ARTICULATEARENA: A METRIC FOR ARTICULATED KINEMATICS

Anonymous authors
Paper under double-blind review

ABSTRACT

Modern methods reconstruct or generate simulation-ready articulated objects, predicting not only their geometry but also how their parts are connected and allowed to move. Evaluating the geometry is straightforward, but evaluating the predicted articulation is not, because articulation specifies a motion rather than a shape, and there is no agreed distance between two motions. More specifically, existing protocols score joint type, axis direction, origin, and motion limits separately, although these parameters jointly describe a single physical motion, and the same motion can be written as different parameter values. As a result, a joint can score maximally wrong against an equivalent encoding of itself, and several component errors are ill-conditioned or undefined exactly where predictions become accurate. We propose **ArticulateArena**, a representation-invariant counterpart of Chamfer distance for articulation that compares the motions one-DOF joints induce rather than the parameters that encode them. It represents each joint by the unordered pair of its Lie-algebra endpoint twists, and we prove that the resulting quotient distance is a metric. It unifies fixed, revolute, prismatic, and helical joints, brings continuous joints into the same score through a compactification, and reads as the RMS motion of the moving part in meters when weighted by its mass distribution. A motion-aware tree edit distance lifts the metric to full kinematic trees, pricing structural errors such as spurious or missing joints in the same motion units as joint errors, and for a fixed inner product it remains a metric on trees up to relabeling. Alongside the metric we release **ArticulateArena-20K**, a new suite of 19,977 articulated objects with verified kinematics, and we re-evaluate published reconstruction methods on it under the new metric.

1 INTRODUCTION

Robots are trained and evaluated in simulation, and simulation benefits from running fast, ideally at many times real time. One way to speed it up is to simplify the world into rigid links with compact collision geometry, connected by joints that the simulator executes as constraints rather than by resolving collisions between objects. The constraint also stands in for the physical mechanism, so the hardware that realizes a joint never has to be modeled or held together by contact forces. Articulated objects are the assets that make this simplification possible, and a rapidly growing body of work therefore reconstructs them automatically, recovering part geometry and joint parameters, up to a full URDF with its kinematic tree of links, from several observed articulation states (Jiang et al., 2022b; Liu et al., 2023a; Weng et al., 2024; Guo et al., 2025; Lin et al., 2025; Wu et al., 2026a) or from one or a few images (Heppert et al., 2023; Chen et al., 2024; Mandi et al., 2025; Le et al., 2025; Yuan et al., 2025; Li et al., 2026b; Wang et al., 2026b; Wu et al., 2026b), while implicit shape representations with an articulation code recover joint angles without a kinematic model (Mu et al., 2021). Complementary systems infer articulation from a static 3D mesh, including ArtLLM, Particulate, and Articulate AnyMesh (Wang et al., 2026a; Li et al., 2026a; Qiu et al., 2025), or generate complete articulated assets, including Articraft and SPARK (Zhou et al., 2026; He et al., 2025). Most of this progress is reported through one evaluation template. For each joint it compares the predicted axis direction, the joint origin, the joint type, and, in the most complete protocols, the motion limits, four errors in four different units whose only aggregate is a thresholded success rate. The recovered geometry, by contrast, is judged by the agreed-upon Chamfer distance and F-score (Fan et al., 2017; Knapitsch et al., 2017). The recovered motion has no comparable standard. As a consequence, an asset that moves exactly like the real object can be reported as a failure, one that misses by a hair is indistinguishable from one that is wildly wrong, and the numbers of different papers cannot be compared, because each protocol scores its own subset of the components (Table 5).

We argue that this template, rather than the reconstruction models, is now the bottleneck of the field. Its difficulty is structural. The parameters of a joint describe one motion only together, so scoring

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

them in isolation misreads the joint. Worse, the same motion can be written down as different parameter values, so a parameter difference need not be a motion difference at all. Reversing the axis direction and negating the limits, for instance, describes the same hinge, yet the two encodings disagree in every component. A score built on the parameters therefore grades the code instead of the motion. It also judges too coarsely, since a joint whose single component slightly misses its tolerance is discarded as completely wrong rather than scored by how far the recovered motion actually is from the truth. Table 3 and Figure 6 make these failures concrete on simple joints, where a score in use charges a nonzero error between a joint and itself, loses conditioning exactly where a prediction becomes accurate, or jumps across a joint-type boundary that is physically continuous (§5.1). What one wants to measure is how far the recovered motion is from the true one, and no score in use is a distance on motions.

Our solution is to compare motions, not parameters. Classical screw theory represents the motion of a one-DOF joint by a single vector, a *twist*, together with the interval the joint can travel (Murray et al., 1994). We encode a joint by the unordered pair of twists at the two ends of that travel. This pair removes the representational redundancy entirely, because every way of writing down the same joint yields the same pair, up to swapping its two elements. The distance between two joints, E, is then simply the distance between their endpoint pairs, which we prove is a metric (§4). Because it compares motions directly, it treats a nearly welded revolute joint as nearly fixed instead of categorically wrong, and it stays well-behaved where the per-component scores blow up. The construction then extends along three independent axes. The norm behind E is a choice, E_α under a dimensionless norm that treats every joint alike, or E_B under a physical one that weights errors by how far they actually move the part, in meters. A compactification E^ϕ brings continuous joints into the same score instead of leaving them unscorable. A lift E^{tree} to kinematic trees prices structural errors, such as a spurious or missing joint, in the same motion units as joint errors. Together these pieces give articulation what Chamfer distance gives geometry, a single standard measure of error.

In detail, we make the following contributions:

(i) We define E, the quotient distance between the unordered pairs of Lie-algebra endpoint twists of one-DOF joints, with its compactified extension E^ϕ for continuous joints and its lift E^{tree} to kinematic trees, and prove that both are metrics for a fixed inner product (§4).

(ii) We show concretely which metric axiom each per-component score in current use violates and where each degenerates, then re-evaluate published methods under the quotient construction, which induces different orderings on their outputs (§5).

(iii) We introduce **ArticulateArena-20K**, a standard suite spanning 15 supercategories, 628 categories, and 19,977 articulated objects with ground-truth kinematics. The library covers all five one-DOF joint types, including the continuous joints that prior protocols cannot score (§4.4).

2 Related work

Articulated objects. Datasets and part-level perception supply the representations on which the reconstruction and generation systems cited in §1 build. PartNet and PartNet-Mobility provide part annotations and simulator-ready joints (Mo et al., 2019; Xiang et al., 2020), and perception methods predict motion parts, openable parts, actionable segments, or category-level part poses and articulation states (Wang et al., 2019; Jiang et al., 2022a; Geng et al., 2023; Liu et al., 2023b; Li et al., 2020; Weng et al., 2021), and factor-graph tracking recovers joint parameters from observed motion without category priors (Heppert et al., 2022). Whatever their input, the reconstruction systems score recovered kinematics with the per-component template that §3.2 examines. We do not propose another reconstructor, but a representation-invariant way to compare the articulated models these systems produce.

Rigid-motion metrics and structured matching. Coordinate-independent norms and metrics for rigid displacements and rotations (Kazerounian & Rastegar, 1992; Park, 1995; Huynh, 2009) compare poses or individual displacements, not joints with bounded or unbounded motion ranges, and the componentwise scores above make no use of them. We instead compare the full admissible motion encoded by each joint. At the object level, tree-edit distances (Zhang & Shasha, 1989; Pawlik & Augsten, 2016) price topology errors and assignment handles part relabeling (Kuhn, 1955); our lift connects them to the joints through a single motion-priced edit cost, rather than treating joint parameters as unrelated fields.

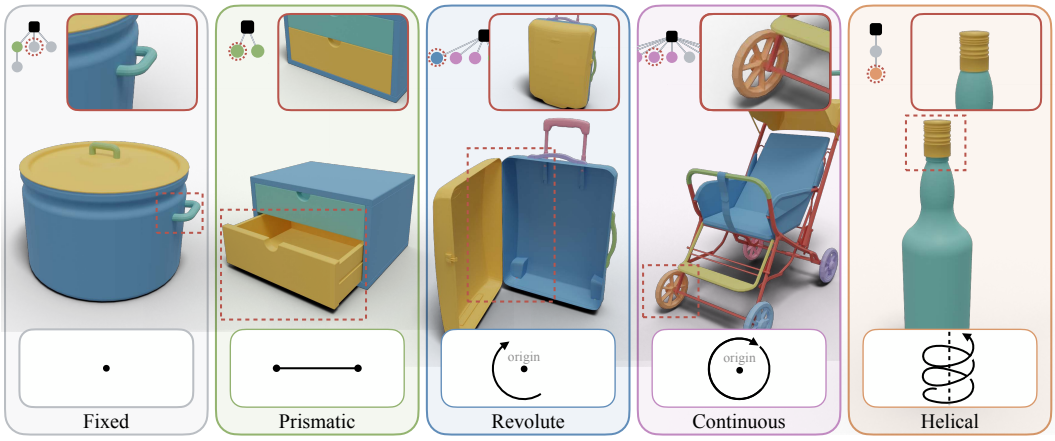


Figure 1: Categories of joints in articulated objects: fixed, prismatic, revolute, continuous, and helical. Each panel shows an object with one part displaced along its joint (dashed box), an inset of the same part at rest, the kinematic tree with that joint highlighted, and an icon of the motion the joint allows. These five one-DOF categories are exactly the joints our metric covers.

3 Preliminaries

3.1 Articulated objects

An articulated object is a set of rigid links connected by joints into a kinematic tree, usually encoded as a URDF. Each link carries a mesh or a set of collision shapes, and each non-root link is attached to its parent by exactly one joint. A one-DOF joint (Appendix A.1) is described by four pieces of information: a *type* t (Figure 1); an *axis* $\boldsymbol{a}$, the unit direction of rotation or translation; an *origin* $\boldsymbol{o}$, a point the axis passes through; and a *limit* $l = [l^-, l^+]$, the range of the joint coordinate. A revolute joint with $l^+ - l^- = \infty$ is *continuous*. The axis and origin fix the *line* along which the part moves, the limits fix the *extent* of the motion, and the type says which kind of motion it is. Together they form the tuple

$$J \;=\; (t,\; \boldsymbol{a},\; \boldsymbol{o},\; l), \tag{1}$$

which a reconstruction method predicts for every joint and an evaluation protocol must score against the ground truth, J_{pred} against J_{gt}.

3.2 Previous evaluation protocols

Table 1: Notation used throughout the paper.

Symbol	Meaning
J	a one-DOF joint
t	joint type
$\boldsymbol{a}$	joint axis (unit)
$\boldsymbol{o}$	joint origin
$l = [l^-, l^+]$	limit interval
q	joint coordinate (rad or m)
$\boldsymbol{\xi} = (\boldsymbol{\omega}, \boldsymbol{v})$	screw twist
$\boldsymbol{\omega}$	rotational part of $\boldsymbol{\xi}$
$\boldsymbol{v}$	translational part of $\boldsymbol{\xi}$
$\hat{\boldsymbol{\xi}}$	matrix form of $\boldsymbol{\xi}$
$\mathfrak{se}(3)$	Lie algebra of $SE(3)$

Existing reconstruction and generation methods evaluate predicted kinematics through one componentwise template, comparing the tuple J of equation 1 entry by entry. The most common form, of which the other protocols in use score subsets (Table 5, Appendix A.4), scores the following components (Le et al., 2025); URDFormer, one of the image-to-URDF pipelines above, reports only object- and part-level category, parent, and discretized position accuracy instead. They carry different units and cannot be summed, so the protocol thresholds each one and reports a *success rate*, the fraction of joints whose every component falls within its own tolerance. Appendix A.4 states this aggregation precisely and discusses what it discards, and §5.1 works through the failures of each component and shows how our metric resolves them.

(i) *Joint type error*, a binary measure of whether the predicted joint type matches the ground truth, $e_{\text{type}} = \mathbf{1}[t_{\text{pred}} \neq t_{\text{gt}}]$.

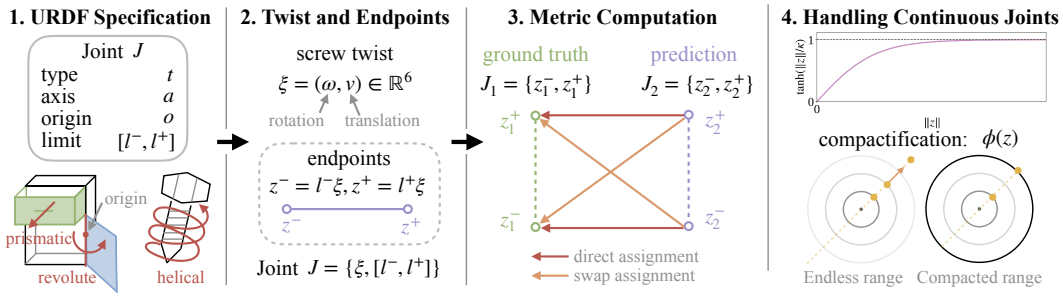


Figure 2: From a URDF joint to our metric. (1) A joint is specified by type, axis, origin, and limits. (2) The twist $\boldsymbol{\xi} = (\boldsymbol{\omega}, \boldsymbol{v})$ absorbs type, axis, and origin, and the limits give the endpoint pair $\boldsymbol{z}^{\pm} = l^{\pm}\boldsymbol{\xi}$. (3) E compares two endpoint pairs under the better of the direct and the swapped assignment. (4) Radial compactification maps unbounded ranges into the unit ball, so continuous joints become boundary points.

(ii) *Joint axis error*, the angle between the predicted and ground-truth joint axes up to sign, $e_{\text{axis}} = \arccos\left(|\boldsymbol{a}_{\text{pred}} \cdot \boldsymbol{a}_{\text{gt}}|/(\|\boldsymbol{a}_{\text{pred}}\|_2 \, \|\boldsymbol{a}_{\text{gt}}\|_2)\right) \in [0, \pi/2]$, where $\boldsymbol{a}$ is the axis of rotation for revolute joints or of translation for prismatic joints.

(iii) *Joint origin error*, the shortest distance between the joint axes for revolute joints and the Euclidean distance between origins for prismatic joints, with $\boldsymbol{p} = \boldsymbol{o}_{\text{pred}} - \boldsymbol{o}_{\text{gt}}$ given by $e_{\text{orig}}^{\text{rev}} = |\boldsymbol{p} \cdot (\boldsymbol{a}_{\text{pred}} \times \boldsymbol{a}_{\text{gt}})|/\|\boldsymbol{a}_{\text{pred}} \times \boldsymbol{a}_{\text{gt}}\|$ and $e_{\text{orig}}^{\text{pris}} = \|\boldsymbol{p}\|_2$.

(iv) *Joint limit error*, comprising two sub-components. With the motion vectors $\boldsymbol{m}_j = \boldsymbol{a}_j \, (l_j^+ - l_j^-)$ for $j \in \{\text{pred}, \text{gt}\}$, where l_j^+ and l_j^- are the upper and lower joint limits, the motion-range and motion-direction differences are $e_{\text{lim}}^{\text{range}} = \|\boldsymbol{m}_{\text{pred}} - \boldsymbol{m}_{\text{gt}}\|_2$ and $e_{\text{lim}}^{\text{dir}} = 1 - \boldsymbol{m}_{\text{pred}} \cdot \boldsymbol{m}_{\text{gt}}/(\|\boldsymbol{m}_{\text{pred}}\|_2 \, \|\boldsymbol{m}_{\text{gt}}\|_2)$, the latter ranging from 0 (identical direction) through 1 (perpendicular) to 2 (opposite).

4 The Quotient Metric

We define a joint by a single *twist* $\boldsymbol{\xi} = (\boldsymbol{\omega}, \boldsymbol{v}) \in \mathbb{R}^6$ together with its limit interval. The twist will absorb the type, the axis, and the origin at once. Here $\boldsymbol{\omega}$ is the rotational part and $\boldsymbol{v}$ the translational part of an infinitesimal rigid motion. The hat map writes $\boldsymbol{\xi}$ as the matrix $\hat{\boldsymbol{\xi}} = \left(\begin{smallmatrix} \hat{\omega} & v \\ 0 & 0 \end{smallmatrix}\right) \in \mathfrak{se}(3)$, with $\hat{\boldsymbol{\omega}}$ the skew-symmetric matrix of $\boldsymbol{\omega}$, and the matrix exponential $\exp(q\,\hat{\boldsymbol{\xi}}) \in SE(3)$ is the rigid motion reached by flowing along $\boldsymbol{\xi}$ for q units (Murray et al., 1994). A drawer slides along a prismatic twist, a door swings about a revolute twist, and a helical joint couples the two motions along one axis. Figure 2 summarizes the construction of this section. We normalize $\boldsymbol{\xi}$ to a *unit* screw, with $\|\boldsymbol{\omega}\| = 1$ when $\boldsymbol{\omega} \neq 0$ and $\|\boldsymbol{v}\| = 1$ otherwise, and call q the *joint coordinate*, the signed amount of motion away from the rest state. This normalization makes q a rotation angle in radians when $\boldsymbol{\omega} \neq 0$ (revolute, helical) and a translation length in meters when $\boldsymbol{\omega} = 0$ (prismatic), and the limits $l^{\pm}$ carry the same unit. The twist packs rotation and translation into one object, and the tuple parameters of §3.1 embed as special cases:

$$\boldsymbol{\xi} = (\boldsymbol{\omega}, \boldsymbol{v}), \qquad \boldsymbol{\xi}_{\text{rev}} = (\boldsymbol{a}, \boldsymbol{o} \times \boldsymbol{a}), \qquad \boldsymbol{\xi}_{\text{pris}} = (0, \boldsymbol{a}), \qquad \boldsymbol{\xi}_{\text{hel}} = (\boldsymbol{a}, \boldsymbol{o} \times \boldsymbol{a} + h\,\boldsymbol{a}), \tag{2}$$

with $\boldsymbol{a}$ the axis, $\boldsymbol{o}$ the origin, and h the pitch. The moment $\boldsymbol{o} \times \boldsymbol{a}$ encodes the axis position without naming a point on it (Appendix A.2 derives these embeddings).

We define a one-DOF joint by its motion, the curve of child-link placements $g(q) = \exp(q\,\hat{\boldsymbol{\xi}})\, g_0$ in $SE(3)$, $q \in [l^-, l^+]$, with $\boldsymbol{\xi} \in \mathfrak{se}(3)$ a unit screw twist and the anchor g_0 the placement of the link at $q = 0$, the state in which the URDF stores its meshes. Prediction and ground truth describe the same object in the same stored state, so the anchor is shared data whose accuracy the shape metrics judge, and the kinematic content of the joint is the pair

$$J = \left(\boldsymbol{\xi}, \, [l^-, l^+]\right) \in \mathfrak{se}(3) \times \overline{\mathbb{R}}^2, \tag{3}$$

replacing the componentwise tuple $(t, \boldsymbol{a}, \boldsymbol{o}, l)$ of §3. The interval $[l^-, l^+]$ is extended-real, so that continuous joints have $l^{\pm} = \pm\infty$. Equivalently we represent the joint by its unordered pair of

endpoint twists

$$\{\boldsymbol{z}^-, \boldsymbol{z}^+\} = \{l^- \boldsymbol{\xi},\ l^+ \boldsymbol{\xi}\} \subset \mathfrak{se}(3), \tag{4}$$

which keeps limits as first-class data and has no log/winding ambiguity. Re-zeroing the coordinate is not a free re-encoding, because $q \mapsto q - \delta$ moves the anchor to $\exp(\delta\,\hat{\boldsymbol{\xi}})\, g_0$, which the stored state pins down. One representational symmetry remains (Proposition B.7). Reversing the screw orientation swaps the endpoints, and the unordered pair absorbs it. The revolute exponential is 2π-periodic, so limits shifted by a full turn generate the same placements, but we keep the joint coordinate as part of the specification and do not quotient this shift, which keeps E continuous in the limits (Remark B.9). Two anchored encodings then have the same endpoint pair iff they agree up to orientation reversal, and the proposed metric is the quotient distance of flat $\mathfrak{se}(3) \times \mathfrak{se}(3)$ under the swap:

$$E(J_1, J_2) \;=\; \min\left\{\|\boldsymbol{z}_1^- - \boldsymbol{z}_2^-\|^2 + \|\boldsymbol{z}_1^+ - \boldsymbol{z}_2^+\|^2,\ \ \|\boldsymbol{z}_1^- - \boldsymbol{z}_2^+\|^2 + \|\boldsymbol{z}_1^+ - \boldsymbol{z}_2^-\|^2\right\}^{1/2} \tag{5}$$

Table 2: The design space. An inner product (rows) combined with a limit treatment or the tree lift (columns). Shaded entries mark the reported variant.

	Finite	Unbounded	Trees
Split norm $\|\cdot\|_\alpha$	E_α	E_α^ϕ	$E_\alpha^{\phi,\mathrm{tree}}$
Kinetic norm $\|\cdot\|_B$	E_B	E_B^ϕ	$E_B^{\phi,\mathrm{tree}}$

a true metric, being a quotient by a finite isometry group (Theorem B.1). The norm $\|\cdot\|$ comes from an inner product on $\mathfrak{se}(3)$ chosen in §4.1, and a subscript names that choice (Table 2): E_α under the split norm and E_B under the kinetic-energy norm. Superscripts mark extensions that apply under either subscript, E^ϕ for the compactified metric of §4.2 and E^{tree} for the tree lift of §4.3. It is zero iff the two encodings agree up to orientation reversal (Proposition B.7), smooth except on a measure-zero set of ties, invariant under screw-orientation reversal and the choice of prismatic origin (Propositions B.2 and B.3), and it identifies the joint from its endpoint pair up to orientation (Proposition B.5). Theorem B.17 further bounds E on both sides by the L^2 discrepancy of the full linearized motion segment, so comparing endpoints already controls the whole path. Appendix B states and proves these properties. E is symmetric, so we write its arguments as J_1 and J_2 rather than J_{pred} and J_{gt}. The hard min makes E itself nondifferentiable at ties. When a differentiable score is wanted, mainly as a training loss, Appendix A.3 gives smooth swap-invariant surrogates that replace the min. Appendix A.3 states the full computation, at the joint and the tree level, in pseudocode.

4.1 Inner product on $\mathfrak{se}(3)$

The norm in equation 5 acts on endpoint-twist differences $\Delta\boldsymbol{z} = (\Delta\boldsymbol{\omega}, \Delta\boldsymbol{v})$, and Theorem B.1 asks only that it come from an inner product on $\mathfrak{se}(3)$. The most intuitive choice is the *split norm*, which weighs rotation against translation through an inverse-length α:

$$\|\Delta\boldsymbol{z}\|_\alpha^2 \;=\; \|\Delta\boldsymbol{\omega}\|^2 + \alpha^2\|\Delta\boldsymbol{v}\|^2. \tag{6}$$

The resulting E_α is dimensionless and satisfies the metric axioms (Theorem B.1), and it compares every joint of every object by one fixed rule. That uniformity is also its flaw, since a 5° axis error scores the same on a 1 m door and on a 2 cm knob, although the door moves fifty times more material.

The *kinetic-energy form* lets the moving link set the scale instead. It is induced by the child link's body B through its mass m, center of mass $\boldsymbol{c}$, and inertia tensor I:

$$\|\Delta\boldsymbol{z}\|_B^2 \;=\; \frac{1}{m}\,\Delta\boldsymbol{\omega}^\top I\,\Delta\boldsymbol{\omega} + \|\Delta\boldsymbol{v} + \Delta\boldsymbol{\omega} \times \boldsymbol{c}\|^2. \tag{7}$$

The resulting E_B is measured in meters and equals the RMS linearized material motion of the child link (Theorem B.15), so the door now scores fifty times worse than the knob. The two forms are close relatives. For a centered link with isotropic inertia $I = mr^2 I_3$, where r is the radius of gyration, equation 7 is exactly r^2 times equation 6 at $\alpha = 1/r$. A global α therefore assumes that every link has the same size $1/\alpha$, and the kinetic-energy form replaces that assumption with the moments of the link that actually moves. Its price is that it needs those moments. The joint must come with a link geometry, the score inherits the object's scale, and the norm varies from joint to joint, which keeps the per-link variant of the tree lift outside the metric proof (§4.3). E_B itself remains a true metric, because the body B is fixed by the ground truth before any prediction is seen, so each joint carries one fixed inner-product norm and Theorem B.1 applies verbatim. We keep both norms, E_α for geometry-free uniform comparison and E_B as the physically calibrated default of our evaluation.

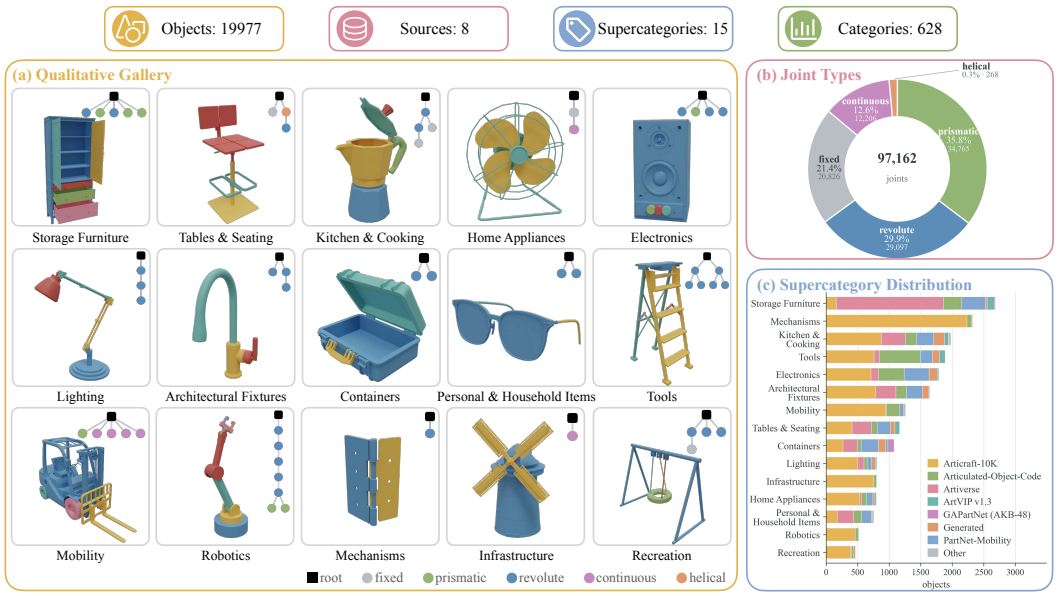


Figure 3: Overview of **ArticulateArena-20K**. (a) One object from each of the 15 supercategories, rendered with its kinematic tree (top right). (b) The distribution of the library's joints over the five joint types. (c) Objects and categories per supercategory, stacked by source library.

4.2 Unbounded joints via radial compactification

For unbounded joints we map $\mathfrak{se}(3)$ into its unit ball by radial compactification and measure the quotient distance there:

$$\phi(\boldsymbol{z}) \;=\; \tanh\big(\|\boldsymbol{z}\|/\kappa\big)\,\frac{\boldsymbol{z}}{\|\boldsymbol{z}\|}, \qquad E^{\phi}(J_1, J_2) \;=\; E\big(\{\phi(\boldsymbol{z}_1^-),\,\phi(\boldsymbol{z}_1^+)\},\;\{\phi(\boldsymbol{z}_2^-),\,\phi(\boldsymbol{z}_2^+)\}\big) \tag{8}$$

A continuous joint becomes the antipodal boundary pair $\{\pm\boldsymbol{\xi}\}$, i.e. the axis direction up to sign. The completion of the space of limited joints contains the continuous joints as boundary points, and E^{ϕ} is continuous in this limit (Proposition B.11). The construction applies under either inner product of §4.1, with $\|\cdot\|$ in ϕ the chosen norm, giving E^{ϕ}_{α} and E^{ϕ}_{B}. Both are dimensionless, including E^{ϕ}_{B}. Dividing by $\|\boldsymbol{z}\|$ inside ϕ gives each compactified endpoint the inverse units of the norm, and the outer norm in E cancels them again, so $\|\phi(\boldsymbol{z})\| = \tanh(\|\boldsymbol{z}\|/\kappa)$ and every distance between compactified pairs is a pure number (Remark B.14). E^{ϕ}_{B} thus trades the meter reading of E_B for boundedness while keeping the body-aware weighting of error directions, and κ carries the unit of the norm. Although motivated by unbounded joints, ϕ is defined on all of $\mathfrak{se}(3)$, so every joint type can be compactified. Evaluating all pairs under E^{ϕ} therefore unifies finite and continuous joints in a single dimensionless score, which §5 uses as the common score over all joint types.

4.3 Lifting to skeleton and geometry

The metric so far covers a single joint, but an object has several joints arranged in a kinematic tree, with links as nodes and joints on the edges. The same object admits many such trees, since parts can be relabeled, the base link re-chosen, and a handle either made its own link welded to the drawer or left part of the drawer mesh. The object-level distance must therefore live on trees modulo these choices. The fixed joint $J_0 = \{0, 0\}$ is a canonical origin, and $E(J, J_0)$ measures how much motion a joint carries. We lift E to trees, under either inner product of §4.1, as an edit distance whose every cost comes from E itself. An edit *substitutes* the joint on an edge at cost $E(J, J')$, *contracts* an edge by merging its two links at cost $E(J, J_0)$, or *expands* a link into two joined by a new joint J at the same cost. The tree distance is the cheapest edit sequence from one tree to the other,

$$E^{\mathrm{tree}}(T_1, T_2) \;=\; \inf_{\pi\,:\,T_1 \rightsquigarrow T_2} \sum_{o \in \pi} c(o), \tag{9}$$

over finite edit sequences π on unrooted trees with unlabeled nodes. No hand-set topology penalty appears, because a spurious or missing joint costs exactly the motion it carries. A nearly welded extra

handle is charged $\epsilon\|\boldsymbol{\xi}\|$ rather than a constant, so topology and joint discrepancies are measured in the same motion space.

On raw URDF trees E^{tree} is only a pseudometric, since a tree with an exact fixed edge and the tree with that edge contracted are at distance zero. Call two trees *representation equivalent* when they differ only by node relabeling, base-link choice, and insertion or deletion of exact fixed edges. On these classes E^{tree} is a metric (Theorem B.20). We compute E^{tree} exactly, certified through its assignment relaxation (Kuhn, 1955) (Appendix B.5). Geometry enters through the inner product. The configuration discrepancy $d(g_1, g_2)$ of a link between two placements, the mass-weighted L^2 displacement of its material points, depends on the mass measure only through its moments $(m, \boldsymbol{c}, I)$ (Kazerounian & Rastegar, 1992; Park, 1995) (Appendix B.4), which gives E_B its geometry-aware reading (§4.1). Instantiating the edit costs per joint with the ground-truth child link's kinetic-energy form gives the physically weighted variant of our evaluation, a metric for a fixed skeleton and correspondence (Theorem B.22), with only the topology edits under these edge-dependent norms outside the theorems (§4.1).

4.4 The ArticulateArena-20K dataset

A metric is most useful when it is computed on a standard suite of objects, so that numbers are comparable across papers, and such a suite needs ground-truth kinematics that are trustworthy across all joint types. We therefore assemble **ArticulateArena-20K**, a suite of 19,977 articulated objects in 628 categories and 15 supercategories that merges public articulated-object libraries with objects we generated ourselves, all normalized into a common URDF convention with verified joint annotations so that every object exposes the tuple of §3.1 and, equivalently, the endpoint-twist representation of this section. It covers the five joint types of Figure 1, continuous joints included, and ranges from single-joint items to assemblies with over a hundred movable joints. Figure 3 summarizes the composition, and Appendix E details the sources, the construction pipeline, and the category taxonomy. The empirical study of §5 evaluates on 200 objects chosen from the library to cover every supercategory and as many categories as possible.

5 Experiments

5.1 Where per-component scores fail and E succeeds

Table 3 collects the canonical cases in which the scores of §3.2 report a nonzero, ill-conditioned, or undefined error while E returns the correct value under either inner product, taking Articulate-Anything as the representative protocol. Appendix A.5 works through each case.

Table 3: Failure cases of per-component scores. Rows 1 and 4 compare two encodings of the same motion, row 2 two different motions that a component score cannot separate, and rows 3 and 5 two motions that converge to each other.

	Case	Per-component score	Ours
1	Axis reversed, limits negated	$e_{\mathrm{lim}}^{\mathrm{dir}} = 2$, maximal against itself	$E = 0$, endpoints swap
2	Interval shifted, same width	$e_{\mathrm{lim}} = 0$ although the configurations differ	$E > 0$
3	Near-parallel axes	$e_{\mathrm{orig}}^{\mathrm{rev}}$ divides by $\|\boldsymbol{a}_{\mathrm{pred}} \times \boldsymbol{a}_{\mathrm{gt}}\| \to 0$, undefined at parallel	no denominator, shrinks continuously
4	Prismatic origin moved along the part	$e_{\mathrm{orig}}^{\mathrm{pris}} > 0$	$E = 0$, origin absent from the twist
5	Revolute range $[0, \epsilon]$ against fixed	$e_{\mathrm{type}} = 1$ for every $\epsilon > 0$	$E = \epsilon\|\boldsymbol{\xi}\| \to 0$
6	Continuous joint	e_{lim} undefined, $\boldsymbol{m}$ infinite	boundary point under ϕ, finite ranges approach it
7	Aggregation	thresholds discard magnitude, no distance	a metric, with a differentiable surrogate

5.2 Ablation

We ablate five design choices of the metric: the inner product, E_α against E_B (Figure 4; Appendices C.3 and C.4), the split-norm weight α (Appendix C.2), the compactification scale κ (Ap-

pendix C.5), the two-endpoint representation against denser sampling (Appendix C.6), and the motion-aware tree lift against a constant topology penalty (Figure 5). Appendix C.1 sweeps every single-parameter perturbation of a ground-truth joint against the prior component scores.

Inner product. Figure 4 calibrates both inner products against an external ground truth, the material-motion discrepancy d of §4.3. The question is not which score is larger, since their units differ, but which one tracks actual material motion. E_B lies on the identity to within a few percent, and its only visible departures are the largest axis errors at large swept angles, exactly the linearization gap that Theorem B.15 delimits. E_α preserves the trend on average but scatters by more than an order of magnitude at a fixed d, so the same physical displacement can score ten times differently depending on the link. That scatter is the empirical cost of assuming every link has size $1/\alpha$ (§4.1).

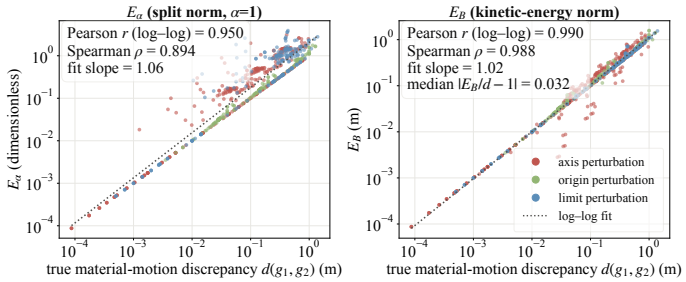


Figure 4: Calibration of E_α (left) and E_B (right) against the true material-motion discrepancy $d(g_1, g_2)$ of the child link, computed on the actual $SE(3)$ displacements rather than on their linearization. Axis, origin, and limit perturbations (colors) of 690 joints from 60 objects whose child links span a wide range of sizes. The dotted line is the least-squares fit in log space.

Compactification scale. The scale κ in equation 8 sets where the transition between finite and continuous ranges happens, not whether it happens. Sweeping κ around the default π changes no qualitative property of E^ϕ, only how resolution is allocated along the range axis, and $\kappa = \pi$ keeps ranges up to a full turn in the responsive part of the curve (Appendix C.5, Figure 10).

Endpoint sampling. Comparing two joints through two endpoint twists might seem to discard the motion in between, yet sampling n points along each linearized segment adds nothing. The squared discrepancy is quadratic along the segment, so every $n \geq 3$ coincides with the $n \to \infty$ limit, and denser sampling only reparameterizes the endpoint metric at $n/2$ times the cost (Appendix C.6, Figure 12).

Tree lift. Figure 5 compares the motion-aware edit costs of §4.3 with the common alternative, a constant penalty λ for every unmatched joint. The constant penalty is discontinuous at every structural boundary. A spurious joint that barely moves, a missing joint that barely moved, and a part split into two along its own screw all cost a full λ the moment they appear, whereas E_α^{tree} starts from zero and grows with the motion involved. The knob then faces a dilemma. Once a joint's own error exceeds 2λ the assignment prefers to leave it unmatched, so the score plateaus while E_α^{tree} keeps growing, and forcing the match instead escapes the cap only by breaking the triangle inequality. A constant penalty is either capped or not a metric. On random corruptions it also piles distinct predictions onto identical scores in bands at multiples of λ, and the ranking it induces changes with the knob itself. Pricing structural edits by the motion they carry removes all of this, and it has no parameter to tune. The figure reports the exact edit distance, and its assignment relaxation returns the same value on every sweep and on nearly all random corruptions (Appendix B.5).

5.3 Comparison

Setup. We re-evaluate eight published methods on 200 objects from **ArticulateArena-20K** (Appendix D.1). Table 4 reports the prior component protocol alongside E under three parameter choices of §4, the object-level $E_\alpha^{\phi,\text{tree}}$, finite-range E_B, and all-joint E_α^ϕ ($\alpha = 1$, $\kappa = \pi$). The tree score is the exact edit distance of §4.3 (Appendix B.5). Gen (%) is computed over all objects; the remaining entries are conditional on a valid articulated output. Methods retain their native input modalities, so this compares kinematic outputs rather than common-input end-to-end performance. The success indicator, coverage, frame normalization, and geometry handling are in Appendices A.4 and D.3.

Results. On these outputs the quotient scores induce different orderings from the thresholded component protocol, especially when axis, origin, and limit errors trade off. This is expected because E_B scores the encoded motion jointly, absorbs representation symmetries, and weights errors by their ef-

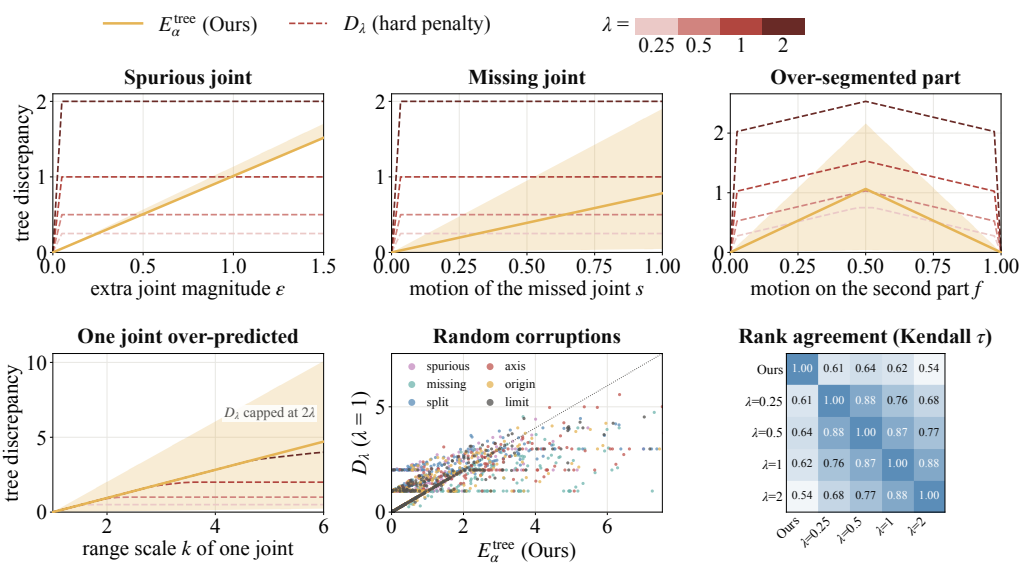


Figure 5: Motion-aware edit costs (E_α^{tree} computed exactly, amber, median and interquartile range over 100 multi-joint trees from ArticulateArena-20K) against a constant topology penalty λ per unmatched joint (D_λ, dashed, $\lambda \in \{0.25, 0.5, 1, 2\}$). Both schemes score joints with the same metric E_α and differ only in what an unmatched joint costs. Top: a spurious joint of growing magnitude, a missing joint whose own motion shrinks to zero, and one part split into two along the same screw. Bottom left: the range of one joint scaled by k. Bottom middle: 2,000 randomly corrupted predictions, E_α^{tree} against D_1, colored by the first edit. Bottom right: Kendall τ rank agreement among the schemes over those predictions.

Table 4: Articulated reconstruction on 200 objects from **ArticulateArena-20K**. Arrows indicate the better direction; shading marks the best defined value; – denotes unavailable entries. The bootstrap intervals of Appendix D show which of these gaps the evaluation set resolves. †Articulate AnyMesh predicts no motion limits, so its E columns impute each joint's limits from the ground truth and are excluded from the shading.

		Prior per-component scores					Ours		
Method	Gen↑ (%)	TypeErr↓ (%)	OriginErr↓ (m)	AxisErr↓ (rad)	LimitErr↓	SuccRate↑ (%)	E_B ↓ (m)	E_α^ϕ ↓	$E_\alpha^{\phi,\text{tree}}$ ↓
Articraft	99.5	12.1	0.466	0.870	1.90 / 1.00	17.1	1.172	0.597	0.476
Articulate AnyMesh†	48.0	16.7	0.264	0.881	–	9.4	1.032	0.581	0.528
Articulate-Anything	98.5	18.3	0.604	0.629	2.19 / 0.91	15.7	1.702	0.550	0.392
ArtLLM	93.0	30.6	0.398	0.918	7.13 / 1.03	3.8	2.868	1.216	0.585
Ditto	100.0	31.5	0.645	0.819	1.75 / 1.05	5.5	1.365	0.608	0.552
Particulate	91.0	7.1	0.414	0.172	0.98 / 0.34	15.4	1.066	0.465	0.419
SPARK	99.5	18.1	0.530	0.800	2.36 / 1.00	9.5	1.543	0.687	0.520
URDFormer	77.0	37.7	0.735	1.019	2.07 / 0.98	0.0	1.964	0.726	0.566

fect on the ground-truth body. Detailed comparisons, qualitative cases (Figure 13), and common-set uncertainty are reported in Appendices D.2 and D.4.

6 Conclusion

We introduced a quotient metric on one-DOF joints that compares the motions the joints generate rather than the parameters encoding them, places all five joint types, continuous included, on one scale, and reads as the RMS motion of the child link in meters under the kinetic-energy inner product. Lifted to kinematic trees, it replaces the per-component scores in current use with a representation-invariant distance that satisfies the metric axioms for a fixed inner product and stays well-conditioned where those scores degenerate. We hope that it, together with ArticulateArena-20K, becomes the default way to report articulated-reconstruction quality. Multi-DOF joints, closed kinematic chains, and sharper swept-motion bounds under the kinetic-energy form are left to future work.

AI USE STATEMENT

We used generative AI tools in three places. First, in the generation of the dataset, where part of ArticulateArena-20K was produced with generative models (Appendix E.1). Second, in the organization of the dataset, where an LLM proposed a category label for every object from its name, description, and renders. Third, to refine text written by the authors, to re-check the mathematical proofs written by the authors, and to help with LaTeX layout. We have reviewed all AI-assisted work. Every generated asset was loaded and checked for validity. Every proposed category label was reviewed and corrected by the authors, so the final categorization and supercategory taxonomy (Appendix E.2) reflect human judgment. All refined text was checked by the authors for accuracy, and every proof was verified by the authors independently of the AI re-check. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

The one aspect of this work the ICLR Code of Ethics speaks to is dataset release. ArticulateArena-20K redistributes assets from publicly released research datasets under their respective licenses, with attribution in §4.4 and Appendix E, alongside synthetic assets we generated ourselves. The library describes household objects and contains no personal data or human-subject material.

REPRODUCIBILITY STATEMENT

The metric is fully specified by equation 5, equation 6, equation 7, equation 8, and equation 9, with the screw-twist embeddings of a URDF joint in Appendix A.2 and the differentiable training surrogates in Appendix A.3. Every theoretical claim is stated with its assumptions and proved in Appendix B. The per-component scores and the success indicator used for comparison are written out in §3.2 and Appendix A.4, evaluation counts and the common-set analysis in Appendix D, and the ablation setups in §5.2 and Appendix C. The construction, normalization, and quality control of ArticulateArena-20K are described in Appendix E. We will release the evaluation code, the ablation scripts, and the library with its URDF assets and verified joint annotations.

REFERENCES

Zoey Chen, Aaron Walsman, Marius Memmel, Kaichun Mo, Alex Fang, Karthikeya Vemuri, Alan Wu, Dieter Fox, and Abhishek Gupta. URDFormer: A pipeline for constructing articulated simulation environments from real-world images. In *Robotics: Science and Systems (RSS)*, 2024.

Haoqiang Fan, Hao Su, and Leonidas J. Guibas. A point set generation network for 3d object reconstruction from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2463–2471, 2017.

Yipeng Gao, Yunhao Ge, Peilin Cai, Daniel Seita, and Laurent Itti. Lam: Language articulated object modelers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7081–7091, 2023.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 27144–27153, 2025.

Yumeng He, Ying Jiang, Jiayin Lu, Yin Yang, and Chenfanfu Jiang. SPARK: Sim-ready part-level articulated reconstruction with VLM knowledge. *arXiv preprint arXiv:2512.01629*, 2025.

Nick Heppert, Toki Migimatsu, Brent Yi, Claire Chen, and Jeannette Bohg. Category-independent articulated object tracking with factor graphs. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3800–3807. IEEE, 2022.

Nick Heppert, Muhammad Zubair Irshad, Sergey Zakharov, Katherine Liu, Rares Andrei Ambrus, Jeannette Bohg, Abhinav Valada, and Thomas Kollar. CARTO: Category and joint agnostic reconstruction of ARTiculated objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21201–21210, 2023.

Du Q. Huynh. Metrics for 3d rotations: Comparison and analysis. *Journal of Mathematical Imaging and Vision*, 35(2):155–164, 2009.

Denys Iliash, Jiayi Liu, Egor Fokin, Qirui Wu, Ali Mahdavi Amiri, Manolis Savva, and Angel X Chang. Artiverse: A diverse and physically grounded dataset for articulated objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8932–8942, 2026.

Hanxiao Jiang, Yongsen Mao, Manolis Savva, and Angel X. Chang. OPD: Single-view 3d openable part detection. In *European Conference on Computer Vision (ECCV)*, pp. 410–426, 2022a.

Zhenyu Jiang, Cheng-Chun Hsu, and Yuke Zhu. Ditto: Building digital twins of articulated objects from interaction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022b.

Kazem Kazerounian and Jahangir Rastegar. Object norms: A class of coordinate and metric independent norms for displacements. In *Proceedings of the ASME 22nd Biennial Mechanisms Conference*, pp. 271–275, 1992.

Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics*, 36(4):78:1–78:13, 2017.

Harold W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1–2):83–97, 1955.

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. In *International Conference on Learning Representations (ICLR)*, 2025.

Ruining Li, Yuxin Yao, Chuanxia Zheng, Christian Rupprecht, Joan Lasenby, Shangzhe Wu, and Andrea Vedaldi. Particulate: Feed-forward 3D object articulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026a.

Xiaolong Li, He Wang, Li Yi, Leonidas J. Guibas, A. Lynn Abbott, and Shuran Song. Category-level articulated object pose estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3706–3715, 2020.

Zizhang Li, Cheng Zhang, Zhengqin Li, Henry Howard-Jenkins, Zhaoyang Lv, Chen Geng, Jiajun Wu, Richard Newcombe, Jakob Engel, and Zhao Dong. ART: Articulated reconstruction transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7468–7479, 2026b.

Shengjie Lin, Jiading Fang, Muhammad Zubair Irshad, Vitor Campagnolo Guizilini, Rares Andrei Ambrus, Greg Shakhnarovich, and Matthew R. Walter. SplArt: Articulation estimation and part-level reconstruction with 3d gaussian splatting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8841–8851, 2025.

Jiayi Liu, Ali Mahdavi-Amiri, and Manolis Savva. PARIS: Part-level reconstruction and motion analysis for articulated objects. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023a.

Minghua Liu, Yinhao Zhu, Hong Cai, Shizhong Han, Zhan Ling, Fatih Porikli, and Hao Su. Part-SLIP: Low-shot part segmentation for 3d point clouds via pretrained image-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21736–21746, 2023b.

Zhao Mandi, Yijia Weng, Dominik Bauer, and Shuran Song. Real2code: Reconstruct articulated objects via code generation. In *International Conference on Learning Representations (ICLR)*, 2025.

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 (CVPR)*, pp. 909–918, 2019.

Jiteng Mu, Weichao Qiu, Adam Kortylewski, Alan Yuille, Nuno Vasconcelos, and Xiaolong Wang. A-SDF: Learning disentangled signed distance functions for articulated shape representation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 13001–13011, 2021.

Richard M. Murray, Zexiang Li, and S. Shankar Sastry. *A Mathematical Introduction to Robotic Manipulation*. CRC Press, 1994.

Frank C. Park. Distance metrics on the rigid-body motions with applications to mechanism design. *ASME Journal of Mechanical Design*, 117(1):48–54, 1995.

Mateusz Pawlik and Nikolaus Augsten. Tree edit distance: Robust and memory-efficient. *Information Systems*, 56:157–173, 2016.

Xiaowen Qiu, Jincheng Yang, Yian Wang, Zhehuan Chen, Yufei Wang, Tsun-Hsuan Wang, Zhou Xian, and Chuang Gan. Articulate AnyMesh: Open-vocabulary 3D articulated objects modeling. *arXiv preprint arXiv:2502.02590*, 2025.

Alessandro Tasora, Radu Serban, Hammad Mazhar, Arman Pazouki, Daniel Melanz, Jonathan Fleischmann, Michael Taylor, Hiroyuki Sugiyama, and Dan Negrut. Chrono: An open source multi-physics dynamics engine. In *international conference on high performance computing in science and engineering*, pp. 19–49. Springer, 2015.

Penghao Wang, Siyuan Xie, Hongyu Yan, Xianghui Yang, Jingwei Huang, Chunchao Guo, and Jiayuan Gu. ArtLLM: Generating articulated assets via 3D LLM. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 34281–34291, 2026a.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8876–8884, 2019.

Xuyang Wang, Zhenyu Li, Jian Ding, Habib Slim, Peter Wonka, Hongdong Li, and Mohamed Elhoseiny. Artic-O: End-to-end articulated object reconstruction via latent geometry learning. *arXiv preprint arXiv:2606.21938*, 2026b.

Yijia Weng, He Wang, Qiang Zhou, Yuzhe Qin, Yueqi Duan, Qingnan Fan, Baoquan Chen, Hao Su, and Leonidas J. Guibas. CAPTRA: Category-level pose tracking for rigid and articulated objects from point clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 13209–13218, 2021.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3141–3150, 2024.

Di Wu, Liu Liu, Anran Huang, Yuyan Liu, Qiaojun Yu, Shaofan Liu, Liangtu Song, and Cewu Lu. REArtGS++: Generalizable articulation reconstruction with temporal geometry constraint via planar gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1177–1186, 2026a.

Zhuangzhe Wu, Yue Xin, Chengkai Hou, Minghao Chen, Yaoxu Lyu, Jieyu Zhang, and Shanghang Zhang. URDF-Anything+: End-to-end generation for simulation-ready articulated assets. *arXiv preprint arXiv:2603.14010*, 2026b.

Fanbo Xiang, Yuzhe Qin, Kaichun Mo, Yikuan Xia, Hao Zhu, Fangchen Liu, Minghua Liu, Hanxiao Jiang, Yifu Yuan, He Wang, Li Yi, Angel X. Chang, Leonidas J. Guibas, and Hao Su. SAPIEN: A SimulAted part-based interactive environment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11097–11107, 2020.

Sylvia Yuan, Ruoxi Shi, Xinyue Wei, Xiaoshuai Zhang, Hao Su, and Minghua Liu. LARM: A large articulated-object reconstruction model. In *ACM SIGGRAPH Asia Conference Papers*, 2025.

Kaizhong Zhang and Dennis E. Shasha. Simple fast algorithms for the editing distance between trees and related problems. *SIAM Journal on Computing*, 18(6):1245–1262, 1989.

Matt Zhou, Ruining Li, Xiaoyang Lyu, Zhaomou Song, Zhening Huang, Chuanxia Zheng, Christian Rupprecht, Andrea Vedaldi, and Shangzhe Wu. Articraft: An agentic system for scalable articulated 3D asset generation. *arXiv preprint arXiv:2605.15187*, 2026.

Appendix contents

A Details of the metric and of the prior protocol **14**
A.1 Scope of one-DOF joints 14
A.2 Screw twists of the joint tuple 15
A.3 Computation of the metric 15
A.4 The per-component protocol 16
A.5 Failure cases in detail 17
B Theoretical properties of our metric **18**
B.1 Metric axioms 18
B.2 Representation invariance and identifiability 18
B.3 Continuity and compactification 20
B.4 Interpretation of the distance 21
B.5 The tree edit metric 22
C Additional ablation studies **24**
C.1 Single-parameter perturbations against the component scores 24
C.2 Split-norm weight 25
C.3 Sweep shapes under both inner products 25
C.4 Object scale 26
C.5 Compactification scale 27
C.6 Endpoint sampling 27
D Additional evaluation results **28**
D.1 The evaluation set 28
D.2 Detailed comparison 29
D.3 Evaluation coverage 30
D.4 Common-set sensitivity 30
E The ArticulateArena-20K library **31**
E.1 Construction 31
E.2 Supercategory taxonomy 32

A Details of the metric and of the prior protocol

Appendix A.1 records why the metric is defined on one-DOF joints, Appendix A.2 derives the twist embedding of the joint tuple, Appendix A.3 states the evaluation in pseudocode together with its differentiable surrogates, Appendix A.4 details the per-component protocol of §3.2, and Appendix A.5 expands the failure cases of Table 3.

A.1 Scope of one-DOF joints

The metric is defined on one-DOF joints, and the library and evaluation follow the same scope. This choice is deliberate, and we record the reasoning here, following the joint taxonomy used by simulation engines such as Chrono (Tasora et al., 2015).

First, one DOF is where the theory is clean. A one-DOF joint is a single curve $q \mapsto \exp(q\,\hat{\boldsymbol{\xi}})$, so its admissible motion is captured by one twist and one interval, the endpoint pair lives in $\mathfrak{se}(3) \times \mathfrak{se}(3)$, and the whole quotient construction of §4 goes through with a single $\mathbb{Z}_2$ swap. A k-DOF joint

sweeps a k-dimensional region of $\mathfrak{se}(3)$, and the analogue of the endpoint pair would be a set-valued object, for instance an ellipsoid-like region covering the admissible twists. Such an extension is possible in principle, and the quotient viewpoint would carry over, but the resulting distance loses the closed form and the compact proofs that make the one-DOF case practical.

Second, one DOF is where the data is. The asset formats in current use, URDF and the MJCF-style XML variants, expose revolute, prismatic, and their fixed and continuous special cases, and reconstruction and generation methods predict exactly these types. Simulators likewise execute higher-DOF connections mostly as compositions of one-DOF primitives. A metric supporting joint classes that no representation, generator, or evaluation target produces would exercise no data.

Third, higher-DOF joints decompose. A cylindrical joint is a revolute and a prismatic joint on the same axis, a universal joint is two revolute joints with intersecting axes, and a planar or spherical joint factors into two or three one-DOF joints through intermediate massless links. The tree lift of §4.3 then compares such compositions joint by joint. The decomposition is less clean than a native multi-DOF treatment, since it introduces intermediate links and is not unique, but it keeps every comparison inside the one-DOF theory. Extending the metric natively to multi-DOF joints is a matter of engineering the set-valued endpoint representation, and we leave it to future work.

A.2 Screw twists of the joint tuple

A body moving along a twist $\boldsymbol{\xi} = (\boldsymbol{\omega}, \boldsymbol{v})$ carries the material point instantaneously at position $\boldsymbol{x}$ with velocity $\boldsymbol{\omega} \times \boldsymbol{x} + \boldsymbol{v}$. In matrix form,

$$\hat{\boldsymbol{\xi}} = \begin{pmatrix} \hat{\boldsymbol{\omega}} & \boldsymbol{v} \\ 0 & 0 \end{pmatrix} \in \mathfrak{se}(3), \qquad \hat{\boldsymbol{\omega}} = \begin{pmatrix} 0 & -\boldsymbol{\omega}_3 & \boldsymbol{\omega}_2 \\ \boldsymbol{\omega}_3 & 0 & -\boldsymbol{\omega}_1 \\ -\boldsymbol{\omega}_2 & \boldsymbol{\omega}_1 & 0 \end{pmatrix}, \tag{10}$$

so that $\hat{\boldsymbol{\omega}}\boldsymbol{x} = \boldsymbol{\omega} \times \boldsymbol{x}$. The embeddings of equation 2 follow directly. A rotation about the axis through $\boldsymbol{o}$ with unit direction $\boldsymbol{a}$ moves the point at $\boldsymbol{x}$ with velocity $\boldsymbol{a} \times (\boldsymbol{x} - \boldsymbol{o}) = \boldsymbol{a} \times \boldsymbol{x} + \boldsymbol{o} \times \boldsymbol{a}$, which is the twist $(\boldsymbol{a},\ \boldsymbol{o} \times \boldsymbol{a})$. Adding a translation of pitch h along the axis appends $h\,\boldsymbol{a}$ to the translational part, and a pure translation is $(0,\ \boldsymbol{a})$, in which no origin appears.

A.3 Computation of the metric

Algorithms 1 and 2 state the evaluation in pseudocode. Algorithm 1 computes the joint distance of equation 5, optionally compactified by equation 8, under either norm of §4.1. Algorithm 2 lifts it to two kinematic trees, using the certificate of Appendix B.5 so that the returned value is the exact edit distance.

Algorithm 1 Joint distance $E(J_1, J_2)$, with the compactified variant E^{ϕ}.

Require: joints $J_i = (t_i, \boldsymbol{a}_i, \boldsymbol{o}_i, l_i)$ for $i \in \{1, 2\}$, a norm $\|\cdot\|$ from equation 6 or equation 7, a scale κ if compactifying

$\boldsymbol{\xi}_i \leftarrow$ unit screw twist of $(t_i, \boldsymbol{a}_i, \boldsymbol{o}_i)$ by equation 2, for $i \in \{1, 2\}$
$\boldsymbol{z}_i^{\pm} \leftarrow l_i^{\pm}\,\boldsymbol{\xi}_i$ {endpoint pairs, equation 4}
if computing E^{ϕ}, which is required as soon as any $l_i^{\pm}$ is infinite and optional otherwise **then**
  for all endpoints $\boldsymbol{z} \in \{\boldsymbol{z}_1^-, \boldsymbol{z}_1^+, \boldsymbol{z}_2^-, \boldsymbol{z}_2^+\}$ **do**
    if $\boldsymbol{z} = \pm\infty\,\boldsymbol{\xi}$ **then**
      $\boldsymbol{z} \leftarrow \pm\boldsymbol{\xi}/\|\boldsymbol{\xi}\|$ {continuous joint, boundary point of the unit ball}
    else
      $\boldsymbol{z} \leftarrow \tanh\big(\|\boldsymbol{z}\|/\kappa\big)\,\boldsymbol{z}/\|\boldsymbol{z}\|$ {equation 8}
    end if
  end for
end if
$d_{\mathrm{id}} \leftarrow \big(\|\boldsymbol{z}_1^- - \boldsymbol{z}_2^-\|^2 + \|\boldsymbol{z}_1^+ - \boldsymbol{z}_2^+\|^2\big)^{1/2}$
$d_{\mathrm{sw}} \leftarrow \big(\|\boldsymbol{z}_1^- - \boldsymbol{z}_2^+\|^2 + \|\boldsymbol{z}_1^+ - \boldsymbol{z}_2^-\|^2\big)^{1/2}$
return $\min(d_{\mathrm{id}}, d_{\mathrm{sw}})$ {quotient by the endpoint swap, equation 5}

Algorithm 2 Tree distance $E^{\mathrm{tree}}(T_1, T_2)$, exact via the certificate.

Require: kinematic trees T_1, T_2 with joint twists in a common canonical frame
1: contract every exact fixed edge of T_1 and of T_2
2: $C_{jk} \leftarrow E(J_j, J_k)$ for all moving joints $J_j \in T_1$, $J_k \in T_2$ {Algorithm 1}
3: $u_j \leftarrow E(J_j, J_0)$ for every moving joint {cost of leaving J_j unmatched}
4: $M \leftarrow$ optimal assignment under C with unassigned costs u {Hungarian}
5: **if** M extends to an isomorphism of the contracted trees **then**
6: **return** $\mathrm{cost}(M)$ {certified exact, Appendix B.5}
7: **end if**
8: **return** $\min\{\mathrm{cost}(M') : M' \text{ realizable}\}$ {search over realizable matchings}

Differentiable surrogates for training and optimization. The metric E of equation 5 combines the two endpoint discrepancies by a hard min over the swap. This is the exact metric and is what every evaluation in this paper uses, but it is not differentiable at ties, where the two assignments cost the same. When a differentiable score is wanted, mainly to use E as a training loss, two smooth alternatives replace the min. The first is a loss on the symmetric invariants $\boldsymbol{s} = \boldsymbol{z}^- + \boldsymbol{z}^+$ and $D = (\boldsymbol{z}^- - \boldsymbol{z}^+)(\boldsymbol{z}^- - \boldsymbol{z}^+)^\top$, which are swap-invariant by construction. The second is a soft-min over the two assignments with a temperature that anneals to the exact metric.

A.4 The per-component protocol

This subsection collects the details of the per-component template of §3.2, which components each published protocol scores and the thresholded success indicator used in Table 4.

Scores reported by previous methods. Table 5 lists which kinematic components each published protocol scores, and is the basis of the template written out in §3.2. Every entry records a quantity the cited method reports in its own evaluation of the reconstructed kinematics, with the source named below. Ditto (Jiang et al., 2022b) reports the axis-angle error for both joint types and the pivot-position error for revolute joints in its Table 1, and joint type accuracy in Appendix D on the Synthetic and Shape2Motion sets, where nearly all methods reach 100%. PARIS (Liu et al., 2023a) reports the axis-angle and revolute pivot-position errors of its Table 2. URDFormer (Chen et al., 2024) reports no joint-level component, only the object- and part-category, parent, and discretized part-position accuracies of its Table II, together with the whole-pipeline real-robot task success rate of its Table I. Real2Code (Mandi et al., 2025) reports the joint type, axis, and revolute position errors of its Table 2. Articulate-Anything (Le et al., 2025) scores all four components and reports the thresholded joint success rate that aggregates them, stating the position and angle tolerances as $50\,\mathrm{mm}$ and $0.25\,\mathrm{rad}$ and leaving the limit tolerances unspecified. SPARK (He et al., 2025) reports the AxisErr, PivotErr, and TypeErr triple of its Table 2.

Table 5: Kinematic evaluation metrics reported by existing articulated-object reconstruction methods. ✓ marks a component the method reports in its own evaluation of the reconstructed kinematics, whether in its main results or in an appendix, and † marks a component reported for revolute joints only. A dash indicates that the method reports no such quantity. §Joint type accuracy is reported only in Appendix D of Jiang et al. (2022b), on the Synthetic and Shape2Motion sets, where nearly all methods reach 100%. ‡URDFormer infers the joint type and the link axis from the predicted mesh category and scores neither, reporting only object- and part-category accuracy, parent accuracy, a discretized part-position error, and a whole-pipeline real-robot task success rate. Geometry-only metrics such as Chamfer distance and F-score are outside the scope of this table.

Method	Type	Axis	Origin	Limit	Success rate
Ditto (Jiang et al., 2022b)	✓§	✓	✓†	–	–
PARIS (Liu et al., 2023a)	–	✓	✓†	–	–
URDFormer (Chen et al., 2024)‡	–	–	–	–	–
Real2Code (Mandi et al., 2025)	✓	✓	✓†	–	–
Articulate-Anything (Le et al., 2025)	✓	✓	✓	✓	✓
SPARK (He et al., 2025)	✓	✓	✓	–	–

Aggregation and the success indicator. The components of §3.2 carry a count, an angle in radians, a length in meters, a range that is a length or an angle depending on type, and a dimensionless direction term in $[0, 2]$. No weighted sum of such units is meaningful, so the most complete protocol aggregates by thresholding. This discards the magnitude of every error, so a prediction just past a tolerance and one that is wildly wrong are the same failure, and in no case is the result a distance on joints. Most protocols drop the limits altogether (Table 5), and the one that scores them folds the interval back onto the axis through $\boldsymbol{m}$, which reintroduces the sign ambiguity that e_{axis} was patched to remove and is undefined for continuous joints. The failures share one root. The four parameters encode a single motion only together, and every score reads them in isolation, so the template grades how the motion is written down rather than the motion itself.

The thresholded aggregation is stated only in prose by Le et al. (2025). For the success-rate column of Table 4 we compute it as the indicator

$$\text{succ}(J_{\text{pred}}, J_{\text{gt}}) \;=\; \mathbb{1}\big[\, e_{\text{type}} = 0 \;\wedge\; e_{\text{axis}} < \tau_{\boldsymbol{a}} \;\wedge\; e_{\text{orig}} < \tau_p \,\big], \tag{11}$$

with the component scores e_{type}, e_{axis}, and e_{orig} of §3.2, the published thresholds $\tau_a = 0.25$ rad and $\tau_p = 50$ mm, and the reported score is the fraction of joints with $\text{succ} = 1$. The original protocol leaves the limit tolerances unspecified, so we threshold only the three well-specified components and drop the limit terms.

A.5 Failure cases in detail

We expand the rows of Table 3 in order; Figure 6 shows each failure as a continuous sweep against E_α.

(i) *A joint against itself under axis reversal.* The joint $(\boldsymbol{\xi}, [l^-, l^+])$ and $(-\boldsymbol{\xi}, [-l^+, -l^-])$ are the same motion (for instance $\boldsymbol{a} = (1, 0, 0)$, $[0, 1]$ versus $\boldsymbol{a} = (-1, 0, 0)$, $[-1, 0]$). The axis error is zero, but the motion vector $\boldsymbol{m} = \boldsymbol{a}\,(l^+ - l^-)$ flips sign, so $e^{\text{dir}}_{\text{lim}} = 2$ is maximal, scoring the joint maximally against itself. Our endpoints $\{\boldsymbol{z}^-, \boldsymbol{z}^+\} = \{l^-\boldsymbol{\xi},\, l^+\boldsymbol{\xi}\}$ only swap, so $E = 0$.

(ii) *Where the interval sits, not only its width.* Since $\boldsymbol{m} = \boldsymbol{a}\,(l^+ - l^-)$ keeps only the length, $[0, 1]$ and $[1, 2]$ both give $\boldsymbol{m} = \boldsymbol{a}$ and zero limit error although they permit different configurations. Our endpoints $\{0, \boldsymbol{\xi}\}$ against $\{\boldsymbol{\xi}, 2\boldsymbol{\xi}\}$ give $E > 0$. This reads the joint coordinate against the shared anchor g_0 of §4, so a genuine interval shift is charged, while an encoding that merely stores the object in another state is re-anchored first (Remark B.9) and is not.

(iii) *Near-parallel axes.* The revolute origin score $e^{\text{rev}}_{\text{orig}}$ of §3.2 divides by $\|\boldsymbol{a}_{\text{pred}} \times \boldsymbol{a}_{\text{gt}}\| \to 0$ as $\boldsymbol{a}_{\text{pred}} \to \boldsymbol{a}_{\text{gt}}$. It is worst-conditioned exactly when the prediction is most accurate, and undefined for parallel axes. Comparing full twists through $\|\boldsymbol{z}_{\text{pred}} - \boldsymbol{z}_{\text{gt}}\|$ has no such denominator and shrinks continuously as the axes align.

(iv) *Prismatic origin.* A prismatic twist $\boldsymbol{\xi} = (0, \boldsymbol{a})$ carries no axis location, so moving the drawn origin leaves the motion unchanged, yet $e^{\text{pris}}_{\text{orig}} = \|\boldsymbol{o}_{\text{pred}} - \boldsymbol{o}_{\text{gt}}\|$ penalizes it. The origin is absent from the twist, and $E = 0$. A rest-pose offset that a URDF stores in the joint origin, when itself under test, is compared separately as a link pose g_0, not through the joint twist.

(v) *Type as a continuum.* A revolute joint with limits $[0, \epsilon]$ becomes physically fixed as $\epsilon \to 0$, yet the categorical e_{type} stays 1. Our endpoints $\{0, \epsilon\boldsymbol{\xi}\}$ against $\{0, 0\}$ give $E = \epsilon\|\boldsymbol{\xi}\| \to 0$. Because the twist carries pitch continuously, E likewise places revolute, prismatic, and helical joints on one scale, where the categorical type has no metric and no slot for helical joints.

(vi) *Continuous joints.* For $l^\pm = \pm\infty$ the vector $\boldsymbol{m} = \boldsymbol{a}\,(l^+ - l^-)$ is undefined, so the limit score cannot handle continuous joints at all. The compactification ϕ of equation 8 sends them to the boundary pair $\{\pm\boldsymbol{\xi}/\|\boldsymbol{\xi}\|\}$, which finite ranges approach continuously, so bounded, large-range, and continuous joints are all measured.

(vii) *One metric, and a loss.* E is a single distance obeying the metric axioms, so it supports ranking, mean and median error, retrieval, clustering, and significance tests, and its smooth $(\boldsymbol{s}, D)$ surrogate (Appendix A.3) can serve as a differentiable training loss where binary type error and hard thresholds cannot.

B THEORETICAL PROPERTIES OF OUR METRIC

We provide formal statements and proofs for the main properties of the endpoint-twist representation and of our metric. Throughout this section, $V = \mathfrak{se}(3) \cong \mathbb{R}^6$ denotes the twist-coordinate space of §4, equipped with a fixed positive-definite inner product $\langle \cdot, \cdot \rangle$ and induced norm $\|\boldsymbol{z}\| = \sqrt{\langle \boldsymbol{z}, \boldsymbol{z} \rangle}$, and E denotes the endpoint metric under this generic inner product. Both norms of §4.1 are instances, the split norm $\|\cdot\|_\alpha$ and, for a fixed body B, the kinetic-energy norm $\|\cdot\|_B$, so every statement below holds verbatim for E_α and E_B. Only the results that carry the subscript B are specific to the kinetic form. A finite-range one-DOF joint is represented by a normalized screw twist $\boldsymbol{\xi} \in V$ and an interval $[l^-, l^+]$. Its endpoint twists are $\boldsymbol{z}^- = l^- \boldsymbol{\xi}$ and $\boldsymbol{z}^+ = l^+ \boldsymbol{\xi}$, and we identify the joint with the unordered endpoint pair $J \equiv \{\boldsymbol{z}^-, \boldsymbol{z}^+\}$. For two joints $J_1 = \{\boldsymbol{z}_1^-, \boldsymbol{z}_1^+\}$ and $J_2 = \{\boldsymbol{z}_2^-, \boldsymbol{z}_2^+\}$, the distance $E(J_1, J_2)$ is that of equation 5, read with the generic norm above.

Appendix B.1 proves the metric axioms, Appendix B.2 shows that E depends on the motion rather than on its encoding, Appendix B.3 treats the fixed and continuous limits, Appendix B.4 interprets the value of E, and Appendix B.5 lifts these results to kinematic trees.

B.1 METRIC AXIOMS

Theorem B.1 (Endpoint quotient metric)**.** *The function E of equation 5 is a metric on the space of unordered endpoint pairs in V.*

Proof. Let $X = V \times V$ carry the product metric $d_X\big((a,b),(c,d)\big) = \big(\|a-c\|^2 + \|b-d\|^2\big)^{1/2}$, a metric because $\|\cdot\|$ is induced by an inner product. The endpoint swap $S(a,b) = (b,a)$ satisfies $d_X(Sx, Sy) = d_X(x,y)$, so it is an isometry of X, and an unordered pair is the equivalence class $[(a,b)] = \{(a,b),(b,a)\}$ under the two-element group $G = \{I, S\}$. The proposed distance is $E([x],[y]) = \min_{g \in G} d_X(x, gy)$. We verify the axioms.

Non-negativity is immediate from $d_X \geq 0$.

Identity of indiscernibles. If $E([x],[y]) = 0$, then $d_X(x, gy) = 0$ for some $g \in G$, so $x = gy$ and $[x] = [y]$. Conversely, $[x] = [y]$ means $x = gy$ for some $g \in G$, so $E([x],[y]) = 0$.

Symmetry. Every $g \in G$ is an isometry with $g^{-1} \in G$, so $E([x],[y]) = \min_{g \in G} d_X(x, gy) = \min_{g \in G} d_X(g^{-1}x, y) = E([y],[x])$.

Triangle inequality. Let $g, h \in G$ attain $E([x],[y]) = d_X(x, gy)$ and $E([y],[z]) = d_X(y, hz)$. Since $gh \in G$ and g is an isometry,

$$E([x],[z]) \leq d_X(x, ghz) \leq d_X(x, gy) + d_X(gy, ghz) = E([x],[y]) + E([y],[z]). \tag{12}$$

□

B.2 REPRESENTATION INVARIANCE AND IDENTIFIABILITY

Proposition B.2 (Screw-orientation invariance)**.** *The endpoint representation and E are invariant under simultaneous reversal of the screw orientation and limits, $(\boldsymbol{\xi}, [l^-, l^+]) \mapsto (-\boldsymbol{\xi}, [-l^+, -l^-])$.*

Proof. The reversal sends $J = \{l^-\boldsymbol{\xi}, l^+\boldsymbol{\xi}\}$ to $J' = \{(-l^+)(-\boldsymbol{\xi}), (-l^-)(-\boldsymbol{\xi})\} = \{l^+\boldsymbol{\xi}, l^-\boldsymbol{\xi}\}$. Since the pair is unordered, $J' = J$ and $E(J, J') = 0$. □

Proposition B.3 (Prismatic-origin invariance)**.** *Two prismatic joints with the same translation direction and limits but different points chosen as their joint origins have identical endpoint-twist representations, hence distance zero under E, holding the rest-pose transform fixed.*

Proof. A prismatic joint with unit translation direction $\boldsymbol{a}$ has screw twist $\boldsymbol{\xi} = (0, \boldsymbol{a})$, which does not depend on a point chosen along the translation axis. Changing the reported origin therefore changes neither $\boldsymbol{\xi}$ nor $\{l^-\boldsymbol{\xi}, l^+\boldsymbol{\xi}\}$. □

Remark B.4. The proposition concerns the freedom to place the origin anywhere on the prismatic motion axis. Moving the rest-pose transform along the motion while shifting the limits to compensate is instead a change of anchor, which the re-anchoring of Remark B.9 removes before evaluation.

We assume a fixed screw-normalization convention with the property that two collinear normalized nonzero screws differ only by sign, so that $\boldsymbol{\xi}_2 = \lambda\boldsymbol{\xi}_1$ implies $\lambda \in \{-1, +1\}$.

Proposition B.5 (Identifiability modulo screw orientation)**.** *For every non-fixed finite-range one-DOF joint, the unordered endpoint pair $\{l^-\boldsymbol{\xi}, l^+\boldsymbol{\xi}\}$ uniquely identifies the normalized joint parameters $(\boldsymbol{\xi}, [l^-, l^+])$ up to the orientation equivalence $(\boldsymbol{\xi}, [l^-, l^+]) \sim (-\boldsymbol{\xi}, [-l^+, -l^-])$.*

Proof. Suppose two non-fixed joints have the same endpoint pair, $\{l_1^-\boldsymbol{\xi}_1, l_1^+\boldsymbol{\xi}_1\} = \{l_2^-\boldsymbol{\xi}_2, l_2^+\boldsymbol{\xi}_2\}$. Because the joints are non-fixed, the shared pair contains a nonzero endpoint, and every nonzero endpoint lies in both $\mathrm{span}(\boldsymbol{\xi}_1)$ and $\mathrm{span}(\boldsymbol{\xi}_2)$. The two spans therefore agree, so $\boldsymbol{\xi}_2 = \lambda\boldsymbol{\xi}_1$ for some nonzero λ, which the normalization forces to ± 1. If $\boldsymbol{\xi}_2 = \boldsymbol{\xi}_1$, equality of the pairs gives $[l_2^-, l_2^+] = [l_1^-, l_1^+]$. If $\boldsymbol{\xi}_2 = -\boldsymbol{\xi}_1$, it forces $l_2^- = -l_1^+$ and $l_2^+ = -l_1^-$, which is exactly the orientation equivalence. □

Remark B.6. For a fixed joint $\{\boldsymbol{z}^-, \boldsymbol{z}^+\} = \{0, 0\}$, the underlying screw direction is intentionally unidentifiable. All such parameterizations describe the same zero-motion joint.

Proposition B.7 (Same motion, same representation)**.** *Let two joints with unit screws $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2$, finite limits $l_i^- < l_i^+$, and a shared anchor g_0 generate the placement curves $g_i(q) = \exp(q\,\hat{\boldsymbol{\xi}}_i)\,g_0$ on $[l_i^-, l_i^+]$. The curves coincide up to a coordinate change $q \mapsto \sigma q + \delta$ with $\sigma \in \{\pm 1\}$ iff the encodings agree up to the orientation reversal of Proposition B.2 and, when the screw is revolute, a shift of both limits by a common $2\pi k$ with $k \in \mathbb{Z}$. Consequently $E(J_1, J_2) = 0$ iff the encodings agree up to orientation reversal, which for a prismatic or helical screw is iff the joints generate the same motion, and for a revolute screw iff they generate the same motion with the same winding of the coordinate.*

Proof. Suppose $\exp((\sigma q + \delta)\,\hat{\boldsymbol{\xi}}_1)\,g_0 = \exp(q\,\hat{\boldsymbol{\xi}}_2)\,g_0$ for all $q \in [l_2^-, l_2^+]$, with the limit intervals matched by the coordinate change. Cancel g_0 and differentiate in q to get $\sigma\,\hat{\boldsymbol{\xi}}_1 X(q) = \hat{\boldsymbol{\xi}}_2 X(q)$ with $X(q) = \exp(q\,\hat{\boldsymbol{\xi}}_2)$ invertible, so $\boldsymbol{\xi}_2 = \sigma\boldsymbol{\xi}_1$. Substituting back leaves $\exp(\delta\,\hat{\boldsymbol{\xi}}_1) = I$. A prismatic screw translates by δ and a helical screw by δ times its nonzero pitch, so both force $\delta = 0$, while a revolute screw rotates by the angle δ about its axis line, so $\delta \in 2\pi\mathbb{Z}$. Hence for $\sigma = 1$ the encodings agree up to a revolute $2\pi k$ limit shift, and for $\sigma = -1$ up to that shift composed with orientation reversal. Conversely, orientation reversal reparameterizes the same curve by $q \mapsto -q$, and a revolute $2\pi k$ limit shift reparameterizes it by $q \mapsto q + 2\pi k$, fixing the curve pointwise by the 2π-periodicity of the revolute exponential. For the consequence, orientation reversal only swaps the endpoint pair, so encodings that agree up to it have $E = 0$. If instead $E = 0$, the endpoint pairs are equal as unordered pairs, and Proposition B.5 recovers the encodings up to orientation reversal. The first part identifies this with the same motion, except for the revolute $2\pi k$ shift, which E does not quotient (Remark B.9). □

Remark B.8. Unit normalization makes the affine form of the coordinate change automatic rather than an assumption. If a smooth monotone reparameterization f matches the two curves, the same differentiation gives $f'(q)\,\boldsymbol{\xi}_1 = \boldsymbol{\xi}_2$ for all q, so f' is a constant of absolute value one. Continuous joints have $l^\pm = \pm\infty$, every shift reparameterizes the same full orbit, and the compactified representation $\{\pm\boldsymbol{\xi}/\|\boldsymbol{\xi}\|\}$ is already shift-invariant. A fixed joint is a constant curve and is represented by $\{0, 0\}$ for every screw.

Remark B.9 (Re-anchoring). An encoding that stores the object in a different state, with anchor $g_0' = \exp(\delta\,\hat{\boldsymbol{\xi}})\,g_0$ and limits $[l^- - \delta,\ l^+ - \delta]$, describes the same motion. Before evaluation it is re-

anchored by shifting its limits by δ, so that both coordinates read zero at the same placement, and a drawer saved half open is not charged a limit error against the same drawer saved closed. The revolute 2π shift is the one re-anchoring that no stored state pins down, since $\exp(2\pi k\,\hat{\boldsymbol{\xi}})\,g_0 = g_0$, and E deliberately leaves it unquotiented. A URDF whose rest state lies a full turn outside its own limits is a different specification for every simulator that does not reduce revolute coordinates, and any choice of a representative from each $2\pi\mathbb{Z}$ orbit is discontinuous at its seam. If one nevertheless wants to identify encodings that differ by a full turn, one may first replace $l^{\pm}$ by $l^{\pm} - 2\pi k$ with $k = \left\lfloor \frac{1}{2\pi}\left(\frac{l^- + l^+}{2} + \pi\right)\right\rfloor$, which wraps the limit midpoint into $[-\pi, \pi)$ and makes $E = 0$ exactly on joints with the same motion; the price is that $[0, 2\pi - \epsilon]$ and $[0, 2\pi]$ are then sent a full turn apart.

B.3 Continuity and compactification

Proposition B.10 (Continuous limit to a fixed joint)**.** *Let $J_0 = \{0, 0\}$ represent a fixed joint and let $J_\epsilon = \{0, \epsilon\boldsymbol{\xi}\}$ be a finite-range one-DOF joint with normalized screw $\boldsymbol{\xi}$. Then $E(J_\epsilon, J_0) = |\epsilon|\,\|\boldsymbol{\xi}\| \to 0$ as $\epsilon \to 0$.*

Proof. Both endpoints of J_0 are zero, so either endpoint assignment yields the same value, $E(J_\epsilon, J_0) = \big(\|0 - 0\|^2 + \|\epsilon\boldsymbol{\xi} - 0\|^2\big)^{1/2} = |\epsilon|\,\|\boldsymbol{\xi}\|$. □

For a scale parameter $\kappa > 0$, let $\phi_\kappa : V \to V$ be the radial map

$$\phi_\kappa(\boldsymbol{z}) = \begin{cases} \tanh\left(\dfrac{\|\boldsymbol{z}\|}{\kappa}\right)\dfrac{\boldsymbol{z}}{\|\boldsymbol{z}\|}, & \boldsymbol{z} \neq 0, \\ 0, & \boldsymbol{z} = 0, \end{cases} \tag{13}$$

whose image is the open unit ball $\mathbb{B}^\circ = \{\boldsymbol{x} \in V : \|\boldsymbol{x}\| < 1\}$. The compactified endpoint metric is $E^\phi(J_1, J_2) = E\big(\{\phi_\kappa(\boldsymbol{z}_1^-), \phi_\kappa(\boldsymbol{z}_1^+)\}, \{\phi_\kappa(\boldsymbol{z}_2^-), \phi_\kappa(\boldsymbol{z}_2^+)\}\big)$.

Proposition B.11 (Convergence to an unbounded joint)**.** *Let $\boldsymbol{u} \in V$ with $\|\boldsymbol{u}\| = 1$, and consider the symmetric finite-range joints $J_L = \{-L\boldsymbol{u}, L\boldsymbol{u}\}$. Under radial compactification, $J_L \to J_\infty \equiv \{-\boldsymbol{u}, \boldsymbol{u}\}$ as $L \to \infty$, with $E^\phi(J_L, J_\infty) = \sqrt{2}\,\big(1 - \tanh(L/\kappa)\big) \to 0$.*

Proof. Since $\|\boldsymbol{u}\| = 1$, we have $\phi_\kappa(\pm L\boldsymbol{u}) = \pm\tanh(L/\kappa)\boldsymbol{u}$, so the compactified endpoint pair is $\{-\tanh(L/\kappa)\boldsymbol{u},\, \tanh(L/\kappa)\boldsymbol{u}\}$. Of the two endpoint assignments against $J_\infty = \{-\boldsymbol{u}, \boldsymbol{u}\}$, the sign-matched one attains the minimum, since the crossed assignment costs $2\big(1 + \tanh(L/\kappa)\big)^2 \geq 2\big(1 - \tanh(L/\kappa)\big)^2$. The matched assignment gives

$$E^\phi(J_L, J_\infty)^2 = \big\|-\tanh(L/\kappa)\boldsymbol{u} + \boldsymbol{u}\big\|^2 + \big\|\tanh(L/\kappa)\boldsymbol{u} - \boldsymbol{u}\big\|^2 = 2\big(1 - \tanh(L/\kappa)\big)^2, \tag{14}$$

and $\tanh(L/\kappa) \to 1$ as $L \to \infty$. □

Remark B.12 (Completion interpretation). The radial map is a bijection from V onto $\mathbb{B}^\circ$, and the metric completion of the open unit ball under the ambient norm is the closed unit ball $\overline{\mathbb{B}} = \{\boldsymbol{x} : \|\boldsymbol{x}\| \leq 1\}$. The antipodal pair $\{-\boldsymbol{u}, \boldsymbol{u}\}$ with $\|\boldsymbol{u}\| = 1$ lies on the boundary of this completion and is the natural limit of finite-range joints whose range tends to infinity along the screw direction $\boldsymbol{u}$.

Proposition B.13 (Compactified metric)**.** *E^ϕ is a metric on joints of all five types, continuous joints included, and it keeps the fixed joint at $J_0 = \{0, 0\}$.*

Proof. Write $\Phi(J) = \{\phi_\kappa(\boldsymbol{z}^-), \phi_\kappa(\boldsymbol{z}^+)\}$ for a finite joint and $\Phi(J) = \{-\boldsymbol{\xi}/\|\boldsymbol{\xi}\|, \boldsymbol{\xi}/\|\boldsymbol{\xi}\|\}$ for a continuous one, so that $E^\phi(J_1, J_2) = E(\Phi(J_1), \Phi(J_2))$. Every image is an unordered pair in V, on which E is a metric by Theorem B.1. The map Φ is injective on unordered endpoint pairs. On finite joints it applies the bijection ϕ_κ to each endpoint, finite joints land in the open ball and

continuous joints on its boundary, and a boundary pair $\{\pm \boldsymbol{u}\}$ fixes the unit screw up to the sign that the unordered pair already absorbs. A metric pulled back along an injective map is a metric, and $\phi_\kappa(0) = 0$ gives $\Phi(J_0) = J_0$. □

Remark B.14 (Scale equivariance and dimensionlessness). Under the kinetic norm, rescaling every length by $s > 0$ maps $\boldsymbol{v} \mapsto s\boldsymbol{v}$ in each endpoint twist, $\boldsymbol{c} \mapsto s\boldsymbol{c}$, and $I \mapsto s^2 I$, so $\|\boldsymbol{z}\|_B \mapsto s\|\boldsymbol{z}\|_B$ and the raw E_B scales by s, which is the precise sense in which it is measured in meters. With κ rescaled alongside, a compactified endpoint $\boldsymbol{u} = \phi_\kappa(\boldsymbol{z})$ transforms as $(\boldsymbol{u}_\omega, \boldsymbol{u}_v) \mapsto (\boldsymbol{u}_\omega/s,\ \boldsymbol{u}_v)$, because the rotational block is divided by the meter-valued norm. Substituting a difference of compactified endpoints into equation 7 then gives

$$\frac{1}{m}\Big(\frac{\Delta \boldsymbol{u}_\omega}{s}\Big)^\top \big(s^2 I\big)\Big(\frac{\Delta \boldsymbol{u}_\omega}{s}\Big) + \Big\|\Delta \boldsymbol{u}_v + \frac{\Delta \boldsymbol{u}_\omega}{s} \times (s\,\boldsymbol{c})\Big\|^2 = \frac{1}{m}\,\Delta \boldsymbol{u}_\omega^\top I\,\Delta \boldsymbol{u}_\omega + \big\|\Delta \boldsymbol{u}_v + \Delta \boldsymbol{u}_\omega \times \boldsymbol{c}\big\|^2, \tag{15}$$

so the s^2 carried by the inertia tensor is cancelled by the two inverse factors of the rotational block, and the cross term cancels likewise. Hence E_B^ϕ is invariant under the rescaling, a pure number in $[0, 2]$, and the body enters it only through the relative weighting of error directions.

B.4 Interpretation of the distance

Material motion under the kinetic norm. The first result gives the exact physical interpretation of the kinetic inner product. Importantly, this result concerns the *linearized material motion field* induced by a twist discrepancy. It does not state that the Lie-algebra norm equals finite Euclidean displacement after exponentiation.

Let B be a rigid body with mass measure $dm(\boldsymbol{x})$, total mass $m = \int_B dm(\boldsymbol{x})$, and center of mass $\boldsymbol{c} = \frac{1}{m}\int_B \boldsymbol{x}\, dm(\boldsymbol{x})$. Its moment of inertia tensor about the center of mass is $I_C = \int_B \big(\|\boldsymbol{x} - \boldsymbol{c}\|^2 I_3 - (\boldsymbol{x} - \boldsymbol{c})(\boldsymbol{x} - \boldsymbol{c})^\top\big)\, dm(\boldsymbol{x})$. A twist discrepancy $\Delta \boldsymbol{z} = (\Delta \boldsymbol{\omega}, \Delta \boldsymbol{v})$ induces the linearized material-motion field $\boldsymbol{w}_{\Delta \boldsymbol{z}}(\boldsymbol{x}) = \Delta \boldsymbol{\omega} \times \boldsymbol{x} + \Delta \boldsymbol{v}$. For two finite placements $g_1, g_2 \in SE(3)$ of the body, the configuration discrepancy of §4.3 is $d(g_1, g_2)^2 = \int_B \|g_1 \boldsymbol{x} - g_2 \boldsymbol{x}\|^2\, dm(\boldsymbol{x})$, which depends on dm only through $(m, \boldsymbol{c}, I_C)$ (Kazerounian & Rastegar, 1992; Park, 1995), and the calibration of Figure 4 computes it on the actual $SE(3)$ displacements.

Theorem B.15 (Material-motion identity)**.** *The mass-weighted squared L_2 magnitude of the material-motion field is*

$$\int_B \|\boldsymbol{w}_{\Delta \boldsymbol{z}}(\boldsymbol{x})\|^2\, dm(\boldsymbol{x}) = \Delta \boldsymbol{\omega}^\top I_C \Delta \boldsymbol{\omega} + m\,\|\Delta \boldsymbol{v} + \Delta \boldsymbol{\omega} \times \boldsymbol{c}\|^2\,. \tag{16}$$

Consequently, the RMS material-motion discrepancy is

$$d_{\mathrm{RMS}}^{\mathrm{lin}}(\Delta \boldsymbol{z}) = \left(\frac{1}{m}\int_B \|\boldsymbol{w}_{\Delta \boldsymbol{z}}(\boldsymbol{x})\|^2\, dm(\boldsymbol{x})\right)^{1/2} = \left[\frac{1}{m}\Delta \boldsymbol{\omega}^\top I_C \Delta \boldsymbol{\omega} + \|\Delta \boldsymbol{v} + \Delta \boldsymbol{\omega} \times \boldsymbol{c}\|^2\right]^{1/2}. \tag{17}$$

Proof. Write $\boldsymbol{x} = \boldsymbol{c} + \boldsymbol{y}$ with $\boldsymbol{y} = \boldsymbol{x} - \boldsymbol{c}$, so that $\int_B \boldsymbol{y}\, dm(\boldsymbol{x}) = 0$ by definition of the center of mass. The field becomes $\boldsymbol{w}_{\Delta \boldsymbol{z}}(\boldsymbol{x}) = \boldsymbol{u} + \Delta \boldsymbol{\omega} \times \boldsymbol{y}$ with $\boldsymbol{u} = \Delta \boldsymbol{v} + \Delta \boldsymbol{\omega} \times \boldsymbol{c}$, and hence

$$\int_B \|\boldsymbol{w}_{\Delta \boldsymbol{z}}(\boldsymbol{x})\|^2\, dm(\boldsymbol{x}) = m\|\boldsymbol{u}\|^2 + 2\boldsymbol{u}^\top\Big(\Delta \boldsymbol{\omega} \times \int_B \boldsymbol{y}\, dm(\boldsymbol{x})\Big) + \int_B \|\Delta \boldsymbol{\omega} \times \boldsymbol{y}\|^2\, dm(\boldsymbol{x}). \tag{18}$$

The middle term vanishes. For the last term, the vector identity $\|\boldsymbol{a} \times \boldsymbol{y}\|^2 = \boldsymbol{a}^\top\big(\|\boldsymbol{y}\|^2 I_3 - \boldsymbol{y}\boldsymbol{y}^\top\big)\boldsymbol{a}$ gives $\int_B \|\Delta \boldsymbol{\omega} \times \boldsymbol{y}\|^2\, dm(\boldsymbol{x}) = \Delta \boldsymbol{\omega}^\top I_C \Delta \boldsymbol{\omega}$, proving equation 16. Dividing by m and taking the square root gives equation 17. □

Remark B.16. If the body measure is normalized so that $m = 1$, equation 17 directly becomes $\|\Delta \boldsymbol{z}\|_B^2 = \Delta\boldsymbol{\omega}^\top I_C \Delta\boldsymbol{\omega} + \|\Delta\boldsymbol{v} + \Delta\boldsymbol{\omega} \times \boldsymbol{c}\|^2$. For an unnormalized physical mass measure, equation 16 is a mass-weighted L_2 norm and equation 17 the corresponding RMS quantity.

Endpoints and the full motion segment. The second result holds for any inner product. For a finite one-DOF joint, the endpoint-twist representation determines a line segment in twist-coordinate space. After selecting the endpoint assignment that realizes E, let $\boldsymbol{\Delta}^- = \boldsymbol{z}_1^- - \boldsymbol{z}_2^-$ and $\boldsymbol{\Delta}^+ = \boldsymbol{z}_1^+ - \boldsymbol{z}_2^+$, and write $\boldsymbol{\Delta}(t) = (1-t)\boldsymbol{\Delta}^- + t\boldsymbol{\Delta}^+$ for $t \in [0,1]$ for the linearly parameterized discrepancy along the two motion segments. Its RMS trajectory discrepancy in twist space is $D_{\mathrm{path}}^2 = \int_0^1 \|\boldsymbol{\Delta}(t)\|^2 \, dt$.

Theorem B.17 (Endpoint-to-path equivalence)**.** *For the optimal endpoint assignment,*

$$\frac{1}{\sqrt{6}} E(J_1, J_2) \le D_{\mathrm{path}} \le \frac{1}{\sqrt{2}} E(J_1, J_2). \tag{19}$$

Thus the endpoint metric and the L_2 discrepancy of the complete Lie-algebra motion segments are equivalent up to universal constants.

Proof. For the selected assignment, $E(J_1, J_2)^2 = \|\boldsymbol{\Delta}^-\|^2 + \|\boldsymbol{\Delta}^+\|^2$. Expanding D_{path}^2 and using $\int_0^1 (1-t)^2 dt = \int_0^1 t^2 dt = \int_0^1 2t(1-t) dt = \frac{1}{3}$,

$$D_{\mathrm{path}}^2 = \frac{1}{3}\left(\|\boldsymbol{\Delta}^-\|^2 + \|\boldsymbol{\Delta}^+\|^2 + \langle \boldsymbol{\Delta}^-, \boldsymbol{\Delta}^+ \rangle\right). \tag{20}$$

By Cauchy–Schwarz, $|\langle \boldsymbol{\Delta}^-, \boldsymbol{\Delta}^+ \rangle| \le \|\boldsymbol{\Delta}^-\| \|\boldsymbol{\Delta}^+\| \le \frac{1}{2}\big(\|\boldsymbol{\Delta}^-\|^2 + \|\boldsymbol{\Delta}^+\|^2\big) = \frac{1}{2} E(J_1, J_2)^2$. Substituting these bounds into equation 20 gives $\frac{1}{6} E(J_1, J_2)^2 \le D_{\mathrm{path}}^2 \le \frac{1}{2} E(J_1, J_2)^2$, and taking square roots proves the claim. □

Remark B.18. Theorem B.17 concerns the line segments traced by the joint representation in twist-coordinate space. It should not be interpreted as an exact equality with finite Euclidean material displacement after applying the exponential map to $SE(3)$.

B.5 The tree edit metric

A kinematic tree is a finite tree $T = (V, \mathcal{E}, \ell)$ with unlabeled nodes (links) and each edge e labeled by a joint $\ell(e) \in \mathcal{J}$, where $(\mathcal{J}, E)$ is the metric space of unordered endpoint pairs of Theorem B.1 or, with continuous joints admitted, its compactified counterpart $(\mathcal{J}, E^\phi)$ of Proposition B.13. The results below use only that the label distance is a metric with J_0 as a point, so they hold for either, and the reported $E_\alpha^{\phi,\mathrm{tree}}$ is the second case. Trees are unrooted, and all twists are expressed in a common canonical object frame, so edge labels do not change under re-rooting and the endpoint quotient already absorbs the per-joint sign. The zero-motion (fixed) joint is $J_0 = \{0, 0\}$.

Three edit operations transform trees, with costs

$$c_{\mathrm{sub}}(J, J') = E(J, J'), \qquad c_{\mathrm{con}}(J) = E(J, J_0), \qquad c_{\mathrm{exp}}(J) = E(J, J_0), \tag{21}$$

where a substitution replaces the label of one edge, a contraction removes an edge and merges its two endpoints, and an expansion splits a node into two nodes joined by a fresh edge labeled J. For finite edit sequences π we set

$$E^{\mathrm{tree}}(T_1, T_2) = \inf_{\pi \,:\, T_1 \rightsquigarrow T_2} \sum_{o \in \pi} c(o). \tag{22}$$

Call T_1 and T_2 *representation equivalent*, written $T_1 \sim T_2$, if they differ only by a tree isomorphism together with insertions or deletions of edges labeled exactly J_0. Contracting every J_0-labeled edge yields a canonical representative $\hat{T}$ of each class, in which every label satisfies $E(\ell(e), J_0) > 0$.

Proposition B.19 (Consistency with the joint metric). *For all $J_1, J_2 \in \mathcal{J}$,*

$$c_{\mathrm{sub}}(J_1, J_2) \le c_{\mathrm{con}}(J_1) + c_{\mathrm{exp}}(J_2). \tag{23}$$

Deleting a joint and inserting another is never cheaper than substituting directly.

Proof. This is the triangle inequality of E through J_0:

$$E(J_1, J_2) \le E(J_1, J_0) + E(J_0, J_2). \tag{24}$$

□

Theorem B.20 (Tree edit metric). *E^{tree} is a pseudometric on kinematic trees, and a metric on the equivalence classes modulo $\sim$.*

Proof. We verify the four properties in turn.

Non-negativity. Every operation has cost ≥ 0, hence so does every edit sequence, and so does the infimum.

Symmetry. The operations are invertible. Contraction and expansion are mutually inverse with the identical cost $E(J, J_0)$, and the inverse of a substitution is a substitution with $E(J', J) = E(J, J')$. Reversing an edit sequence therefore preserves its total cost, so the two infima agree.

Triangle inequality. Fix $\delta > 0$ and choose sequences $\pi_{12} : T_1 \rightsquigarrow T_2$ and $\pi_{23} : T_2 \rightsquigarrow T_3$ with

$$C(\pi_{12}) \le E^{\mathrm{tree}}(T_1, T_2) + \delta/2, \qquad C(\pi_{23}) \le E^{\mathrm{tree}}(T_2, T_3) + \delta/2. \tag{25}$$

Their concatenation edits T_1 into T_3 at cost $C(\pi_{12}) + C(\pi_{23})$, so

$$E^{\mathrm{tree}}(T_1, T_3) \le E^{\mathrm{tree}}(T_1, T_2) + E^{\mathrm{tree}}(T_2, T_3) + \delta, \tag{26}$$

and $\delta \to 0$ gives the claim.

Invariance under $\sim$. Inserting or deleting a J_0-labeled edge is an expansion or contraction of cost $E(J_0, J_0) = 0$, and isomorphisms are free because nodes carry no labels. Hence $T_1 \sim T_1'$ implies $E^{\mathrm{tree}}(T_1, T_2) = E^{\mathrm{tree}}(T_1', T_2)$ for every T_2, and E^{tree} descends to the quotient.

Identity of indiscernibles on the quotient. If $[T_1] = [T_2]$, the invariance step gives $E^{\mathrm{tree}}(T_1, T_2) = 0$. For the converse, invariance lets us take $T_1 = \hat{T}_1$ and $T_2 = \hat{T}_2$. Along an edit sequence $\pi : \hat{T}_1 \rightsquigarrow \hat{T}_2$, call an edge *original* if it descends, possibly relabeled, from an edge of $\hat{T}_1$ that no contraction has removed, and *new* if an expansion created it. Contracting the new edges of the current tree yields $\hat{T}_1$ with the original edges removed so far contracted, the surviving original edges corresponding. This holds before the first operation, and every operation preserves it. A substitution leaves the structure alone, an expansion or the contraction of a new edge leaves the new-edge contraction unchanged, and the contraction of an original edge contracts both sides by the same edge. At the end, let M pair each surviving edge of $\hat{T}_1$ with its final edge in $\hat{T}_2$, and let U_1 and U_2 collect the unmatched edges of $\hat{T}_1$ and $\hat{T}_2$. The invariant says that M is an isomorphism of the unlabeled trees $\hat{T}_1/U_1$ and $\hat{T}_2/U_2$. A matched edge moves from $\ell(e)$ to $\ell(e')$ by substitutions, an edge of U_1 is relabeled and then contracted, and an edge of U_2 is expanded and then relabeled, so the triangle inequality of E gives

$$C(\pi) \;\ge\; \sum_{(e,e') \in M} E\big(\ell(e), \ell(e')\big) + \sum_{e \in U_1 \cup U_2} E\big(\ell(e), J_0\big). \tag{27}$$

Call a triple (M, U_1, U_2) with this isomorphism property *admissible*. There are finitely many, so the right-hand side has a minimum over them that does not depend on π. Suppose $[T_1] \ne [T_2]$. If $U_1 \cup U_2$ is nonempty, the bound pays $E(\ell(e), J_0) > 0$ for an unmatched edge, since canonical representatives carry no J_0 label. If it is empty, M is an isomorphism of $\hat{T}_1$ onto $\hat{T}_2$ as unlabeled trees, and it cannot preserve every label, since a label-preserving isomorphism would give $[T_1] = [T_2]$, so some matched pair pays $E(\ell(e), \ell(e')) > 0$. Every admissible triple therefore has a positive bound, their minimum is a positive constant, and $E^{\mathrm{tree}}(T_1, T_2) > 0$. □

Remark B.21 (Fixed-boundary continuity at the tree level). Let T_0 contain an exact fixed edge between two links and let T_ϵ be the same tree with that edge relabeled $J_\epsilon = \{0, \epsilon\boldsymbol{\xi}\}$. Contracting the edge of T_ϵ costs $E(J_\epsilon, J_0) = |\epsilon| \, \|\boldsymbol{\xi}\|$ and produces a tree equivalent to T_0, so

$$E^{\mathrm{tree}}([T_\epsilon], [T_0]) \leq |\epsilon| \, \|\boldsymbol{\xi}\| \longrightarrow 0. \tag{28}$$

The continuity of Proposition B.10 therefore survives the lift. A scheme that charges a constant λ per unmatched joint instead jumps by λ when a joint's range crosses zero.

Exact computation. Contracting exact fixed edges and then solving the optimal assignment of moving joints, with E as the pairwise cost and $E(J, J_0)$ as the cost of leaving a joint unassigned, drops the requirement that the matching be realizable by an edit path. This can only lower the total, so the relaxation lower-bounds E^{tree}, and whenever its optimal assignment follows an isomorphism of the contracted trees the bound is tight, which certifies the exact value at the cost of one linear-time check. Kinematic trees are small, so the uncertified cases yield to a search over realizable matchings in milliseconds per pair. On the trees and corruptions of Figure 5 the certificate covers 91% of predictions, and the exact and relaxed values agree on 92% of them and on every structured sweep. The remaining gaps have median relative size 7%, and the two scores rank predictions with Kendall $\tau = 0.994$. The relaxation also keeps the score polynomial-time for arbitrary inputs, where exact unordered tree edit distance is NP-hard in the worst case.

Theorem B.22 (Fixed-skeleton weighted tree metric)**.** *Fix a ground-truth object with movable joints indexed by $k = 1, \ldots, K$ and ground-truth child bodies $B_1, \ldots, B_K$. On candidate assignments $\boldsymbol{J} = (J_1, \ldots, J_K)$ of one joint per ground-truth edge, each factor taken modulo the endpoint swap, define*

$$E_B^{\text{fixed-tree}}(\boldsymbol{J}, \boldsymbol{J}')^2 \;=\; \sum_{k=1}^{K} E_{B_k}(J_k, J_k')^2. \tag{29}$$

Then $E_B^{\text{fixed-tree}}$ is a metric on the product space of candidate assignments.

Proof. Each B_k is fixed by the ground truth, so $\|\cdot\|_{B_k}$ is a fixed inner-product norm and E_{B_k} is a metric by Theorem B.1. Symmetry and identity of indiscernibles hold coordinatewise. For the triangle inequality, $E_{B_k}(J_k, J_k'') \leq E_{B_k}(J_k, J_k') + E_{B_k}(J_k', J_k'')$ for every k, and Minkowski's inequality on the K-vectors of per-edge distances gives the claim. □

Remark B.23 (Weighted variant with topology edits). Theorem B.20 assumes the single fixed inner product of this section, and Theorem B.22 covers the per-edge weighting when the ground truth fixes the skeleton and the correspondence. The remaining case, topology-editing distances whose edit costs use edge-dependent bodies, falls outside both theorems, since contracting or expanding an edge changes which body weighs which joint. We use this last variant only as a score, without claiming the metric axioms.

C Additional ablation studies

This section expands the ablations summarized in §5.2.

C.1 Single-parameter perturbations against the component scores

Figure 6 sweeps one parameter of one ground-truth joint at a time and scores every step under E_α and under the component template. E_α responds smoothly and monotonically to every physical error and is exactly zero on pure gauge changes (the fixed-joint axis and origin sweeps, the prismatic origin sweep); on a prismatic axis error it follows the chord, $E_\alpha = \alpha\sqrt{(l^-)^2 + (l^+)^2} \cdot 2\sin(\theta/2)$. Each failure of Table 3 is visible as its own panel. The revolute origin score diverges as the axes align, so the success indicator is false even at zero error; the type error is a 0/1 cliff that E_α crosses continuously in the fixed → revolute column; the limit scores couple to the axis error through the motion vector $\boldsymbol{m} = \boldsymbol{a}\,(l^+ - l^-)$; and a fixed joint has a zero axis vector, for which the axis score of §3.2 is $0/0$, so once the denominator is regularized it evaluates to $\arccos 0 = \pi/2$ against itself

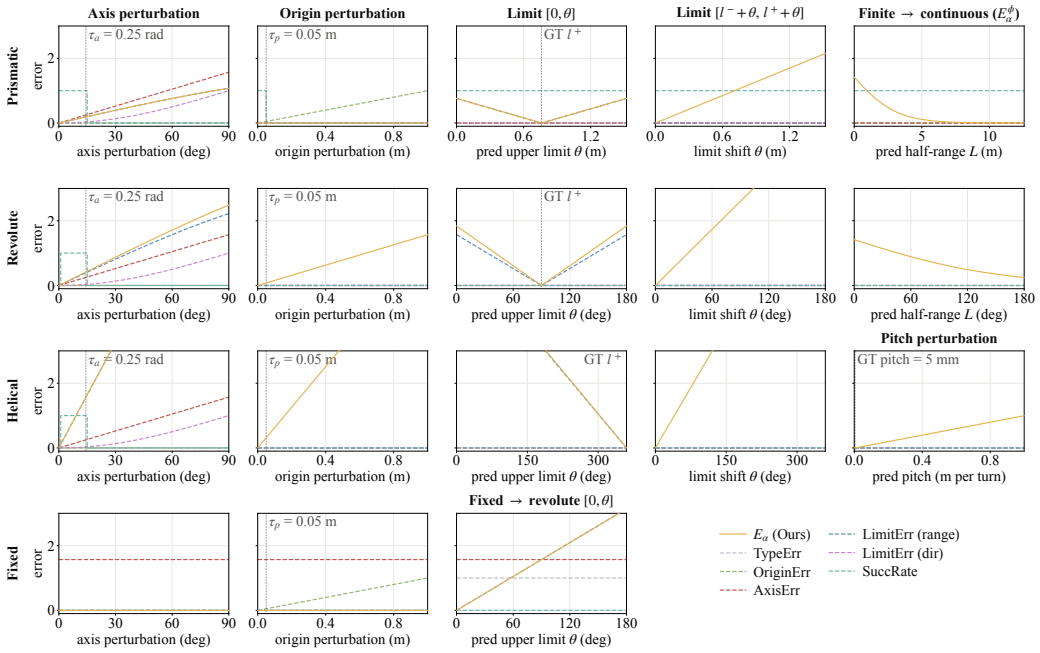


Figure 6: Controlled single-parameter perturbations of ground-truth library joints, one row block per joint type; the helical row composes the prismatic + continuous encoding of a screwed bulb into one screw (pitch 5 mm per turn). Solid amber is E_α ($\alpha = 1$; the finite → continuous column uses E_α^ϕ, $\kappa = \pi$), dashed curves are the component scores of §3.2 with the published thresholds $\tau_a = 0.25$ rad, $\tau_p = 0.05$ m.

and fixed joints can never be "successful". The pitch column has no baseline curve left to fail: the tuple taxonomy cannot express a helical joint, both sides reduce to the same revolute tuple, and every component score sits at zero for arbitrarily wrong pitch, while E_α grows exactly as the pitch discrepancy per turn.

C.2 Split-norm weight

Figure 7 documents how the choice of α moves the split-norm scores. The weight rescales the translational contribution: prismatic errors scale linearly with α, while revolute axis errors are nearly α-independent, their endpoint difference being rotation-dominated. The prismatic origin panel is flat at zero for the whole family, so the gauge invariance of Proposition B.3 does not depend on the choice of α. The compactified curves are bounded and saturate, and they preserve the raw curves' ordering in α, so the compactification does not scramble the underlying comparisons. No choice of α makes the split norm track physical size; that is the role of the kinetic-energy form, calibrated in Figure 4.

C.3 Sweep shapes under both inner products

Figure 8 repeats the perturbation sweeps under both inner products of §4.1. Both are smooth, monotone, and gauge-invariant on every sweep, so the qualitative properties behind Table 2 do not depend on the inner-product choice. On purely translational endpoint differences the two coincide exactly, since $\|(0, \Delta \boldsymbol{v})\|_B = \|\Delta \boldsymbol{v}\| = \|(0, \Delta \boldsymbol{v})\|_\alpha$ at $\alpha = 1$, which is why the prismatic panels overlap. The curves separate where rotation enters: the kinetic norm weighs a rotational error by the moving link's moments about the axis, the split norm weighs it identically for every link, and the two agree exactly on a centered link with isotropic inertia of gyration radius $1/\alpha$ (§4.1), the formal sense in which a global α assumes every link shares one characteristic size.

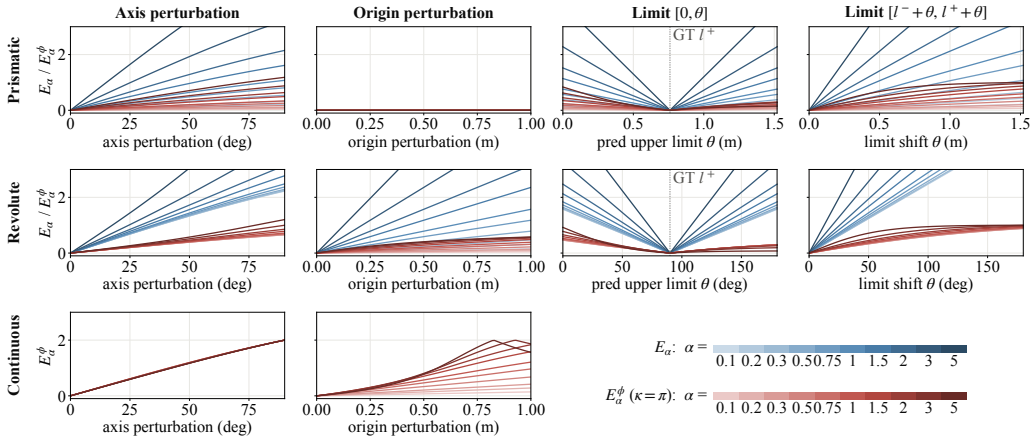


Figure 7: Sensitivity to the split-norm weight $\alpha \in \{0.1, \ldots, 5\}$, two decades around the default $\alpha = 1$. Blue curves are the raw E_α, red curves the compactified E^ϕ_α ($\kappa = \pi$), lighter for smaller α. Rows: representative prismatic, revolute, and continuous joints. Columns: axis rotation ($0..90°$), origin translation ($0..1$ m), the predicted upper limit $[0, \theta]$, and a limit shift $[l^- + \theta,\ l^+ + \theta]$; the continuous row admits only the compactified score.

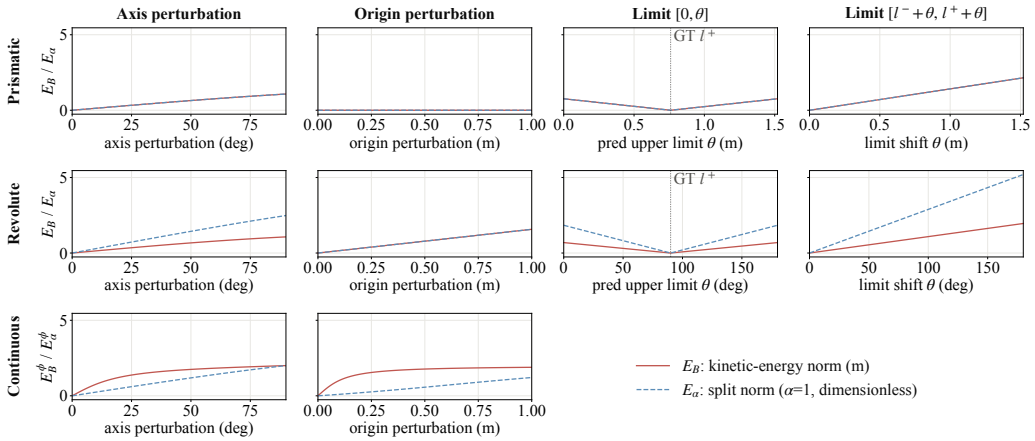


Figure 8: The sweeps of Figure 7 scored under both inner products: solid brick is the kinetic-energy E_B (meters), dashed blue the split-norm E_α ($\alpha = 1$, dimensionless). The two carry different units, so only shapes may be compared, not heights. The continuous row uses the compactified score, with ϕ normalized by each curve's own norm.

C.4 Object scale

Figure 9 isolates object size. The kinetic-energy score is exactly scale-covariant, $E_B(s) = s\,E_B(1)$, a slope-one line in every panel (uniform scaling sends $m \to s^3 m$, $\boldsymbol{c} \to s\,\boldsymbol{c}$, $I \to s^5 I$), matching its reading as RMS material displacement: the same angular error moves twice the material twice as far on a twice-as-large object. The split norm is not scale-covariant. Its rotational term is scale-free, so E_α flattens wherever the error has an angular component and follows s only where it is purely translational. A fixed global α therefore under-weights errors on large objects and over-weights them on small ones relative to physical displacement, the controlled counterpart of the calibration gap in Figure 4.

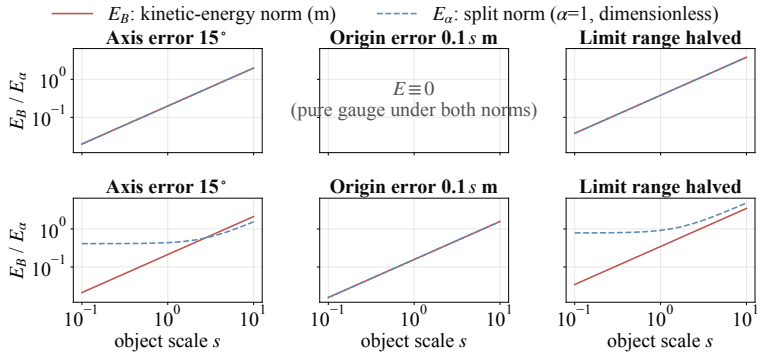


Figure 9: The same representation-level errors applied to the same objects uniformly scaled by $s \in [0.1, 10]$ (log–log axes): a fixed 15° axis rotation, an origin offset of $0.1\,s$ m, and a halved motion range. Solid brick is E_B, dashed blue E_α. The prismatic origin panel carries no curves because that error is pure gauge under both norms at every scale.

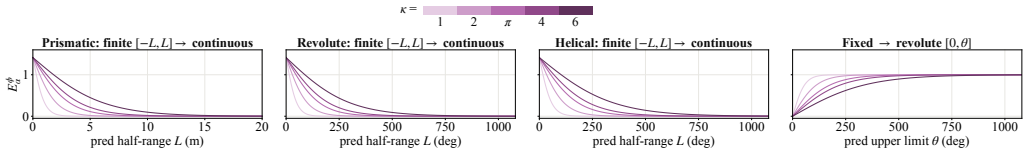


Figure 10: Sweep of the compactification scale κ. Curves show E_α^ϕ ($\alpha = 1$) for $\kappa \in \{1, 2, \pi, 4, 6\}$, lighter for smaller κ. Left three panels: a ground-truth prismatic, revolute, or helical joint of unbounded range ($l^\pm = \pm\infty$, the continuous case of each type) against a finite prediction $[-L, L]$ as L grows. Right: a welded ground-truth joint against a revolute prediction whose range $[0, \theta]$ opens from zero.

C.5 Compactification scale

Figure 10 sweeps κ around the default π. Against an unbounded ground truth the closed form is $E^\phi = \sqrt{2}\,(1 - \tanh(L\|\boldsymbol{\xi}\|/\kappa))$ (Proposition B.11). Every curve starts at $\sqrt{2}$, where the finite joint is the welded joint, and decays monotonically to zero, so an unbounded joint is literally the limit of a growing finite range, with no formula switch. The fourth panel shows the other boundary, $E^\phi = \tanh(\theta\|\boldsymbol{\xi}\|/\kappa)$ rising from zero, the compactified form of Proposition B.10. Across all panels κ acts as a pure horizontal dilation, with half decay near $L\|\boldsymbol{\xi}\| \approx 0.55\,\kappa$, and the default $\kappa = \pi$ keeps ranges up to a full turn in the responsive part of the curve.

Figure 11 checks the default against the benchmark's own joints. The map ϕ reserves the unit sphere for continuous joints, so finite joints must stay clearly inside it, and at $\kappa = \pi$ they do: no finite ground-truth joint saturates under either norm, with 95th-percentile radii of 0.40 (kinetic) and 0.78 (split) on revolute joints and 0.24 on prismatic ones, whereas $\kappa = 1$ already pushes 18% of the split-norm revolute joints past 0.99. The default keeps every finite joint in the responsive part of $\tanh$ without crowding the boundary reserved for continuous joints.

C.6 Endpoint sampling

Figure 12 evaluates denser sampling of the linearized motion segment against the two-endpoint representation. In every panel the curves collapse. Because $\|\boldsymbol{\Delta}(t)\|^2$ is quadratic in t, three points already integrate it exactly, and $n \geq 3$ is indistinguishable from the $n \to \infty$ limit D_{path}. Pooled over all sweeps, the rank correlation with E_2 is at least 0.996 for every n, and the ratio D_{path}/E_2 stays inside the band $[1/\sqrt{3},\ 1]$ of Theorem B.17. Denser sampling is a bounded monotone reparameterization of the endpoint metric that costs $n/2$ times more and adds no discriminative information. A continuous joint has no segment to sample, which is what the compactification handles.

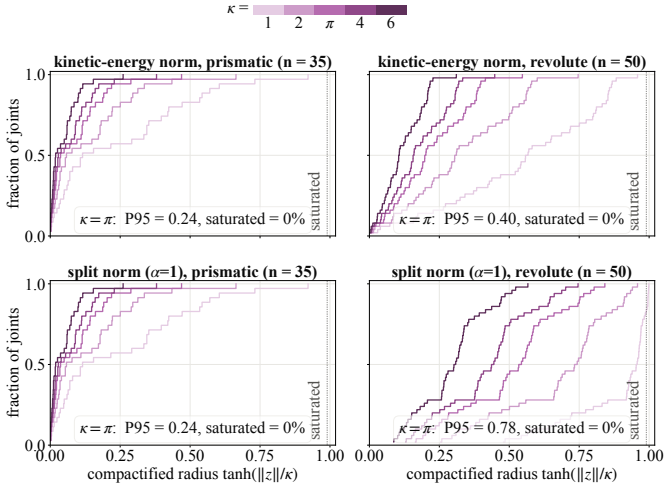


Figure 11: Compactified endpoint radius $\tanh(\|\boldsymbol{z}\|/\kappa)$ of the finite ground-truth joints of the evaluation set (per joint the larger of its two endpoints), as a CDF for each $\kappa \in \{1, 2, \pi, 4, 6\}$, lighter for smaller κ. Rows: the norm inside ϕ (kinetic-energy, split); columns: joint type. The dotted line marks radius 0.99, where $\tanh$ saturation would begin to collapse resolution; each panel is annotated with the $\kappa = \pi$ statistics.

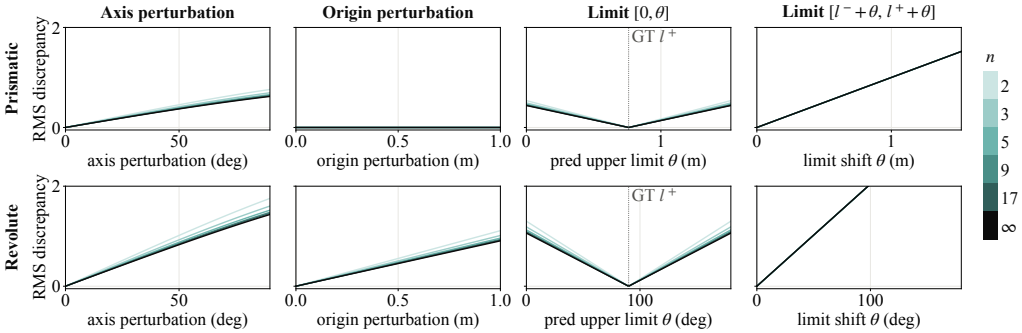


Figure 12: E_n compares n uniform samples of the linearized segments under the same $\mathbb{Z}_2$ quotient, in RMS convention so that $E_2 = E_\alpha/\sqrt{2}$ is the endpoint metric. Rows: prismatic and revolute joints. Columns: axis, origin, and limit sweeps. Darker curves use more samples; the dashed line is the closed-form $n \to \infty$ limit D_{path} of Theorem B.17.

D Additional evaluation results

This section records the composition of the evaluation set (Appendix D.1), discusses the entries of Table 4 in detail (Appendix D.2), reports the evaluation counts behind them (Appendix D.3), and repeats the comparison on the objects shared by all methods (Appendix D.4).

D.1 The evaluation set

The 200 objects of Table 4 are chosen from **ArticulateArena-20K** independently of every method.

Coverage. The evaluation set spans all 15 supercategories of Table 9 and 113 of the 628 categories, so no part of the taxonomy is left out, and within that coverage it concentrates on the supercategories that robots interact with most often. Kitchen & Cooking, Electronics, Containers, Architectural Fixtures, and Tools contribute 28, 26, 25, 22, and 20 objects, together more than half of the set, because doors, drawers, appliances, containers, and hand tools are the articulated objects a manipulation policy meets first, while the two smallest supercategories contribute three objects each.

Size. The selection was made before any of the eight methods was run, and the same 200-object list is presented to all of them. Each method is scored on the objects it generates, with coverage reported in Table 7 and controlled for on the common set in Appendix D.4. Two hundred objects support the claims Table 4 makes, as the bootstrap intervals of the next paragraph show, while keeping the comparison affordable, since one row means running a published method end to end and several of the methods are generation pipelines. The library is released with verified kinematics, so larger or differently stratified evaluations can be built on it.

What the evaluation set separates. The claims read off Table 4 are coarse, that the quotient scores induce different orderings on these outputs than the component protocol and that Articulate-Anything, fifth on E_B among the methods that predict their own motion limits, has the lowest object-level $E_\alpha^{\phi,\mathrm{tree}}$. Table 6 reports the object-level score with 95% percentile bootstrap intervals over objects. The intervals have half-widths of 0.06–0.09 on means of 0.39–0.59, the interval of the leading method overlaps those of three of the other seven and is disjoint from the remaining four, and a paired bootstrap leaves the gap between the top two undecided. The order of the three leading methods is therefore left open, and the difference in orderings does not rest on any difference below the resolution of the evaluation set.

Table 6: The object-level score $E_\alpha^{\phi,\mathrm{tree}}$ with 95% percentile bootstrap intervals over objects, on the 200-object evaluation set. The resampling unit is the object, not the joint, and each method is resampled within its own valid outputs. A paired bootstrap over the 180 objects that the two leading methods both score puts their difference at -0.039 with a 95% interval of $[-0.091, 0.012]$. Articulate AnyMesh† carries the ground-truth-imputed limits of Table 4 and covers 96 objects, so its interval is wider than the others.

Method	$E_\alpha^{\phi,\mathrm{tree}}$
Articraft	0.476 [0.415, 0.539]
Articulate AnyMesh†	0.528 [0.436, 0.624]
Articulate-Anything	0.392 [0.337, 0.449]
ArtLLM	0.585 [0.520, 0.651]
Ditto	0.552 [0.488, 0.617]
Particulate	0.419 [0.363, 0.476]
SPARK	0.520 [0.459, 0.583]
URDFormer	0.566 [0.493, 0.642]

D.2 Detailed comparison

Among the methods that predict their own motion limits in Table 4, Particulate has the lowest E_B (1.066 m), while Articraft has the highest success rate (17.1%). The two orderings differ most for Ditto, whose success rate is only 5.5% yet whose E_B (1.365 m) is third, behind Particulate and Articraft but ahead of SPARK (1.543 m) and Articulate-Anything (1.702 m).

Particulate has the lowest type, axis, and limit errors and the third-highest success rate, and it also has the lowest E_B among methods that predict their own limits; its origin error is larger than Articulate AnyMesh's, but for revolute joints an offset axis changes the induced motion of the ground-truth body, which E_B measures together with axis direction and range, so the component view and the joint metric need not agree. The two views disagree more sharply in the other direction. Articulate-Anything has the fourth-largest type error and no component-level lead, yet it has the lowest object-level $E_\alpha^{\phi,\mathrm{tree}}$ (0.392). Under the geometry-free split norm, Particulate has the lowest joint-level E_α^{ϕ} (0.465). Articulate AnyMesh has the lowest origin error but predicts no motion limits, so its quotient entries in Table 4 impute the ground-truth range and read as geometry-only diagnostics rather than competitive scores; its imputed E_B falls below Particulate's precisely because the range error is given away. These cases illustrate why component-wise scores do not induce a canonical ordering. Figure 13 shows three representative cases, two of which are objects on which URDFormer produced no valid output.

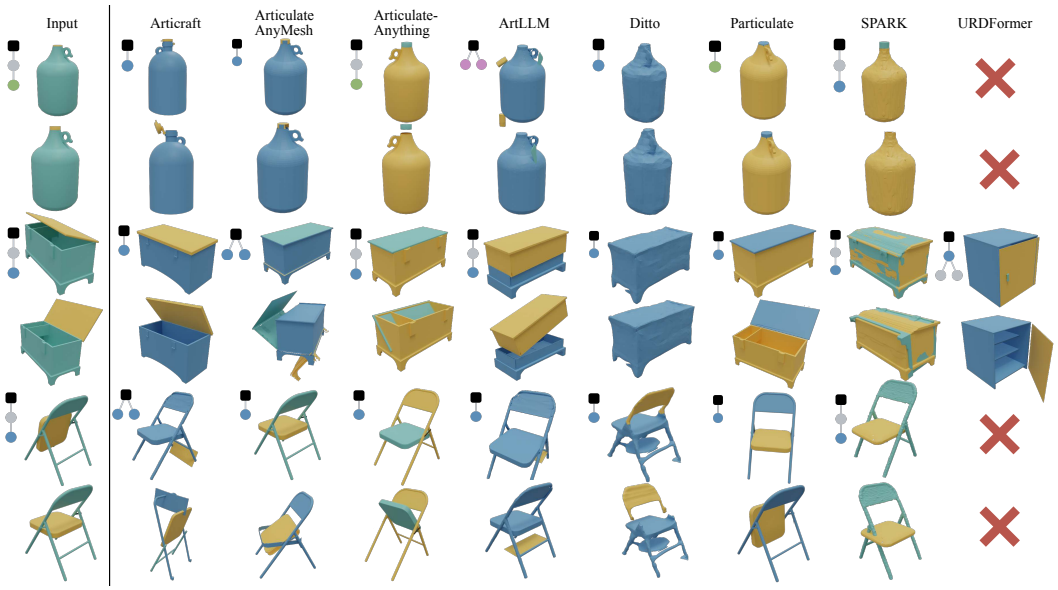


Figure 13: Qualitative comparison on three objects from the evaluation set of Table 4, one per row: a bottle, a box with a hinged lid, and a folding chair. The first column is the input and the remaining columns are the methods. In each panel the top render is the rest configuration and the bottom render actuates the joints, the ground-truth ones in the input column, and the kinematic tree a method returns is drawn beside its render, as in Figure 1. A red cross marks a generation failure, so URDFormer returns no valid asset for two of the three objects. The table aggregates all valid outputs, not only these three examples.

D.3 Evaluation coverage

Table 7 reports the evaluation counts omitted from the main table. Gen counts objects with a valid articulated output, and Pair counts the joint pairs we pair with finite predicted and ground-truth limits, which is where E_B is defined; the gap between the columns is the continuous joints and the limit-free predictions that the restriction drops. For each matched ground-truth joint, E_B uses the same ground-truth child-link geometry in the relevant comparison frame; predicted meshes never enter the norm. Particulate and ArtLLM use this geometry in their preprocessed frame with a geometry-only scale for canonical meters, while URDFormer uses the benchmark's camera registration.

Table 7: Coverage of the 200-object evaluation in Table 4. Gen counts objects, and Pair counts the joint pairs with finite predicted and ground-truth limits. †Counts under the ground-truth-imputed limits of Table 4.

Method	Gen	Pair
Articraft	199	158
Articulate AnyMesh†	96	79
Articulate-Anything	197	162
ArtLLM	186	105
Ditto	200	165
Particulate	182	150
SPARK	199	164
URDFormer	154	125

D.4 Common-set sensitivity

The main table reports each method on the objects for which it produces a valid output. To check whether this coverage difference changes the quotient ranking, we repeat the comparison on the 67 objects for which all eight methods produce a valid output. Table 8 reports means and percentile 95% bootstrap intervals over these object IDs, with the resampling unit the object rather than the

individual joint. Each score is reported where it is defined, E_B on the pairs whose predicted and ground-truth limits are both finite, a count that varies by method, and the three compactified scores on all 67 objects, since they admit continuous ground-truth joints. The ordering is stable across all four scores at the ends and unresolved in the middle. Among the methods that predict their own motion limits, Articulate-Anything is in the top two under every score, joined there by Ditto under the two kinetic scores and by Particulate under the two split-norm scores. ArtLLM is last under all four, and under the three joint-level scores its interval is disjoint from every other method; between the remaining methods the spacing is below the resolution of the intervals, so we do not read a rank off it. Articulate AnyMesh† is reported under the ground-truth-imputed limits of Table 4, so the orderings above cover only the methods that predict their own motion limits.

Table 8: Sensitivity analysis on the 67-object intersection shared by all eight methods. Entries are means with percentile 95% bootstrap intervals in brackets. E_B is computed on the pairs with finite predicted and ground-truth limits; the other three scores are dimensionless and defined on all 67 objects. †Articulate AnyMesh predicts no motion limits, so its limits are imputed from the ground truth as in Table 4 and its row reads as a geometry-only diagnostic rather than a competitive score.

Method	E_B (m)	E_B^ϕ	E_α^ϕ	$E_\alpha^{\phi,\mathrm{tree}}$
Articraft	1.217 [0.762, 1.927]	0.564 [0.420, 0.715]	0.702 [0.567, 0.846]	0.539 [0.423, 0.661]
Articulate AnyMesh†	1.039 [0.548, 1.836]	0.510 [0.360, 0.671]	0.590 [0.450, 0.739]	0.556 [0.436, 0.682]
Articulate-Anything	1.096 [0.549, 2.044]	0.412 [0.309, 0.527]	0.458 [0.358, 0.567]	0.303 [0.223, 0.389]
ArtLLM	3.249 [2.885, 3.616]	1.046 [0.954, 1.136]	1.183 [1.074, 1.283]	0.582 [0.468, 0.697]
Ditto	0.932 [0.512, 1.644]	0.460 [0.338, 0.590]	0.602 [0.496, 0.717]	0.544 [0.428, 0.665]
Particulate	1.176 [0.870, 1.500]	0.515 [0.409, 0.629]	0.465 [0.373, 0.556]	0.422 [0.335, 0.511]
SPARK	1.121 [0.592, 2.064]	0.461 [0.349, 0.584]	0.691 [0.589, 0.798]	0.498 [0.393, 0.607]
URDFormer	1.409 [0.983, 2.113]	0.567 [0.458, 0.685]	0.730 [0.636, 0.826]	0.557 [0.442, 0.677]

E THE ARTICULATEARENA-20K LIBRARY

This section documents how the library is assembled, while §4.4 summarizes its composition and statistics.

E.1 CONSTRUCTION

Collected assets. The library merges Articraft-10K (Zhou et al., 2026), Artiverse (Iliash et al., 2026), Articulated-Object-Code (Gao et al., 2026), PartNet-Mobility (Xiang et al., 2020), GAPartNet-AKB48, and a tail of smaller sources with a collection we generated ourselves. Figure 3(c) breaks each supercategory down by source. The public sources arrive with different directory layouts, mesh formats, and metadata schemas. Each asset is re-packaged into a common layout with one URDF per object and its meshes resolved, and every object receives a preview rendered with a shared camera and lighting, so that downstream tooling touches a single format.

Generated assets. The generated collection consists of objects we produced with several generative models and, in part, by hand. Each object is authored directly as parts connected by joints, so its kinematics is specified at creation rather than annotated afterwards, and it enters the library through the same normalization and verification as every collected asset.

Categorization. Every object carries a category label, read from per-object metadata where the source provides one and derived from the descriptive object name for Articraft-10K by truncating at connectives and dropping generic modifiers. Labels are normalized to a shared lowercase form so that variant spellings of one class merge across sources, yielding the 628 categories, which we group into the 15 supercategories of Table 9 (Appendix E.2). An LLM assisted this step by proposing a label for every object from its name, description, and renders, and the proposals were reviewed and corrected by the authors, so every final label passed a human check.

Quality control. We audit the library in a browser-based viewer that renders each object, drives every movable joint through its limits, and records per-object issue reports of missing textures, broken meshes or joints, and wrong orientation or scale. We use these reports to repair or exclude assets.

E.2 SUPERCATEGORY TAXONOMY

The categories of **ArticulateArena-20K** are grouped into 15 supercategories. Table 9 summarizes the mapping, and Figure 14 shows the long-tailed category distribution. The full category list ships with the library.

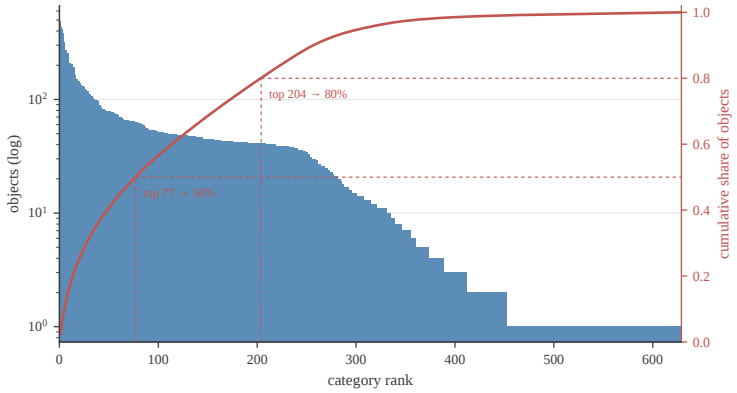


Figure 14: Category sizes of **ArticulateArena-20K** by rank (bars, log scale) and the cumulative share of objects (curve). The largest 77 categories hold half of the objects and the largest 204 hold 80%.

Table 9: Supercategories of **ArticulateArena-20K**, ordered by size, with the three largest categories of each and their object counts.

Supercategory	Objects	Categories	Largest categories
Storage Furniture	2,689	31	nightstand (490), chest_of_drawers (430), storage_furniture (381)
Mechanisms	2,324	80	branching_tree (77), prismatic_slider (63), yaw_axis_module (62)
Kitchen & Cooking	1,983	48	refrigerator (270), oven (210), dishwasher (190)
Tools	1,890	125	scissors (192), pliers (106), stapler (103)
Electronics	1,789	43	laptop (415), remote_control (138), switch (128)
Architectural Fixtures	1,645	48	faucet (254), door (207), window (150)
Mobility	1,262	70	cart (82), kick_scooter (79), car (54)
Tables & Seating	1,167	33	desk (321), table (132), chair (131)
Containers	1,080	22	bottle (163), bucket (143), trash_can (117)
Lighting	806	20	floor_lamp (112), task_lamp (82), ceiling_light_fixture (78)
Infrastructure	805	26	ferris_wheel (80), clock_tower (51), elevator (50)
Home Appliances	798	20	fan (98), washing_machine (88), ceiling_fan (77)
Personal & Household Items	756	22	eyeglasses (272), globe (118), clock (78)
Robotics	517	15	robotic_arm (92), opposed_twinslide_gripper (47), robotic_leg (47)
Recreation	466	25	fidget_toy (54), dj_equipment (53), playground_swing (42)
Total	19,977	628	