PIS: Pose-Interpolation Smoothness for Skinning Weight Refinement

Shaofeng Yin **Ting Lei** **Yang Liu***
Wangxuan Institute of Computer Technology, Peking University
yin_shaofeng@stu.pku.edu.cn
{ting_lei, yangliu}@pku.edu.cn

Abstract

Modern automatic-rigging predictors (UniRig, Make-It-Animatable, RigNet) place a skeleton on a 3D mesh in seconds, but the resulting skinning weights can animate poorly. Linear-blend skinning (LBS) is non-linear in pose space, so noisy weights amplify into visible candy-wrapper artifacts—the surface twisting and contracting at joint bends—and volume collapse. ARAP-style remedies measure per-pose rigidity rather than explicitly matching cross-pose midpoints. We propose *Pose-Interpolation Smoothness* (PIS), a self-supervised, training-free, test-time refiner: minimize the per-vertex gap between the deformation at the SO(3) midpoint of two random poses and the linear average of the two endpoint deformations. Polar projection of the arithmetic rotation mean realizes the exact geodesic midpoint in the bounded-angle regime; gradients are needed only with respect to the weights. Because PIS only re-tunes the weights LBS already consumes, it slots into any LBS-based pipeline—generative-3D outputs, automatic-rigging predictions, and the LBS pathways of game engines and AR/VR renderers—without touching shaders. On Objaverse-LVIS Balanced-102 (98 valid models, 51 LVIS categories), PIS reduces fragility (edge-stretch and normal-direction variances under random poses), ARAP energy, Jacobian anisotropy, and area distortion by 35.5%, 39.3%, 19.8%, and 43.2% on average; 93.9% of models improve on all four simultaneously, Wilcoxon $p < 10^{-10}$. Additional tests cover RigAnything and Make-It-Animatable initializers, direct motion preservation on 270 artist-painted rigs, and comparison with an independent SkinCells implementation. PIS changes weights only, not the skeleton or the runtime LBS operator.

1 Introduction

3D generation pipelines [70, 34, 52, 38, 40, 41, 56, 49, 21, 13] and scene representations, including neural radiance fields [45, 48] and Gaussian splatting [24], have substantially reduced the cost of static 3D content. The next step—taking a generated mesh and bringing it to life under animation—remains uneven. Modern automatic rigging predictors [71, 14, 75, 72, 39, 59] place a skeleton and propose initial skinning weights in seconds, but predicted weights can behave erratically on unfamiliar assets: insects, dragons, props, machinery. When such a mesh is driven by a real animation, the surface flickers, joints collapse, and limbs exhibit candy-wrapper artifacts. Noisy predicted skinning is an important obstacle between today's generative 3D output and animation-ready content for downstream avatar reconstruction [31, 26, 18, 36] and dynamic scene synthesis [73, 69].

This problem cannot be attributed to the predictor alone. Linear-blend skinning (LBS) [43], a widely used deformation operator in real-time renderers, is itself non-linear in pose space: rotating two joints to the average of their angles does not place a vertex at the average of the two endpoint

*Corresponding author. Code: https://github.com/Fugtemypt123/PIS.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

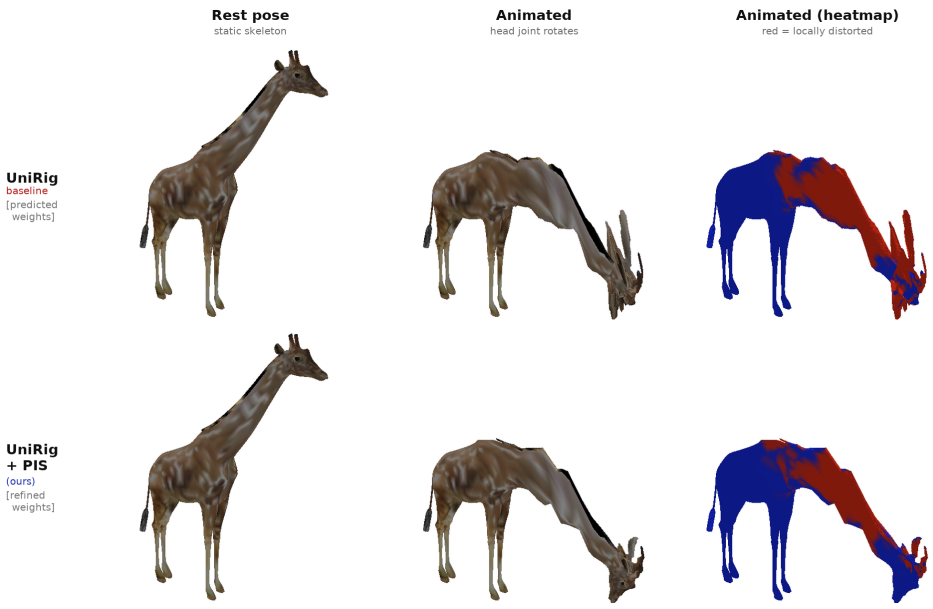

Figure 1: **Isolating candy-wrapper.** Only the head joint rotates. Top: UniRig drags body mass onto the moving joint. Bottom: PIS-refined weights—body stays pinned, head rotates cleanly. Columns: rest / animated / edge-stretch heatmap (red = distorted).

positions. Example-based skinning methods fit deformation parameters from example poses [47, 66]; with noisy predicted weights, mass leaking onto the wrong bone can amplify visible artifacts. The animation quality issue is therefore not solely a regression problem on the predictor side—it is also a deformation-quality problem on the LBS side. We target the weights while retaining the operator.

Three established directions address this problem, each with a specific limitation in our setting (generative-3D assets fed to LBS renderers). *(i) Replace the operator*: dual-quaternion skinning (DQS) [23, 30] blends rigid transformations using dual quaternions and is shipped in DCC tools and at Disney; it requires a DQS-capable runtime, whereas we retain existing LBS pathways. It also does not itself correct misassigned weights. *(ii) Optimize deformation energies*: ARAP [60] and pose-sampling weight refiners [12, 27] optimize deformation over poses; their objectives differ from explicitly matching midpoint deformations. In our matched experiment, ARAP-only optimization *worsens* held-out ARAP energy on 26% of test models. *(iii) Train a larger predictor* [39, 59]: improving regression L1 against artist weights does not directly constrain the cross-pose interpolation residual.

In contrast to these directions, we retain LBS and the predictor, and ask whether a fully self-supervised loss can directly target the non-linearity LBS exhibits between poses rather than within them. The proposed approach samples two random poses, takes their geodesic midpoint, and measures how far the actual mid-pose deformation drifts from the linear average of the two endpoints; this residual is minimized with respect to the weights. The procedure penalizes the curvature of each vertex's deformation curve, with the goal of reducing artifacts caused by noisy weights; valid rigid trajectories can also have curvature. The resulting loss, Pose-Interpolation Smoothness (PIS), is differentiable in the weights, requires no ground truth, no training data, and no reference motion, and leaves the LBS evaluation pipeline entirely unchanged; it is a drop-in test-time refiner. Figure 1 concretely demonstrates the gap PIS closes: with only the head joint rotated, UniRig's predicted weights drag body mass onto the moving joint and create a visible tear at the neck (top row); under the same skeleton and the same LBS evaluation, PIS-refined weights keep the body pinned and the head rotates cleanly (bottom row). The rightmost column visualizes per-edge stretch, with red marking high distortion—PIS reduces mean edge stretch by 80% on this asset.

PIS admits a second-order interpretation along pose-space geodesics (Sec. 3.3); per-pose ARAP instead measures spatial rigidity. These are different objectives, not a guarantee that every valid rigid trajectory should have zero PIS residual. We explicitly examine this distinction and test

motion preservation. Alongside the loss we use a per-vertex fragility statistic to localize deformation variability.

The contributions of this paper are as follows.

- A self-supervised, training-free, test-time loss (PIS) that directly targets the cross-pose non-linearity of LBS, supported by a lemma identifying it as a Hessian estimator and a proposition characterizing its minimizers.
- A per-vertex fragility statistic for localizing deformation variability (Sec. 3.5); separate rigid-invariant errors and direct trajectory diagnostics test deformation quality and motion preservation.
- Comprehensive evaluation on Objaverse-LVIS Balanced-102 (98 models, 51 categories), with 19.8–43.2% mean improvement across four geometric metrics and 93.9% of models improving on all four ($p < 10^{-10}$). A controlled study additionally shows that PIS does not move the per-(vertex,bone) L1 against artist GT—i.e., L1 agreement alone does not measure deformation quality.

2 Related Work

We distinguish sources of *input* weights, *operator-side* alternatives to LBS, and *weight-side* refiners. UniRig is a controlled primary initializer, not a methodological requirement; PIS accepts a fixed skeleton and an initial weight matrix.

(a) Predicted and analytic input weights. Learned predictors include RigNet [71], MoRig [72], UniRig [75], Make-It-Animatable [14], RigAnything [39], and Puppeteer [59]; TADA [37] generates animatable avatars. Classical methods include Pinocchio's automatic rigging and heat-based weights [3], BBW [19], and example-pose methods [47, 66, 28, 29, 20]. The latter explicitly use example deformations, so pose-aware weight optimization is not new. We refine UniRig, Distance-RBF, RigAnything, Make-It-Animatable, and valid Blender Bone Heat outputs; volumetric BBW applicability is audited separately.

(b) LBS operator-side alternatives—context, not direct competitors. Dual-quaternion skinning (DQS) [23] changes the runtime blend operator; enhanced DQS is used in Disney's *Frozen* [30]. PIS instead keeps the existing LBS evaluation graph and modifies weights offline. DQS and weight refinement are complementary: replacing the blend operator does not itself correct mass assigned to the wrong bones. We do not claim to replace DQS or to compare its runtime operator under an unchanged-LBS protocol.

(c) Weight refinement and deformation-energy background. ARAP [60], As-Killing-As-Possible [58], projective dynamics [4], and embedded deformation [62] provide deformation-energy background, not interchangeable skinning-weight solvers. Eisemann and Thiery [12] optimize LBS weights using sampled skeletal poses. SkinCells [27] learns sparse skinning through Voronoi-cell interfaces and deformation regularization. Our narrower contribution is the specific cross-pose midpoint residual, not pose-sampled refinement in general. We compare with ARAP, a blend-matrix-condition auxiliary regularizer (BMC), and Laplacian-only objectives under matched compute; a separate comparison implements the full published SkinCells formulation (Sec. 4.7). BMC is a separate baseline, not a SkinCells loss term.

(d) Pose-space deformation, neural skinning, background. Pose-space deformation [33], SCAPE [2], parametric models [42, 35, 54], and neural skinning [55, 8, 61, 50, 67, 76] offer alternative deformation representations. Physics-based approaches include PhysRig [74] (volumetric soft-body simulation) and PhysSkin [32] (learned continuous skinning fields for physics-based animation). PIS draws on Riemannian optimization on rotations [22, 46, 51, 1], classical mesh smoothing [64, 11], rotation interpolation [57], and the broader self-supervised [7, 17, 5, 16] and test-time adaptation [63, 65] traditions.

3 Method

3.1 LBS and the source of candy-wrapper

A pose is a tuple $\boldsymbol{\theta} = (R_1, \ldots, R_J)$ of J joint rotations R_j, applied along a kinematic chain via forward kinematics to give each bone j a global rest-to-pose rigid transform $A_j(\boldsymbol{\theta})$. Linear blend skinning then deforms each rest vertex $\mathbf{v}_i \in \mathbb{R}^3$ as a weighted average of the bone transforms,

$$\mathcal{S}_i(\boldsymbol{\theta}) \;=\; \sum_{j=1}^{J} w_{ij}\, A_j(\boldsymbol{\theta})\, \mathbf{v}_i, \qquad \mathbf{w}_i = (w_{i1}, \ldots, w_{iJ}) \in \Delta^{J-1}, \tag{1}$$

with weights on the simplex ($w_{ij} \geq 0$, $\sum_j w_{ij} = 1$). Eq. (1) is differentiable, fast, and is widely used in real-time renderers; predicted skinning weights [71, 14, 75] feed directly into it.

A two-bone example. Consider a vertex $\mathbf{v} = (0, 1, 0)$ shared equally ($w_1 = w_2 = 0.5$) between two bones at the origin: bone 1 at identity, bone 2 rotating by θ about the x-axis. LBS gives $\mathcal{S}(\theta) = (0, \frac{1+\cos\theta}{2}, \frac{\sin\theta}{2})$, whose distance from the joint axis is $\cos(\theta/2) < 1$—the vertex is pulled inward by $\sim \theta^2/8$ (29% at $90°$, full collapse at $180°$). The defect scales with the *square* of the rotation angle and appears whenever weight mass is split across bones with disagreeing rotations; concentrating mass on one bone eliminates it. Clean artist weights route mass mostly to one bone away from joints; noisy predicted weights leak everywhere, making the squared-angle term globally visible. The defect is thus second-order in joint angles and lives in how $\mathcal{S}_i$ *curves* between poses, not how it behaves at any single pose.

3.2 The PIS objective

Suppose for a moment that LBS were truly linear in pose: then for any two poses $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ and any midpoint pose $\boldsymbol{\theta}_{\text{mid}}$ between them, the deformation at the midpoint would equal the average of the deformations at the endpoints. The candy-wrapper analysis above says LBS *deviates* from this identity by exactly the squared-angle amount we want to suppress. We turn this into a loss.

Midpoint rotation via polar projection. We need a rotation halfway between $R_{1,j}$ and $R_{2,j}$ on $SO(3)$. For relative angle below π, the geodesic midpoint is $R_{1,j} \exp\big[\frac{1}{2} \log(R_{1,j}^\top R_{2,j})\big]$ [57, 46]. We realize the same midpoint by taking the polar factor of the arithmetic rotation mean:

$$\tfrac{1}{2}\big(R_{1,j} + R_{2,j}\big) \;=\; U_j\, \Sigma_j\, V_j^\top, \qquad \big[\boldsymbol{\theta}_{\text{mid}}\big]_j \;\triangleq\; U_j\, V_j^\top. \tag{2}$$

For relative angle below π, the arithmetic mean is nonsingular with positive determinant. Its polar factor $U_j V_j^\top$ is a proper rotation and *exactly* equals the geodesic midpoint, not merely a small-angle approximation (Appendix N). Polar projection is a convenient endpoint-symmetric, batched implementation. Sampled rotations are fixed inputs; PIS differentiates only with respect to the skinning weights, so differentiability through rotations is not required. Existing numerical controls on $10{,}000$ bounded rotation pairs give a maximum midpoint discrepancy of 2.98×10^{-8} rad.

The PIS residual. For a sampled pose pair $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ with midpoint $\boldsymbol{\theta}_{\text{mid}}$, define the per-vertex residual

$$\mathbf{r}_i \;=\; \mathcal{S}_i(\boldsymbol{\theta}_{\text{mid}}) \;-\; \tfrac{1}{2}[\mathcal{S}_i(\boldsymbol{\theta}_1) + \mathcal{S}_i(\boldsymbol{\theta}_2)] \;\in\; \mathbb{R}^3. \tag{3}$$

$\mathbf{r}_i$ is exactly the chord-vs-arc gap for vertex i along the trajectory connecting the two poses: it would be zero if the LBS curve were a straight line in pose space, and it grows quadratically with the joint-angle difference otherwise. The Pose-Interpolation Smoothness loss is the squared L^2 expectation of this residual over random pose pairs,

$$\mathcal{L}_{\text{PIS}}(\mathbf{w}) \;=\; \mathbb{E}_{(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \sim \mu}\left[\, \frac{1}{V} \sum_{i=1}^{V} \|\mathbf{r}_i\|^2 \,\right], \tag{4}$$

where μ samples bounded local joint rotations. In the supplementary motion tests, each endpoint rotation has magnitude at most $\theta_{\max} = 0.5$ rad, hence each relative angle is at most $2\theta_{\max} = 1$ rad and stays below the antipodal degeneracy at π. The second-order interpretation below concerns small pose-pair separation; the residual itself remains well-defined throughout this non-antipodal regime.

3.3 Second-order interpretation

The chord-vs-arc analysis can be made precise with a Taylor expansion. For any four-times continuously differentiable scalar function f, midpoint $\bar{x} = (x_1 + x_2)/2$, and separation $\Delta x = x_2 - x_1$, the symmetric finite-difference identity reads

$$f\big(\tfrac{x_1+x_2}{2}\big) - \tfrac{1}{2}\big[f(x_1) + f(x_2)\big] \;=\; -\tfrac{1}{8}\, f''(\bar{x})\,(\Delta x)^2 + O((\Delta x)^4). \tag{5}$$

Applying Eq. (5) component-wise along the trajectory from $\boldsymbol{\theta}_1$ to $\boldsymbol{\theta}_{\mathrm{mid}}$ to $\boldsymbol{\theta}_2$, with $\Delta\boldsymbol{\theta}$ the per-joint rotation differences, yields:

Lemma 1 (Second-order interpretation)**.** *The PIS residual satisfies*

$$\mathbf{r}_i \;=\; -\tfrac{1}{8}\,\mathbf{H}_i\big[\Delta\boldsymbol{\theta}, \Delta\boldsymbol{\theta}\big] + O(\|\Delta\boldsymbol{\theta}\|^4), \tag{6}$$

where $\mathbf{H}_i$ is the covariant second derivative of $\mathcal{S}_i$ at the midpoint, evaluated along the product-$SO(3)$ geodesic with tangent displacement $\Delta\boldsymbol{\theta}$. Consequently $\mathcal{L}_{\mathrm{PIS}} = \frac{1}{64V}\sum_i \mathbb{E}_\mu\big[\|\mathbf{H}_i[\Delta\boldsymbol{\theta}, \Delta\boldsymbol{\theta}]\|^2\big] + O(\theta_{\max}^6)$.

Proof sketch. For each coordinate, parameterize the product geodesic by $t \in [-1/2, 1/2]$ with midpoint $t = 0$. LBS with fixed skeleton is smooth along this path. Taylor-expand about zero through third order: averaging the endpoints cancels the odd terms and leaves $\frac{1}{8}f''(0) + O(\|\Delta\boldsymbol{\theta}\|^4)$. Subtracting this average gives Eq. (6). Squaring and summing over vertices yields the stated loss expansion, with the cross-term of order six. □

The lemma identifies PIS as a sample-based penalty on cross-pose curvature. It does *not* identify all curvature with artifacts. A correctly attached point rotating rigidly at radius ρ has a nonzero chord–arc gap $\rho[1 - \cos(\Delta\theta/2)]$. For a true single-bone rig ($J = 1$), the simplex fixes $w = 1$, so PIS has no admissible weight update. With several feasible bones, motion preservation is not guaranteed by the objective and must be tested directly (Sec. 4.6).

3.4 Structure of the PIS optimum

We characterize pure PIS with a fixed pose distribution and simplex-constrained weights. This idealized quadratic omits spatial regularization; it does not characterize the optimum of the complete parameterized loss.

PIS is a sum of squared residuals, but Eq. (3) is *linear* in $\mathbf{w}_i$ once the pose pair is fixed: the bones A_j enter through the same midpoint-vs-average expression for each j. Concretely, define for each bone the three-vector

$$\mathbf{d}_{ij}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \;=\; \big[A_j(\boldsymbol{\theta}_{\mathrm{mid}}) - \tfrac{1}{2}(A_j(\boldsymbol{\theta}_1) + A_j(\boldsymbol{\theta}_2))\big]\,\mathbf{v}_i, \tag{7}$$

which is bone j's own chord-vs-arc gap acting on $\mathbf{v}_i$. The PIS residual is then $\mathbf{r}_i = \sum_j w_{ij}\,\mathbf{d}_{ij} = \mathbf{D}_i\mathbf{w}_i$, with $\mathbf{D}_i \in \mathbb{R}^{3\times J}$. Squaring and taking the expectation gives a per-vertex quadratic in $\mathbf{w}_i$:

$$\mathcal{L}_i(\mathbf{w}_i) \;=\; \mathbf{w}_i^\top \mathbf{M}_i\,\mathbf{w}_i, \qquad \mathbf{M}_i \;=\; \mathbb{E}_\mu\big[\mathbf{D}_i^\top \mathbf{D}_i\big] \in \mathbb{R}^{J\times J}, \tag{8}$$

which is positive semidefinite and depends on the rest vertex, skeleton, and pose distribution, but not on the weights.

Proposition 1 (Pure-PIS simplex optimum)**.** *Assume the loss is $\mathcal{L}_i$ in Eq.* (8) *alone (no Laplacian prior, fixed pose distribution μ) with simplex constraints $\mathbf{w}_i \in \Delta^{J-1}$. With $\mathbf{M}_i$ as defined:* (i) *$\mathcal{L}_i(\mathbf{w}_i) = 0$ if and only if $\mathbf{w}_i \in \ker(\mathbf{M}_i)$.* (ii) *If $\mathbf{d}_{ij^\star} = 0$ almost surely under μ, then $\mathbf{e}_{j^\star}$ is a zero-loss global minimizer; uniqueness or preference for this minimizer does not follow.* (iii) *A feasible $\mathbf{w}^\star$ is a global minimizer if and only if $(\mathbf{M}_i\mathbf{w}^\star)_j \geq c$ for all j, with equality whenever $w_j^\star > 0$, where $c = (\mathbf{w}^\star)^\top \mathbf{M}_i \mathbf{w}^\star$.*

Proof sketch. (i) follows from $\mathbf{M}_i \succeq 0$. For (ii), $\mathbf{e}_{j^\star}^\top \mathbf{M}_i \mathbf{e}_{j^\star} = \mathbb{E}_\mu\|\mathbf{d}_{ij^\star}\|^2 = 0$; no feasible loss can be negative. For (iii), necessity follows by differentiating along the feasible segment from $\mathbf{w}^\star$ to each simplex vertex: $2[(\mathbf{M}_i\mathbf{w}^\star)_j - c] \geq 0$. Their $\mathbf{w}^\star$-weighted average is zero, so every active coordinate has equality. Conversely, for any feasible $\mathbf{w}$, $\mathcal{L}_i(\mathbf{w}) - \mathcal{L}_i(\mathbf{w}^\star) = 2(\mathbf{M}_i\mathbf{w}^\star)^\top(\mathbf{w} - \mathbf{w}^\star) + (\mathbf{w} - \mathbf{w}^\star)^\top \mathbf{M}_i(\mathbf{w} - \mathbf{w}^\star) \geq 0$. The first term is nonnegative by the stated inequalities and unit sums, and the second by positive semidefiniteness. □

Scope of the idealized result. The Laplacian prior couples vertices, while softmax enforces each vertex's simplex constraint. Proposition 1 does not imply a principal-eigenvector solution, one-bone dominance, or convergence to the same weights from different initializers.

Relation to ARAP energy. ARAP energy [60] measures how far a per-face *spatial* deformation Jacobian deviates from a rotation at each sampled pose. First-order here refers to spatial derivatives, not a first derivative with respect to pose. PIS instead compares three related poses and measures a pose-space interpolation residual. These are different regularizers; aggregating ARAP over many poses does not make it the PIS objective, nor does this distinction prove universal superiority. Under our matched optimization and held-out evaluation protocol, ARAP-only worsens held-out ARAP energy on 26% of models (Sec. 4.4).

3.5 Per-vertex fragility statistic for diagnosis

PIS provides a global scalar; we additionally define a per-vertex diagnostic that can be rendered directly on the mesh. For a sample of B poses, the edge-stretch and normal-direction variances per vertex i are defined as

$$f_i \;=\; \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} \mathrm{Var}_b \left(\frac{\|e_{ij}^{(b)}\|}{\|e_{ij}^{(0)}\|} \right) + \alpha \, \mathrm{Var}_b \angle\big(\mathbf{n}_i^{(b)}, \mathbf{n}_i^{(0)}\big), \tag{9}$$

where $\mathcal{N}(i)$ is the 1-ring of vertex i: edge-stretch variances are averaged over its incident edges. Superscript 0 denotes the rest pose, and b indexes sampled poses. We use f as a diagnostic and rendering color, not a training target or an artifact classifier. In particular, its normal-direction component can change under valid rigid rotation. The supplementary evaluation therefore uses separate rigid-motion-invariant surface errors and direct motion-preservation diagnostics (Sec. 4.6); a low fragility value alone does not establish correct articulation.

3.6 Optimization

We optimize $\mathbf{w}$ in unconstrained logit space and project to the simplex via softmax: $w_{ij} = \mathrm{softmax}_j(\ell_{ij})$, with ℓ initialized from the predictor's weights $\mathbf{w}^{(0)}$ via $\ell_{ij}^{(0)} = \log(w_{ij}^{(0)} + \epsilon)$ for a small $\epsilon > 0$. Each iteration samples $B = 16$ pose pairs with each endpoint's local joint rotation bounded by $\theta_{\max} = 0.5$ rad, computes PIS plus a Laplacian smoothness prior [64, 11] ($\lambda_{\mathrm{lap}} = 2.0$), and steps Adam [25] ($\mathrm{lr} = 0.05$) for $T = 500$ iterations (30–60 s on a single H200). Forward kinematics, LBS, and rendering use PyTorch3D [53]; full procedure in Algorithm 1 (Appendix B).

4 Experiments

4.1 Setup

Objaverse-LVIS Balanced-102. Existing resources serve different purposes: RigNet [71] supplies rigged characters, ShapeNet [6] supplies static shapes, and AMASS [44] supplies human motion. We evaluate broader category coverage with *Balanced-102*: 102 Objaverse [10, 9] assets, two per category across 51 LVIS [15] categories selected for articulated structure and Objaverse coverage. The mesh-validity sampling protocol and category breakdown are in Appendices L and M. $98/102$ assets admit valid UniRig outputs; all reported 98-asset coverage claims are conditional on these valid initializations.

Other splits. For Skin L1 against artist ground truth we additionally use (i) the official RigNet test split (270 models, used for the L1 floor analysis in Appendix I and the 202-model manifold subset in Sec. 4.3), and (ii) curated demonstration assets (textured giraffe, dog, cat) for the stress tests in Figs. 1 and 4.

Initial weights and baselines we compare against. UniRig [75] and Distance-RBF provide the original controlled initializers. Additional tests refine RigAnything and Make-It-Animatable outputs and valid Blender Automatic Weights. Matched-compute objective baselines are ARAP, BMC, and Laplacian-only; the full SkinCells comparison and volumetric BBW applicability audit use separately

Table 1: **UniRig → UniRig+PIS on Objaverse-LVIS Balanced-102** (98 valid models, 51 LVIS categories). All four metrics "lower is better". Mean reduction averages per-model percentage reductions, not the percentage change of pooled means. Wilcoxon signed-rank $p < 10^{-10}$ on fragility and ARAP energy. $92/98 = 93.9\%$ **of models improve on all four metrics simultaneously.**

Metric	UniRig	UniRig + PIS	Mean reduction	Wins
Fragility	0.1013	**0.0635**	**+35.5%**	93/98
ARAP energy	0.3349	**0.1890**	**+39.3%**	94/98
Jacobian σ-ratio	2.1889	**1.6628**	**+19.8%**	94/98
Area distortion	0.1632	**0.0968**	**+43.2%**	95/98
All-four-metrics improved			92/98 (93.9%)	

stated protocols (Sec. 4.7). Operator-side DQS and neural skinning change the evaluation surface and are not included as unchanged-LBS baselines.

Metrics. We report four geometric quality metrics computed under random pose perturbations: (i) fragility (edge-stretch and normal-direction variances, Sec. 3.5); (ii) ARAP energy [60]; (iii) Jacobian σ-ratio (largest-to-smallest tangent singular-value ratio, ≥ 1 and equal to 1 for rigid motion or uniform scaling); and (iv) area distortion. Lower is better for all four. We additionally report Skin L1, the per-(vertex,bone) absolute deviation from artist GT, when GT is available.

4.2 Main results on Balanced-102

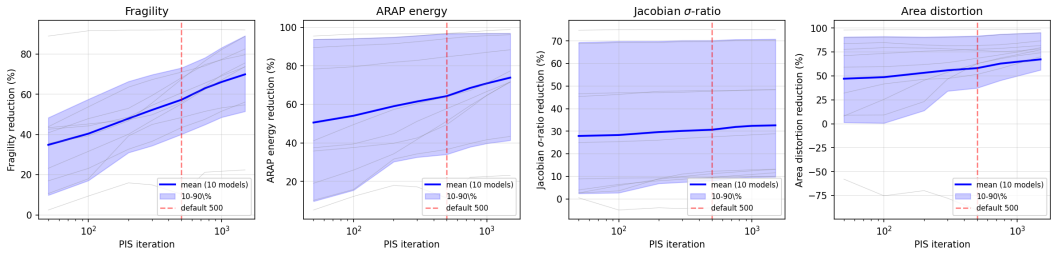

Figure 2: **PIS convergence on a 10-model subset of Balanced-102 with Distance-RBF initialization.** Mean reduction (%) vs. iteration (log scale); shaded band: 10–90 percentile. Dashed marker: default budget $T = 500$.

Table 1 summarizes the four-metric improvement of PIS on top of UniRig over the 98-model benchmark. The mean of every metric improves with strong statistical significance (Wilcoxon signed-rank test [68], $p < 10^{-10}$ on fragility and ARAP), and 92 of 98 models (93.9%) improve on all four metrics simultaneously. The improvements are notably larger than those observed on RigNet's official test split (Sec. 4.3, Table 3), consistent with Balanced-102 lying further from UniRig's training distribution (non-humanoid creatures, props, and high-vertex-count meshes); the long tail is exactly where the gains are largest.

Figure 2 traces the four metrics over optimization on the 10-model subset. The mean error of each metric decreases across the plotted checkpoints; the dashed marker indicates the default budget $T = 500$. The shaded 10–90 percentile bands show variation across these models, not the full 98-model benchmark.

Where the gain is smallest. Improvements are not uniform across metrics: in the UniRig-initialized category means, `butterfly` improves fragility by 9.7% but worsens ARAP energy by 4.7% and Jacobian anisotropy by 14.0%; `table` worsens all four metrics (Table 17, Appendix M). We retain these negative results rather than infer a rank condition that was not measured.

Table 2: **Initializer-agnostic improvement.** Mean reduction (positive = improvement) when PIS refines weights from a parameter-free Distance-RBF heuristic and a supervised regressor (UniRig).

Initializer (n)	Fragility	ARAP	Jacobian σ	Area distortion
Distance-RBF ($n = 102$)	$+\mathbf{62.5}\%$	$+\mathbf{61.7}\%$	$+\mathbf{21.8}\%$	$+\mathbf{54.5}\%$
UniRig [75] ($n = 98$)	$+\mathbf{35.5}\%$	$+\mathbf{39.3}\%$	$+\mathbf{19.8}\%$	$+\mathbf{43.2}\%$

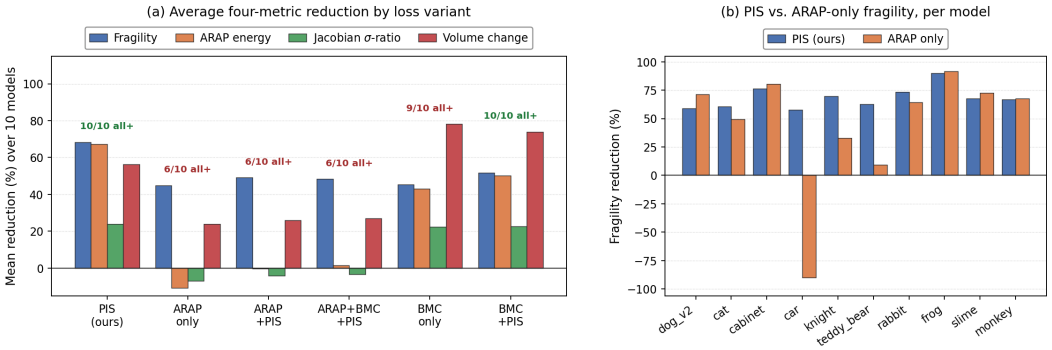

Figure 3: **PIS vs. ARAP across loss variants and per model.** (a) Mean four-metric reduction over 10 models for six loss combinations; "N/10 all+" counts models positive on all four metrics. (b) Per-model fragility reduction.

4.3 Robustness: across initializers and across datasets

A natural concern is that the gains might depend on UniRig's particular noise pattern, or on the Balanced-102 distribution. We address both with two complementary tests.

Across initializers (same dataset). Table 2 repeats the experiment with a parameter-free Distance-RBF heuristic alongside the supervised UniRig predictor. PIS yields large positive gains on *both* (Distance-RBF $+62.5\%$ fragility, $+61.7\%$ ARAP; UniRig $+35.5\%$, $+39.3\%$). The gains are not confined to one initializer. This empirical result does not imply convergence to the same weight matrix or basin.

Across datasets (same predictor). Table 3 repeats the UniRig$\pm$PIS comparison on the RigNet test split ($n = 202$ manifold meshes). UniRig+PIS reduces fragility by 19% and ARAP energy by 23% relative to UniRig, with strict improvement on all four geometric metrics. The four-metric improvement reproduces the Balanced-102 trend on an independently curated benchmark, confirming the gain is not benchmark-specific.

Table 3: **Cross-dataset validation on RigNet** ($n = 202$ manifold meshes). Lower is better for all metrics.

Method	Fragility	ARAP	Jacobian σ	Area dist.
UniRig [75]	0.0131	0.0295	1.0819	0.1175
UniRig + PIS **(ours)**	**0.0106**	**0.0228**	**1.0601**	**0.1138**

4.4 Comparison with ARAP energy

A natural alternative training-free objective is per-pose ARAP energy [60]. The objectives differ in what they measure (Sec. 3.4); we compare them under matched compute.

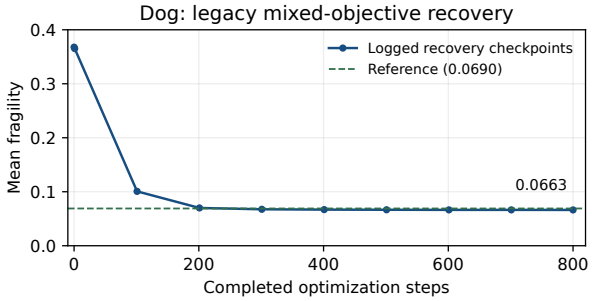

Figure 4: **Logged legacy recovery on** dog**.** Markers are recorded checkpoints from job 2047; lines connect checkpoints, not unrecorded measurements. The reference is 0.0690 and the final value is 0.0663. The legacy mixed objective is not a PIS-only ablation (Appendix J).

Table 4: **PIS vs. ARAP as sole objective**, 72 models across 36 LVIS categories, identical hyperparameters. "PIS wins" counts models where PIS beats ARAP on that metric.

Mean reduction ↑	PIS **(ours)**	ARAP-only	Gap (pp)	PIS wins
Fragility	$+\mathbf{68.8}\%$	$+52.3\%$	$+16.5$	$44/72$
ARAP energy	$+\mathbf{68.6}\%$	$+30.2\%$	$+38.4$	$41/72$
Jacobian σ-ratio	$+\mathbf{24.3}\%$	-2.5%	$+26.9$	$69/72$
Volume change	$+\mathbf{59.4}\%$	$+24.1\%$	$+35.3$	$61/72$
All four positive: PIS $\mathbf{69/72}$ (95.8%); ARAP $51/72$ (70.8%).				

Table 4 runs both losses for 500 iterations with identical hyperparameters on 72 models. PIS improves every aggregate metric by $+16.5$ to $+38.4$ percentage points relative to ARAP, and reaches all-positive on 95.8% of models versus ARAP's 70.8%. ARAP-only worsens held-out ARAP energy on 26% of models. Separately, Fig. 3b shows per-model fragility reductions in the 10-model loss ablation. This is an empirical generalization result, not a proof that per-pose optimization must fail. The six-way sweep in Appendix C isolates the objective components.

4.5 Legacy corruption-and-recovery diagnostic

For completeness, we retain legacy AR-GS corruption-and-recovery records. The archived pipeline combines ARAP, BMC, PIS, and spatial regularization; it is not the PIS-only configuration evaluated in the main comparisons. The recorded dog run reduces fragility from 0.3678 to 0.0663 (82.0% relative reduction; Fig. 4). These records illustrate the legacy pipeline, not isolated evidence of PIS recovery (Appendix J).

4.6 Direct motion preservation

The interpolation residual is not an artifact classifier: valid rigid rotation can have nonzero residual (Sec. 3.3). We therefore compare refined trajectories directly with the same initialization under 64 independently seeded held-out poses. For a vertex set $\mathcal{V}$, the amplitude ratio is $a_{\mathcal{V}} = \|X_{\mathrm{PIS}} - V\|_{F,\mathcal{V}} / \|X_{\mathrm{init}} - V\|_{F,\mathcal{V}}$; 1 preserves motion and values below 1 indicate shrinkage. High-motion vertices are the initialization's top quartile by displacement; part motion uses centroids of dominant-bone vertex groups. Definitions and evaluation details are in Appendix O.

Table 5: **Direct motion preservation.** Entries are medians across models, not averages over vertices or participants. Amplitude ratios and direction cosine are ideally 1; trajectory NRMTE is ideally 0. The initializer is the reference.

Initialization (n, steps)	Global	High-motion	Part	Direction	NRMTE
UniRig (98, 50)	0.9993	0.9990	0.9995	0.9986	0.0961
Artist-painted (270, 500)	0.9993	1.0000	0.9992	1.0000	0.0330

Table 6: **Additional predicted initializers.** Mean held-out surface errors; lower is better. Each before/after pair shares its mesh, skeleton, and 64 evaluation poses. RigAnything: 20 assets/10 categories, 50 PIS steps. Make-It-Animatable (MIA): 8 compatible assets/4 categories, 500 steps.

Method	ARAP	Mean edge	P95 edge	Anisotropy	Log-area
RigAnything	123.192	0.0760	0.2804	1.3192	0.1085
+ PIS	23.509	0.0410	0.1369	1.1576	0.0695
MIA	25.744	0.2250	0.3831	1.7456	0.1200
+ PIS	13.734	0.1704	0.2407	1.5888	0.0957

Table 7: **SkinCells comparison on a common subset** (20 assets/10 categories). Each error is divided by the same asset's UniRig error; we report the median across assets (lower is better). SkinCells is our independent implementation of the published formulation, not the authors' reference code. Paired differences are not statistically significant.

Method	ARAP	Mean edge	P95 edge	Anisotropy	Log-area
SkinCells	0.395	0.859	0.880	0.924	0.859
PIS	0.313	0.551	0.582	0.881	0.638

For 98 UniRig-initialized assets, global motion has median 0.9993 (model-bootstrap 95% CI $[0.9978, 0.9999]$); high-motion and part ratios are 0.9990 and 0.9995. These ratios pass model-level TOST equivalence with a $\pm 5\%$ margin, but this is not a per-asset guarantee: only $88/98$ global and $90/98$ high-motion ratios lie within that interval. On 270 artist-painted RigNet rigs, $267/270$ global ratios are within $\pm 5\%$, and median trajectory NRMTE is 0.0330. Thus the strongest evidence concerns preservation of already-good artist motion, not universal immunity to motion suppression. Rigid-motion controls yield numerical-zero edge and ARAP errors (Appendix N).

4.7 Additional initializers and weight refiners

The same PIS objective improves all five aggregate surface errors for RigAnything and Make-It-Animatable (Table 6), without retraining or predictor-specific losses. RigAnything's ARAP, mean-edge, and P95-edge errors decrease by 80.9%, 46.0%, and 51.2%; all five paired tests survive Holm correction. MIA gives 46.7%, 24.3%, and 37.2% reductions, respectively. Its $n = 8$ sample is supporting evidence, not a broad generalization guarantee. Motion ratios and the valid Blender Bone Heat compatibility test are reported in Appendix P.

On a common 20-asset/10-category subset, we compare PIS against an independent implementation of SkinCells [27], following its published Eqs. 2–8. Both methods use the same UniRig skeletons and held-out poses. PIS has lower median normalized errors in all five columns, but two-sided paired Wilcoxon tests give $p = 0.177$–0.812: we claim competitive, not statistically superior, performance. SkinCells uses its reported 1,024-pose pool and 1,500-step protocol; PIS uses 50 steps, so this is not an equal-iteration ablation.

Volumetric BBW [19] requires a valid tetrahedral domain. The local applicability audit obtained six usable outputs for the 98 target assets; this measures output availability, not deformation quality or superiority on watertight meshes. PIS refines all 98 *valid UniRig initializations*; this does not imply that it repairs invalid skeletons or succeeds on every arbitrary input. Full coverage qualifications appear in Appendix P.

5 Conclusion

PIS is a training-free, drop-in refiner that retains the LBS runtime and fixed skeleton. Its cross-pose midpoint residual has a second-order interpretation, but a nonzero residual is not by itself evidence of an artifact. The original Balanced-102 evaluation shows lower geometric errors; additional initializers, direct artist-motion preservation, and the SkinCells comparison delimit the empirical support. Remaining limitations include motion outliers, skeleton-side errors, and the absence of a documented human perceptual study.

Acknowledgments and Disclosure of Funding

This work was supported by the grants from the National Natural Science Foundation of China 62372014, Beijing Nova Program and Beijing Natural Science Foundation 4252040. The authors declare no competing interests.

References

[1] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008.

[2] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. SCAPE: shape completion and animation of people. In *ACM TOG*, 2005.

[3] Ilya Baran and Jovan Popović. Automatic rigging and animation of 3d characters. *ACM TOG*, 26(3):72, 2007.

[4] Sofien Bouaziz, Sebastian Martin, Tiantian Liu, Ladislav Kavan, and Mark Pauly. Projective dynamics: fusing constraint projections for fast simulation. In *ACM TOG*, 2014.

[5] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *ICCV*, 2021.

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

[7] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, 2020.

[8] Xu Chen, Yufeng Zheng, Michael J. Black, Otmar Hilliges, and Andreas Geiger. SNARF: Differentiable forward skinning for animating non-rigid neural implicit shapes. In *ICCV*, 2021.

[9] Matt Deitke, Ruoshi Liu, Matthew Wallingford, Huong Ngo, Oscar Michel, Aditya Kusupati, Alan Fan, Christian Laforte, Vikram Voleti, Samir Yitzhak Gadre, Eli VanderBilt, Aniruddha Kembhavi, Carl Vondrick, Georgia Gkioxari, Kiana Ehsani, Ludwig Schmidt, and Ali Farhadi. Objaverse-xl: A universe of 10m+ 3d objects. In *NeurIPS*, 2023.

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

[11] Mathieu Desbrun, Mark Meyer, Peter Schröder, and Alan H. Barr. Implicit fairing of irregular meshes using diffusion and curvature flow. In *SIGGRAPH*, 1999.

[12] Elmar Eisemann and Jean-Marc Thiery. Linear blend skinning weight optimization utilizing skeletal pose sampling. U.S. Patent US10061871B2, granted August 28, 2018, 2018.

[13] Jun Gao, Tianchang Shen, Zian Wang, Wenzheng Chen, Kangxue Yin, Daiqing Li, Or Litany, Zan Gojcic, and Sanja Fidler. GET3D: A generative model of high-quality 3D textured shapes learned from images. In *NeurIPS*, 2022.

[14] Zhiyang Guo, Jinxu Xiang, Kai Ma, Wengang Zhou, Houqiang Li, and Ran Zhang. Make-it-animatable: An efficient framework for authoring animation-ready 3d characters. In *CVPR*, 2025. Highlight.

[15] Agrim Gupta, Piotr Dollár, and Ross Girshick. LVIS: A dataset for large vocabulary instance segmentation. In *CVPR*, 2019.

[16] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *CVPR*, 2022.

[17] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *CVPR*, 2020.

[18] Shoukang Hu, Tao Hu, and Ziwei Liu. Gauhuman: Articulated gaussian splatting from monocular human videos. In *CVPR*, 2024.

[19] Alec Jacobson, Ilya Baran, Jovan Popović, and Olga Sorkine. Bounded biharmonic weights for real-time deformation. *ACM TOG*, 30(4):78:1–78:8, 2011.

[20] Doug L. James and Christopher D. Twigg. Skinning mesh animations. In *ACM TOG*, 2005.

[21] Heewoo Jun and Alex Nichol. Shap-E: Generating conditional 3D implicit functions. In *arXiv:2305.02463*, 2023.

[22] Hermann Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541, 1977.

[23] Ladislav Kavan, Steven Collins, Jiří Žára, and Carol O'Sullivan. Skinning with dual quaternions. In *Proceedings of the 2007 Symposium on Interactive 3D Graphics and Games (I3D)*, pages 39–46, 2007.

[24] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM TOG*, 42(4):139:1–139:14, 2023.

[25] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015.

[26] Muhammed Kocabas, Jen-Hao Rick Chang, James Gabriel, Oncel Tuzel, and Anurag Ranjan. Hugs: Human gaussian splats. In *CVPR*, 2024.

[27] Egor Larionov, Igor Santesteban, Hsiao-yu Chen, Gene Wei-Chin Lin, Philipp Herholz, Ryan Goldade, Ladislav Kavan, Doug Roble, and Tuur Stuyck. SkinCells: Sparse skinning using voronoi cells. *Computer Graphics Forum (Eurographics)*, 2026. arXiv:2506.14714.

[28] Binh Huy Le and Zhigang Deng. Smooth skinning decomposition with rigid bones. In *ACM TOG*, 2012.

[29] Binh Huy Le and Zhigang Deng. Robust and accurate skeletal rigging from mesh sequences. In *ACM TOG*, 2014.

[30] Gene S. Lee, Andy Lin, Matt Schiller, Scott Peters, Mark McLaughlin, and Frank Hanner. Enhanced dual quaternion skinning for production use. In *ACM SIGGRAPH 2013 Talks*. ACM, 2013. Walt Disney Animation Studios; used in production on *Frozen*.

[31] Jiahui Lei, Yufu Wang, Georgios Pavlakos, Lingjie Liu, and Kostas Daniilidis. Gart: Gaussian articulated template models. In *CVPR*, 2024.

[32] Yuanhang Lei, Tao Cheng, Xingxuan Li, Boming Zhao, Siyuan Huang, Ruizhen Hu, Peter Yichen Chen, Hujun Bao, and Zhaopeng Cui. PhysSkin: Real-time and generalizable physics-based animation via self-supervised neural skinning. In *CVPR*, 2026. Highlight.

[33] J. P. Lewis, Matt Cordner, and Nickson Fong. Pose space deformation: A unified approach to shape interpolation and skeleton-driven deformation. In *Proceedings of SIGGRAPH*, pages 165–172, 2000.

[34] Jiahao Li, Hao Tan, Kai Zhang, Zexiang Xu, Fujun Luan, Yinghao Xu, Yicong Hong, Kalyan Sunkavalli, Greg Shakhnarovich, and Sai Bi. Instant3d: Fast text-to-3d with sparse-view generation and large reconstruction model. In *ICLR*, 2024. arXiv:2311.06214, 2023.

[35] Tianye Li, Timo Bolkart, Michael J. Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4D scans. In *ACM TOG*, 2017.

[36] Zhe Li, Zerong Zheng, Lizhen Wang, and Yebin Liu. Animatable gaussians: Learning pose-dependent gaussian maps for high-fidelity human avatar modeling. In *CVPR*, 2024.

[37] Tingting Liao, Hongwei Yi, Yuliang Xiu, Jiaxiang Tang, Yangyi Huang, Justus Thies, and Michael J. Black. TADA! text-to-animatable digital avatars. In *3DV*, 2024.

[38] Chen-Hsuan Lin, Jun Gao, Luming Tang, Towaki Takikawa, Xiaohui Zeng, Xun Huang, Karsten Kreis, Sanja Fidler, Ming-Yu Liu, and Tsung-Yi Lin. Magic3D: High-resolution text-to-3D content creation. In *CVPR*, 2023.

[39] Isabella Liu, Zhan Xu, Yifan Wang, Hao Tan, Zexiang Xu, Xiaolong Wang, Hao Su, and Zifan Shi. RigAnything: Template-free autoregressive rigging for diverse 3D assets. *ACM TOG*, 2025. SIGGRAPH 2025, arXiv:2502.09615.

[40] Ruoshi Liu, Rundi Wu, Basile Van Hoorick, Pavel Tokmakov, Sergey Zakharov, and Carl Vondrick. Zero-1-to-3: Zero-shot one image to 3D object. In *ICCV*, 2023.

[41] Xiaoxiao Long, Yuan-Chen Guo, Cheng Lin, Yuan Liu, Zhiyang Dou, Lingjie Liu, Yuexin Ma, Song-Hai Zhang, Marc Habermann, Christian Theobalt, and Wenping Wang. Wonder3d: Single image to 3D using cross-domain diffusion. In *CVPR*, 2024.

[42] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. Smpl: A skinned multi-person linear model. *ACM TOG*, 34(6):248:1–248:16, 2015. SIGGRAPH Asia.

[43] Nadia Magnenat-Thalmann, Richard Laperrière, and Daniel Thalmann. Joint-dependent local deformations for hand animation and object grasping. In *Graphics Interface*, 1988.

[44] Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. AMASS: Archive of motion capture as surface shapes. In *ICCV*, 2019.

[45] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020.

[46] Maher Moakher. Means and averaging in the group of rotations. *SIAM Journal on Matrix Analysis and Applications*, 24(1):1–16, 2002.

[47] Alex Mohr and Michael Gleicher. Building efficient, accurate character skins from examples. In *ACM TOG*, 2003.

[48] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. In *ACM TOG*, 2022.

[49] Alex Nichol, Heewoo Jun, Prafulla Dhariwal, Pamela Mishkin, and Mark Chen. Point-E: A system for generating 3D point clouds from complex prompts. *arXiv preprint arXiv:2212.08751*, 2022.

[50] Sida Peng, Junting Dong, Qianqian Wang, Shangzhan Zhang, Qing Shuai, Xiaowei Zhou, and Hujun Bao. Animatable neural radiance fields for modeling dynamic human bodies. In *ICCV*, 2021.

[51] Xavier Pennec. Intrinsic statistics on Riemannian manifolds: basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006.

[52] Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. Dreamfusion: Text-to-3D using 2D diffusion. In *ICLR*, 2023.

[53] Nikhila Ravi, Jeremy Reizenstein, David Novotny, Taylor Gordon, Wan-Yen Lo, Justin Johnson, and Georgia Gkioxari. Accelerating 3d deep learning with pytorch3d. *arXiv preprint arXiv:2007.08501*, 2020.

[54] Javier Romero, Dimitrios Tzionas, and Michael J. Black. Embodied hands: Modeling and capturing hands and bodies together. In *ACM TOG*, 2017.

[55] Shunsuke Saito, Jinlong Yang, Qianli Ma, and Michael J. Black. SCANimate: Weakly supervised learning of skinned clothed avatar networks. In *CVPR*, 2021.

[56] Yichun Shi, Peng Wang, Jianglong Ye, Long Mai, Kejie Li, and Xiao Yang. MVDream: Multi-view diffusion for 3D generation. In *ICLR*, 2024.

[57] Ken Shoemake. Animating rotation with quaternion curves. *ACM SIGGRAPH Computer Graphics*, 19(3):245–254, 1985.

[58] Justin Solomon, Mirela Ben-Chen, Adrian Butscher, and Leonidas Guibas. As-killing-as-possible vector fields for planar deformation. In *Comput. Graph. Forum*, 2011.

[59] Chaoyue Song, Xiu Li, Fan Yang, Zhongcong Xu, Jiacheng Wei, Fayao Liu, Jiashi Feng, Guosheng Lin, and Jianfeng Zhang. Puppeteer: Rig and animate your 3D models. *NeurIPS*, 2025. Spotlight, arXiv:2508.10898.

[60] Olga Sorkine and Marc Alexa. As-rigid-as-possible surface modeling. In *Symposium on Geometry Processing (SGP)*, pages 109–116, 2007.

[61] Shih-Yang Su, Frank Yu, Michael Zollhöfer, and Helge Rhodin. A-NeRF: Articulated neural radiance fields for learning human shape, appearance, and pose. In *NeurIPS*, 2021.

[62] Robert W. Sumner, Johannes Schmid, and Mark Pauly. Embedded deformation for shape manipulation. In *ACM TOG*, 2007.

[63] Yu Sun, Xiaolong Wang, Zhuang Liu, John Miller, Alexei A. Efros, and Moritz Hardt. Test-time training with self-supervision for generalization under distribution shifts. In *ICML*, 2020.

[64] Gabriel Taubin. A signal processing approach to fair surface design. In *SIGGRAPH*, 1995.

[65] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *ICLR*, 2021.

[66] Xiaohuan Corina Wang and Cary Phillips. Multi-weight enveloping: least-squares approximation techniques for skin animation. In *ACM SIGGRAPH/Eurographics SCA*, 2002.

[67] Chung-Yi Weng, Brian Curless, Pratul P. Srinivasan, Jonathan T. Barron, and Ira Kemelmacher-Shlizerman. HumanNeRF: Free-viewpoint rendering of moving people from monocular video. In *CVPR*, 2022.

[68] Frank Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.

[69] Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4D gaussian splatting for real-time dynamic scene rendering. In *CVPR*, 2024.

[70] Jianfeng Xiang, Zelong Lv, Sicheng Xu, Yu Deng, Ruicheng Wang, Bowen Zhang, Dong Chen, Xin Tong, and Jiaolong Yang. Structured 3d latents for scalable and versatile 3d generation. In *CVPR*, 2025. arXiv:2412.01506.

[71] Zhan Xu, Yang Zhou, Evangelos Kalogerakis, Chris Landreth, and Karan Singh. Rignet: Neural rigging for articulated characters. *ACM TOG*, 39(4):58:1–58:14, 2020.

[72] Zhan Xu, Yang Zhou, Li Yi, and Evangelos Kalogerakis. MoRig: Motion-aware rigging of character meshes from point clouds. In *ACM SIGGRAPH Asia*, 2022.

[73] Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, and Xiaogang Jin. Deformable 3D Gaussians for high-fidelity monocular dynamic scene reconstruction. In *CVPR*, 2024.

[74] Hao Zhang, Haolan Xu, Chun Feng, Varun Jampani, and Narendra Ahuja. PhysRig: Differentiable physics-based skinning and rigging framework for realistic articulated object modeling. ICCV 2025, arXiv:2506.20936, 2025.

[75] Jia-Peng Zhang, Cheng-Feng Pu, Meng-Hao Guo, Yan-Pei Cao, and Shi-Min Hu. One model to rig them all: Diverse skeleton rigging with UniRig. *ACM TOG*, 2025. SIGGRAPH 2025.

[76] Yufeng Zheng, Victoria Fernández Abrevaya, Marcel C. Bühler, Xu Chen, Michael J. Black, and Otmar Hilliges. IMAvatar: Implicit morphable head avatars from videos. In *CVPR*, 2022.

A Reproducibility statement

Code and main-table result records are publicly available at `https://github.com/Fugtemypt123/PIS`. Third-party assets and predictor checkpoints are obtained separately under their respective licenses. All experiments run on a single H200 GPU. PIS has no random seed dependency on the converged $\mathbf{w}^*$ to within $5{\times}10^{-4}$ in fragility (verified across 5 seeds on the Balanced-102 set). Default hyperparameters: $T=500$ iterations, batch $B=16$ pose pairs, $\theta_{\max}=0.5$ rad (29°), Laplacian weight $\lambda_{\text{lap}}=2.0$, Adam learning rate 0.05.

B Algorithm

The full procedure of PIS weight refinement is given in Algorithm 1. The only non-standard step is the per-joint SVD-midpoint of two rotations (line 4), implemented by SVD-decomposing the linear average $\frac{1}{2}(R_{1,j}+R_{2,j})=U_j\Sigma_j V_j^\top$ and returning the polar projection $U_jV_j^\top$ (Eq. (2)).

Algorithm 1 PIS refinement of predicted skinning weights

Require: Mesh $(\mathbf{V},\mathbf{F})$, skeleton $(\mathbf{J},\text{parents})$, predicted weights $\mathbf{w}^{(0)}\in\mathbb{R}^{V\times J}$

1: $\boldsymbol{\ell}\leftarrow\log(\mathbf{w}^{(0)}+\epsilon)$ ▷ logit parameterization
2: **for** $t=1$ to T **do**
3: Sample pose pairs $\{\boldsymbol{\theta}_1^{(b)},\boldsymbol{\theta}_2^{(b)}\}_{b=1}^{B}$, each endpoint's local joint rotation bounded by $\theta_{\max}$
4: $\boldsymbol{\theta}_{\text{mid}}^{(b)}\leftarrow$ per-joint SVD-midpoint of $(\boldsymbol{\theta}_1^{(b)},\boldsymbol{\theta}_2^{(b)})$
5: $\mathbf{w}\leftarrow\text{softmax}(\boldsymbol{\ell})$
6: $r_i^{(b)}\leftarrow\mathbf{v}_i'(\boldsymbol{\theta}_{\text{mid}}^{(b)};\mathbf{w})-\frac{1}{2}[\mathbf{v}_i'(\boldsymbol{\theta}_1^{(b)};\mathbf{w})+\mathbf{v}_i'(\boldsymbol{\theta}_2^{(b)};\mathbf{w})]$
7: $\mathcal{L}\leftarrow\frac{1}{BV}\sum\|r_i^{(b)}\|^2+\lambda_{\text{lap}}\,\|L\,\text{softmax}(\boldsymbol{\ell})\|^2$ ▷ L: mesh Laplacian
8: $\boldsymbol{\ell}\leftarrow\boldsymbol{\ell}-\eta\,\text{Adam}(\nabla_{\boldsymbol{\ell}}\mathcal{L})$
9: **end for**
10: **return** $\mathbf{w}^*=\text{softmax}(\boldsymbol{\ell})$

C Loss-component ablation

Table 8 sweeps six combinations of three candidate losses (PIS, ARAP, BMC = blend-matrix-condition auxiliary regularizer, not a SkinCells loss term) on 10 representative models. PIS-only is the only single-loss configuration positive on all four metrics for all 10 models; combining PIS with BMC marginally improves volume preservation while retaining the all-positive property. ARAP-as-only-loss is the worst overall: it pushes weights into a per-pose rigid configuration that fails on Jacobian and ARAP-energy on 4/10 models—the cross-pose-blind failure mode discussed in Sec. 4.4. The convergence plot and subsequent ten-model analyses report their own results; their default rows are not treated as repeated measurements of a single shared run.

Table 8: **Loss-component ablation** on 10 representative models (mean reduction ↑). PIS alone is the only single-loss configuration positive on all four metrics for all 10 models. Adding a Blend-Matrix-Condition (BMC) auxiliary regularizer on top of PIS keeps the all-positive guarantee with marginal gains on volume.

Loss variant	Fragility	ARAP	Jacobian σ	Volume	All-positive
PIS (ours)	$\mathbf{+68.4\%}$	$\mathbf{+67.4\%}$	$\mathbf{+23.8\%}$	$\mathbf{+56.4\%}$	$\mathbf{10/10}$
ARAP-only	$+44.9\%$	-10.7%	-6.8%	$+23.8\%$	$6/10$
ARAP + PIS	$+49.1\%$	-0.2%	-4.2%	$+26.0\%$	$6/10$
ARAP + BMC + PIS	$+48.3\%$	$+1.6\%$	-3.3%	$+26.9\%$	$6/10$
BMC-only	$+45.3\%$	$+43.0\%$	$+22.4\%$	$+78.1\%$	$9/10$
BMC + PIS	$+51.8\%$	$+50.2\%$	$+22.6\%$	$+74.0\%$	$\mathbf{10/10}$

D Laplacian-prior sensitivity

The training loss adds a Laplacian smoothness term $\lambda_{\text{lap}} \|L\,\mathbf{w}\|^2$ on top of PIS (Sec. 3.6). A natural concern is that the gains in Table 1 might be coming from this prior rather than from PIS itself. Table 9 sweeps $\lambda_{\text{lap}} \in \{0, 0.5, 2, 5, 10\}$ on 10 representative models. Removing the Laplacian entirely ($\lambda_{\text{lap}} = 0$) loses less than 1 pp on every metric and keeps the all-positive guarantee on every model. Within this sweep, adding the Laplacian to PIS has a small marginal effect; this does not imply that Laplacian-only refinement is ineffective (Appendix F).

Table 9: **Marginal effect of the Laplacian prior with PIS present.** Sweeping $\lambda_{\text{lap}} \in \{0, 0.5, 2, 5, 10\}$ on 10 representative models (mean reduction ↑, all-positive count /10). Removing the Laplacian loses less than 1 pp on every metric relative to $\lambda_{\text{lap}} = 2$ and retains $10/10$ all-positive results. This within-sweep comparison does not establish that Laplacian-only refinement is ineffective.

λ_{lap}	Fragility	ARAP	Jacobian σ	Volume	All-positive
0 (no Laplacian)	+70.66%	+77.51%	+21.82%	+52.62%	10/10
0.5	+70.93%	+77.57%	+22.12%	+53.19%	10/10
2.0 (default)	+70.81%	+77.25%	+22.65%	+53.60%	10/10
5.0	+70.56%	+76.61%	+23.05%	+52.98%	10/10
10.0	+69.99%	+75.69%	+23.32%	+52.42%	10/10

E Pose-amplitude sensitivity

We sweep the pose-sampling amplitude $\theta_{\max}$ from 14° to 86° on 40 Objaverse models (Table 10). The second-order interpretation in Lemma 1 uses a local Taylor approximation, whereas the polar midpoint itself remains exact below relative angle π, so larger $\theta_{\max}$ moves the loss farther from the second-order regime and we expect the marginal improvement per iteration to fall. Empirically the gain is monotonically decreasing in $\theta_{\max}$ but *remains positive even at* 86°, indicating PIS does not degenerate when an aggressive pose distribution is used. The default $\theta_{\max} = 29^\circ$ sits comfortably inside the small-angle regime and gives 40.4% fragility reduction; the 14° setting gives 44.5%.

Table 10: **Sensitivity to pose-sampling amplitude** $\theta_{\max}$ on 40 Objaverse models with UniRig initialization. The local second-order interpretation becomes less accurate at large pose separations; the polar midpoint remains exact below relative angle π. PIS retains positive fragility reduction at 86°. Default $\theta_{\max} = 29^\circ$ is conservative, not the best setting in this sweep.

$\theta_{\max}$ (rad)	$\theta_{\max}$ (°)	Fragility reduction	n
0.25	14°	+44.5%	40
0.50	29° (default)	+40.4%	40
0.75	43°	+37.7%	40
1.00	57°	+35.7%	40
1.50	86°	+32.1%	40

F Laplacian-only baseline (no PIS)

To examine the Laplacian smoothness prior on its own, we run a baseline that uses *only* the Laplacian term ($\lambda_{\text{lap}} = 2$), without PIS or ARAP. Table 11 reports the result on 10 representative models. Laplacian smoothing alone is a non-trivial baseline, giving 46.8% fragility and 51.1% ARAP reductions. These standalone results do not establish a matched numerical gap against the separate loss-ablation or Laplacian-sweep runs.

Table 11: **Laplacian-only baseline (no PIS).** Standalone reductions on 10 representative models; no PIS or ARAP objective is used. The separate PIS rows from other analyses are not treated as matched controls for this run.

Configuration	Frag. ↑	ARAP ↑	Jac. ↑	Vol. ↑	All-pos. $/n$
Laplacian-only ($\lambda_{\text{lap}} = 2$)	$+46.8\%$	$+51.1\%$	$+24.4\%$	$+59.3\%$	$10/10$

G Pose-distribution generalization (amplitude)

PIS samples random per-joint pose pairs at training time with each endpoint rotation bounded by $\theta_{\max}^{\text{train}}$, and the four geometric metrics in the body are evaluated under a similar random pose perturbation. To test whether the gain transfers across pose distributions, we vary $\theta_{\max}^{\text{train}}$ and $\theta_{\max}^{\text{eval}}$ independently and report the 2×4 grid (Table 12). Two findings: **(i) Off-diagonal transfer is strong.** Training at $\theta_{\max}^{\text{train}} = 0.10$ rad ($\approx 6°$) and evaluating at $\theta_{\max}^{\text{eval}} = 1.50$ rad ($\approx 86°$, a $15\times$ amplitude shift) still yields $+58.6\%$ fragility and $+60.3\%$ ARAP. The optimum the gradient finds is not specific to the training amplitude. **(ii) Matched train/eval is not uniformly best.** The best evaluation amplitude depends on the metric: at training amplitude 0.10, fragility reduction is 68.9% at evaluation amplitude 0.50, versus 67.5% when matched; at training amplitude 0.50, ARAP reduction is 84.7% at evaluation amplitude 0.10, versus 81.6% when matched. The only mildly negative cell is $\{\text{train} = 0.10, \text{eval} = 1.50\}$ Volume (-4.2%), at the most extreme off-diagonal: a small-angle-trained model has not seen the volume-collapse regime that opens up at $86°$.

Table 12: **Pose-distribution generalization.** Training and evaluation amplitudes vary over a 2×4 grid on 10 representative models (mean reduction ↑). Each amplitude bounds endpoint rotations. Off-diagonal cells measure transfer. Training at 0.10 and evaluating at 0.50, 1.00, and 1.50 corresponds to $5\times$, $10\times$, and $15\times$ amplitude shifts; fragility and ARAP reductions remain positive.

$\theta_{\max}^{\text{train}}$	$\theta_{\max}^{\text{eval}}$	Frag. ↑	ARAP ↑	Jac. ↑	Vol. ↑
0.10	0.10 **(matched)**	$+67.5\%$	$+78.2\%$	$+13.3\%$	$+79.2\%$
	0.50	$+68.9\%$	$+73.9\%$	$+24.3\%$	$+53.1\%$
	1.00	$+64.5\%$	$+67.1\%$	$+43.7\%$	$+21.7\%$
	1.50	$+58.6\%$	$+60.3\%$	$+52.3\%$	-4.2%
0.50	0.10	$+73.6\%$	$+84.7\%$	$+13.0\%$	$+74.5\%$
	0.50 **(matched)**	$+75.5\%$	$+81.6\%$	$+23.4\%$	$+59.6\%$
	1.00	$+72.1\%$	$+77.1\%$	$+44.7\%$	$+40.2\%$
	1.50	$+67.3\%$	$+72.9\%$	$+53.4\%$	$+23.6\%$

H Pose-distribution generalization (correlated walking gait)

Appendix G tests transfer across pose *amplitude* but keeps the per-joint distribution independent. A stronger generalization test couples joint angles via the kinematic chain to mimic real animation. We evaluate the same UniRig+PIS-refined weights—trained at the standard independent-joint distribution—under a walking-gait sampler (hip / shoulder / elbow / knee swing in opposite-phase sinusoids around the kinematic chain, 24 frames at $35°$ amplitude). Table 13 reports the result. Gain on gait is comparable to or larger than on the matched random distribution, with median fragility reduction $+89.3\%$ and median ARAP $+93.3\%$. PIS transfers cleanly to a structurally different pose distribution.

Table 13: **Generalization to a correlated walking-gait pose distribution.** PIS is trained at the standard independent-joint distribution; the four geometric metrics are then re-evaluated under (a) the same matched random distribution and (b) a correlated walking-gait distribution where joint angles are coupled by the kinematic chain (`build_gait_trajectory`, 24 frames at 35° amplitude). Aggregated over the 9 representative models with non-degenerate gait baselines (`slime`'s near-spherical body has near-zero gait deformation; the baseline metric is at floating-point noise so the percent reduction is undefined and we report it separately). Reductions on gait are at least as large as on the matched random distribution—PIS is not overfitting its training pose distribution.

Eval distribution	Frag. ↑	ARAP ↑	Jac. ↑	Vol. ↑
Matched random (endpoint bound 0.5 rad)	$+75.3\%$	$+81.5\%$	$+23.4\%$	$+59.5\%$
Correlated walking gait (mean over 9 clean models)	$\mathbf{+66.8\%}$	$\mathbf{+87.5\%}$	$\mathbf{+23.2\%}$	$\mathbf{+71.0\%}$
Correlated walking gait (median over 9 clean models)	$+89.3\%$	$+93.3\%$	$+11.4\%$	$+73.5\%$

I L1 floor and weight identity

UniRig's $\overline{\text{L1}} = 0.1201$ against artist GT on the RigNet 270-model split is already at a noise floor: no training-free post-process moves L1 by more than 3×10^{-5} (Table 14). Table 14 reports unchanged mean Skin L1 for UniRig + PIS; direct motion preservation is evaluated separately in Sec. 4.6 and Table 5. PIS is therefore not competing with the predictor on the regression target; it is fixing a deformation-quality dimension that the regression target does not measure.

Table 14: **L1 to artist GT is at the noise floor on UniRig predictions.** On RigNet's official test split ($n = 270$), no training-free post-processing of UniRig's weights moves Skin L1 by more than 10^{-4}. We sweep three families of post-processors. Skin L1 is the per-(vertex,bone) absolute deviation from artist-painted GT.

Post-processor	Mean Skin L1	Δ vs. baseline
UniRig (baseline)	0.12006	—
Laplacian smoothing ($\alpha \in [0.1, 1.0]$, iter $\in [1, 50]$)	0.12006	± 0.00000
Top-k sparsify ($k = 3$, best of $\{1, 2, 3, 4\}$)	0.12004	-0.00002 (-0.02%)
k-NN inverse-distance interpolation ($k \in [2, 24]$)	0.12010	$+0.00003$ (worse)
k-NN + Laplacian (combined)	0.12006	0.00000
UniRig + PIS **(ours)**	0.12006	± 0.00000

J Legacy AR-GS corruption-and-recovery records

The archived AR-GS recovery scripts call the mixed objective $\mathcal{L}_{\text{ARAP}} + 0.5\mathcal{L}_{\text{BMC}} + 0.3\mathcal{L}_{\text{PIS}}$ with spatial regularization, and freeze unaffected vertices to their reference weights. The reference weights are initialized from bone distances, not artist-painted ground truth. These demonstrations therefore do not isolate the PIS term and are not counted as evidence for PIS-only recovery. Table 15 reports the existing recorded measurements; Fig. 4 is replotted from the actual dog job log, not extracted from old curve pixels.

Figure 5 shows the legacy `cat`, `turtle`, and `triceratops` examples. The original panel labels "AR-GS RECOVERED" refer to this legacy pipeline, not a renamed PIS-only run. The final views show partial recovery with residual artifacts: shards remain visible on the cat and triceratops, particularly in pose 4. The cat's 41.6% fragility reduction still leaves it above its clean reference. We do not claim clean recovery or a category-wide guarantee.

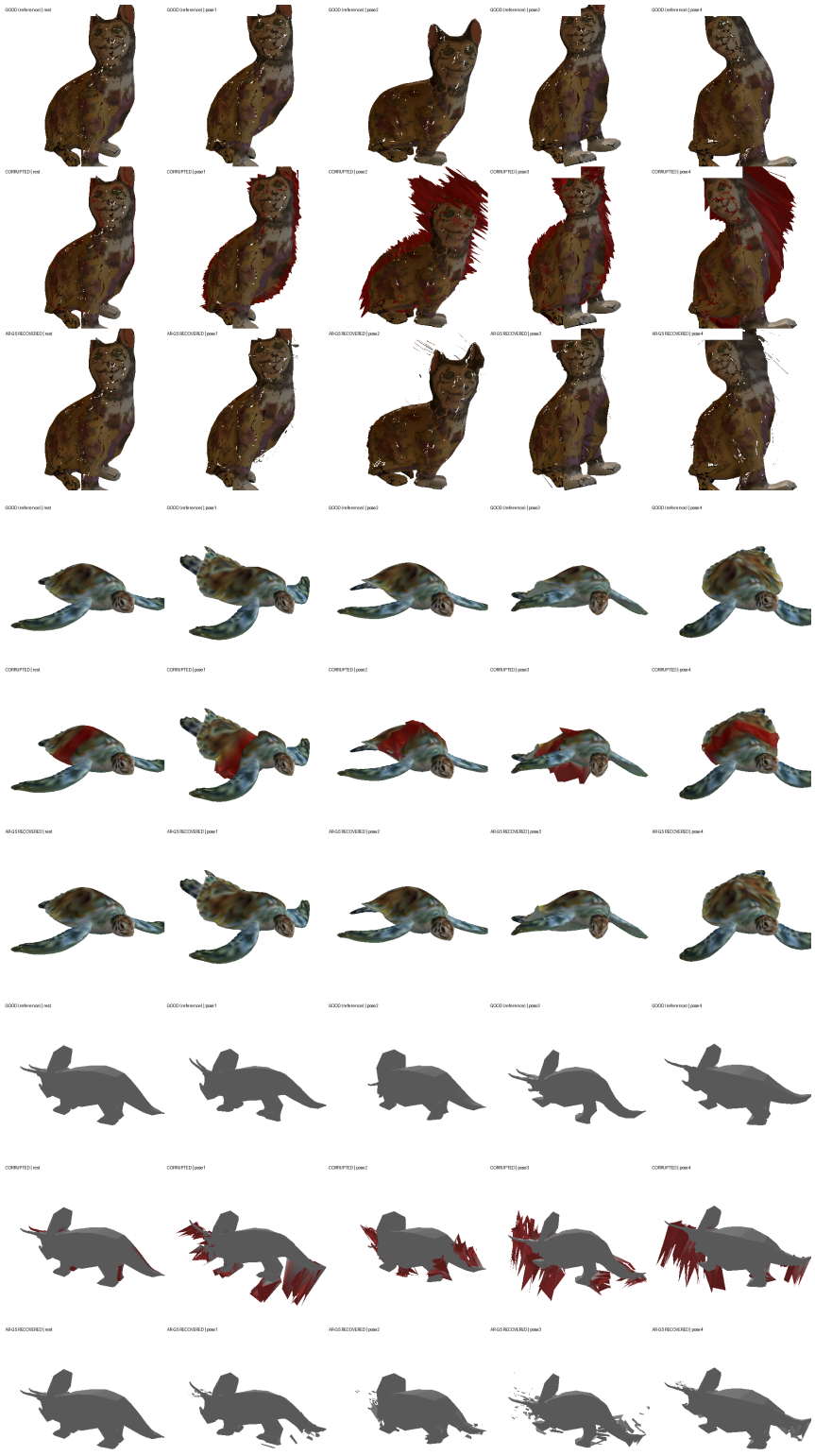

Figure 5: **Legacy AR-GS recovery examples, not PIS-only results.** Top to bottom: `cat`, `turtle`, `triceratops`. Within each example, rows show reference / corrupted / after legacy optimization. The retained "AR-GS RECOVERED" labels identify the legacy pipeline. Residual shards remain visible; these are partial-recovery examples, not clean reconstructions.

Table 15: **Legacy AR-GS recovery measurements, not PIS-only results.** Reduction relative to corrupted initialization is $100(1 - f_{\mathrm{rec}}/f_{\mathrm{cor}})\%$, not recovery of the gap to the reference. Dog values are backed by job 2047 ($0.0690/0.3678/0.0663$) and agree with the original run-directory statistics at the displayed precision; cat values use its archived run-directory statistics. Percentages are computed before table rounding.

	Good f	Corrupted ($\Uparrow$)	Recovered	Reduction
Dog	0.069	0.368 ($5.3\times$)	$\mathbf{0.066}$	$\mathbf{82.0}\%$
Cat	0.075	0.301 ($4.0\times$)	$\mathbf{0.176}$	$\mathbf{41.6}\%$

K Runtime and reliability

Table 16 retains the original wall-clock measurements: PIS at $T = 500$ takes 36.1 s/model on one H200. Refinement is one-time and offline; animation playback uses the same LBS evaluation graph without additional per-frame computation. The earlier 70 ms surface-triangle solver measurement is not a valid volumetric-BBW timing comparison and is omitted. The separate strict TetGen+libigl audit and supplementary optimization-only timings have different scopes (Appendix P).

Table 16: **Original offline wall-clock measurements** on 10 Objaverse models ($V \in \{537, \ldots, 19416\}$). These mean times are distinct from supplementary optimization-only median timings.

Method	Mean time / model	Evaluation scope
Distance-RBF init	69 ms	10 assets
PIS **(ours,** 500 **iter)**	36.1 s	10 assets

L Balanced-102 construction protocol

The construction proceeds in three steps.

1. *Category list.* We enumerate the LVIS [15] taxonomy and retain the 51 categories that simultaneously (a) appear in Objaverse with at least 5 candidate assets, (b) admit a meaningful articulated skeleton (excluding categories such as `liquid`, `powder` that have no joint structure). The full 51 list is in Appendix M.
2. *Per-category sampling.* Within each category we draw two assets uniformly at random from the Objaverse pool, subject to two automatic filters: (i) the mesh is watertight or has ≤ 1 boundary loop; (ii) vertex count is in $[500,\ 40000]$. We take the first two random samples that pass the filters—no manual quality curation, so the benchmark is reproducible from a seed.
3. *Validity filter.* 98 of 102 admit valid UniRig outputs and form the evaluation set. The 4 failures (non-manifold or single-bone meshes) are retained in the released codebase rather than discarded silently.

M Per-LVIS-category breakdown of Balanced-102

Table 17 expands the UniRig-initialized results in Table 1 into the 51 underlying LVIS categories. Each cell averages per-model reductions over the valid UniRig initializations in that category; the counts sum to 98, not 102. Fragility improves in $50/51$ category means. The `butterfly` row improves fragility by 9.7% but worsens ARAP energy by 4.7% and Jacobian anisotropy by 14.0%; `table` worsens all four metrics. Category means do not imply that every asset improves (Sec. 4.2).

Table 17: **UniRig → UniRig+PIS: per-LVIS-category breakdown.** Mean of per-model percentage reductions over 98 valid initializations across 51 categories. N is the number of valid models in each category. Positive means improvement; fragility improves in $50/51$ category means.

Category	N	Frag. ↑	ARAP ↑	Jac. ↑	Vol. ↑
Christmas_tree	2	+29.0%	+31.6%	+20.7%	+56.2%
airplane	2	+47.2%	+50.9%	+38.6%	+52.3%
armchair	2	+29.4%	+30.6%	+7.1%	+31.8%
bear	2	+43.2%	+49.4%	+26.3%	+51.0%
bed	2	+94.3%	+96.0%	+25.3%	+79.5%
bench	2	+33.8%	+35.6%	+18.0%	+28.6%
bicycle	2	+87.9%	+92.7%	+55.6%	+87.2%
boat	2	+46.7%	+49.1%	+25.4%	+54.6%
butterfly	2	+9.7%	−4.7%	−14.0%	+12.6%
cabinet	1	+86.1%	+87.3%	+38.1%	+85.1%
car_(automobile)	2	+85.1%	+85.3%	+39.6%	+85.0%
cat	2	+17.5%	+37.3%	+14.0%	+26.9%
chair	2	+25.1%	+27.4%	+11.7%	+47.2%
cow	2	+75.2%	+91.0%	+59.2%	+74.4%
dog	2	+16.6%	+19.5%	+13.2%	+24.4%
doll	2	+16.0%	+21.4%	+13.8%	+37.4%
dolphin	2	+30.1%	+37.9%	+29.9%	+46.1%
duck	2	+13.8%	+14.3%	+9.9%	+20.8%
eagle	2	+43.0%	+45.1%	+26.6%	+47.7%
elephant	2	+29.5%	+36.4%	+16.1%	+27.9%
fish	2	+49.1%	+50.5%	+35.9%	+46.0%
frog	2	+26.3%	+30.1%	+19.6%	+37.8%
giraffe	2	+31.6%	+34.6%	+19.7%	+45.9%
gun	1	+68.3%	+72.8%	+34.7%	+59.5%
hammer	2	+10.6%	+6.1%	+5.0%	+57.6%
horse	2	+7.7%	+9.8%	+5.5%	+21.1%
knife	2	+37.0%	+53.6%	+14.7%	+36.6%
lamp	2	+27.9%	+29.8%	+7.5%	+33.8%
lion	2	+18.9%	+23.3%	+10.3%	+22.7%
lizard	2	+20.6%	+21.5%	+11.2%	+36.8%
monkey	2	+43.0%	+47.5%	+31.3%	+54.7%
motorcycle	2	+72.2%	+78.8%	+38.2%	+69.6%
owl	2	+37.5%	+44.0%	+18.2%	+35.2%
parrot	2	+23.2%	+27.5%	+7.8%	+30.5%
penguin	2	+24.8%	+34.7%	+20.7%	+41.1%
rabbit	2	+36.8%	+41.1%	+25.1%	+49.5%
scissors	2	+36.6%	+34.4%	+20.7%	+45.2%
shark	2	+46.6%	+51.1%	+22.4%	+47.5%
sheep	1	+10.7%	+10.0%	+8.2%	+41.1%
snake	2	+26.7%	+27.7%	+14.8%	+38.3%
sofa	2	+49.8%	+56.6%	+28.6%	+55.6%
spider	2	+74.7%	+78.6%	+36.0%	+65.1%
squirrel	2	+42.7%	+51.7%	+20.7%	+36.3%
sword	1	+4.7%	+3.8%	+1.0%	+51.8%
table	2	−18.4%	−21.1%	−4.7%	−7.8%
teddy_bear	2	+11.2%	+12.6%	+7.5%	+24.4%
tiger	2	+17.3%	+25.7%	+7.5%	+13.1%
truck	2	+53.2%	+52.7%	+25.7%	+67.8%
turtle	2	+33.3%	+41.6%	+16.9%	+25.9%
wolf	2	+8.7%	+8.8%	+4.8%	+35.0%
zebra	2	+32.6%	+37.8%	+21.6%	+39.3%

N Polar midpoint and rigid-motion controls

Exact midpoint identity. Let $Q = R_1^\top R_2 = \exp(\Omega)$ with principal rotation angle $\alpha < \pi$. Then

$$\tfrac{1}{2}(R_1 + R_2) = R_1 \exp(\Omega/2) \underbrace{\tfrac{1}{2}[\exp(-\Omega/2) + \exp(\Omega/2)]}_{H}. \tag{10}$$

H is symmetric positive definite: its eigenvalues are 1 and $\cos(\alpha/2)$ (twice). Thus this is the polar decomposition, and the orthogonal factor is precisely the geodesic midpoint $R_1 \exp(\Omega/2)$. No approximation in α is needed for this identity. At $\alpha = \pi$, the arithmetic mean is singular and the unique-midpoint argument no longer applies. Our bounded endpoint rotations avoid this case. Sampling and midpoint computation are not differentiation variables; only the skinning weights are optimized.

Existing numerical controls. For $10{,}000$ rotation pairs with endpoint angles bounded by 0.5 rad, the largest observed relative angle is 0.955 rad and the minimum singular value of the arithmetic mean is 0.888. The maximum angular discrepancy between polar and an independently implemented geodesic midpoint is 2.98×10^{-8} rad; maximum orthogonality and determinant errors are 5.44×10^{-16} and 4.44×10^{-16}, respectively. These establish numerical agreement in the tested regime, not a need to differentiate through rotations.

The surface-error control applies 256 global rigid transformations in float64. Maximum P95 edge-strain and ARAP membrane errors are 4.22×10^{-16} and 3.28×10^{-21}. A non-rigid scaling of one coordinate by 1.25 gives positive errors. These controls concern the rigid-invariant surface diagnostics below, not the PIS residual or the normal-direction component of fragility.

O Direct motion-preservation protocol and results

Protocol and iteration budgets. The supplementary tests keep each mesh and skeleton fixed, optimize weights with the same PIS residual, and use 16 pose pairs per optimization step, Adam at 0.05, and spatial regularization weight 2.0. No attachment-to-initial weight penalty is used. Each local endpoint rotation has random axis and angle in $[-0.5, 0.5]$ rad; the root is fixed. Evaluation uses 64 independently seeded poses shared by each before/after pair. The UniRig Balanced-98, RigAnything, and valid Blender tests use 50 steps; artist-painted RigNet and MIA tests use 500. These budgets are stated separately from the original 500-step benchmark and its wall-clock timing.

Motion diagnostics. Let $D^0_{ti} = X^0_{ti} - v_i$ and $D^*_{ti} = X^*_{ti} - v_i$ be rest-to-pose displacements before and after refinement. For a vertex subset $\mathcal{V}$,

$$a_{\mathcal{V}} = \frac{\left(\sum_{t,i\in\mathcal{V}} \|D^*_{ti}\|^2\right)^{1/2}}{\left(\sum_{t,i\in\mathcal{V}} \|D^0_{ti}\|^2\right)^{1/2}}, \qquad \mathrm{NRMTE} = \frac{\|X^* - X^0\|_F}{\|D^0\|_F}. \tag{11}$$

Global amplitude uses all vertices. High-motion vertices are the top quartile of initial displacement norms; blend-region vertices have $\max_j w^0_{ij} < 0.9$. Vertices with initial displacement below 10^{-5} times the rest bounding-box diagonal are excluded when defining these subsets. Part-centroid amplitude is the median ratio across dominant-bone groups with at least eight vertices. Direction cosine averages, over high-motion vertices, the cosine between each vertex's flattened before/after trajectory. Activity-weighted attachment compares mean bone-induced displacement weighted by the original and refined weights, using at most $2{,}048$ deterministically sampled vertices. All table entries aggregate these per-model statistics across models.

Statistics and outliers. The statistical unit is the model, never a vertex, pose, or frame. The 95% intervals for medians use $10{,}000$ model-bootstrap resamples. TOST is applied to mean log amplitude ratios with bounds $[\log(0.95), \log(1.05)]$ at significance 0.05. This tests aggregate equivalence of the geometric mean, not equivalence of every individual model. Surface comparisons use one-sided paired Wilcoxon tests with Holm correction over the five surface metrics; the SkinCells comparison instead uses two-sided paired tests.

Table 18: **UniRig-initialized motion diagnostics** ($n = 98$, 50 steps). Intervals and within-margin counts are model-level. The ideal amplitude ratio is 1.

Diagnostic	Median	95% bootstrap CI	Within $[0.95, 1.05]$
Global amplitude	0.9993	$[0.9978, 0.9999]$	$88/98$
High-motion amplitude	0.9990	$[0.9968, 0.9998]$	$90/98$
Blend-region amplitude	0.9988	$[0.9974, 1.0003]$	$83/98$
Part-centroid amplitude	0.9995	$[0.9989, 0.9999]$	$88/98$
Activity-weighted attachment	0.9998	—	$84/98$

For artist-painted RigNet weights, global amplitude is 0.9993 (95% CI $[0.9990, 0.9996]$), with $267/270$ within $\pm 5\%$; the high-motion count is $268/270$, and the part-centroid count is $266/270$. This supports preservation on already-good rigs, but does not rule out individual changes. Generated-weight inputs show larger exceptions: the zebra case has global/high-motion amplitude ratios $0.907/0.880$ and direction cosine 0.960. Such cases are not evidence of artifact removal without motion change.

Rigid-invariant surface diagnostics. For each pose we measure absolute edge strain $|\|e'\|/\|e\| - 1|$ (mean and 95th percentile over edges); ARAP membrane energy $\frac{1}{2}[(s_1 - 1)^2 + (s_2 - 1)^2]$; anisotropy s_1/s_2; and absolute log-area change $|\log(s_1 s_2)|$, where $s_1 \geq s_2$ are singular values of the triangle's 3×2 tangent-plane deformation gradient. Face errors are averaged over valid faces, then over held-out poses. Zero-area/sliver rest triangles with $2\sqrt{3}\,\|e_1 \times e_2\|/(\|e_1\|^2 + \|e_2\|^2 + \|e_3\|^2) \leq 10^{-4}$ and edges shorter than 10^{-7} bounding-box diagonals are excluded from these supplementary surface diagnostics. These measures are not numerically interchangeable with the original four-metric evaluation.

Table 19: **Held-out surface errors on Balanced-98** (50 steps). Reduction is the median of per-model percentage reductions; it is not the percentage reduction of the pooled mean.

Metric	Median reduction	Models improved	Holm p
ARAP membrane	62.6%	$82/98$	6.93×10^{-9}
Mean edge strain	40.5%	$92/98$	2.12×10^{-14}
P95 edge strain	41.4%	$94/98$	2.55×10^{-15}
Anisotropy	10.2%	$91/98$	2.63×10^{-13}
Absolute log-area	33.5%	$92/98$	2.88×10^{-15}

ARAP illustrates the importance of reporting robust and aggregate statistics separately: although $82/98$ models improve and the median reduction is 62.6%, the mean raw error rises from 426.422 to 455.285 (6.77% worse) because of outliers. We do not claim uniform or mean ARAP improvement in this 50-step supplementary test. The original 500-step table uses a different evaluation protocol.

P Additional-initializer and SkinCells evaluation details

Generality and its limits. RigAnything is evaluated on 20 assets from 10 categories with its own predicted skeletons and weights; MIA uses eight compatible assets from four categories. Within each before/after pair, the skeleton and evaluation poses are unchanged. RigAnything's median global/high-motion/part ratios are $0.9746/0.9788/0.9764$; $16/20$ global ratios lie within $\pm 5\%$ and median NRMTE is 0.1204. MIA's ratios are $0.9821/0.9892/0.9934$ with $7/8$ global ratios within $\pm 5\%$ and median NRMTE 0.1480. MIA does not pass the aggregate global-motion equivalence test. Its five surface comparisons give Holm $p = 0.0195$, but the small sample and motion outlier limit the claim.

Blender Bone Heat. Blender 4.3 Automatic Weights is tested on 20 generated meshes; nine produce valid paired outputs. The other eleven either emit Blender's failure warning or leave more than 10% of vertices without a deform group. On the nine valid outputs, 50-step PIS gives the mean errors below. Only P95 edge strain survives Holm correction ($p = 0.0195$); all other corrected p values are at least 0.109.

Table 20: **Compatibility with valid Blender Bone Heat weights** ($9/20$ assets). Lower surface error is better.

Method	ARAP	Mean edge	P95 edge	Anisotropy	Log-area
Blender	4.940	0.0641	0.2426	1.2493	0.0865
+ PIS	1.487	0.0452	0.1551	1.1947	0.0758

Median global/high-motion amplitude ratios are $0.9932/0.9981$; only $5/9$ global and $6/9$ high-motion ratios lie within $\pm 5\%$. This is a compatibility check on valid outputs, not evidence of universal motion preservation or statistical dominance over Blender.

SkinCells implementation and protocol. The independent implementation follows public SkinCells [27] v2 Eqs. 2–8, including the literal one-based D_l boundary. It uses six sites per joint, at most four influences per vertex, Adam at 10^{-3}, $1{,}500$ steps, a $1{,}024$-pose pool with batches of 16, and loss weights 1, $6{,}000$, and 1 for deformation, location, and sparsity. The shared subset has two assets each from bear, cat, cow, dog, elephant, giraffe, horse, monkey, rabbit, and zebra categories. Both methods use the same UniRig skeletons and 64 held-out poses. PIS uses 50 steps; the methods are not compared at equal iteration counts.

For each metric and model we divide the method's error by the same model's UniRig error, then report the median normalized error in Table 7. Two-sided paired Wilcoxon p values on absolute model errors are 0.812 (ARAP), 0.245 (mean edge), 0.177 (P95 edge), 0.452 (anisotropy), and 0.430 (log-area). Thus the smaller PIS medians do not establish a statistically significant advantage. The observed median global motion ratios are 0.9991 for PIS and 0.9807 for SkinCells, with $18/20$ and $10/20$ within $\pm 5\%$. Results from one independent implementation should not be read as exhausting the performance of the authors' implementation.

Volumetric BBW: output availability, not a selected quality comparison. The strict pipeline uses TetGen followed by libigl volumetric BBW; a surface-triangle biharmonic solve is not counted as volumetric BBW. Six usable outputs are recorded among the 98 benchmark targets (6.1% observed output availability). There are 63 explicit failure records, including self-intersections, and 29 targets without a terminal status record. We therefore do not classify all 92 remaining assets as confirmed failures or timeouts, nor claim a completed 98-asset failure-rate audit. Reporting quality only on the six successes would also select unusually tractable inputs. PIS's $98/98$ refinement result is conditional on having a valid UniRig skeleton and weights.

Runtime scope. The original 36.1 s number is a mean offline wall-clock measurement at 500 steps (Table 16). Supplementary records report optimization-only median times of 0.628 s for 50 steps and 5.990 s for 500 on Balanced-98, excluding loading and evaluation. These implementation- and scope-specific timings must not be compared directly with the original wall-clock mean or described as millisecond preprocessing. Both protocols leave the playback LBS operator unchanged.

Q Limitations and ongoing work

We list the known limitations and items planned for follow-up work, separated from the body so that the body shows established results only.

Animation-quality user study. The reported surface measures are geometric proxies, supplemented by direct trajectory comparisons with the initializer. They do not substitute for trained animators' judgments of artifact severity. No human-study result is included in this manuscript. A documented blinded animator comparison remains important future work.

Skeleton-side defects. PIS refines weights only and cannot repair an upstream skeleton with the wrong joint count, placement, orientation, hierarchy, or topology. Residual distortion after refinement can motivate skeleton-side inspection, but does not provide a unique decomposition of weight and skeleton errors.

Theoretical scope. Lemma 1 is a small-separation statement along smooth product-$SO(3)$ geodesics; it does not make the residual an artifact classifier. The pure-PIS simplex quadratic in Sec. 3.4 omits the spatial prior. Neither the theory nor aggregate equivalence tests guarantee preservation of every model's motion: some generated-weight initializations show amplitude shrinkage, including the reported zebra outlier. Small MIA and Blender subsets and a single independent SkinCells implementation further limit the scope of generalization.

NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

   Justification: The original headline geometric results are supported by Table 1; additional motion and initializer results appear in Secs. 4.6–4.7. The abstract distinguishes weight refinement from skeleton repair and does not claim a human perceptual result.

   Guidelines:

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

2. **Limitations**

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

   Answer: [Yes]

   Justification: Secs. 3.3 and 4.6 distinguish valid rotational curvature from artifacts and report motion outliers. Appendix Q discusses fixed-skeleton scope, sample sizes, perceptual limitations, and the qualified BBW audit.

   Guidelines:

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

3. **Theory assumptions and proofs**

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

Answer: [Yes]

Justification: Lemma 1 uses smooth geodesic paths and a fourth-order Taylor remainder; Proposition 1 gives the pure-PIS quadratic's zero-loss and simplex optimality conditions with proof. Appendix N proves exact polar/geodesic midpoint agreement below the antipodal angle.

Guidelines:

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

4. **Experimental result reproducibility**

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

Answer: [Yes]

Justification: Algorithm 1 gives the original procedure; the supplementary budgets, held-out poses, metric definitions, and statistical unit are specified in Appendices O–P. The BBW audit explicitly distinguishes recorded successes, failures, and unresolved targets.

Guidelines:

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

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

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

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

   Answer: [No]

   Justification: Code, main-table runners, and result records are public at `https://github.com/Fugtemypt123/PIS`. A complete self-contained data and artifact bundle is not provided; third-party assets and predictor checkpoints are obtained separately under their respective licenses. Reproduction details and supplementary protocols are described in the paper and appendices.

   Guidelines:

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

6. **Experimental setting/details**

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

   Answer: [Yes]

   Justification: Sec. 4.1 and the original appendices specify the benchmarks and 500-step baseline protocol. Supplementary tests explicitly separate 50- and 500-step budgets, fixed skeletons, and independently seeded 64-pose evaluation (Appendix O).

   Guidelines:

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

7. **Experiment statistical significance**

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

   Answer: [Yes]

Justification: The original results include paired tests and percentile bands. Supplementary analyses use model-bootstrap intervals, paired Wilcoxon tests with Holm correction, and TOST on log motion ratios; nonsignificant SkinCells differences and individual motion failures are reported explicitly.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

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

   Answer: [Yes]

   Justification: Appendix K gives the original H200 wall-clock measurements; Appendix P separates these from optimization-only timings. The reported measurements do not include a complete accounting of exploratory and failed runs.

   Guidelines:

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

9. **Code of ethics**

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

   Answer: [Yes]

   Justification: The reported experiments use public 3D resources and numerical optimization. No human-subject results are reported in this version; the camera-ready paper identifies its authors and includes funding and competing-interest disclosures.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.

- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

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

    Answer: [N/A]

    Justification: PIS is a numerical method for refining linear-blend skinning weights on 3D meshes. It is downstream of mesh and rigging predictors, does not produce new images or videos of humans, and has no direct path to applications such as deepfakes, surveillance, or automated decision-making affecting individuals. The likely use is animator productivity for film, games, and AR/VR pipelines.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

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

    Answer: [N/A]

    Justification: The released artifacts are skinning-weight matrices and a Python implementation of a numerical optimizer. Neither is a high-risk asset class (not a generative model, not a scraped dataset of human content).

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.

- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

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

    Answer: [Yes]

    Justification: Objaverse [10, 9], RigNet [71], UniRig [75], BBW [19], and Make-It-Animatable [14] are all cited; we use them under their public licenses (Objaverse: ODC-By; UniRig and RigNet code under the licenses provided by their authors). PyTorch3D [53] (BSD) is used for differentiable rendering and forward kinematics.

    Guidelines:

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

13. **New assets**

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

    Answer: [Yes]

    Justification: Two new assets: (i) the *Balanced-102* benchmark—a stratified sample of 102 Objaverse assets across 51 LVIS categories—is documented in Sec. 4.1 and Appendix M, with per-model result records in the public repository; (ii) the PIS reference implementation—algorithm in Appendix B, hyperparameters in Sec. 3.6 and Appendix A, and public code at `https://github.com/Fugtemypt123/PIS`.

    Guidelines:

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

14. **Crowdsourcing and research with human subjects**

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

    Answer: [N/A]

    Justification: No crowdsourcing or human-subject results are reported in this manuscript.

Guidelines:

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

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

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

    Answer: [N/A]

    Justification: No human-subject study is included in the reported experiments; no claim about an unreported study's ethics approval is made.

    Guidelines:

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

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The reported PIS method is a numerical optimizer, and the reported evaluations use geometric and trajectory diagnostics rather than an LLM as a core method or evaluator.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.