ARF: Artistic Radiance Fields

Kai Zhang[1] Nick Kolkin[2] Sai Bi[2] Fujun Luan[2]
Zexiang Xu[2] Eli Shechtman[2] Noah Snavely[1]

[1]Cornell University [2]Adobe Research

Abstract. We present a method for transferring the artistic features of an arbitrary style image to a 3D scene. Previous methods that perform 3D stylization on point clouds or meshes are sensitive to geometric reconstruction errors for complex real-world scenes. Instead, we propose to stylize the more robust radiance field representation. We find that the commonly used Gram matrix-based loss tends to produce blurry results lacking in faithful style detail. We instead utilize a nearest neighbor-based loss that is highly effective at capturing style details while maintaining multi-view consistency. We also propose a novel deferred back-propagation method to enable optimization of memory-intensive radiance fields using style losses defined on full-resolution rendered images. Our evaluation demonstrates that, compared to baselines, our method transfers artistic appearance in a way that more closely resembles the style image. Please see our project webpage for video results and an open-source implementation: https://www.cs.cornell.edu/projects/arf/.

Keywords: Style transfer, neural radiance fields, 3D content creation

1 Introduction

Creating art in a specific style can require significant time and expertise. Extending an artwork to dimensions beyond the 2D image plane, such as time (in the case of animation), or 3D space (in the case of sculptures or virtual environments), involves further constraints and challenges. Hence, the styles employed by artists when moving their work beyond a static 2D canvas are constrained by the effort required to create a consistent visual experience.

We propose Artistic Radiance Fields (ARF), a new approach to transferring the artistic features from a single 2D image to a full, real-world 3D scene that yields high-quality, artistic free-viewpoint renderings. Our method converts a photorealistic radiance field [33,1,4] reconstructed from multiple images of complex, real-world scenes into a new, stylized radiance field that can be consistently rendered from different viewpoints, as shown in Fig. 1. The quality of these renderings is in contrast to that of prior 3D stylization work [16,14,34] that often suffers from geometrically inaccurate reconstructions of point clouds or triangle meshes and can lack style detail.

We formulate the stylization of radiance fields as an optimization problem: we render images of the radiance fields from multiple viewpoints in a differentiable

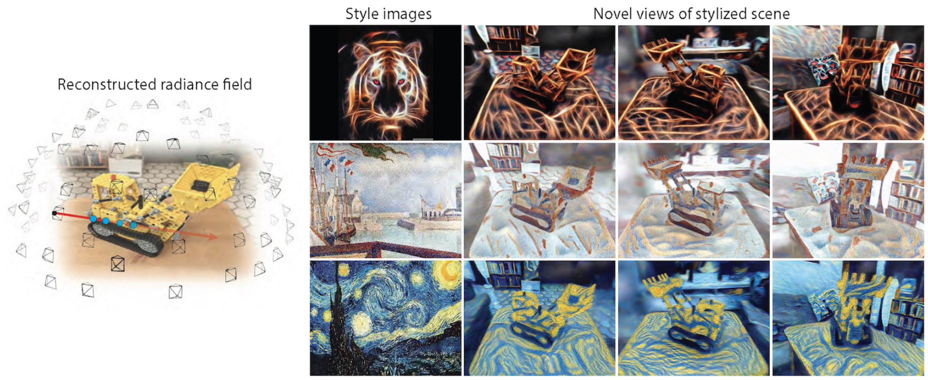

Fig. 1. We propose ARF, a new approach to 3D stylization. ARF takes a reconstructed radiance field of a real scene (1st column) and converts it into an artistic radiance field by matching features extracted from an input 2D style image (2nd column), leading to high-quality stylized novel views (3rd-5th columns). Our approach produces consistent results across viewpoints, as can be seen more clearly in the supplementary video.

manner, and minimize a content loss between the rendered stylized images and the original captured images, and a style loss between the rendered images and the style image. While prior methods [16,14,34] apply the commonly-used Gram matrix–based style loss for 3D stylization, we observe that this type of loss leads to averaged-out style details that degrade the quality of the stylized renderings.

This limitation motivates us to apply a novel style loss based on Nearest Neighbor Feature Matching (NNFM) that is better suited to the creation of high-quality 3D artistic radiance fields. In particular, for each feature vector in the VGG feature map of a rendered image, we find its nearest neighbor (NN) feature vector in the style image's VGG feature map, and seek to minimize the distance between these two feature vectors. In contrast to a Gram matrix describing global feature statistics across the entire image, NN feature matching focuses on local image descriptions, better capturing distinctive local details. Coupled with our style loss, we also enforce a VGG feature-based content loss – which balances stylization and content preservation – as well as a color transfer technique that improves the color match between our final renderings and the input style image.

One challenge is that volumetric radiance field rendering consumes significant memory. Such rendering can often only regress sparsely sampled pixels during training, and not the full images needed to compute VGG features used in many style losses. We propose a practical innovation called *deferred back-propagation* that allows us to perform optimization on high-resolution images. Deferred back-propagation enables memory-efficient auto-differentiation of scene parameters with image losses computed on full-resolution images (e.g., VGG-based style losses) by accumulating cached gradients in a patch-wise fashion.

We demonstrate that ARF can robustly transfer detailed artistic features from diverse 2D style exemplars to a variety of complex 3D scenes, yielding significantly better visual quality compared to prior methods, which tend to yield over-smoothed and blurry stylized novel views (see Figures 4, 5, and 6). In our user studies, our method is also consistently preferred over baselines.

In summary, our contributions are:

– A new radiance field-based approach to 3D scene stylization that can faithfully transfer detailed style features from a 2D image to a 3D scene, and which produces consistent stylized novel views of high visual quality.
– A finding that Nearest Neighbor Feature Matching (NNFM) loss better preserves details in the style images than the Gram matrix–based loss commonly used in prior 3D stylization work.
– A deferred back-propagation method for differentiable volumetric rendering, allowing for computation of losses on full-resolution images while significantly reducing the GPU memory footprint.

2 Related Work

In this section, we review related work to provide context for our work.

Image style transfer. Since Gatys et al. [11] introduced neural style transfer, significant progress has been made towards image stylization [25,22], image harmonization [50,29,41], color matching [44,43,28], texture synthesis [37,23,13] and beyond [18]. These style transfer approaches leverage features extracted by a pre-trained convolutional neural network (e.g., VGG-19 [40]) and optimize for a set of loss functions (typically a content loss capturing an input photo's features and a style loss matching a style image's feature statistics, e.g., as captured by a Gram matrix). Depending on whether the style transfer is achieved via iterative optimization or a single forward pass of a deep network, existing methods can be categorized as optimization-based and feed-forward-based:

Optimization-based style transfer. Gatys et al. [11] perform style transfer via iterative optimization to minimize content and style losses. Many follow-up works [7,22,37,12,21,30,25,30] have investigated alternative style loss formulations to further improve semantic consistency and high-frequency style details like brushstrokes. Unlike style transfer methods that encode the statistics of style features with a single Gram matrix, Chen and Schmidt [7], CNNMRF [22], Deep Image Analogy [25] and NNST [20] propose to search for nearest neighbors and minimize distances between features extracted from corresponding content and style patches in a coarse-to-fine fashion. These methods achieve impressive 2D stylization quality when provided with source and target images that share similar semantics. Our approach draws inspiration from this line of work and is the first to introduce nearest neighbor feature matching (NNFM) for 3D stylization. Our NNFM loss is most similar to that proposed in [20] for 2D style transfer. However, when stylizing 3D radiance fields, we find that we can achieve the same level of stylistic detail more efficiently by only applying stylization at a single coarse scale (as opposed to coarse-to-fine) and by skipping the style image augmentations (rotation and/or scaling) used in [22,20]. Compared with the contextual loss [31], our NNFM loss is simpler and avoids the need for distance and similarity normalizations.

Feed-forward style transfer. Rather than performing iterative optimization, feed-forward approaches [17,2,9,36,45,39,24] train neural networks that can transfer the style of an exemplar image to an input image using a single forward pass. While fast, these methods often struggle to faithfully reproduce style details like brushstrokes, and yield lower visual quality compared to optimization-based techniques. In the pursuit of high-quality results, we did not pursue this direction in our work.

Video style transfer. Stylizing a video by separately processing each frame with a 2D style transfer method often leads to flickering artifacts in the resulting stylized video. Video style transfer [38] methods address this problem by enforcing an additional temporal coherency loss across frames [6,15,38,42]. Alternative approaches rely on aligning and fusing style features according to their similarity to content features [10,27] to maintain temporal consistency. Despite sharing the similar challenge of consistency across views, stylizing a 3D scene is a distinct problem from video stylization, because it requires synthesizing novel views while maintaining style consistency, which is best achieved through stylization in 3D world space rather than in 2D image space.

3D style transfer. 3D style transfer aims to transform the appearance of a 3D scene so that its renderings from different viewpoints match the style of a desired image. Prior methods represent real world scenes using point clouds [16,34] or triangle meshes [46,32]. For example, Huang et al. [16] and Mu et al. [34] use featurized 3D point clouds modulated with the style image, followed by a 2D CNN renderer to produce stylized renderings. Yin et al. [46] create novel geometric and texture variations of 3D meshes by transferring the shape and texture style from one textured mesh to another. The performance of such methods is limited by the quality of these point clouds or meshes, which frequently contain noticeable artifacts when reconstructed from complex real-world scenes. In contrast, we perform style transfer on radiance fields [33,26,48,49,5] which can more faithfully reproduce the appearance of real world scenes. Work closely relevant to ours is that of Chiang et al. [8], who use neural radiance fields as a scene representation and apply pre-trained style hypernetworks for appearance stylization. However, their method produces over-smoothed and blurry stylization results, and cannot capture detailed structures in the style image such as brushstrokes, due to the limitations of pre-trained feed-forward models. Our approach can more faithfully capture distinctive details in the style exemplar while preserving recognizable scene content.

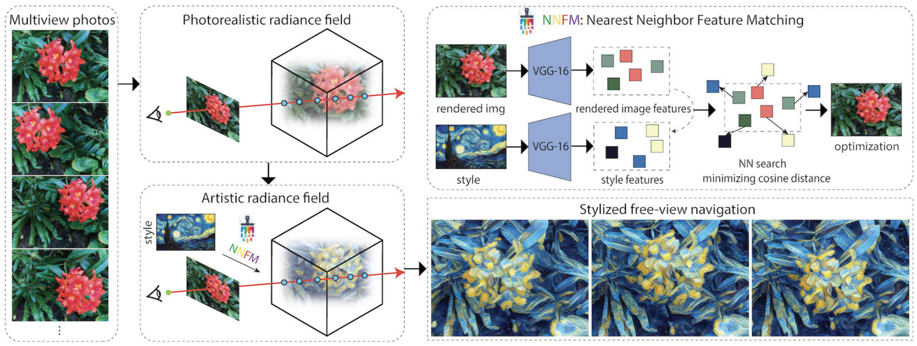

Fig. 2. Overview of our method. We first reconstruct a photo-realistic radiance field from multiple photos. We then stylize this reconstruction using an exemplar style image through the use of a nearest neighbor feature matching (NNFM) style loss. Once this stylization is done, we can obtain consistent free-viewpoint stylized renderings. We invite readers to watch the videos on our project page to better appreciate our results.

3 Background on Radiance Fields

NeRF [33] proposes neural radiance fields to model and reconstruct real scenes, achieving photo-realistic novel view synthesis results. In general, the radiance field representation can be seen as a 5D function that maps any 3D location $\mathbf{x}$ and viewing direction $\mathbf{d}$ to volume density σ and RGB color $\mathbf{c}$:

$$\sigma, \mathbf{c} = \textsc{RadianceField}(\mathbf{x}, \mathbf{d}). \tag{1}$$

This representation can be rendered from any viewpoint via differentiable volume rendering, and hence can be optimized to fit a collection of photos capture from multiple viewpoints, then later used to synthesize photo-realistic novel views. We move beyond photo-realism and add an artistic feel to the radiance field by stylizing it using an exemplar style image, such as a painting or sketch.

4 Stylizing Radiance Fields

Given a photo-realistic radiance field reconstructed from photos of a real scene, our approach transforms it into an artistic one by stylizing the 3D scene appearance with a 2D style image. We achieve this by fine-tuning the radiance field using a novel nearest neighbor feature matching style loss (Sec. 4.1) that can transfer detailed local style structures. We also introduce a deferred back-propagation technique that enables radiance field optimization with full-resolution images (Sec. 4.2) in the face of limited GPU memory. We apply a view-consistent color transfer technique to further enhance visual quality (Sec. 4.3).

4.1 Style transfer losses

Art often features unique visual details. For instance, Van Gogh's *The Starry Night* is characterized by long and curvy brushstrokes. Neural features produced by pre-trained neural networks (like VGG) can effectively capture such details and have been widely used for 2D style transfer. However, transferring such rich visual details to 3D scenes using prior VGG-based style losses is a challenge, since the style information measured by such losses are often based on global statistics that do not necessarily capture local details accurately in a view-consistent way.

To address this issue, we propose to use the ***N**earest **N**eighbor **F**eature **M**atching* (NNFM) loss to transfer complex high-frequency visual details from a 2D style image to a 3D scene (parameterized by a radiance field) in a way that yields consistency across multiple viewpoints. In particular, let $\boldsymbol{I}_{\text{style}}$ denote the style image, and $\boldsymbol{I}_{\text{render}}$ denote an image rendered from the radiance field at a selected viewpoint. We extract VGG feature maps $\boldsymbol{F}_{\text{style}}$ and $\boldsymbol{F}_{\text{render}}$ from $\boldsymbol{I}_{\text{style}}$ and $\boldsymbol{I}_{\text{render}}$, respectively. Let $\boldsymbol{F}_{\text{render}}(i,j)$ denote the feature vector at pixel location (i,j) of the feature map $\boldsymbol{F}_{\text{render}}$. Our NNFM loss can be written as:

$$\ell_{\text{nnfm}}(\boldsymbol{F}_{\text{render}}, \boldsymbol{F}_{\text{style}}) = \frac{1}{N} \sum_{i,j} \min_{i',j'} D\big(\boldsymbol{F}_{\text{render}}(i,j), \boldsymbol{F}_{\text{style}}(i',j')\big), \tag{2}$$

where N is the number of pixels in $\boldsymbol{F}_{\text{render}}$, and $D(\boldsymbol{v}_1, \boldsymbol{v}_2)$ computes the cosine distance between two vectors $\boldsymbol{v}_1, \boldsymbol{v}_2$:

$$D(\boldsymbol{v}_1, \boldsymbol{v}_2) = 1 - \frac{< \boldsymbol{v}_1, \boldsymbol{v}_2 >}{\|\boldsymbol{v}_1\|_2 \|\boldsymbol{v}_2\|_2}. \tag{3}$$

In short, for each feature in $\boldsymbol{F}_{\text{render}}$, ℓ_{nnfm} minimizes its cosine distance (Eq. (3)) to its nearest neighbor in the style image's VGG feature space ($\boldsymbol{F}_{\text{style}}$). The NNFM loss can also be viewed as a variant of the Chamfer distance for comparing high-dimensional feature sets, similar to the Chamfer L1 distance widely used for comparing 2D images and 3D point clouds.

Note that our loss does not rely on per-view global statistics. This grants more flexibility to the optimization process and focuses it on adjusting local 3D scene appearance to perceptually match patches of the style exemplar.

Controlling stylization strength. Using our NNFM loss alone can sometimes lead to overly strong stylization, making the content harder to recognize. To address this issue, we add an additional content-preserving loss penalizing the ℓ_2 difference between the feature maps of rendered and content images:

$$\ell = \ell_{\text{nnfm}}(\boldsymbol{F}_{\text{render}}, \boldsymbol{F}_{\text{style}}) + \lambda \cdot \ell_2(\boldsymbol{F}_{\text{render}}, \boldsymbol{F}_{\text{content}}), \tag{4}$$

where λ is a weight controlling stylization strength: a larger λ better preserves content, while a smaller λ leads to stronger stylization. $\boldsymbol{F}_{\text{render}}$, $\boldsymbol{F}_{\text{style}}$ and $\boldsymbol{F}_{\text{content}}$ are feature maps of a rendered image, the style image and a content image, respectively; they are all extracted using exactly the same VGG feature extractor. (See Sec. 4.4 for more detail.)

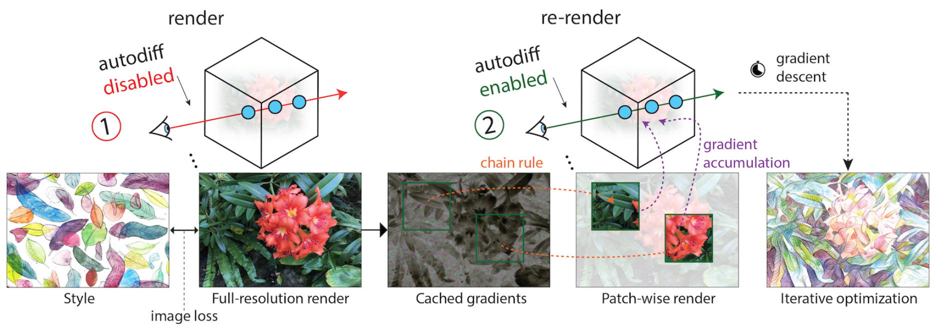

Fig. 3. Illustration of deferred back-propagation. We first disable auto-differentiation to render a full-resolution image, then compute the image loss (e.g., a style loss defined by NNFM or by a Gram matrix), and cache its gradients with respect to pixel values of the full-resolution image. Next, we back-propagate the cached gradients to scene parameters and accumulate in a patch-wise manner: For each patch, we re-render it with auto-differentiation enabled, and apply the chain rule to back-propagate the corresponding cached patch gradients to scene parameters for accumulation. This way, we correctly compute the gradients of a loss imposed on the full-resolution rendered image with respect to the scene parameters, with the same GPU memory footprint of rendering a single small patch differentiably.

4.2 Deferred back-propagation

We stylize a radiance field by minimizing our loss (Eqn. 4) imposed on images rendered using differentiable volume rendering. Such rendering is very memory-inefficient in practice, because the color of each pixel is composited from a large number of samples along the corresponding ray. As a result, rather than rendering a full-resolution image at each optimization step, many methods randomly sample a sparse set of pixels for rendering. While such sparse pixel sampling is a reasonable strategy when minimizing a loss computed independently per-pixel, such as an ℓ_1/ℓ_2 loss, it does not work for complex CNN-based losses, such as our NNFM loss or a Gram-matrix style loss, which require full-resolution renderings.

We propose a simple technique we call *deferred back-propagation* that can directly optimize on full-resolution images, allowing for more sophisticated and powerful image losses to be used in conjunction with radiance fields. As shown by Fig. 3, we first render a full-resolution image with auto-differentiation disabled, then compute the image loss and its gradient with respect to the rendered image's pixel colors, which produces a cached gradient image. Finally, in a *patch-wise* manner, we re-render the pixel colors with auto-differentiation enabled, and back-propagate the cached gradients to the scene parameters for accumulation. In this way, gradient back-propagation is deferred from the full-resolution image rendering stage to the patch-wise re-rendering stage, reducing the GPU memory cost from that of rendering a full-resolution image to that of rendering a small patch. While this technique is general, we apply it to our stylization task when optimizing our style loss, and also when optimizing the Gram-matrix loss in our experimental comparisons (see Fig. 7).

4.3 View-consistent color transfer

While our style and content losses can perceptually transfer styles and preserve the original content, we find they can lead to color mismatches between rendered images and the style image. We devise a simple technique to address this issue that leads to enhanced visual quality (see Fig. 7). We first recolor the training views via 2D color transfer from the style image. These recolored images are used to pre-optimize our artistic radiance field as initialization for our stylization optimization; they are also used for our content preservation loss (Eq. 4). Additionally, after the 3D stylization process, we compute another color transfer transform on images rendered to the training viewpoints, and apply this same color transform to the color values produced from rendering the radiance fields.

For color transfer we adopt a simple affine transformation of colors in RGB space, the parameters of which are estimated by matching color statistics of an image set to those of an image. Specifically, let $\{\boldsymbol{c}_i\}_{i=1}^m$ be the set of all pixel colors in an image set to be recolored, and let $\{\boldsymbol{s}_i\}_{i=1}^n$ be the set of all pixel colors of the style image. We analytically solve for an affine transformation $\boldsymbol{A}$ such that $\mathrm{E}[\boldsymbol{A}\boldsymbol{c}] = \mathrm{E}[\boldsymbol{s}]$ and $\mathrm{Cov}[\boldsymbol{A}\boldsymbol{c}] = \mathrm{Cov}[\boldsymbol{s}]$, i.e., the mean and covariance of the color-transformed image set should match those of the style image. We refer the readers to our supplemental document for derivation of $\boldsymbol{A}$.

4.4 Implementation details

To represent a radiance field, our work primarily uses Plenoxels [47] for its fast reconstruction and rendering speed. However, our framework is agnostic to the radiance field representation. To demonstrate this flexibility, we also apply ARF to stylize NeRF [33] and TensoRF [47] representations, and in each case achieve high visual quality with faithful style transfer, as shown in Fig. 8.

During stylization, we fix the density component of the initial photorealistic radiance field, and only optimize the appearance component when converting to an artistic radiance field. We also discard any view-dependent appearance modelling. To extract the feature maps $\boldsymbol{F}_{\text{render}}$, $\boldsymbol{F}_{\text{style}}$, and $\boldsymbol{F}_{\text{content}}$ in Eq. 4, we use a pretrained VGG-16 network that consists of 5 layer blocks: `conv1`, `conv2`, `conv3`, `conv4`, `conv5`. We use the `conv3` block as the feature extractor, because we empirically find that it captures style details better than the other blocks. We set the content-preserving weight $\lambda = 0.001$ in Eq. 4 for all forward-facing captures, and $\lambda = 0.005$ for all 360° captures. We refer the readers to our supplemental document for more implementation details.

5 Experiments

We evaluate our method by performing both qualitative and quantitative comparisons to baseline methods. We show stylization results for various real-world scenes guided by different style images. Our experimental results show that our method significantly outperforms baseline methods by generating stylized renderings that are more faithful to the input style image, while maintaining the recognizable semantic and geometric features of the original scene. We invite readers to our project page for better assessment of 3D stylization quality.

Datasets. We conduct extensive experiments on multiple real-world scenes including four forward-facing captures: *Flower*, *Orchids*, *Horns*, and *Trex*, from [33], and seven 360° captures: *Family*, *Horse*, *Playground*, *Truck*, *M60*, and *Train* from the Tanks and Temples dataset [19], as well as the *Real Lego* dataset from [1]. All scenes contain complex structures and intricate details that are difficult to reconstruct with prior triangle mesh or point cloud-based methods. We also experiment with a diverse set of style images including a neon tiger, Van Gogh's *The Starry Night*, sketches, etc., to test our method's ability to handle a diverse range of style exemplars.

Baselines. We compare our method to state-of-the-art methods [16,8] for 3D style transfer quality. Specifically, Huang et al. [16] adopt point clouds featurized by VGG features averaged across views as a scene representation, and transform the pointwise features by modulating them with the encoding vector of a style image for stylization. Chiang et al. [8] use implict MLPs as in NeRF++ [49] to reconstruct a radiance field for a scene, then update the weights of the radiance prediction branch using a hypernetwork that takes a style image as input. For both methods, we use their released code and pre-trained models. We did not compare to off-the-shelf video stylization methods, because prior work has demonstrated that they are less competitive compared to 3D style transfer approaches [16,8].

Qualitative comparisons. We show visual comparisons between methods in Fig. 4 (forward-facing captures) and Fig. 5 (360° captures). Visually, we see that our results exhibit a better style match to the exemplar image compared to the baselines. For instance, in the *Flower* scene in Fig. 4, our method faithfully captures both the color tone and the brushstrokes of *The Starry Night*, while the baseline method of Huang et al. [16] generates over-smoothed results without detailed structures. Moreover, Huang et al. also fails to recover complex geometric structures such as plant leaves due to inaccuracies in the reconstructed meshes. In comparison, our method effectively reconstructs and preserves the geometric and semantic content of the original scene, thanks to the more robust radiance fields representation.

Chiang et al. [8] only transfers the overall color tone of the style image to the scene and fails to recover the rich details that our method does. For example, in the *Family* statue scene in Fig. 5, our method captures the subtle textural details of the watercolor feather style image, and reproduces them in the stylized renderings. In contrast, the method of Chiang et al. [8] generates blurry results without such intricate structures, because their hypernetwork is trained on a fixed dataset of style images and can fail to capture the details of an unseen style input. Our method benefits from both the optimization-based framework as well as our NNFM style loss, which greatly boost 3D stylization quality.

We show additional results from our method in Fig. 6. Our method is robust to different scenes with varying levels of complexity and generates consistently superior results under a variety of styles.

User study. We also perform a user study to compare our method to prior work. A user is presented with a sequence of stylization results, where for each result the user is shown a style image, a video of the original scene, and two corresponding stylized videos, one produced with our method and one by a baseline. The user is then asked to select the result whose style better matches that of the given style image. In total, we collected ratings covering 25 randomly selected (scene, style) pairs. We divided the questions into 5 batches, each with 5 questions, and asked a group of users to rate a randomly selected batch. We collected an average of $\sim$12 ratings for each individual pair. We found that users preferred our method over the method of Huang et al. [16] 86.8% of the time, and over that of Chiang et al. [8] 94.1% of the time. These results show a clear preference for our method.

Ablations. We perform ablation studies to justify our design choices. We first compare our NNFM loss to the prior Gram matrix–based and CNNMRF losses. As we can see from Fig. 7, our NNFM loss generates significantly better results and more faithfully preserves the style details of the example images compared to the other two losses. In Fig. 7, we also validate the necessity of the color transfer stage. Without color transfer, the generated results tend to have different color tones compared to the style image, leading to a degraded style match. Our color transfer method effectively addresses this issue. Finally, we perform an ablation of using the feature map at different layers of the VGG-16 network for computing the NNFM loss. We find that our choice of the `conv3` layer block preserves stylistic details better than other layers. We refer the readers to our supplemental document for the results of this ablation experiment, as well as an ablation of using different color spaces for the color transfer method in Sec. 4.3.

Limitations. Our method has a few limitations. First, geometric artifacts, e.g., floaters, in the radiance fields can cause artifacts in both the photorealistic and our stylized renderings. Such floaters can be removed by adding additional regularizers on volume density to the loss during optimization [3,35]. Second, although our artistic radiance fields can be rendered in real time once optimized, a relatively time-consuming optimization procedure is still required for every style image ($\sim$3 mins for forward-facing captures, and $\sim$20 mins for 360 captures on a single NVIDIA RTX 3090 GPU). Third, our reconstructed artistic radiance fields do not support manual editing. Enabling artists to interactively edit them is highly desirable for the sake of facilitating creativity.

6 Conclusion

We have presented a method to create artistic radiance fields from photorealistic radiance fields given user-specified style exemplars. The reconstructed artistic radiance fields can then be used to render high-quality stylized novel views that faithfully mimic the input style image in terms of color tone and style details like

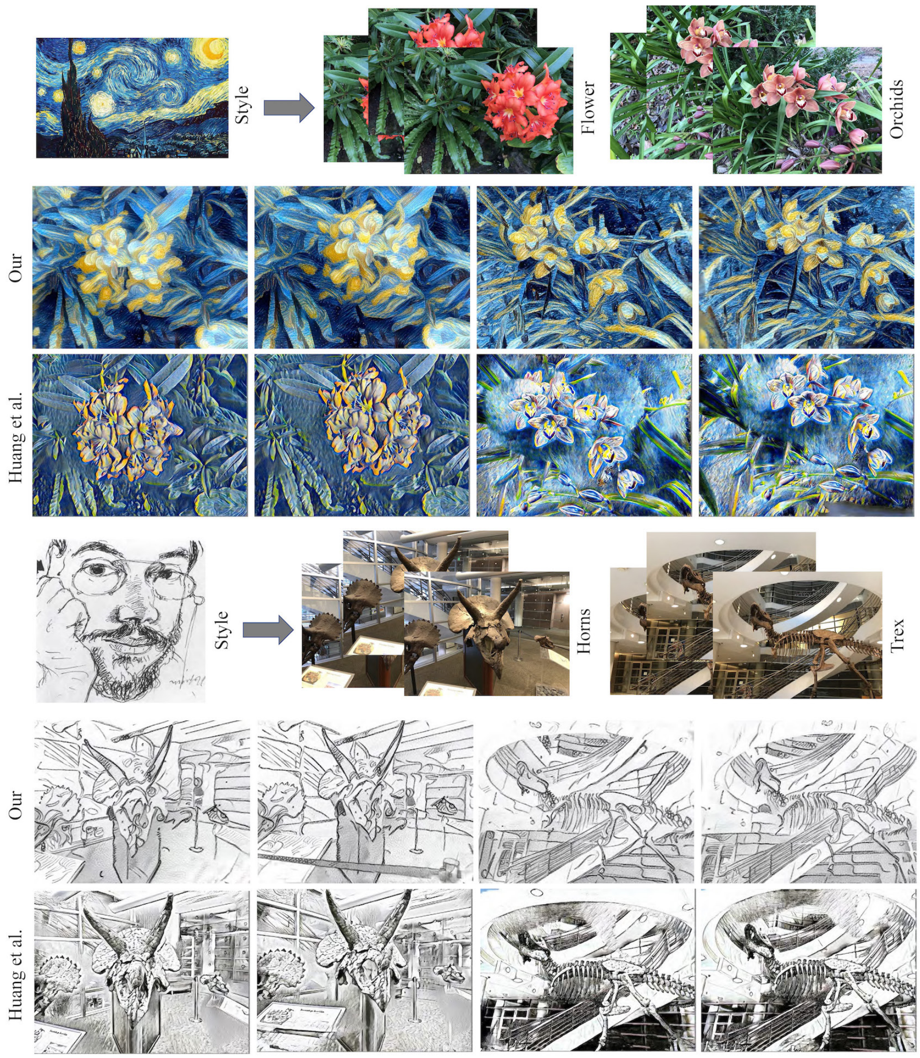

Fig. 4. Comparison with the baseline method Huang et al. [16] on real-world forward-facing data. Our stylized novel views contain significantly more faithful style details. Additionally, Huang et al. [16] requires reconstructing meshes from images, an error-prone process. Such errors can impact the quality of stylized novel views, as can be seen in the leaves in the *Flower* and *Orchids* scenes, and in ceiling of *Trex*. In contrast, our method, based on radiance fields, exhibits many fewer geometric artifacts.

Fig. 5. Comparison with the baseline methods Huang et al. [16] and Chiang et al. [8] on real-world Tanks and Temples data. Our results match both the colors and details of the style image most faithfully, while preserving sharp and recognizable content.

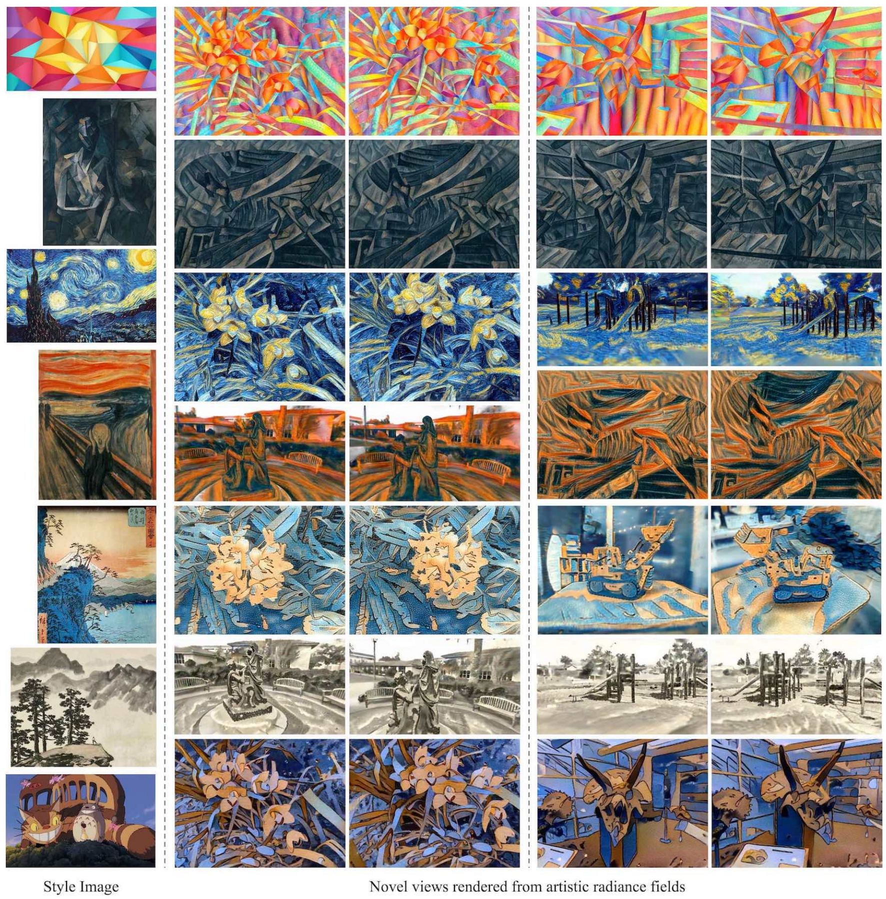

Fig. 6. Our method can generate compelling results for a wide range of (real-world scene, style image) pairs. The leftmost image in each row is the style image, and the rest are stylized novel views rendered from corresponding artistic radiance fields (two novel views are shown for each artistic radiance field).

Fig. 7. Ablation studies of color transfer and NNFM loss. Without color transfer, there is a noticeable color mismatch between the synthesized views and the style image (shown as insets in the first column). Replacing the NNFM loss with the commonly-used Gram loss [11] leads to less compelling results with many more artifacts. Our NNFM loss also generates more faithful 3D stylization results than the prior CNNMRF loss [22].

Fig. 8. Applicability across different radiance field representations. Our ARF method is applicable to a variety of radiance fields representations, including Plenoxels [47], NeRF [33] and TensoRF [4], in each case producing high-quality 3D stylization results.

brushstrokes. This enables an immersive experience of artistic 3D scenes. Key to our method's success is the proposed coupling of the nearest neighbor featuring matching loss and view-consistent color transfer, rather than the commonly-used Gram loss. We demonstrate that our method achieves superior 3D stylization quality over baselines through evaluations across various 3D scenes and 2D styles.

Acknowledgements. We would like to thank Adobe artist Daichi Ito for helpful discussions about 3D artistic styles.

References

1. Alex Yu and Sara Fridovich-Keil, Tancik, M., Chen, Q., Recht, B., Kanazawa, A.: Plenoxels: Radiance fields without neural networks (2021)
2. An, J., Huang, S., Song, Y., Dou, D., Liu, W., Luo, J.: Artflow: Unbiased image style transfer via reversible neural flows. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 862–871 (2021)
3. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Mip-nerf 360: Unbounded anti-aliased neural radiance fields. CVPR (2022)
4. Chen, A., Xu, Z., Geiger, A., Yu, J., Su, H.: Tensorf: Tensorial radiance fields. arXiv preprint arXiv:2203.09517 (2022)
5. Chen, A., Xu, Z., Zhao, F., Zhang, X., Xiang, F., Yu, J., Su, H.: MVSNeRF: Fast generalizable radiance field reconstruction from multi-view stereo. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14124–14133 (2021)
6. Chen, D., Liao, J., Yuan, L., Yu, N., Hua, G.: Coherent online video style transfer. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1105–1114 (2017)
7. Chen, T.Q., Schmidt, M.: Fast patch-based style transfer of arbitrary style. arXiv preprint arXiv:1612.04337 (2016)
8. Chiang, P.Z., Tsai, M.S., Tseng, H.Y., Lai, W.S., Chiu, W.C.: Stylizing 3d scene via implicit representation and hypernetwork. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 1475–1484 (2022)
9. Chiu, T.Y., Gurari, D.: Iterative feature transformation for fast and versatile universal style transfer. In: European Conference on Computer Vision. pp. 169–184. Springer (2020)
10. Deng, Y., Tang, F., Dong, W., Huang, H., Ma, C., Xu, C.: Arbitrary video style transfer via multi-channel correlation. arXiv preprint arXiv:2009.08003 (2020)
11. Gatys, L.A., Ecker, A.S., Bethge, M.: Image style transfer using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2414–2423 (2016)
12. Gu, S., Chen, C., Liao, J., Yuan, L.: Arbitrary style transfer with deep feature reshuffle. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8222–8231 (2018)
13. Heitz, E., Vanhoey, K., Chambon, T., Belcour, L.: A sliced wasserstein loss for neural texture synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9412–9420 (2021)
14. Höllein, L., Johnson, J., Niessner, M.: StyleMesh: Style transfer for indoor 3D scene reconstructions. arXiv preprint arXiv:2112.01530 (2021)
15. Huang, H., Wang, H., Luo, W., Ma, L., Jiang, W., Zhu, X., Li, Z., Liu, W.: Real-time neural style transfer for videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 783–791 (2017)
16. Huang, H.P., Tseng, H.Y., Saini, S., Singh, M., Yang, M.H.: Learning to stylize novel views. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13869–13878 (2021)
17. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE international conference on computer vision. pp. 1501–1510 (2017)
18. Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., Song, M.: Neural style transfer: A review. IEEE transactions on visualization and computer graphics **26**(11), 3365–3385 (2019)

19. Knapitsch, A., Park, J., Zhou, Q.Y., Koltun, V.: Tanks and temples: Benchmarking large-scale scene reconstruction. ACM Transactions on Graphics **36**(4) (2017)
20. Kolkin, N., Kucera, M., Paris, S., Sykora, D., Shechtman, E., Shakhnarovich, G.: Neural neighbor style transfer. arXiv e-prints pp. arXiv–2203 (2022)
21. Kolkin, N., Salavon, J., Shakhnarovich, G.: Style transfer by relaxed optimal transport and self-similarity. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10051–10060 (2019)
22. Li, C., Wand, M.: Combining markov random fields and convolutional neural networks for image synthesis. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2479–2486 (2016)
23. Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., Yang, M.H.: Diversified texture synthesis with feed-forward networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3920–3928 (2017)
24. Li, Y., Fang, C., Yang, J., Wang, Z., Lu, X., Yang, M.H.: Universal style transfer via feature transforms. Advances in neural information processing systems **30** (2017)
25. Liao, J., Yao, Y., Yuan, L., Hua, G., Kang, S.B.: Visual attribute transfer through deep image analogy. ACM Trans. Graph. (2017)
26. Liu, L., Gu, J., Zaw Lin, K., Chua, T.S., Theobalt, C.: Neural sparse voxel fields. Advances in Neural Information Processing Systems **33**, 15651–15663 (2020)
27. Liu, S., Lin, T., He, D., Li, F., Wang, M., Li, X., Sun, Z., Li, Q., Ding, E.: Adaattn: Revisit attention mechanism in arbitrary neural style transfer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6649–6658 (2021)
28. Luan, F., Paris, S., Shechtman, E., Bala, K.: Deep photo style transfer. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4990–4998 (2017)
29. Luan, F., Paris, S., Shechtman, E., Bala, K.: Deep painterly harmonization. Computer Graphics Forum **37**(4), 95–106 (2018)
30. Mechrez, R., Talmi, I., Zelnik-Manor, L.: The contextual loss for image transformation with non-aligned data. In: Proceedings of the European conference on computer vision (ECCV). pp. 768–783 (2018)
31. Mechrez, R., Talmi, I., Zelnik-Manor, L.: The contextual loss for image transformation with non-aligned data. In: ECCV (2018)
32. Michel, O., Bar-On, R., Liu, R., Benaim, S., Hanocka, R.: Text2mesh: Text-driven neural stylization for meshes. arXiv preprint arXiv:2112.03221 (2021)
33. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: European conference on computer vision. pp. 405–421. Springer (2020)
34. Mu, F., Wang, J., Wu, Y., Li, Y.: 3d photo stylization: Learning to generate stylized novel views from a single image. arXiv preprint arXiv:2112.00169 (2021)
35. Niemeyer, M., Barron, J.T., Mildenhall, B., Sajjadi, M.S.M., Geiger, A., Radwan, N.: Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2022)
36. Park, D.Y., Lee, K.H.: Arbitrary style transfer with style-attentional networks. In: proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5880–5888 (2019)
37. Risser, E., Wilmot, P., Barnes, C.: Stable and controllable neural texture synthesis and style transfer using histogram losses. arXiv preprint arXiv:1701.08893 (2017)
38. Ruder, M., Dosovitskiy, A., Brox, T.: Artistic style transfer for videos and spherical images. International Journal of Computer Vision **126**(11), 1199–1219 (2018)

39. Sheng, L., Lin, Z., Shao, J., Wang, X.: Avatar-net: Multi-scale zero-shot style transfer by feature decoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8242–8250 (2018)
40. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
41. Tsai, Y.H., Shen, X., Lin, Z., Sunkavalli, K., Lu, X., Yang, M.H.: Deep image harmonization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3789–3797 (2017)
42. Wang, W., Xu, J., Zhang, L., Wang, Y., Liu, J.: Consistent video style transfer via compound regularization. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 12233–12240 (2020)
43. Xia, X., Xue, T., Lai, W.s., Sun, Z., Chang, A., Kulis, B., Chen, J.: Real-time localized photorealistic video style transfer. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 1089–1098 (2021)
44. Xia, X., Zhang, M., Xue, T., Sun, Z., Fang, H., Kulis, B., Chen, J.: Joint bilateral learning for real-time universal photorealistic style transfer. In: European Conference on Computer Vision. pp. 327–342. Springer (2020)
45. Yao, Y., Ren, J., Xie, X., Liu, W., Liu, Y.J., Wang, J.: Attention-aware multi-stroke style transfer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1467–1475 (2019)
46. Yin, K., Gao, J., Shugrina, M., Khamis, S., Fidler, S.: 3dstylenet: Creating 3d shapes with geometric and texture style variations. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12456–12465 (2021)
47. Yu, A., Li, R., Tancik, M., Li, H., Ng, R., Kanazawa, A.: Plenoctrees for real-time rendering of neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5752–5761 (2021)
48. Yu, A., Ye, V., Tancik, M., Kanazawa, A.: pixelnerf: Neural radiance fields from one or few images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4578–4587 (2021)
49. Zhang, K., Riegler, G., Snavely, N., Koltun, V.: Nerf++: Analyzing and improving neural radiance fields. arXiv preprint arXiv:2010.07492 (2020)
50. Zhang, L., Wen, T., Shi, J.: Deep image blending. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 231–240 (2020)