

DensiCrafter: Physically-Constrained Generation and Fabrication of Self-Supporting Hollow Structures

Shengqi Dang^{1,2,3}, Fu Chai^{2,3}, Jiaxin Li², Chao Yuan^{1,3}, Wei Ye^{1,2,3}, Nan Cao^{1,2,3*}

¹ Shanghai Research Institute for Intelligent Autonomous Systems

² Shanghai Innovation Institute

³ Tongji University

dangsq123@tongji.edu.cn, nan.cao@gmail.com

Abstract

The rise of 3D generative models has enabled automatic 3D geometry and texture synthesis from multimodal inputs (e.g., text or images). However, these methods often ignore physical constraints and manufacturability considerations. In this work, we address the challenge of producing 3D designs that are both lightweight and self-supporting. We present DensiCrafter, a framework for generating lightweight, self-supporting 3D hollow structures by optimizing the density field. Starting from coarse voxel grids produced by Trellis, we interpret these as continuous density fields to optimize and introduce three differentiable, physically constrained, and simulation-free loss terms. Additionally, a mass regularization penalizes unnecessary material, while a restricted optimization domain preserves the outer surface. Our method seamlessly integrates with pretrained Trellis-based models (e.g., Trellis, DSO) without any architectural changes. In extensive evaluations, we achieve up to 43% reduction in material mass on the text-to-3D task. Compared to state-of-the-art baselines, our method could improve the stability and maintain high geometric fidelity. Real-world 3D-printing experiments confirm that our hollow designs can be reliably fabricated and could be self-supporting.

Code — <https://github.com/idvxlab/DensiCrafter>

Extended version — <https://arxiv.org/abs/2511.09298>

Introduction

Recent advances in 3D generative models, such as Trellis (Xiang et al. 2025) and CLAY (Zhang et al. 2024), have enabled the automatic synthesis of high-quality surface geometry and textures from diverse inputs (e.g., text or images). These capabilities dramatically lower the barrier for designers and creators to prototype rich 3D assets, supporting applications in virtual reality, gaming, and digital art. In addition to generating virtual assets, such 3D generative models can also serve as the front-end for downstream fabrication workflows, such as 3D printing and robotic assembly.

However, bringing these generated assets into the physical world demands more than visual plausibility. It requires

adherence to real-world physics, including material properties, structural mechanics, load-bearing behavior, and overall manufacturability, which current pretrained 3D generative models seldom consider. Recent works begin to close this gap. Static stability under gravity (i.e., self-supporting) is often the first constraint imposed: Atlas3D (Chen et al. 2024) uses simulation-based optimization to progressively stabilize generated meshes, while DSO (Li et al. 2025) further finetunes the Trellis to generate more stable structures. However, previous methods assume the generated 3D assets as solid interiors and only adjust surface meshes. In reality, the mass distribution is equally critical to physical performance. Moreover, efficient material usage remains essential for fabrication and sustainability. However, to the best of our knowledge, there are no solutions that integrate density-based optimization directly into 3D generative models.

From the perspective of manufacturability and physical constraints, we raise the question: *can we automatically generate lightweight, self-supporting structures while preserving the high-quality outer surface produced by modern 3D generators?* Even when restricted to rigid bodies, achieving this is highly non-trivial and presents several key challenges: (1) Most 3D generative models focus solely on surface geometry, making it difficult to incorporate internal mass distribution into their pipelines. (2) Enforcing self-supporting behavior typically requires rigid-body dynamics simulations, which are computationally expensive and poorly suited to continuous mass distribution. (3) A careful balance must be struck between optimizing internal mass and preserving the original surface geometry. Modifying the mass distribution should not compromise the visual fidelity of the generated mesh, which is not an easy task.

To address these challenges, we present *DensiCrafter*, a method for generating lightweight, self-supporting hollow structures from multimodal inputs. Our approach builds on the pretrained state-of-the-art 3D generative model Trellis (Xiang et al. 2025), which employs a two-stage pipeline: (1) synthesizing a voxel grid to capture object occupancy, (2) reconstructing the high-resolution mesh and textures based on the voxel grid. Our method represents Trellis’s voxel grid output as a continuous density field and performs online optimization to incorporate physically-aware internal structure into the generative process. To promote self-supporting behavior without resorting to expensive simulations, we in-

*Nan Cao is the corresponding author.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

roduce three differentiable, simulation-free constraints: (1) shifting the center of mass into the support region, (2) maximizing the bottom contact area, and (3) minimizing the vertical position of the mass center. To ensure material efficiency while preserving visual quality, we penalize total mass and constrain optimization to the object’s interior and a thin bottom region, preserving the outer surface. The optimized density field is converted back into a voxel grid and fed into Trellis to reconstruct the high-resolution outer mesh, while the internal hollow geometry is extracted from the optimized density. Our method achieves up to a 43% reduction in material mass and improves upright stability over all baselines, with minimal impact on inference time. The generated structures retain high semantic alignment and geometric fidelity. Real-world FDM 3D printing further confirms the reliability and manufacturability of the resulting designs (Fig. 8, 9). The main contributions are :

- We introduce a novel task of generating lightweight, self-supporting 3D structures, and propose DensiCrafter, a framework that integrates density field optimization seamlessly into the existing 3D generation pipeline.
- We design a set of fully differentiable, simulation-free, and physically-constrained losses that guide the optimization by restricting the center of mass and the bottom contact surface, ensuring static stability.
- We incorporate two regularization terms, including a mass penalty and a spatially restricted optimization domain, achieving material efficiency with surface fidelity.
- We validate our method through extensive evaluations, and real-world FDM 3D printing experiments demonstrate substantial reductions in material mass, improved upright stability, and practical manufacturability.

Related Work

We review the related works from the following two aspects.

3D Generative Model

Current 3D generative models (Xiang et al. 2025; Zhang et al. 2024; Zhao et al. 2025) can produce high-quality 3D representations, such as NeRF (Mildenhall et al. 2021), 3DGS (Kerbl et al. 2023), CAD instructions (Wu, Xiao, and Zheng 2021), or Tetrahedron (Guo et al. 2025), conditioned on various inputs. Due to the distinct nature of geometry and texture, researchers often decouple their generation processes (Zhang et al. 2024; Zhao et al. 2025). Trellis (Xiang et al. 2025) introduces a structured latent variable approach, dividing generation into two stages: coarse voxel generation followed by fine-grained geometry and material synthesis. Leveraging Trellis’s strong generative performance, we can optimize the coarse voxel structure to adjust density for desired physical properties, and then generate textures and detailed geometry conditioned on the refined voxel structure.

Physics-aware Generation

Recently, there has been growing interest in endowing 3D generative models with physically plausible behavior. Some

works directly infer an object’s material properties and geometry from input images or video (Standley et al. 2017; Zhong et al. 2024; Zhai et al. 2024; Li et al. 2023), or reconstruct a complete 3D shape and then subject it to physics-based post-processing under scene constraints (Xie et al. 2024; Feng et al. 2024). Others first optimize existing 3D representations to synthesize new training datasets (Shu et al. 2020), and train a generative model that inherently respects physical requirements (Wang and He 2019). Depending on the application, different physical attributes are targeted. For example, Atlas3D (Chen et al. 2024), PhysComb (Guo et al. 2024), and DSO (Li et al. 2025) focus on ensuring that generated shapes remain statically stable under gravity. However, these methods typically adjust only the surface geometry of the 3D object, assuming the object is a solid of uniform density.

We enhance Trellis’s workflow by integrating topology optimization concepts (Bendsøe 1989; Kang and Wang 2011; Bendsøe and Sigmund 1999; Musialski et al. 2016; Bäcker et al. 2014; Hafner, Ly, and Wojtan 2024). The traditional topology optimization requires explicit problem specification—including boundary conditions, loads, and mechanical constraints—to optimize material distribution, and is typically applied as a post-processing step after shape design. In contrast, our method incorporates topology-aware considerations directly into the content generation process, without requiring predefined mechanical constraints.

Methodology

As shown in Fig. 1, DensiCrafter builds a lightweight, self-supporting 3D model for fabrication via two key steps: (a) *Density Field Optimization*: We first convert the solid object O (a set of points within a N^3 voxelized cubic domain Ω) generated by Trellis into a continuous density field $\tilde{\rho}$. This field is then optimized to enhance upright stability and reduce material usage, subject to physical constraints, while preserving the original surface geometry. (b) *Hollow Structure Generation*: Given the optimized density field $\tilde{\rho}^{\text{opti}}$, we extract a high-resolution outer surface mesh along with a complementary inner surface mesh, resulting in a lightweight, self-supporting structure ready for fabrication. The following sections detail each of these steps.

Density Field Optimization

Our method casts a 3D shape as a continuous density field and optimizes the density field by solving

$$\min_{\theta(\mathbf{x}) : \mathbf{x} \in \Omega^{\text{opti}}} \mathcal{L}_{\text{phy}} + \lambda_{\text{mass}} \mathcal{L}_{\text{mass}} \quad (1)$$

where $\theta(\mathbf{x})$ is the field of parameters that defines the density field $\tilde{\rho}$ at each spatial point \mathbf{x} . The term \mathcal{L}_{phy} is the physical loss to enforce self-supporting behavior and upright stability. The term $\mathcal{L}_{\text{mass}}$ is a loss to minimize the amount of materials used for fabrication, weighted by λ_{mass} . The optimization domain Ω^{opti} restricts updates to the object’s interior and a thin basal layer, preserving the geometry of the outer surface. Next, we first introduce the use of $\theta(\mathbf{x})$ to represent the continuous density field. We then detail the physical loss \mathcal{L}_{phy} used to enforce self-supporting behavior. Finally, we

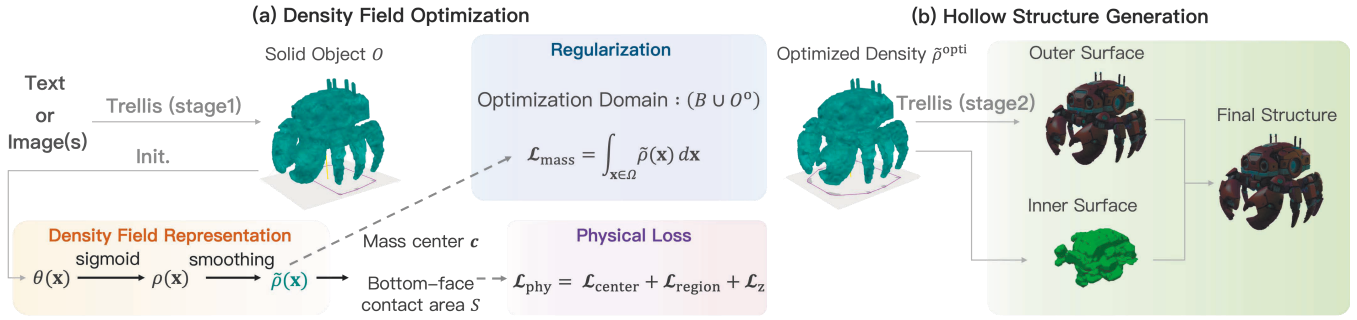


Figure 1: Pipeline of our method. We optimize this density field using differentiable losses that embed simulation-free physical constraints, yielding a material-efficient, self-supporting result.

describe two regularization strategies: the mass penalty term $\mathcal{L}_{\text{mass}}$ and the restriction of the optimization domain O^{opti} .

Density Field Representation. The mass distribution is represented within the cubic domain Ω , as a continuous density field $\tilde{\rho} : \mathbf{x} \rightarrow [0, 1]$, where $\tilde{\rho}(\mathbf{x}) \approx 1$ indicates solid material and $\tilde{\rho}(\mathbf{x}) \approx 0$ denotes empty space. This density field can be parameterized by:

$$\rho(\mathbf{x}) = \text{sigmoid}(\theta(\mathbf{x})) = \frac{1}{1 + \exp(-\theta(\mathbf{x}))} \quad (2)$$

where $\theta : \mathbf{x} \rightarrow \mathbb{R}$ is the scale field. We initialize θ to preserve the surface structure of O :

$$\theta(\mathbf{x}) \stackrel{\text{init.}}{=} \begin{cases} +\infty & \mathbf{x} \in \partial O \\ -\infty & \mathbf{x} \in \Omega \setminus \partial O \end{cases} \quad (3)$$

where the point set ∂O denotes the boundary of O . To promote spatial coherence and avoid isolated material regions, we apply a local smoothing operator to ρ , resulting in $\tilde{\rho}$ that is used for defining our loss functions:

$$\tilde{\rho}(\mathbf{x}) = \begin{cases} \rho(\mathbf{x}) & \mathbf{x} \in \partial O \\ (K * \rho)(\mathbf{x}), & \mathbf{x} \in \Omega^{\text{opti}} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

In our implementation, the smoothing operator applies a $3 \times 3 \times 3$ averaging kernel K to Ω^{opti} , while preserving the object’s original boundary.

Physical Loss. The physical loss \mathcal{L}_{phy} comprises three differentiable, simulation-free terms that enable the self-supporting behavior: (1) aligning the center of mass \mathbf{c} within the support region, i.e., the center loss ($\mathcal{L}_{\text{center}}$), (2) maximizing the bottom contact region S , i.e., the region loss ($\mathcal{L}_{\text{region}}$), and (3) lowering the height of the center of mass, i.e., the height loss (\mathcal{L}_z), which is formally defined as:

$$\mathcal{L}_{\text{phy}} = \mathcal{L}_{\text{center}} + \mathcal{L}_{\text{region}} + \mathcal{L}_z \quad (5)$$

Here, the bottom contact region S and the mass center \mathbf{c} can be calculated as:

$$S = \{\mathbf{s} \in \Omega : s_z = z_{\min}, \tilde{\rho}(\mathbf{s}) > 0.5\} \quad (6)$$

$$\mathbf{c} = \frac{\int_{\Omega} \mathbf{x} \tilde{\rho}(\mathbf{x}) dV}{\int_{\Omega} \tilde{\rho}(\mathbf{x}) dV} \quad (7)$$

where z_{\min} means the ground-level of O and dV denotes the differential volume element.

Centering Loss. Directly constraining the center of mass to lie inside the support region (i.e., convex hull of S) is not easily differentiable. Instead, we observe that the centroid of S always lies within the support region (proof provided in the appendix). We penalize the projection of \mathbf{c} onto the bottom plane close to the centroid of S :

$$\mathcal{L}_{\text{center}} = \frac{1}{|S|} \int_{\mathbf{s} \in S} \| (s_x, s_y) - (c_x, c_y) \|_2 dA \quad (8)$$

where dA denotes the differential area element.

Region Loss. A larger support region increases the object’s ability to withstand larger perturbations. Since for any two sets $S_1 \subset S_2$ we have $\text{ConvexHull}(S_1) \subset \text{ConvexHull}(S_2)$, expanding the bottom-face contact area directly enlarges the support region. We thus encourage a large contact area by minimizing its negative:

$$\mathcal{L}_{\text{region}} = -|S| = - \int_{\mathbf{s} \in S} dA \quad (9)$$

Height Loss. Lowering the center of mass increases the critical overturning angle, thereby enhancing stability against rotational perturbations. Therefore, we penalize c_z :

$$\mathcal{L}_z = c_z \quad (10)$$

Regularization. To minimize the total mass while preserving the overall shape, we introduce two regularizations: the mass penalty and the restricted optimization domain.

Mass Penalty. We directly penalize the total mass of the density field via the loss term $\mathcal{L}_{\text{mass}}$:

$$\mathcal{L}_{\text{mass}} = \int_{\Omega} \tilde{\rho}(\mathbf{x}) dV \quad (11)$$

Restricted Optimization Domain. Instead of optimizing the entire voxel grid, we restrict updates to the object’s interior O° and a thin basal neighborhood B , thereby preserving the external surface structure. The optimization domain Ω^{opti} is defined as

$$\Omega^{\text{opti}} = (O^\circ \cup B) \quad (12)$$

where $O^\circ = O \setminus \partial O$ is the interior of the original solid object O , and

$$B = \{\mathbf{x} \in \Omega \setminus O : z \in [z_{\min}, z_{\min} + \epsilon], \exists \mathbf{x}' \in O, \text{s.t.} \|\mathbf{x} - \mathbf{x}'\|_1 < \epsilon\} \quad (13)$$

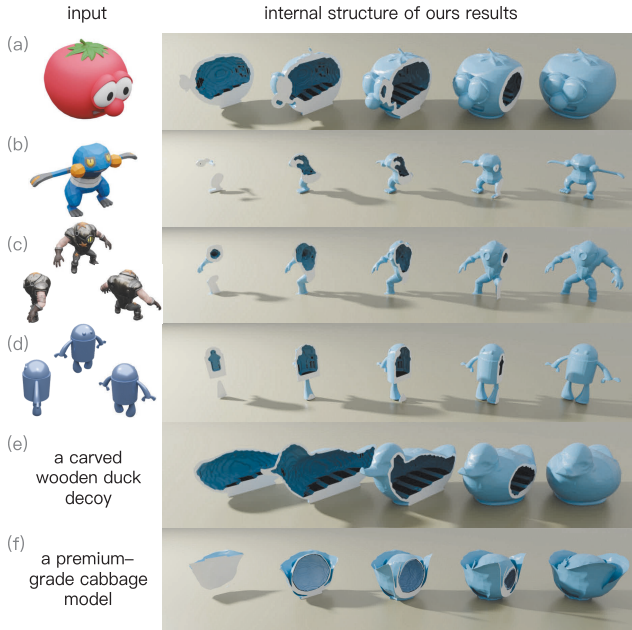


Figure 2: Self-supporting structures generated from diverse inputs by optimizing the internal mass distribution; all examples remain upright under simulation.

is a thin layer above the ground plane that lies beneath existing material, essential for the supporting structure.

Hollow Structure Generation

After optimization has converged, we generate the final hollow structure in three stages. (1) We compute the solid and hollow regions, based on the optimized density field $\tilde{\rho}$. Specifically, the solid region is defined as $V = \{\mathbf{x} \in \Omega \mid \tilde{\rho}(\mathbf{x}) > 0.5\}$, and the hollow region as $H = \{\mathbf{x} \in O \mid \tilde{\rho}(\mathbf{x}) \leq 0.5\}$. (2) We extract the outer surface mesh $\mathcal{M}_{\text{outer}}$ from V using Trellis (Xiang et al. 2025), and the inner surface mesh $\mathcal{M}_{\text{inner}}$ from H using Marching Cubes (Lorenson and Cline 1987). (3) We invert the normals of $\mathcal{M}_{\text{inner}}$ and combine it with $\mathcal{M}_{\text{outer}}$ to construct the final structure. Since the Trellis-generated outer surface aligns precisely with the voxels, the final outer surface resides strictly within the voxel cells corresponding to ∂O . This precise alignment facilitates non-intersection between the inner and outer surfaces. In our experiments, we did not observe any intersections or perforations between the inner and outer surfaces.

Experiments

In this section, we present our experimental results, including baseline comparisons, ablation study and real-world 3D-printing experiments.

Experimental Settings

We present the experimental settings, covering implementation, baselines, dataset details, and metrics.

Implementation. We performed our model on a NVIDIA H100 GPU on text-to-3D and image-to-3D (single-view and

	Atlas3D	Trellis	DSO/DPO	Ours
physically constrained	✓	✗	✓	✓
material efficient	✗	✗	✗	✓
efficient generation	✗	✓	✓	✓
without finetuning	✓	✓	✗	✓
high-quality surface	✗	✓	✓	✓

Table 1: Comparison of baselines and our method across key criteria.

multi-view) tasks. Consistent with Trellis, we set $N = 64$. For hyperparameters, we used $\lambda_{\text{mass}} = 100$ and $\epsilon = 2$ found on a small validation set (analysis provided in the appendix) and 42 as the global random seed. The optimization employed the Adam optimizer for 2000 steps with a learning rate of $1e^{-2}$. All the objects were scaled to the range of $[-0.5, 0.5]^3$ for comparison.

Baselines. We compare our method against SOTA baselines, including both generative models and physically-aware optimization: (1) **Trellis** (Xiang et al. 2025) is a leading 3D generative model that produces high-quality solid geometry from text or images. (2) **Atlas3D** (Chen et al. 2024) optimizes a differentiable simulation objective to encourage self-supporting geometry during synthesis on the text-to-3D task. (3) **DPO** (Wallace et al. 2024) improves physical plausibility by learning from paired supervision (i.e., plausible vs. implausible shapes). (4) **DSO** (Li et al. 2025) avoids paired data by optimizing generation through direct physical reward signals derived from differentiable simulations. We compare our method with the baselines in Table 1. For text-to-3D, we compare with **Trellis** and **Atlas3D**. For image-to-3D (including single- and multi-view), we include **Trellis**, **DPO**, and **DSO**. Our method is fully compatible with Trellis-based models. Accordingly, we also evaluate its integration with DPO and DSO.

Dataset. For the text-to-3D testset, we assembled 150 concise prompts generated by GPT, each describing a target 3D object (e.g., “a detailed dragon fruit model”). For the image-to-3D testset, we collected rendering images of unsupported, stability-challenging 3D models, from a carefully selected subset of the Objaverse-xl (Deitke et al. 2023) dataset. 150 objects were collected for rendering, spanning characters, animals, and various unsupported decorative objects. For single-view image input, only a frontal rendering was used. For multi-view image inputs, three renderings from distinct viewpoints were used.

Metrics. We evaluate our method and baselines from the following four aspects, including material efficiency, physical stability, semantic consistency, and geometric fidelity:

- **Material Efficiency (Mas).** We compute the average material mass of all generated objects (Mas).
- **Physical Stability (Stable, Rot).** We run simulations in Mujoco (Todorov, Erez, and Tassa 2012) to test and report the proportion of stable outcomes (Stable) and the average angular deviation from the upright pose (Rot).

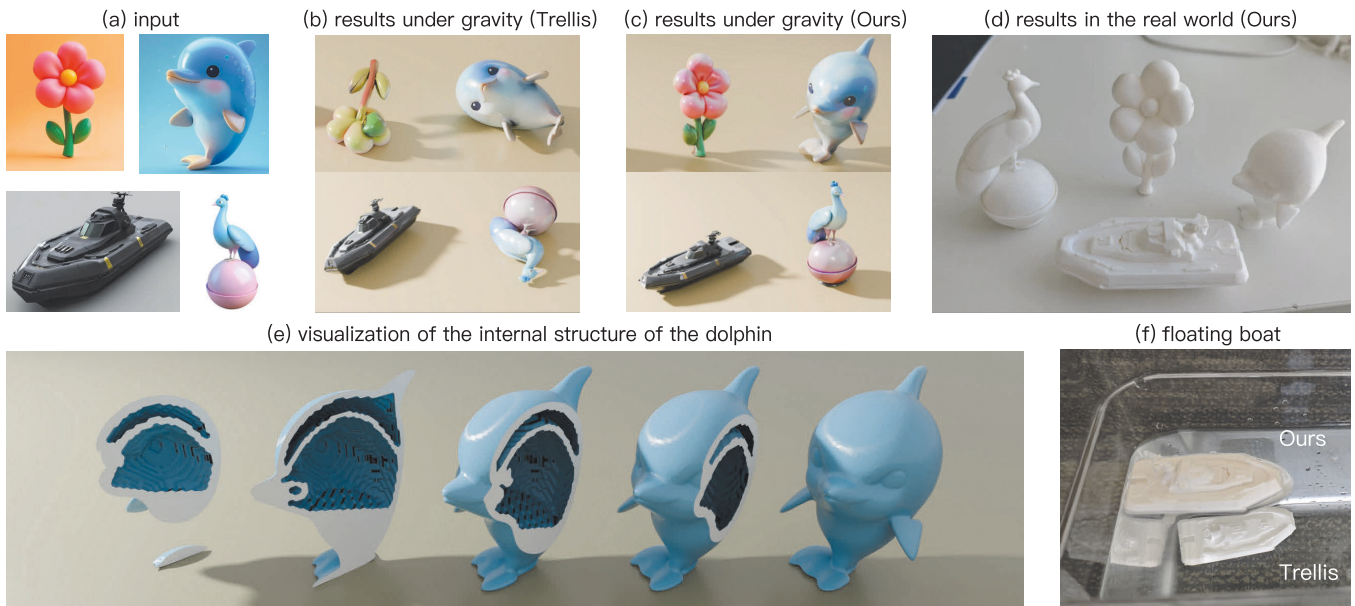


Figure 3: Our method produces self-supporting, material-efficient hollow structures.

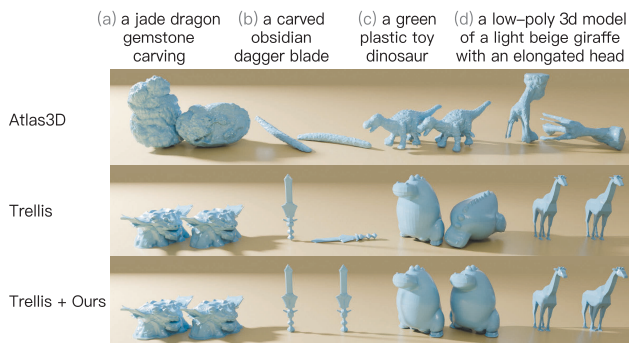


Figure 4: Text-to-3D generation results (each pair: left: without gravity; right: under gravity).

For hollowed structures, we recompute the inertia and center of mass for simulation.

- **Semantic Consistency (CLIP).** We measure how consistently the output matches the input by computing the CLIPScore (Hessel et al. 2021) between multi-view renderings of the generated object and the input (CLIP).
- **Geometric Fidelity (F-score, CF).** To evaluate how much the generated surface mesh deviates from the original mesh generated by Trellis, we use the Chamfer Distance (CF) and F-Score between the results of each baseline and Trellis.

Results and Comparisons

We first showcase representative examples, and then provide both qualitative and quantitative comparisons.

Results. Fig. 2 illustrates how our method adaptively hollows out internal regions to improve stability while preserving the visual fidelity of the input across different modalities.

		Mas↓	Stable↑	Rot ↓	CLIP↑
text	Trellis	0.214	84%	14.39°	24.4
	+Ours	0.121 (-43%)	92.7%	5.48°	24.4
single view	Trellis	0.046	68.0%	32.23°	84.7
	+Ours	0.036 (-22%)	83.3%	15.80°	84.4
image	DPO	0.055	83.3%	15.41°	84.3
	+Ours	0.044 (-19%)	88.7%	10.53°	84.1
	DSO	0.079	91.3%	8.28°	85.0
	+Ours	0.057 (-28%)	94%	5.64°	84.5
multi view images	Trellis	0.047	71.3%	26.49°	84.6
	+Ours	0.035 (-26%)	87.3%	11.59°	84.6
	DPO	0.052	83.3%	14.95°	84.6
	+Ours	0.039 (-24%)	90.7%	8.19°	84.4
	DSO	0.078	89.3%	10.61°	84.6
	+Ours	0.055 (-29%)	93.3%	6.60°	84.7

Table 2: Quantitative evaluation of material efficiency, physical stability, and semantic consistency.

The hollowing pattern typically targets the upper regions of the object, where cavities can be introduced without compromising structural support. This process results in a substantial volume reduction, especially in objects with large internal mass (e.g., Fig 2 (a,e,f)), where extensive and continuous internal voids are carved. In contrast, for objects with relatively less internal volume (e.g., Fig 2 (b)), the optimization selectively hollows out parts of the torso while preserving denser lower-body regions to enhance standing stability.

Qualitative Comparison. For the text-to-3D generation (Fig. 4), we observe that Atlas3D struggles to generate 3D objects that reflect the input text (e.g. dinosaur with five legs and the incomplete giraffe model). Additionally,

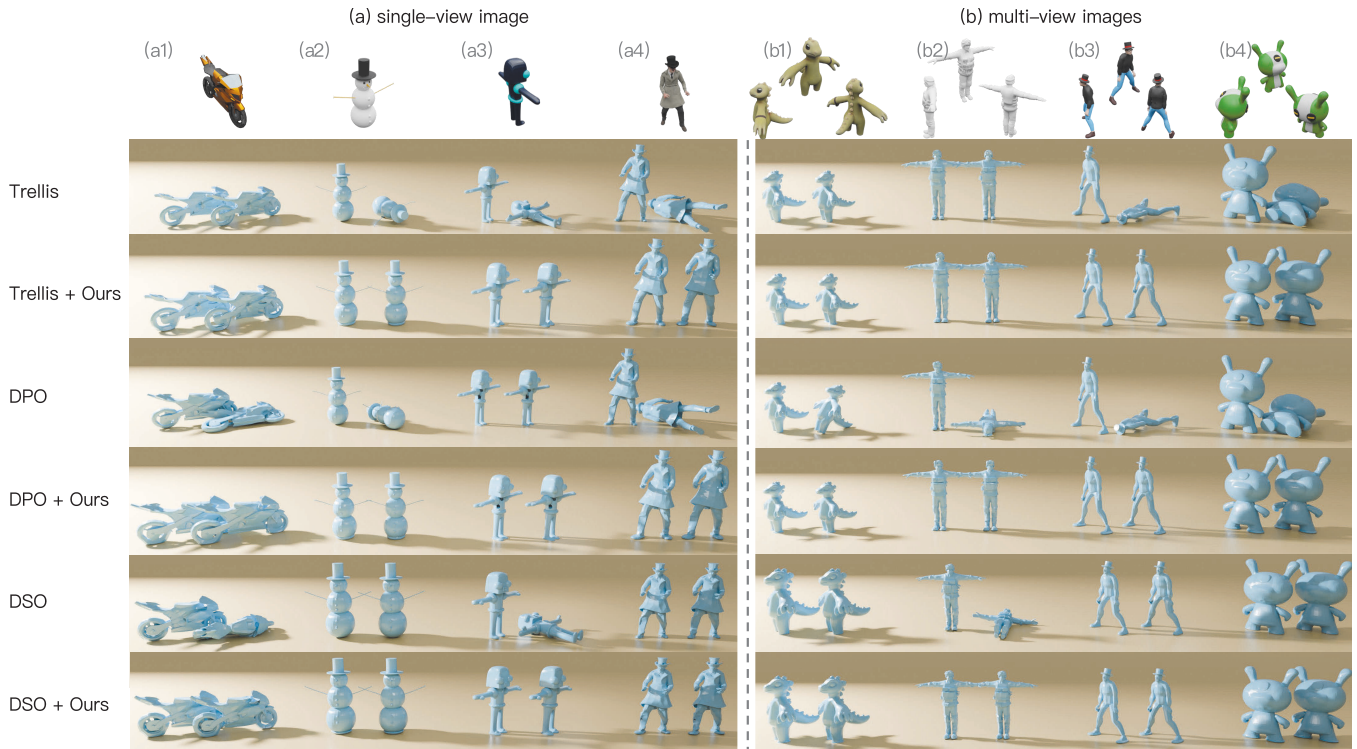


Figure 5: Image-to-3D generation results (each pair: left: without gravity; right: under gravity).

		F-score ↑	CD ↓
text	Trellis + Ours	0.9988	0.0081
single-view image	Trellis + Ours	0.9986	0.0056
	DPO	0.9517	0.0143
	DSO	0.7996	0.0336
multi-view images	Trellis + Ours	0.9995	0.0054
	DPO	0.9719	0.0100
	DSO	0.8137	0.0319

Table 3: Geometric fidelity of different generation methods.



Figure 6: Surface meshes and normal maps generated by Atlas3D and our method. Atlas3D is almost unable to produce a smooth and high-quality surface mesh.

Atlas3D	Trellis	DPO	DSO	Trellis + Ours
29min	24.3s	28.5s	28.8s	29.1s

Table 4: Wall-clock time of generating a 3D object.

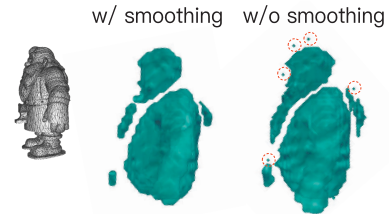


Figure 7: Ablation of density-field smoothing. Without smoothing (right), the optimized cavities contain isolated, tiny spaces (red circle).

the surface quality of Atlas3D’s outputs is poor, with noticeable roughness (Fig. 6). In contrast, Trellis generates high-quality objects with smooth surfaces, but faces challenges in stability (e.g., Fig. 4 (b,c)). Our method ensures both high generation quality and self-supporting structures. For the image-to-3D generation, we present results in Fig. 5 for both single-view and multi-view images. All baselines produce high-quality 3D models. Interestingly, DSO-generated objects tend to have significantly larger volumes compared to the others (Fig. 5 (a2, b1, b4)). In contrast, our method excels at preserving a lighter volume while still maintaining high stability, as shown by the improved stability and smaller volumes in our results. Moreover, when integrated into Trellis, DPO, and DSO, our method consistently enhances the upright stability of their generated objects.

Quantitative Comparison. As shown in Table 2, our

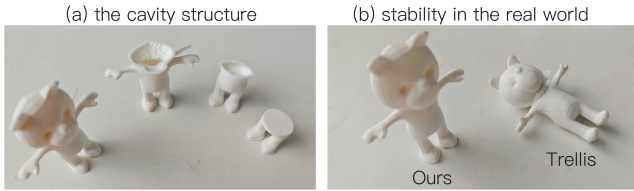


Figure 8: Real-world 3D print results: (a) the hollowed interior (b) reducing mass while maintaining upright stability under gravity, unlike solid Trellis outputs.

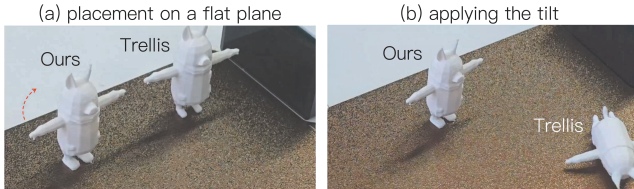


Figure 9: Stability under lean. To test robustness, we gradually lean the supporting platform.

method consistently improves both material efficiency and physical stability without degrading semantic consistency. For text-to-3D generation, our approach reduces the average mass by 43.3%, while increasing the stability success rate to 92.7% and significantly lowering the average tilt angle to 5.48°. For image-to-3D generation, our method improves over Trellis and further enhances DPO and DSO. On top of DSO, our optimization reduces mass by 28.4% while improving stability to 94.0% and reducing tilt by 2.6°. Our method introduces minimal distortion while enhancing functionality. Table 3 shows that our method (best F-Score and CD) achieves the highest geometric fidelity of the original Trellis outputs. Table 4 reports the wall-clock time to generate a 3D design. Atlas3D requires nearly 29 minutes, whereas Trellis and related methods (DPO, DSO) finish in under 30 seconds. Our density-field optimization adds only a few seconds of overhead beyond Trellis.

Ablation Study

We performed an ablation study under the multi-view images input setting.

Density Smoothing. As shown in Fig. 7, we visualize the inner surfaces of the carved cavities. With smoothing,

	Mas↓	Stable↑	Rot↓
Ours	0.035	87.3%	11.59°
w/o $\mathcal{L}_{\text{mass}}$	0.037	86.0%	13.17°
w/o B	0.032	74.7%	24.28°
w/o $\mathcal{L}_{\text{center}}$	0.035	86.0%	12.26°
w/o $\mathcal{L}_{\text{area}}$	0.033	76.7%	21.26°
w/o \mathcal{L}_z	0.030	85.3%	13.72°

Table 5: Ablation of different loss terms and regularization.

the hollow regions form large, continuous pockets; without smoothing, the inner surface contains small, isolated areas that are fragile and easily lost during fabrication.

Different Physical Constraints. Table 5 shows that each physics-based loss contributes uniquely. Removing $\mathcal{L}_{\text{center}}$ slightly degrades stability, and removing $\mathcal{L}_{\text{area}}$ causes a more pronounced drop in upright success. Interestingly, disabling \mathcal{L}_z yields marginally better mass reduction but at the expense of a higher average tilt angle, revealing a trade-off between material efficiency and static stability.

Regularization. As shown in Table 5, removing $\mathcal{L}_{\text{mass}}$ slightly increases mass and reduces stability, confirming its role in promoting lightweight, balanced designs. In contrast, removing B drastically degrades stability, highlighting the importance of reinforcing ground contact regions.

Real World 3D-Printing

We validated the manufacturability by printing several optimized models on a standard FDM printer (Fig. 3 (d), Fig. 8, Fig. 9). As shown in Fig. 8, our method automatically carved internal cavities and reduced mass, yet the printed prototype remained upright. In contrast, the solid Trellis model failed. To assess robustness, we placed the 3D-printings of both Trellis-generated and our generated structures upright on a flat platform and increased the lean angle (Fig. 9). Trellis model began to fail at relatively small inclinations, while our hollow design remained stable.

Discussion and Conclusion

We have presented DensiCrafter, a density-field optimization method that generates lightweight, self-supporting 3D structures by carving continuous internal cavities under fully differentiable, simulation-free physical constraints. Extensive evaluations and real-world 3D-printing experiments confirm that our method delivers significant material savings, enhanced upright stability, and reliable manufacturability without compromising visual fidelity and computational efficiency. Failure cases and corresponding analyses are provided in the appendix. By integrating the generative model with practical stability and fabrication requirements, DensiCrafter paves the way for deployable, physics-aware 3D content creation across text-to-3D and image-to-3D generation pipelines. Relative to Spin-it Faster (Hafner, Ly, and Wojtan 2024), which segments density fields via quadric surface fitting, DensiCrafter directly optimizes the continuous voxel field. This direct approach more naturally leverages Trellis’s structure latent representation.

While DensiCrafter produces lightweight, self-supporting hollow structures effectively, it assumes all objects behave as rigid bodies and does not model internal stress distributions or deformations under load. Extending our method to handle elastic or compliant materials would require mechanics-aware constraints or differentiable finite element methods. DensiCrafter may also inherit distortions from Trellis, and enhancing base model quality is valuable future work. Another promising future direction is to learn the prior of mass distribution from large-scale data, enabling the development of world models (Ding et al. 2025) that can reason about the real-world physical laws.

Acknowledgments

Nan Cao is the corresponding author. This work was sponsored by Natural Science Foundation of Shanghai (25ZR1401336).

References

- Bächer, M.; Whiting, E.; Bickel, B.; and Sorkine-Hornung, O. 2014. Spin-it: Optimizing moment of inertia for spinnable objects. *ACM Transactions on Graphics.*, 33(4): 1–10.
- Bendsøe, M. P. 1989. Optimal shape design as a material distribution problem. *Structural optimization*, 1: 193–202.
- Bendsøe, M. P.; and Sigmund, O. 1999. Material interpolation schemes in topology optimization. *Archive of applied mechanics*, 69: 635–654.
- Chen, Y.; Xie, T.; Zong, Z.; Li, X.; Gao, F.; Yang, Y.; Wu, Y. N.; and Jiang, C. 2024. Atlas3D: Physically Constrained Self-Supporting Text-to-3D for Simulation and Fabrication. In *Proceedings of Conference on Neural Information Processing Systems*.
- Deitke, M.; Liu, R.; Wallingford, M.; Ngo, H.; Michel, O.; Kusupati, A.; Fan, A.; Laforte, C.; Voleti, V.; Gadre, S. Y.; et al. 2023. Objaverse-xl: A universe of 10m+ 3d objects. *Advances in Neural Information Processing Systems*, 36: 35799–35813.
- Ding, J.; Zhang, Y.; Shang, Y.; Zhang, Y.; Zong, Z.; Feng, J.; Yuan, Y.; Su, H.; Li, N.; Sukiennik, N.; et al. 2025. Understanding world or predicting future? a comprehensive survey of world models. *ACM Computing Surveys*, 58(3): 1–38.
- Feng, Y.; Shang, Y.; Li, X.; Shao, T.; Jiang, C.; and Yang, Y. 2024. Pie-nerf: Physics-based interactive elastodynamics with nerf. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4450–4461.
- Guo, M.; Wang, B.; He, K.; and Matusik, W. 2025. TetSphere Splatting: Representing High-Quality Geometry with Lagrangian Volumetric Meshes. In *Proceedings of International Conference on Learning Representations*.
- Guo, M.; Wang, B.; Ma, P.; Zhang, T.; Owens, C. E.; Gan, C.; Tenenbaum, J. B.; He, K.; and Matusik, W. 2024. Physically Compatible 3D Object Modeling from a Single Image. *Advances in Neural Information Processing Systems*, 119260–119282.
- Hafner, C.; Ly, M.; and Wojtan, C. 2024. Spin-it faster: Quadrics solve all topology optimization problems that depend only on mass moments. *ACM Transactions on Graphics*, 43(4): 1–13.
- Hessel, J.; Holtzman, A.; Forbes, M.; Le Bras, R.; and Choi, Y. 2021. Clipscore: A reference-free evaluation metric for image captioning. In *Proceedings of the conference on empirical methods in natural language processing*, 7514–7528.
- Kang, Z.; and Wang, Y. 2011. Structural topology optimization based on non-local Shepard interpolation of density field. *Computer methods in applied mechanics and engineering*, 200(49-52): 3515–3525.
- Kerbl, B.; Kopanas, G.; Leimkühler, T.; and Drettakis, G. 2023. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics.*, 42(4): 139–1.
- Li, R.; Zheng, C.; Rupprecht, C.; and Vedaldi, A. 2025. DSO: Aligning 3D Generators with Simulation Feedback for Physical Soundness. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6772–6783.
- Li, X.; Qiao, Y.-L.; Chen, P. Y.; Jatavallabhula, K. M.; Lin, M.; Jiang, C.; and Gan, C. 2023. PAC-NeRF: Physics Augmented Continuum Neural Radiance Fields for Geometry-Agnostic System Identification. In *Proceedings of International Conference on Learning Representations*.
- Lorensen, W. E.; and Cline, H. E. 1987. Marching cubes: A high resolution 3D surface construction algorithm. *ACM SIGGRAPH Computer Graphics*, 163–169.
- Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99–106.
- Musialski, P.; Hafner, C.; Rist, F.; Birsak, M.; Wimmer, M.; and Kobbelt, L. 2016. Non-linear shape optimization using local subspace projections. *ACM Transactions on Graphics.*, 35(4): 1–13.
- Shu, D.; Cunningham, J.; Stump, G.; Miller, S. W.; Yukish, M. A.; Simpson, T. W.; and Tucker, C. S. 2020. 3D Design Using Generative Adversarial Networks and Physics-Based Validation. *Journal of Mechanical Design*, 142(7): 071701.
- Standley, T.; Sener, O.; Chen, D.; and Savarese, S. 2017. image2mass: Estimating the mass of an object from its image. In *Proceedings of Conference on Robot Learning*, 324–333. PMLR.
- Todorov, E.; Erez, T.; and Tassa, Y. 2012. Mujoco: A physics engine for model-based control. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 5026–5033.
- Wallace, B.; Dang, M.; Rafailov, R.; Zhou, L.; Lou, A.; Purushwalkam, S.; Ermon, S.; Xiong, C.; Joty, S.; and Naik, N. 2024. Diffusion model alignment using direct preference optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8228–8238.
- Wang, J.; and He, Y. 2019. Physics-aware 3d mesh synthesis. In *Proceedings of IEEE International Conference on 3D Vision (3DV)*, 502–512.
- Wu, R.; Xiao, C.; and Zheng, C. 2021. Deepcad: A deep generative network for computer-aided design models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6772–6782.
- Xiang, J.; Lv, Z.; Xu, S.; Deng, Y.; Wang, R.; Zhang, B.; Chen, D.; Tong, X.; and Yang, J. 2025. Structured 3D Latents for Scalable and Versatile 3D Generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 21469–21480.
- Xie, T.; Zong, Z.; Qiu, Y.; Li, X.; Feng, Y.; Yang, Y.; and Jiang, C. 2024. Physgaussian: Physics-integrated 3d gaussians for generative dynamics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4389–4398.

Zhai, A. J.; Shen, Y.; Chen, E. Y.; Wang, G. X.; Wang, X.; Wang, S.; Guan, K.; and Wang, S. 2024. Physical property understanding from language-embedded feature fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 28296–28305.

Zhang, L.; Wang, Z.; Zhang, Q.; Qiu, Q.; Pang, A.; Jiang, H.; Yang, W.; Xu, L.; and Yu, J. 2024. CLAY: A Controllable Large-scale Generative Model for Creating High-quality 3D Assets. *ACM Transactions on Graphics.*, 43(4): 1–20.

Zhao, Z.; Lai, Z.; Lin, Q.; Zhao, Y.; Liu, H.; et al. 2025. Hunyuan3D 2.0: Scaling Diffusion Models for High Resolution Textured 3D Assets Generation. arXiv:2501.12202.

Zhong, L.; Yu, H.-X.; Wu, J.; and Li, Y. 2024. Reconstruction and simulation of elastic objects with spring-mass 3d gaussians. In *Proceedings of European Conference on Computer Vision*, 407–423.