

Deforming Radiance Fields with Cages

Tianhan Xu¹ and Tatsuya Harada^{1,2}

¹ The University of Tokyo

² RIKEN

{tianhan.xu, harada}@mi.t.u-tokyo.ac.jp

Abstract. Recent advances in radiance fields enable photorealistic rendering of static or dynamic 3D scenes, but still do not support explicit deformation that is used for scene manipulation or animation. In this paper, we propose a method that enables a new type of deformation of the radiance field: free-form radiance field deformation. We use a triangular mesh that encloses the foreground object called *cage* as an interface, and by manipulating the cage vertices, our approach enables the free-form deformation of the radiance field. The core of our approach is cage-based deformation which is commonly used in mesh deformation. We propose a novel formulation to extend it to the radiance field, which maps the position and the view direction of the sampling points from the deformed space to the canonical space, thus enabling the rendering of the deformed scene. The deformation results of the synthetic datasets and the real-world datasets demonstrate the effectiveness of our approach. Project page: <https://xth430.github.io/deforming-nerf/>.

Keywords: Scene representation · Radiance field · Scene manipulation · Cage-based deformation · Free-form deformation

1 Introduction

Photorealistic free-view rendering has recently received increasing attention for its various real-world applications such as virtual reality, augmented reality, games, and movies. Recently, neural scene representations [23, 28, 30, 40] have shown better capability to capture both geometry and appearance that exceed traditional structure-from-motion [13, 44] or image-based rendering [4, 10]. The most representative work is Neural Radiance Field (NeRF) [28], which represents the static 3D scene as a radiance field and uses a neural network to encode the volume density and the view-dependent radiance color. With volume rendering [19], NeRF can achieve photorealistic rendering from an arbitrary viewpoint. Subsequent works extended NeRF to support modeling dynamic scenes [35, 36, 39, 43], dark scenes [27], multi-scale rendering [1]. Manipulable or editable scene rendering is one of the directions of NeRF extensions that received attention for its numerous applications such as scene animation or new scene generation. However, the above-mentioned works focus on modeling the existing scenes and thus cannot generate scenes that are unseen during the training.

For some specific object categories, such as the human body or articulated objects, recent studies [24, 32, 33, 37, 38, 41, 46] enable the generation of the unseen scene by controlling the body shape or bone pose. Besides, some works utilize the idea of compositionality to separate foreground objects in the scene during training, thus allowing the scaling or moving of objects in the scene [16, 48]. However, the manipulation in these approaches only allows affine transformations of objects. Although the above methods attempted to develop for radiance field manipulation, they have a common and clear limitation: they cannot perform explicit scene manipulation with details (e.g. torsion or local scaling) for arbitrary categories of objects.

To address the above issues, we propose a new approach for manipulating the optimized radiance field. Our method allows free-form deformation of the radiance field, thus enabling explicit object-level scene deformation or animation. Our idea is an extension of cage-based deformation (CBD), which is originally proposed for mesh deformation [17, 18, 22]. Specifically, the deformation of a fine mesh, or the displacement of its vertices, is driven by manipulating the vertices of the coarse triangular mesh called *cage* that enclosed the fine mesh (e.g. Fig. 3(c)). Such a mesh deformation method is also known as *free-form deformation*. Extending cage-based deformation to radiance field deformation while maintaining the properties of the radiance field such as volumetric representation and view-dependent radiance is non-trivial and yet unexplored. In this paper, we derive a novel formulation for applying CBD to the radiance field that satisfies the properties of the radiance field. However, we find that simply applying the proposed formulation to achieve radiance field deformation brings a new issue: the volume rendering process of the radiance field usually requires a huge number of sampling points [28], and CBD is usually accompanied by a high-dimensional tensor computation, these facts lead to impractical deformation computation times. To address this specific issue, we also propose a discretization method specifically suitable for the radiance field that significantly reduces the computation time of CBD.

We conducted extensive experiments with various types of CBD algorithms using synthetic datasets and real-world datasets. Reasonable deformation and photorealistic rendering quality demonstrate the effectiveness of our approach.

In summary, our contributions are listed as follows:

- We proposed a new approach to explicitly manipulate the radiance fields using a coarse triangular mesh called *cage*, allowing free-form deformation of the scene while maintaining photorealistic rendering quality.
- We proposed a discretization method for cage coordinate computation specifically adapted for the radiance field rendering, which achieves a speedup of several orders of magnitude compared to the naive computation.
- We conducted extensive experiments to deform the radiance field and the rendering results demonstrate the soundness and effectiveness of our method.

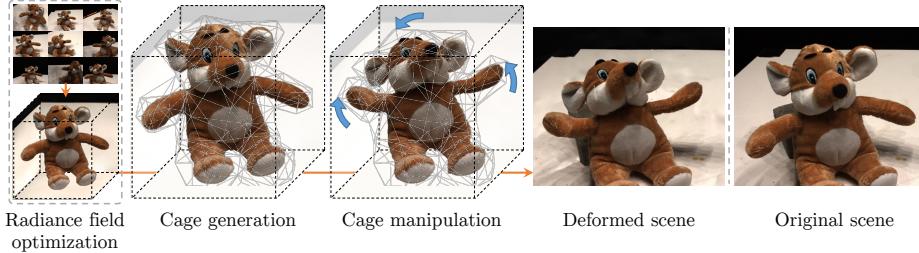


Fig. 1. An overview of our approach. Our method takes multi-view images capturing a static scene as input and uses an off-the-shelf algorithm to optimize a radiance field. Then, we automatically and/or manually generate a cage based on the optimized radiance field. By manipulating the vertices of the cage, the radiance field can be deformed accordingly. Finally, through volume rendering, the free-view rendering of the deformed scene can be achieved.

2 Related Work

Neural scene representation. Recently, neural scene representation, which uses a neural network to encode the 3D scenes, has received a lot of attention due to its high quality of geometry and appearance modeling compared to standard 3D representation including voxel [9, 47], point clouds [6, 11] or textured-mesh [20, 21]. The most representative work is Neural Radiance Field (NeRF) [28], which shows that representing static scenes with volumetric density and view-dependent radiance can capture high-resolution geometry and support photorealistic novel view rendering. An obvious limitation of the original NeRF is that it can only model static scenes. Subsequent work relaxed this limitation and enabled the dynamic scene modeling by simultaneously learning the deformation fields [35, 39, 43] or introducing high-dimensional representation [36]. While these methods achieved the capture of dynamic scenes, none of them can generate new dynamics that are unseen in the training.

Manipulable neural scene rendering. Recent work attempted to incorporate controllability into NeRF to achieve scene manipulation or new scene generation. For the specific task of human body modeling, various works proposed to combine NeRF with a parametric human model to enable human body reposing [37, 38], shape control [24] or even clothing changes [46]. For the articulated object, [33, 41] proposed to build NeRF on the local coordinates of the pre-defined skeleton thus allowing the rendering of the re-posed object, and [32] proposed to learn the unknown skeleton structure along with NeRF. However, the above approaches are limited to specific categories of objects and thus cannot be generalized to the modeling and manipulating of arbitrary objects.

In addition to the above methods of using human model or skeleton to assist in modeling, another direction of manipulatable scene modeling methods utilize an idea of compositionality [12, 16, 34, 48]. Specifically, these methods treat the 3D scene as a composition of multiple objects or backgrounds. By modeling each

object independently, the movement or scaling of each object can be achieved. However, the controllability of such methods focuses on the location or size of objects w.r.t. the whole scene, and cannot achieve detailed deformation of the shape or appearance for individual objects. In contrast to all the above approaches, our method focuses on object-level deformation for detailed shape and appearance manipulation.

Concurrent work [50] uses a similar idea of mesh-based deformation for geometry editing of NeRF, which takes extracted fine mesh as an interface.

Cage-based deformation. Cage-based deformation (CBD) is a volumetric deformation method that is typically used for fine mesh deformation by manipulating the corresponding cage vertices. Here, *cage* denotes a watertight mesh that encloses the target fine mesh to be deformed. The core of CBD is *cage coordinates*, a generalized form of barycentric coordinates, which is used to represent the relative positions of spatial points w.r.t. the cage. The new position of a spatial point can be computed from its cage coordinates and the deformed cage. Previous works proposed several cage coordinates with different properties, including mean value coordinates (MVC) [7, 18], harmonic coordinates (HC) [5, 17], green coordinates (GC) [22], etc. For example, the computation of MVC and GC have closed-formulation and thus can be computed in a feedforward manner, while HC does not have a closed-formulation and therefore its computation requires loop optimization. Specifically, the computation of HC discretizes the space into grid points and updates the HC value for each grid point by performing laplacian smoothing with certain boundary conditions. More comparisons and mathematical preliminaries can be found in [31]. In addition to the traditional CBD algorithm, recent works proposed to combine CBD with deep learning algorithm to achieve high-quality mesh deformation [14, 49].

All of the above methods are focused on using CBD for mesh deformation. Our method aims to extend the CBD to the deformation of the radiance field.

3 Method

Our goal is to deform the optimized radiance field by manipulating the corresponding cage vertices, thus achieving a photorealistic rendering of the new deformed scene. An overview of our approach is shown in Fig. 1. The first step is to optimize a radiance field from the multi-view images (Sec. 3.1). Then, a cage enclosing the foreground object is generated based on the optimized radiance field (Sec. 3.2). With our proposed cage-based deformation formulation for the radiance field, the free-form deformation of the radiance field can be achieved (Sec. 3.3, Sec. 3.4).

3.1 Radiance fields revisited

Neural radiance field (NeRF) [28] uses a neural network to encode the 3D scene as a continuous neural representation, which receives the spatial position $\mathbf{x} \in \mathbb{R}^3$

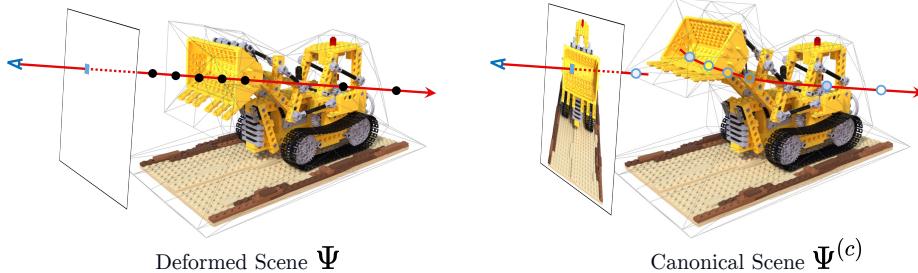


Fig. 2. Rendering process of the deformed scene. To perform volume rendering for the deformed radiance field Ψ , we map the sampling points on the ray to the canonical space through cage-based deformation and query the color and density in the canonical radiance field $\Psi^{(c)}$.

and view direction $\mathbf{d} \in \mathbb{R}^3$ as inputs and computes the RGB color $\mathbf{c} \in \mathbb{R}^3$ and density $\sigma \in \mathbb{R}$ of that point. With volume rendering, photorealistic rendering of NeRF from an arbitrary viewpoint can be achieved. Recently, some variants of radiance field representation have been proposed, for example, Plenoxels [8] use grid representation and directly optimize radiance field without using neural networks. Without loss of generality, we refer to the 3D scene representation that can be formulated as $\Psi : (\mathbf{x}, \mathbf{d}) \rightarrow (\mathbf{c}, \sigma)$ as *radiance field*.

As explored in previous studies, given a static scene, a radiance field $\Psi^{(c)}$ can be optimized from a set of multi-view images with calibrated camera parameters. Here c stands for “canonical”, which denotes the original static scene, to distinguish it from the later deformed scene.

3.2 Cage generation from radiance fields

In this paper, *cage* refers to a coarse 3D triangular mesh that strictly encloses the foreground object. We demonstrate a method for automatically and/or manually generating a cage from the optimized radiance field. Specifically, the first step is to convert the radiance field into a fine mesh using surface extraction methods such as marching cubes [26] (Fig. 3(b)). The second step is to create the corresponding cage for the generated mesh (Fig. 3(c)). For scenes containing only foreground objects (such as those optimized using masked images), we use [45] to compute the corresponding cage. For scenes containing backgrounds, we use Blender [3] to manually split the foreground objects from the reconstructed fine mesh and then apply [45] to compute the cage. However, some cage predictions may be inaccurate due to the complex shapes or fine details of the scenes. For these cases, we manually resolve them based on the automatically generated cage for better manipulation performance. Alternatively, if only a simple deformation (or moving, scaling) is needed, we can also use 3D software to manually build a simple cage, such as rectangle or cylinder, by referring to the extracted fine mesh.

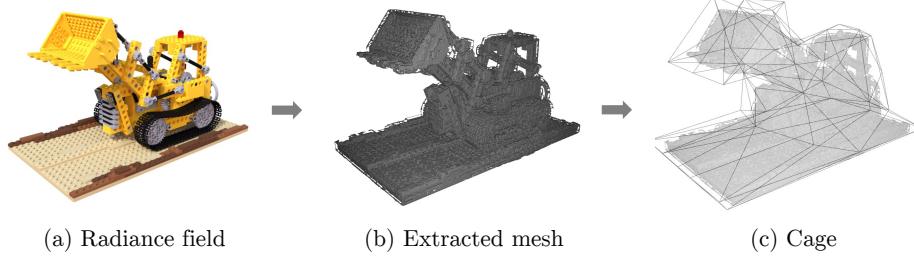


Fig. 3. Cage generation process. After extracting a fine mesh from the optimized radiance field, a cage can be generated automatically and/or manually according to the fine mesh.

3.3 Cage-based deformation

Cage-based deformation (CBD) is originally proposed for deforming a fine mesh using the cage, which calculates the vertex displacement of the fine mesh caused by the cage manipulation. Specifically, given a cage \mathcal{C} with vertices $\{\mathbf{v}_j\}$, points $\mathbf{x} \in \mathbb{R}^3$ inside \mathcal{C} can be identified with cage coordinates $\{\omega_j\}$ which represent the relative position of \mathbf{x} w.r.t. \mathcal{C} . Formally, the position of point \mathbf{x} is weighted by the cage vertices as:

$$\mathbf{x} = \sum_j \omega_j(\mathbf{x}) \mathbf{v}_j. \quad (1)$$

Consider that we manipulate the vertices of \mathcal{C} and deform it to cage \mathcal{C}' with vertices $\{\mathbf{v}'_j\}$. Using the calculated cage coordinate, the deformed position of \mathbf{x} for the deformed cage \mathcal{C}' can be calculated as:

$$\mathbf{x}' = \sum_j \omega_j(\mathbf{x}) \mathbf{v}'_j. \quad (2)$$

Previous studies [17, 18, 22] proposed several kinds of cage coordinates and achieved promising results on the mesh deformation.

Note that although the above formulation seems simple, the actual derivation and computation of the cage coordinate $\{\omega_j\}$ is complicated and usually accompanied by a large tensor computation. For detailed computation, we recommend referring to the original papers of these cage coordinates [17, 18, 22].

3.4 Deforming radiance fields

In this section, we introduce a novel formulation that extends the application of CBD from mesh to the radiance field. Remind that our goal is to deform the optimized radiance field $\Psi^{(c)}$ for the free-view rendering of the deformed scenes. Suppose we have a cage $\mathcal{C}^{(c)}$ that accompanies $\Psi^{(c)}$ that encloses the foreground object. Consider that we manipulate the vertices of $\mathcal{C}^{(c)}$ and deform it to a new cage \mathcal{C} , and denote the desired radiance field after deformation as Ψ .

To achieve volume rendering of Ψ , the sampling points are required to be mapped from the deformed space to the canonical space for color and density computations, as shown in Fig 2. To describe such deformed-to-canonical mapping, we reversely treat the cage deformation process as: the new cage \mathcal{C} is deformed to the canonical cage $\mathcal{C}^{(c)}$. While contrary to the actual cage manipulation process, such a convention allows us to map the points in the deformed space back to the canonical space. Specifically, we denote the deformed-to-canonical mapping of spatial position and view direction as:

$$\phi_{\mathbf{x}} : \mathbf{x} \rightarrow \mathbf{x}^{(c)}, \quad \phi_{\mathbf{d}} : (\mathbf{x}, \mathbf{d}) \rightarrow \mathbf{d}^{(c)} \quad (3)$$

Here, $\mathbf{x}^{(c)}$ can be simply derived from Eq. (2) and $\mathbf{d}^{(c)}$ can be derived from difference approximation as $\mathbf{d}^{(c)} = \text{norm}((\phi_{\mathbf{x}}(\mathbf{x} + \Delta t \mathbf{d}) - \phi_{\mathbf{x}}(\mathbf{x})) / \Delta t)$. Δt denotes a small constant and $\text{norm}(\cdot)$ normalizes the vector length to 1. Note that the above mappings are derived from the simple CBD computation without any learnable components.

The deformed radiance field Ψ can be divided into three parts depending on the space that: (1) outside both the canonical cage and deformed cage (2) inside the canonical cages but outside the deformed cages, (3) inside the deformed cages. Specifically, it can be formulated as follows:

$$\Psi(\mathbf{x}, \mathbf{d}) = \begin{cases} \Psi^{(c)}(\mathbf{x}, \mathbf{d}), & \mathbf{x} \in \mathbb{R}^3 \setminus (\mathbb{V}^{(c)} \cup \mathbb{V}) \\ (\mathbf{0}, 0), & \mathbf{x} \in \mathbb{V}^{(c)} \setminus (\mathbb{V}^{(c)} \cap \mathbb{V}) \\ \Psi^{(c)}(\phi_{\mathbf{x}}(\mathbf{x}), \phi_{\mathbf{d}}(\mathbf{x}, \mathbf{d})), & \mathbf{x} \in \mathbb{V} \end{cases} \quad (4a)$$

$$(4b) \quad (4c)$$

Here, $\mathbb{V}^{(c)}, \mathbb{V} \subset \mathbb{R}^3$ denotes the space enclosed by $\mathcal{C}^{(c)}$ and \mathcal{C} , respectively. Eq. (4a) indicates that the radiance field remains unchanged before and after deformation for the position outside the cages. For the points inside the canonical cage, we clear them, namely setting the color and density to zero, as in Eq. (4b). For the points inside the deformed cage, we map the spatial position and view direction to the canonical space through Eq. (3) and then query the color and density from $\Psi^{(c)}$, as in Eq. (4c).

4 Implementation details

4.1 Faster cage coordinates computation

Technically, the rendering of the deformed scene can be achieved by computing the deformed-to-canonical mapping in Eq. (4) for all the sampling points on all the rays. However, because of the huge number of sampling points and the fact that the computation of cage coordinates is usually accompanied by a high-dimensional tensor computation (as discussed in Sec. 3.3), the above brute-force computation is usually impractical either in terms of time or memory capacity. A rough estimation can be given as, rendering images of size (h, w) from N different viewpoints, with M points sampled on each ray, the number of points

that require the cage coordinate computation is about $h \times w \times N \times M$ in order of magnitude¹. For instance, for $M = 512$, rendering 200 images of size (800, 800) requires about $\sim 10^{10}$ orders of magnitude in the times of cage coordinates computation.

Inspired by the computation process of harmonic coordinates (HC), we propose to discretize the space into $n \times n \times n$ grid points for cage coordinates computation, even for the cage coordinates that have their closed-formulations, i.e. MVC and GC (briefly discussed in Sec. 2). At the inference time, we pre-compute the cage coordinates for each grid and use trilinear interpolation to calculate the cage coordinates for arbitrary points. We surprisingly find that such a simple discretization, however, brings great benefits for the specific nature of volume rendering of the radiance field. Note that such discretization makes the number of cage coordinates computation independent of h, w, N, M given above, that is, once the pre-computation of grid points is completed, there is no increase in the computation of cage coordinates when we want to render the scene from the additional new viewpoint or with different image resolution. The only thing to consider here is the size of n , which requires a trade-off between discretization resolution and computation speed. Here, the number of points that require cage coordinates computation is about n^3 in order of magnitude¹.

We practically use $n = 128$ in our experiments which gives about $\sim 10^6$ orders of magnitude in the times of computation.

4.2 Cage refinement

The computation complexity of cage coordinates is proportional to the number of cage vertices. For fast inference, we control the hyperparameters (e.g. discrete voxel size) in cage generation algorithm [45] to ensure that the number of vertices is in the range of $30 \sim 200$. For scenes with complex shapes or details, we first generate a cage with a larger number of vertices (~ 1000) and then manually decimate the vertices using Blender [3].

4.3 Radiance fields representation

We use Plenoxels [8] as radiance fields representation, which supports very fast scene optimization and rendering. Note that our method is not dependent on specific radiance field representations and thus can be directly applied to other representations such as NeRF [28] or the latest faster radiance field representations [2, 29, 42].

5 Experiments

In this section, we evaluate the effectiveness of our approach through a variety of scenes, including synthetic dataset and real-world dataset. We show the results of extensive ablation studies and then discuss the limitations of our approach.

¹ In fact, the actual number is smaller than this approximation since we only compute for points inside the cage.

Unless otherwise specified, the deformations of the results are performed with harmonic coordinate with discretization resolution $n = 128$. For the canonical scene optimization, we follow the default setting used in [8]. We use one Nvidia A100 GPU for all the experiments.

5.1 Datasets

NeRF and NSVF synthetic dataset. We use synthetic dataset in original NeRF [28] and Neural Sparse Voxel Fields (NSVF) [23] papers. These scenes contain only foreground objects, and the images are captured from multiple cameras placed on the hemisphere. We follow the train/test split as in the original papers.

DTU MVS dataset. We use the real-world DTU MVS dataset [15], which contains a variety of static objects, and each scene uses 49 or 64 cameras to capture high-resolution images. We use all available cameras for training, and create test camera trajectories from camera interpolation for evaluation.

5.2 Results

Since our approach is the first to use a coarse cage as an interface for free-form deformation of the radiance field, there is no existing method for a direct comparison. The ground truth of the deformed scene is also not available, therefore, we show the qualitative results before and after the deformation for evaluation.

We use Blender [3] to manually deform the generated cage of the canonical scene with various types of deformations such as bending, stretching, torsion, scaling, etc. Novel view synthesis results of original/deformed scene on synthetic dataset and DTU dataset are shown in Fig 4 and Fig 5, respectively. As shown in the results, our proposed radiance field deformation approach enables explicit manipulation of the scene while maintaining photorealistic rendering quality. In addition to the free-form deformation of the entire object with the generated cage, our approach also allows for local manipulation of the object as in the last two figures in Fig. 5: all you need is to create a simple cage and deform it, which can be done with little effort using almost any existing 3D software.

The above features of our approach also support simple radiance field manipulation achieved by existing works [25], such as object movement, duplication, and scaling. Moreover, as shown in Fig. 6, our method also supports applications of generating continuous free-view animation from a static scene by cage interpolation.

5.3 Ablation study

We discuss the impact of different discretization resolutions and different cage coordinates on the synthesis quality. As introduced in 2, we use three commonly used cage coordinates for comparison: mean value coordinates (MVC) [18], harmonic coordinates [17] and green coordinates [22]. We observed that the impact

Table 1. Computation time in seconds for rendering an image for three cage coordinates with different discretization resolutions. “Precise” means not using discretization, i.e., computing precise cage coordinates for all the sampling points on the rays. Here, “MVC” means mean value coordinates [18], “HC” means harmonic coordinates [17], and “GC” means green coordinates [22]. Please also refer to Sec. 5.3 and Fig. 7.

	MVC [18]	HC [17]	GC [22]
64^3	0.31	0.23	0.35
128^3	0.98	0.90	2.49
256^3	5.71	6.32	19.69
Precise	102	N/A	243

of the above two factors on the synthesis quality is subtle, we choose the synthetic “Lego” scene with relatively obvious distinction for ablation. The computation time for rendering an image and the synthesis results are shown in Tab. 1 and Fig 7, respectively.

We use the same settings as assumptions in Sec. 4.1 except for the number of rendered images, i.e. $h = w = 800, M = 512, n = 128$ and $N = 1$. The cage we used for the synthetic “Lego” scene has 42 vertices.

Impact of discretization resolution. As shown in Tab. 1 and Fig 7, although 64^3 resolution has a faster computation speed, significant artifacts can be observed in the synthesized results (e.g. blurry or fake shadow). This indicates that low discretization resolution brings a large error in the deformed-to-canonical mapping of the sampled points. For 128^3 resolution, the artifact is lightened with an acceptable computation time increase. For 256^3 resolutions, it can be seen that the improvement in synthesized quality is limited, but causes a larger increase in computation time as well as memory cost. For cage coordinates that have a closed-formulation (i.e., MVC and GC), although it is impractical due to the extremely long computation time (about 2 ~ 4 minutes per scene), we show the results of the precise computation of cage coordinates without using discretization as an upper limit for comparison.

Impact of different cage coordinates. Comparison are also shown in Fig 7. MVC and HC show similar synthesized quality. GC shows a more reasonable deformation for the long-striped parts in the center of the image, however, the loss of detail for small parts is also observed.

5.4 Limitations

Cage generation. As noticed in the cage-based mesh deformation, the quality of the cage greatly affects the deformation quality. However, the method we use for cage generation shows some difficulties in representing detailed cage shapes while keeping a small number of cage vertices. For objects with difficult shapes,

manual refinement of the cage comes necessary, especially for real scenes with backgrounds. We conducted an extensive survey on the automatic cage generation from 3D scenes, and to our surprise, this task seems to still be unexplored. Especially for real scenes, to the best of our knowledge, there is no effective way to generate a cage automatically. We believe that the automatic generation of the cage from 3D scenes is a promising direction for future work.

Failure cases. We report some failure cases of our approach in Fig. 8. The first typical failure case occurs when a part of the scene (usually the background) is not well modeled due to the occlusion. When deforming or moving the foreground objects so that the under-modeled part is exposed, obvious artifacts will be observed. However, addressing this issue is very challenging because the traditional optimization method of the radiance field cannot handle the part unseen during the training, which makes other scene manipulation methods [48] also suffer from the same issue. We assume that the use of occlusion-aware scene modeling methods or scene completion techniques may help to alleviate this issue.

The second typical failure case may occur when a part of the object gets drastically deformed, the irrelevant parts may also be affected thus causing artifacts. This is also a long-standing issue for cage-based mesh deformation. As discussed in the previous works of CBD [31], we believe that this issue might be alleviated by generating cages with higher accuracy as mentioned above or by choosing appropriate cage coordinates.

6 Conclusion

We presented a new method that enables free-form deformation of the radiance field. We derived a novel formulation to extend the application of cage-based deformation to the radiance field. By manipulating the vertices of the cage, we can explicitly perform free-form deformation of the radiance field while maintaining photorealistic rendering quality. To address the issue of impractical deformation computation time that arises in a naive implementation, we propose to use a discretization method specifically adapted for the radiance field and succeed in reducing the computation time by several orders of magnitude. Currently, the quality of the scene deformation is still largely influenced by the quality of the generated cage, this leaves us with a trade-off between the effort of manual cage refinement and the deformation quality. A better automatic cage generation algorithm would be a promising direction for future work.

Acknowledgements We would like to thank Daisuke Kasuga, Ryosuke Sasaki, Tomoyuki Takahata, Haruo Fujiwara, and Atsuhiro Noguchi for comments and discussions. This work was partially supported by JST AIP Acceleration Research JPMJCR20U3, Moonshot R&D Grant Number JPMJPS2011, CREST Grant Number JPMJCR2015, JSPS KAKENHI Grant Number JP19H01115 and Basic Research Grant (Super AI) of Institute for AI and Beyond of the University of Tokyo.

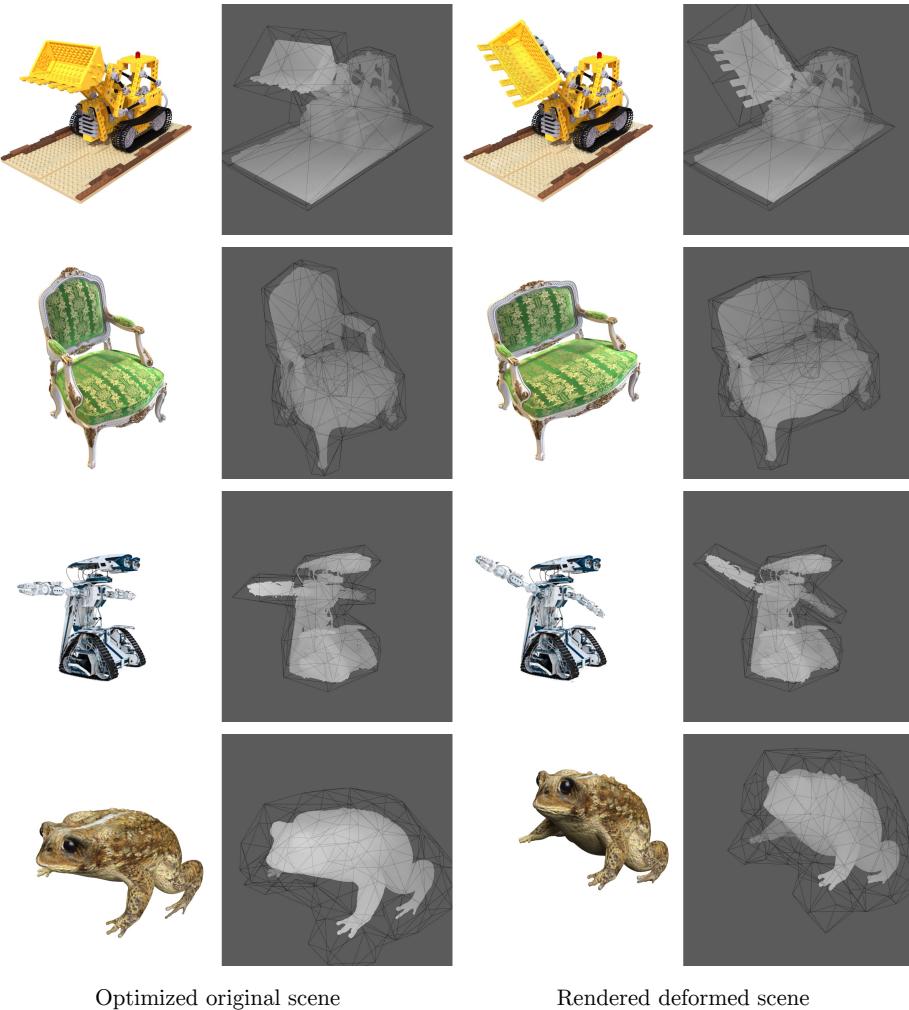


Fig. 4. Qualitative results on NeRF and NSVF synthetic dataset. The original scene (left) and the deformed scene (right) are rendered from a novel viewpoint. Disparity map and corresponding cage are also presented.

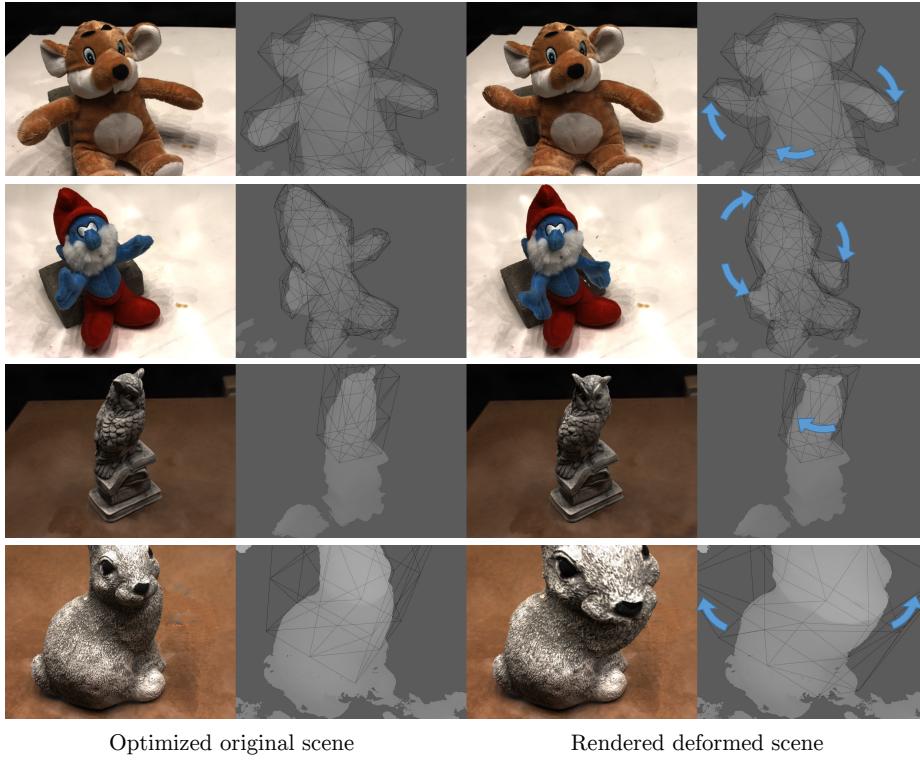


Fig. 5. Qualitative results on DTU dataset. The original scene (left) and the deformed scene (right) are rendered from a novel viewpoint. Disparity map of the foreground object and corresponding cage are also presented. Arrows illustrate the manipulation of the cage.



Fig. 6. Qualitative results of cage interpolation. Our approach can generate continuous free-view animation by interpolating the starting and ending cages.

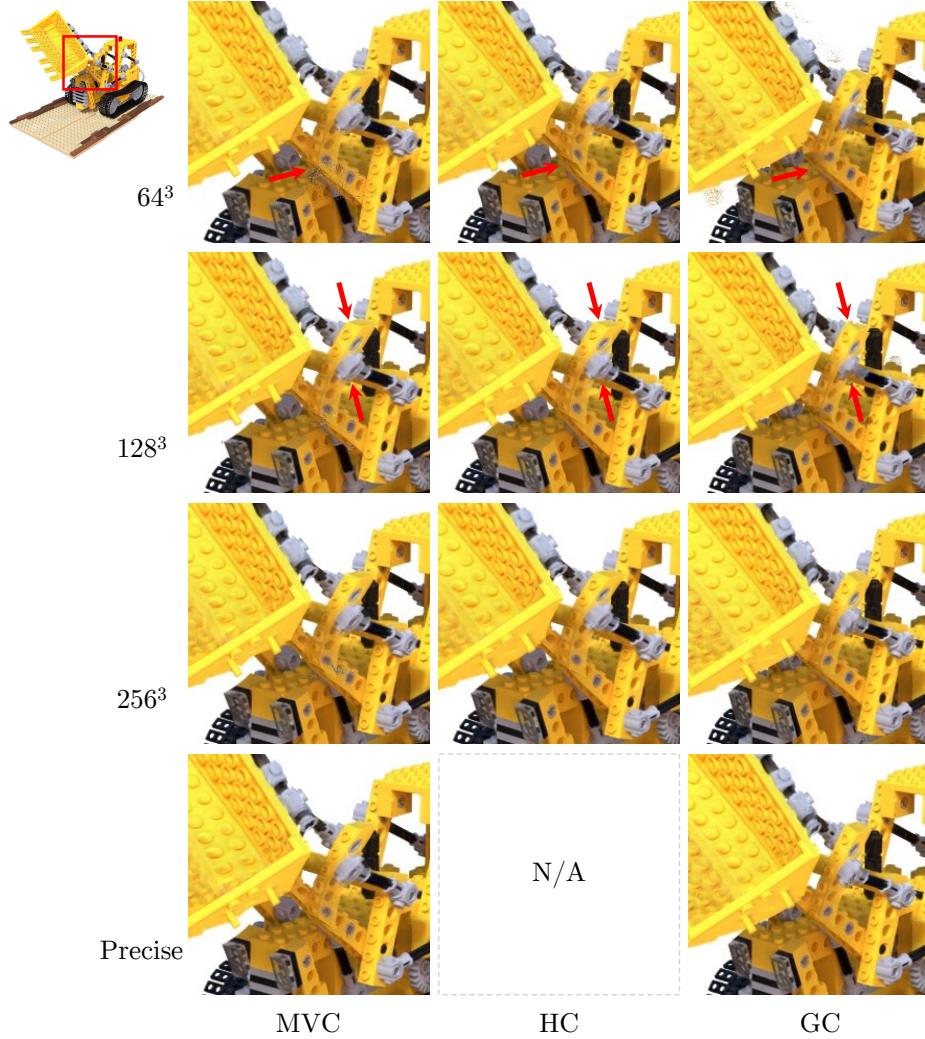


Fig. 7. Ablation on different cage coordinates and discretization resolution on synthetic “Lego” dataset. For more details please refer to Sec. 5.3 and Tab. 1.



Fig. 8. Failure cases. Left: moving the foreground object results in exposing the parts of the scene that are under-modeled due to occlusion. Right: drastic cage manipulation may cause the artifacts.

References

1. Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P.P.: Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5855–5864 (2021)
2. Chen, A., Xu, Z., Geiger, A., Yu, J., Su, H.: Tensorf: Tensorial radiance fields. European Conference on Computer Vision (2022)
3. Community, B.O.: Blender - a 3D modelling and rendering package. Blender Foundation, Stichting Blender Foundation, Amsterdam (2018), <http://www.blender.org>
4. Davis, A., Levoy, M., Durand, F.: Unstructured light fields. In: Computer Graphics Forum. vol. 31, pp. 305–314. Wiley Online Library (2012)
5. DeRose, T., Meyer, M.: Harmonic coordinates. In: Pixar Technical Memo 06-02, Pixar Animation Studios (2006)
6. Fan, H., Su, H., Guibas, L.J.: A point set generation network for 3d object reconstruction from a single image. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 605–613 (2017)
7. Floater, M.S.: Mean value coordinates. Computer aided geometric design (2003)
8. Fridovich-Keil, S., Yu, A., Tancik, M., Chen, Q., Recht, B., Kanazawa, A.: Plenoxtels: Radiance fields without neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5501–5510 (2022)
9. Girdhar, R., Fouhey, D.F., Rodriguez, M., Gupta, A.: Learning a predictable and generative vector representation for objects. In: European Conference on Computer Vision. pp. 484–499. Springer (2016)
10. Gortler, S.J., Grzeszczuk, R., Szeliski, R., Cohen, M.F.: The lumigraph. In: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. pp. 43–54 (1996)
11. Guo, M.H., Cai, J.X., Liu, Z.N., Mu, T.J., Martin, R.R., Hu, S.M.: Pct: Point cloud transformer. Computational Visual Media (2021)
12. Guo, M., Fathi, A., Wu, J., Funkhouser, T.: Object-centric neural scene rendering. arXiv preprint arXiv:2012.08503 (2020)
13. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge university press (2003)
14. Jakab, T., Tucker, R., Makadia, A., Wu, J., Snavely, N., Kanazawa, A.: Keypoint-deformer: Unsupervised 3d keypoint discovery for shape control. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12783–12792 (2021)
15. Jensen, R., Dahl, A., Vogiatzis, G., Tola, E., Aanæs, H.: Large scale multi-view stereopsis evaluation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 406–413 (2014)
16. Jiakai, Z., Xinhang, L., Xinyi, Y., Fuqiang, Z., Yanshun, Z., Minye, W., Yingliang, Z., Lan, X., Jingyi, Y.: Editable free-viewpoint video using a layered neural representation. In: ACM SIGGRAPH (2021)
17. Joshi, P., Meyer, M., DeRose, T., Green, B., Sanocki, T.: Harmonic coordinates for character articulation. ACM Transactions on Graphics (TOG) (2007)
18. Ju, T., Schaefer, S., Warren, J.: Mean value coordinates for closed triangular meshes. In: ACM Siggraph 2005 Papers, pp. 561–566 (2005)
19. Kajiya, J.T., Von Herzen, B.P.: Ray tracing volume densities. ACM SIGGRAPH computer graphics (1984)

20. Kanazawa, A., Tulsiani, S., Efros, A.A., Malik, J.: Learning category-specific mesh reconstruction from image collections. In: European Conference on Computer Vision. pp. 371–386 (2018)
21. Kato, H., Ushiku, Y., Harada, T.: Neural 3d mesh renderer. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3907–3916 (2018)
22. Lipman, Y., Levin, D., Cohen-Or, D.: Green coordinates. ACM Transactions on Graphics (TOG) (2008)
23. 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)
24. Liu, L., Habermann, M., Rudnev, V., Sarkar, K., Gu, J., Theobalt, C.: Neural actor: Neural free-view synthesis of human actors with pose control. ACM Trans. Graph.(ACM SIGGRAPH Asia) (2021)
25. Liu, S., Zhang, X., Zhang, Z., Zhang, R., Zhu, J.Y., Russell, B.: Editing conditional radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5773–5783 (2021)
26. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3d surface construction algorithm. ACM siggraph computer graphics (1987)
27. Mildenhall, B., Hedman, P., Martin-Brualla, R., Srinivasan, P.P., Barron, J.T.: Nerf in the dark: High dynamic range view synthesis from noisy raw images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16190–16199 (2022)
28. 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)
29. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. ACM Trans. Graph. (2022)
30. Niemeyer, M., Mescheder, L., Oechsle, M., Geiger, A.: Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3504–3515 (2020)
31. Nieto, J.R., Susín, A.: Cage based deformations: a survey. In: Deformation models, pp. 75–99. Springer (2013)
32. Noguchi, A., Iqbal, U., Tremblay, J., Harada, T., Gallo, O.: Watch it move: Unsupervised discovery of 3d joints for re-posing of articulated objects. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3677–3687 (2022)
33. Noguchi, A., Sun, X., Lin, S., Harada, T.: Neural articulated radiance field. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5762–5772 (2021)
34. Ost, J., Mannan, F., Thuerey, N., Knodt, J., Heide, F.: Neural scene graphs for dynamic scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2856–2865 (2021)
35. Park, K., Sinha, U., Barron, J.T., Bouaziz, S., Goldman, D.B., Seitz, S.M., Martin-Brualla, R.: Nerfies: Deformable neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 5865–5874 (2021)
36. Park, K., Sinha, U., Hedman, P., Barron, J.T., Bouaziz, S., Goldman, D.B., Martin-Brualla, R., Seitz, S.M.: Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields. ACM Trans. Graph. (2021)

37. Peng, S., Dong, J., Wang, Q., Zhang, S., Shuai, Q., Zhou, X., Bao, H.: Animatable neural radiance fields for modeling dynamic human bodies. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14314–14323 (2021)
38. Peng, S., Zhang, Y., Xu, Y., Wang, Q., Shuai, Q., Bao, H., Zhou, X.: Neural body: Implicit neural representations with structured latent codes for novel view synthesis of dynamic humans. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9054–9063 (2021)
39. Pumarola, A., Corona, E., Pons-Moll, G., Moreno-Noguer, F.: D-nerf: Neural radiance fields for dynamic scenes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10318–10327 (2021)
40. Sitzmann, V., Zollhöfer, M., Wetzstein, G.: Scene representation networks: Continuous 3d-structure-aware neural scene representations. Advances in Neural Information Processing Systems **32** (2019)
41. Su, S.Y., Yu, F., Zollhöfer, M., Rhodin, H.: A-nerf: Articulated neural radiance fields for learning human shape, appearance, and pose. Advances in Neural Information Processing Systems **34**, 12278–12291 (2021)
42. Sun, C., Sun, M., Chen, H.T.: Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5459–5469 (2022)
43. Tretschk, E., Tewari, A., Golyanik, V., Zollhöfer, M., Lassner, C., Theobalt, C.: Non-rigid neural radiance fields: Reconstruction and novel view synthesis of a dynamic scene from monocular video. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12959–12970 (2021)
44. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment—a modern synthesis. In: International workshop on vision algorithms. pp. 298–372. Springer (1999)
45. Xian, C., Lin, H., Gao, S.: Automatic generation of coarse bounding cages from dense meshes. In: IEEE International Conference on Shape Modeling and Applications (2009)
46. Xu, T., Fujita, Y., Matsumoto, E.: Surface-aligned neural radiance fields for controllable 3d human synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15883–15892 (2022)
47. Yan, X., Yang, J., Yumer, E., Guo, Y., Lee, H.: Perspective transformer nets: Learning single-view 3d object reconstruction without 3d supervision. Advances in neural information processing systems **29** (2016)
48. Yang, B., Zhang, Y., Xu, Y., Li, Y., Zhou, H., Bao, H., Zhang, G., Cui, Z.: Learning object-compositional neural radiance field for editable scene rendering. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 13779–13788 (2021)
49. Yifan, W., Aigerman, N., Kim, V.G., Chaudhuri, S., Sorkine-Hornung, O.: Neural cages for detail-preserving 3d deformations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 75–83 (2020)
50. Yuan, Y.J., Sun, Y.T., Lai, Y.K., Ma, Y., Jia, R., Gao, L.: Nerf-editing: geometry editing of neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18353–18364 (2022)