This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Artiverse: A Diverse and Physically Grounded Dataset for Articulated Objects

Denys Iliash[1] Jiayi Liu[1] Egor Fokin[1] Qirui Wu[1]
Ali Mahdavi-Amiri[1] Manolis Savva[1] Angel X. Chang[1,2]
[1]Simon Fraser University [2]Canada-CIFAR AI Chair, Amii
3dlg-hcvc.github.io/artiverse/

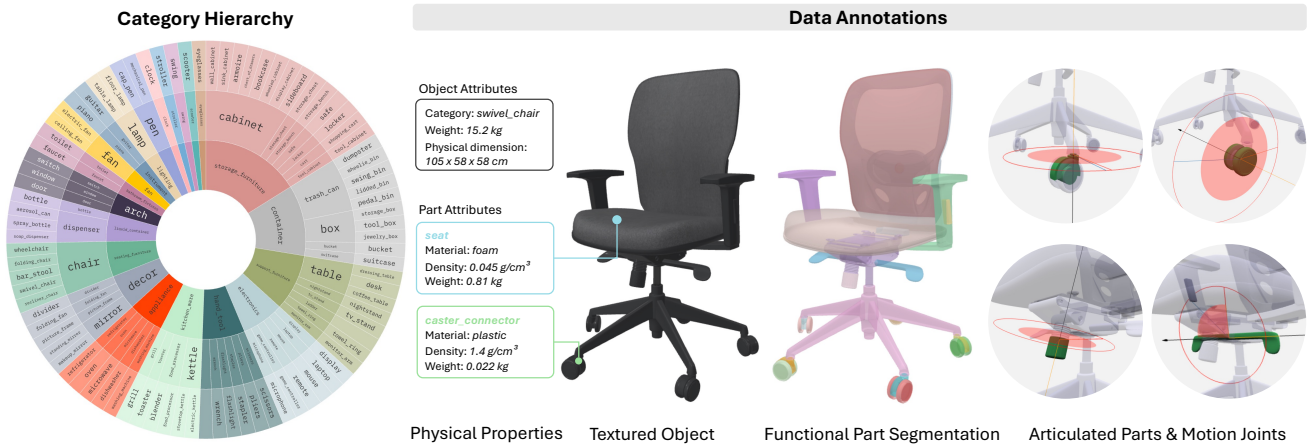


Figure 1. We introduce Artiverse, a physically grounded 3D dataset of articulated objects across diverse categories, richly annotated with functional parts, motion joints for articulation, and physical properties including metric scale, per-part material, and mass.

Abstract

We present Artiverse, a diverse and physically grounded dataset of high-quality articulated 3D objects designed for realistic functional modeling and simulation. Artiverse contains 5.4K human-authored objects across a broad range of 88 categories, aggregated from multiple 3D static repositories. Objects are annotated with functional parts, interior structures, realistic kinematic relationships and articulated joints including multi-DoF joints, and physical attributes such as metric scale, material, and mass. We develop a semi-automated annotation pipeline that combines few-shot segmentation, geometric reasoning, and multi-stage human verification to achieve high-quality and efficient annotation, reducing manual annotation time by over 30%. We demonstrate the value of Artiverse on tasks of part mobility analysis, articulated object generation, and physics-based interaction. Artiverse provides a data resource to advance functional understanding for articulated objects.

1. Introduction

Objects are not merely static shapes. They are functional entities that serve a purpose in the world. A tabletop provides support, a cup holds liquids, and a hook enables hanging items. Beyond these general purposes implied by geometry, many everyday objects express their functionality through interaction: a drawer slides open to provide storage, a door swings to allow passage, and a faucet handle rotates to release water. Understanding these objects requires modeling their *functional interactivity*, which captures how objects are organized into movable parts, regions for interaction, and exhibit physically plausible behavior.

The notion of functional interactivity can be characterized along three complementary dimensions. *Functional parts* capture the structural decomposition of objects into components that contribute to their function, encompassing both movable and interactable elements. *Kinematic relationships* describe how articulated parts connect and move under joint constraints, and functional dependencies that enable coupled behaviors (e.g., open the microwave door by pushing a button). *Physical grounding and realism*, such

as metric scale, mass and material, as well as detailed geometry and textures, ensure that articulated motions behave plausibly in simulation, faithfully reflecting the behavior of objects in the physical world.

Despite substantial progress in 3D dataset construction, existing resources fall short in capturing this level of functional realism. Large-scale object repositories [5, 9, 10] provide diverse geometry but mostly contain static shapes. Articulated object datasets [34, 53] annotate part mobility but are limited in joint complexity (e.g., mostly single-DoF motions), diversity in part structure and category, and visual realism (e.g., simplified texture and missing interior geometries). Physical properties are rarely modeled until recently, with ArtVIP [21] releasing physical digital replicas for 206 articulated objects, and PhysX-3D [4] annotating physical attributes on PartNet mostly for rigid objects. Thus, there are limited assets with sufficient geometric quality, diversity, and physical realism for studying functionally complete articulated objects at scale.

To bridge this gap, we introduce **Artiverse**, a dataset of 5,402 human-authored articulated objects across 88 diverse categories, richly annotated with functional parts, articulations, and physical attributes. Objects are carefully filtered from multiple repositories (see Table 1). Artiverse advances articulation modeling by supporting complex motion joints (e.g, cylindrical, universal joints) and modeling both kinematic and functional dependencies among parts and joints, enabling the study of chained and interdependent motions. To ensure functional realism, we complete interior structures, scale objects to metric units, and estimate per-part mass. All assets are standardized in orientation and released in consistent formats (e.g., GLB and USD), making them readily usable in simulation environments.

Building such a dataset at scale requires addressing the high annotation cost and the expertise needed for 3D geometry and kinematics. We therefore design a semi-automated annotation pipeline integrating few-shot segmentation models and geometric reasoning to initialize proposals for segmentation, motion parameters, and physical attributes. Automated proposals are validated and corrected by human annotators in multiple stages throughout the pipeline, ensuring consistent and high-quality annotations, as illustrated in Figure 2. In practice, our pipeline reduces manual annotation time by 32.0% and 33.5% for segmentation and motion annotation respectively, with 50.12% of the automatic annotations free of human correction, demonstrating efficiency and scalability.

Through its combination of kinematic complexity, geometric diversity, and physical grounding, Artiverse enables a broad range of research directions. It offers reusable assets to build interactive scenes and immersive virtual environments for everyday activity understanding or interactive design. It provides a rich resource for learning-based methods to advance functional understanding of articulated objects. It supports generative modeling of articulated and physically grounded objects by providing denser coverage of the data distribution. It can be used to train and evaluate embodied agents in physics-based simulators to facilitate policy learning for interactive tasks such as object manipulation. In this paper, we illustrate some of these possibilities through benchmarks for part mobility analysis, experiments on generative modeling of articulated objects, and demonstrations of realistic interactions in simulation. In summary, we make the following contributions:

- We release Artiverse, a diverse 3D dataset of human-authored articulated objects, sourced from multiple high-quality shape repositories, and richly annotated with functional parts, articulation, and physical properties.
- We design a semi-automated annotation pipeline combining learning-based and geometric algorithms with human verification to efficiently produce high-quality annotations at scale.
- We show the value of Artiverse with benchmarks for part mobility analysis and generation of articulated objects, and demonstrations of realistic interactions in simulation.

2. Related work

Simulatable articulated assets. Existing articulated object datasets vary widely in geometry type, data source, annotation granularity, and scale, as summarized in Table 1. Real scan datasets [34, 35, 39] reflect the realism of real-world objects but suffer from reconstruction noise and limited scalability due to the per-part capture effort. More datasets are built on synthetic objects [15, 47, 56]. PartNet-Mobility [53] remains the most widely used. ACD [19] introduces more detailed objects and annotations but is limited to openable objects. These datasets primarily focus on part mobility labels and overlook realistic textures or physical attributes. Recent efforts begin to incorporate physical properties but remain limited in scale and coverage of articulated objects, with ArtVIP [21] releasing 206 articulated objects, and PhysX-3D [4] mostly focus on rigid objects. Building on these efforts, Artiverse contributes a comprehensive collection of articulated assets with broader category diversity, richer part structures, more complex motion patterns, and per-part materials and masses.

Data annotation pipelines. Annotating articulated assets is highly labor-intensive, requiring expertise to ensure accurate and consistent motion specification. Challenges include identifying movable part boundaries, defining plausible motion joints and ranges, and specifying coherent kinematic hierarchies. Most existing efforts rely on fully manual annotation [15, 21, 34, 47, 53], which limits scalability to diverse objects and categories. Recently, PhysX-3D [4] proposed a human-in-the-loop pipeline that uses vision-language models (VLMs) to infer physical attributes from

Dataset	Type	Data Source	Tex	Kin	Fun	Mat	Met	Mas	# Obj	# Part	# Cat
RBO [39]	Scan	Self-scanned	✓	✗	✗	✗	✓	✗	14	21	14
ReArt-48 [35]	Scan	Self-scanned	✓	✗	✗	✗	✓	✗	48	-	5
AKB-48 [34]	Scan	Self-scanned	✓	✗	✗	✗	✓	✓	2,037	-	48
RPMNet [56]	Synth	ShapeNet [5], 3D Warehouse	✗	✗	✗	✗	✗	✗	969	1,420	43
Shape2Motion [47]	Synth	ShapeNet [5], 3D Warehouse	✗	✗	✗	✗	✗	✗	2,440	6,762	45
PartNet-Mobility [53]	Synth	PartNet [40], 3D Warehouse	✓	✓	✗	✗	✗	✗	2,346	11,753	46
GAPartNet [15]	Both	PartNet-Mobility [47], AKB-48 [34]	✓	✗	✗	✗	✗	✗	1,166	8,489	27
ACD [19]	Synth	ABO [6], 3D-Future [13], HSSD [24]	✓	✓	✗	✗	✗	✗	354	1,350	21
ArtVIP [21]	Synth	Artist-created digital twins	✓	✓	✓*	✗	✓	✓	205	705	29
Artiverse (ours)	Synth	ShapeNet [5], ABO [6], 3D-Future [13], HSSD [24], Objaverse [9, 10], TexVerse [58], 3DComPaT200 [2], PartNeXt [46], SSDB [12]	✓	✓	✓	✓	✓	✓	**5,402**	**24,607**	**88**

Table 1. Comparison of existing datasets containing 3D articulated objects. Columns indicate: **Tex**tured objects, **Kin**ematic and **Fun**ctional dependencies, per-part **Mat**erials, **Met**ric scale, per-part **Mas**ses, and numbers of articulated **obj**ects, **part**s (before completion), and **cat**egories. "-" indicates information not reported in the original paper, and "*" means partially available. Artiverse offers a comprehensive combination of functional part annotations, kinematic modeling, and physical attributes across diverse data sources and categories at scale.

2D part renderings based on the fine-grained PartNet [40] hierarchy. Without access to 3D geometry, purely 2D VLM-based annotation can be unreliable for tasks requiring spatial precision, such as segmenting parts on meshes and specifying joint axes for articulation. To improve the accuracy and robustness of automatic proposals, we integrate multi-view few-shot models and geometry-based heuristics beyond pure VLMs to exploit the full 3D structure of each model. This design allows our system to assist human annotators with minimal intervention while maintaining high-quality, consistent annotations at scale.

Articulated object modeling and applications. Articulated assets support a spectrum of research directions by providing realistic priors for learning, standardized benchmarks for evaluation, and interactive components for scene generation [52, 57], simulation [23, 38], and embodied AI tasks [42, 48]. Approaches for part mobility analysis focus on understanding the movable part structures and their articulation patterns from static 3D shapes [16, 19, 47, 56] or 2D observations [20, 44], but achieving robustness across diverse objects remains challenging. Reconstruction methods recover 3D articulated models from multi-view and/or multi-state images [11, 17, 31, 37, 49, 50], though handling occlusion and motion ambiguity remains an open problem. Generative methods learn structural and motion prior from 3D [14, 26, 29] or metadata [27, 32, 33, 51] to synthesize interactive parts, but are often constrained by limited diversity and realism of available training data. Procedural methods synthesize articulated objects through parametric or functional recombination of parts [18, 22, 30, 43], though how to ensure the functional validity and consistency with real data distribution remains non-trivial. These advances highlight the growing demand for large-scale, realistic, and physically-grounded articulated datasets.

3. Data Annotation

3.1. Design Desiderata

The construction of Artiverse is guided by the following considerations shaping our data selection and annotation.

Object coverage. Artiverse focuses on human-made articulated objects for daily activities, such as furniture, tools, and appliances. We prioritize categories appearing frequently in human environments (e.g., household, office, and outdoor) to ensure relevance for real-world applications. We organize the objects in three levels of semantic categories: super-category, main category, and sub-category. Super-categories are broad and general (e.g., appliance, lighting, kitchenware). Main categories are specific to the object function (e.g., toaster, window, dispenser). Sub-categories capture fine-grained distinctions in the usage context (e.g, office chair vs. bar stool).

Functional part specification. We define a unified taxonomy of functional part labels that capture both articulated (e.g., levers, lids) and interaction-related (e.g., handles, backrest, shelves, racks) components. These labels are designed to be category-consistent and reflect the functional roles of parts in enabling articulation and interaction.

Annotation consistency. To maintain coherent semantics across the dataset, we use category-level templates that specify part labels, motion types and dependencies for each object family. The templates are defined by human experts and are flexible to accommodate extra information discovered by annotators. These templates guide both automated inference and human verification, ensuring consistent labeling and complete articulation coverage.

3.2. Semi-automated Workflow

We design a semi-automated annotation workflow, which includes the steps of model selection and preprocessing, part segmentation, interior completion, motion annotation, physical property estimation, and multi-stage human verifi-

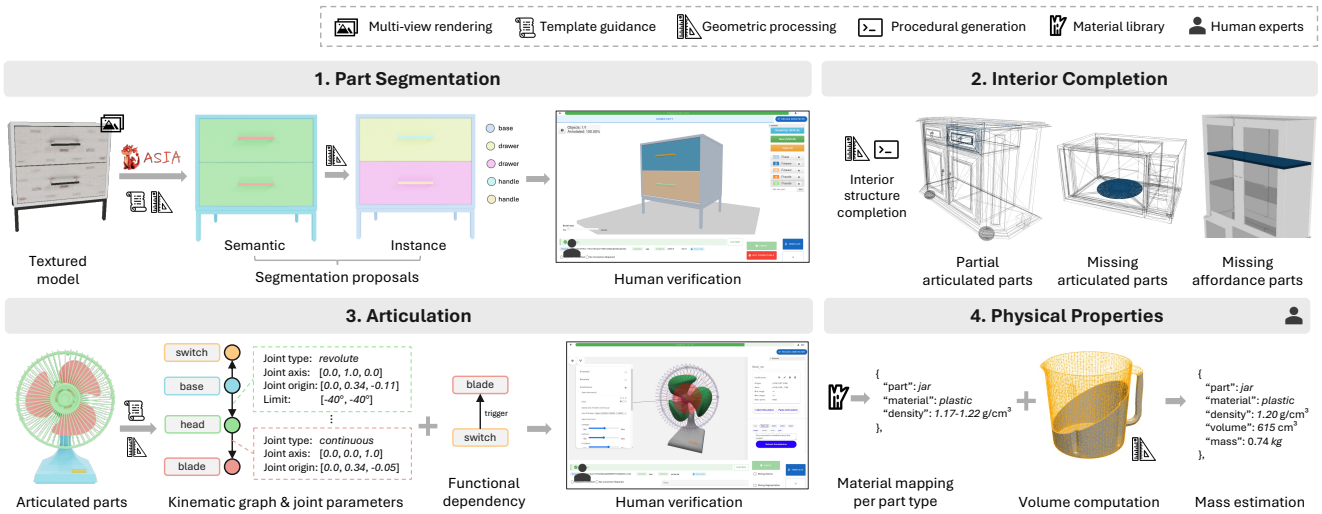


Figure 2. Overview of the key modules in our semi-automated annotation framework. Each selected object progresses through four stages sequentially after preprocessing: 1) functional part segmentation, 2) optional interior completion for objects with missing internal structures, 3) annotation of articulation mechanisms, and 4) estimation of physical properties. Human verification is incorporated at the end of the segmentation and articulation stages to correct errors and ensure that accurate annotations are passed to subsequent modules. After a final verification at the end of the pipeline, the output is a fully annotated 3D articulated object ready for simulation.

cation in between. The key steps of the pipeline are illustrated in Figure 2. In the following sections, we discuss how we design each key module. Please refer to the supplement for more implementation details.

3.2.1. Part Segmentation

Each object in Artiverse is decomposed into parts that are associated to an articulation hierarchy and interaction affordance. For every object category, human experts design a template listing possible part labels shared across instances in the category. These templates may be refined during the human verification stage if additional parts are discovered. Leveraging the template, our part annotation strategy follows a two-stage process: we first generate initial part proposals using learning-based models and geometric algorithms, and then refine them through human verification.

Initial proposals. Our part labels are designed to be semantically meaningful. However, unlike general semantic segmentation, part boundaries are determined by functional behavior rather than purely visual or semantic similarity, leading to subtle but critical distinctions. Therefore, it is essential to employ a segmentation model that can incorporate contextual definitions of each label, making few-shot models [25, 36, 41, 45] a better fit than non-adaptable 3D segmentors [1, 7, 8, 59]. We opt for the state-of-the-art model ASIA [41] in our pipeline, as it supports label customization in a few-shot manner via visual exemplars. We train ASIA on groups of categories that share similar part structures and labels to minimize training redundancy. We manually annotate a dozen representative shapes in 3D and render part masks from multiple views to serve as training data. Once trained, ASIA produces label-aware part masks on multi-view images, reflecting our functional context.

Map proposals to 3D part instances. To project 2D segmentation results onto 3D meshes, we first precompute an oversegmentation based on the mesh's topological connectivity. Each over-segment aggregates label votes from multi-view projections according to the number of covered pixels. Unlabeled internal surfaces are assigned part labels using a distance-based propagation that considers the physical proximity of sampled surface points. Once semantic segmentation converges, part instances are decomposed from each semantic group. We use a union-find algorithm to obtain disjoint connected sets of mesh segments, each forming an independent part instance, determined by local geometric continuity and spatial distance.

Human verification. At the end of the segmentation stage, human verification checks the accuracy of part instances and labels proposed by the automated process above. We develop a web interface for annotators to check whether corrections are required, and correct any mistakes with a 3D part annotation UI, refining part labels if needed (see the example interface in Figure 2). As the last step, human experts verify the annotations from each annotator to ensure clean and consistent part segmentation.

3.2.2. Interior Completion

For some categories such as appliances and furniture, the interior serves as a critical structure enabling the intended functionality. However, the vast majority of objects in static datasets lack sufficiently detailed interior geometry. To provide assets that behave realistically in physical simulation,

we complete the interiors of these objects where necessary. We focus on three types of interiors to complete.

Partial articulated parts. Some articulated parts are partially modeled, with interior geometry omitted because it is not visible when the part is in the resting state. This issue is typical for components exhibiting translational motion, where hidden regions are revealed only during articulation. For example, many drawers are modeled with only the front panel, while the inner box is missing. We design a geometry-informed completion algorithm for such parts, extending the geometry consistently with the motion constraints to ensure plausibility in full motion range.

Missing articulated parts. Certain appliances contain movable internal components that are entirely absent in many 3D assets. Examples include dishwashers with baskets for dishes, microwaves with rotating turntables, and refrigerators with movable storage compartments. To address this, we implement a procedural algorithm that populates these missing elements by referencing similar parts from other objects within the same category, ensuring functional completeness and structural realism.

Missing interaction affordance parts. Furniture and appliances that provide storage space are often modeled as hollow shells, resulting in unrealistic behavior when interacting with other objects. To restore functional realism, we add interior structures such as shelves, racks, and dividers that correspond to the intended use of the object, e.g., refrigerator shelves for food placement or wardrobe racks for hanging clothes. These interior completions are generated procedurally following category-specific templates that ensure geometric consistency and plausible spatial layout.

3.2.3. Motion Annotation

At this stage, models have verified part segmentations and complete geometry, including interiors. We follow a similar strategy as in part segmentation, where automated motion proposals are first generated and later verified by human experts to ensure accuracy and physical plausibility.

Initial proposal. We design algorithms to predict the joint parameters, including joint type, motion axis, and joint limits. In particular, we establish general rules that use PCA-aligned oriented bounding boxes for each part and sampled points around contact regions to infer motion parameters for different motion types, constrained by geometric contact and collision relationships. Kinematic dependencies between articulated parts are then inferred based on spatial connectivity and dependency options defined in the per-category templates. The inference rules are reusable across categories and are registered for different part types, requiring only minor adjustments for specialized components.

Human verification. In the motion verification stage, human annotators validate two key aspects: 1) whether the completed interior geometry is realistic; and 2) whether the proposed motions are physically plausible. We develop a web-based interface that enables annotators to visualize and adjust articulations interactively, modify joint parameters, and edit kinematic graphs as needed (see Figure 2 for an example page). To improve efficiency, the interface supports features such as motion copying between similar parts and instant motion previews, allowing rapid correction and validation of proposed motion parameters.



Figure 3. Visual comparison of assets from Artiverse and PartNet-Mobility [53]. We manually assemble sub-scenes using assets from both datasets with the original object scales preserved. Artiverse objects exhibit higher geometric fidelity, realistic textures, and consistent metric scale, resulting in assets that are visually richer and more suitable for realistic simulation and scene design.

3.2.4. Physical Properties

To support physics-based simulation, we annotate objects with metric scale, and parts with materials and masses.

Metric scale. The metric scale is either obtained from the original data source or estimated during preprocessing. We ask LLMs to infer a reasonable range of physical dimensions for each sub-category, and then sample within the range to rescale the object accordingly. It is critical to estimate metric scale in early stages of the pipeline, as our geometric algorithms for part instance decomposition and interior completion rely on this for robustness.

Per-part mass. Each part's mass is estimated as the product of its approximated volume and the density sampled within the range of an estimated material. We predefine a list of material labels and query LLMs for a reasonable density range for each material. We assign each part label a default material label as the initial proposal, and later verify it manually in the end. Accurate volume estimation is non-trivial for surface meshes, as our data is mixed with solid parts enclosed by shells and hollow parts lacking modeled thickness. Thus, we design different processing strategies for hollow and solid parts, guided by their semantic labels. For solid parts, we extract a volumetric mesh by tetrahderalization and compute the volume directly. For hollow parts, we offset the surface inwards along the normal to create a shell, and then tetrahedralize to approximate the volume.

Dataset	# obj	Category		# Func. Parts		# Arti. Parts		# Joints	
		Total	Avg # obj	Total	Avg	Total	Avg	1-DoF	2-DoF
PM	2,346	46	**51.0**	14,100	6.0	11,753	5.0	11,753	0
ArtVIP	205	29	7.1	1,784	8.8	705	3.4	705	0
Artiverse	**5,402**	**88**	49.4	**38,608**	**8.9**	**24,607**	**5.4**	**23,640**	**480**

Table 2. Comparison of articulation annotation statistics. Artiverse exhibits broader category coverage, richer functional part structure, and more complex articulation on average.

4. Artiverse

4.1. Data Statistics

Artiverse contains 5,402 human-made articulated objects sourced from 10 repositories, covering 20 super-categories, 61 main categories, and 88 sub-categories, encompassing a wide range of furniture, appliances, tools, and everyday household objects. Figure 4 summarizes the distribution of data sources, object counts per super-category, and different joint types. Table 2 reports a detailed breakdown of our annotation statistics in comparison with two prior datasets: PartNet-Mobility [53] (PM) and ArtVIP [21]. Artiverse provides broader category coverage with more objects overall, more functional parts, and more varied and complex articulations per object. Together, these statistics highlight the diversity and functional complexity captured in Artiverse, which can benefit downstream applications. In Figure 3, we qualitatively compare our assets with PM by manually assembling sub-scenes, showing that Artiverse provides data with higher geometric fidelity, realistic textures, and consistent metric scale, resulting in assets that are visually richer and more suitable for realistic simulation and scene design.

4.2. Data Schema

Each object in Artiverse is represented as a textured 3D mesh enriched with a hierarchical annotation schema that captures its structural, functional, and physical properties. Annotations are organized in two levels: (1) the *object level*, which stores global metadata such as unique identifier, category labels, dataset source, and metric scale; (2) the *part level*, which encodes articulated and interaction-related components, their semantic labels, motion parameters, material properties including label and density, and mass.

All assets are released in multiple interoperable formats. The textured geometry is provided in glTF binary (GLB) for visualization and rendering, Universal Scene Description (USD) and URDF for simulation compatibility, and JSON metadata for easy parsing and programmatic analysis. Each object has a standardized orientation, consistent scale, and unified coordinate convention, enabling direct integration into physics simulators. Section 5.3 showcases the usage of Artiverse assets in simulators.

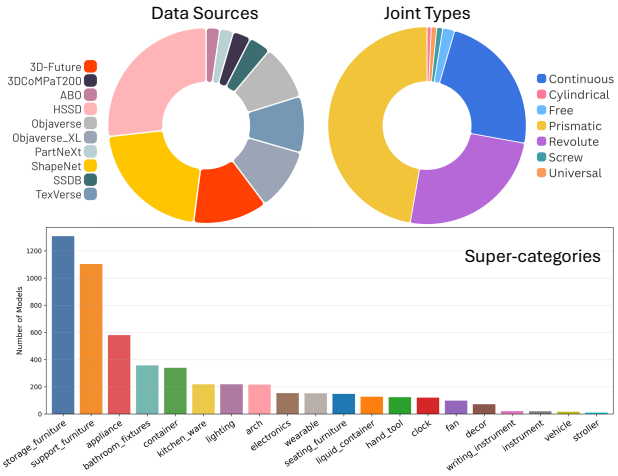


Figure 4. We visualize the distribution of objects across super-categories and across data sources, and the frequency of different joint types. Artiverse spans a broad range of object families with balanced coverage and aggregates assets from multiple repositories to increase model and category diversity. The distribution of joint types highlights the presence of common 1-DoF and complex multi-DoF joints (e.g., universal), reflecting motion diversity.

4.3. Annotation Efficiency

We evaluate the efficiency of our semi-automated annotation pipeline by comparing it against a fully manual annotation process for the two main annotation stages: part segmentation and motion annotation. For each category, we sample 5 representative objects and record the time used by annotators to complete each stage manually, and compare it with the human-correction time logged through our annotation interface. Using this protocol, our automatic proposals can save 32.0% and 33.5% of the manual annotation time for segmentation and motion annotation, respectively. In our semi-automated pipeline, the average human correction time is reduced to 1.5 minutes for part segmentation and 1.3 minutes for motion annotation. We also record that 50.12% of the parts on average do not require human adjustment afterwards. To ensure consistency across annotators, an expert verification pass is performed on the corrected results, which adds an additional 0.8 minutes per object on average.

These measurements show that our semi-automated pipeline achieves a substantial reduction in manual annotation effort, making it feasible to annotate thousands of objects while maintaining high-quality and consistent outputs. The modular design of the pipeline also allows the integration of improved automated components in the future, enabling gains in scalability and annotation reliability.

Eval	Train	Segmentation			Motion		
		P↑	R↑	F1↑	+M↑	+MA↑	+MAO↑
PM	PM	81.8	46.0	54.2	22.1	**17.0**	**14.3**
	Artiverse	**82.2**	**47.9**	**55.8**	**22.4**	15.9	8.7
Artiverse	PM	72.8	31.7	40.6	7.2	2.5	1.2
	Artiverse	**77.0**	**43.0**	**50.7**	**22.4**	**16.6**	**10.8**

Table 3. We benchmark the task of part mobility analysis with a recent method FPNGroupMot [19] in a cross-dataset evaluation between Artiverse and PartNet-Mobility (PM), training on each dataset respectively. The results show that training on Artiverse consistently outperforms training on PM in both segmentation and motion prediction metrics. This demonstrates that the superior diversity and quality of Artiverse enhances model generalization.

Eval	Methods	Reconstruction quality ↓				Collision
		RS-d_{gIoU}	AS-d_{gIoU}	RS-d_{cDist}	AS-d_{cDist} ↓	AOR↓
PM	SG [32]	**0.756**	**0.768**	**0.178**	**0.229**	**0.022**
	AA [27]	1.172	1.179	0.425	0.780	0.024
Artiverse	SG [32]	**0.810**	**0.822**	**0.105**	**0.158**	**0.009**
	AA [27]	1.250	1.258	0.385	0.856	0.042

Table 4. We benchmark the task of image-conditioned articulated object generation with Articulate-Anything [27] that purely relies on VLMs, and an adapted version of SINGAPO [32] re-trained on Artiverse. Across both evaluation sets, SINGAPO achieves consistently higher performance, indicating the limitations of VLM priors for modeling 3D articulated structures and highlighting the benefit of Artiverse data in distribution modeling.

5. Experiments

To demonstrate the utility and versatility of Artiverse, we evaluate its effectiveness across multiple downstream applications that rely on 3D articulated objects. We focus on two representative benchmark tasks: *part mobility analysis* and *articulated object generation*. These tasks capture complementary aspects of reasoning and synthesis that both benefit from dataset scale and diversity. Part mobility analysis (Section 5.1) assesses how well a method can infer motion types, joint parameters, and articulated structures from geometric cues, reflecting the informativeness and precision of our articulation annotations. Articulated object generation (Section 5.2) evaluates the ability to synthesize functionally consistent articulated shapes, highlighting the dataset's diversity and annotation completeness. Finally, we demonstrate the simulation readiness of Artiverse assets through physics-based simulation (Section 5.3), illustrating their usability for embodied AI and interactive environments.

5.1. Part Mobility Analysis

Task definition. This task measures both geometric understanding and the ability to recover motion patterns from static geometry. Given a static 3D shape, the goal is to decompose the object into articulated part segmentation and predict the joint parameters associated with each part.

Setup. We use a recent method FPNGroupMot proposed in S2O [19] to benchmark the task performance trained on Artiverse against the prior dataset PartNet-Mobility [53] (PM). The models take as input a point cloud sampled from the object surface and output instance segmentation of articulated parts and their corresponding motion parameters, including motion type, joint axis, and origin. We split both datasets into train/val/test sets with an 80/10/10 ratio. The model is trained on each dataset respectively, and evaluated on both test sets to show the generalization capability.

Metrics. We break down the task evaluation into articulated part segmentation and motion prediction. We follow the evaluation metrics used in S2O. Specifically, we use precision (P), recall (R), and F1 score as segmentation metrics with area-based instance matching macro-averaged over the object. For motion estimation, we report macro F1 scores with matches further refined by motion type (+M), motion type and axis (+MA) and origin (+MAO).

Discussion. The quantitative results are in Table 3. When benchmarking with PM, both methods achieve significantly better performance when trained on Artiverse than when trained on PM, demonstrating that the diversity and comprehensive annotations of our dataset enhance generalization of existing models. However, when evaluated on the Artiverse test set, performance drops noticeably compared to prior benchmarks, indicating that the higher geometric fidelity, richer articulation patterns, and complex multi-DoF or dependent joints in Artiverse present new challenges. This suggests that while current approaches can exploit larger data for improved learning, they still struggle to reason over more complex and variable geometric and motion structures. We believe Artiverse therefore not only serves as a stronger training resource but also establishes a more demanding benchmark that exposes the limitations of existing methods and motivates the development of models capable of unified reasoning over articulation behaviors.

5.2. Articulated Object Generation

Task definition. The goal of articulated object generation is to synthesize new 3D shapes that are both geometrically realistic and functionally coherent, producing plausible articulated parts, motion joints, and their kinematic dependencies. We benchmark this task under a conditional generation setting, where the model receives a single-view image of an object as input and generates corresponding articulated shapes that are structurally and geometrically consistent with the input. This setup provides a strong test of generalization, as it requires the model to infer hidden geometry, articulation, and motion behavior in a constrained setting from limited visual information, particularly for in-the-wild or previously unseen inputs.

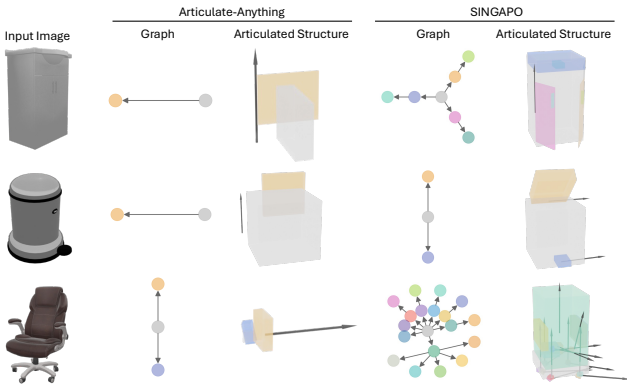


Figure 5. Qualitative comparison on the task of image-conditioned generation of articulated objects by showing the generated kinematic graph and structures. SINGAPO trained on Artiverse outperforms Articulate-Anything that relies on the VLM knowledge only across different categories.

Setup. We benchmark two recent open-sourced methods on the image-conditioned generation task: SINGAPO [32] and Articulate-Anything [27]. SINGAPO learns high-level articulation structure from training data, while Articulate-Anything leverages VLMs and retrieval of objects and parts from a database. We report SINGAPO performance trained on Artiverse against PM to show the effect of the more diverse data in Artiverse. Articulate-Anything serves as a complementary baseline that shows how well a VLM prior can generalize articulation understanding without fine-tuning on annotated data. The original SINGAPO method leverages a diffusion model to generate part layout, kinematic structure, and part labels to retrieve part meshes from the library to produce the final 3D geometry. We retrain SINGAPO on the full Artiverse, excluding objects with more than 32 parts, and evaluate on PM and our test split.

Metrics. We follow the evaluation setup from SINGAPO to evaluate the structural and geometric consistency of generated objects against ground truth data. The metrics are defined using 3D bounding box IoU (gIoU) parts at resting state (RS-) or across multiple articulated states (AS-) averaged over objects. Collision rate during articulation is measured using the averaged overlap ratio (AOR).

Discussion. The qualitative results are shown in Figure 5 and quantitative results are shown in Section 5.1. We observe that the SINGAPO model trained on Artiverse generalizes well overall across the test set. From the metrics of AS-gIoU and AOR, we see that Articulate-Anything still lags behind SINGAPO trained on Artiverse, indicating the limitation of VLM priors in capturing fine-grained articulation details without direct supervision. Overall, these results suggest that Artiverse serves as a valuable resource for training generative models that can produce functionally plausible articulated objects, and that learning from structured data priors remains crucial for detailed articulation synthesis.

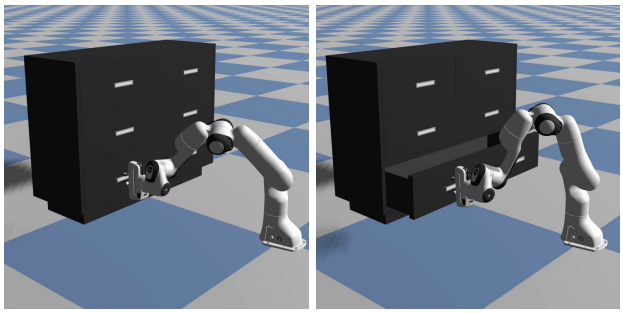


Figure 6. Loading an example Artiverse asset into Genesis [3] to show realistic and accurate joint modeling.

5.3. Artiverse in Simulation

Artiverse provides simulation-ready articulated assets in both URDF and USD formats, enabling seamless integration with a wide range of physics engines and robotics simulators. To demonstrate the practical simulation readiness, physical realism, and articulation correctness of our assets, we load them directly into the Genesis [3] and train a policy to interact with the example assets. Figure 6 illustrates an interaction with a cabinet from Artiverse and successfully opens one of its doors, showcasing accurate joint modeling.

6. Conclusion

We presented Artiverse, a comprehensive dataset of physically grounded and articulated 3D objects. It collects human-made everyday objects across diverse categories, each annotated with detailed functional part structure, articulation mechanisms, and physical attributes such as material, mass, and metric scale. To construct the dataset efficiently, we developed a semi-automated annotation pipeline that integrates geometric reasoning, few-shot segmentation, and human verification in multiple stages. This pipeline provides a scalable and flexible framework for producing high-quality annotations across heterogeneous 3D assets. Through benchmarks on part mobility analysis and articulated object generation, we show that Artiverse improves the robustness and generalization of existing approaches while also revealing new challenges posed by complex multi-DoF joints and intricate kinematic relationships. We hope Artiverse will serve as a valuable resource for future research in functional 3D modeling, generative object design, and physically grounded reasoning, ultimately enabling richer benchmarks and inspiring new directions in articulable object understanding. Looking ahead, we plan to steadily grow Artiverse in scale, diversity, and physical realism to support an ever wider range of applications.

Acknowledgments

This work was funded in part by a CIFAR AI Chair, a Canada Research Chair, and NSERC Discovery grants, and enabled by support from the Digital Research Alliance of Canada. We thank Weikun Peng, Han-Hung Lee, and Tristan Engst for helpful discussions, and Yiming Zhang for web interface support. We thank our annotators for their dedication in ensuring the data quality: Arya Faghihy, Austin Wang, Daniel Chang, Derek Pun, Dingdong Yang, Dongchen Yang, George Katsadze, Han-Hung Lee, Ivan Tam, Jenna Lee, Jill Ference, Jiyeon Han, Junghun Byun, Max Vernikovskiy, Mingrui Zhao, Nina Dang, Qirui Wu, Saba Azzarouk, Simon Jin, Sonia Raychaudhuri, Stella Lin, Tommaso Galliena, Weikun Peng, Xiaohao Sun, Xingguang Yan, Yuefan Wu.

References

[1] Ahmed Abdelreheem, Ivan Skorokhodov, Maks Ovsjanikov, and Peter Wonka. SATR: Zero-shot semantic segmentation of 3D shapes. In *ICCV*, pages 15166–15179, 2023. 4

[2] Mahmoud Ahmed, Xiang Li, Arpit Prajapati, and Mohamed Elhoseiny. 3DCoMPaT200: Language-grounded compositional understanding of parts and materials of 3d shapes. *arXiv preprint arXiv:2501.06785*, 2025. 3, 1

[3] Genesis Authors. Genesis: A generative and universal physics engine for robotics and beyond, 2024. 8

[4] Ziang Cao, Zhaoxi Chen, Linag Pan, and Ziwei Liu. PhysX-3D: Physical-grounded 3D asset generation. *arXiv preprint arXiv:2507.12465*, 2025. 2

[5] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3D model repository. *arXiv preprint arXiv:1512.03012*, 2015. 2, 3

[6] Jasmine Collins, Shubham Goel, Kenan Deng, Achleshwar Luthra, Leon Xu, Erhan Gundogdu, Xi Zhang, Tomas F Yago Vicente, Thomas Dideriksen, Himanshu Arora, et al. ABO: Dataset and benchmarks for real-world 3D object understanding. In *CVPR*, pages 21126–21136, 2022. 3

[7] Dale Decatur, Itai Lang, and Rana Hanocka. 3D highlighter: Localizing regions on 3D shapes via text descriptions. In *CVPR*, pages 20930–20939, 2023. 4

[8] Dale Decatur, Itai Lang, Kfir Aberman, and Rana Hanocka. 3D paintbrush: Local stylization of 3D shapes with cascaded score distillation. In *CVPR*, pages 4473–4483, 2024. 4

[9] Matt Deitke, Dustin Schwenk, Jordi Salvador, Luca Weihs, Oscar Michel, Eli VanderBilt, Ludwig Schmidt, Kiana Ehsani, Aniruddha Kembhavi, and Ali Farhadi. Objaverse: A universe of annotated 3D objects. In *CVPR*, pages 13142–13153, 2023. 2, 3

[10] Matt Deitke, Ruoshi Liu, Matthew Wallingford, Huong Ngo, Oscar Michel, Aditya Kusupati, Alan Fan, Christian Laforte, Vikram Voleti, Samir Yitzhak Gadre, et al. Objaverse-XL: A universe of 10M+ 3D objects. *NeurIPS*, 36, 2024. 2, 3

[11] Jianning Deng, Kartic Subr, and Hakan Bilen. PAOLI: Pose-free articulated object learning from sparse-view images. *arXiv preprint arXiv:2509.04276*, 2025. 3

[12] Matthew Fisher, Daniel Ritchie, Manolis Savva, Thomas Funkhouser, and Pat Hanrahan. Example-based synthesis of 3D object arrangements. *ACM TOG*, 31(6):1–11, 2012. 3

[13] Huan Fu, Rongfei Jia, Lin Gao, Mingming Gong, Binqiang Zhao, Steve Maybank, and Dacheng Tao. 3D-Future: 3D furniture shape with texture. *International Journal of Computer Vision*, 129:3313–3337, 2021. 3

[14] Daoyi Gao, Yawar Siddiqui, Lei Li, and Angela Dai. MeshArt: Generating articulated meshes with structure-guided transformers. In *CVPR*, pages 618–627, 2025. 3

[15] Haoran Geng, Helin Xu, Chengyang Zhao, Chao Xu, Li Yi, Siyuan Huang, and He Wang. GAPartNet: Cross-category domain-generalizable object perception and manipulation via generalizable and actionable parts. In *CVPR*, pages 7081–7091, 2023. 2, 3

[16] Pradyumn Goyal, Dmitry Petrov, Sheldon Andrews, Yizhak Ben-Shabat, Hsueh-Ti Derek Liu, and Evangelos Kalogerakis. GEOPARD: Geometric pretraining for articulation prediction in 3D shapes. *arXiv preprint arXiv:2504.02747*, 2025. 3

[17] Junfu Guo, Yu Xin, Gaoyi Liu, Kai Xu, Ligang Liu, and Ruizhen Hu. ArticulatedGS: Self-supervised digital twin modeling of articulated objects using 3d gaussian splatting. In *CVPR*, pages 27144–27153, 2025. 3

[18] Minghao Guo, Victor Zordan, Sheldon Andrews, Wojciech Matusik, Maneesh Agrawala, and Hsueh-Ti Derek Liu. Kinematic kitbashing for modeling functional articulated objects. *arXiv preprint arXiv:2510.13048*, 2025. 3

[19] Denys Iliash, Hanxiao Jiang, Yiming Zhang, Manolis Savva, and Angel X Chang. S2O: Static to openable enhancement for articulated 3D objects. *arXiv preprint arXiv:2409.18896*, 2024. 2, 3, 7, 4

[20] Hanxiao Jiang, Yongsen Mao, Manolis Savva, and Angel X Chang. OPD: Single-view 3d openable part detection. In *ECCV*, pages 410–426, 2022. 3

[21] Zhao Jin, Zhengping Che, Zhen Zhao, Kun Wu, Yuheng Zhang, Yinuo Zhao, Zehui Liu, Qiang Zhang, Xiaozhu Ju, Jing Tian, et al. ArtVIP: Articulated digital assets of visual realism, modular interaction, and physical fidelity for robot learning. *arXiv preprint arXiv:2506.04941*, 2025. 2, 3, 6

[22] Abhishek Joshi, Beining Han, Jack Nugent, Yiming Zuo, Jonathan Liu, Hongyu Wen, Stamatis Alexandropoulos, Tao Sun, Alexander Raistrick, Gaowen Liu, et al. Infinigen-Sim: Procedural generation of articulated simulation assets. *arXiv preprint arXiv:2505.10755*, 2025. 3

[23] Justin Kerr, Chung Min Kim, Mingxuan Wu, Brent Yi, Qianqian Wang, Ken Goldberg, and Angjoo Kanazawa. Robot see robot do: Imitating articulated object manipulation with monocular 4d reconstruction. *arXiv preprint arXiv:2409.18121*, 2024. 3

[24] Mukul Khanna, Yongsen Mao, Hanxiao Jiang, Sanjay Haresh, Brennan Shacklett, Dhruv Batra, Alexander Clegg, Eric Undersander, Angel X Chang, and Manolis Savva. Habitat synthetic scenes dataset (HSSD-200): An analysis

of 3D scene scale and realism tradeoffs for objectgoal navigation. In *CVPR*, pages 16384–16393, 2024. 3, 1

[25] Hyunjin Kim and Minhyuk Sung. PartSTAD: 2D-to-3D part segmentation task adaptation. In *ECCV*, pages 422–439, 2024. 4

[26] Jens U Kreber and Joerg Stueckler. Guiding diffusion-based articulated object generation by partial point cloud alignment and physical plausibility constraints. In *ICCV*, pages 3206–3214, 2025. 3

[27] Long Le, Jason Xie, William Liang, Hung-Ju Wang, Yue Yang, Yecheng Jason Ma, Kyle Vedder, Arjun Krishna, Dinesh Jayaraman, and Eric Eaton. Articulate-Anything: Automatic modeling of articulated objects via a vision-language foundation model. *arXiv preprint arXiv:2410.13882*, 2024. 3, 7, 8

[28] Han-Hung Lee, Yiming Zhang, and Angel X Chang. Duoduo CLIP: Efficient 3D understanding with multi-view images. In *ICLR*, 2025. 1

[29] Jiahui Lei, Congyue Deng, Bokui Shen, Leonidas Guibas, and Kostas Daniilidis. Nap: Neural 3d articulation prior. *arXiv preprint arXiv:2305.16315*, 2023. 3

[30] Xinyu Lian, Zichao Yu, Ruiming Liang, Yitong Wang, Li Ray Luo, Kaixu Chen, Yuanzhen Zhou, Qihong Tang, Xudong Xu, Zhaoyang Lyu, et al. Infinite mobility: Scalable high-fidelity synthesis of articulated objects via procedural generation. *arXiv preprint arXiv:2503.13424*, 2025. 3

[31] Jiayi Liu, Ali Mahdavi-Amiri, and Manolis Savva. PARIS: Part-level reconstruction and motion analysis for articulated objects. In *ICCV*, pages 352–363, 2023. 3

[32] Jiayi Liu, Denys Iliash, Angel X Chang, Manolis Savva, and Ali Mahdavi-Amiri. SINGAPO: Single image controlled generation of articulated parts in objects. *arXiv preprint arXiv:2410.16499*, 2024. 3, 7, 8

[33] Jiayi Liu, Hou In Ivan Tam, Ali Mahdavi-Amiri, and Manolis Savva. CAGE: Controllable articulation generation. In *CVPR*, pages 17880–17889, 2024. 3

[34] Liu Liu, Wenqiang Xu, Haoyuan Fu, Sucheng Qian, Qiaojun Yu, Yang Han, and Cewu Lu. AKB-48: A real-world articulated object knowledge base. In *CVPR*, pages 14809–14818, 2022. 2, 3

[35] Liu Liu, Han Xue, Wenqiang Xu, Haoyuan Fu, and Cewu Lu. Toward real-world category-level articulation pose estimation. *IEEE Transactions on Image Processing*, 31:1072–1083, 2022. 2, 3

[36] Minghua Liu, Yinhao Zhu, Hong Cai, Shizhong Han, Zhan Ling, Fatih Porikli, and Hao Su. PartSLIP: Low-shot part segmentation for 3D point clouds via pretrained image-language models. In *CVPR*, pages 21736–21746, 2023. 4

[37] Yu Liu, Baoxiong Jia, Ruijie Lu, Junfeng Ni, Song-Chun Zhu, and Siyuan Huang. Building interactable replicas of complex articulated objects via gaussian splatting. *ArXiv*, abs/2502.19459, 2025. 3

[38] Rundong Luo, Haoran Geng, Congyue Deng, Puhao Li, Zan Wang, Baoxiong Jia, Leonidas Guibas, and Siyuan Huang. PhysPart: Physically plausible part completion for interactable objects. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 12386–12393, 2025. 3

[39] Roberto Martín-Martín, Clemens Eppner, and Oliver Brock. The RBO dataset of articulated objects and interactions. *The International Journal of Robotics Research*, 38(9):1013–1019, 2019. 2, 3

[40] Kaichun Mo, Shilin Zhu, Angel X Chang, Li Yi, Subarna Tripathi, Leonidas J Guibas, and Hao Su. PartNet: A large-scale benchmark for fine-grained and hierarchical part-level 3D object understanding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 909–918, 2019. 3

[41] Sai Raj Kishore Perla, Aditya Vora, Sauradip Nag, Ali Mahdavi-Amiri, and Hao Zhang. ASIA: Adaptive 3D segmentation using few image annotations. *ArXiv*, abs/2509.24288, 2025. 4, 1

[42] Pengzhen Ren, Min Li, Zhen Luo, Xinshuai Song, Ziwei Chen, Weijia Liufu, Yixuan Yang, Hao Zheng, Rongtao Xu, Zitong Huang, et al. InfiniteWorld: A unified scalable simulation framework for general visual-language robot interaction. *arXiv preprint arXiv:2412.05789*, 2024. 3

[43] Jianhua Sun, Yuxuan Li, Jiude Wei, Longfei Xu, Nange Wang, Yining Zhang, and Cewu Lu. Arti-PG: A toolbox for procedurally synthesizing large-scale and diverse articulated objects with rich annotations. In *ICCV*, pages 6396–6405, 2025. 3

[44] Xiaohao Sun, Hanxiao Jiang, Manolis Savva, and Angel Chang. OPDMulti: Openable part detection for multiple objects. In *2024 International Conference on 3D Vision (3DV)*, pages 169–178, 2024. 3

[45] Anh Thai, Weiyao Wang, Hao Tang, Stefan Stojanov, James M Rehg, and Matt Feiszli. 3× 2: 3D object part segmentation by 2D semantic correspondences. In *European Conference on Computer Vision*, pages 149–166. Springer, 2024. 4

[46] Penghao Wang, Yiyang He, Xin Lv, Yukai Zhou, Lan Xu, Jingyi Yu, and Jiayuan Gu. PartNeXt: A next-generation dataset for fine-grained and hierarchical 3D part understanding. *arXiv preprint arXiv:2510.20155*, 2025. 3

[47] Xiaogang Wang, Bin Zhou, Yahao Shi, Xiaowu Chen, Qinping Zhao, and Kai Xu. Shape2motion: Joint analysis of motion parts and attributes from 3d shapes. In *CVPR*, pages 8876–8884, 2019. 2, 3

[48] Xinjie Wang, Liu Liu, Yu Cao, Ruiqi Wu, Wenkang Qin, Dehui Wang, Wei Sui, and Zhizhong Su. EmbodiedGen: Towards a generative 3d world engine for embodied intelligence. *ArXiv*, abs/2506.10600, 2025. 3

[49] Yijia Weng, Bowen Wen, Jonathan Tremblay, Valts Blukis, Dieter Fox, Leonidas Guibas, and Stan Birchfield. Neural implicit representation for building digital twins of unknown articulated objects. In *CVPR*, pages 3141–3150, 2024. 3

[50] Di Wu, Liu Liu, Linli Zhou, Anran Huang, Liangtu Song, Qiaojun Yu, Qi Wu, and Cewu Lu. REArtGS: Reconstructing and generating articulated objects via 3D gaussian splatting with geometric and motion constraints. *ArXiv*, abs/2503.06677, 2025. 3

[51] Ruiqi Wu, Xinjie Wang, Liu Liu, Chunle Guo, Jiaxiong Qiu, Chongyi Li, Lichao Huang, Zhizhong Su, and Ming-Ming Cheng. DIPO: Dual-state images controlled articulated ob-

ject generation powered by diverse data. *arXiv preprint arXiv:2505.20460*, 2025. 3

[52] Hongchi Xia, Entong Su, Marius Memmel, Arhan Jain, Raymond Yu, Numfor Mbiziwo-Tiapo, Ali Farhadi, Abhishek Gupta, Shenlong Wang, and Wei-Chiu Ma. DRAWER: Digital reconstruction and articulation with environment realism. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 21771–21782, 2025. 3

[53] Fanbo Xiang, Yuzhe Qin, Kaichun Mo, Yikuan Xia, Hao Zhu, Fangchen Liu, Minghua Liu, Hanxiao Jiang, Yifu Yuan, He Wang, et al. SAPIEN: A simulated part-based interactive environment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11097–11107, 2020. 2, 3, 5, 6, 7, 1

[54] Jianfeng Xiang, Zelong Lv, Sicheng Xu, Yu Deng, Ruicheng Wang, Bowen Zhang, Dong Chen, Xin Tong, and Jiaolong Yang. Structured 3d latents for scalable and versatile 3d generation. In *CVPR*, 2025. 4

[55] Xinhao Yan, Jiachen Xu, Yang Li, Changfeng Ma, Yunhan Yang, Chunshi Wang, Zibo Zhao, Zeqiang Lai, Yunfei Zhao, Zhuo Chen, and Chunchao Guo. X-part: high fidelity and structure coherent shape decomposition. *ArXiv*, abs/2509.08643, 2025. 4

[56] Zihao Yan, Ruizhen Hu, Xingguang Yan, Luanmin Chen, Oliver van Kaick, Hao Zhang, and Hui Huang. RPM-Net: Recurrent prediction of motion and parts from point cloud. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)*, 38(6):240:1–240:15, 2019. 2, 3

[57] Yandan Yang, Baoxiong Jia, Peiyuan Zhi, and Siyuan Huang. PhyScene: Physically interactable 3d scene synthesis for embodied ai. In *CVPR*, pages 16262–16272, 2024. 3

[58] Yibo Zhang, Li Zhang, Rui Ma, and Nan Cao. TexVerse: A universe of 3D objects with high-resolution textures. *arXiv preprint arXiv:2508.10868*, 2025. 3

[59] Ziming Zhong, Yanyu Xu, Jing Li, Jiale Xu, Zhengxin Li, Chaohui Yu, and Shenghua Gao. Meshsegmenter: Zero-shot mesh semantic segmentation via texture synthesis. In *ECCV*, pages 182–199, 2024. 4