

---

# Physically Compatible 3D Object Modeling from a Single Image

---

Minghao Guo<sup>1</sup>, Bohan Wang<sup>1\*</sup>, Pingchuan Ma<sup>1</sup>, Tianyuan Zhang<sup>1</sup>, Crystal Elaine Owens<sup>1</sup>,  
Chuang Gan<sup>2,3</sup>, Joshua B. Tenenbaum<sup>1,4,5</sup>, Kaiming He<sup>1</sup>, Wojciech Matusik<sup>1</sup>

<sup>1</sup>MIT CSAIL, <sup>2</sup>UMass Amherst, <sup>3</sup>MIT-IBM Watson AI Lab, <sup>4</sup>MIT BCS,  
<sup>5</sup>Center for Brains, Minds and Machines

<https://gmh14.github.io/phys-comp/>

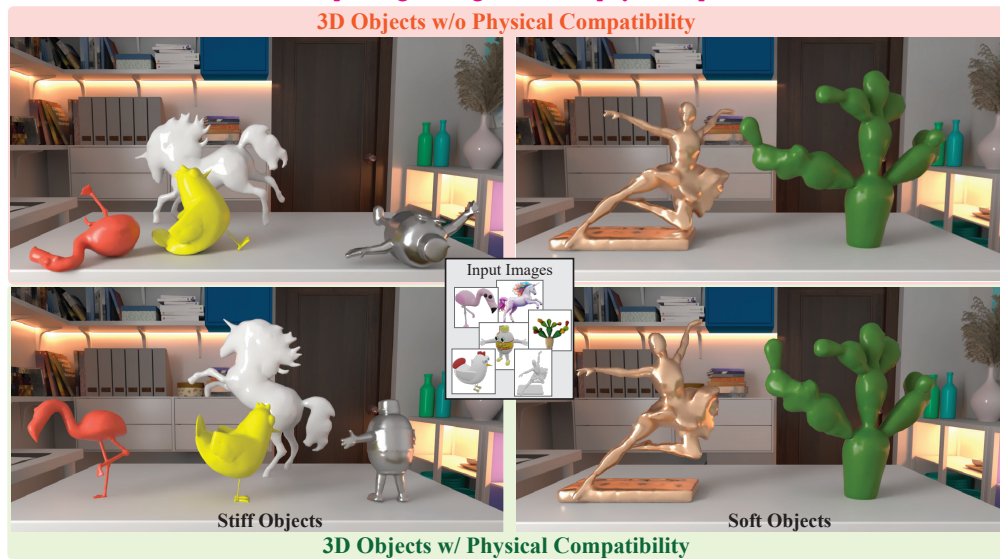


Figure 1: Existing methods for single-view reconstruction often result in objects that, when subjected to real-world physical forces (such as gravity) and user-required mechanical materials, exhibit problematic behaviors such as toppling over (top left) and undesirable deformation (top right), diverging from their intended depiction in the input images. In contrast, our approach produces physical objects that maintain stability (bottom left) and mirror the objects’ static equilibrium state captured in the input images (bottom right).

## Abstract

We present a computational framework that transforms single images into 3D physical objects. The visual geometry of a physical object in an image is determined by three orthogonal attributes: mechanical properties, external forces, and rest-shape geometry. Existing single-view 3D reconstruction methods often overlook this underlying composition, presuming rigidity or neglecting external forces. Consequently, the reconstructed objects fail to withstand real-world physical forces, resulting in instability or undesirable deformation – diverging from their intended designs as depicted in the image. Our optimization framework addresses this by embedding physical compatibility into the reconstruction process. We explicitly decompose the three physical attributes and link them through static equilibrium, which serves as a hard constraint, ensuring that the optimized physical shapes exhibit desired physical behaviors. Evaluations on a dataset collected from Objaverse demonstrate that our framework consistently enhances the physical realism of 3D models over existing methods. The utility of our framework extends to practical applications in dynamic simulations and 3D printing, where adherence to physical compatibility is paramount.

---

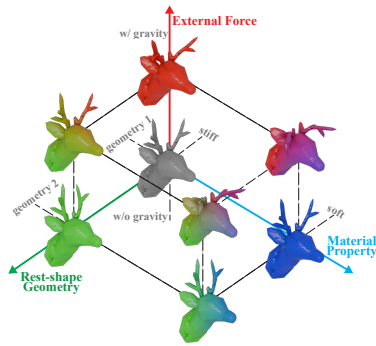
\*Corresponding author.

# 1 Introduction

The field of single-image 3D shape modeling has experienced significant advancements over the past years, largely propelled by advances in single-view reconstruction techniques. These methods, ranging from generating multi-view consistent images for per-scene 3D reconstruction [22, 23, 21, 24, 32, 20], to employing large reconstruction models (LRMs) for feedforward inference [12, 41, 44, 50, 47, 40], have enhanced the geometric quality and visual fidelity of the 3D shapes to unprecedented levels.

However, reconstructing a 3D shape from an image often aims to be beyond a mere visualization. These generated objects find applications in virtual environments such as filming and gaming, as well as in tangible fields like industrial design and engineering. Despite their diverse applications, a common oversight in many current single-view reconstruction methods is the negligence of physical principles. As shown in the top row of Fig. 1, when subjected to real-world physics such as gravity, these 3D objects produced from these techniques exhibit issues such as instability and undesired deformation, diverging from their depiction in the input images. Such inconsistency can significantly undermine the practical utility of the models, as they fail to meet the functional and aesthetic expectations set by the original image.

Fundamentally, an image is more than a visual representation of an object: It captures a physical snapshot of the object in a state of static equilibrium, under the influence of real-world forces. In this context, the geometry seen in an image is determined by three orthogonal attributes: *mechanical properties*, *external forces*, and *rest-shape geometry*. As shown in the inset figure, these attributes collectively model the spectrum of potential static configurations that a physical object might adopt. Reconstructing such an object from an image is essentially an ill-posed problem, since multiple combinations of these attributes can result in identical static geometry. Current methods, however, often overlook this underlying composition; they typically assume objects are rigid or neglect the impact of external forces. The reconstructed objects thus merely replicate the visual geometry without considering the three physical attributes.



To bridge this gap, we explicitly decompose these attributes for reconstructing a physical object from a single image. Our framework holistically takes mechanical properties and external forces as predefined inputs, reflecting typical user specifications in real-world applications like 3D printing and simulations. The output is the rest-shape geometry tailored to these inputs. These attributes are integrated through the principles of static equilibrium physics. This explicit decomposition imposes two stringent physical constraints in object modeling: static equilibrium is enforced as a *hard constraint*, and the physical object must conform to user-specified material properties and external forces. These resulting physical objects are stable, robust under real-world physics, and are high-fidelity replicas inferred from the input images, as shown in the bottom row of Fig. 1.

More specifically, we propose *physical compatibility* optimization, which is a physically constrained optimization with rest-shape geometry as the variable. In this setup, the objective is for the modeled physical object to exhibit desired behaviors, such as matching the geometry depicted in the input image under external forces and maintaining stability under gravity. The constraint is the equation of static equilibrium simulation, ensuring that during optimization, the physical object remains in the equilibrium state, with internal forces generated by deformation from the rest shape balancing the external forces. We parameterize the rest-shape geometry using a plastic deformation field and solve this hard-constrained optimization problem by using implicit differentiation with gradient descent.

For evaluation, we introduce five metrics designed to comprehensively assess the physical compatibility of the modeled 3D objects under simulation. These metrics include image loss between the rendered image of the modeled physical object and the input image, stability under gravity, as well as measures from finite element analysis, such as integrity and structural robustness. Our framework’s versatility is demonstrated by its integration with five distinct single-view reconstruction methods, each employing unique geometry representations. Results on a dataset collected from Objaverse [9], consisting of 100 shapes, show that our framework consistently produces 3D objects with enhanced physical compatibility. Furthermore, we demonstrate the practical utility of our framework through applications in dynamic simulations and 3D printing fabrication.

## 2 Related Work

**Single-view 3D reconstruction.** Recent strides in single-view 3D reconstruction have mainly been fueled by data-driven methods, paralleled by advancements in 3D geometry representation, including NeRF [27], NeuS [43], triplanes [33], Gaussian splatting [16], surface meshes [29], and tet-spheres [11]. These developments have significantly enhanced the geometric quality and visual fidelity of the reconstructed 3D shapes. There are primarily two types of single-view reconstruction methods: 1) Test-time optimization-based methods [31, 23, 39, 45], use multiview diffusion models [21] and iteratively reconstruct 3D scenes using these diffusion priors. 2) Feedforward methods [13, 48, 38, 7, 44, 50] leverage large datasets and learn general 3D priors for shape reconstruction to enable efficient one-step 3D reconstruction from single or sparse views. Unlike the aforementioned methods, our work emphasizes the integration of physical modeling into the reconstruction process. This integration distinguishes our work by ensuring that the resulting 3D shapes are not only visually accurate but also physically plausible under real-world conditions.

**Physics-based 3D modeling.** There has been an increasing interest in incorporating physics into 3D shape modeling. While many approaches utilize video input, which offers a richer temporal context for inferring physical properties such as material parameters [51] and geometry [19], others approach the problem by first reconstructing an object’s geometry from multi-view images and subsequently applying physical simulations [10, 46, 26, 25]. Additionally, several studies have explored extracting physical information from static images [49, 3, 37], using data-driven techniques to estimate properties like shading, mass, and material. In contrast, our work incorporates physical principles, specifically static equilibrium, as hard constraints within the reconstruction process. This integration allows for the optimization of 3D models that adhere to desired physical behaviors depicted by the image.

**Fabrication-aware shape design.** Originating from the computer graphics community, fabrication-aware shape design systems enable designers to specify higher-level objectives – such as structural integrity, deformation, and appearance – with the final shape as the output of the computational system [4]. Related methodologies in this domain, particularly those addressing static equilibrium, include inverse elastic shape design [8] and sag-free initialization [14]. However, these approaches typically require a manually created initial geometry, whereas our work aims to construct the physical object directly from a single input image.

## 3 Approach

Our objective is to create 3D objects from a single image that are physically compatible, ensuring that they align with the input image in the static equilibrium state while also meeting the stability requirements. Governed by universal physical principles, the physical behavior of an object is determined by its mechanical properties, external forces, and rest-shape geometry. Our framework treats the rest-shape geometry as the optimization variable, assuming that the mechanical properties and external forces are predefined as inputs. Fig. 2 illustrates the overall pipeline.

### 3.1 Formulation of Physical Compatibility

In our approach, we treat the entity depicted in the input image as a solid object. We employ Finite Element Method (FEM) for robust solid simulation. The object is represented by a volumetric mesh, denoted as  $\mathcal{M} = (\mathbf{x}, \mathbf{T})$ . Here,  $\mathbf{x} \in \mathbb{R}^{3N}$  represents the 3D positions of the vertices, with  $N$  denoting the total number of vertices.  $\mathbf{T} \in \mathbb{N}^{Z \times K}$  describes the mesh connectivity, where  $Z$  represents the total number of elements and  $K$  indicates the number of vertices per element. The mesh in its *rest-shape geometry*, which is the state without any internal or external forces applied, is represented as  $\mathcal{M}_{\text{rest}} = (\mathbf{X}_{\text{rest}}, \mathbf{T})$ . The input image depicts the *static geometry*, which is the deformed geometry of the object under static equilibrium<sup>1</sup>, denoted as  $\mathcal{M}_{\text{static}} = (\mathbf{x}_{\text{static}}, \mathbf{T})$ . In accordance with Newton’s laws,  $\mathbf{x}_{\text{static}}$  adheres to the following equation:

$$\mathbf{f}_{\text{int}}(\mathbf{x}_{\text{static}}, \mathbf{X}_{\text{rest}}; \Theta) = \mathbf{f}_{\text{ext}}(\mathbf{x}_{\text{static}}), \quad (1)$$

<sup>1</sup>Although our implementation employs *quasi-static equilibrium*, we use the term *static equilibrium* across the paper for consistency.

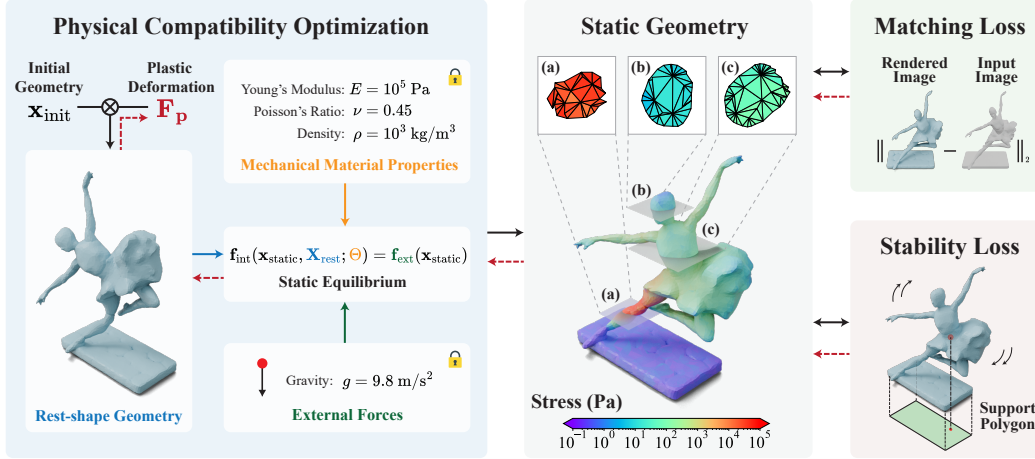


Figure 2: Overall pipeline. Given predefined mechanical properties and external forces, our pipeline optimizes the rest-shape geometry to ensure that the shape, when in a state of static equilibrium, aligns with the target image and meets stability criteria. We visualize the stress distribution of the static geometry using a colored heat map, illustrating the spatially varying deformation of the physical object under static equilibrium.

where  $\mathbf{f}_{\text{int}}(\cdot, \cdot; \Theta) : \mathbb{R}^{3N} \times \mathbb{R}^{3N} \rightarrow \mathbb{R}^{3N}$  denotes the internal forces exerted by deformed objects transitioning from  $\mathbf{X}_{\text{rest}}$  to  $\mathbf{x}_{\text{static}}$ ,  $\mathbf{f}_{\text{ext}}(\cdot) : \mathbb{R}^{3N} \rightarrow \mathbb{R}^{3N}$  embodies the external interaction forces such as gravity, and  $\Theta$  represents the mechanical material properties, such as the stiffness of the object. Eq. 1 reveals that  $\Theta$  (mechanical properties),  $\mathbf{f}_{\text{ext}}$  (external forces), and  $\mathbf{X}_{\text{rest}}$  (the rest-shape geometry) collectively determine the static geometry  $\mathbf{x}_{\text{static}}$ .

Given  $\Theta$  and  $\mathbf{f}_{\text{ext}}(\cdot)$ , the goal of physically compatible modeling is to ensure that the rest-shape geometry  $\mathcal{M}_{\text{rest}}$  conforms to given objectives under static equilibrium. This is formulated as the following optimization problem:

$$\begin{aligned} \min_{\mathbf{X}_{\text{rest}}, \mathbf{x}_{\text{static}}} \quad & \mathcal{J}(\mathbf{X}_{\text{rest}}, \mathbf{x}_{\text{static}}) = \mathcal{L}(\mathbf{x}_{\text{static}}) + \mathcal{L}_{\text{reg}}(\mathbf{X}_{\text{rest}}) \\ \text{s.t.} \quad & \mathbf{f}_{\text{int}}(\mathbf{x}_{\text{static}}, \mathbf{X}_{\text{rest}}; \Theta) = \mathbf{f}_{\text{ext}}(\mathbf{x}_{\text{static}}). \end{aligned} \quad (2)$$

Here,  $\mathcal{J}(\mathbf{X}_{\text{rest}}, \mathbf{x}_{\text{static}})$  is the objective function, consisting of  $\mathcal{L}(\mathbf{x}_{\text{static}})$ , which measures the alignment of the geometry  $\mathbf{x}_{\text{static}}$  with the specified target.  $\mathcal{L}_{\text{reg}}(\mathbf{X}_{\text{rest}})$  regularizes the rest-shape geometry  $\mathbf{X}_{\text{rest}}$ , with more details discussed in Section 3.2.

Within the scope of this work, two tasks for  $\mathcal{L}(\mathbf{x}_{\text{static}})$  are considered: 1)  $\mathbf{x}_{\text{static}}$  replicates the geometry depicted in the input image; and 2)  $\mathbf{x}_{\text{static}}$  maintains stability and inherently remains upright without toppling. In the first scenario, the loss function is  $\mathcal{L}(\mathbf{x}_{\text{static}}) = \|\mathbf{x}_{\text{static}} - \mathbf{X}_{\text{target}}\|_2^2$  which measures the point-wise Euclidean distance between the static shape and the target geometry  $\mathcal{M}_{\text{target}} = (\mathbf{X}_{\text{target}}, \mathbf{T})$ . In the second scenario, the loss function is  $\mathcal{L}(\mathbf{x}_{\text{static}}) = \|\text{proj}_z(\mathcal{C}(\mathbf{x}_{\text{static}})) - \hat{\mathbf{C}}\|$ , where  $\mathcal{C}(\cdot)$  computes the center of mass of  $\mathcal{M}_{\text{static}}$ ,  $\text{proj}_z(\cdot)$  denotes the projection of the center onto the  $z$ -plane in world coordinates, and  $\hat{\mathbf{C}}$  represents the target position for the center of mass to guarantee stability. Minimization of this function ensures the structural stability of  $\mathcal{M}_{\text{static}}$ .

It is crucial to highlight that the variables  $\mathbf{X}_{\text{rest}}$  and  $\mathbf{x}_{\text{static}}$  are tightly coupled through a hard constraint in our problem formulation. This constraint, which ensures that the object remains static equilibrium, is essential to achieving physical compatibility. Enforcing this configuration guarantees that the 3D physical object conforms strictly to external forces such as gravity, thereby ensuring the system adheres to the inherent physical constraints.

### 3.2 Parameterization of Rest-shape Geometry

To solve the optimization problem defined Eq. 2, one might consider a straightforward approach by directly treating  $\mathbf{X}_{\text{rest}}$  as the optimization variable. However, this brings challenges in maintaining the physical validity of the rest-shape geometry, i.e., there shall be no inversions or inside-out elements.

This non-inversion requirement is typically enforced through nonlinear inequality constraints [11, 36], leading to intractable optimization. Drawing inspiration from natural modeling processes [42], we propose a parameterization of  $\mathbf{X}_{\text{rest}}$  by treating it as the result of plastic deformation applied to an initial configuration. A *plastic deformation* can transform objects without the volume preservation constraint [1]. Specifically, we denote the initial configuration of the rest-shape geometry as  $\mathcal{M}_{\text{init}} = (\mathbf{X}_{\text{init}}, \mathbf{T})$ .  $\mathbf{X}_{\text{rest}}$  is implicitly parameterized by the plastic deformation field  $\mathbf{F}_{\mathbf{p}}$  as

$$\mathbf{X}_{\text{rest}} := \phi(\mathbf{F}_{\mathbf{p}}; \mathbf{X}_{\text{init}}), \quad \text{with} \quad \mathbf{f}_{\text{int}}(\mathbf{X}_{\text{rest}}, \mathbf{X}_{\text{init}}; \Theta) = \mathbf{0}. \quad (3)$$

Intuitively, this equation suggests that  $\mathbf{X}_{\text{rest}}$  results from applying plastic strain field  $\mathbf{F}_{\mathbf{p}}$  to  $\mathbf{X}_{\text{init}}$  without any external forces. The plastic strain field  $\mathbf{F}_{\mathbf{p}}$  is the collection of transformations, with each transformation is an  $\mathbb{R}^{3 \times 3}$  matrix applied to each material point. Throughout this paper, we also represent plastic deformation in its vector form as  $\mathbf{F}_{\mathbf{p}} \in \mathbb{R}^{9Z}$ , which corresponds to the flattened vector form of the  $\mathbb{R}^{3 \times 3}$  transformation collection. For a detailed explanation of the computation of  $\mathbf{X}_{\text{rest}}$  from  $\mathbf{F}_{\mathbf{p}}$  and its integration into the static equilibrium, we refer the reader to Appendix B.

There are several benefits using  $\mathbf{F}_{\mathbf{p}}$  for parameterizing rest-shape geometry: It exhibits invariance to translation, which ensures that the spatial positioning of  $\mathbf{X}_{\text{init}}$  does not affect the deformation outcomes. Moreover, the non-inversion requirement can be efficiently satisfied by constraining the singular values of  $\mathbf{F}_{\mathbf{p}}$ , thereby avoiding the need for complicated inequality constraints. Appendix B provides a comprehensive analysis of these advantages.

By substituing Eq. 3, we reformulate the optimization problem Eq. 2 as follows:

$$\begin{aligned} \min_{\mathbf{F}_{\mathbf{p}}, \mathbf{x}_{\text{static}}} \quad & \mathcal{J}(\mathbf{F}_{\mathbf{p}}, \mathbf{x}_{\text{static}}) = \mathcal{L}(\mathbf{x}_{\text{static}}) + \mathcal{L}_{\text{reg}}(\mathbf{F}_{\mathbf{p}}) \\ \text{s.t.} \quad & \mathbf{f}_{\text{int}}(\mathbf{x}_{\text{static}}, \phi(\mathbf{F}_{\mathbf{p}}; \mathbf{X}_{\text{init}}); \Theta) = \mathbf{f}_{\text{ext}}(\mathbf{x}_{\text{static}}). \end{aligned} \quad (4)$$

Here, the optimization variables are  $\mathbf{F}_{\mathbf{p}}$ , where the initial geometry configuration  $\mathbf{X}_{\text{init}}$  is treated as a constant. The regularization term  $\mathcal{L}_{\text{reg}}(\mathbf{F}_{\mathbf{p}})$  is defined as the smoothness of plastic deformation using bi-harmonic energy [5], represented as  $\mathcal{L}_{\text{reg}}(\mathbf{F}_{\mathbf{p}}) = \|\mathbf{L}\mathbf{F}_{\mathbf{p}}\|_2^2$ , where  $\mathbf{L} \in \mathbb{R}^{9Z \times 9Z}$  denotes the graph Laplacian matrix, encapsulating the connectivity of the volumetric mesh elements.

### 3.3 Implicit Differentiation-based Optimization

Solving the optimization problem in Eq. 4 is non-trivial due to its nonlinear objective and the nonlinear hard constraint. A straightforward approach is incorporating the constraint directly into the objective as an additional loss term; however, this method may lead to imperfect satisfaction of the constraint, which undermines the fundamental goal of ensuring physical compatibility.

We resort to implicit differentiation, a technique used in sensitivity analysis [6], to compute the gradient of the objective function  $\mathcal{J}$  with respect to the variable  $\mathbf{F}_{\mathbf{p}}$ . This approach effectively reduces the dimensionality of the optimization variables since we only need to calculate the gradient with respect to  $\mathbf{F}_{\mathbf{p}}$  and also ensures that the gradient direction takes into account the hard constraint. Specifically, the gradient is computed as follows:

$$\frac{\partial \mathcal{J}}{\partial \mathbf{F}_{\mathbf{p}}} = - \left( \frac{\partial \mathcal{L}}{\partial \mathbf{x}_{\text{static}}} \right) \left[ \frac{\partial \mathbf{f}_{\text{net}}}{\partial \mathbf{x}_{\text{static}}} \right]^{-1} \frac{\partial \mathbf{f}_{\text{net}}}{\partial \mathbf{F}_{\mathbf{p}}} + \frac{\partial \mathcal{L}_{\text{reg}}}{\partial \mathbf{F}_{\mathbf{p}}}, \quad (5)$$

where  $\mathbf{f}_{\text{net}} = \mathbf{f}_{\text{int}} - \mathbf{f}_{\text{ext}}$  represents the net forces. A comprehensive derivation of this gradient formula is provided in Appendix C. By utilizing this gradient, the optimization can be solved using standard optimization tools, such as the Adam optimizer [17]. This facilitates the integration of our method into existing single-view reconstruction pipelines.

### 3.4 Implementation Details

Given an input image, we initially utilize off-the-shelf single-view reconstruction models to obtain the 3D object’s target geometry, ensuring alignment with the input image. The output of these reconstruction models varies depending on the geometric representation used. For instance, methods employing tetrahedral representations, such as TetSphere [11], yields volumetric meshes that can be directly used as  $\mathcal{M}_{\text{target}}$ . Conversely, methods that output surface meshes [44] or point clouds [40], which are often non-volumetric and typically non-manifold, require additional processing steps to be suitable for our computational pipeline. We use TetWild [15], a robust tetrahedral meshing algorithm,

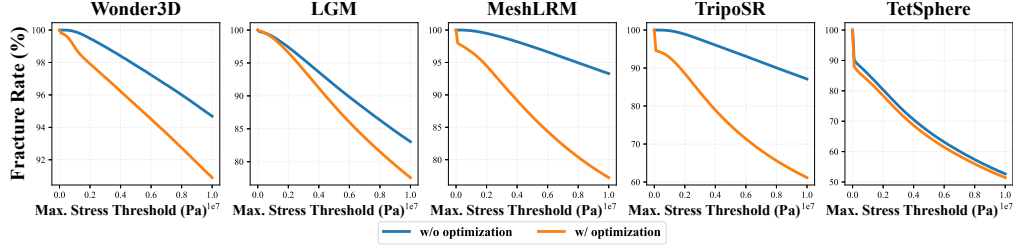


Figure 3: Quantitative results on fracture rate. We plot the relationship between the fracture rate and the maximum stress threshold across five single-image reconstruction methods. The shapes optimized with our framework exhibit a consistently lower fracture rate compared to those shapes obtained without our pipeline. MeshLRM and TripoSR feature prevalent thin structures in their reconstructed shapes, whereas our approach significantly reduces the fracture rate in both cases.

to convert these unstructured outputs into high-quality tetrahedral meshes, resulting in volumetric mesh  $\mathcal{M}_{\text{target}}$ . For initiating the optimization process, we set  $\mathcal{M}_{\text{init}} = \mathcal{M}_{\text{target}}$ , assuming that  $\mathcal{M}_{\text{target}}$  is a reasonably good initial approximation for the optimization. Note that  $\mathcal{M}_{\text{init}}$  is not strictly confined to  $\mathcal{M}_{\text{target}}$ ; any volumetric mesh could potentially serve as the initial approximation, given the flexibility of  $\mathbf{F}_{\mathbf{p}}$  to accommodate spatially varying deformations.

For the material constitutive model, we use isotropic Neo-Hookean material as detailed in [35]. The mechanical properties  $\Theta$ , including Young’s modulus  $E$ , Poisson’s ratio  $\nu$ , and mass density  $\rho$ , are set by users. These values can be specified directly through numerical input or chosen from a collection of pre-established material options, such as plastic or rubber. We consider gravity and fixed attachment forces as options for external forces. Gravity is always included to reflect its omnipresence in the real world. The use of fixed attachment forces depends on the specific needs of the application, for instance, anchoring an object at a designated site. Detailed formulations for both force types are provided in Appendix F.

## 4 Evaluation

In this section, we present evidence that our approach enhances the physical compatibility of 3D objects produced using state-of-the-art single-view reconstruction techniques. We conduct a series of quantitative evaluations using five metrics (Sec. 4.1) to compare the physical compatibility of shapes optimized by our framework against those produced by existing methods without our method (Sec. 4.2). We also provide qualitative comparisons to demonstrate to the effectiveness of our approach (Sec. 4.3). Furthermore, we explore the practical applications of our method by illustrating how it enables the reconstruction of diverse 3D shapes with different material properties from the same single image, and by demonstrating that our optimized shapes are readily adaptable for dynamic simulations and fabrication (Sec. 4.4).

### 4.1 Baselines and Evaluation Protocol

Existing metrics for evaluating single-view reconstruction methods primarily focus on the visual appearance of the objects. Measures such as PSNR and SSIM are used to assess image fidelity, while chamfer distance and volume IoU evaluate geometric quality. However, these metrics do not consider the underlying physics principles that govern the behavior of 3D objects. Consequently, they are insufficient for evaluating the physical compatibility of reconstructed shapes, a crucial aspect for applications requiring accurate physical interactions and structural stability.

**Metrics.** To address this oversight, we draw inspiration from the field of finite element analysis [2] and introduce five novel metrics specifically designed to assess the physical compatibility of 3D models comprehensively. These metrics are tailored to ensure a more thorough evaluation of method performance in real-world scenarios with rich physics:

- **Number of Connected Components (#CC.)** evaluates the structural integrity of the object. Physical objects should not have floating or disconnected structures, ideally consisting of one single connected component.

Table 1: Quantitative results on four metrics evaluating physical compatibility. We apply our pipeline to five single-image reconstruction techniques and assess our metrics on both the initial shapes from these methods (Baseline) and the optimized shapes from the integration of our framework with each baseline (Ours). Our method demonstrates quantitative improvements in mean stress, stability rate, and image fidelity across all benchmarks. Among all methods, TetSphere integrated with our framework achieves superior performance across all evaluation metrics. This can be attributed to the explicit volumetric representation used in TetSphere. The mean and standard deviation are calculated across all examples for each method. A higher deviation in Mean Stress suggests a larger variance in structural thickness and curvature, while a higher deviation in Img. Loss indicates a larger variance in static shape deformation.

Method		Init. Geo.	#CC. ↓	Mean Stress ↓ (kPa)	Standable. ↑ (%)	Img. Loss ↓
<b>Wonder3D</b>	Baseline	NeuS	2.54 ± 2.64	10.68 ± 17.47	6.9	0.073 ± 0.063
	Ours			0.45 ± 0.96	72.4	0.069 ± 0.048
<b>LGM</b>	Baseline	Gaussian splatting	2.67 ± 2.13	1.14 ± 2.03	20.3	0.121 ± 0.091
	Ours			1.01 ± 1.34	<u>85.9</u>	0.116 ± 0.065
<b>MeshLRM</b>	Baseline	surface mesh	1.55 ± 2.13	0.54 ± 1.41	28.6	0.065 ± 0.042
	Ours			0.38 ± 1.05	73.8	0.064 ± 0.042
<b>TripoSR</b>	Baseline	NeRF	1.43 ± 1.12	0.29 ± 1.28	24.2	0.066 ± 0.047
	Ours			<u>0.22 ± 0.94</u>	80.6	<u>0.059 ± 0.039</u>
<b>TetSphere</b>	Baseline	tet-sphere	<b>1.00 ± 0.00</b>	<u>0.22 ± 0.51</u>	32.8	0.061 ± 0.045
	Ours			<b>0.19 ± 0.78</b>	<b>92.4</b>	<b>0.057 ± 0.040</b>

- **Mean Stress** calculates the average von Mises stress [28] across all tetrahedra of all objects. It measures the extent of physical deformation. Under the same external interactions, higher mean stress indicates a greater likelihood of fracture and the existence of unrealistic thin structures.
- **Percentage of Standability (Standable.)** assesses whether the object can maintain stability under gravity, remaining upright without toppling. A standable object is one that effectively supports itself against gravitational forces.
- **Matching loss (Img. Loss)** calculates the  $l_1$  difference between the rendered image of the object after applying gravity and the input target image, quantifying the deviation of the physical object from the desired shape due to physical influences.
- **Fracture Rate** measures the number of tetrahedral elements that exceed a predefined stress threshold, potentially leading to fractures. The resilience of a method against physical stresses is quantified using a degradation curve, with more physically reliable methods exhibiting a smaller area under the curve for the fracture rate.

**Baselines.** We consider five single-view reconstruction baselines in our evaluation, each associated with a distinct geometry representation: Wonder3D [23] with NeuS, LGM [40] with Gaussian splatting, MeshLRM [44] with surface mesh, TripoSR [41] with NeRF triplane, and TetSphere [11] with tetrahedral spheres. For the baseline results, we used the publicly available inference code to reconstruct the 3D objects.<sup>2</sup> To demonstrate the versatility of our method, we integrated our physical compatibility optimization framework with all five baseline models and reported the results to ensure a fair comparison. The implementation details of our framework are provided in Appendix D.

**Evaluation Datasets.** The evaluation dataset was sourced from Objaverse [9]. We initially randomly selected approximately 200 shapes from the categories of plants, animals, and characters – categories that demand greater physical compatibility. Single-view images were rendered using the publicly released code by the authors of Objaverse<sup>3</sup>. Subsequently, these images were used to reconstruct 3D objects using the baseline methods mentioned earlier. We filtered out shapes of extremely poor quality, specifically those with more than 8 connected components. This process resulted in a final set of 100 shapes for detailed evaluation.

Despite these shapes being a part of the training data for most baseline methods, our evaluation focuses on assessing the physical compatibility – a factor overlooked by these methods. The results

<sup>2</sup>For MeshLRM, since the pre-trained model is not publicly available yet, we obtained the reconstructed shapes directly from the authors for use in our study.

<sup>3</sup><https://github.com/allenai/objaverse-rendering>

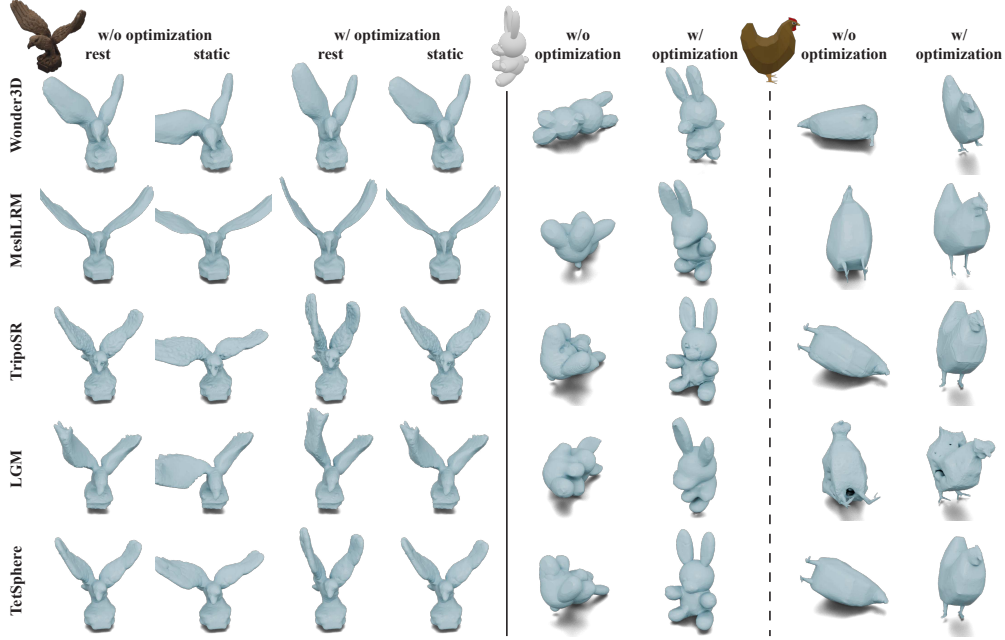


Figure 4: Qualitative results on physical compatibility optimization. Left: Rest shapes optimized using our approach result in static shapes that closely match the input images when subjected to gravity. In contrast, shapes without the optimization fail to replicate the geometry in the input image. Right: our optimization process ensures that the optimized shapes are capable of supporting themselves, whereas the baseline methods fail to achieve this stability.

obtained from this dataset provide valuable insights and observations on the physical compatibility of each method, demonstrating the practical effectiveness of our approach.

## 4.2 Quantitative Results

Table 1 shows the quantitative results for four out of five metrics evaluated for both baselines and those integrated with our physical compatibility optimization. Fig. 3 shows the curve of fracture rate.

Our quantitative analysis yields several observations: 1) The underlying geometry representation significantly impacts the structural integrity of reconstructed shapes, as evidenced by the number of connected components (#CC.). LGM, using a point cloud representation, exhibits the poorest structural integrity, often resulting in floating structures due to its inability to differentiate the interior from the exterior of a 3D object. In contrast, TetSphere, with its volumetric representation, maintains the most integral structure. 2) Both MeshLRM and TripoSR generally produce more physically stable 3D objects, as indicated by Mean Stress and Standability (Standable.) metrics. However, they tend to diverge under gravity, as shown by the Matching Loss metric (Img. Loss), compared to TetSphere. 3) Notably, our method consistently enhances the physical compatibility performance across all baselines. The improvement is particularly significant for Wonder3D and MeshLRM. Wonder3D typically generates multi-view images before reconstructing the 3D shape, which can lead to thin structures due to inconsistencies across the views. Similarly, MeshLRM’s reliance on surface mesh could often result in thin structures. Our method strengthens the physical robustness for both cases. 4) Our method also enhances the structure robustness to fracture, as demonstrated in Fig. 3. It notably improves the performance of both MeshLRM and TripoSR in reducing fracture rates.

## 4.3 Qualitative Results

Fig. 4 and more qualitative results in Appendix 4 illustrate the effectiveness of our physical compatibility optimization. Without optimization, the static shapes behave undesirably under general physical principles: they either sag excessively under gravity, diverging from the geometry depicted in the input image, or fail to remain upright, toppling over. Our optimization method incorporates



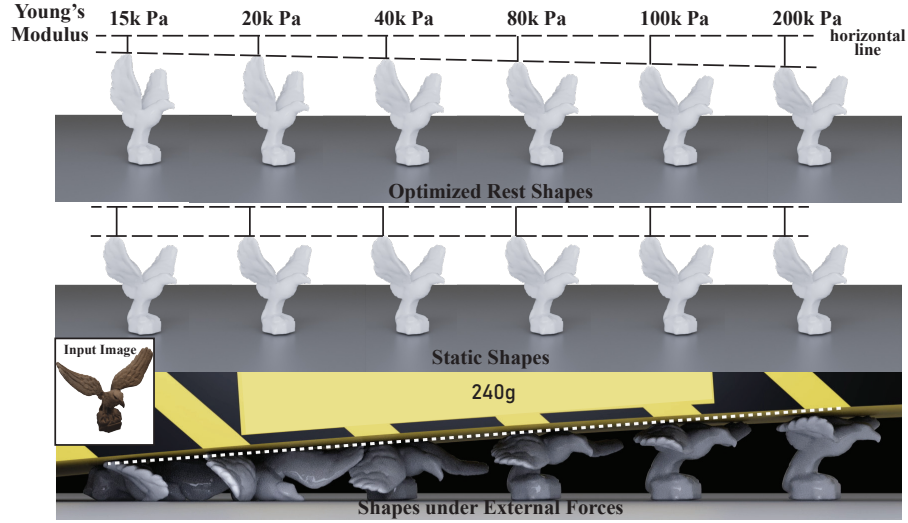


Figure 5: Ablation study on Young’s modulus. By changing the material properties, our method can produce various rest-shape geometries (top), which all result in the same static shapes that match the input image (middle). Although these static shapes appear identical under static equilibrium, they exhibit different deformation when subjected to the same compression forces exerted by the yellow block, attributable to the differences in their material properties (bottom).

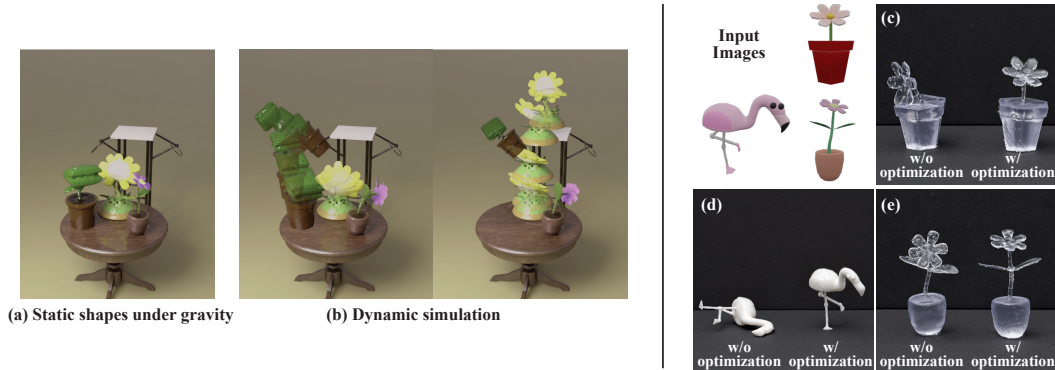


Figure 6: Applications of physically compatible objects. Left: Our optimized physical objects is simulation-ready and can be seamlessly integrated into dynamic simulation pipeline to produce complex dynamics and motions. Right: Real-world validation using 3D printing shows that shapes optimized using our method closely replicate the input images, demonstrating the practical effectiveness of our method in manufacturing.

physical principles to ensure that the optimized rest shapes are self-supporting and stable, and match the input images under static equilibrium.

#### 4.4 Analysis

**Ablation study on Young’s Modulus.** We investigate the influence of predefined mechanical material properties, particularly Young’s modulus, on the optimized rest shapes and their physical behaviors. Using the same input image, we obtained six optimized rest shapes with varying Young’s modulus values within our framework with TetSphere. As shown in Fig. 5, although the optimized rest-shape geometries vary, they all deform to the same static geometry under the influence of gravity, matching the input image. Moreover, the physical responses to identical external forces, such as compression by a box, differ due to the variations in material properties. These results highlight how the explicit decomposition of physical attributes in our framework expands the controllability of object modeling, allowing for diverse physical behaviors under uniform external forces.

**Application to dynamic simulation.** The immediate output of our method is a simulation-ready rest-shape geometry, which can be seamlessly integrated into a simulation pipeline to produce complex dynamics and motions. Fig. 6 (left) and the accompanying video in the Supplementary Material illustrate three plants modeled using our framework, demonstrating their behavior under gravity and complex interactions. Implementation details of this simulation are provided in Appendix F. These examples underscore the practical utility of our method for generating physically realistic dynamics and simulations.

**Application to fabrication.** We further evaluate our method in real world by fabricating three shapes using 3D printing, both with and without optimization. The results, shown in Fig. 6 (right), with detailed implementation procedures available in Appendix E, demonstrate that the 3D printed shapes align with our computational results. These real-world experiments demonstrate the practical effectiveness and validate the physical realism of the objects produced by our method.

## 5 Conclusion

In this work, we introduced physical compatibility optimization for reconstructing a physical object from a single image. Our method decomposes three orthogonal attributes governing physical behavior: mechanical properties, external forces, and rest-shape geometry. Unlike existing methods that often ignore one or more dimensions, our framework holistically considers all three factors, allowing for diverse rest-shape geometries from the same input image by varying object stiffness and external forces. We formulate physical compatibility optimization as a constrained optimization problem by integrating static equilibrium as a hard constraint. Our approach produces physical objects that match the geometry depicted in the input image under external forces and remain stable under gravity. Both quantitative and qualitative evaluations demonstrated improvements in physical compatibility over existing baselines. Our method’s versatility is evident through its integration with various single-view reconstruction methods and its practical applications in dynamic simulations and 3D printing.

**Limitations and Future Work** One limitation of our framework is its reliance on predefined material properties and external forces as inputs. Although this provides controllability of the final optimized rest-shape geometry, automating the extraction of these parameters from a single image presents a potential avenue for future work. Moreover, our method relies on the use of a tetrahedral mesh, which is derived by tetrahedralizing the output geometry produced by baseline methods. A natural extension of our work is the development of a differentiable converter that can transform any geometric representation into a tetrahedral mesh. This would enable future research where our physical compatibility optimization could be integrated into a pre-trained large reconstruction model, which could then be fine-tuned to directly produce physically compatible 3D objects. Lastly, our current methodology focuses solely on physical objects in a state of static equilibrium. Exploring the reconstruction of 3D objects undergoing dynamics captured from video is an intriguing prospect for future research.

## References

- [1] E. C. Aifantis. The physics of plastic deformation. *International journal of plasticity*, 3(3): 211–247, 1987.
- [2] G. Allaire. *Numerical analysis and optimization: an introduction to mathematical modelling and numerical simulation*. OUP Oxford, 2007.
- [3] S. Bell, K. Bala, and N. Snavely. Intrinsic images in the wild. *ACM Transactions on Graphics (TOG)*, 33(4):1–12, 2014.
- [4] A. H. Bermano, T. Funkhouser, and S. Rusinkiewicz. State of the art in methods and representations for fabrication-aware design. In *Computer Graphics Forum*, volume 36, pages 509–535. Wiley Online Library, 2017.
- [5] M. Botsch and O. Sorkine. On linear variational surface deformation methods. *IEEE transactions on visualization and computer graphics*, 14(1):213–230, 2007.

- [6] T. Burczyński, J. Kane, and C. Balakrishna. Comparison of shape design sensitivity analysis formulations via material derivative-adjoint variable and implicit differentiation techniques for 3-d and 2-d curved boundary element. *Computer methods in applied mechanics and engineering*, 142(1-2):89–109, 1997.
- [7] D. Charatan, S. Li, A. Tagliasacchi, and V. Sitzmann. pixelsplat: 3d gaussian splats from image pairs for scalable generalizable 3d reconstruction. *arXiv preprint arXiv:2312.12337*, 2023.
- [8] X. Chen, C. Zheng, W. Xu, and K. Zhou. An asymptotic numerical method for inverse elastic shape design. *ACM Transactions on Graphics (TOG)*, 33(4):1–11, 2014.
- [9] M. Deitke, D. Schwenk, J. Salvador, L. Weihs, O. Michel, E. VanderBilt, L. Schmidt, K. Ehsani, A. Kembhavi, and A. Farhadi. Objaverse: A universe of annotated 3d objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13142–13153, 2023.
- [10] Y. Feng, Y. Shang, X. Li, T. Shao, C. Jiang, and Y. Yang. Pie-nerf: Physics-based interactive elastodynamics with nerf. *arXiv preprint arXiv:2311.13099*, 2023.
- [11] M. Guo, B. Wang, K. He, and W. Matusik. Tetsphere splatting: Representing high-quality geometry with lagrangian volumetric meshes.
- [12] Z. He and T. Wang. Openlrm: Open-source large reconstruction models. <https://github.com/3DTopia/OpenLRM>, 2023.
- [13] Y. Hong, K. Zhang, J. Gu, S. Bi, Y. Zhou, D. Liu, F. Liu, K. Sunkavalli, T. Bui, and H. Tan. LRM: Large reconstruction model for single image to 3D. Nov. 2023.
- [14] J. Hsu, N. Truong, C. Yuksel, and K. Wu. A general two-stage initialization for sag-free deformable simulations. *ACM Transactions on Graphics (TOG)*, 41(4):1–13, 2022.
- [15] Y. Hu, Q. Zhou, X. Gao, A. Jacobson, D. Zorin, and D. Panozzo. Tetrahedral meshing in the wild. *ACM Trans. Graph.*, 37(4):60:1–60:14, July 2018. ISSN 0730-0301. doi: 10.1145/3197517.3201353. URL <http://doi.acm.org/10.1145/3197517.3201353>.
- [16] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 2023.
- [17] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *ICLR*, 2015.
- [18] M. Li, Z. Ferguson, T. Schneider, T. R. Langlois, D. Zorin, D. Panozzo, C. Jiang, and D. M. Kaufman. Incremental potential contact: intersection-and inversion-free, large-deformation dynamics. *ACM Trans. Graph.*, 39(4):49, 2020.
- [19] X. Li, Y.-L. Qiao, P. Y. Chen, K. M. Jatavallabhula, M. Lin, C. Jiang, and C. Gan. Pac-nerf: Physics augmented continuum neural radiance fields for geometry-agnostic system identification. 2022.
- [20] M. Liu, C. Xu, H. Jin, L. Chen, V. T. Mukund, Z. Xu, and H. Su. One-2-3-45: Any single image to 3D mesh in 45 seconds without Per-Shape optimization. June 2023.
- [21] R. Liu, R. Wu, B. V. Hoorick, P. Tokmakov, S. Zakharov, and C. Vondrick. Zero-1-to-3: Zero-shot one image to 3d object, 2023.
- [22] Y. Liu, C. Lin, Z. Zeng, X. Long, L. Liu, T. Komura, and W. Wang. SyncDreamer: Generating multiview-consistent images from a single-view image. Sept. 2023.
- [23] X. Long, Y.-C. Guo, C. Lin, Y. Liu, Z. Dou, L. Liu, Y. Ma, S.-H. Zhang, M. Habermann, C. Theobalt, and W. Wang. Wonder3D: Single image to 3D using Cross-Domain diffusion. Oct. 2023.
- [24] L. Melas-Kyriazi, C. Rupprecht, I. Laina, and A. Vedaldi. RealFusion: 360° reconstruction of any object from a single image. Feb. 2023.

- [25] M. Mezghanni, T. Bodrito, M. Boulkenafed, and M. Ovsjanikov. Physical simulation layer for accurate 3d modeling. In *CVPR*, 2021.
- [26] M. Mezghanni, M. Boulkenafed, A. Lieutier, and M. Ovsjanikov. Physically-aware generative network for 3d shape modeling. In *CVPR*, 2021.
- [27] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *ECCV*, 2020.
- [28] R. v. Mises. Mechanik der festen körper im plastisch-deformablen zustand. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1913: 582–592, 1913.
- [29] B. Nicolet, A. Jacobson, and W. Jakob. Large steps in inverse rendering of geometry. *ACM Transactions on Graphics (TOG)*, 40(6):1–13, 2021.
- [30] K. Perlin. Improving noise. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 681–682, 2002.
- [31] B. Poole, A. Jain, J. T. Barron, and B. Mildenhall. Dreamfusion: Text-to-3d using 2d diffusion. *arXiv preprint arXiv:2209.14988*, 2022.
- [32] G. Qian, J. Mai, A. Hamdi, J. Ren, A. Siarohin, B. Li, H.-Y. Lee, I. Skorokhodov, P. Wonka, S. Tulyakov, and B. Ghanem. Magic123: One image to high-quality 3d object generation using both 2d and 3d diffusion priors. *arXiv preprint arXiv:2306.17843*, 2023.
- [33] J. R. Shue, E. R. Chan, R. Po, Z. Ankner, J. Wu, and G. Wetzstein. 3d neural field generation using triplane diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20875–20886, 2023.
- [34] E. Sifakis and J. Barbic. Fem simulation of 3d deformable solids: a practitioner’s guide to theory, discretization and model reduction. In *Acm siggraph 2012 courses*, pages 1–50. 2012.
- [35] B. Smith, F. D. Goes, and T. Kim. Stable neo-hookean flesh simulation. *ACM Transactions on Graphics (TOG)*, 37(2):1–15, 2018.
- [36] J. Smith and S. Schaefer. Bijective parameterization with free boundaries. *ACM Transactions on Graphics (TOG)*, 34(4):1–9, 2015.
- [37] T. Standley, O. Sener, D. Chen, and S. Savarese. image2mass: Estimating the mass of an object from its image. In *Conference on Robot Learning*, pages 324–333. PMLR, 2017.
- [38] S. Szymanowicz, C. Rupprecht, and A. Vedaldi. Splatter image: Ultra-fast single-view 3d reconstruction. 2024.
- [39] J. Tang, J. Ren, H. Zhou, Z. Liu, and G. Zeng. DreamGaussian: Generative gaussian splatting for efficient 3D content creation. Sept. 2023.
- [40] J. Tang, Z. Chen, X. Chen, T. Wang, G. Zeng, and Z. Liu. LGM: Large Multi-View gaussian model for High-Resolution 3D content creation. Feb. 2024.
- [41] D. Tochilkin, D. Pankratz, Z. Liu, Z. Huang, A. Letts, Y. Li, D. Liang, C. Laforte, V. Jampani, and Y.-P. Cao. Triposr: Fast 3d object reconstruction from a single image. *arXiv preprint arXiv:2403.02151*, 2024.
- [42] B. Wang, G. Matcuk, and J. Barbič. Modeling of personalized anatomy using plastic strains. *ACM Transactions on Graphics (TOG)*, 40(2):1–21, 2021.
- [43] P. Wang, L. Liu, Y. Liu, C. Theobalt, T. Komura, and W. Wang. NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. June 2021.
- [44] X. Wei, K. Zhang, S. Bi, H. Tan, F. Luan, V. Deschaintre, K. Sunkavalli, H. Su, and Z. Xu. Meshlm: Large reconstruction model for high-quality mesh. *arXiv preprint arXiv:2404.12385*, 2024.

- [45] R. Wu, B. Mildenhall, P. Henzler, K. Park, R. Gao, D. Watson, P. P. Srinivasan, D. Verbin, J. T. Barron, B. Poole, et al. Reconfusion: 3d reconstruction with diffusion priors. *arXiv preprint arXiv:2312.02981*, 2023.
- [46] T. Xie, Z. Zong, Y. Qiu, X. Li, Y. Feng, Y. Yang, and C. Jiang. Physgaussian: Physics-integrated 3d gaussians for generative dynamics. *arXiv preprint arXiv:2311.12198*, 2023.
- [47] Y. Xu, Z. Shi, W. Yifan, H. Chen, C. Yang, S. Peng, Y. Shen, and G. Wetzstein. Grm: Large gaussian reconstruction model for efficient 3d reconstruction and generation. *arXiv preprint arXiv:2403.14621*, 2024.
- [48] A. Yu, V. Ye, M. Tancik, and A. Kanazawa. pixelNeRF: Neural radiance fields from one or few images. 2021.
- [49] A. J. Zhai, Y. Shen, E. Y. Chen, G. X. Wang, X. Wang, S. Wang, K. Guan, and S. Wang. Physical property understanding from language-embedded feature fields. *arXiv preprint arXiv:2404.04242*, 2024.
- [50] K. Zhang, S. Bi, H. Tan, Y. Xiangli, N. Zhao, K. Sunkavalli, and Z. Xu. Gs-lrm: Large reconstruction model for 3d gaussian splatting. *arXiv preprint arXiv:2404.19702*, 2024.
- [51] L. Zhong, H.-X. Yu, J. Wu, and Y. Li. Reconstruction and simulation of elastic objects with spring-mass 3d gaussians. *arXiv preprint arXiv:2403.09434*, 2024.

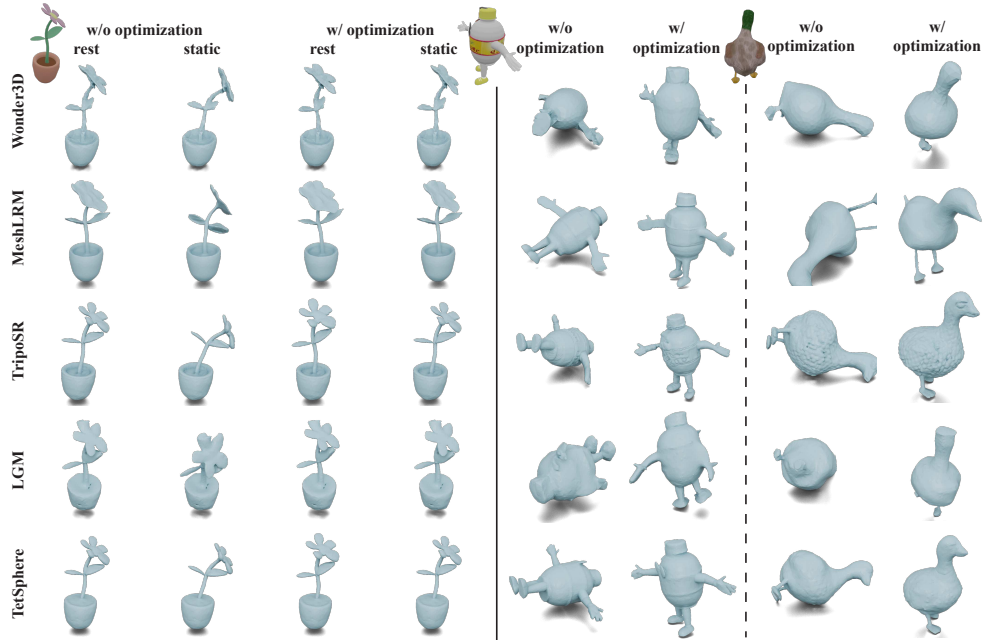


Figure 7: Additional qualitative results of physical compatibility optimization (part 1/2).

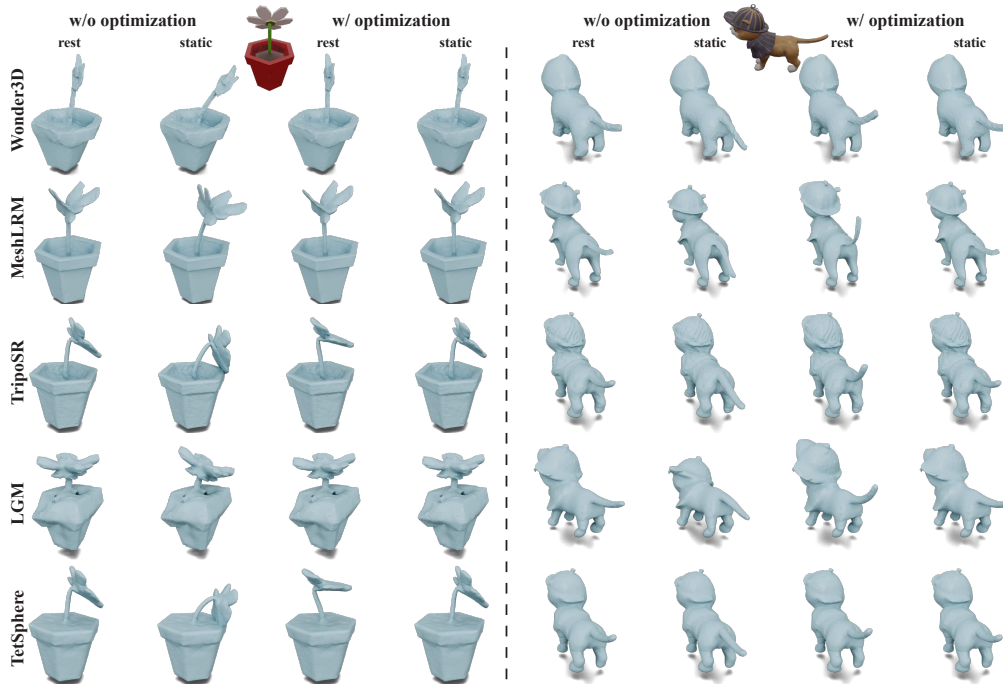


Figure 8: Additional qualitative results of physical compatibility optimization (part 2/2).

## A Additional Qualitative Results

Figure 7 and 8 show additional results of our physical compatibility optimization.

## B Plastic Strain Field $F_p$

To enhance the understanding of our framework without compromising generalizability, let us consider  $\mathcal{M}_{init}$  to be a tetrahedral mesh composed of a single element and four vertices. When

subject to static equilibrium influenced by gravity, the object adheres to the equation:

$$\mathbf{f}_{\text{int}}(\mathbf{x}, \phi(\mathbf{F}_{\mathbf{p}}; \mathbf{X}_{\text{init}}); \Theta) = \mathbf{M}\mathbf{g}, \quad (6)$$

where  $\mathbf{f}_{\text{int}}(\cdot, \cdot)$  denotes the elastic force (internal force),  $\mathbf{M}$  is the mass matrix, and  $\mathbf{g}$  denotes the gravity acceleration. To compute this force, we first consider the elastic energy  $\mathcal{E}$ . The definition of elastic energy unfolds as follows:

$$\begin{aligned} \mathcal{E}(\mathbf{F}_{\mathbf{e}}, \mathbf{F}_{\mathbf{p}}; \Theta) &= V(\mathbf{F}_{\mathbf{p}})\Phi(\mathbf{F}_{\mathbf{e}}; \Theta), \\ V(\mathbf{F}_{\mathbf{p}}) &= V_{\text{init}}\det(\mathbf{F}_{\mathbf{p}}), \\ \mathbf{F}_{\mathbf{e}} &= \mathbf{F}\mathbf{F}_{\mathbf{p}}^{-1}, \\ \mathbf{F} &= \partial\mathbf{x}/\partial\mathbf{X}_{\text{init}}, \end{aligned}$$

where  $V(\mathbf{F}_{\mathbf{p}})$  represents the volume of the element under plastic strain,  $V_{\text{init}}$  is the initial volume of the element,  $\mathbf{F}_{\mathbf{e}}$  denotes the elastic deformation gradient,  $\mathbf{F}$  represent the total deformation gradient, and  $\Phi(\cdot; \Theta)$  the elastic energy density function. This deformation gradient  $\mathbf{F}$  is computed through standard methodology [34].

Consequently, the derivation of the elastic force encapsulates the computation of the first-order partial derivative of the elastic energy with respect to the vertex positions:

$$\begin{aligned} \mathbf{f}_{\text{int}}(\mathbf{x}, \phi(\mathbf{F}_{\mathbf{p}}; \mathbf{X}_{\text{init}}); \Theta) &:= \frac{\partial\mathcal{E}(\mathbf{F}_{\mathbf{e}}(\mathbf{x}), \mathbf{F}_{\mathbf{p}}; \Theta)}{\partial\mathbf{x}} \\ &= V(\mathbf{F}_{\mathbf{p}}) \frac{\partial\Phi}{\partial\mathbf{F}_{\mathbf{e}}} : \frac{\partial\mathbf{F}}{\partial\mathbf{x}} \mathbf{F}_{\mathbf{p}}^{-1}. \end{aligned}$$

Notably, given the linear dependence of  $\mathbf{F}$  on  $\mathbf{x}$ ,  $\frac{\partial\mathbf{F}}{\partial\mathbf{x}}$  remains constant.

Given  $\mathbf{F}_{\mathbf{p}}$  and  $\mathbf{X}_{\text{init}}$  as inputs, the solution to Eq. 6 is the static shape,  $\mathbf{x} = \mathbf{x}_{\text{static}}$ . Likewise, to calculate  $\mathbf{X}_{\text{rest}}$  from  $\mathbf{F}_{\mathbf{p}}$  and  $\mathbf{X}_{\text{init}}$  in Eq. 3, we solve a similar equation with zero external force.

$$\mathbf{f}_{\text{int}}(\mathbf{x}, \phi(\mathbf{F}_{\mathbf{p}}; \mathbf{X}_{\text{init}}); \Theta) = \mathbf{0},$$

where the solution to this equation is  $\mathbf{X}_{\text{rest}}$ .

Considering the elastic energy, the translation of  $\mathbf{X}_{\text{init}}$  does not alter the deformation gradient  $\mathbf{F}$ . Consequently,  $\mathbf{F}_{\mathbf{p}}$  remain unaffected and exhibit translation invariance. In terms of the elastic force, it maintains translation invariance as well, since  $\mathbf{F}$  is not affected by any shift in  $\mathbf{X}_{\text{init}}$ .

Finally, by using isotropic materials, our approach enables a further reduction in the DOFs of  $\mathbf{F}_{\mathbf{p}}$ . Let us denote  $\mathbf{F}_{\mathbf{p}}$  as  $\mathbf{F}_{\mathbf{p}} = \mathbf{R}\mathbf{S}$ . The elastic deformation gradient is then derived as  $\mathbf{F}_{\mathbf{e}} = \mathbf{F}(\mathbf{R}\mathbf{S})^{-1} = \mathbf{F}\mathbf{S}^{-1}\mathbf{R}^{-1}$ . Given the invariance property  $\Phi(\mathbf{F}_{\mathbf{e}}; \theta) = \Phi(\mathbf{F}_{\mathbf{e}}\mathbf{R}; \theta)$ , which constantly holds for isotropic materials, the rotation component  $\mathbf{R}$  becomes redundant and can be excluded from the formulation. This simplification implies that the only requirement for  $\mathbf{F}_{\mathbf{p}}$  is to be a symmetric matrix. During the optimization process, this property facilitates the prevention of the inversion: In order to ensure that  $\det(\mathbf{F}_{\mathbf{p}}) > 0$ , we can simply adjust the eigenvalues of  $\mathbf{F}_{\mathbf{p}}$  to make them remain positive. This adjustment is crucial for the rest mesh  $\mathbf{X}_{\text{rest}}$  to maintain in the non-inverted state.

## C Computation of Gradient

By differentiating the constraint in Eq. 4 with respect to  $\mathbf{F}_{\mathbf{p}}$ , we obtain

$$\frac{\partial\mathbf{f}_{\text{net}}}{\partial\mathbf{F}_{\mathbf{p}}} + \frac{\partial\mathbf{f}_{\text{net}}}{\partial\mathbf{x}_{\text{static}}} \frac{\partial\mathbf{x}_{\text{static}}}{\partial\mathbf{F}_{\mathbf{p}}} = 0. \quad (7)$$

Then, we have

$$\frac{\partial\mathbf{x}_{\text{static}}}{\partial\mathbf{F}_{\mathbf{p}}} = -\left[\frac{\partial\mathbf{f}_{\text{net}}}{\partial\mathbf{x}_{\text{static}}}\right]^{-1} \frac{\partial\mathbf{f}_{\text{net}}}{\partial\mathbf{F}_{\mathbf{p}}}. \quad (8)$$

Substituting the result into the objective in Eq. 4, we get

$$\begin{aligned} \frac{\partial\mathcal{J}}{\partial\mathbf{F}_{\mathbf{p}}} &= \frac{\partial\mathcal{L}}{\partial\mathbf{F}_{\mathbf{p}}} + \frac{\partial\mathcal{L}_{\text{reg}}}{\partial\mathbf{F}_{\mathbf{p}}} \\ &= \frac{\partial\mathcal{L}}{\partial\mathbf{x}_{\text{static}}} \frac{\partial\mathbf{x}_{\text{static}}}{\partial\mathbf{F}_{\mathbf{p}}} + \frac{\partial\mathcal{L}_{\text{reg}}}{\partial\mathbf{F}_{\mathbf{p}}} \\ &= -\frac{\partial\mathcal{L}}{\partial\mathbf{x}_{\text{static}}} \left[\frac{\partial\mathbf{f}_{\text{net}}}{\partial\mathbf{x}_{\text{static}}}\right]^{-1} \frac{\partial\mathbf{f}_{\text{net}}}{\partial\mathbf{F}_{\mathbf{p}}} + \frac{\partial\mathcal{L}_{\text{reg}}}{\partial\mathbf{F}_{\mathbf{p}}}, \end{aligned} \quad (9)$$

which is the gradient with respect to  $\mathbf{F}_p$ . In practice,  $\frac{\partial \mathbf{f}_{\text{net}}}{\partial \mathbf{x}_{\text{static}}}$  and  $\frac{\partial \mathbf{f}_{\text{net}}}{\partial \mathbf{F}_p}$  are stored as sparse matrices and computed based on [42]. Considering about the performance, we first compute  $\frac{\partial \mathcal{L}}{\partial \mathbf{x}_{\text{static}}} \left[ \frac{\partial \mathbf{f}_{\text{net}}}{\partial \mathbf{x}_{\text{static}}} \right]^{-1}$  using sparse linear solver. This results in a dense vector with size  $3N$ . We then multiply it with  $\frac{\partial \mathbf{f}_{\text{net}}}{\partial \mathbf{F}_p}$ .

## D Implementation Details of Evaluation

To evaluate the physical compatibility of baseline methods, which often produce shapes comprising multiple connected components, we first extract the largest connected component from each mesh. All meshes are then normalized to the unit cube. Notably, the reconstructed shapes from TripoSR and Wonder3D are not axis-aligned; thus, we manually rotate these shapes to ensure the head points towards the  $z$ -axis in the world coordinate space. For integrating our physical compatibility framework, We use two sets of Young’s modulus,  $E = 5 \times 10^4 \text{Pa}$  and  $E = 5 \times 10^5 \text{Pa}$ , which are selected based on whether the shape would become overly soft, potentially leading to static equilibrium failure due to excessive stress causing numerical bounds to be exceeded. Poisson’s ratio  $\nu = 0.45$  and mass density  $\rho = 1000 \text{kg/m}^3$  are consistent across all meshes. Evaluation metrics require solving for static equilibrium Eq. 1. We employ the Newton-Raphson solver with line search, setting the maximum number of iterations to be 200. For optimizing Eq. 4, we use gradient descent and allow up to 1000 iterations. Our experiments run on a desktop PC with an AMD Ryzen 9 5950X 16-core CPU and 64GB RAM. The average runtime for this optimization process is approximately 80 seconds.

## E Implementation Details of 3D Printing

The selected model shapes were 3D printed using stereolithography (Form3; Formlabs, 100  $\mu\text{m}$  layer thickness) to create the flexible designs (using Flexible 80A, tensile modulus  $<3 \text{MPa}$ , 100% strain to failure) and rigid designs (using White Resin V4; tensile modulus 1.6 GPa), both without post-curing. The flexible flowers are 55 and 65 mm in height and the rigid goose is 50 mm in length. Shapes with and without optimization were printed with similar support structures designed to preserve delicate features.

## F Dynamic Simulation of Deformable Objects

We model each solid deformable object using FEM with hyperelastic materials for dynamic simulation. Then, we solve the standard partial differential equation (PDE) for dynamic FEM simulation:

$$M\ddot{x} + D(x)\dot{x} + f_{\text{elastic}}(x) + f_{\text{attachment}}(x) + f_{\text{contact}}(x) = Mg, \quad (10)$$

where  $x$  represents the node positions within the finite element meshes – we use tetrahedral meshes – of the objects,  $M$  denotes the mass matrix,  $D$  is the Rayleigh damping matrix,  $f_{\text{elastic}}(\cdot)$  is the hyperelastic forces,  $f_{\text{attachment}}(x)$  is the attachment forces that constrain the objects to a specific location, and  $f_{\text{contact}}(\cdot)$  denotes the contact forces between surfaces. We employ the implicit backward Euler method for time discretization, transforming the PDE into:

$$A^n x^{n+1} + b^n + f_{\text{elastic}}(x^{n+1}) + f_{\text{attachment}}(x^{n+1}) + f_{\text{contact}}(x^{n+1}) = 0, \quad (11)$$

where  $x^{n+1}$  is the position vector at timestep  $(n + 1)$ ,  $A^n$  and  $b^n$  is a constant matrix and vector, respectively, derived from values at timestep  $n$ . Finally, we solve this nonlinear equation using Newton’s method at each timestep.

The hyperelastic material selected for the deformable objects is the same as the one used for the rest shape optimization [35] in Sec. 3. Attachment forces are modeled as spring forces  $f_{\text{attachment}}(x) = k_a(Sx - \bar{x}(t))$ , where  $k_a$  is the stiffness of the spring, the selection matrix  $S$  selects the attached vertices, and  $\bar{x}(t)$  denotes the target attachment locations at time  $t$ . Contact forces are generated from penalizing any vertex penetration into the contact surface, expressed as  $f = k_c d$ , where  $k_c$  represents the contact stiffness and  $d$  denotes the penetration depth, with  $d = 0$  in the absence of contact. This gives the normal contact forces. Friction forces are computed following the methods outlined in [18]. Then, the total contact force  $f_{\text{contact}}$  is the sum of normal contact forces and friction forces.

For the dynamic simulation in Figure 7, the attachment of each plant is defined as the bottom part of each pot. We keyframe-animate the trajectory of attachment vertices  $\bar{x}(t)$ . Gravity is enabled



throughout the entire simulation. At the end of the sequence, we apply wind forces to the plants, computed using 4D Perlin Noise [30].

## **G Broader Impacts**

Our research presents a computational framework for reconstructing physical objects from single images. This advancement holds significant potential for various applications, including dynamic simulations, 3D printing, virtual reality, and industrial design. By ensuring that the reconstructed objects adhere to real-world physical laws, our method can enhance the realism and functionality of virtual environments, improve the precision of 3D printed objects, and contribute to the development of more reliable industrial designs.

There are mainly two potential negative societal impacts: Improved 3D reconstruction capabilities could potentially be misused to create highly realistic fake objects or environments for disinformation purposes. This could include generating deceptive media content that appears authentic. As the framework automates the reconstruction process, there is a potential risk of it being used in automated systems without sufficient oversight, potentially leading to unintended and harmful outcomes due to errors or misuse. Developing systems to monitor the use of the technology and ensure accountability for its applications, as well as providing comprehensive guidelines and training for users to promote ethical use and awareness of potential misuse, will address these potential negative impacts.

## NeurIPS Paper Checklist

### 1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [\[Yes\]](#)

Justification: See Section 3 and 4.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

### 2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: See Section 5.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

### 3. Theory Assumptions and Proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [\[Yes\]](#)

Justification: See Section 3, B, and C.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

#### 4. Experimental Result Reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: See Section 3, 4, and D.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

#### 5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: Project page is available at <https://gmh14.github.io/phys-comp/>.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

## 6. Experimental Setting/Details

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: See Section 4 and D.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

## 7. Experiment Statistical Significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: See Table 1 and Figure 3.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

## 8. Experiments Compute Resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: See Section D.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

## 9. Code Of Ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines?>

Answer: [Yes]

Justification: The authors fully acknowledged the code of ethics and commit to adhering to its guidelines.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

## 10. Broader Impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: See Section G.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

## 11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The paper poses no such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

## 12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: See Section 4.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, [paperswithcode.com/datasets](https://paperswithcode.com/datasets) has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset’s creators.

### 13. **New Assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: The paper does not release new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

### 14. **Crowdsourcing and Research with Human Subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

### 15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.