

ParticleNeRF: Particle Based Encoding for Online Neural Radiance Fields in Dynamic Scenes

Jad Abou-Chakra¹Feras Dayoub²Niko Sünderhauf¹¹ Queensland University of Technology² University of Adelaide

Abstract

Neural Radiance Fields (NeRFs) are coordinate-based implicit representations of 3D scenes that use a differentiable rendering procedure to learn a representation of an environment from images. This paper extends NeRFs to handle dynamic scenes in an online fashion. We do so by introducing a particle-based parametric encoding, which allows the intermediate NeRF features – now coupled to particles in space – to be moved with the dynamic geometry. We backpropagate the NeRF’s photometric reconstruction loss into the position of the particles in addition to the features they are associated with. The position gradients are interpreted as particle velocities and integrated into positions using a position-based dynamics (PBS) physics system. Introducing PBS into the NeRF formulation allows us to add collision constraints to the particle motion and creates future opportunities to add other movement priors into the system such as rigid and deformable body constraints. We show that by allowing the features to move in space, we incrementally adapt the NeRF to the changing scene.

1. Introduction

Neural Radiance Fields [19] are 3D scene representations that are trained from images using a differentiable rendering process. They generate photorealistic images from unseen viewpoints and have had tremendous success at modelling static scenes in close to realtime [20]. In contrast, NeRF formulations that model dynamic scenes are much slower to train. Most approaches to date attempt to represent the entire evolution of a 3D scene within the NeRF [23–25]. Encoding this temporal aspect within the implicit representation can be excessive for certain applications in robotics where the latest state of the 3D world is of most interest.

The slower training times in the NeRF architectures used to model dynamic scenes are caused by two fundamental issues. First, they use encodings that are non-

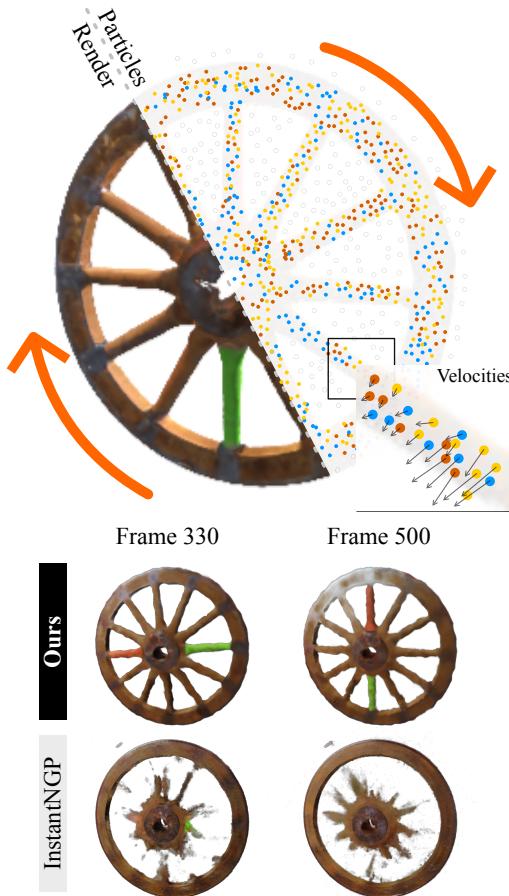


Figure 1. Our particle encoding significantly improves the NeRFs ability to adapt to changing scenes when compared to the top performing parametric encoding - InstantNGP. The particle encoding fills the NeRF space with particles at random locations with each particle being paired to a feature. Both the positions of the particles and their associated features are optimized during training. As the wheel spins, the particles move to maintain a low reconstruction error.

parametric [2, 3, 5, 19, 33], and as a consequence require large neural network backbones which are expensive to ren-

der and train. In contrast, NeRFs that model static scenes in the order of seconds rely on parametric encodings which are more computationally efficient [4, 7, 20, 29]. Second, the challenge of modelling a changing 3D scene is made harder by requiring the NeRF to remember the state of the world at every timestep in the dataset. While this may be required for some applications like capturing 3D videos in an implicit representation [6, 8, 23, 24, 28], it is not required in some robotic applications where NeRFs can be used to model the last state of the environment.

We tackle the challenge of learning a changing scene incrementally where the NeRF does not have access to future frames during training and is required to adapt to changes in the order of milliseconds. We construct experiments with moving objects and show how the fastest implementation of NeRF to date [20] behaves when the training data constantly changes. We propose a particle-based encoding which we hypothesize is better suited for moving objects. Our key insight is that once a NeRF has been learnt, the intermediate features stored within its embedding layer could themselves be moved rather than relearnt – Figure 2. Our particle encoding associates particles in space with feature vectors. We formulate the NeRF photometric reconstruction loss so that it depends on both the associated features and the positions of the particles. The NeRF gradients flow into the positions of the particles. This consequently provides the means of moving features in space in response to changes in the underlying scene. Given a NeRF of an incrementally changing environment, optimising the positions of these particles can be faster than relearning the associated features.

In contrast to the deformation networks used in [6, 8, 23, 24, 28], which are designed for offline training and are conditioned on time, our method incrementally adapts an existing NeRF after a new set of images is received – making it more suitable for online usage.

In addition to the particle encoding, we build a physics system to manipulate the particle positions in a controlled manner. We interpret the position gradients as scaled velocity vectors that are added to the velocities of the particle that are already present. The physics system provides a powerful mechanism for adding constraints and prior knowledge of the motion. In this paper, we add collision constraints to the particles to prevent them from accumulating at a location and to create a soft upper limit on the number of neighbours that need to be searched over when performing the training and the rendering. However, we envision that future work can use the constraint system to add other priors that regularize the NeRF loss by inducing rigid body motion, articulated motion, or soft body motion.

In summary, our contributions are:

1. A particle encoding that allows the NeRF’s intermediate features to move with the geometry they are representing.

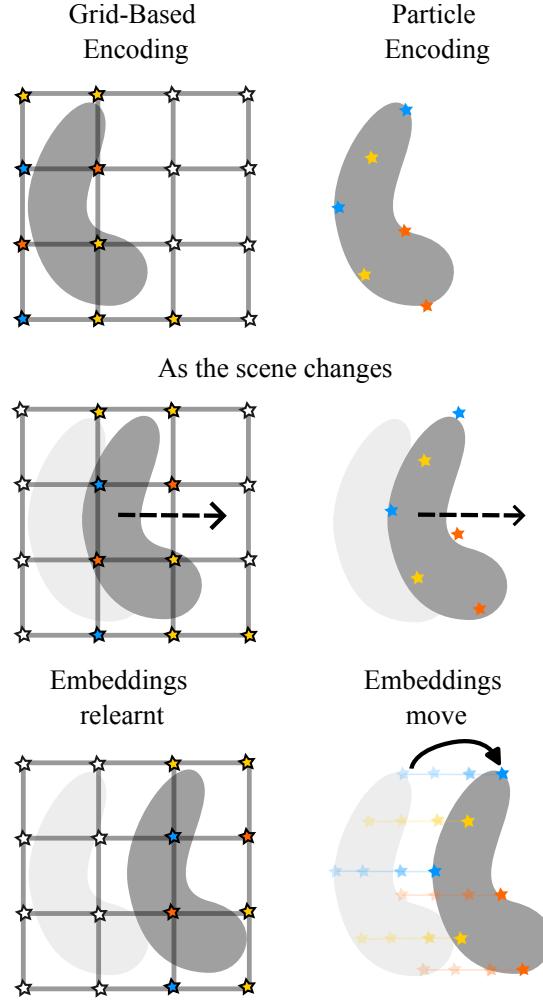


Figure 2. Illustration of the key difference between a grid-based parametric encoding and our particle encoding. Each node in a grid is associated with a feature. As the object moves, all the features on the object need to be relearnt. With a particle encoding, the particles can simply move to account for the motion.

2. The introduction of a physics system that is incorporated into the NeRF formulation to update the motion of the particles while preventing collisions.

We apply these changes as extensions to InstantNGP [20], the fastest implementation of NeRF available and make our code publicly available. Our experiments demonstrate that while the particle encoding has inferior reconstruction quality on static environments, it is significantly superior to InstantNGP at quickly adapting to changes in the scene.

2. Related Work

We review Neural Radiance Fields in the context of (i) how they have been used to tackle dynamic environments,

(ii) how certain encodings allow close to realtime training and rendering, and lastly (iii) how our particle encoding is different to three other point-based encodings suggested to date.

Dynamic NeRFs NeRFs are representations of a 3D scene that can be learnt from a set of posed images. They were formulated to tackle the challenge of rendering photo-realistic images from unseen viewpoints. They are a form of coordinate-based implicit representations that map points in space to density and color. Although, originally, they could only represent static scenes, several papers extended the architecture to encompass moving scenes [9, 15, 23–25, 34]. The distinguishing factor between the static and dynamic NeRF formulations is that the first represents a single static scene while the latter encodes the evolution of a 3D scene over time. During training, these dynamic NeRFs can access images taken at any point in time. They are therefore designed to be trained after a sequence of images has already been taken. To date, all these works take hours to train and are not close to being realtime capable. Since these dynamic NeRF approaches use the complete set of images during training, we refer to them as “offline” methods. Alternatively, we seek to create a NeRF at every timestep and only capture the latest state of the scene. In this context, our method is “online” or “incremental”. For many applications, representing a time-varying scene within the NeRF itself is unnecessary. A robot, for example, is usually only interested in the latest state of the world. Learning a NeRF incrementally is therefore the challenge of learning a NeRF at a timestep t given a NeRF at a timestep $t - 1$ and a set of images taken at time t . Our particle encoding is formulated to tackle this problem.

Embeddings In general, a NeRF is composed of two main parts: the encoding, and a multi layered perceptron (MLP). An encoding transforms the spatial input of the NeRF to an alternative feature space where learning can be made easier [31]. The MLP transforms the resultant feature into values that represent color and geometry. Mildenhall et al. [19] introduced the first version of NeRFs with a Fourier encoding which is a simple trigonometric map. The Fourier encoding does not have any parameters that are learnt as part of the training and is not associated with a discrete datastructure. It is, therefore, referred to as non-parametric, or functional. This version of NeRF takes hours to train because it requires a large MLP to represent the scene. Later works [4, 20, 29] show that by storing a large set of features in a discrete datastructure, the spatial input could be used to index into and interpolate between features in that set. The computed features become the embeddings that are consumed by the NeRF’s neural network. Since the features in the set are stored as parameters and are updated dur-

ing training, these are referred to as “parametric” encodings by [20]. By storing features in memory, the computational burden on the MLP is reduced and training speed is significantly increased. The strategy used to index and interpolate into the feature set distinguishes the various types of parametric encodings. Almost all the parametric encodings conceptually rely on a fixed grid datastructure [4, 20, 29] with the exception of [35]. The grid encompasses the workspace of the NeRF and each node within is associated with a feature. Therefore, while the features associated with the nodes can be learnt, they cannot be moved. In this paper, we show that if the features are allowed to be both learnt and moved, then NeRFs can be made to incrementally adapt to dynamic scenes. We introduce a particle encoding, where features are associated with moving points in space. We show that in a changing scene, the feature-holding particles move in accordance with the scene – Figure 2.

Point-based NeRFs There are three other works that use points as part of their formulation. PointNet and SPIDR [16, 35] similarly use points as embeddings but crucially they do not allow their points to move as part of their optimization. Furthermore, they takes in the order of minutes to train and require depth data as well as a more involved feature extraction pipeline. In contrast, our feature extraction pipeline is simpler and invariant to both rotations and translations. Lastly, they do not target the dynamic scene use case. NeuroFluid [11] uses particles and NeRFs to specifically model the behaviour of fluids. Their encoding is non-parametric in that each particle does not hold a high dimensional feature. Instead they use a large MLP and as a consequence their method is not realtime capable. Lastly, our approach is conceptually different to [11] since they model particles as physical components of the fluid, whereas we model them as a latent descriptors of the embedding space.

2.1. NeRF Preliminaries

A NeRF is a continuous representation of a 3D scene that maps a point $\mathbf{x} \in \mathbb{R}^3$ and a unit-norm view direction $\mathbf{d} \in \mathbb{R}^3$ to a color $\mathbf{c} \in \mathbb{R}^3$ and a density value $\sigma \in \mathbb{R}$. During training, rays are cast from camera centers through their image plane. Each ray \mathbf{r} is associated with a direction vector \mathbf{d} and a set of points \mathbf{x}_i progressively sampled from the ray’s length [19]. The points \mathbf{x}_i and the direction vector \mathbf{d} are mapped by the NeRF to \mathbf{c}_i and σ_i . Defining $\delta_i = \|\mathbf{x}_{i+1} - \mathbf{x}_i\|$, the expected color of the ray $\hat{\mathcal{C}}(\mathbf{r})$ is given by:

$$\hat{\mathcal{C}}(\mathbf{r}) = \sum_{i=1}^N w_i \mathbf{c}_i \quad (1)$$

where

$$w_i = T_i(1 - \exp(-\sigma_i \delta_i)) \quad (2)$$

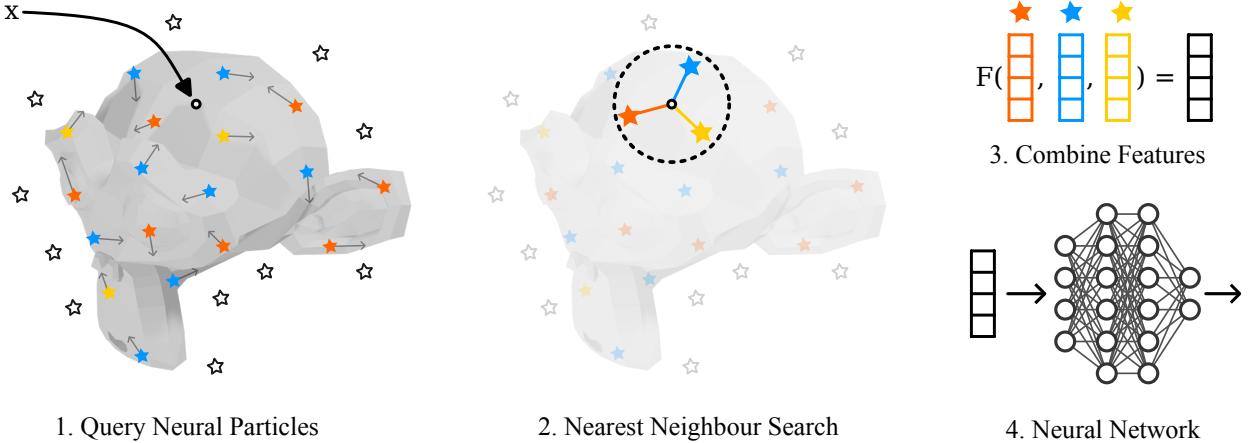


Figure 3. Illustration of our particle encoding. (1) A query point x is sampled in space. (2) The features and positions of the particles within a search radius are retrieved. (3) The features and distances from the query point are used to interpolate the feature at the query point x . (4) The resulting feature is evaluated by the neural network to give color and density. To train the encoding, the loss gradients are backpropagated through the network (4), the query feature (3), and finally into the positions and features of the particles in (2).

and

$$T_i = \exp \left(- \sum_{j=1}^{i-1} \sigma_j \delta_j \right) \quad (3)$$

A photometric loss is defined as

$$L_{\text{rgb}} = \sum_{r \in \mathcal{R}} \|\hat{C}(r) - c_{\text{gt}}(r)\|_2 \quad (4)$$

The ground-truth color $c_{\text{gt}}(r)$ is the color of the pixel in the training image that the ray r intersects. \mathcal{R} is the set of all rays that pass through the training images. Refer to [18, 30] for an in-depth derivation.

Encodings Given a point \mathbf{x}_i and a direction vector \mathbf{d}_i , a NeRF first maps each to an embedding space $E_x(\mathbf{x}_i)$ and $E_d(\mathbf{d}_i)$. The embeddings are subsequently consumed by a multilayered perceptron to output color and density. Not including an embedding layer critically deteriorates ability of the NeRF to reconstruct scenes [31]. The choice of the embedding layer can significantly reduce the size of the MLP required [4, 20, 35] and in some cases removes the need for one entirely [7]. Close to realtime performance is possible due to a particular class of encodings that the authors of [20] refer to as “parametric”. These encodings are characterized by a discrete datastructure that can be queried by the incoming coordinate \mathbf{x}_i to produce a set of features \mathbf{f}_j . The features \mathbf{f}_j are then combined to produce an embedding f_i . By storing the features in memory, a large MLP is no longer required and significant savings are made in both rendering and training times when compared to other formulations that use functional encodings [2, 3, 5, 19, 33].

3. ParticleNeRF

Our particle encoding is formulated as a neural point-cloud $\mathcal{P} = \{(\mathbf{x}_i, \mathbf{v}_i, \mathbf{f}_i) : i = 1, 2, \dots, M\}$, where $\mathbf{x}_i \in \mathbb{R}^3$ is the position of a latent particle in the system, $\mathbf{v}_i \in \mathbb{R}^3$ is the velocity of the particle, $\mathbf{f}_i \in \mathbb{R}^m$ is its associated feature, and M is the total number of particles. The pointcloud \mathcal{P} is an irregular grid of latent features that defines the map F from a coordinate $\mathbf{x}_j \in \mathbb{R}^3$ to f_j .

$$f_j = F(\mathbf{x}_j, \mathcal{P}) = \sum_{(\mathbf{x}_i, \mathbf{f}_i) \in \mathcal{P}} w(\|\mathbf{x}_j - \mathbf{x}_i\|_2) \mathbf{f}_i \quad (5)$$

where w is the bump function – a compactly supported radial basis function (RBF) – given by:

$$w(r) = \begin{cases} \exp \left(-\frac{s^2}{s^2 - r^2} \right), & \text{for } r \in (-s^2, s^2) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

The search radius s controls the influence of a particle on a region of space. Using a compact RBF allows the efficient calculation of f_j with a fixed-radius nearest neighbour search algorithm [32] where the radius is set to s .

ParticleNeRF uses the same architecture as InstantNGP (a 3 layer MLP with 64 neurons each), but replaces its hash encoding with our particle encoding. Our encoding is invariant to rotations and translations since a feature at a query point x is calculated based on the distances to neighboring particles and not their positions. The invariance is important since rigidly transforming a set of neighbouring particles should not affect their underlying geometry. Note that

if a query point finds no neighbouring particles, the feature is set to 0.

We optimize the MLP parameters Φ , the particle features \mathbf{f}_i and the particle positions \mathbf{x}_i relative to the NeRF loss L_{rgb} defined in Eq. (4):

$$\Phi^*, \{\mathbf{f}_i\}^*, \{\mathbf{x}_i\}^* = \arg \min_{\Phi, \{\mathbf{f}_i\}, \{\mathbf{x}_i\}} L_{\text{rgb}} \quad (7)$$

The gradients of the NeRF loss are used to update the position of the particles. This allows them to move with the geometry in the scene as it changes and enables better performance on online dynamic scenes.

3.1. Position Based Dynamics

We observe that propagating the NeRF loss into the particle positions can in some instances lead to multiple particles accumulating in a small region of space. Although this does not adversely affect the reconstruction quality, it is not desirable because it needlessly increases the number of neighbours that each query point has to process during training and evaluation. To address this problem in a structured way, we build a position-based physics system [17,21] into InstantNGP that resolves the dynamics of our particles. Position-based dynamics (PBD) is a simple, robust, and fast physics system that can evolve the motion of our particles while resolving the constraints that we put on them. We interpret the gradients $\frac{dL_{\text{rgb}}}{d\mathbf{x}_i}$ as scaled velocity vectors that are added to the particles' most recent velocity. Using PBD not only provides a means of limiting the number of neighbours a particle has, but it also creates future opportunities to add other constraints into the system. For example, if an object is known to be a rigid object, or part of an articulated body, the particles can be suitably constrained.

Algorithm 1 PBD Physics Step

```

1: for all particles  $i$  do
2:    $\mathbf{v}_i \leftarrow \gamma \mathbf{v}_i - \alpha \frac{dL_{\text{rgb}}}{d\mathbf{x}_i}$ 
3:    $\mathbf{p}_i \leftarrow \mathbf{x}_i$ 
4:    $\mathbf{x}_i \leftarrow \mathbf{x}_i + \Delta t \mathbf{v}_i$ 
5: for all collision pairs  $i, j$  do
6:    $l \leftarrow \|\mathbf{x}_2 - \mathbf{x}_1\|$ 
7:    $\mathbf{x}_i \leftarrow \mathbf{x}_i + 0.5(l - \delta) \frac{\mathbf{x}_2 - \mathbf{x}_1}{l}$ 
8: for all particles  $i$  do
9:    $\mathbf{v}_i \leftarrow (\mathbf{x}_i - \mathbf{p}_i)/\Delta t$ 

```

In all our experiments, we choose a damping factor $\gamma = 0.96$, a timestep $\Delta t = 0.01$, a minimum distance of $\delta = 0.01$ and a gradient scaling factor $\alpha = 10$.

3.2. Implementation

We build our particle encoding as an extension of the InstantNGP implementation released by Müller *et al.* [20]. The base InstantNGP implementation transforms a given scene so that it fits within a unit cube. Our minimum distance δ in the physics system and the search radius s that is used to query neighbouring particles are given in the units of the unit cube. InstantNGP uses an occupancy grid as an acceleration structure. For both ParticleNeRF and InstantNGP experiments, the structure is updated on every training step.

We implement the sorting-based fixed-radius nearest search algorithm described by [10] and the position-based dynamics physics system described by [21] as additional CUDA kernels [22] within InstantNGP.

Two different Adam optimizers [13] are used. The first is for the parameters of the MLP Φ and the second is for the particle features \mathbf{f}_i . Both are parameterized by $\beta_1 = 0.9$, $\beta_2 = 0.99$, $\epsilon = 10^{-10}$ with the learning rate set at 0.01.

4. Experiments

Our particle encoding is intended to be used on dynamic scenes in an incremental fashion. In Section 4.1, we show how the particle encoding performs on a dataset that is commonly used in the NeRF literature. Section 4.2 uses a dataset that is specifically designed to ablate the performance of the encoding.

For all the experiments involving the particle encoding, we initialize the unit cube containing the scene with a grid of uniformly spaced particles. Each particle is associated with a 4 dimensional feature and is initialized randomly between $\pm 10^{-2}$.

4.1. Animated Blender Dataset

We compare ParticleNeRF against InstantNGP [20] since it is the fastest implementation of NeRF available and is therefore most suitable for online applications [12]. It is also the implementation of choice for many NeRF-based robotic applications that would benefit from an online dynamic NeRF [1,12,14,26,27]. To date, there are no dynamic methods [9,15,23–25,34] that train at the speeds required for an online application. We baseline our method using the standard “Blender” dataset [19]. Each scene in the dataset has a central object that is being looked upon by 100 cameras distributed on the top half of the surface of a sphere. The Blender dataset, however, is only designed for static scenes and in its released form is not appropriate for evaluating the online challenge. To introduce movement, we animate each of the scenes over 100 frames. Two types of motions are synthesized. In the first set of experiments, we rotate the central object around its up axis by a fixed amount per animation frame that ranges between 1° per frame to 4°

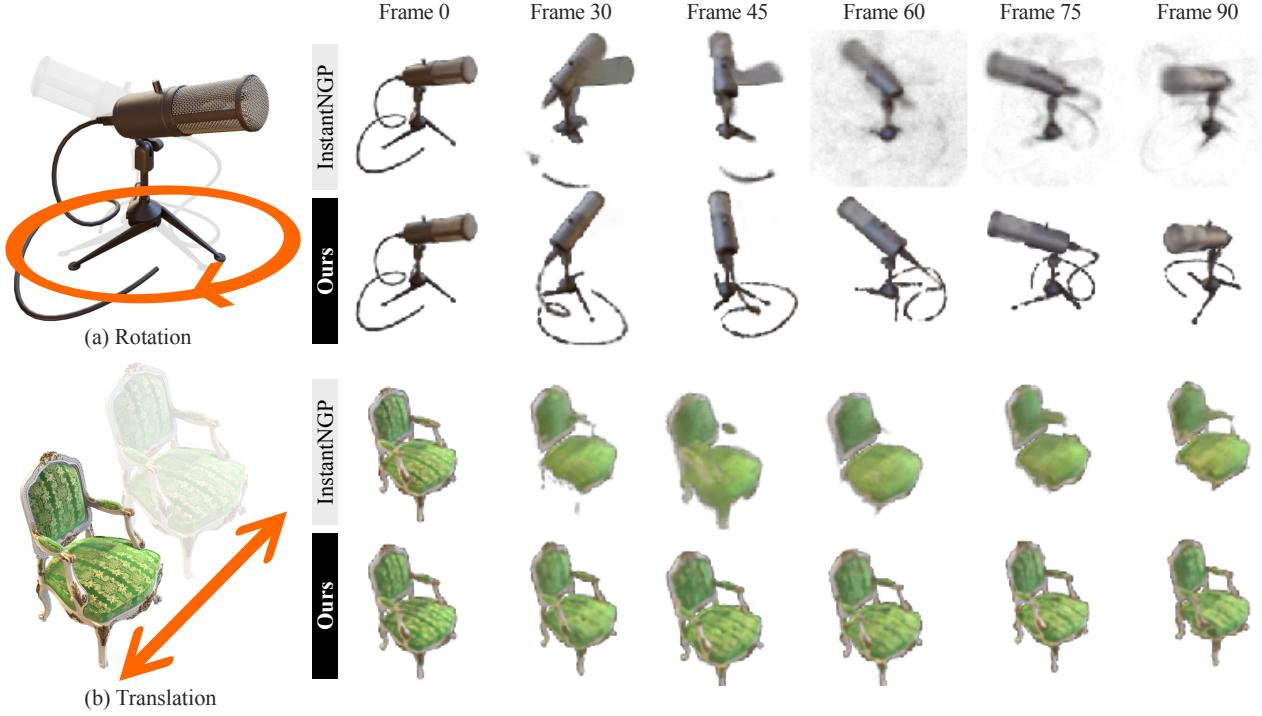


Figure 4. We test our encoding on an animated version of the Blender dataset [19]. **(a)** shows an object rotating around its up axis. **(b)** shows an object translating from side to side. InstantNGP cannot learn features fast enough to maintain a high quality reconstruction. ParticleNeRF is able to move its features in space and maintain the structure of the object.

Table 1. Performance of InstantNGP and ParticleNerf on the Animated Blender Dataset reported through the mean and standard deviation of the photometric PSNR over 100 frames

	Step per Frame	Model Encoding	Chair	Drums	Ficus	Hotdog	Lego	Mic	Ship
Static		Hash	26.98	22.49	22.37	30.80	27.63	29.91	22.56
		Particle	23.91	18.44	19.50	26.74	22.41	26.39	21.30
Rotation	1°	Hash	20.29 ± 2.4	16.74 ± 1.2	19.20 ± 0.8	26.32 ± 1.3	20.15 ± 1.2	23.49 ± 1.3	22.31 ± 2.0
		Particle	23.04 ± 1.3	18.00 ± 1.1	19.49 ± 0.8	26.59 ± 0.7	21.77 ± 0.8	25.13 ± 0.7	21.73 ± 1.8
	2°	Hash	16.60 ± 1.8	15.38 ± 1.1	18.48 ± 0.8	24.11 ± 1.5	18.06 ± 1.3	19.35 ± 2.0	21.40 ± 2.2
		Particle	21.88 ± 1.3	17.36 ± 1.1	19.05 ± 0.8	25.79 ± 0.7	20.91 ± 0.8	23.66 ± 0.8	21.55 ± 1.8
Translation	3°	Hash	15.98 ± 1.7	15.03 ± 1.1	17.92 ± 0.9	23.51 ± 1.3	17.00 ± 1.3	18.18 ± 1.9	20.99 ± 2.3
		Particle	20.97 ± 1.4	16.82 ± 1.1	18.50 ± 0.8	25.16 ± 0.8	20.28 ± 0.8	22.82 ± 0.8	21.05 ± 1.9
	4°	Hash	15.73 ± 1.7	13.91 ± 1.2	17.67 ± 0.9	23.10 ± 1.4	16.05 ± 1.5	16.60 ± 2.0	20.92 ± 2.3
		Particle	19.81 ± 1.7	16.15 ± 1.1	17.67 ± 0.9	24.20 ± 0.9	19.41 ± 1.0	21.44 ± 1.1	20.95 ± 1.9
Translation	1 cm	Hash	22.19 ± 2.1	17.31 ± 1.4	18.98 ± 0.8	25.07 ± 1.9	20.71 ± 1.5	24.68 ± 1.4	20.32 ± 2.1
		Particle	23.12 ± 1.3	18.09 ± 1.1	19.38 ± 0.8	25.05 ± 1.5	21.78 ± 0.9	25.15 ± 0.7	20.71 ± 1.9
	2 cm	Hash	18.16 ± 2.0	15.82 ± 1.3	18.21 ± 0.9	22.62 ± 1.8	19.10 ± 1.6	19.52 ± 2.2	18.76 ± 2.2
		Particle	22.03 ± 1.3	17.46 ± 1.1	18.42 ± 0.7	24.22 ± 1.2	20.99 ± 0.9	23.72 ± 0.8	20.79 ± 1.8
	3 cm	Hash	17.01 ± 1.9	15.18 ± 1.3	17.94 ± 0.9	20.37 ± 1.9	18.42 ± 1.7	17.36 ± 2.6	18.81 ± 2.0
		Particle	20.77 ± 1.4	16.77 ± 1.1	17.52 ± 0.8	23.11 ± 1.3	20.15 ± 0.9	22.40 ± 1.0	20.05 ± 1.8

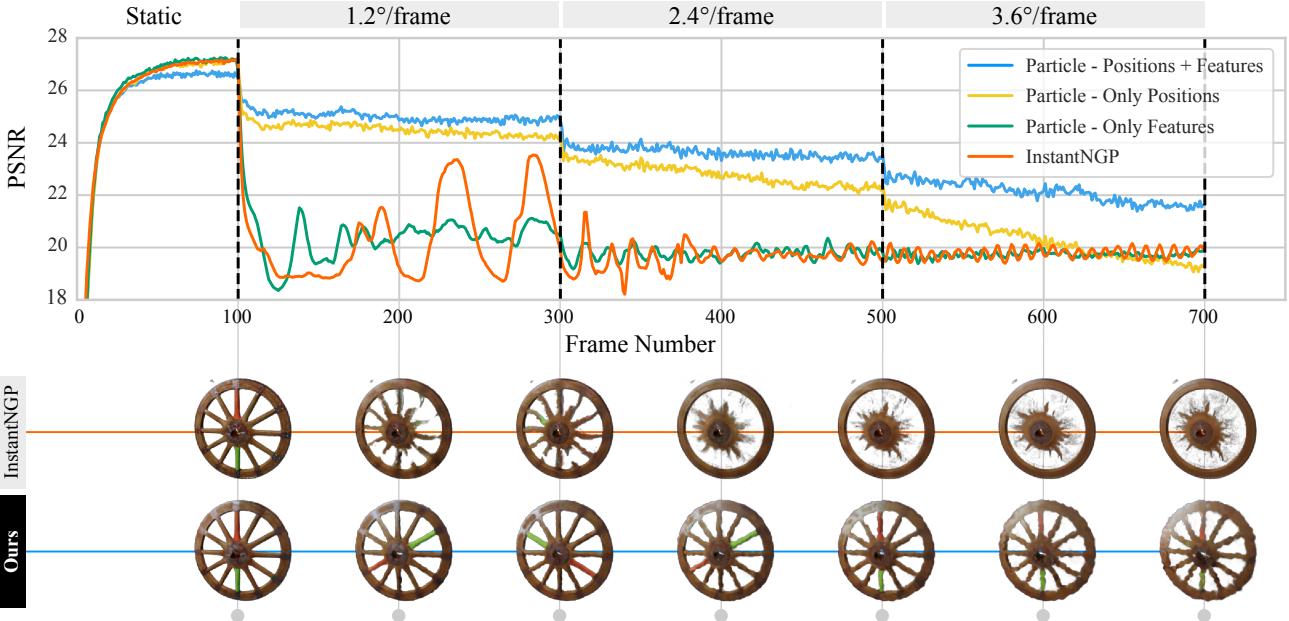


Figure 5. On the wheel dataset, our method succeeds at maintaining the structure of the wheel as it rotates. We show how our particle encoding performs in three settings: (i) When both the features and the positions of the particles are optimized (blue). (ii) When only the positions of the particles are optimized (yellow). (iii) When the only the particle features are optimized (green). Lastly we show how InstantNGP performs (orange). We visually display the results of the reconstruction from an unseen viewpoint at the frames indicated for both InstantNGP and (i). A visual inspection reveals that even at the lowest PSNR at frame 700, our method preserves the wheel’s structure.

per frame. In the second set, we translate the objects by a fixed amount per frame along a line that is 10cm long. The steps range between 1 cm per frame and 3 cm per frame. The two type of motions are illustrated in Fig. 4a and Fig. 4b respectively.

For each experiment, the scene is held static for 1500 training steps. The results of the static reconstruction at the end of this phase are reported in the first section of Table 1. Following the static phase, the animation begins. At every animation frame, the central object is moved. Both InstantNGP and ParticleNeRF are allowed only 5 training iterations before the next frame is loaded. At the end of every frame, we evaluate the photometric reconstruction quality (PSNR) from 7 unseen views. When the animation has completed, we record the mean and standard deviation of the PSNR across all the frames. The results are shown in Table 1. For the ParticleNeRF experiments, we initialize the scene with 100,000 particles with a search radius s of 0.04.

Table 1 shows that while InstantNGP has superior reconstruction quality on static scenes, its performance deteriorates significantly under motion. ParticleNeRF’s reconstruction ability gracefully declines with the speed of the motion until the particles can no longer move as fast as the object. At the end of each phase of motion, we render the

current state of the wheel from an unseen viewpoint and display the results in Figure 4. The visualizations clearly show that InstantNGP cannot learn features fast enough. In contrast, ParticleNeRF maintains the structure of the object despite the movement because it is capable of both relearning and relocating features in the embedding space to explain the changes in the scene.

4.2. Wheel Dataset and Ablation

To more thoroughly probe the dynamic representational ability of our particle encoding, we introduce and publish another dataset comprised of a single object – a wheel with spokes shown in Figure 1 – that is made to rotate at various speeds along its main axle. The wheel is kept stationary for the first 100 frames and then made to rotate at successively faster speeds over the next 600 frames. Both InstantNGP and ParticleNeRF are given only 5 training steps per frame and are configured with the same parameters from Section 4.1. Ten cameras are distributed on the top half of a sphere looking at the center of the wheel. Every frame, the cameras produce new training images which replace the previous training images given to ParticleNeRF and InstantNGP. Evaluation occurs at every frame using 10 unseen viewpoints. We include 2 ablations that analyze the effect

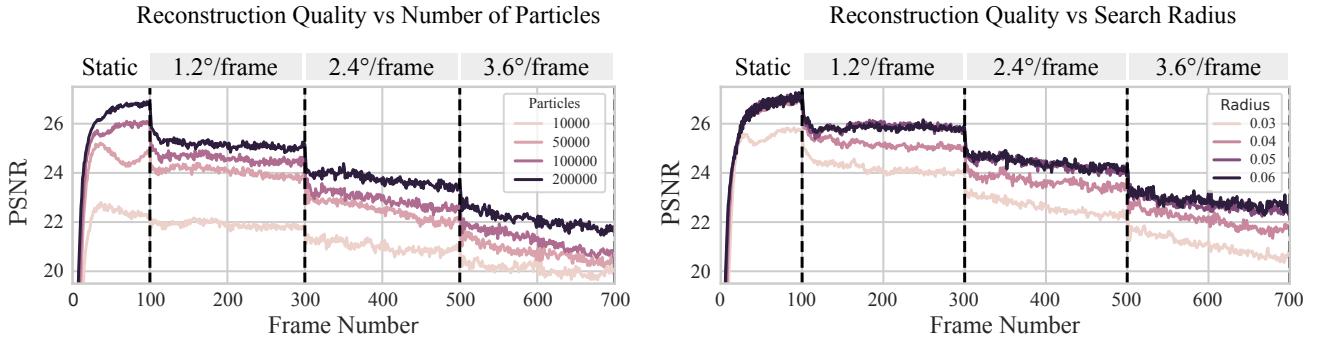


Figure 6. An ablation showing the effect of the number of particles (left) and the particle search radius (right) on the photometric reconstruction quality. When ablating the number of particles, search radius is set at 0.04. When ablating the search radius, number of particles is set at 100,000. In both cases, there is a diminishing effect.

of exclusively optimizing on either particle positions or particle features. In all experiments, we allow the features to train in the static phase and only disable the optimization in the dynamic phase. In contrast, when position optimization is disabled, it is done from the start of training.

The quantitative and qualitative results in Figure 5 highlight the significant advantage of our model over InstantNGP in dynamic settings. Although InstantNGP is superior at capturing detailed textures in static scenes, it is not capable of adapting to a moving one. We observe a ghosting effect where InstantNGP quickly recovers structure upon the wheel’s return to its original configuration. This is manifested as oscillations in the PSNR metric in Figure 5. ParticleNeRF exhibits high performance and the influence of backpropagating into the particle positions is clearly shown as the primary contributor. Allowing features to be optimized simultaneously reduces the accumulation of errors over frames. At high speeds, ParticleNeRF’s reconstruction ability begins to deteriorate. However, the deterioration is not catastrophic – as is evidenced by the visual render taken at the 700th frame in Figure 5. Instead it can be likened to a reduction in resolution caused by particles gradually leaving the structure. We have observed that once a particle develops a strong feature with no neighbours, it can no longer be reincorporated into a larger corpus of particles. This is the fundamental cause of the degradation and is a topic for future work which we believe can be addressed by adding a cohesive constraint through the physics system or alternatively by decaying the feature vectors associated with lone particles. Finally, we ablate our system against a varying number of particles and search radii and we observe diminishing returns with increased values – Figure 6.

5. Limitations and Future Work

Towards Realtime On an Nvidia RTX3090, InstantNGP averages 7 ms per training step at a batch size of 262,144.

Conversely, our particle encoding averages 41 ms per training step. We believe this gap can be closed because the bulk of the time is not spent on computations but rather on waiting for CUDA atomic operations to complete when backpropagating gradients into the particle positions and features. Multiple queries use the same neighbouring particles and create a resource contention which is exacerbated over larger search radii. Future work can overcome this challenge by creating algorithms that minimize this contention.

Physics System We incorporate a physics system as a means of controlling particle dynamics. We add collision constraints that prevent the accumulation of particles in any one location. We believe that the addition of other constraints are an excellent means of adding prior knowledge of the scene into the reconstruction. For example, when an object is known to be articulated, the particles can be constrained to move along a skeleton. This provides a non visual prior to complement the NeRF photometric loss and will likely allow the NeRF to maintain a high quality reconstruction at higher speeds.

Static Reconstruction While our particle encoding is better suited for online dynamic scenes, it can only express blurred textures. For many applications – where the scene is constantly changing and the geometry is of more interest than the texture – this may be an acceptable trade.

6. Conclusion

We have introduced the challenge of building an online NeRF of a dynamic scene. We created a new parametric encoding that is particle-based. By backpropagating the NeRF loss into the particle positions, we demonstrate that moving learnt features is better than relearning them in dynamic settings. We think this work makes significant progress towards using NeRFs as realtime state representations for dynamic environments.

References

- [1] Jad Abou-Chakra, Feras Dayoub, and Niko Sünderhauf. Implicit object mapping with noisy data. *arXiv preprint arXiv:2204.10516*, 2022. 5
- [2] Jonathan T. Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P. Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields, 2021. 1, 4
- [3] 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*, pages 5470–5479, 2022. 1, 4
- [4] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *European Conference on Computer Vision (ECCV)*, 2022. 2, 3, 4
- [5] Shin-Fang Chng, Sameera Ramasinghe, Jamie Sherrah, and Simon Lucey. Garf: Gaussian activated radiance fields for high fidelity reconstruction and pose estimation. *arXiv e-prints*, pages arXiv–2204, 2022. 1, 4
- [6] Jiemin Fang, Taoran Yi, Xinggang Wang, Lingxi Xie, Xiaopeng Zhang, Wenyu Liu, Matthias Nießner, and Qi Tian. Fast dynamic radiance fields with time-aware neural voxels. *arxiv:2205.15285*, 2022. 2
- [7] 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, June 2022. 2, 4
- [8] Wanshui Gan, Hongbin Xu, Yi Huang, Shifeng Chen, and Naoto Yokoya. V4d: Voxel for 4d novel view synthesis. *arXiv preprint arXiv:2205.14332*, 2022. 2
- [9] Chen Gao, Ayush Saraf, Johannes Kopf, and Jia-Bin Huang. Dynamic view synthesis from dynamic monocular video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5712–5721, October 2021. 3, 5
- [10] Simon Green. Particle simulation using cuda. *NVIDIA whitepaper*, 2010. 5
- [11] Shanyan Guan, Huayu Deng, Yunbo Wang, and Xiaokang Yang. Neurofluid: Fluid dynamics grounding with particle-driven neural radiance fields. In *ICML*, 2022. 3
- [12] Justin Kerr, Letian Fu, Huang Huang, Jeffrey Ichnowski, Matthew Tancik, Yahav Avigal, Angjoo Kanazawa, and Ken Goldberg. Evo-neRF: Evolving neRF for sequential robot grasping. In *6th Annual Conference on Robot Learning*, 2022. 5
- [13] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5
- [14] Simon Klenk, Lukas Koestler, Davide Scaramuzza, and Daniel Cremers. E-nerf: Neural radiance fields from a moving event camera. *arXiv preprint arXiv:2208.11300*, 2022. 5
- [15] Tianye Li, Mira Slavcheva, Michael Zollhöfer, Simon Green, Christoph Lassner, Changil Kim, Tanner Schmidt, Steven Lovegrove, Michael Goesele, Richard Newcombe, and Zhaoyang Lv. Neural 3d video synthesis from multi-view video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5521–5531, June 2022. 3, 5
- [16] Ruofan Liang, Jiahao Zhang, Haoda Li, Chen Yang, and Nandita Vijaykumar. Spidr: Sdf-based neural point fields for illumination and deformation. *arXiv preprint arXiv:2210.08398*, 2022. 3
- [17] Miles Macklin, Matthias Müller, and Nuttapong Chentanez. Xpb: Position-based simulation of compliant constrained dynamics. In *Proceedings of the 9th International Conference on Motion in Games, MIG ’16*, page 49–54, New York, NY, USA, 2016. Association for Computing Machinery. 5
- [18] N. Max. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 1(2):99–108, 1995. 4
- [19] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 1, 3, 4, 5, 6
- [20] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4):102:1–102:15, July 2022. 1, 2, 3, 4, 5
- [21] Matthias Müller, Bruno Heidelberger, Marcus Hennix, and John Ratcliff. Position based dynamics. *Journal of Visual Communication and Image Representation*, 18(2):109–118, 2007. 5
- [22] NVIDIA, Péter Vingelmann, and Frank H.P. Fitzek. Cuda, release: 10.2.89, 2020. 5
- [23] Keunhong Park, Utkarsh Sinha, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Steven M. Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. *ICCV*, 2021. 1, 2, 3, 5
- [24] Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Ricardo Martin-Brualla, and Steven M. Seitz. Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields. *ACM Trans. Graph.*, 40(6), dec 2021. 1, 2, 3, 5
- [25] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. *arXiv preprint arXiv:2011.13961*, 2020. 1, 3, 5
- [26] Antoni Rosinol, John J Leonard, and Luca Carlone. Nerf-slam: Real-time dense monocular slam with neural radiance fields. *arXiv preprint arXiv:2210.13641*, 2022. 5
- [27] Nur Muhammad Mahi Shafullah, Chris Paxton, Lerrel Pinto, Soumith Chintala, and Arthur Szlam. Clip-fields: Weakly supervised semantic fields for robotic memory. *arXiv preprint arXiv:2210.05663*, 2022. 5
- [28] Liangchen Song, Anpei Chen, Zhong Li, Zhang Chen, Lele Chen, Junsong Yuan, Yi Xu, and Andreas Geiger. Nerf-player: A streamable dynamic scene representation with decomposed neural radiance fields, 2022. 2
- [29] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct voxel grid optimization: Super-fast convergence for radiance fields

- reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5459–5469, 2022. 2, 3
- [30] Andrea Tagliasacchi and Ben Mildenhall. Volume rendering digest (for nerf). *arXiv preprint arXiv:2209.02417*, 2022. 4
- [31] Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *NeurIPS*, 2020. 3, 4
- [32] Matthias Teschner, Bruno Heidelberger, Matthias Müller, Danat Pomeranets, and Markus Gross. Optimized spatial hashing for collision detection of deformable objects. *VMV'03: Proceedings of the Vision, Modeling, Visualization*, 3, 12 2003. 4
- [33] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T. Barron, and Pratul P. Srinivasan. Ref-NeRF: Structured view-dependent appearance for neural radiance fields. *CVPR*, 2022. 1, 4
- [34] Wenqi Xian, Jia-Bin Huang, Johannes Kopf, and Changil Kim. Space-time neural irradiance fields for free-viewpoint video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9421–9431, June 2021. 3, 5
- [35] Qiangeng Xu, Zexiang Xu, Julien Philip, Sai Bi, Zhixin Shu, Kalyan Sunkavalli, and Ulrich Neumann. Point-nerf: Point-based neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5438–5448, 2022. 3, 4