

Dynamic Appearance Particle Neural Radiance Field

Ancheng Lin,
Jun Li,

University of Technology Sydney

Abstract

Neural Radiance Fields (NeRFs) have shown great potential in modelling 3D scenes. Dynamic NeRFs extend this model by capturing time-varying elements, typically using deformation fields. The existing dynamic NeRFs employ a similar Eulerian representation for both light radiance and deformation fields. This leads to a close coupling of appearance and motion and lacks a physical interpretation. In this work, we propose Dynamic Appearance Particle Neural Radiance Field (DAP-NeRF), which introduces particle-based representation to model the motions of visual elements in a dynamic 3D scene. DAP-NeRF consists of superposition of a static field and a dynamic field. The dynamic field is quantised as a collection of *appearance particles*, which carries the visual information of a small dynamic element in the scene and is equipped with a motion model. All components, including the static field, the visual features and motion models of the particles, are learned from monocular videos without any prior geometric knowledge of the scene. We develop an efficient computational framework for the particle-based model. We also construct a new dataset to evaluate motion modelling. Experimental results show that DAP-NeRF is an effective technique to capture not only the appearance but also the physically meaningful motions in a 3D dynamic scene.

1 Introduction

Recent advancements in neural radiance fields (NeRFs) (Mildenhall et al. 2020) have shown remarkable success in modelling 3D scenes with a continuous representation. These models facilitate high-fidelity rendering from novel viewpoints without requiring explicit geometry. One of the most promising extensions is Dynamic NeRFs (Pumarola et al. 2021; Liu et al. 2022), which aims to model dynamic scenes by incorporating a time-varying radiance field. There are two primary schemes: directly adding a time dimension or introducing a deformation field before the canonical radiance field. The deformation-field scheme separately represents the motion and appearance of a dynamic scene and has gained popularity due to its effectiveness.

Existing models employ an Eulerian formulation for both the deformation field and the appearance field. A learnable field model, such as a group of neural networks, uses the 3D Euclidean coordinates as its input query and outputs the desired physical quantities (field variables). This scheme is

effective when the field needs to be defined throughout the entire Euclidean space, such as when considering the light scattering characteristics over a three-dimensional scene. However, the Eulerian field representation can be problematic when the quantities of interest are confined to specific regions or supported on a sub-manifold within the Euclidean space, for instance within a solid object or on the surface. It might waste model capacity and make the subsequent use of the model inconvenient. For example, existing methods that apply dynamic NeRF to explore the interaction between objects and the environment require manual preprocessing like shape extraction (Qiao, Gao, and Lin 2022) and inverse deformation estimation (Chen et al. 2022).

To address these limitations, we propose to employ particle-based representation for moving or deforming objects within dynamic scenes. In our method, particles represent a finite approximation of the distribution of physical quantities that determine appearance. More specifically, in a volumetric rendering scheme, if the light properties, colour, and scattering probability at a location $x \in \mathcal{X}$ are determined by a view-independent physical feature f and the viewpoint, then the spatial distribution of f over \mathcal{X} becomes the primary focus of modeling. For this purpose, particles are used to quantise the space \mathcal{X} , with the advantage that the distribution of f can be made time-varying by using a movement model for the particles. This particle-based representation can be integrated with an existing Eulerian appearance field that represents static elements. Therefore, we achieve a hybrid (static-Eulerian, dynamic-Lagrangian) NeRF model that can learn from monocular videos using only photometric supervision.

In summary, this work introduces the Dynamic Appearance Particle Neural Radiance Field (**DAP-NeRF**). The major contributions are as follows:

- We have designed a hybrid framework of radiance field models. The dynamic elements of the fields are described using a particle-based representation, which corresponds to a Lagrangian approach to field models. The introduction of the Lagrangian model complements the widely adopted Eulerian dynamic NeRFs, which specify the static elements of the scene in our framework. The hybrid framework serves as more than an adequate appearance model of dynamic scenes. The employed particles also provide an explicitly interpretable and physically mean-

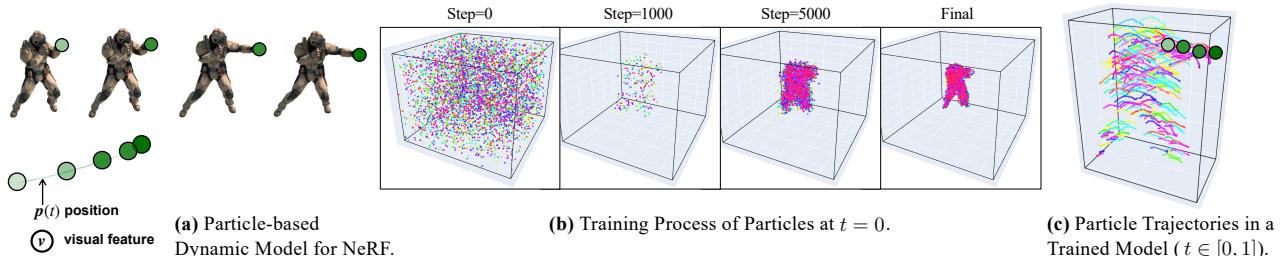


Figure 1: Method Overview. (a) Particles represent observed movements. A particle p is associated with a small volume of material corresponding to a semantically meaningful part of a moving object. $p(t)$ forms an explicit dynamic model of the object part. v denotes the visual feature of the particle. (See Sec. 3.1) (b) Panes show the particles at $t = 0$ during different stages of training. The particles are corresponding to $t = 0$. Notice particle removal in early training and re-sampling in the latter stages. (See Sec. 3.2) (c) Particle trajectories in a trained dynamic model. Each trajectory is generated by some particle- i $\{p^i(t)|t \in [0, 1]\}$.

ingful description of the motions.

- We have developed an efficient and effective computational structure for the proposed hybrid framework. This structure has the capability to automatically identify dynamic and static elements.
- We have constructed a dataset and introduced a new metric to evaluate the motion modelling of dynamic NeRFs.

Empirical studies have shown that our framework achieves state-of-the-art performance in novel view synthesis tasks. We've demonstrated that our framework produces particles that effectively capture the dynamics of moving objects, facilitating scene decoupling and motion editing. Moreover, the learned particle motion model shows superior quantitative results in motion modelling, surpassing existing methods that deform a canonical field.

2 Related Work

2.1 Dynamic NeRF

Neural Radiance Field (NeRF) is a powerful technique that can learn high-quality 3D scene representations from a set of posed images. Due to its rendering quality and flexibility, NeRF has been extended to various areas, including generative models (Jiang et al. 2023) and SLAM (Li et al. 2023a).

Recently, there has been growing interest in applying NeRFs to scenarios where objects are moving or undergoing deformation. Dynamic NeRFs can represent dynamic scenes using synchronised multi-view videos (Liu et al. 2022; Lombardi et al. 2019; Wang et al. 2022) and even monocular videos (Pumarola et al. 2021; Tretschk et al. 2021; Fang et al. 2022). There are mainly two categories of methods. The first class directly learns a time-conditional radiance field by adding a time dimension (Gao et al. 2021; Xian et al. 2021; Li et al. 2021; Gan et al. 2022), which usually requires additional geometry regularisation, such as scene depth and optical flow. Another series of methods (Pumarola et al. 2021; Liu et al. 2022; Tretschk et al. 2021; Fang et al. 2022; Guo et al. 2022; Park et al. 2021a) decouples the model into a deformation field and a canonical radiance field. The deformation field captures the geometry changes and maps 3D

points into a canonical space for radiance querying. However, this approach can struggle with motion inconsistency in topologically varying scenes, leading to the development of alternative methods (Park et al. 2021b; Song et al. 2022).

2.2 Efficient Distributed Representation

The original NeRF (Mildenhall et al. 2020) mapping raw spatial coordinates to light radiance can be time-consuming due to a large number of MLP forward passes. A recent advance is to distribute the field representation to individual voxels. DVGO (Sun, Sun, and Chen 2022) employs a dense voxel grid and a small MLP to significantly accelerate NeRF training and rendering. Instant-NGP (Müller et al. 2022) uses a hashing technique to reduce voxel grid storage costs and improve optimisation speed and efficiency.

The representations introduced to static NeRFs have also been adopted by dynamic NeRFs. (Fang et al. 2022) are the first to use voxel grids in the canonical radiance field, showing high training efficiency. (Guo et al. 2022) improve the deformation field with a hybrid representation (voxel grid and MLP). (Park et al. 2023) and (Cao and Johnson 2023) propose using 4D hash grids to decrease memory usage. (Fridovich-Keil et al. 2023) introduce a factorisation approach to the 4D grid, achieving a decomposition into static and dynamic components.

2.3 Point-based Representation

This sub-area aims to discretise a radiance field using 3D points. Point-NeRF (Xu et al. 2022) was an early attempt using a point cloud representation encoding local geometry and appearance. However, it is restricted to static scenes and requires depth data. NeuroFluid (Guan et al. 2022) integrates particles into NeRF for fluid modelling, but lacks realistic texture and only works on simulated scenes. More closely related to the presented work, Particle-NeRF (Abou-Chakra, Dayoub, and Sünderhauf 2022) encodes a scene completely in particles. The particles represent a moment of a 3D scene and can be incrementally adapted to the next moment. There isn't a motion model for the entire life cycle of particles. Two concurrent works (Li et al. 2023b; Uzolas, Eisemann, and Kellnhofer 2023) focus on scenes that conform to specific

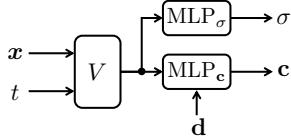


Figure 2: Commonly adopted structure of dynamic NeRFs.

physical models. They only address foreground dynamic objects with pre-extracted initial shapes.

3 Dynamic Appearance Particle NeRF

Consider the dynamic light scattering field model: A pixel of the camera frame at time t is computed by casting a ray $\mathbf{r}(s) = \mathbf{o} + s\mathbf{d}$, $s \in \mathbb{R}^+$ from the camera centre \mathbf{o} to the pixel on the image plane:

$$\mathbf{C}(\mathbf{r}, t) = \int_{s_n}^{s_f} T(s)\sigma(\mathbf{r}(s), t)\mathbf{c}(\mathbf{r}(s), \mathbf{d}, t)ds \quad (1)$$

where $\mathbf{C}(\mathbf{r}, t)$ denotes the pixel colour rendered at time $t \in \mathbb{R}^+$, s_n, s_f denote the near and far bounds of the ray. $\mathbf{r}(\cdot)$ returns a 3D location along the ray. $T(s) = \exp(-\int_{s_n}^s \sigma(\mathbf{r}(s'), t)ds')$ is the accumulated transmittance. Given a location and time t (and direction \mathbf{d}), $\sigma(\cdot)$ stands for the volume density, and $\mathbf{c}(\cdot)$ stands for the emitted radiance.

An effective strategy to specify the model of $\sigma(\cdot)$ and $\mathbf{c}(\cdot)$ is to learn small MLPs and localised feature vectors distributed at grid nodes in the 3D region of interest. Fig. 2 displays the computational structure. The generic formulation of the components are as follows:

$$V(\mathbf{x}, t) : \mathbb{R}^3 \times \mathbb{R} \mapsto \mathbb{R}^C \quad (2)$$

$$\text{MLP}_\sigma(\cdot) : \mathbb{R}^C \mapsto \mathbb{R} \quad (3)$$

$$\text{MLP}_c(\cdot) : \mathbb{R}^C \times \mathbb{R}^3 \mapsto \mathbb{R}^3 \quad (4)$$

where the majority of the field information is modelled by V . We adopt this framework and embed the Lagrangian dynamic model in V . However, it should be noted that the proposed technique is not tightly coupled to a specific overarching architecture. It is possible to integrate the proposed Lagrangian representation into alternative frameworks, such as Instant-NGP mentioned in Sec. 2.

We model the feature field of the light radiance V by the following superposition:

$$V(\mathbf{x}, t) := (1 - w_t(\mathbf{x}))V^s(\mathbf{x}) + w_t(\mathbf{x})V_t^d(\mathbf{x}) \quad (5)$$

where $V^s(\mathbf{x})$ and $V_t^d(\mathbf{x})$ represent the static and dynamic components of the field, respectively. The static field component, $V^s(\mathbf{x})$, employs a grid-based model as in (2), but with the difference that $V^s(\mathbf{x})$ is time-independent. This aligns with the canonical feature field in (Fang et al. 2022). The particle-based dynamic field component, $V_t^d(\mathbf{x})$, is proposed by this work and will be introduced in details in the following subsection. The superposition is determined by $w_t(\mathbf{x})$, which takes value 1 if the dynamic field is effective at (\mathbf{x}, t) and 0 otherwise. The support of the dynamic field

(where it is effective) is implied by the particle representation of $V_t^d(\mathbf{x})$. Fig. 3 (a) shows this superposition scheme.

It is helpful to notice a denotation convention due to the hierarchy of concepts that are adopted in this work. We denote time t on the right hand side of (5) in subscripts, as opposed to an independent argument to the functions. This is to clarify the fact that the dynamic component $V_t^d(\cdot)$ consists of an ensemble of particles. When considering the composition of particles making $V_{t_0}^d(\cdot)$, the focus is on the discretisation of a field at a certain time t_0 . The time argument is fixed and does not affect the construction of the instantaneous status of $V_{t_0}^d(\cdot)$. The dynamics of the system is encoded in the individual particle models.

3.1 Particle-based Dynamic Model

The dynamic field V_t^d is a continuous function, as a component of a NeRF model. We are concerned with integrating V_t^d in finite volumes. Therefore, we formulate V_t^d as

$$\begin{aligned} V_t^d(\mathbf{x}) &= (V_t^d * \delta)(\mathbf{x}) \\ &= \int V_t^d(\mathbf{x}')\delta(\mathbf{x} - \mathbf{x}')d\nu(\mathbf{x}') \end{aligned} \quad (6)$$

where $\delta(\cdot)$ is the Dirac function $\delta(\mathbf{r}) = 0$, $\mathbf{r} \neq 0$ and $\int \delta(\mathbf{r})d\nu = 1$ and $\nu(\cdot)$ is the volume integration variable.

Replacing the $\delta(\cdot)$ by a kernel with finite support, $\delta \rightarrow W$, where W is a kernel locally supported close to 0 ($W(\mathbf{x}) = 0$ when $\|\mathbf{x}\|$ exceeds a small radius and $\int W(\mathbf{x})d\nu(\mathbf{x}) = 1$), the finite-width kernel approximation leads to a quantisation scheme of a physical field using the notion of *particles* (Gingold and Monaghan 1977). A particle represents the interesting physical quantities within a small spatial extent. In the NeRF problem, it is the *visual features* that are of the central interest, i.e. the vector quantity produced by the module V in (2). The visual feature is fed into subsequent MLPs to output the light radiance properties (density and colour). In our model, the visual feature field is represented via smoothed particles. When querying a location \mathbf{x} at time t ,

$$V_t^d(\mathbf{x}) \approx \sum_{i=1}^{N_p} \mathbf{v}^i W(\mathbf{x} - \mathbf{p}^i(t)) \quad (7)$$

where i specifies one of the N_p particles. As shown in Fig. 3 (b), each particle represents the movements of a small finite volume of certain visual characteristics, consisting of $\mathbf{p}^i(t)$ and \mathbf{v}^i . $\mathbf{p}^i(t)$ is a 3D trajectory, mapping time t to a 3D position. The latter, \mathbf{v}^i , is a time invariant appearance feature.

Due to how the particles contribute to the NeRF model, we call them *appearance particles* and will discuss their computational details in the next subsection.

3.2 Appearance Particle Model

Recall that an individual particle carries two pieces of dynamics information, a visual feature \mathbf{v} and time-varying position $\mathbf{p}(t)$. In DAP-NeRF, \mathbf{v}^i is a per-particle learnable feature vector of appearance characteristics. $\mathbf{p}^i(\cdot)$ is implemented using a small neural network,

$$\mathbf{p}^i(t) := \text{NN}_{\text{motion}}(t, \mathbf{s}^i) + \mathbf{s}^i \quad (8)$$

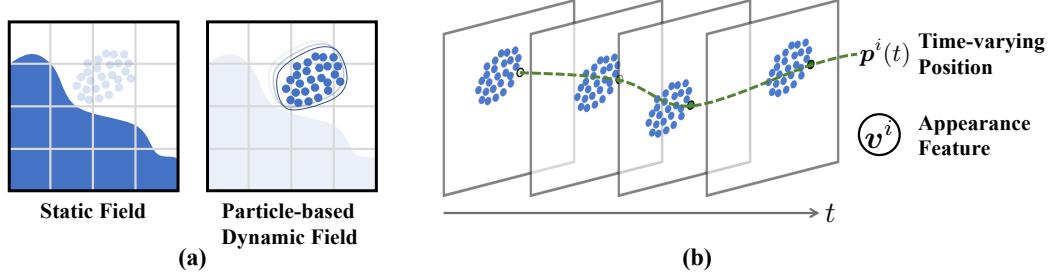


Figure 3: Overview of the proposed hybrid representation. (a) Superposition of dynamic and static field modelled by particle-based and Eulerian (voxel-grid-based) representations. (b) Two main attributes represented by particles. The particle trajectory $p^i(t)$ is time-varying, while the appearance feature v^i is time invariant.

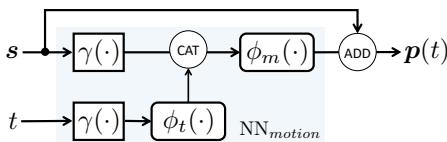


Figure 4: Computational structure for time-varying position of a particle. γ is the positional encoding function (Mildenhall et al. 2020), ϕ_m and ϕ_t are 3 and 2-layer MLP, ‘CAT’ is concatenate operation and ‘ADD’ is element-wise addition.

where $\text{NN}_{\text{motion}}$ is the neural network that computes the particle motion as the offset from the starting point s^i . The computation is illustrated in Fig. 4. The technical details are provided in the appendix.

Therefore, the learning model of particles consists of i) per-particle visual feature v^i , ii) per-particle starting point s^i and iii) the network parameters of $\text{NN}_{\text{motion}}$ that are shared by all particles.

Initialisation. We initially place a set of particles by randomly distributing them within the pre-defined bounding box encompassing the scene. For each placed particle, we initialise s^i as the spatial coordinate. The setup is illustrated in the first pane of Fig. 1 (b).

Particle Removal and Re-sampling. To ensure that particles represent only the dynamic components of a scene, we introduce a run-time strategy that automatically removes and re-samples particles during training. Specifically, we

i) remove particles that are located in the known free space or rarely move throughout the whole timeline. Free spaces are identified using the occupancy mask technique described in (Sun, Sun, and Chen 2022). The immobility of a particle i is identified by the length of its trajectory, i.e., $\int_0^1 \|\frac{\partial p^i(r)}{\partial r}\|_2 dr$ not exceeding a threshold. The removal step is performed every few training steps from the beginning. The second pane of Fig. 1 (b) shows the result of removal.

ii) re-sample particles in nearby areas of the remaining particles to replace the removed particles. The re-sampling step uses the same interval as removal. The third pane of Fig. 1 (b) displays the first re-sampling, and the fourth pane is the final result of alternate removal and re-sampling.

Specific configurations are detailed in the appendix.

3.3 Efficient Computation

Dynamic Feature Field. The computation of (7) requires nearest neighbour search, which is expensive when the number of particles N_p is large. We propose a computational scheme referring to the grid structure, which is inspired by the strategy of (Sulsky, Chen, and Schreyer 1993).

Specifically, we employ a grid \mathbf{G}^D of $N_x \times N_y \times N_z$ nodes and each node is associated with a C -dimension feature. The node feature is of the same format as particle visual feature v . Then the computation of $V_t^d(x)$ consists of two steps of i) propagating appearance information from particles to grid nodes and ii) passing the information to the query location.

At time t , we update features at n -th node of \mathbf{G}^D by:

$$\mathbf{G}^D(n) = \sum_{i \in \mathcal{N}(n)} w_{i \rightarrow n} v^i \quad (9)$$

where $\mathcal{N}(n)$ represents the particles in one of the cells adjacent to n . The *distribution weight* $w_{i \rightarrow n}$ is the tri-linear interpolation weight for particle i in the grid cell containing n . Having appearance features propagated to \mathbf{G}^D , $V_t^d(x)$ is computed via standard tri-linear interpolation

$$\text{Interp}(x, \mathbf{G}^D) : (\mathbb{R}^3, \mathbb{R}^{C \times N_x \times N_y \times N_z}) \mapsto \mathbb{R}^C \quad (10)$$

We query this field by directly accessing neighbour grid nodes, instead of searching neighbours over all particles like the previous method (Abou-Chakra, Dayoub, and Sünderhauf 2022). This approach reduces the computational complexity from $O(N_p)$ to $O(1)$.

We also note that the computational grid for particles is consistent with the static field that we have employed (Sun, Sun, and Chen 2022). The dynamic field is straightforwardly integrated with the static field. Hence we complete the computational model of (5).

3.4 Optimisation

Following TiNeuVox (Fang et al. 2022), our training process involves selecting camera rays randomly, querying the radiance field, and applying standard volumetric rendering (Mildenhall et al. 2020) to compute ray colours and losses. Our framework consists of three loss functions: regular photometric loss L_{photo} of NeRF (Mildenhall et al. 2020), background entropy loss L_{bg} and per-point rgb loss L_{ptrgb} following DVGO (Sun, Sun, and Chen 2022).

4 Experiments

After testing the validity of the proposed DAP-NeRF in novel view synthesis tasks, we demonstrate the capacity and advantage of the particle-based representation in motion modelling, as well as related applications. This section also explores ablation studies on particle quantity effects.

4.1 Experimental Settings

Dataset. The experiments are conducted on eight *360° Synthetic Scenes* (Pumarola et al. 2021), four real scenes (Park et al. 2021b) and our specifically constructed testing scenarios. For a fair comparison, we follow the settings of (Fang et al. 2022; Gan et al. 2022; Park et al. 2021b) where images are trained and rendered at 400×400 and 960×540 pixels for synthetic and real datasets, respectively.

Additionally, to comprehensively evaluate models in motion modelling, we’ve designed three test scenes and built a dataset of 300 images per scene. The dataset also includes data for ground-truth material motion at specific locations and times. We will provide further details in Sec. 4.3.

Model and optimisation configuration. We implement our framework using PyTorch (Paszke et al. 2019). Following (Fang et al. 2022), the resolution of the voxel grids used both in static field $V^s(\cdot)$ and dynamic field $V^d(\cdot)$ are set to 160^3 . The number of particles is 200k. The dimensions C of particle and voxel features are both 12. The network width is 256. The frequency number of positional encoding is set to 10 for (x, y, z) , 4 for view direction d , 8 for time t .

We use Adam (Kingma and Ba 2015) as the optimiser with a minibatch of 4,096 rays. The learning rates are set to 0.005 for particle feature, 0.0005 for motion predictor network ϕ_m . Other learning rates for static field $V^s(\cdot)$, radiance network $MLP_\sigma(\cdot)$ and $MLP_c(\cdot)$ are following (Fang et al. 2022). The total number of training iterations is 60k, and we perform particle removal and resampling every 2k iterations.

4.2 Novel View Synthesis

We conduct experiments on the view synthesis task using the *360° Synthetic Scenes* dataset and compare the results of DAP-NeRF with the state-of-the-art methods. Tab. 1 presents the results, evaluated based on peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) (Wang et al. 2004). Results demonstrate that DAP-NeRF matches or surpasses the performance of state-of-the-art models, despite the fact that the particles are designated for motion modelling rather than appearance accuracy. In Fig. 5 (a), one can visually inspect the reconstructed 3D scenes at different moments. Fig. 5 (b) and (c) show the learned particles and their trajectories, which captures the physical essences of the motions. Fig. 6 qualitatively shows that our model can accurately capture the geometry at each moment.

For the real dynamic scenes, our method also achieves competitive results (see Tab. 2). However, the results are not as dramatic as in *360° Synthetic Scenes*. We deduce that the forward-facing setting could result in insufficient geometric constraints, which might affect the performance of the explicit motion modelling. We plan to explore regularization techniques for improved performance in future work.

4.3 Evaluation of Motion Modelling

To further evaluate how well our models capture the physically relevant dynamics of a scene, we developed a dataset with explicit object motion information. As shown in Fig. 7, each scene contains a moving object of interest (a ball), demonstrating varying levels of motion complexity.

Since the scene and object trajectories are constructed in a 3D design software (Blender-Community 2018), the ground-truth motions are known. Specifically, for any 3D coordinate (x, y, z) at time t , we can determine if it’s occupied and, if so, consider the occupying object’s velocity as the velocity at (x, y, z, t) . With the ground-truth velocity, we design a metric to quantitatively evaluate the motion modelling. Formally, *Motion Field Error* (MFE) is defined as the average Euclidean-norm difference between velocities:

$$\begin{aligned} MFE(\vec{F}_1, \vec{F}_2) &:= \frac{1}{V} \int_V \|\vec{F}_1(\mathbf{x}) - \vec{F}_2(\mathbf{x})\|_2 dV \\ &\approx \frac{1}{N} \sum_n^N \|\vec{F}_1(\mathbf{x}_n) - \vec{F}_2(\mathbf{x}_n)\|_2 \end{aligned} \quad (11)$$

where \vec{F}_1, \vec{F}_2 are the velocity fields to be compared, V is the volume of the region of interest. The integration (11) is approximated by traversing the N voxels, where \mathbf{x}_n represents the center of a voxel. i) For existing deformable dynamic NeRF models, e.g. TiNeuVox (Fang et al. 2022), the velocity at a voxel is approximately implied the deformation field

$$\vec{F}_{df}(\mathbf{x}) \approx \frac{df(\mathbf{x}, t) - df(\mathbf{x}, t + \delta t)}{\delta t} \quad (12)$$

where df stands for the deformation field component of a deformable dynamic NeRF. Given a time t , df specifies a 3D offset from a 3D location to a location in the canonical 3D model $df : \mathbb{R}^3 \times \mathbb{R}^+ \mapsto \mathbb{R}^3$. Note that we need to zero-out the velocities in empty areas where deformable NeRFs are likely to produce incorrect values, as discussed in Sec. 4.4. ii) In DAP-NeRF, the velocity field can be computed directly using the particle motions. Specifically, we computing the average velocity of particles that affect the voxel region at time t . And the velocity of a particle is $v_p = \frac{p_{t+\delta t} - p_t}{\delta t}$.

We compare the our dynamic model with TiNeuVox (Fang et al. 2022), a representative method for deformation fields. We calculate the mean MFE metric between the motion fields estimated by each model and the corresponding ground truth for sampled time steps $t = [0.1, 0.3, 0.5, 0.7, 0.9]$. As demonstrated in Tab. 3, particle-based dynamic model in DAP-NeRF effectively captures motion with lower error in dynamic scenes.

4.4 Decoupling of Scene Components

To illustrate how different models manage dynamics, we introduce artificial grid lines parallel to the axes in the canonical (first) frame, as shown in Fig. 8 (a). Note that the grid lines are solely for the visualisation of modelled motions; the models are NOT aware of the added lines during training.

The rendering results are shown in Fig. 8 (b), which demonstrate that the deforming-based method causes continuous deformation across all space, including empty ar-

Method	PSNR ↑								
	Hell Warrior	Mutant	Hook	Bouncing Balls	Lego	T-Rex	Stand Up	Jumping Jacks	Mean
T-NeRF (Pumarola et al. 2021)	23.19	30.56	27.21	37.81	23.82	30.19	31.24	32.01	29.50
D-NeRF (Pumarola et al. 2021)	25.02	31.29	29.25	38.93	21.64	31.75	32.79	32.80	30.43
NDVG (Guo et al. 2022)	25.53	35.53	29.80	34.58	25.23	30.15	34.05	29.45	30.54
TiNeuVox (Fang et al. 2022)	28.17	33.61	31.45	40.73	25.02	32.70	35.43	34.23	32.67
V4D (Gan et al. 2022)	27.03	36.27	31.04	42.67	25.62	34.53	37.20	35.36	33.72
K-Planes (Fridovich-Keil et al. 2023)	24.81	32.59	28.13	40.33	25.27	30.75	33.17	31.64	30.84
DAP-NeRF (ours)	29.51	35.75	32.69	41.29	25.43	34.07	37.86	35.90	34.06

Method	SSIM ↑								
	Hell Warrior	Mutant	Hook	Bouncing Balls	Lego	T-Rex	Stand Up	Jumping Jacks	Mean
T-NeRF (Pumarola et al. 2021)	0.93	0.96	0.94	0.98	0.90	0.96	0.97	0.97	0.951
D-NeRF (Pumarola et al. 2021)	0.95	0.97	0.96	0.98	0.83	0.97	0.98	0.98	0.953
NDVG (Guo et al. 2022)	0.95	0.99	0.97	0.97	0.93	0.97	0.98	0.96	0.965
TiNeuVox (Fang et al. 2022)	0.97	0.98	0.97	0.99	0.92	0.98	0.99	0.98	0.973
V4D (Gan et al. 2022)	0.96	0.99	0.97	0.99	0.95	0.99	0.99	0.99	0.979
K-Planes (Fridovich-Keil et al. 2023)	0.95	0.97	0.95	0.99	0.94	0.97	0.98	0.97	0.965
DAP-NeRF (ours)	0.97	0.99	0.98	0.99	0.94	0.98	0.99	0.99	0.979

Table 1: Per-scene quantitative comparisons on synthetic dynamic scenes.

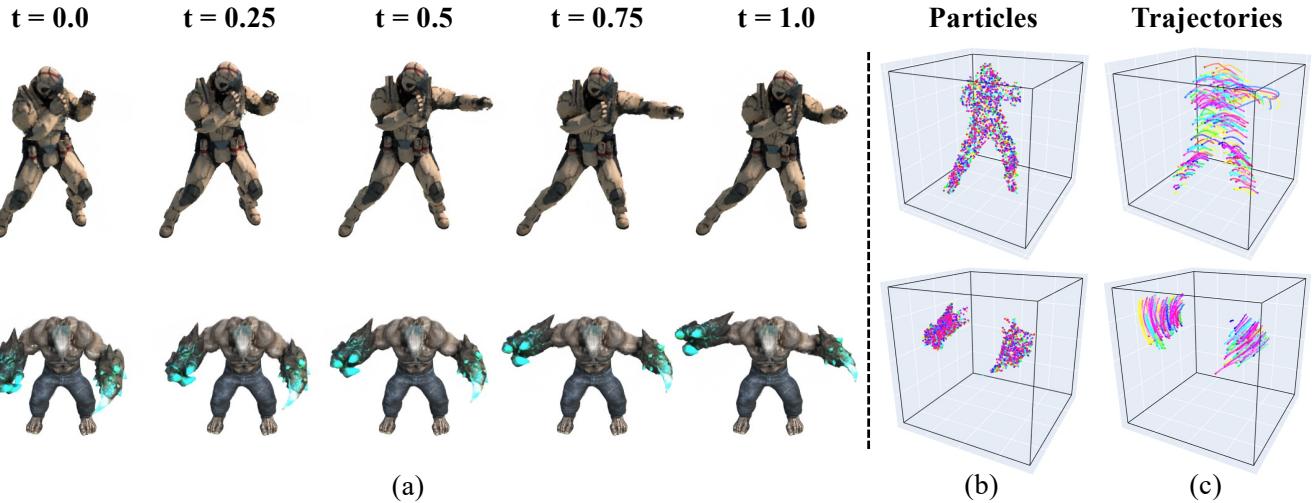


Figure 5: Synthesised New Views of Dynamic Scenes and Particle-based Dynamic Models. (a) displays the dynamic scenes at different moments rendered using the learned DAP-NeRF models. (b) shows the particles at time $t = 0$. To enhance visual clarity, we only render 2k randomly sampled particles. (c) shows the corresponding learned trajectory of particles, where only 200 randomly sampled particles are rendered.

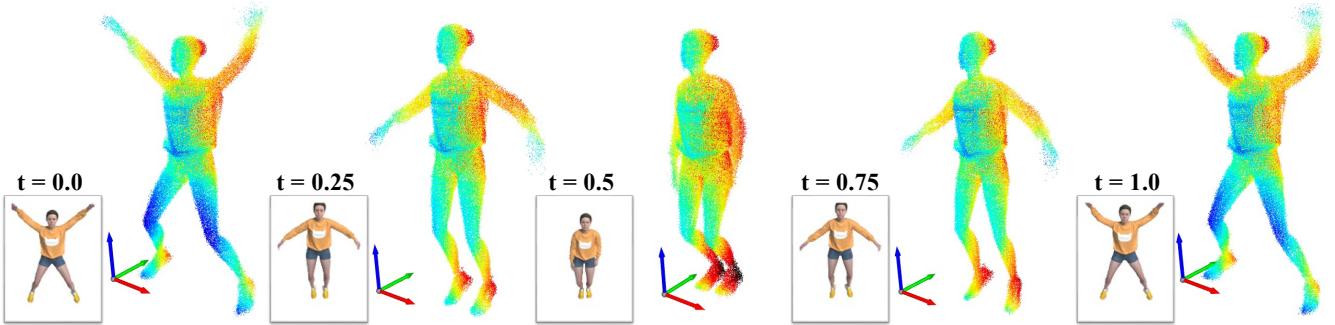


Figure 6: Particle positions learned by DAP-NeRF at different frames. The color of each point is determined by its y-axis value.

Method	PSNR \uparrow	SSIM \uparrow
HyperNeRF (Park et al. 2021b)	22.4	0.814
TiNeuVox (Fang et al. 2022)	24.3	0.837
V4D (Gan et al. 2022)	24.8	0.832
DAP-NeRF (ours)	23.7	0.827

Table 2: Quantitative comparisons on real dynamic scenes. We present the average metric values over the four scenes.

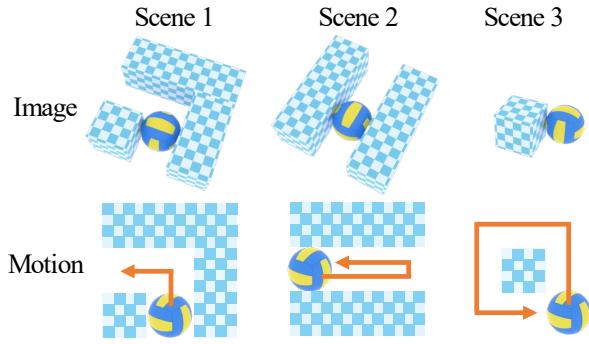


Figure 7: The dataset for evaluating motion modelling.

Method	Scene 1	Scene 2	Scene 3
TiNeuVox	0.0197	0.0363	0.0376
DAP-NeRF (ours)	0.0029	0.0104	0.0136

Table 3: Motion Field Error (MFE) for evaluating motion modelling. The definition of MFE can be found in (11).

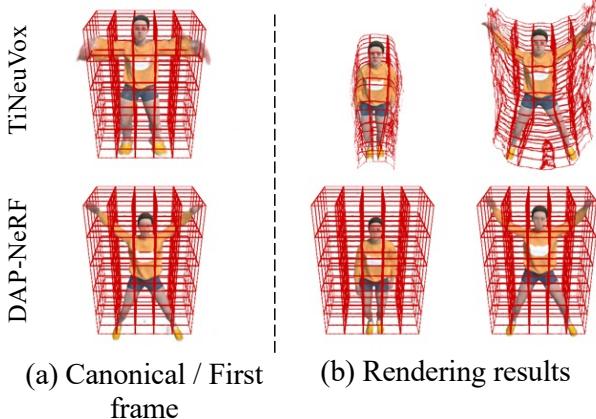


Figure 8: Comparison in dealing with dynamics. We add artificial grids (in red) aligned with the coord-axes at the canonical/first frame for TiNeuVox and our method.



Figure 9: Decomposition of scene components. The ‘full’ image is rendered using the final superpositional feature field. The other two images are rendered using only static and dynamic field respectively.

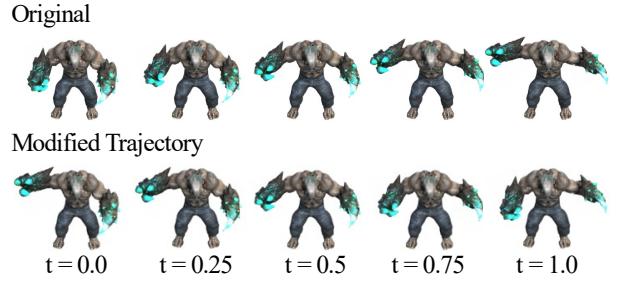


Figure 10: Application on editing dynamic scene. We reverse the original trajectories of particles that are responsible for the shape’s right hand.

eas (e.g., air). In comparison, our method decouples the actual moving objects from the static components and only models the motion for objects. Fig. 9 shows that a trained DAP-NeRF enables a direct and high-quality decomposition. Fig. 10 illustrates how this feature can be helpful in practical applications, such as scene editing.

4.5 Ablation Studies

Particle Number	PSNR \uparrow	SSIM \uparrow
200k	34.06	0.979
100k	33.77	0.979
50k	33.22	0.974
10k	32.52	0.970

Table 4: Ablation study of particle number.

We examine the impact of particle quantity through experiments, employing synthetic scenes from (Pumarola et al. 2021) for evaluation. Averaged metrics across all scenes are displayed in Tab. 4, which illustrate that the model maintains robust performance even with reduced particle count.

5 Conclusion

We have presented the Dynamic Appearance Particle Neural Radiance Field (DAP-NeRF), a novel framework that introduces Lagrangian particles to construct a superpositional radiance field. The proposed appearance particles can not only carry local light radiance information but also capture object motions in an explicitly interpretable and physically meaningful manner. DAP-NeRF is effective and efficient, requir-

ing only monocular video photometric supervision. We have demonstrated that DAP-NeRF performs well in conventional novel view synthesis and excels in motion modeling tasks. One potential future work is the applications on scenes containing fluids.

References

- Abou-Chakra, J.; Dayoub, F.; and Sünderhauf, N. 2022. ParticleNeRF: A Particle-Based Encoding for Online Neural Radiance Fields in Dynamic Scenes. *CoRR*, abs/2211.04041.
- Blender-Community. 2018. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam.
- Cao, A.; and Johnson, J. 2023. HexPlane: A Fast Representation for Dynamic Scenes. *CVPR*.
- Chen, H.; Tretschk, E.; Stuyck, T.; Kadlecak, P.; Kavan, L.; Vouga, E.; and Lassner, C. 2022. Virtual Elastic Objects. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, 15806–15816. IEEE.
- Fang, J.; Yi, T.; Wang, X.; Xie, L.; Zhang, X.; Liu, W.; Nießner, M.; and Tian, Q. 2022. Fast Dynamic Radiance Fields with Time-Aware Neural Voxels. In Jung, S. K.; Lee, J.; and Bargteil, A. W., eds., *SIGGRAPH Asia 2022 Conference Papers, SA 2022, Daegu, Republic of Korea, December 6-9, 2022*, 11:1–11:9. ACM.
- Fridovich-Keil, S.; Meanti, G.; Warburg, F.; Recht, B.; and Kanazawa, A. 2023. K-Planes: Explicit Radiance Fields in Space, Time, and Appearance. *CoRR*, abs/2301.10241.
- Fridovich-Keil, S.; Meanti, G.; Warburg, F. R.; Recht, B.; and Kanazawa, A. 2023. K-Planes: Explicit Radiance Fields in Space, Time, and Appearance. In *CVPR*.
- Gan, W.; Xu, H.; Huang, Y.; Chen, S.; and Yokoya, N. 2022. V4D: Voxel for 4D Novel View Synthesis. *CoRR*, abs/2205.14332.
- Gao, C.; Saraf, A.; Kopf, J.; and Huang, J. 2021. Dynamic View Synthesis from Dynamic Monocular Video. In *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*, 5692–5701. IEEE.
- Gingold, R. A.; and Monaghan, J. J. 1977. Smoothed particle hydrodynamics: theory and application to non-spherical stars. *Monthly notices of the royal astronomical society*, 181(3): 375–389.
- Guan, S.; Deng, H.; Wang, Y.; and Yang, X. 2022. NeuroFluid: Fluid Dynamics Grounding with Particle-Driven Neural Radiance Fields. In Chaudhuri, K.; Jegelka, S.; Song, L.; Szepesvári, C.; Niu, G.; and Sabato, S., eds., *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, 7919–7929. PMLR.
- Guo, X.; Chen, G.; Dai, Y.; Ye, X.; Sun, J.; Tan, X.; and Ding, E. 2022. Neural Deformable Voxel Grid for Fast Optimization of Dynamic View Synthesis. *CoRR*, abs/2206.07698.
- Jiang, Z.; Lu, G.; Liang, X.; Zhu, J.; Zhang, W.; Chang, X.; and Xu, H. 2023. 3D-TOGO: Towards Text-Guided Cross-Category 3D Object Generation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(1): 1051–1059.
- Kingma, D. P.; and Ba, J. 2015. Adam: A Method for Stochastic Optimization. In Bengio, Y.; and LeCun, Y., eds., *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.
- Li, H.; Gu, X.; Yuan, W.; Yang, L.; Dong, Z.; and Tan, P. 2023a. Dense RGB SLAM With Neural Implicit Maps. In *Proceedings of the International Conference on Learning Representations*.
- Li, X.; Qiao, Y.; Chen, P. Y.; Jatavallabhula, K. M.; Lin, M.; Jiang, C.; and Gan, C. 2023b. PAC-NeRF: Physics Augmented Continuum Neural Radiance Fields for Geometry-Agnostic System Identification. In *ICLR*. OpenReview.net.
- Li, Z.; Niklaus, S.; Snavely, N.; and Wang, O. 2021. Neural Scene Flow Fields for Space-Time View Synthesis of Dynamic Scenes. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, 6498–6508. Computer Vision Foundation / IEEE.
- Liu, J.; Cao, Y.; Mao, W.; Zhang, W.; Zhang, D. J.; Keppo, J.; Shan, Y.; Qie, X.; and Shou, M. Z. 2022. DeVRF: Fast Deformable Voxel Radiance Fields for Dynamic Scenes. *CoRR*, abs/2205.15723.
- Lombardi, S.; Simon, T.; Saragih, J. M.; Schwartz, G.; Lehrmann, A. M.; and Sheikh, Y. 2019. Neural volumes: learning dynamic renderable volumes from images. *ACM Trans. Graph.*, 38(4): 65:1–65:14.
- Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In Vedaldi, A.; Bischof, H.; Brox, T.; and Frahm, J., eds., *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part I*, volume 12346 of *Lecture Notes in Computer Science*, 405–421. Springer.
- Müller, T.; Evans, A.; Schied, C.; and Keller, A. 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Trans. Graph.*, 41(4): 102:1–102:15.
- Park, K.; Sinha, U.; Barron, J. T.; Bouaziz, S.; Goldman, D. B.; Seitz, S. M.; and Martin-Brualla, R. 2021a. Nerfies: Deformable Neural Radiance Fields. In *ICCV*, 5845–5854. IEEE.
- Park, K.; Sinha, U.; Hedman, P.; Barron, J. T.; Bouaziz, S.; Goldman, D. B.; Martin-Brualla, R.; and Seitz, S. M. 2021b. HyperNeRF: a higher-dimensional representation for topologically varying neural radiance fields. *ACM Trans. Graph.*, 40(6): 238:1–238:12.
- Park, S.; Son, M.; Jang, S.; Ahn, Y. C.; Kim, J.; and Kang, N. 2023. Temporal Interpolation Is All You Need for Dynamic Neural Radiance Fields. *CoRR*, abs/2302.09311.
- Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Köpf, A.; Yang, E. Z.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang,

- L.; Bai, J.; and Chintala, S. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Wallach, H. M.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E. B.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, 8024–8035.
- Pumarola, A.; Corona, E.; Pons-Moll, G.; and Moreno-Noguer, F. 2021. D-NeRF: Neural Radiance Fields for Dynamic Scenes. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, 10318–10327. Computer Vision Foundation / IEEE.
- Qiao, Y.-L.; Gao, A.; and Lin, M. C. 2022. NeuPhysics: Editable Neural Geometry and Physics from Monocular Videos. In *Conference on Neural Information Processing Systems (NeurIPS)*.
- Song, L.; Chen, A.; Li, Z.; Chen, Z.; Chen, L.; Yuan, J.; Xu, Y.; and Geiger, A. 2022. NeRFPlayer: A Streamable Dynamic Scene Representation with Decomposed Neural Radiance Fields. *arXiv preprint arXiv:2210.15947*.
- Sulsky, D.; Chen, Z.; and Schreyer, H. L. 1993. A particle method for history-dependent materials. *Computer Methods in Applied Mechanics and Engineering*, 118: 179–196.
- Sun, C.; Sun, M.; and Chen, H. 2022. Direct Voxel Grid Optimization: Super-fast Convergence for Radiance Fields Reconstruction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, 5449–5459. IEEE.
- Tretschk, E.; Tewari, A.; Golyanik, V.; Zollhöfer, M.; Lassner, C.; and Theobalt, C. 2021. Non-Rigid Neural Radiance Fields: Reconstruction and Novel View Synthesis of a Dynamic Scene From Monocular Video. In *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*, 12939–12950. IEEE.
- Uzolas, L.; Eisemann, E.; and Kellnhofer, P. 2023. Template-free Articulated Neural Point Clouds for Reposable View Synthesis. *CoRR*, abs/2305.19065.
- Wang, L.; Zhang, J.; Liu, X.; Zhao, F.; Zhang, Y.; Zhang, Y.; Wu, M.; Yu, J.; and Xu, L. 2022. Fourier PlenOctrees for Dynamic Radiance Field Rendering in Real-time. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, 13514–13524. IEEE.
- Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4): 600–612.
- Xian, W.; Huang, J.; Kopf, J.; and Kim, C. 2021. Space-Time Neural Irradiance Fields for Free-Viewpoint Video. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, 9421–9431. Computer Vision Foundation / IEEE.
- Xu, Q.; Xu, Z.; Philip, J.; Bi, S.; Shu, Z.; Sunkavalli, K.; and Neumann, U. 2022. Point-NeRF: Point-based Neural Radiance Fields. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, 5428–5438. IEEE.