

Virtual Elastic Objects

Hsiao-yu Chen^{1,3†}, Edith Tretschk^{2,3†}, Tuur Stuyck³, Petr Kadlec³, Ladislav Kavan³, Etienne Vouga¹, Christoph Lassner³
¹University of Texas at Austin, ²Max Planck Institute for Informatics, ³Meta Reality Labs Research

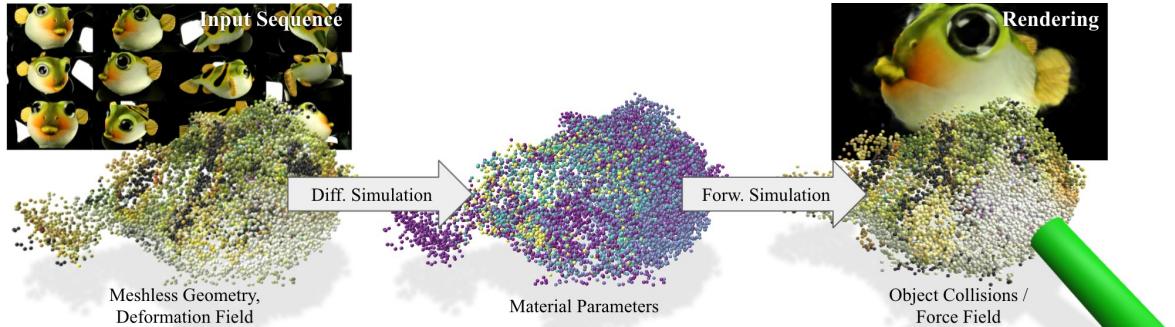


Figure 1. *Method overview*. We use a multi-view capture system to record objects deforming under the influence of external forces. Our method reconstructs a meshless geometry and deformation field from these sequences. Using a differentiable simulator, we optimize the material parameters of the objects to match the observations. These parameters allow us to find novel, plausible object configurations in response to new force fields or collision constraints due to user interactions. Finally, we re-render the deformed state.

Abstract

We present Virtual Elastic Objects (VEOs): virtual objects that not only look like their real-world counterparts but also behave like them, even when subject to novel interactions. Achieving this presents multiple challenges: not only do objects have to be captured including the physical forces acting on them, then faithfully reconstructed and rendered, but also plausible material parameters found and simulated. To create VEOs, we built a multi-view capture system that captures objects under the influence of a compressed air stream. Building on recent advances in model-free, dynamic Neural Radiance Fields, we reconstruct the objects and corresponding deformation fields. We propose to use a differentiable, particle-based simulator to use these deformation fields to find representative material parameters, which enable us to run new simulations. To render simulated objects, we devise a method for integrating the simulation results with Neural Radiance Fields. The resulting method is applicable to a wide range of scenarios: it can handle objects composed of inhomogeneous material, with very different shapes, and it can simulate interactions with other virtual objects. We present our results using a newly collected dataset of 12 objects under a variety of force fields, which will be made available upon publication.

†Work was done while Hsiao-yu (partially) and Edith were interning at Reality Labs Research.

1. Introduction

3D reconstruction is one of the fundamental problems of computer vision and a cornerstone of augmented and virtual reality. A recent achievement in this direction is the discovery of a fairly general formulation for representing radiance fields [5, 29, 33, 35, 39, 49, 50, 53, 57, 66, 68]. Neural radiance fields are remarkably versatile for reconstructing real-world objects with high-fidelity *geometry* and *appearance*. But static appearance is only the first step: it ignores how an object moves and interacts with its environment. 4D reconstruction tackles this problem in part by incorporating the time dimension: with more intricate capture setups and more data, we can reconstruct objects over time—but can only re-play the captured sequences. Today, in the age of mixed reality, a photo-realistically reconstructed object might still destroy immersion if it is not “physically realistic” because the object cannot be interacted with. (For example, if a soft object appears as rigid as the rocks next to it when stepped on.)

By building on advances in computer vision and physics simulation, we begin to tackle the problem of physically-realistic reconstruction and create *Virtual Elastic Objects*: virtual objects that not only look like their real-world counterparts but also behave like them, even when subject to novel interactions. For the first time, this allows for full-loop reconstruction of deforming elastic objects: from capture, to reconstruction, to simulation, to interaction, to re-rendering.

Our core observation is that with the latest advances in 4D reconstruction using neural radiance fields, we can both capture radiance and deformation fields of a moving object over time, and re-render the object given novel deformation fields. The remaining challenge is to capture an object’s physics from observations of its interactions with the environment. With the right representation that jointly encodes an object’s geometry, deformation, and material behavior, compatible with both differentiable physical simulation and the deformation fields provided by 4D reconstruction algorithms, we can use these deformation fields to provide the necessary supervision to learn the material parameters.

But even with this insight, multiple challenges remain to create Virtual Elastic Objects. We list them together with our technical contributions:

1) Capture. To create VEOs, we need to collect data that not only contains visual information but also information about physical forces. We present the new **PLUSH** dataset¹ containing occlusion-free 4D recordings of elastic objects deforming under known controlled force fields. To create this dataset, we built a multi-camera capture rig that incorporates an air compressor with a movable, tracked nozzle. More details can be found in Sec. 3.1.

2) Reconstruction. VEOs do not require any prior knowledge about the geometry of the object to be reconstructed; the reconstruction thus must be template-free and provide full 4D information (*i.e.*, a 3D reconstruction and deformation information over time). We extend Non-rigid Neural Radiance Fields [56] with novel losses, and export point clouds and point correspondences to create the data required to supervise learning material behavior using physical simulation. We provide further details in Sec. 3.2.

3) Simulation. Crucially for creating realistic interactive objects, a physical simulation is required, both to optimize for an unknown object’s physical parameters and to generate deformations of that object in response to novel interactions. We implement a differentiable quasi-static simulator that is particle-based and is compatible with the deformation field data provided by our 4D reconstruction algorithm. We present the differentiable simulator and explain how we use it to obtain physical parameters in Sec. 3.3, and describe simulations of novel interactions in Sec. 3.4.

4) Rendering. Since we convert from a neural representation of the captured object’s geometry to a point cloud reconstructing the object’s physical properties, we require a function that allows rendering the object given new simulated deformations of the point cloud. We introduce a mapping function that enables us to use deformed point clouds instead of continuous deformation fields to alter the ray casting for the Neural Radiance Fields we used for the original reconstruction. Further details on re-rendering can be found in Sec. 3.5.

¹<https://hsiaoyu.github.io/VEO/>

2. Related Work

Our work integrates together multiple areas of computer vision, computer graphics, and simulation.

Recovering Elastic Parameters for 3D Templates. A number of prior works estimate material parameters of a pre-scanned 3D template by tracking the object over time from depth input. Wang *et al.* [59] were among the first to tackle tracking, rest pose estimation, and material parameter estimation from multi-view depth streams. They adopt a gradient-free downhill simplex method for parameter fitting, and can only optimize a limited number of material parameters. Objects built from multiple types of materials cannot be faithfully captured without manual guidance or prior knowledge of a part decomposition. Hahn *et al.* [17] learn an inhomogeneous viscoelastic model from recordings of motion markers covering the object. Recently, Weiss *et al.* [62] infer homogeneous linear material properties by tracking deformations of a given template with a single depth camera. In contrast to these methods, ours jointly reconstructs not just object deformations and physics *without a need for depth input or markers* but also geometry and appearance *without a need for a template*. Our formulation can model inhomogeneous, nonlinear materials without prior knowledge or annotations.

3D/4D Reconstruction. Representing static scenes remains an open problem, with recent mesh-based [16, 60] and neural approaches [7, 34, 46]. Reconstructing non-rigid objects from a video sequence is an equally long-standing computer vision and graphics problem [58, 69]. Shape-from-Template methods deform a provided template using RGB [67] or RGB-D data [71]. DynamicFusion [38] is a model-free, real-time method for reconstructing general scenes from a single RGB-D video. When reliable 2D correspondences are available from optical flow, non-rigid structure-from-motion (NRSfM) can be used to reconstruct the 3D geometry [1, 24], perhaps even using physics-based priors [2]. There are also image-based approaches that do not yield a true 3D scene [4, 65]. Recently, reconstruction using neural representations have become more common. Whereas OccupancyFlow [40] requires 3D supervision, Neural Volumes [30] reconstructs a dynamic scene from multi-view input only, but does not compute temporal correspondences. See a recent survey on neural rendering [55] for more.

Neural Radiance Fields [35], the seminal work of Mildenhall *et al.*, lays the groundwork for several follow-up reconstruction methods that extend it to dynamic scenes [3, 9, 11, 26, 27, 31, 42, 43, 45, 63]. In this work, we assume multi-view RGB video input with known camera parameters and foreground segmentation masks and so extend Non-Rigid Neural Radiance Fields (NR-NeRF) [56].

Data-Driven Physics Simulation. Much recent research has explored the potential of machine learning to enhance or even replace traditional physics-based simulation. Learning

natural laws from data without any priors has been shown possible for a few simple physics systems [48], but the computational cost scales exponentially with the complexity of the system, and remains intractable for real-world problems. For simulating elastic objects specifically, one line of work replaces traditional mesh kinematics with a learned deformation representation to improve performance: Fulton *et al.* [10] use an autoencoder to learn a nonlinear subspace for elastic deformation, and Holden *et al.* [18] train a neural network to predict the deformation of cloth using a neural subspace. Some methods use neural networks to augment coarse traditional simulations with fine details [14, 25].

Another line of work uses data to fit a parameterized material model to observed deformations. This idea has been successfully applied to muscle-actuated biomechanical systems such as human faces [22, 51], learning the rest pose of an object in zero gravity [6], the design of soft robotics [19, 20], and motion planning with frictional contacts [8, 13]. Yang *et al.* [64] learn physical parameters for cloth by analysing the wrinkle patterns in video. While all of these methods learn physical parameters from data, our method is unique in requiring no template or other prior knowledge about object geometry to reconstruct and re-render novel deformations of an object.

Meshless Simulation. Meshless physics-based simulation emerged as a counter-part to traditional mesh-based methods [37] and is ideal for effects such as melting or fracture [37, 44]. These methods have been later extended to support oriented particles and skinning [15, 32, 36]. Another extension of point-based simulations consists in incorporating a background Eulerian grid, which enables more efficient simulation of fluid-like phenomena [21, 52].

3. Method

3.1. Capture

To create a physically accurate representation of an object, we first need to record visual data of its deformation under known physical forces. For recording, we use a static multi-view camera setup consisting of 19 OpenCV AI-Kit Depth (OAK-D) cameras², each containing an RGB and two grey-scale cameras (note that VEO does not use the stereo camera data to infer classical pairwise stereo depth). They represent an affordable, yet surprisingly powerful solution for volumetric capture. In particular, their on-board H265 encoding capability facilitates handling the amount of data produced during recording (5.12GB/s uncompressed). For details on temporal synchronization of the cameras, please see the supplemental material. Since the cameras lack a lens system with zoom capabilities, we keep them close to the object to optimize the pixel coverage and re-configure the system depending on object size. The maxi-

mum capture volume has a size of roughly 30cm³. We put a black sheet around it to create a dark background with the exception of five stage lights that create a uniform lighting environment. A visualization of the camera layout and capture system can be found in the supplementary material.

In addition to the images, we also need to record force fields on the object surface. This raises a problem: if a prop is used to exert force on the capture subject, the prop becomes an occluder that interferes with photometric reconstruction. We solved this problem when capturing our **PLUSH** dataset by actuating the object using transparent fishing line and a compressed air stream; see Sec. 4.1 for further details.

3.2. 4D Reconstruction

Given the captured video of an object deforming under external forces, we need 4D reconstruction to supply a temporally-coherent point cloud that can be used to learn the object material properties. To that end, we use NR-NeRF [56], which extends the static reconstruction method NeRF [35] to the temporal domain. NeRF learns a volumetric scene representation: a coordinate-based Multi-Layer Perceptron (MLP) $\mathbf{v}(\mathbf{x}) = (o, \mathbf{c})$ that regresses geometry (opacity $o(\mathbf{x}) \in \mathbb{R}$) and appearance (RGB color $\mathbf{c}(\mathbf{x}) \in \mathbb{R}^3$) at each point \mathbf{x} in 3D space. At training time, the weights of \mathbf{v} are optimized through 2D supervision by RGB images with known camera parameters: for a given pixel of an input image, the camera parameters allow us to trace the corresponding ray $\mathbf{r}(s)$ through 3D space. We then sample the NeRF at $|S|$ points $\{\mathbf{r}(s) \in \mathbb{R}^3\}_{s \in S}$ along the ray, and use a volumetric rendering equation to accumulate the samples front-to-back via weighted averaging: $\tilde{\mathbf{c}} = \sum_{s \in S} \alpha_s \mathbf{c}(\mathbf{r}(s))$ (*i.e.*, alpha blending with alpha values $\{\alpha_s \in \mathbb{R}\}_s$ derived from the opacities $\{o_s\}_s$). A reconstruction loss encourages the resulting RGB value $\tilde{\mathbf{c}}$ to be similar to the RGB value of the input pixel.

On top of the static geometry and appearance representation \mathbf{v} (*the canonical model*), NR-NeRF models deformations explicitly via a jointly learned ray-bending MLP $\mathbf{b}(\mathbf{x}, \mathbf{l}_t) = \mathbf{d}$ that regresses a 3D offset \mathbf{d} for each point in space at time t . (\mathbf{l}_t is an auto-decoded latent code that conditions \mathbf{b} on the deformation at time t .) When rendering a pixel at time t with NR-NeRF, \mathbf{b} is queried for each sample $\mathbf{r}(s)$ on the ray in order to deform it into the canonical model: $(o, \mathbf{c}) = \mathbf{v}[\mathbf{r}(s) + \mathbf{b}(\mathbf{r}(s), \mathbf{l}_t)]$. Unlike NR-NeRF’s monocular setting, we have a multi-view capture setup. We thus disable the regularization losses of NR-NeRF and only use its reconstruction loss.

Extensions. We improve NR-NeRF in several ways to adapt it to our setting. The input videos contain background, which we do not want to reconstruct. We obtain foreground segmentations for all input images via image matting [28] together with a hard brightness threshold. During training,

²<https://store.opencv.ai/products/oak-d>

we use a background loss $L_{background}$ to discourage geometry along rays of background pixels. When later extracting point clouds, we need opaque samples on the inside of the object as well. However, we find that $L_{background}$ leads the canonical model to prefer empty space even inside the object. We counteract this effect with a density loss $L_{density}$ that raises the opacity of point samples of a foreground ray that are ‘behind’ the surface, while emptying out the space in front of the surface with $L_{foreground}$. During training, we first build a canonical representation by pretraining the canonical model on a few frames and subsequently using it to reconstruct all images. Our capture setup not only provides RGB streams but also grey-scale images. We use these for supervision as well. In practice, we use a custom weighted combination of these techniques for each sequence to get the best reconstruction.

Point Cloud Extraction In order to extract a temporally-consistent point cloud from this reconstruction, we require a forward deformation model, which warps from the canonical model to the deformed state at time t . However, NR-NeRF’s deformation model \mathbf{b} is a backward warping model: it deforms each deformed state into the canonical model. We therefore jointly train a coordinate-based MLP \mathbf{w} to approximate the inverse of \mathbf{b} . After training, we need to convert the reconstruction from its continuous MLP format into an explicit point cloud. To achieve that, we cast rays from all input cameras and extract points from the canonical model that are at or behind the surface and whose opacity exceeds a threshold. These points can then be deformed from the canonical model into the deformed state at time t via \mathbf{w} . See the supplemental material for further details. We thus obtain a 4D reconstruction in the form of a 3D point cloud’s evolving point positions $\{P_t\}_t$, which are in correspondence across time. To keep the computational cost of the subsequent reconstruction steps feasible, we downsample the point cloud to 9-15k points if necessary.

3.3. Learning Material Parameters

Before we can simulate novel interactions with a captured object, we need to infer its physical behavior. Given that we have no prior knowledge of the object, we make several simplifying assumptions about its mechanics, with an eye towards minimizing the complexity of the physical model while also remaining flexible enough to capture heterogeneous objects built from multiple materials.

First, we assume a *spatially varying, isotropic nonlinear Neo-Hookean material model* for the object. Neo-Hookean elasticity well-approximates the behavior of many real-world materials, including rubber and many types of plastic, and is popular in computer graphics applications because its nonlinear stress-strain relationship guarantees that no part of the object can invert to have negative volume, even if the object is subjected to arbitrary large and non-

linear deformations. Finally, Neo-Hookean elasticity admits a simple parameterization: a pair of Lamé parameters $(\mu_i, \lambda_i) \in \mathbb{R}^2$ at each point i of the point cloud P .

Second, we assume that the object deforms *quasistatically* over time: that at each point in time, the internal elastic forces exactly balance gravity and applied external forces. The quasistatic assumption greatly simplifies learning material parameters, and is valid so long as inertial forces in the captured video sequences are negligible (or equivalently, so long as external forces change sufficiently slowly over time that there is no secondary motion, which is true for the air stream and string actuations in our **PLUSH** dataset).

Overview. We first formulate a differentiable, mesh-free *forward* physical simulator that is tailored to work directly with the (potentially noisy) reconstructed point cloud. This forward simulator maps from the point cloud P_0 of the object in its *reference pose* (where it is subject to no external forces besides gravity), an assignment of Lamé parameters to every point, and an assignment of an external force $\mathbf{f}_i \in \mathbb{R}^3$ to each point on the object surface, to the deformed position $\mathbf{y}_i \in \mathbb{R}^3$ of every point in the point cloud after the object equilibrates against the applied forces.

Next, we learn the Lamé parameters that match the object’s observed behavior by minimizing a loss function \mathbf{L} that sums, over all times t , the distance between \mathbf{y}_i and the corresponding target position of the point in the 4D point cloud P_t .

Quasistatic Simulation. To compute the equilibrium positions \mathbf{y}_i of the points in P for given external loads and material parameters, we solve the variational problem

$$\arg \min_{\mathbf{y}} \mathbf{E}(\mathbf{y}), \quad (1)$$

where \mathbf{E} is the total energy of the physical system, capturing both the elastic energy of deformation as well as work done on the system by external forces. In what follows, we derive the expression for \mathbf{E} , and discuss how to solve Eq. 1.

Following Müller *et al.* [37], we adopt a mesh-free, point-based discretization of elasticity to perform forward simulation. For every point \mathbf{x}_i in the reference point cloud P_0 , we define a neighborhood \mathcal{N}_i containing the 6 nearest neighbors of \mathbf{x}_i in P_0 . For any given set of deformed positions \mathbf{y}_j of the points in \mathcal{N}_i , we estimate strain within the neighborhood in the least-squares sense. More specifically, the local material deformation gradient $\mathbf{F}_i \in \mathbb{R}^3$ maps the neighborhood \mathcal{N}_i from the reference to the deformed state:

$$\mathbf{F}_i(\mathbf{x}_i - \mathbf{x}_j) \approx \mathbf{y}_i - \mathbf{y}_j \quad \forall \mathbf{x}_j \in \mathcal{N}_i. \quad (2)$$

For neighborhoods larger than three, Eq. 2 is over-determined, and we hence solve for \mathbf{F}_i in the least-squares

sense, yielding the closed-form solution:

$$\mathbf{F}_i = \mathbf{Y}_i \mathbf{W}_i \mathbf{X}_i^T (\mathbf{X}_i \mathbf{W}_i \mathbf{X}_i^T)^{-1}, \quad (3)$$

where the j -th column of \mathbf{X}_i and \mathbf{Y}_i are $\mathbf{x}_i - \mathbf{x}_j$ and $\mathbf{y}_i - \mathbf{y}_j$, respectively, and \mathbf{W}_i is a diagonal matrix of weights depending on the distance from \mathbf{x}_j to \mathbf{x}_i [37].

The elastic energy of the object can be computed from the classic Neo-Hookean energy density [41]:

$$\Psi_{NH}^i = \frac{\mu_i}{2}(I_c - 3) - \mu_i \log J + \frac{\lambda_i}{2}(J - 1)^2, \quad (4)$$

where I_c is the trace of the right Cauchy-Green tensor $\mathbf{F}_i^T \mathbf{F}_i$, and J is the determinant of \mathbf{F}_i . μ_i and λ_i are the Lamé parameters assigned to point i . The total elastic energy is then:

$$\mathbf{E}_{NH} = \sum_i V_i \Psi_{NH}^i, \quad (5)$$

where $V_i \in \mathbb{R}$ approximates the volume of \mathcal{N}_i .

We also need to include the virtual work done by the external force field to Eq. 1:

$$\mathbf{E}_W = \sum_i \mathbf{f}_i \cdot \mathbf{y}_i, \quad (6)$$

where \mathbf{f}_i is the force applied to point i (the force of the air stream on the boundary). If we measured the tension in the fishing lines, we could also include the forces they exert on the object in Eq. 6. But since a fishing line is effectively inextensible relative to the object we are reconstructing, we instead incorporate the fishing lines as soft constraints on the positions of the points $Q \subset P$ attached to the lines: we assume that at time t , points in Q should match their observed positions in P_t , and formulate an attraction energy:

$$\mathbf{E}_A = \alpha \sum_{q \in Q} \|\mathbf{y}_q - \mathbf{x}_q^*\|^2, \quad (7)$$

where \mathbf{x}_q^* is the position of the point corresponding to \mathbf{y}_q in P_t , and α is a large penalty constant. We found that this soft constraint formulation works better in practice than alternatives such as enforcing $\mathbf{y}_q = \mathbf{x}_q^*$ as a hard constraint; see the supplemental material for more discussion.

The total energy in Eq. 1 is thus $\mathbf{E} = \mathbf{E}_{NH} + \mathbf{E}_W + \mathbf{E}_A$, which we minimize using Newton's method. Since Newton's method can fail when the Hessian \mathbf{H} of \mathbf{E} is not positive-definite, we perform a per-neighborhood eigen-decomposition of \mathbf{H} and replace all eigenvalues that are smaller than a threshold $\epsilon > 0$ with ϵ ; note that this is a well-known technique to improve robustness of physical simulations [54]. We also make use of a line search to ensure stability and handling of position constraints at points where the capture subject touches the ground; see the supplemental material for further implementation details.

Material Reconstruction. Given the 4D point cloud P_t and forces acting on the object $\{\mathbf{f}_i\}_i$, we use our forward simulator to learn the Lamé parameters that best explain the observed deformations. More specifically, at each time t we define the loss:

$$\mathbf{L}_t = \sum_{i \in \partial\Omega} \|\mathbf{y}_{t,i} - \mathbf{x}_{t,i}^*\|^2 \quad (8)$$

where $\mathbf{x}_{t,i}^*$ is the position of point i in P_t , and $\mathbf{y}_{t,i}$ is the output of the forward simulation. We use an ℓ_2 loss to penalize outliers strongly.

We choose a training subsequence T of 20-50 frames from the input where the impact of the air stream roughly covers the surface so that we have some reference for each part of the object, and compute the desired Lamé parameters by minimizing the sum of the loss over all $t \in T$ using the gradient-based Adam optimizer [23]:

$$\mu^*, \lambda^* = \arg \min_{\mu, \lambda} \sum_{t \in T} \mathbf{L}_t. \quad (9)$$

It is not trivial to back-propagate through the Newton solve for $\mathbf{y}_{t,i}$, even if we ignore the line search and assume a fixed number of Newton iterations K . The gradient of \mathbf{y} with respect to the Lamé parameters (μ for instance) can be computed using the chain rule:

$$\frac{\partial \mathbf{L}}{\partial \mu} = \frac{\partial \mathbf{L}}{\partial \mathbf{y}^K} \frac{\partial \mathbf{y}^K}{\partial \mu}, \quad (10)$$

and, for any $1 \leq k \leq K$,

$$\begin{aligned} \frac{\partial \mathbf{y}^k}{\partial \mu} = & \frac{\partial \mathbf{y}_{k-1}}{\partial \mu} - \left(\frac{\partial \mathbf{H}_{k-1}^{-1}}{\partial \mu} + \frac{\partial \mathbf{H}_{k-1}^{-1}}{\partial \mathbf{y}_{k-1}} \frac{\partial \mathbf{y}_{k-1}}{\partial \mu} \right) \nabla \mathbf{E}_{k-1} \\ & - \mathbf{H}_{k-1}^{-1} \left(\frac{\partial \nabla \mathbf{E}_{k-1}}{\partial \mu} + \frac{\partial \nabla \mathbf{E}_{k-1}}{\partial \mathbf{y}_{k-1}} \frac{\partial \mathbf{y}_{k-1}}{\partial \mu} \right). \end{aligned} \quad (11)$$

To avoid an exponentially-large expression tree, we approximate the derivative of the k th Newton iterate \mathbf{y}^k by neglecting the higher-order derivative of the Hessian and of the gradient of the energy with respect to the previous position update:

$$\frac{\partial \mathbf{y}^k}{\partial \mu} \approx \frac{\partial \mathbf{y}_{k-1}}{\partial \mu} - \frac{\partial \mathbf{H}_{k-1}^{-1}}{\partial \mu} \nabla \mathbf{E}_{k-1} - \mathbf{H}_{k-1}^{-1} \frac{\partial \nabla \mathbf{E}_{k-1}}{\partial \mu}$$

Although it is not guaranteed that the higher-order terms are always negligible, this approximation provides a sufficiently high-quality descent direction for all examples we tested. To improve performance and to capture hysteresis in cases where \mathbf{E} has multiple local minima at some times t , we warm-start the Newton optimization at time t using the solution from time $t - 1$. See the supplementary material for more details.

3.4. Novel Interactions

Given a reconstructed VEO, we can use the same physical simulator used for material inference to re-simulate the captured object subject to novel interactions. New force fields can easily be introduced by modifying \mathbf{f}_i in the energy \mathbf{E}_W . Other possible interactions include changing the direction of gravity, adding contact forces to allow multiple objects to mutually interact, or to allow manipulation of the object using mixed-reality tools, etc.

We demonstrate the feasibility of re-simulating novel interactions by implementing a simple penalty energy to handle contact between a VEO and a secondary object, represented implicitly as a signed distance field $d : \mathbb{R}^3 \rightarrow \mathbb{R}$. The penalty energy is given by:

$$\Psi_c(\mathbf{y}) = \begin{cases} \alpha_c d(\mathbf{y})^2 & \text{if } d(\mathbf{y}) < 0 \\ 0 & \text{otherwise,} \end{cases} \quad (12)$$

$$\mathbf{E}_c = \sum_i V_i \Psi_c(\mathbf{y}_i), \quad (13)$$

where α_c is chosen large enough to prevent visually-noticeable penetration of the VEO by the secondary object.

3.5. Rendering

We are able to interact freely with the VEO in a physically plausible manner. Hence, we can close the full loop and realistically render the results of simulated novel interactions using neural radiance fields. While we used \mathbf{b} for deformations during the reconstruction, we are now given a new deformed state induced by a discrete point cloud: a canonical reference point cloud $P_0 = \{\mathbf{x}_s^0\}_s$ and its deformed version $S_d = \{\mathbf{y}_s^d\}_s$. We need to obtain a continuous backward-warping field from that point cloud in order to replace \mathbf{b} , which bends straight rays into the canonical model. To that end, we interpolate the deformation offsets $\mathbf{d}_s^b = \mathbf{x}_s^0 - \mathbf{y}_s^d$ at a 3D sample point \mathbf{p}^d in deformed space using inverse distance weighting (IDW):

$$\mathbf{p}^c = \mathbf{p}^d + \sum_{s \in \mathcal{N}} \frac{w_s}{\sum_{s' \in \mathcal{N}} w_{s'}} \mathbf{d}_s^b, \quad (14)$$

where \mathcal{N} are the $K = 5$ nearest neighbors of \mathbf{p}^d in S_d , and $w_s = w'_s - \min_{s' \in \mathcal{N}} w'_{s'}$, with $w'_s = \|\mathbf{p}^d - \mathbf{y}_s^d\|^{-1}$. We can then sample the canonical model at \mathbf{p}^c as before: $(o, \mathbf{c}) = \mathbf{v}(\mathbf{p}^c)$. To remove spurious geometry that o might show, we set $o(\mathbf{x}) = 0$ for \mathbf{x} that are further than some threshold from S_d . Thus, we can now bend straight rays into the canonical model and render the interactively deformed state of the object in a realistic fashion.

When needed, we can upsample the point cloud from the simulation to make it denser. Unlike for rendering, we need to consider forward warping for this case. See the supplemental material for details.

Object	average (mm)	95% (mm)	max (mm)
Baby Alien	3.8	14.4	29.3
Fish	1.1	6.6	18.5
Leaf	0.4	1.1	9.8
Mr. Seal	0.4	1.9	171.9
Pillow	1.5	7.8	18.35
Dog	1.7	7.5	28.8
Sponge	0.2	1.8	15.8
Dino Rainbow	4.0	14.6	171.4
Dino Blue	5.5	56.0	105.8
Dino Green	6.2	68.4	132.0
Pony	21.1	164.3	204.9
Serpentine	7.5	43.1	94.7
Average*	2.5	18.0	70.2
Average	4.4	32.3	83.5

Table 1. ℓ_2 distance of simulated point clouds compared with reconstructed point clouds on the test set. We record the average distance per point per frame, the 95th percentile of average point distances of all frames, and the maximum distance of all points. Average* excludes the data from Pony and Serpentine.

4. Results

4.1. Dataset

The **PLUSH** dataset consists of 12 soft items encountered in everyday life (see Fig. 2): a pillow, a sponge, and various plush toys. We chose items that are composed of soft (and in some cases, heterogeneous) material, complex geometry, and rich texture and color to enable successful background subtraction, 4D reconstruction and tracking. We provide purchase links for all objects in the supplementary material to enable other researchers to reproduce our experiments. Our strategy for applying external forces is based on the observation that our chosen objects consist of *bulk volumes* (such as the body of a plush toy) along with *flexible extremities* (ears and fingers of the toy). We move object extremities by using transparent fishing line, and we use a stream of compressed air to exert force on bulk volumes. The nozzle position and stream direction must be tracked during video capture to provide the direction and magnitude of forces acting on the object at every point in time. Of the 19 cameras in our capture rig, we use three to track the nozzle using an attached ArUco marker [12, 47]. Using this system, we generate multi-part video sequences for each capture subject, where we sequentially actuate the fishing lines (when applicable) followed by sweeping the air stream over the object. See the supplemental material for details of our methodology for applying and recording external forces. We record between 32s and 67s of video for each object, at a frame rate of 40FPS.

4.2. Virtual Elastic Objects

For each of the 12 examples, we create a VEO using 20-50 frames from the reconstruction and evaluate on the remaining 500-1500 frames. We sample 100k points from NR-NeRF and down-sample to 10k-30k points for simula-



Figure 2. The **PLUSH** dataset consists of 12 items from everyday life: a pillow, a sponge and several plushies. || indicates that we recorded extremity motion for the object, * indicates that the recording has significant second order motion. We additionally provide the mass and recording duration for each object. **Lower right:** Lamé parameter visualizations for Baby Alien and Pony. Colors tending towards purple show a softer region, colors tending towards green and yellow a harder region. Our method clearly identifies different material properties on the objects, for example the arms and ears for the Baby Alien, and the mane and tail of the Pony.

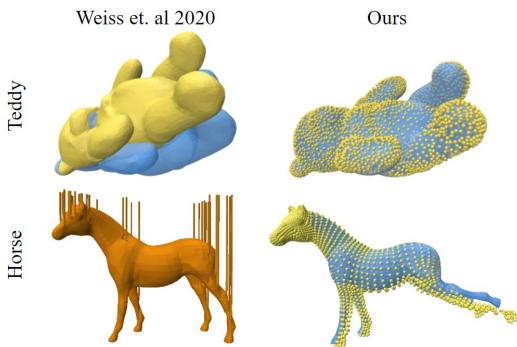


Figure 3. Comparison with Weiss et al. [62] Comparison on two examples. Blue meshes are the ground truth, simulation results are shown in yellow. [62] fails at reconstructing the horse (orange); our heterogeneous model produces overall more reliable results.

tion, depending on the object. We use the ℓ_2 distance between the surface points of the VEO to the reconstructed point cloud from the captured data to evaluate the quality of the reconstructed parameters. For all examples except for the Baby Alien we use the external force field data obtained using the air stream. For the Baby Alien, we specifically use the arm and ear motion to demonstrate the versatility of our method in this scenario. We present the results in Tab. 1.

The error is relatively small for all objects, which shows that our method is applicable to objects with different geometries, and can learn the corresponding material parameters even for heterogeneous objects. Larger errors are observed for objects with a thin and tall component (see the last 4 rows of the table). This error is largely caused by tracking inaccuracies of the nozzle: even slight inaccuracies can cause large errors when, for example, the neck of the dinosaur moves while the recorded air stream direction does not, or barely, touch the object.

Inhomogeneous Material. An important feature of our method is that it can identify different material parameters

for different parts of the object (c.t. Fig. 2, lower right). This is crucial for building a detailed physics model with no prior knowledge of the object. Even more, our method can reliably learn ‘effective’ softness of the material even in places with unreliable tracking, for example thin geometrical structures close to joints. In case of Baby Alien, our method learns that the ears and arms are softer compared to the other body parts; the mane and tail of the Pony are softer, even though these regions are very hard to track. Both reconstructions match the properties of their real counterparts.

We compare our method with the mesh-based work of Weiss et al. [62] (which requires a mesh template). We use the teddy mesh from their paper and simulate it with a heterogeneous material under gravity. We provide 200 depth images and the template to [62], and use the equilibrium point cloud under gravity as input to our method. We then compare both under a novel gravity force; see Fig. 3. Our method is able to estimate the material parameters much better, due to the use of a heterogeneous material model as opposed to the homogeneous model in [62]. We also use a horse model in the same setup. Weiss et al. fail at reconstructing the parts with more detailed geometry (behavior has been confirmed as correct by the authors of [62]) and cannot simulate the object due to the severe artifacts.

Generalization to Novel Poses. The strength of the underlying physics simulator is the ability to generalize to scenarios that are not encountered in the training set. We show different simulated poses of the Baby Alien in Fig. 4, such as pulling the ears in opposite directions, and moving just one single arm. This deformation is particularly challenging for purely data-driven methods since both ears and arms only move synchronously in the training data.

Interaction with Virtual Objects. The physical model of the object enables interactions with all kinds of different virtual items. Fig. 5 shows the one-way coupled interaction of the learned elastic objects with other virtual items.

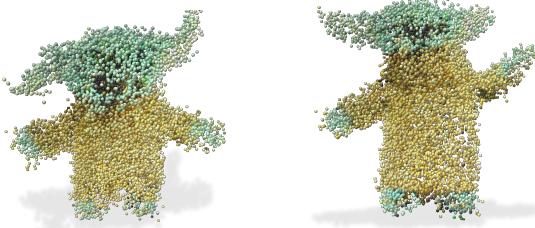


Figure 4. *Simulation of Baby Alien in poses unseen in the dataset.* Using the material model and simulator our method generalizes well to these asymmetric postures for ears and arms; we only observe symmetric forward and backward motions during training.

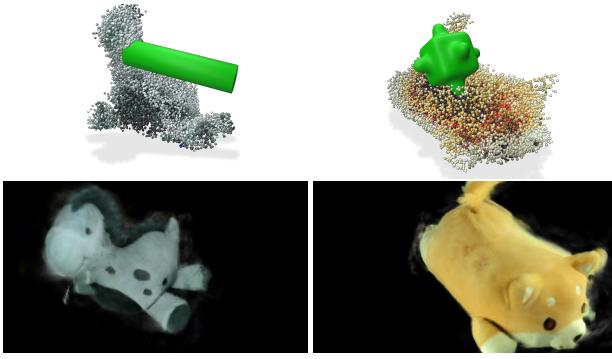


Figure 5. *Rendering of the Dino Blue and Dog VEOs during interactions with secondary objects.* The dinosaur neck bends correctly, and dents are forming on the Dog’s back.

Rendering. Our pipeline ends with re-rendering an object under novel interactions not seen during training. Fig. 5 contains renderings of the Dino Blue and Dog objects, including interactions with two virtual objects. For additional qualitative results, we refer to the supplemental video. Tab. 2 contains quantitative results, where we compare the renderings obtained from the reconstructed point clouds (which are used for supervision when learning the material parameters) and the simulated point clouds.

5. Limitations

Artifacts. Unlike in the simple static setting, we require reliable long-term correspondences on top of the visual reconstruction, which for NeRF only [42, 45, 56] provide. They differ mainly in how they handle the respective monocular problem setting. Our reconstruction method can be seen as a multi-view extension of any of them to our problem setting. Several remaining artifacts arise from the setup. Due to the sparse camera setup (16 cameras for 360 degree coverage), we found NeRF unable to reconstruct viewpoint dependent effects, leading to artifacts around specular regions like eyes. Furthermore, the air compressor leads to quickly oscillating surfaces (*e.g.*, the fins of the fish), which pose a challenge for reconstruction and material parameter estimation, and impacts calibration. These issues impact the

Object	Simulated				Reconstructed											
	Not Masked	PSNR	SSIM	LPIPS	Masked	PSNR	SSIM	LPIPS	Not Masked	PSNR	SSIM	LPIPS	Masked	PSNR	SSIM	LPIPS
Baby Alien	18.40	0.734	0.255	21.17	0.840	0.174	18.75	0.747	0.249	21.92	0.853	0.167				
Fish	19.75	0.692	0.239	22.55	0.808	0.173	20.03	0.701	0.235	22.96	0.818	0.169				
Leaf	25.14	0.901	0.091	27.32	0.935	0.065	25.19	0.901	0.091	27.37	0.935	0.065				
Mr. Seal	20.61	0.697	0.240	24.03	0.801	0.180	20.65	0.698	0.239	24.11	0.802	0.180				
Pillow	21.45	0.743	0.223	23.18	0.806	0.174	21.92	0.760	0.218	23.84	0.823	0.169				
Dog	18.98	0.751	0.206	24.68	0.904	0.104	19.05	0.757	0.203	25.24	0.912	0.100				
Sponge	21.94	0.846	0.130	26.99	0.925	0.079	21.92	0.846	0.130	27.01	0.925	0.070				
Dino Rainbow	18.64	0.754	0.302	23.87	0.839	0.232	20.22	0.778	0.281	26.21	0.859	0.213				
Dino Blue	18.48	0.702	0.244	20.70	0.848	0.160	19.56	0.726	0.227	22.06	0.871	0.143				
Dino Green	18.94	0.779	0.190	21.49	0.863	0.135	20.46	0.794	0.180	23.59	0.879	0.121				
Pony	16.54	0.758	0.245	19.20	0.859	0.163	19.31	0.798	0.200	24.65	0.906	0.108				
Serpentine	18.22	0.798	0.181	21.39	0.903	0.111	19.95	0.813	0.162	23.14	0.916	0.091				
Average*	20.23	0.760	0.212	23.60	0.857	0.145	20.78	0.771	0.205	24.43	0.868	0.140				
Average	19.76	0.763	0.212	23.05	0.861	0.147	20.58	0.777	0.201	24.34	0.875	0.133				

Table 2. *Rendering evaluation.* We report the classic error metrics PSNR and SSIM [61] (−1 to +1), where higher is better for both, and the learned perceptual metric LPIPS [70] (0 is best). We use deformed point clouds to render deformed states of the canonical model, see Sec. 3.5. We use both, the point cloud P_t that the reconstruction (Sec. 3.2) provides directly (‘Reconstructed’) or the point cloud that the simulator provides after learning the material parameters (Sec. 3.3, ‘Simulated’). We report two versions: we either apply the segmentation masks of the input images to the rendered image to remove all artifacts that spill over onto the background (‘Masked’) or we do not (‘Not Masked’). Note, that the values on the reconstructed point cloud are a (soft) upper bound for what the simulator can achieve. The simulated results are close to the reconstructed results, demonstrating that the learned material parameters yield deformation fields that allow to re-render the object as well as the reconstruction can.

extracted point clouds as well as the final renderings (artifacts visible in Fig. 5). The physical simulator turned out to be remarkably robust towards noise and can run with any point cloud with temporal correspondences.

Known Forces. The simulator requires the forces impacting the object during capture to be known. This limits the variety of forces that can be applied and hence the kind of objects that are compatible with the presented method. We expect an extension handling unknown forces an exciting direction for future work. Finding good force priors could be a viable approach in this direction.

6. Conclusion

We introduced a novel, holistic problem setting: estimating physical parameters of a general deformable object from RGB input and known physical forces, and rendering its physically plausible response to novel interactions realistically. We further proposed Virtual Elastic Objects as a solution and demonstrated their ability to synthesize deformed states that greatly differ from observed deformations. Our method leverages a physical simulator that is able to estimate plausible physical parameters from a 4D reconstruction of a captured object. Finally, we showed that these deformed states can be re-rendered with high quality. We hope that the presented results and the accompanying dataset will inspire and enable future work on reconstructing and re-rendering *interactive* objects.

References

- [1] Antonio Agudo, J. M. M. Montiel, Lourdes de Agapito, and Begoña Calvo. Online dense non-rigid 3d shape and camera motion recovery. In *BMVC*, 2014. ²
- [2] Antonio Agudo, Francesc Moreno-Noguer, Begoña Calvo, and José María Martínez Montiel. Sequential non-rigid structure from motion using physical priors. *IEEE transactions on pattern analysis and machine intelligence*, 38(5):979–994, 2015. ²
- [3] Benjamin Attal, Eliot Laidlaw, Aaron Gokaslan, Changil Kim, Christian Richardt, James Tompkin, and Matthew O’Toole. Tōrf: Time-of-flight radiance fields for dynamic scene view synthesis. In *Neural Information Processing Systems (NeurIPS)*, 2021. ²
- [4] Mojtaba Bemana, Karol Myszkowski, Hans-Peter Seidel, and Tobias Ritschel. X-fields: Implicit neural view-, light- and time-image interpolation. *ACM Transactions on Graphics (Proc. SIGGRAPH Asia 2020)*, 39(6), 2020. ²
- [5] Sai Bi, Zexiang Xu, Pratul Srinivasan, Ben Mildenhall, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. Neural reflectance fields for appearance acquisition. *arXiv preprint arXiv:2008.03824*, 2020. ¹
- [6] Xiang Chen, Changxi Zheng, Weiwei Xu, and Kun Zhou. An asymptotic numerical method for inverse elastic shape design. *ACM Transactions on Graphics*, 33(4), 2014. ³
- [7] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *Computer Vision and Pattern Recognition*, pages 5939–5948, 2019. ²
- [8] Tao Du, Kui Wu, Pingchuan Ma, Sébastien Wah, Andrew Spielberg, Daniela Rus, and Wojciech Matusik. Diffpd: Differentiable projective dynamics with contact. *arXiv preprint arXiv:2101.05917*, 2021. ³
- [9] Yilun Du, Yinan Zhang, Hong-Xing Yu, Joshua B. Tenenbaum, and Jiajun Wu. Neural radiance flow for 4d view synthesis and video processing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021. ²
- [10] Lawson Fulton, Vismay Modi, David Duvenaud, David I. W. Levin, and Alec Jacobson. Latent-space dynamics for reduced deformable simulation. *Computer Graphics Forum*, 2019. ³
- [11] Chen Gao, Ayush Saraf, Johannes Kopf, and Jia-Bin Huang. Dynamic view synthesis from dynamic monocular video. *arXiv preprint arXiv:2105.06468*, 2021. ²
- [12] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and R. Medina-Carnicer. Generation of fiducial marker dictionaries using Mixed Integer Linear Programming. *Pattern Recognition*, 51:481–491, mar 2016. ⁶
- [13] Moritz Geilinger, David Hahn, Jonas Zehnder, Moritz Bächer, Bernhard Thomaszewski, and Stelian Coros. Add: Analytically differentiable dynamics for multi-body systems with frictional contact. *ACM Transactions on Graphics (TOG)*, 39(6), 2020. ³
- [14] Zhenglin Geng, Daniel Johnson, and Ronald Fedkiw. Coercing machine learning to output physically accurate results. *Journal of Computational Physics*, 406:109099, Apr 2020. ³
- [15] Benjamin Gilles, Guillaume Bousquet, Francois Faure, and Dinesh K Pai. Frame-based elastic models. *ACM transactions on graphics (TOG)*, 30(2):1–12, 2011. ³
- [16] Thibault Groueix, Matthew Fisher, Vladimir G Kim, Bryan C Russell, and Mathieu Aubry. A papier-mâché approach to learning 3d surface generation. In *Computer Vision and Pattern Recognition*, pages 216–224, 2018. ²
- [17] David Hahn, Pol Banzet, James M. Bern, and Stelian Coros. Real2sim: Visco-elastic parameter estimation from dynamic motion. *ACM Transactions on Graphics (TOG)*, 38(6), 2019. ²
- [18] Daniel Holden, Bang Chi Duong, Sayantan Datta, and Derek Nowrouzezahrai. Subspace neural physics: Fast data-driven interactive simulation. In *Proceedings of the 18th annual ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, New York, NY, USA, 2019. Association for Computing Machinery. ³
- [19] Yuanming Hu, Luke Anderson, Tzu-Mao Li, Qi Sun, Nathan Carr, Jonathan Ragan-Kelley, and Frédéric Durand. Difftaichi: Differentiable programming for physical simulation, 2020. ³
- [20] Yuanming Hu, Jiancheng Liu, Andrew Spielberg, Joshua B Tenenbaum, William T Freeman, Jiajun Wu, Daniela Rus, and Wojciech Matusik. Chainqueen: A real-time differentiable physical simulator for soft robotics. In *2019 International conference on robotics and automation (ICRA)*, pages 6265–6271. IEEE, 2019. ³
- [21] Chenfanfu Jiang, Theodore Gast, and Joseph Teran. Anisotropic elastoplasticity for cloth, knit and hair frictional contact. *ACM Transactions on Graphics (TOG)*, 36(4):1–14, 2017. ³
- [22] Petr Kadlec and Ladislav Kavan. Building accurate physics-based face models from data. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 2(2):1–16, 2019. ³
- [23] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations, ICLR*, 2015. ⁵
- [24] Suryansh Kumar, Anoop Cherian, Yuchao Dai, and Hongdong Li. Scalable dense non-rigid structure-from-motion: A grassmannian perspective. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 254–263, 06 2018. ²
- [25] Zorah Lähner, Daniel Cremers, and Tony Tung. Deepwrinkles: Accurate and realistic clothing modeling. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Computer Vision – ECCV 2018*, pages 698–715. Springer International Publishing, 2018. ³
- [26] Tianye Li, Mira Slavcheva, Michael Zollhoefer, Simon Green, Christoph Lassner, Changil Kim, Tanner Schmidt, Steven Lovegrove, Michael Goesele, and Zhaoyang Lv. Neural 3D Video Synthesis. 2021. ²
- [27] Zhengqi Li, Simon Niklaus, Noah Snavely, and Oliver Wang. Neural scene flow fields for space-time view synthesis of dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6498–6508, 2021. ²

- [28] Shanchuan Lin, Andrey Ryabtsev, Soumyadip Sen-gupta, Brian Curless, Steve Seitz, and Ira Kemelmacher-Shlizerman. Real-Time High-Resolution Background Matting. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8762–8771, dec 2021. 3
- [29] Lingjie Liu, Jiatao Gu, Kyaw Zaw Lin, Tat-Seng Chua, and Christian Theobalt. Neural sparse voxel fields. *NeurIPS*, 2020. 1
- [30] Stephen Lombardi, Tomas Simon, Jason Saragih, Gabriel Schwartz, Andreas Lehrmann, and Yaser Sheikh. Neural volumes: Learning dynamic renderable volumes from images. *ACM Trans. Graph.*, 38(4):65:1–65:14, July 2019. 2
- [31] Stephen Lombardi, Tomas Simon, Gabriel Schwartz, Michael Zollhoefer, Yaser Sheikh, and Jason Saragih. Mixture of volumetric primitives for efficient neural rendering. *ACM Trans. Graph.*, 40(4), July 2021. 2
- [32] Miles Macklin, Matthias Müller, Nuttapong Chentanez, and Tae-Yong Kim. Unified particle physics for real-time applications. *ACM Transactions on Graphics (TOG)*, 33(4):1–12, 2014. 3
- [33] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7210–7219, 2021. 1
- [34] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2
- [35] 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, 2, 3
- [36] Matthias Müller and Nuttapong Chentanez. Solid simulation with oriented particles. In *ACM SIGGRAPH 2011 Papers*, SIGGRAPH ’11, New York, NY, USA, 2011. Association for Computing Machinery. 3
- [37] M. Müller, R. Keiser, A. Nealen, M. Pauly, M. Gross, and M. Alexa. Point based animation of elastic, plastic and melting objects. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA ’04, page 141–151, 2004. 3, 4, 5
- [38] Richard A. Newcombe, Dieter Fox, and Steven M. Seitz. Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015. 2
- [39] Michael Niemeyer and Andreas Geiger. Giraffe: Representing scenes as compositional generative neural feature fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11453–11464, 2021. 1
- [40] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Occupancy flow: 4d reconstruction by learning particle dynamics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6229–6238, 2021. 1
- [41] R.W. Ogden. *Non-linear Elastic Deformations*. Ellis Horwood series in mathematics and its applications. E. Horwood, 1984. 5
- [42] 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. 2, 8
- [43] 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. *arXiv preprint arXiv:2106.13228*, 2021. 2
- [44] Mark Pauly, Richard Keiser, Bart Adams, Philip Dutré, Markus Gross, and Leonidas J. Guibas. Meshless animation of fracturing solids. *ACM Trans. Graph.*, 24(3):957–964, July 2005. 3
- [45] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-NeRF: Neural Radiance Fields for Dynamic Scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. 2, 8
- [46] Edoardo Remelli, Artem Lukoianov, Stephan Richter, Benoit Guillard, Timur Bagautdinov, Pierre Baque, and Pascal Fua. Meshsdf: Differentiable iso-surface extraction. In *Advances in Neural Information Processing Systems*, volume 33, 2020. 2
- [47] Francisco J. Romero-Ramirez, Rafael Muñoz-Salinas, and Rafael Medina-Carnicer. Speeded up detection of squared fiducial markers. *Image and Vision Computing*, 76:38–47, aug 2018. 6
- [48] Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *science*, 324(5923):81–85, 2009. 3
- [49] Katja Schwarz, Yiyi Liao, Michael Niemeyer, and Andreas Geiger. Graf: Generative radiance fields for 3d-aware image synthesis. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. 1
- [50] Pratul P Srinivasan, Boyang Deng, Xiuming Zhang, Matthew Tancik, Ben Mildenhall, and Jonathan T Barron. Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7495–7504, 2021. 1
- [51] Sangeetha Grama Srinivasan, Qisi Wang, Junior Rojas, Gergely Klár, Ladislav Kavan, and Eftychios Sifakis. Learning active quasistatic physics-based models from data. *ACM Transactions on Graphics (TOG)*, 40(4), 2021. 3
- [52] Alexey Stomakhin, Craig Schroeder, Lawrence Chai, Joseph Teran, and Andrew Selle. A material point method for snow simulation. *ACM Transactions on Graphics (TOG)*, 32(4):1–10, 2013. 3
- [53] Edgar Sucar, Shikun Liu, Joseph Ortiz, and Andrew J Davison. imap: Implicit mapping and positioning in real-time. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6229–6238, 2021. 1

- [54] Joseph Teran, Eftychios Sifakis, Geoffrey Irving, and Ronald Fedkiw. Robust quasistatic finite elements and flesh simulation. In *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 181–190, 2005. 5
- [55] A. Tewari, O. Fried, J. Thies, V. Sitzmann, S. Lombardi, K. Sunkavalli, R. Martin-Brualla, T. Simon, J. Saragih, M. Nießner, R. Pandey, S. Fanello, G. Wetzstein, J.-Y. Zhu, C. Theobalt, M. Agrawala, E. Shechtman, D. B Goldman, and M. Zollhöfer. State of the Art on Neural Rendering. *Computer Graphics Forum (EG STAR 2020)*, 2020. 2
- [56] Edgar Treitschke, Ayush Tewari, Vladislav Golyanik, Michael Zollhöfer, Christoph Lassner, and Christian Theobalt. Non-rigid neural radiance fields: Reconstruction and novel view synthesis of a dynamic scene from monocular video. In *IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2021. 2, 3, 8
- [57] Alex Trevithick and Bo Yang. Grf: Learning a general radiance field for 3d scene representation and rendering. *IEEE International Conference on Computer Vision (ICCV)*, 2021. 1
- [58] Tony Tung, Shohei Nobuhara, and Takashi Matsuyama. Complete multi-view reconstruction of dynamic scenes from probabilistic fusion of narrow and wide baseline stereo. In *2009 IEEE 12th International Conference on Computer Vision*, pages 1709–1716, 09 2009. 2
- [59] Bin Wang, Longhua Wu, KangKang Yin, Uri M Ascher, Libin Liu, and Hui Huang. Deformation capture and modeling of soft objects. *ACM Transactions on Graphics*, 34(4), 2015. 2
- [60] Nanyang Wang, Yinda Zhang, Zhuwen Li, Yanwei Fu, Wei Liu, and Yu-Gang Jiang. Pixel2mesh: Generating 3d mesh models from single rgb images. In *European Conference on Computer Vision (ECCV)*, pages 52–67, 2018. 2
- [61] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 8
- [62] S. Weiss, R. Maier, D. Cremers, R. Westermann, and N. Thuerey. Correspondence-free material reconstruction using sparse surface constraints. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, June 2020. 2, 7
- [63] 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, 2021. 2
- [64] Shan Yang, Junbang Liang, and Ming C. Lin. Learning-based cloth material recovery from video. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 4393–4403, 2017. 3
- [65] Jae Shin Yoon, Kihwan Kim, Orazio Gallo, Hyun Soo Park, and Jan Kautz. Novel view synthesis of dynamic scenes with globally coherent depths from a monocular camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5336–5345, 2020. 2
- [66] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelnerf: Neural radiance fields from one or few images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4578–4587, 2021. 1
- [67] Rui Yu, Chris Russell, Neill DF Campbell, and Lourdes Agapito. Direct, dense, and deformable: Template-based non-rigid 3d reconstruction from rgb video. In *Proceedings of the IEEE international conference on computer vision*, pages 918–926, 2015. 2
- [68] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*, 2020. 1
- [69] Li Zhang, Brian Curless, and Steven M. Seitz. Spacetime stereo: shape recovery for dynamic scenes. *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, 2:II–367, 2003. 2
- [70] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 8
- [71] Michael Zollhöfer, Matthias Nießner, Shahram Izadi, Christoph Rehmann, Christopher Zach, Matthew Fisher, Chenglei Wu, Andrew Fitzgibbon, Charles Loop, Christian Theobalt, et al. Real-time non-rigid reconstruction using an rgb-d camera. *ACM Transactions on Graphics (ToG)*, 33(4):1–12, 2014. 2