Locally Stylized Neural Radiance Fields – Supplementary Material –

Anonymous ICCV submission

Paper ID 12483

Abstract

In this supplemental document, we provide additional experiment results of our proposed method. In particular, we provide a user study, and more qualitative results on LLFF and Replica dataset.

1. User study

We conduct a user study to verify the performance of our proposed method compared with ARF [3]. The study consists of 20 questions presented in random order; each question consists of two images rendered from a scene stylized by our method; as well as the two ARF-generated images with the same scene and camera pose. The order of the two choices are randomized. In addition, the corresponding ground truth training image and the style images are also shown. The user is asked to select the choice that preserves the content of the ground truth image, and simultaneously appears similar to that of the style image. Out of the 20 questions, 12 of them correspond to images rendered from the `trex`, `room` and `fern` scenes from the LLFF [1] dataset; and 8 of them correspond to images rendered from the `office3` and `frl_apartment3` scenes in the Replica dataset [2].

We collected a total of 23 replies and the percentage of picking each method is summarized in Figure 1. The study shows that on average our method is picked at a higher percentage than ARF on both the LLFF and Replica datasets.

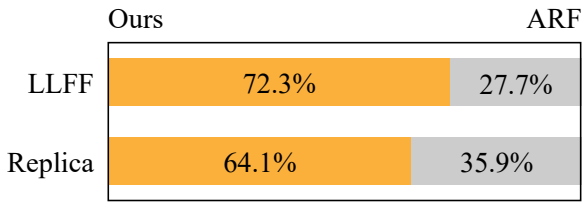

Figure 1. User study results.

2. Further qualitative results

We provide further qualitative results for the LLFF dataset in Figure 2, and for the Replica dataset in Figure 3. Both figures illustrate that our method is less likely to transfer similar, repetitive patterns to the NeRF scene. This is especially the case in low-frequency regions, e.g. concrete walls and bare surfaces.

Our algorithm for matching content-style regions work by assuming that the regions can be paired up in a meaningful sense. However, even in cases where there is little to no correlation between regions from the NeRF scene and style image, our method is able to transfer local patterns on the scene.

Finally, we provide two further examples of modifying the pairing between content and style regions in Figure 4. We demonstrate that our method can achieve diverse and customizable stylization results via adjusting the pairing.

References

[1] Ben Mildenhall, Pratul P. Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis with prescriptive sampling guidelines. *ACM Transactions on Graphics (TOG)*, 2019. 1

[2] Julian Straub, Thomas Whelan, Lingni Ma, Yufan Chen, Erik Wijmans, Simon Green, Jakob J. Engel, Raul Mur-Artal, Carl Ren, Shobhit Verma, Anton Clarkson, Mingfei Yan, Brian Budge, Yajie Yan, Xiaqing Pan, June Yon, Yuyang Zou, Kimberly Leon, Nigel Carter, Jesus Briales, Tyler Gillingham, Elias Mueggler, Luis Pesqueira, Manolis Savva, Dhruv Batra, Hauke M. Strasdat, Renzo De Nardi, Michael Goesele, Steven Lovegrove, and Richard Newcombe. The Replica dataset: A digital replica of indoor spaces. *arXiv preprint arXiv:1906.05797*, 2019. 1

[3] Kai Zhang, Nick Kolkin, Sai Bi, Fujun Luan, Zexiang Xu, Eli Shechtman, and Noah Snavely. Arf: Artistic radiance fields. In *European Conference on Computer Vision*, pages 717–733. Springer, 2022. 1

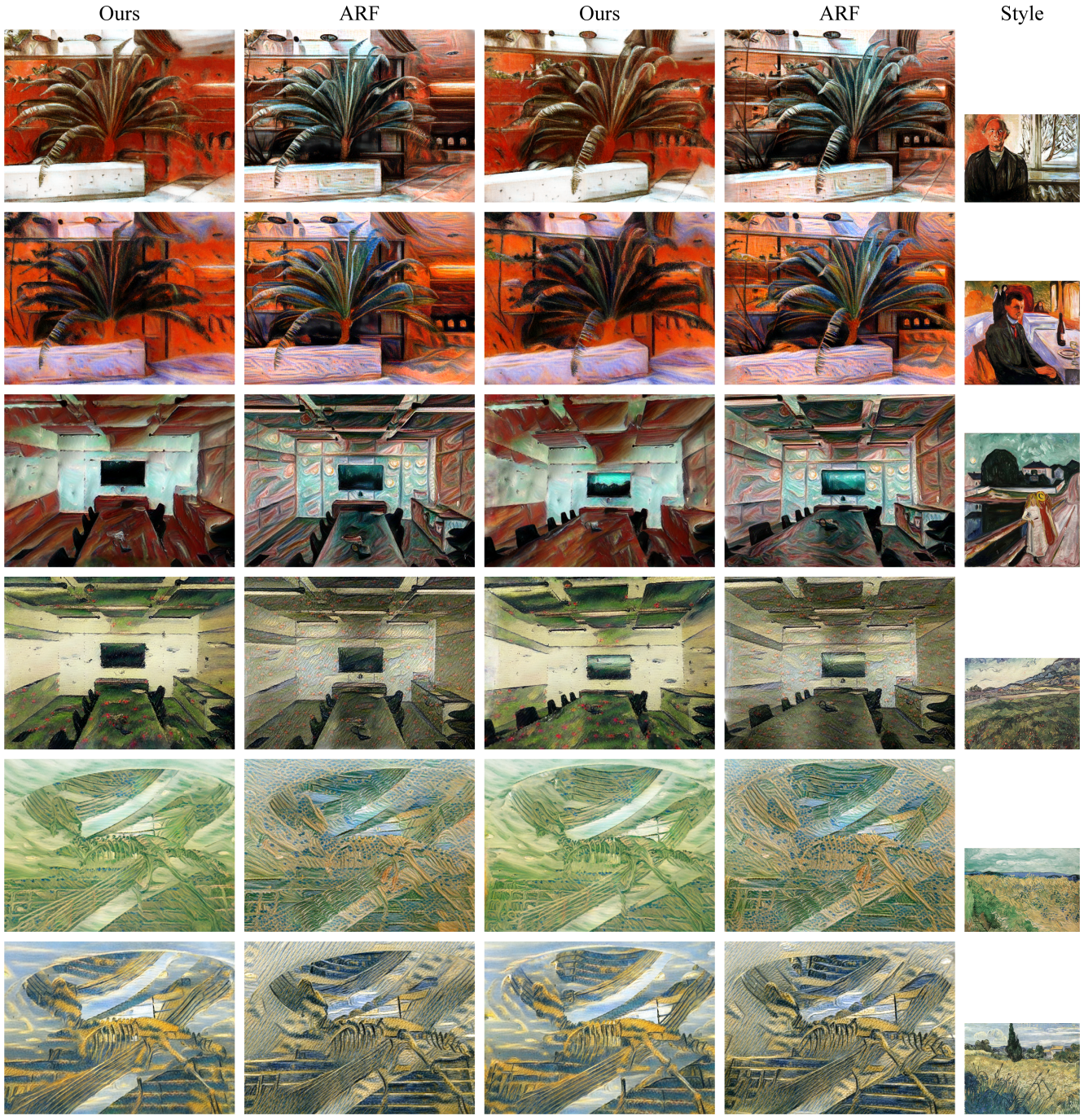

Figure 2. Further qualitative comparison on LLFF dataset.

Figure 3. Further qualitative comparison on Replica dataset.

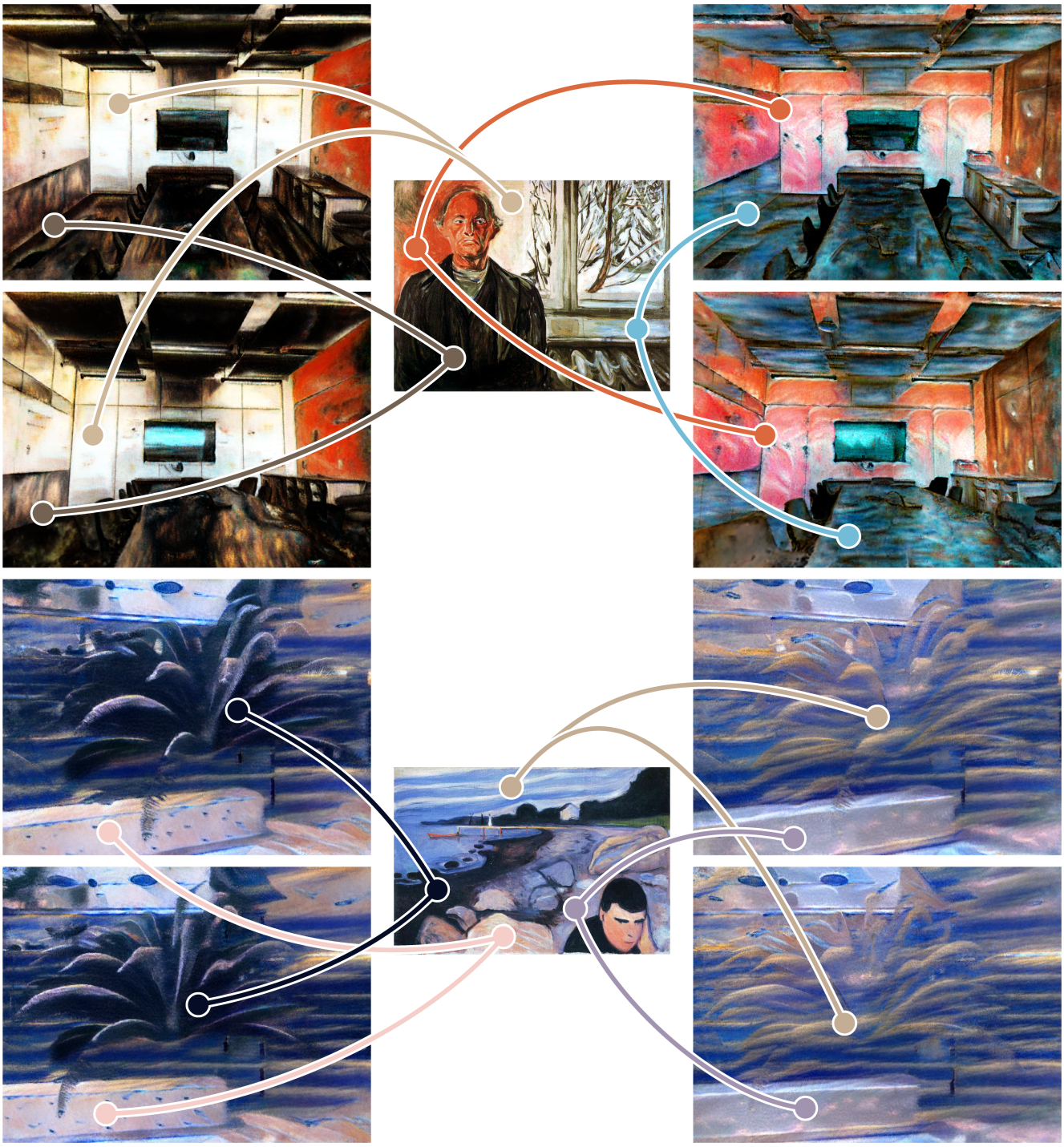

Figure 4. Effect of modifying the pairing between content and style regions. In each example, two content regions has been mapped to two different style regions in the style image (middle column), leading to two completely different stylization results (left and right columns).