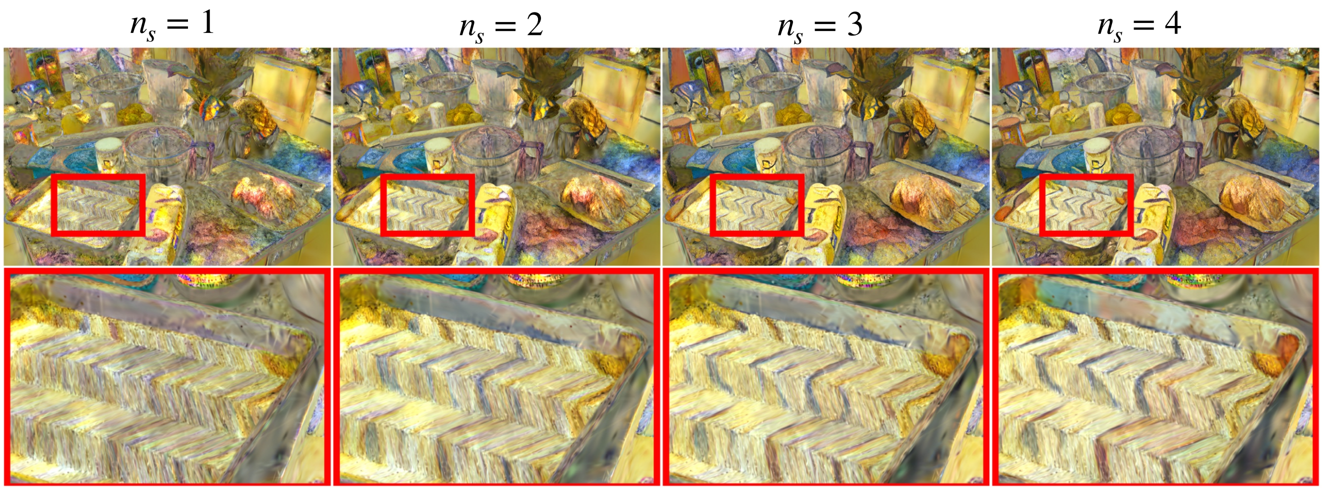

Figure 9. **Ablation of the number of scales** of the SOS loss. Style transfer results using different number of scales (starting from the same initialization obtained by 10k iterations using coarsest scale for all). High-quality style transfer is only achieved when using all scales ($n_\mathrm{s} = 4$).

comparison we decided to downscale images by a factor 4 for the high-resolution scene as a preprocess for ARF.

Although it has been shown that NNFM is superior to Gram feature matrix optimization for NeRF style transfer when optimizing for a single VGG layer [6], our results show that optimizing for a (slightly corrected [3]) Gram-based loss using several image scales and five VGG layers for each scale is an effective solution for applying high quality style transfer at UHR.

5. Details on the perceptual study

As described in the main paper, a comparative perceptual study was conducted using the 40 3D style transfer experiments presented in Section 3 (Figures 18 to 26). For each experiment, they were asked to pick the image that appeared to be the most faithful to the style image among the three displayed results. Each participant was shown ten random experiments and participation was voluntary. Figure 28 is an example of such an experiment displaying the style image (top left image) and three views of the scenes stylized by SGSST, ARF and StyleGaussian respectively and displayed in random order. To choose between one of the three results, the participant had to press the left arrow key to select the bottom left result, the up arrow key to select the top right result and the down arrow key to select the bottom right result.

References

[1] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5470–5479, 2022. 1

[2] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5501–5510, 2022. 1

[3] Bruno Galerne, Lara Raad, José Lezama, and Jean-Michel Morel. Scaling Painting Style Transfer. *Computer Graphics Forum*, 2024. 1, 2

[4] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkuehler, and George Drettakis. 3d Gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4), 2023. 1

[5] Kunhao Liu, Fangneng Zhan, Muyu Xu, Christian Theobalt, Ling Shao, and Shijian Lu. StyleGaussian: Instant 3d style transfer with Gaussian splatting, 2024. 1, 18

[6] Kai Zhang, Nick Kolkin, Sai Bi, Fujun Luan, Zexiang Xu, Eli Shechtman, and Noah Snavely. ARF: Artistic radiance fields. In *ECCV 2022*, 2022. 1, 2, 18

Figure 10. Full view display of the example 1/8 of the teaser figure with the style image (size 4244×3361) displayed at the same scale as the rendered image (size 5187×3361). Images have been downscaled by a factor 2 and compressed using jpeg.

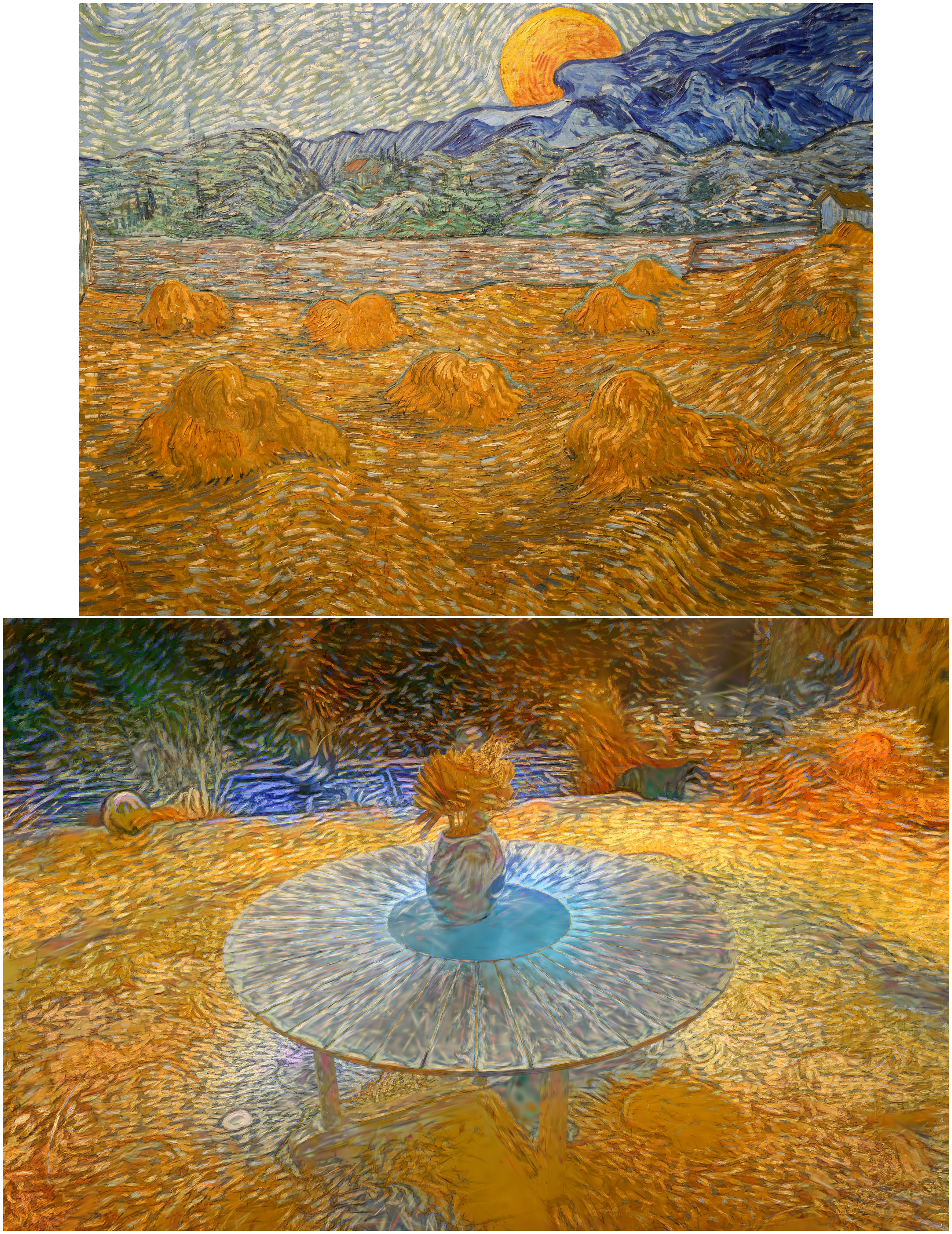

Figure 11. Full view display of the example 2/8 of the teaser figure with the style image (size 4351×3361) displayed at the same scale as the rendered image (size 5187×3361). Images have been downscaled by a factor 2 and compressed using jpeg.

Figure 12. Full view display of the example 3/8 of the teaser figure with the style image (size 4398×3361) displayed at the same scale as the rendered image (size 5187×3361). Images have been downscaled by a factor 2 and compressed using jpeg.

Figure 13. Full view display of the example 4/8 of the teaser figure with the style image (size 4398×3361) displayed at the same scale as the rendered image (size 5187×3361). Images have been downscaled by a factor 2 and compressed using jpeg.

Figure 14. Full view display of the example 5/8 of the teaser figure with the style image (size 5433×3361) displayed at the same scale as the rendered image (size 5187×3361). Images have been downscaled by a factor 2 and compressed using jpeg.

Figure 15. Full view display of the example 6/8 of the teaser figure with the style image (size 700×692) displayed at the same scale as the rendered image (size 5187×3361). Images have been downscaled by a factor 2 and compressed using jpeg.

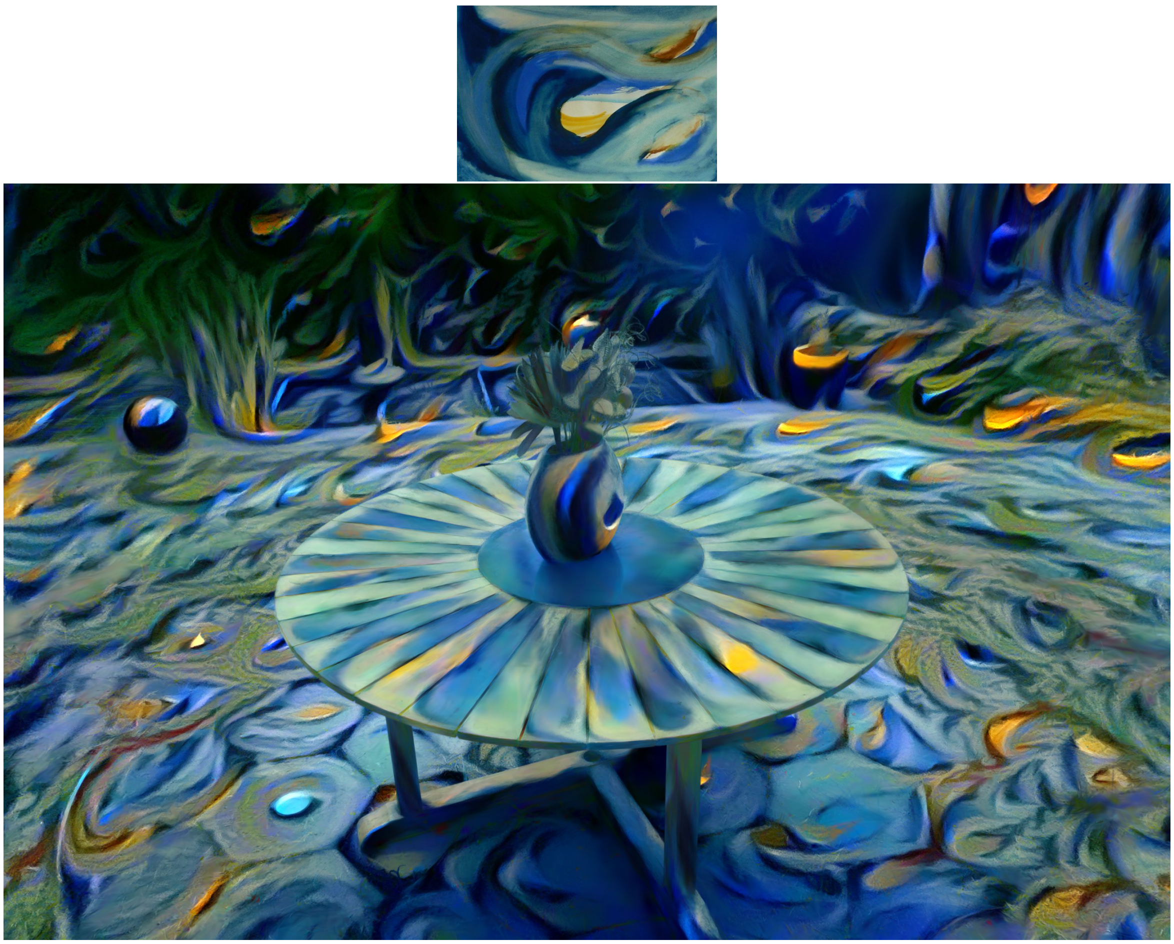

Figure 16. Full view display of the example 7/8 of the teaser figure with the style image (size 1152×781) displayed at the same scale as the rendered image (size 5187×3361). Images have been downscaled by a factor 2 and compressed using jpeg.

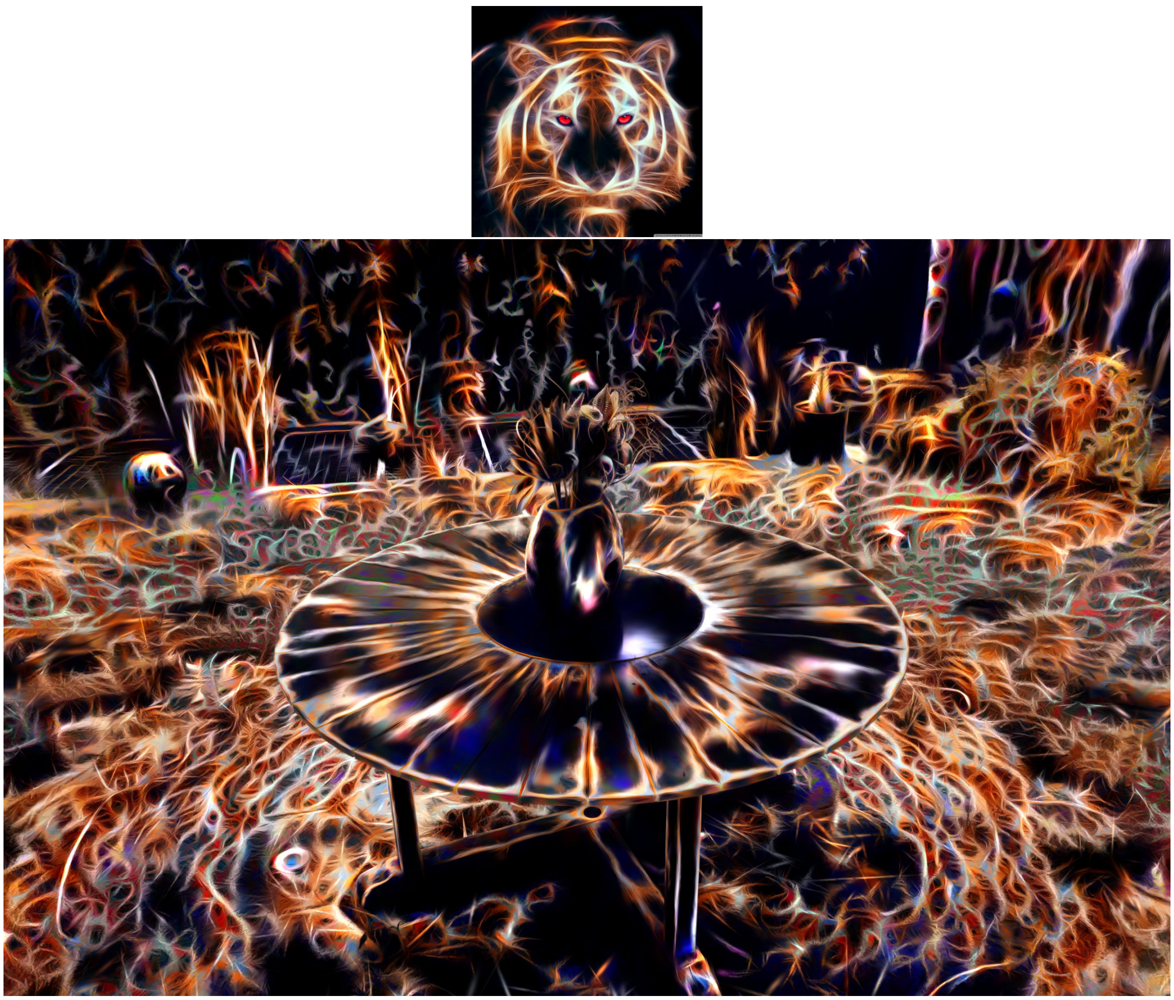

Figure 17. Full view display of the example 8/8 of the teaser figure with the style image (size 1024×1024) displayed at the same scale as the rendered image (size 5187×3361). Images have been downscaled by a factor 2 and compressed using jpeg.

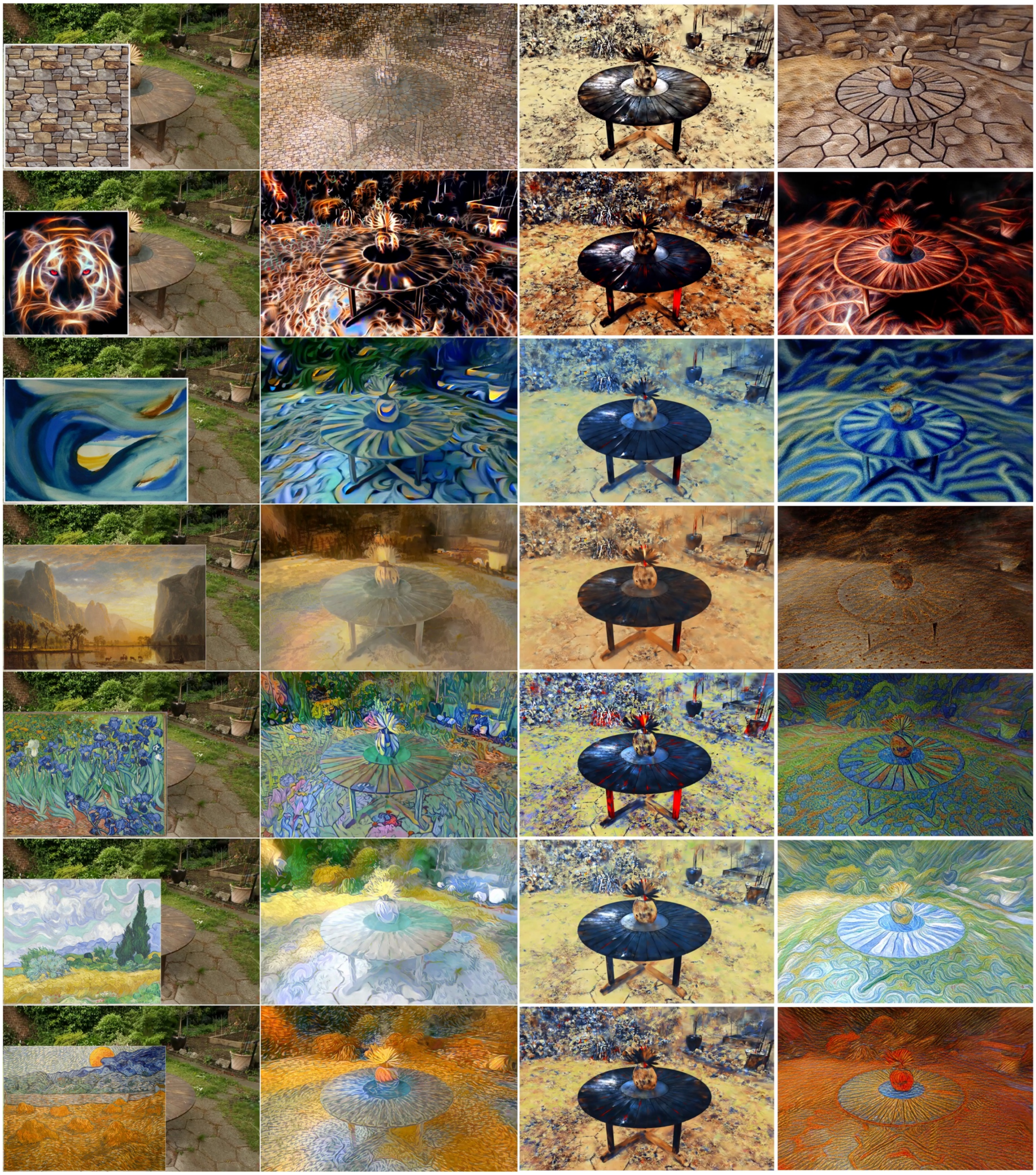

Figure 18. Comparative experiments using the garden scene. From left to right: Content and style, SGSST (ours), StyleGaussian, ARF. Content image size is 5187×3361.

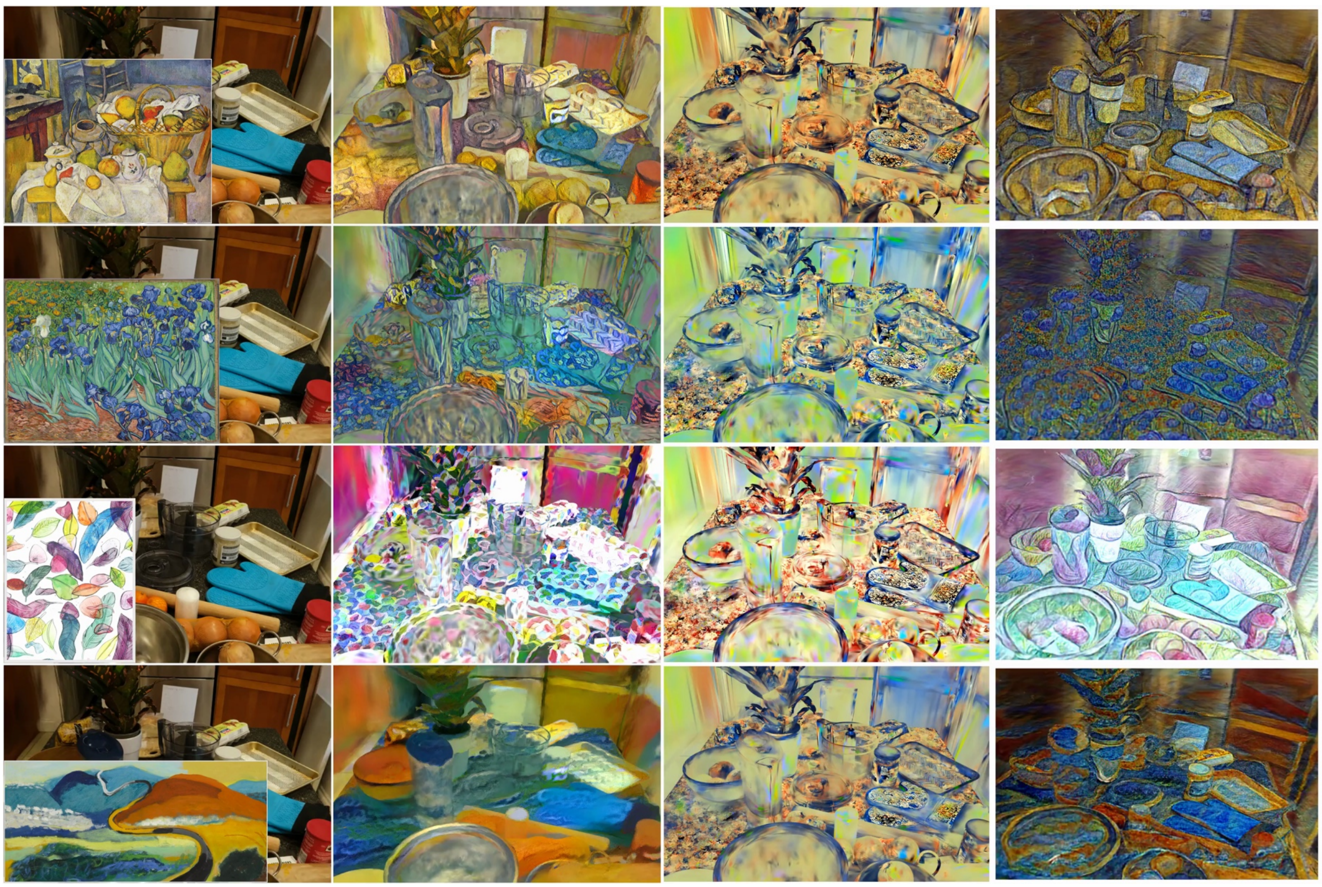

Figure 19. Comparative experiments using the counter scene. From left to right: Content and style, SGSST (ours), StyleGaussian, ARF. Content image size is 3115×2076.

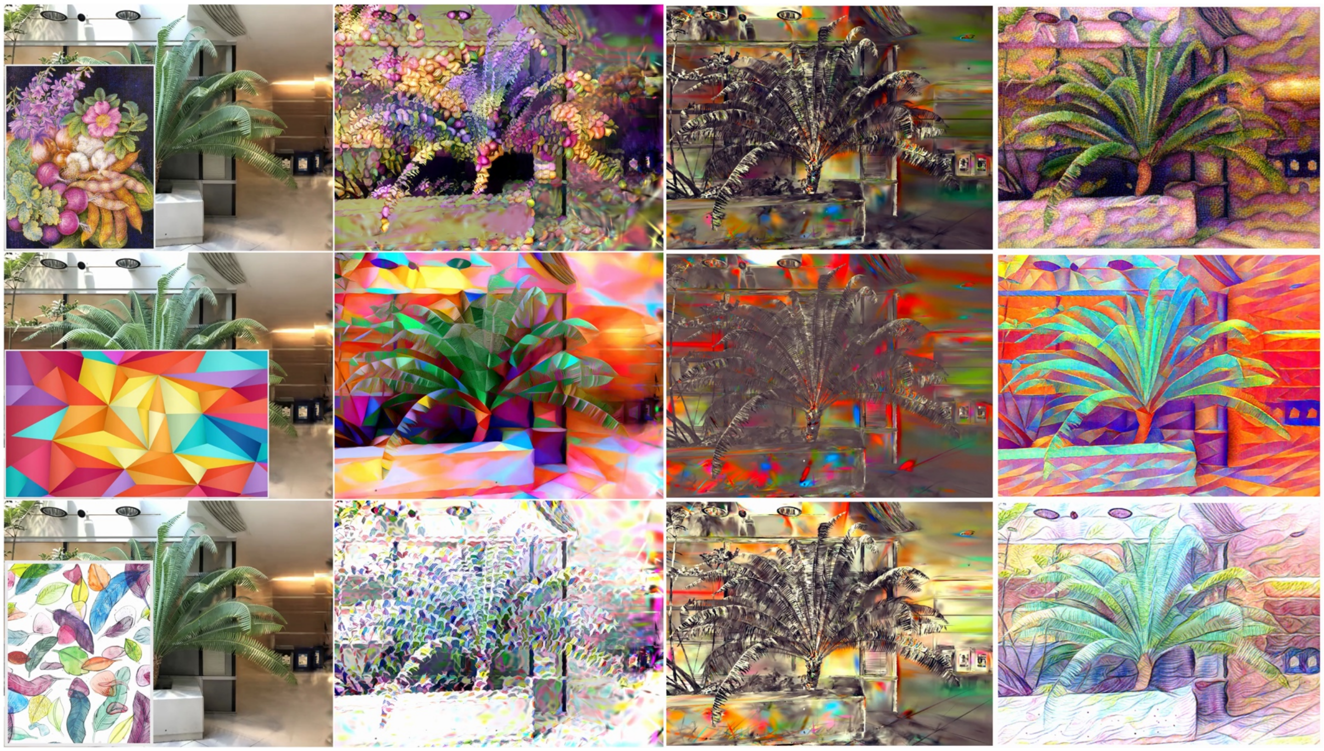

Figure 20. Comparative experiments using the fern scene. From left to right: Content and style, SGSST (ours), StyleGaussian, ARF. Content image size is 4032×3024.

Figure 21. Comparative experiments on the t-rex scene. From left to right: Content and style, SGSST (ours), StyleGaussian, ARF. Content image size is 4032×3024.

Figure 22. Comparative experiments on the kitchen scene. From left to right: Content and style, SGSST (ours), StyleGaussian, ARF. Content image size is 3115×2078.

Figure 23. Comparative experiments on the family scene. From left to right: Content and style, SGSST (ours), StyleGaussian, ARF. Content image size is 977×544.

Figure 24. Comparative experiments on the horse scene. From left to right: Content and style, SGSST (ours), StyleGaussian, ARF. Content image size is 976×544.

Figure 25. Comparative experiments on the train scene. From left to right: Content and style, SGSST (ours), StyleGaussian, ARF. Content image size is 980×545.

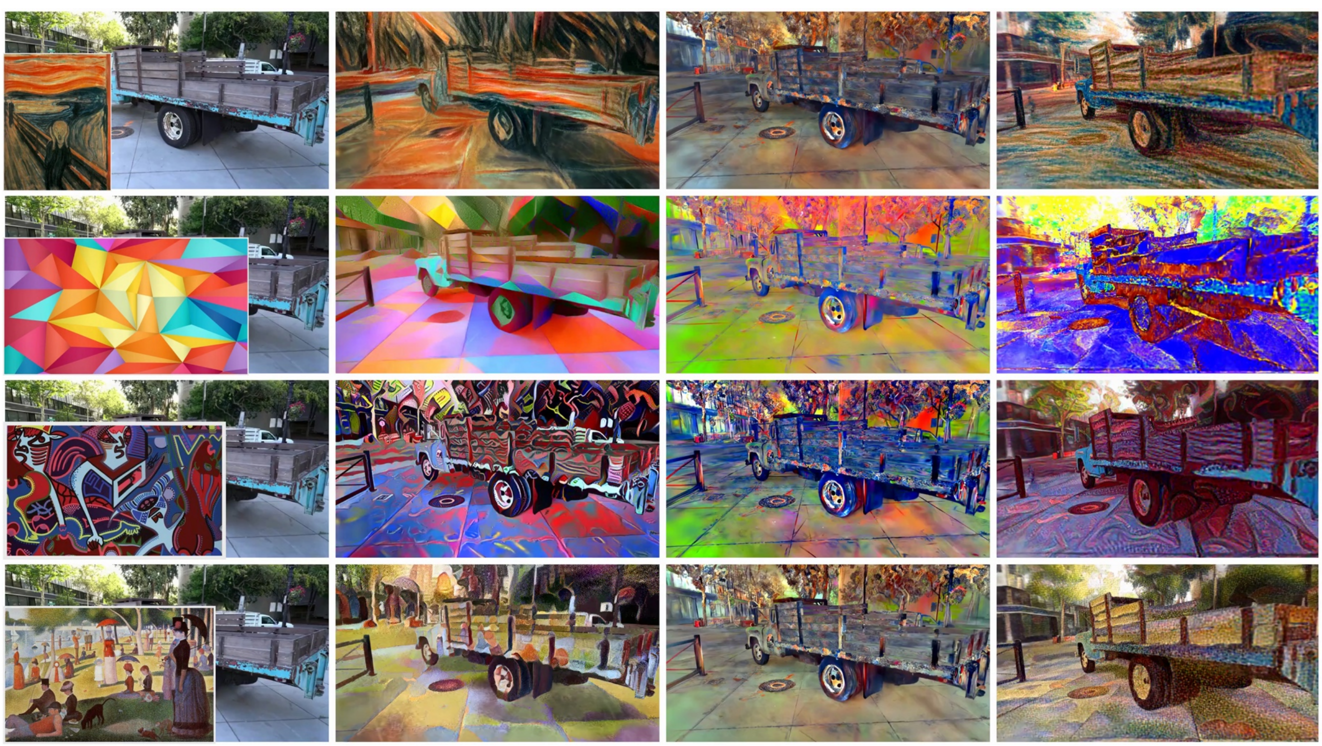

Figure 26. Comparative experiments on the truck scene. From left to right: Content and style, SGSST (ours), StyleGaussian, ARF. Content image size is 979×546.

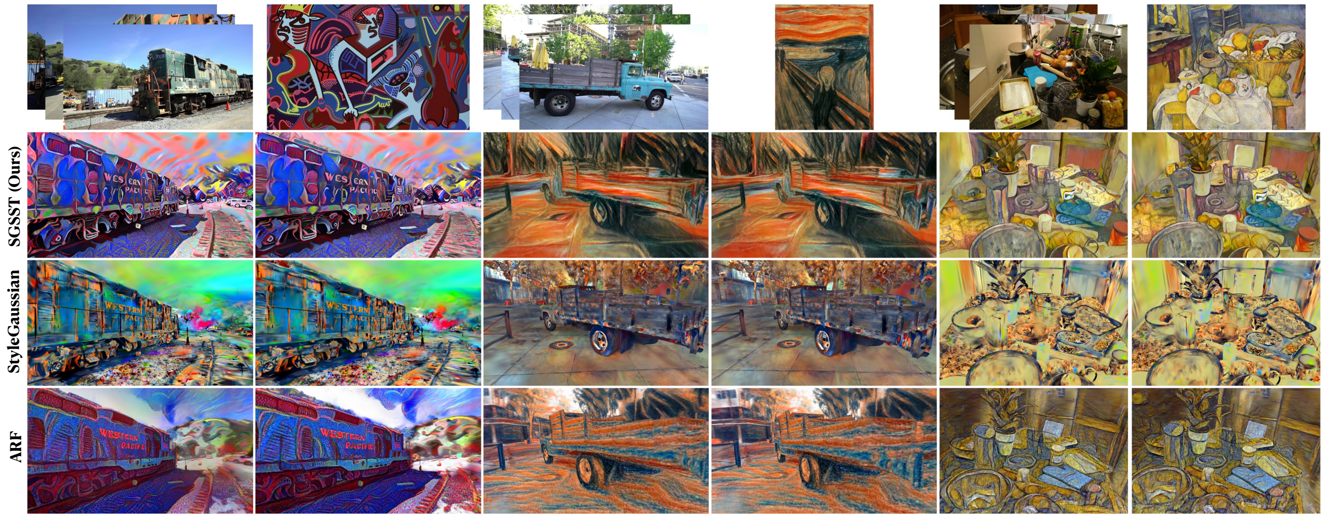

Figure 27. **Comparison of SGSST (ours, top) with StyleGaussian [5] (middle) and ARF [6] (bottom) with short range views.** From left to right the content resolutions are 980×545 (train), 979×546 (truck), and 3115×2076 (counter). For the first two examples, the various outputs keep the resolution of the content, but for the HR counter scene, the output sizes are 3115×2076 for SGSST, 1600×1066 for StyleGaussian and 779×519 for ARF (see supp. mat. for ARF results without downscaling). Thanks to its multiscale global VGG statistics, SGSST is the most faithful method regarding style consistency.

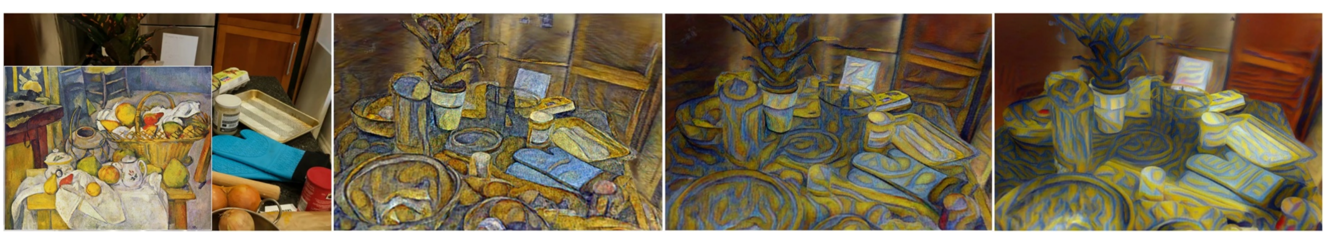

Figure 28. ARF outputs for the HR style transfer: example of the main paper with various downscaling factors. ARF produces good stylization results for inputs of moderate resolution only. From left to right: Scene and style (input size is 3115×2076), ARF result with input downscaled by 4 (size 779×519), ARF result with input downscaled by 2 (size 1557×1038), ARF result with original HR (size 3115×2076).

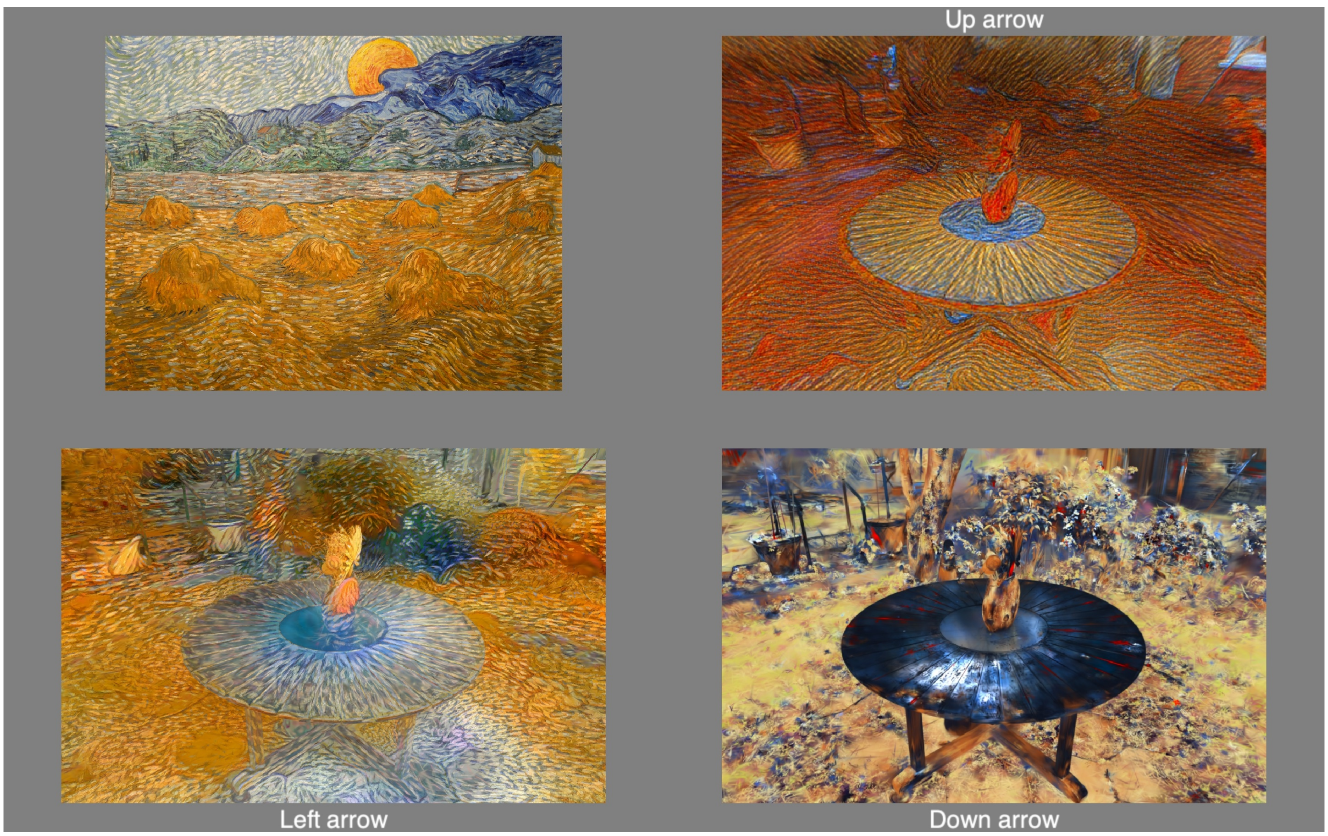

Figure 29. Perceptual study. The style input image is presented on the top left and the results of each stylization algorithm (SGSST, ARF and StyleGaussian) are presented in a random order. To select the best result, the participant has to press the key indicated next to it.