PROXYBOROS: MOTION-FAITHFUL DYNAMIC SELF-INTERSECTION REPAIR FOR RIGGED MESH ANIMATIONS

Anonymous authors
Paper under double-blind review

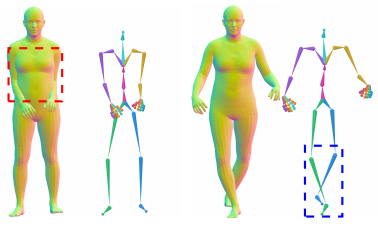
ABSTRACT

Physically plausible animation of rigged shapes is essential for realistic and high-quality 3D content creation. However, motion sequences produced by existing generation models often exhibit surface collisions and self-intersections, resulting in physically implausible geometry and noticeable visual artifacts. Resolving these invalid surface geometries while faithfully preserving the spatiotemporal characteristics of the original motion remains a challenging yet under-explored problem. To address this, we propose *Proxyboros*, an optimization framework for dynamic self-intersection repair that alternates mesh-space repair with rig-space pose adaptation to progressively guide the evolving rig toward temporally coherent skeletal motion. Within each step, *Mesh-Space Proxy Exploration* first extracts intermediate geometric proxies that capture local repair tendencies before independent repairs crystallize into temporally inconsistent static solutions. *Proxy-Guided Pose Adaptation* then incorporates this guidance into a temporally coupled skeletal trajectory using surface-aware proxy guidance, absorbing stable surface-separation cues while relaxing ambiguous local displacements. Experiments on SMPL-H sequences and diverse non-SMPL rigged models demonstrate favorable overall performance over state-of-the-art methods in repair quality, pose preservation, and motion fidelity, with demonstrated applicability across heterogeneous mesh and skeleton topologies. See our anonymous project page for more video demos: https://anonymousiclr2724.github.io/Proxyboros.

1 INTRODUCTION

The rapid progress of generative 3D models is shifting animated content creation from labor-intensive work toward increasingly automated pipelines. Beyond producing plausible shapes and motions, high-quality animation also requires the underlying geometry to remain valid throughout motion. However, current motion generation and editing models often drive body parts into unintended contact, leaving visible self-intersections and penetrations in the resulting meshes. This poses an emerging requirement of dynamic self-intersection repair for rigged mesh animations.

This problem is challenging for two main reasons. ***First***, self-intersection repair requires fine-grained local geometric corrections, whereas the skeleton-based deformation of a rigged shape is globally coupled due to its kinematic structure. Self-intersections typically occur between a small number of mesh triangles and thus require localized adjustments to separate the colliding surfaces. However, directly incorporating such local corrections into the articulated motion space can propagate their influence through the kinematic chain, inducing unnecessary deformation in non-colliding regions and causing the repaired shape to deviate substantially from the original motion (refer to Fig. 1). ***Second***, eliminating self-intersections on individual frames does not necessarily ensure temporal consistency across the sequence. Independently correcting each frame may result in substantially different adjustments over time, introducing temporal jitter



(a) Input **(b)** Rig-space ISIR

Figure 1: Kinematic coupling in rig-space repair. Correcting the self-intersection (red) may induce unintended deformation elsewhere (blue).

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

and motion discontinuities. Conversely, enforcing strong temporal constraints can restrict the local deformations necessary for collision resolution, leading to an inherent conflict between frame-wise geometric validity and sequence-level motion coherence.

The lack of large-scale, high-quality collision-free motion data further compounds these challenges. Existing large-scale 3D datasets with articulated animations often contain a mixture of valid and self-intersecting motion sequences, such as (Deitke et al., 2023b;a), where paired examples of corrupted and collision-free motions are largely unavailable. This makes it difficult to directly learn reliable collision correction from data using conventional supervised approaches. Therefore, developing a generalizable framework that can resolve local self-intersections without relying on collision-free training data, while faithfully preserving the original motion and its temporal coherence, remains a challenging and important problem.

To address these challenges, we propose *Proxyboros*, a general framework for dynamic self-intersection repair of rigged mesh animations. The core idea is to represent local repair tendencies with intermediate mesh proxies and progressively reconcile them with skeletal motion in rig space. To achieve this, we introduce an alternating optimization scheme between mesh and rig spaces. We first perform *Mesh-Space Proxy Exploration* to extract geometric guidance from local self-intersection repair before frame-wise optimization commits to potentially inconsistent static solutions. The resulting proxies are then incorporated by *Proxy-Guided Pose Adaptation*, which projects them into a temporally continuous rig space to recover coherent skeletal motion consistent with the repair guidance. We parameterize this continuous rig space with cubic B-spline pose corrections, coupling neighboring frames through shared temporal control variables. Since independently explored proxies may still contain uncertain or locally inconsistent correction directions, we further introduce surface-aware alignment with proxy guidance to selectively retain separation cues supported by the local surface geometry while relaxing ambiguous local displacements. This allows the skeletal motion to preserve the collision-resolving tendency revealed by the proxies while providing sufficient spatial flexibility to reconcile conflicting frame-wise corrections, without unnecessarily altering collision-free regions. By progressively bridging the local mesh space and the temporally coherent rig space, Proxyboros effectively reconciles fine-grained geometric repair with globally coupled, temporally coherent skeletal motion. It requires neither collision-free training data nor specific body models or learned motion priors, supporting broad applicability across diverse geometries, rig structures, and motion sequences. Our contributions are summarized as follows.

- We propose *Proxyboros*, a dynamic self-intersection repair framework for diverse rigged mesh animations that progressively reconciles local geometric repairs with coherent skeletal motion via intermediate mesh proxies.
- We introduce a shared B-spline pose parameterization and surface-aware proxy guidance to integrate frame-wise repair cues into coherent skeletal motion while accommodating geometric ambiguity, improving motion fidelity without compromising repair effectiveness.
- Extensive experiments demonstrate state-of-the-art overall performance regarding repair quality, pose preservation, and motion fidelity, with demonstrated applicability across diverse geometries, skeleton topologies, and rig configurations.

2 Related Work

Collision-Aware Deformable Simulation. Collision handling has been well studied in deformable-surface simulation. Early work developed robust collision response for cloth dynamics (Bridson et al., 2002) and asynchronous contact handling with non-penetration constraints (Harmon et al., 2009). More recent barrier-based methods, including incremental potential contact (IPC) (Li et al., 2020) and CoDim-IPC (Li et al., 2021), preserve intersection-free trajectories by preventing contacting primitives from crossing. Geometric repulsion energies such as tangent-point energy (TPE) are utilized to maintain surface embeddedness (Yu et al., 2021) and collision-free deformation (Sassen et al., 2024). These methods primarily preserve an initially valid configuration during deformation, whereas our setting addresses self-intersections already present in a given animated sequence.

Static Self-Intersection Repair. Repairing pre-existing self-intersections is inherently ambiguous because collision history and valid reference configurations are typically unavailable. Early methods infer valid separation through explicit geometric constructions, including intersection-contour minimization (Volino & Magnenat-Thalmann, 2006) and discrete penetration stencils (Ye et al., 2017).

Volumetric approaches lift surface intersections to auxiliary volume representations, using tetrahedral discretizations (Sacht et al., 2013), immersed volumes (Li & Barbič, 2018), or differentiable voxelization (Djuren et al., 2026) to guide geometric untangling. Learning-based approaches further approximate collision constraints or collision handling strategies from data (Tan et al., 2021; 2022). Recent methods instead optimize the surface geometry directly. ISIR (Jang et al., 2025) evaluates signed TPE over intersecting triangle pairs and propagates sparse collision gradients across the mesh, while Mesh-Utg (Minarčík et al., 2026) combines global repulsion with local mesh penalties for complex untangling. Despite strong performance on static meshes, these methods do not address temporal and rig consistency in dynamic sequences. Independent static repairs may select different separation directions for temporally related collisions, whereas unconstrained vertex deformation can move the repaired geometry outside the original rig's deformation space. These limitations necessitate novel repair formulations that jointly preserve temporal coherence and skeletal realizability.

Self-Collision Resolution in Dynamic Meshes. Existing work on dynamic self-collision has mainly focused on collision-aware motion generation. Herrmann et al. (2025) introduced a sphere-based body proxy and self-intersection loss to suppress penetrations during motion synthesis, while post-hoc repair of existing motion sequences remains less explored. Most closely related to our work, PoseShield (Li et al., 2026) learns an Eikonal-regularized collision field in the SMPL-H (Loper et al., 2015; Romero et al., 2017) pose space and combines it with a pretrained motion prior for sequence-level correction. However, its formulation relies on a predefined body representation, collision-labeled training data, and an external motion prior. In contrast, our method derives repair guidance directly from the observed geometry and optimizes within the original rig deformation space, enabling repair across heterogeneous mesh and skeleton topologies.

3 Proposed Method

3.1 Problem Statement and Method Overview

We consider a rigged mesh animation $\{\mathcal{M}_t\}_{t=1}^{T}$ of arbitrary length T. The underlying rig consists of a rest-pose triangular mesh with vertices $\boldsymbol{V}_{\text{rest}} \in \mathbb{R}^{N_v \times 3}$ and face connectivity $\boldsymbol{F}$ containing N_f triangular faces, a kinematic skeleton $\boldsymbol{S}$ with J joints, and the corresponding skinning weights $\boldsymbol{W}_{\text{skin}} \in \mathbb{R}^{N_v \times J}$. Mesh connectivity and skeleton topology may vary across animations but remain fixed within each animation. At frame t, $\mathcal{M}_t = \{\boldsymbol{V}_t, \boldsymbol{\Theta}_t, \boldsymbol{F}\}$, where $\boldsymbol{\Theta}_t = \{\boldsymbol{\theta}_{t,j}\}_{j=1}^{J} \in \mathbb{R}^{J \times 3}$ collects the parent-relative local axis-angle rotation vectors of the skeleton joints, and $\boldsymbol{V}_t \in \mathbb{R}^{N_v \times 3}$ denotes the posed vertex positions obtained through Linear Blend Skinning (LBS) (Magnenat-Thalmann et al., 1989; Capell et al., 2002), written as $\boldsymbol{V}_t = \mathcal{D}_{\text{LBS}}(\boldsymbol{V}_{\text{rest}}, \boldsymbol{\Theta}_t; \boldsymbol{S}, \boldsymbol{W}_{\text{skin}})$. Skeletal motion may induce self-intersections in the posed geometry. Given such an animation, we aim to estimate relative pose corrections $\Delta\Theta_{1:T} = \{\Delta\boldsymbol{\Theta}_t\}_{t=1}^{T}$, where $\Delta\boldsymbol{\Theta}_t = \{\Delta\boldsymbol{\theta}_{t,j}\}_{j=1}^{J} \in \mathbb{R}^{J \times 3}$, such that the repaired pose trajectory $\widehat{\Theta}_{1:T}$ removes motion-induced self-intersections while faithfully preserving the characteristics of the original motion. The repaired geometry $\widehat{\boldsymbol{V}}_t = \mathcal{D}_{\text{LBS}}(\boldsymbol{V}_{\text{rest}}, \widehat{\boldsymbol{\Theta}}_t; \boldsymbol{S}, \boldsymbol{W}_{\text{skin}})$ of frame $\widehat{\mathcal{M}}_t$ remains entirely within the original rig deformation space.

To jointly resolve self-intersections and preserve coherent skeletal motion, we propose *Proxyboros*, a recurrent optimization framework that alternates between geometric repair in mesh space and pose adaptation in rig space, as illustrated in Fig. 2. The key idea is to use intermediate proxy meshes as a bridge between the two spaces, carrying local collision-resolving guidance from mesh-space exploration into rig-space optimization, while ensuring that its incorporation remains compatible with the kinematic and temporal structure of skeletal motion. Through interaction between the two spaces, Proxyboros reconciles local geometric correction with the rig's motion constraints, reducing self-intersections while preserving the spatiotemporal characteristics of the original animation.

3.2 Progressive Mesh-Rig Interaction in Proxyboros

Direct rig-space correction entangles localized collision resolution with globally coupled skeletal motion. A repair signal originating from a small intersecting region can propagate through the skeleton articulation and induce unnecessary deformation beyond the collision area. Optimizing in mesh-space avoids this kinematic coupling and provides greater freedom for local geometric correction. However, static self-intersection repair is generally non-unique, with neighboring frames converging toward different valid separation directions. When such independently converged repairs

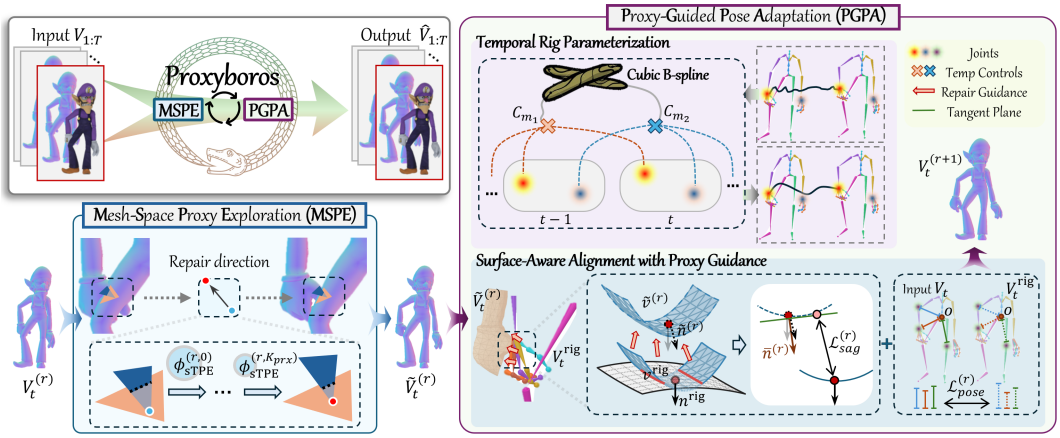


Figure 2: Overview of *Proxyboros*. Each iteration first explores self-intersection repair directions and retains the intermediate geometry as a proxy. The proxy then drives pose adaptation under temporal B-spline controls and surface-aware geometric guidance. The resulting rigged sequence initiates the next round, gradually converting local repair directions into consistent skeletal motion.

are imposed on the rig, their frame-wise discrepancies become inconsistent geometric guidance for skeletal optimization, potentially distorting the original motion or leaving residual self-intersections.

We represent local repair guidance with intermediate proxy meshes before frame-wise optimization converges to a certain static solution. These proxies encode locally effective separation directions while remaining adaptable to the evolving skeletal motion. At iteration r, let $\boldsymbol{V}_{1:T}^{(r)}$ denote the current rigged sequence, with $\boldsymbol{V}_{1:T}^{(0)} = \boldsymbol{V}_{1:T}$. Each iteration comprises two coupled steps:

- *Mesh-Space Proxy Exploration*: To expose a local repair direction without immediately propagating its effect through the skeleton, we first explore the self-intersection correction in the ambient vertex space. Starting from $\boldsymbol{X}_t^{(r,0)} = \boldsymbol{V}_t^{(r)}$, we perform a short repair descent in the mesh space (detailed in Appendix) using signed TPE ϕ_{sTPE} together with a local structure constraint (Jang et al., 2025) to search the potential repair direction. The optimization is truncated after K_{prx} steps to avoid fully repairing, so that the proxy mesh $\widetilde{\boldsymbol{V}}_t^{(r)} = \boldsymbol{X}_t^{(r,K_{\text{prx}})}$ encodes a locally favorable collision-resolving direction. Applying this exploration independently across frames produces $\widetilde{\boldsymbol{V}}_{1:T}^{(r)}$, which provides geometric guidance for the subsequent pose adaptation.
- *Proxy-Guided Pose Adaptation*: The local repair states carried by the proxy sequence may not lie on a temporally coherent trajectory realizable by the original rig. Directly reproducing these frame-wise corrections in rig space would reintroduce their ambiguity into the coupled skeletal motion. To address this, we formulate pose adaptation as a regularized projection of the proxy sequence onto a temporally continuous rig space, retaining repair cues that can be consistently expressed by a coherent pose trajectory while suppressing frame-wise variations incompatible with the coupled skeletal motion. We first define the feasible trajectory space by parameterizing the pose trajectory using shared temporal controls that couple neighboring pose corrections over time. Within this temporally coupled space, proxy-guided surface-aware alignment allows the rig to follow geometry-supported surface-separation cues from the proxies without forcing it to reproduce ambiguous local displacements. A root-relative joint regularization further limits excessive deviation from the original skeletal articulation. Optimizing these constraints yields the updated pose trajectory $\Theta_{1:T}^{(r+1)}$ and the geometry $\boldsymbol{V}_t^{(r+1)} = \mathcal{D}_{\text{LBS}}(\boldsymbol{V}_{\text{rest}}, \boldsymbol{\Theta}_t^{(r+1)}; \boldsymbol{S}, \boldsymbol{W}_{\text{skin}})$ which initializes the next round of mesh-space exploration.

3.3 Proxy-Guided Pose Adaptation

The goal of Proxy-Guided Pose Adaptation is to project the geometric guidance encoded by $\widetilde{\boldsymbol{V}}_{1:T}^{(r)}$ into a coherent skeletal motion within the original rig deformation space. Given the frame-wise ambiguity of the proxy guidance, pose adaptation must maintain temporal consistency without overcommitting to uncertain local corrections. We address these requirements by representing the pose trajectory using shared temporal controls and introducing a proxy-guided surface-aware alignment

objective that regulates the entry of local repair guidance into the rig-space optimization. A root-relative joint regularization further limits excessive deviation from the original skeletal articulation.

Temporal Rig Parameterization. To model temporally smooth and coherent motion sequences, we represent the pose corrections using a shared set of temporal control variables. Specifically, we parameterize the relative pose offsets with a cubic B-spline. Let K denote the number of temporal control variables and $\boldsymbol{C} \in \mathbb{R}^{K\times J\times 3}$ their learnable pose offsets. For frame t, we define the normalized timestamp as $\tau_t = (t-1)/(T-1) \in [0,1]$. Over this normalized temporal domain, we construct a clamped uniform cubic B-spline (De Boor, 1972) with the knot vector

$$\boldsymbol{\xi} = \left[0,0,0,0,\ \frac{1}{K-3}, \frac{2}{K-3}, \ldots, \frac{K-4}{K-3},\ 1,1,1,1\right]. \tag{1}$$

The repeated boundary knots clamp the spline at both endpoints, and the $K-4$ interior knots are uniformly distributed over $(0,1)$. Let $\beta_m(\tau)$ denote the m-th cubic B-spline basis function associated with $\boldsymbol{\xi}$. Each β_m is a locally supported piecewise cubic polynomial. Evaluating these K basis functions at all frame timestamps yields a fixed temporal basis matrix $\boldsymbol{B} \in \mathbb{R}^{T\times K}$, with $\boldsymbol{B}_{t,m} = \beta_m(\tau_t)$, where $\boldsymbol{B}_{t,m}$ specifies the contribution of the m-th temporal control variable to frame t. The relative pose correction is then interpolated as

$$\Delta\boldsymbol{\Theta}_t = \sum_{m=1}^{K} \boldsymbol{B}_{t,m}\boldsymbol{C}_m. \tag{2}$$

Due to the shared local supports of the cubic basis functions, adjacent frames depend on overlapping subsets of temporal control variables, enabling coupling of their pose corrections within local temporal neighborhoods. This coupling reconciles frame-wise variations in the proxy guidance while retaining sufficient flexibility for geometric correction. The interpolated offset $\Delta\boldsymbol{\Theta}_t = \{\Delta\boldsymbol{\theta}_{t,j}\}_{j=1}^{J}$ provides the relative joint-rotation corrections for frame t, which are composed with the original skeletal rotations. Let $\mathcal{R}(\cdot)$ denote the mapping from an axis-angle vector to its corresponding rotation matrix in $\mathrm{SO}(3)$. The adapted joint rotations satisfy

$$\mathcal{R}\left(\boldsymbol{\theta}_{t,j}^{\mathrm{rig}}\right) = \mathcal{R}(\Delta\boldsymbol{\theta}_{t,j})\,\mathcal{R}(\boldsymbol{\theta}_{t,j}), \qquad j = 1,\ldots,J, \tag{3}$$

where $\boldsymbol{\Theta}_t^{\mathrm{rig}} = \{\boldsymbol{\theta}_{t,j}^{\mathrm{rig}}\}_{j=1}^{J}$ denotes the adapted skeletal pose. We fix the root-joint correction to zero, preserving the original root motion and restricting the optimization to relative skeletal articulation. The adapted mesh geometry is obtained through the LBS model: $\boldsymbol{V}_t^{\mathrm{rig}} = \mathcal{D}_{\mathrm{LBS}}(\boldsymbol{V}_{\mathrm{rest}}, \boldsymbol{\Theta}_t^{\mathrm{rig}}; \boldsymbol{S}, \boldsymbol{W}_{\mathrm{skin}})$. With this design, the shared controls $\boldsymbol{C}$ provide temporal coordination of pose corrections while preserving the rig-space validity of the resulting geometry.

Surface-Aware Alignment with Proxy Guidance. At each iteration, the intermediate proxy meshes provide geometric guidance for updating the skeletal motion. We then project this guidance back into the space of temporally continuous skeletal motions by searching for a rig-space sequence that best matches the proxy geometry. A straightforward solution is to minimize the vertex-wise ℓ_2 distance between the rigged meshes and their corresponding proxies. However, such point-to-point alignment is overly restrictive. The proxy meshes are obtained via local geometric repair. They are not necessarily exactly realizable by the underlying skeletal motion, resulting in an inherent gap between the mesh space and the rig space. Moreover, the frame-wise repair directions can vary over time, so that enforcing exact correspondence may conflict with the temporal regularity imposed by continuous skeletal motion. Importantly, the proxy mesh should be regarded as a feasible geometric guidance rather than a unique target configuration. We therefore seek an alignment objective that preserves the essential collision-resolving tendency of the proxy while allowing sufficient flexibility to accommodate the kinematic and temporal constraints of the rig. To this end, we adopt the symmetric point-to-plane residual (Rusinkiewicz, 2019), which has been successfully used for both rigid and non-rigid registration algorithms (Rusinkiewicz, 2019; Yao et al., 2025). Let $(\boldsymbol{v}_{t,i}^{\mathrm{rig}}, \boldsymbol{n}_{t,i}^{\mathrm{rig}})$ and $(\tilde{\boldsymbol{v}}_{t,i}^{(r)}, \tilde{\boldsymbol{n}}_{t,i}^{(r)})$ denote the corresponding vertices and unit normals on $\boldsymbol{V}_t^{\mathrm{rig}}$ and $\tilde{\boldsymbol{V}}_t^{(r)}$, respectively. We define the surface-aware guidance objective as

$$\mathcal{L}_{\mathrm{sag}}^{(r)} = \frac{1}{TN_v}\sum_{t=1}^{T}\sum_{i=1}^{N_v}\left[\left(\boldsymbol{n}_{t,i}^{\mathrm{rig}} + \tilde{\boldsymbol{n}}_{t,i}^{(r)}\right)\cdot\left(\boldsymbol{v}_{t,i}^{\mathrm{rig}} - \tilde{\boldsymbol{v}}_{t,i}^{(r)}\right)\right]^2. \tag{4}$$

Compared with vertex-wise point-to-point alignment, this formulation mainly constrains displacement along the locally averaged normal direction while allowing tangential deviation. This property

Table 1: Quantitative comparison with existing mesh self-intersection repair methods. We leave SRS unreported for shape-space methods, as they do not guarantee valid skeletal motion. The best and second-best results are **bolded** and underlined, respectively.

Methods	Optimization Space	Repair Quality		Pose Preservation		Motion Fidelity	
		PVR (%)↑	IFR (%)↑	SRS (%)↑	VD↓ ($\times 10^2$)	VAD↓ ($\times 10^3$)	VAM↓ ($\times 10^3$)
ISIR (Jang et al., 2025)	Mesh	94.24	89.85	–	6.692	4.727	5.620
Mesh-Utg (Minarčík et al., 2026)	Mesh	93.54	88.41	–	5.673	4.916	5.871
ISIR + skeleton fitting	Rig	93.27	91.72	73.91	6.515	4.539	5.439
Mesh-Utg + skeleton fitting	Rig	89.63	**92.42**	64.75	11.408	24.816	25.324
PoseShield (Li et al., 2026)	Rig	92.03	87.11	88.61	5.837	3.071	2.646
Ours	Rig	**94.64**	89.73	**93.46**	**5.630**	**0.607**	**2.202**

is particularly suitable for collision repair, where intersecting surfaces are typically separated primarily along their normal directions. Preserving this normal-direction correction retains the proxy's essential geometric tendency, whereas allowing tangential sliding provides additional freedom for the rig-space optimization to satisfy temporal consistency and kinematic constraints. In this way, the proxy mesh serves as a soft geometric guidance rather than an exact frame-wise target.

On the other hand, when $\left(\boldsymbol{n}_{t,i}^{\mathrm{rig}} + \tilde{\boldsymbol{n}}_{t,i}^{(r)}\right) \neq \mathbf{0}$, the distance term can be rewritten as

$$\left[\left(\boldsymbol{n}_{t,i}^{\mathrm{rig}} + \tilde{\boldsymbol{n}}_{t,i}^{(r)}\right) \cdot \left(\boldsymbol{v}_{t,i}^{\mathrm{rig}} - \tilde{\boldsymbol{v}}_{t,i}^{(r)}\right)\right]^2 = s_{t,i}^2 \cdot \left[\overline{\boldsymbol{n}}_{t,i} \cdot \left(\boldsymbol{v}_{t,i}^{\mathrm{rig}} - \tilde{\boldsymbol{v}}_{t,i}^{(r)}\right)\right]^2 \tag{5}$$

where

$$\overline{\boldsymbol{n}}_{t,i} = \frac{1}{s_{t,i}}\left(\boldsymbol{n}_{t,i}^{\mathrm{rig}} + \tilde{\boldsymbol{n}}_{t,i}^{(r)}\right) \text{ and } s_{t,i} = \left\|\boldsymbol{n}_{t,i}^{\mathrm{rig}} + \tilde{\boldsymbol{n}}_{t,i}^{(r)}\right\|_2. \tag{6}$$

The factor $s_{t,i}^2$ reflects the local orientation change induced by the proxy correction. Near-aligned normals indicate that the correction changes the surface position while largely preserving its local orientation, providing a more stable separation cue. Larger discrepancies imply stronger local reorientation, which may arise from unresolved or ambiguous repair and are correspondingly attenuated. This adaptive weighting behavior naturally matches the geometric characteristics of local collision repair: intersecting surface patches are typically separated mainly along their normal directions, while their local surface orientations remain largely unchanged after repair.

Since $\mathcal{L}_{\mathrm{sag}}^{(r)}$ does not explicitly constrain pose deviation from the input motion, we further regularize the root-relative joint positions to prevent excessive articulation drift. Let $\boldsymbol{p}_{t,j}$ and $\boldsymbol{p}_{t,j}^{\mathrm{rig},(r)}$ denote the original and adapted global positions of joint j, respectively, and o the root joint. We define

$$\mathcal{L}_{\mathrm{pose}}^{(r)} = \frac{1}{T(J-1)} \sum_{t=1}^{T} \sum_{j \neq o} \left\| \left(\boldsymbol{p}_{t,j}^{\mathrm{rig},(r)} - \boldsymbol{p}_{t,0}^{\mathrm{rig},(r)}\right) - \left(\boldsymbol{p}_{t,j} - \boldsymbol{p}_{t,o}\right) \right\|_2^2. \tag{7}$$

Finally, the objective for proxy-guided pose adaptation at marching step r is formulated as

$$\mathcal{L}_{\mathrm{rig}}^{(r)} = \lambda_{\mathrm{sag}} \mathcal{L}_{\mathrm{sag}}^{(r)} + \lambda_{\mathrm{pose}} \mathcal{L}_{\mathrm{pose}}^{(r)}, \tag{8}$$

where λ_{sag} and λ_{pose} control the strength of proxy alignment and pose preservation, respectively.

4 Experiments

4.1 Experiment Settings

Dataset. We use the motion subset of PoseShield (Li et al., 2026) as our primary benchmark, comprising 100 SMPL-H (Romero et al., 2017) sequences with diverse motion-induced self-intersections. To assess applicability beyond SMPL-based humans, we further evaluate on non-SMPL rigged animations from Objaverse (Deitke et al., 2023b;a) and Truebone Zoo (Truebones Motions Animation Studios, 2022), covering humanoid cartoon characters and animals with varied geometries and skeletal structures.

Baselines. We compare with ISIR (Jang et al., 2025), Mesh-Utg (Minarčík et al., 2026), and PoseShield (Li et al., 2026). ISIR and Mesh-Utg repair each frame directly in vertex space and therefore do not provide corresponding skeletal motion. To enable comparison in the rigged-animation setting, we additionally construct skeleton-fitted variants by optimizing the original SMPL-H rig to repro-

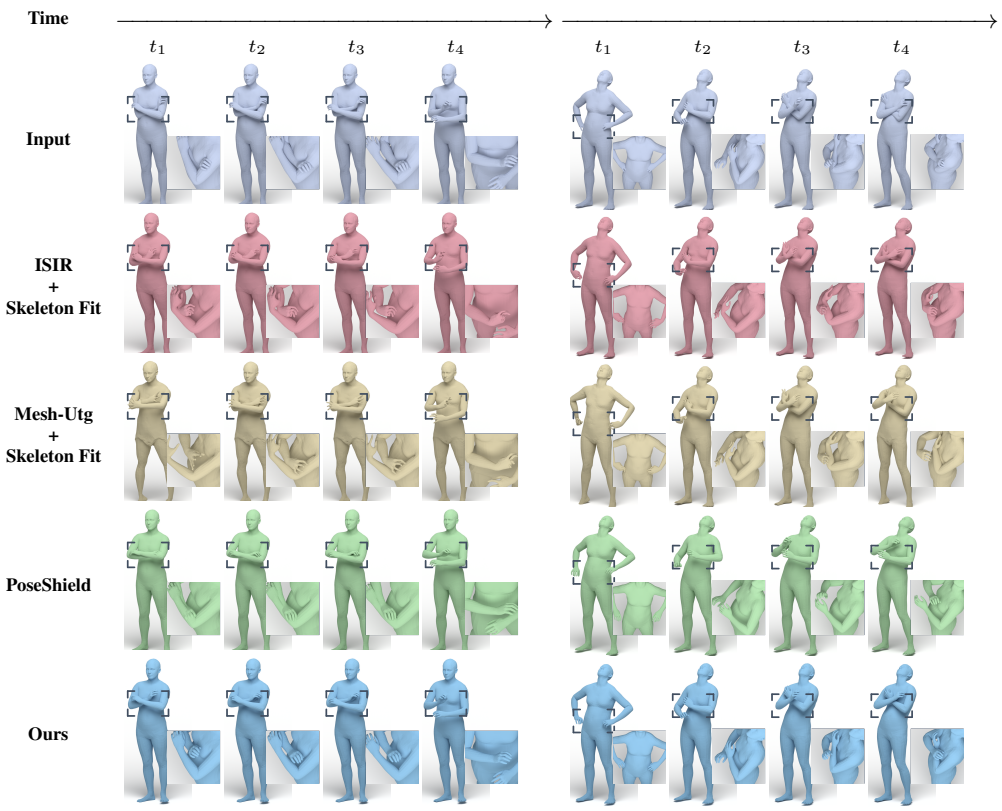


Figure 3: Qualitative comparison on the PoseShield benchmark. Each row shows representative frames from two sample sequences, with self-intersection regions additionally visualized from an alternative viewpoint for assessing repair quality and pose preservation. Our method achieves more complete self-intersection repair while better preserving the original motion characteristics.

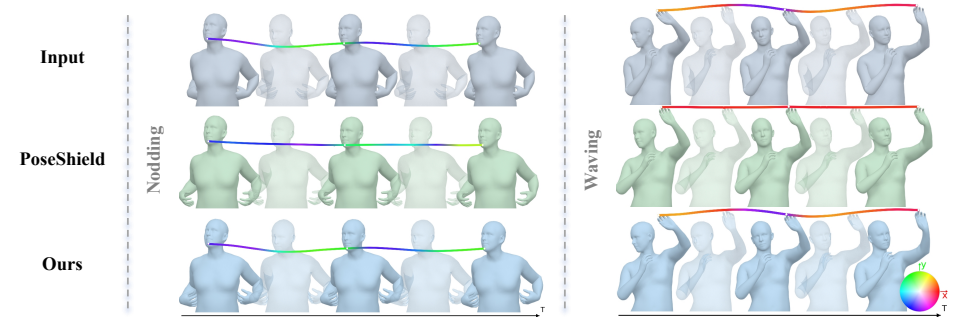


Figure 4: Qualitative comparison of motion preservation under a shared viewpoint. Trajectory hue encodes the instantaneous motion direction according to the HSV reference, while its excursion reflects the motion magnitude. Please zoom in for a clearer view.

duce their repaired geometry. The implementation details are provided in the Appendix. PoseShield and our method directly optimize skeletal motion in pose space.

Metrics. We evaluate the performance from three aspects: repair quality, pose preservation, and motion fidelity. For repair quality, we report *Intersecting Face Resolution* (IFR) (Jang et al., 2025) and *Penetration Volume Reduction* (PVR). IFR measures the reduction in intersecting faces, whereas PVR accounts for penetration severity through the reduction in inter-part penetration volume, distinguishing shallow intersections from deep penetrations that may involve similar numbers of intersecting faces. Pose preservation is measured by *Skeleton Rotation Similarity* (SRS) and *Vertex Discrepancy* (VD). Motion fidelity is assessed by *Vertex Acceleration Difference* (VAD), complemented by *Vertex Acceleration Magnitude* (VAM) as a measure of temporal smoothness. Detailed definitions are provided in the Appendix.



Figure 5: Qualitative comparison on Objaverse. Textured and normal-direction renderings are shown side by side, with all methods evaluated from identical viewpoints. Red boxes mark input self-intersections. Please zoom in for a clearer comparison.

Table 2: Ablation of the main components. PMRO and TRP denote Progressive Mesh-Rig Optimization and Temporal Rig Parameterization, respectively.

Variants	Sub-components				Repair Quality		Pose Preservation		Motion Fidelity	
	PMRO	TRP	$\mathcal{L}_{\text{sag}}$	$\mathcal{L}_{\text{pose}}$	PVR (%)↑	IFR (%)↑	SRS (%)↑	VD↓ ($\times 10^2$)	VAD↓ ($\times 10^3$)	VAM↓ ($\times 10^3$)
Vanilla					**96.10**	**95.22**	62.60	23.063	121.605	121.676
V1	✓				95.52	91.23	**94.81**	6.660	6.632	7.420
V2	✓	✓			95.52	91.64	94.47	6.699	0.663	2.227
V3	✓	✓	✓		95.72	91.69	93.19	7.910	0.789	2.247
V4	✓	✓		✓	90.83	86.25	93.12	**3.579**	**0.434**	**2.144**
V5	✓		✓	✓	94.57	89.59	94.20	5.577	6.300	5.444
Full	✓	✓	✓	✓	94.64	89.73	93.46	5.630	0.607	2.202

4.2 Results and Comparisons

Qualitative Analysis. We first evaluate the methods on the SMPL-H models from the PoseShield benchmark, which are watertight, have complete surface connectivity, and consistent mesh-skeleton topologies. Proxyboros achieves more thorough self-intersection resolution while better preserving the input skeletal motion. As shown in Fig. 3, all baselines exhibit visible intersections in challenging contact regions, whereas Proxyboros consistently separates the colliding surfaces with only minor pose changes. In addition, PoseShield may alter the intended motion pattern after resolving self-collision body parts, *e.g.*, reversing the original front-to-back order of the crossed arms in the right example. Fig. 4 further highlights its limitation in subtle-motion preservation. PoseShield's over-smoothness suppresses small but characteristic temporal variations, hindering both motion fluency and local directional changes. In contrast, Proxyboros better retains these fine-scale dynamics, producing trajectories that faithfully reflect the local motion pattern. The performance of Proxyboros remains stable on non-SMPL-rigged models with diverse topologies and skeletal structures, including open boundaries, independently assembled components, and pre-existing interpenetrations between disconnected parts, as shown in Fig. 5. More results are shown on our project page.

Quantitative Analysis. As shown in Table 1, our method achieves the best PVR and the strongest pose and motion preservation, while remaining competitive in IFR. Mesh-Utg improves face-level repair after skeleton fitting but suffers clear degradation in PVR and motion fidelity, suggesting that its static corrections are not readily realizable within the original rig space. Our method avoids this mismatch and better preserves the input skeletal motion during repair.

Ablation Study. Table 2 and Fig. 6 ablate the main components, with vertex-wise ℓ_2 alignment replacing $\mathcal{L}_{\text{sag}}$ when it is disabled. Introducing proxy-guided mesh-rig optimization substantially reduces the pose and motion distortion caused by direct rig-space repair under kinematic coupling (Vanilla vs. V1). The progressive design itself is further examined against an equal-budget one-shot variant in the Appendix. Temporal rig parameterization improves motion fidelity, producing coherent trajectories without sacrificing repair performance (V1 vs. V2, V5 vs. Full). Surface-aware guidance is critical for resolving severe self-intersections that remain under vertex-wise ℓ_2 alignment (V4 vs. Full), as highlighted in Fig. 6. Root-relative joint regularization further improves pose preservation (V3 vs. Full).

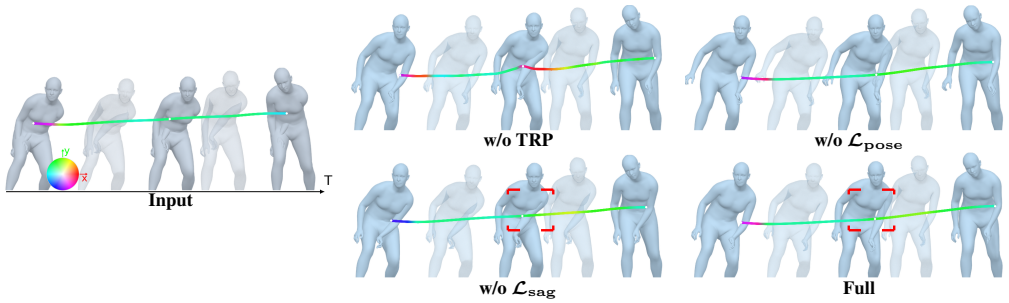


Figure 6: Qualitative ablation of the proposed components. Colored trajectories trace the elbow's movement over time, with red corner boxes highlighting representative self-intersection regions.

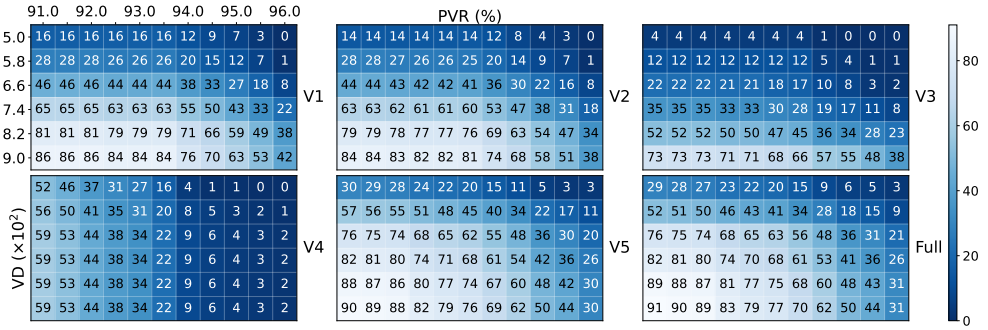


Figure 7: Ablation comparison under joint PVR and VD thresholds. Each cell reports the percentage of samples with PVR above and VD below the corresponding thresholds, with lighter colors indicating a higher proportion of samples that satisfy both criteria.

Moreover, we also explore the trade-off between repair quality and pose preservation in Fig. 7 using joint PVR-VD thresholds. As the PVR requirement becomes more stringent, the full model retains more samples at low VD thresholds, indicating a better balance among well-repaired cases and avoiding the bias of VD-only comparisons, where insufficient repair can trivially yield small deformation. Besides, the reduction of PVR in Table 2 reflects the limited expressiveness of LBS-constrained deformation. Severe folds and contacts in the mesh may require local non-rigid changes that skeletal transformations cannot represent, while relaxing rig constraints can remove penetration through excessive articulation. Within this realizable rig space, our full model achieves a favorable balance between repair effectiveness and motion preservation.

User Study. We conduct a user study on 30 repair cases: 15 SMPL-H sequences from PoseShield, 10 Objaverse humanoids, and 5 Truebone Zoo animals. For each case, we construct pairwise questions by comparing our result with each baseline (75 questions in total), while randomizing the question and display orders within each pair. Participants select the better result for repair quality (RQ) and motion similarity (MS), separately. We received 55 valid responses in total and report our pairwise win rate and tie rate in Table 3. More details are provided in the Appendix.

Table 3: User study on PoseShield and non-SMPL models, reported as *Win rate (tie rate)*, with ties accounting for "both equally good" cases.

Methods		SMPL-H		Non-SMPL	
		RQ (%)↑	**MS (%)↑**	**RQ (%)↑**	**MS (%)↑**
Ours *vs.*	ISIR	78.8 (20.0)	79.4 (19.4)	81.8 (14.5)	84.5 (12.7)
	Mesh-Utg	98.2 (1.2)	100 (0.0)	96.4 (2.7)	100 (0.0)
	PoseShield	43.6 (49.7)	68.5 (27.9)	–	–

5 Conclusion

We proposed *Proxyboros*, a dynamic self-intersection repair framework for rigged mesh animations that preserves the underlying skeletal motion. Proxyboros uses intermediate mesh proxies to progressively coordinate local geometric repair with rig-space pose adaptation, allowing repair directions discovered in mesh space to be absorbed into the evolving skeletal motion. A temporally coupled pose parameterization and surface-aware proxy guidance further regulate this process, balancing effective intersection removal with pose and motion preservation. Experiments on SMPL-H sequences and diverse non-SMPL rigs demonstrate strong repair performance and consistent motion fidelity across varied geometries and skeletal structures.

AI USE STATEMENT

In this work, we used generative AI tools to assist with translation, dataset cleaning, and reformatting. We did not use generative AI tools to develop theoretical models or conceptual frameworks, formulate mathematical claims, design or provide feedback on the research methodology or experiments, implement the proposed methods, support qualitative or thematic data analysis, or interpret experimental results. Other required-disclosure tasks, including generating synthetic datasets, proposing or refining hypotheses, providing critical ingredients for mathematical proofs, and assisting in the writing of proofs, were not applicable to this work.

Additionally, we used generative AI tools for software code creation and editing, information sourcing and searching, improving the manuscript's readability, and identifying potentially relevant literature. AI-assisted code was reviewed and tested by the authors for correctness and reliability. All AI-assisted manuscript edits were reviewed and revised by the authors as appropriate. Information and literature identified through AI-assisted searches were independently verified against the original and authoritative sources before use.

We have reviewed all AI-assisted work and take full responsibility for the final content of this work, including any text, claims, code, or other artifacts produced with the aid of generative AI.

REFERENCES

Robert Bridson, Ronald Fedkiw, and John Anderson. Robust treatment of collisions, contact and friction for cloth animation. *ACM Trans. Graph.*, 21(3):594–603, 2002.

Steve Capell, Seth Green, Brian Curless, Tom Duchamp, and Zoran Popović. Interactive skeleton-driven dynamic deformations. *ACM Trans. Graph.*, 21(3):586–593, 2002.

Carl De Boor. On calculating with b-splines. *J. Approx. Theory*, 6(1):50–62, 1972.

Matt Deitke, Ruoshi Liu, Matthew Wallingford, Huong Ngo, Oscar Michel, Aditya Kusupati, Alan Fan, Christian Laforte, Vikram Voleti, Samir Yitzhak Gadre, et al. Objaverse-xl: A universe of 10m+ 3d objects. In *Adv. Neural Inform. Process. Syst.*, pp. 35799–35813, 2023a.

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 *IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 13142–13153, 2023b.

Tobias Djuren, Ugo Finnendahl, Markus Worchel, Hendrik Meyer, and Marc Alexa. Differentiable voxelization of surface representations. In *ACM SIGGRAPH Conf.*, pp. 1–12, 2026.

David Harmon, Etienne Vouga, Breannan Smith, Rasmus Tamstorf, and Eitan Grinspun. Asynchronous contact mechanics. In *ACM SIGGRAPH Conf.*, pp. 1–12. 2009.

Pascal Herrmann, Maarten Bieshaar, Dennis Mack, Paul Robert Herzog, and Juergen Gall. Self-intersection-aware 3d human motion generation using an efficient human sphere proxy. In *Brit. Mach. Vis. Conf.*, volume 36, Sheffield, UK, 2025.

Takeo Igarashi, Tomer Moscovich, and John F Hughes. As-rigid-as-possible shape manipulation. *ACM Trans. Graph.*, 24(3):1134–1141, 2005.

Wonjong Jang, Yucheol Jung, Gyeongmin Lee, and Seungyong Lee. Instant self-intersection repair for 3d meshes. *ACM Trans. Graph.*, 44(4):1–14, 2025.

Tero Karras. Maximizing parallelism in the construction of bvhs, octrees, and k-d trees. In *Proc. ACM SIGGRAPH/Eurographics Conf. High-Perform. Graph.*, pp. 33–37, 2012.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Int. Conf. Learn. Represent.*, pp. 1–15, 2014.

Minchen Li, Zachary Ferguson, Teseo Schneider, Timothy Langlois, Denis Zorin, Daniele Panozzo, Chenfanfu Jiang, and Danny M Kaufman. Incremental potential contact: intersection-and inversion-free, large-deformation dynamics. *ACM Trans. Graph.*, 39(4):1–49, 2020.

Minchen Li, Danny M Kaufman, and Chenfanfu Jiang. Codimensional incremental potential contact. *ACM Trans. Graph.*, 40(4):1–24, 2021.

Yijing Li and Jernej Barbič. Immersion of self-intersecting solids and surfaces. *ACM Trans. Graph.*, 37(4):1–14, 2018.

Zhengyuan Li, Zeyun Deng, Yifan Shen, Liangyan Gui, Miaolan Xie, Joseph Campbell, Xifeng Gao, Kui Wu, Zherong Pan, and Aniket Bera. Poseshield: Neural collision fields for human self-collision resolution. In *Eur. Conf. Comput. Vis.*, 2026.

Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. Graph.*, 34(6):248:1–248:16, 2015.

Nadia Magnenat-Thalmann, Richard Laperrière, and Daniel Thalmann. Joint-dependent local deformations for hand animation and object grasping. In *Proc. Graph. Interface*, pp. 26–33, 1989.

Jiří Minarčík, Michael Liu, Keenan Crane, and Minchen Li. Untangling surfaces via shape and mesh repulsion. *ACM Trans. Graph.*, 45(4):1–15, 2026.

Baptiste Nicolet, Alec Jacobson, and Wenzel Jakob. Large steps in inverse rendering of geometry. *ACM Trans. Graph.*, 40(6):1–13, 2021.

Soon-Yong Park and Murali Subbarao. An accurate and fast point-to-plane registration technique. *Pattern Recognit. Lett.*, 24(16):2967–2976, 2003.

Javier Romero, Dimitrios Tzionas, and Michael J. Black. Embodied hands: Modeling and capturing hands and bodies together. *ACM Trans. Graph.*, 36(6), 2017.

Szymon Rusinkiewicz. A symmetric objective function for icp. *ACM Trans. Graph.*, 38(4), July 2019. ISSN 0730-0301. doi: 10.1145/3306346.3323037. URL https://doi.org/10.1145/3306346.3323037.

Leonardo Sacht, Alec Jacobson, Daniele Panozzo, Christian Schüller, and Olga Sorkine-Hornung. Consistent volumetric discretizations inside self-intersecting surfaces. In *Comput. Graph. Forum*, volume 32, pp. 147–156. Wiley Online Library, 2013.

Josua Sassen, Henrik Schumacher, Martin Rumpf, and Keenan Crane. Repulsive shells. *ACM Trans. Graph.*, 43(4):1–22, 2024.

Nicholas Sharp and Keenan Crane. A Laplacian for Nonmanifold Triangle Meshes. *Comput. Graph. Forum*, 39(5), 2020.

Qingyang Tan, Zherong Pan, and Dinesh Manocha. Lcollision: Fast generation of collision-free human poses using learned non-penetration constraints. In *Proc. AAAI Conf. Artif. Intell.*, volume 35, pp. 3913–3921, 2021.

Qingyang Tan, Zherong Pan, Breannan Smith, Takaaki Shiratori, and Dinesh Manocha. N-penetrate: Active learning of neural collision handler for complex 3d mesh deformations. In *Int. Conf. Mach. Learn.*, pp. 21037–21049, 2022.

Truebones Motions Animation Studios. Truebones, 2022. URL https://truebones.gumroad.com/. Accessed: 2022-1-15.

Dimitrios Tzionas, Luca Ballan, Abhilash Srikantha, Pablo Aponte, Marc Pollefeys, and Juergen Gall. Capturing hands in action using discriminative salient points and physics simulation. *Int. J. Comput. Vis.*, 118(2):172–193, 2016.

Pascal Volino and Nadia Magnenat-Thalmann. Resolving surface collisions through intersection contour minimization. *ACM Trans. Graph.*, 25(3):1154–1159, 2006.

Yuxin Yao, Bailin Deng, Junhui Hou, and Juyong Zhang. Spare: Symmetrized point-to-plane distance for robust non-rigid 3d registration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2025.

Juntao Ye, Guanghui Ma, Liguo Jiang, Lan Chen, Jituo Li, Gang Xiong, Xiaopeng Zhang, and Min Tang. A unified cloth untangling framework through discrete collision detection. In *Comput. Graph. Forum*, volume 36, pp. 217–228. Wiley Online Library, 2017.

Chris Yu, Caleb Brakensiek, Henrik Schumacher, and Keenan Crane. Repulsive surfaces. *ACM Trans. Graph.*, 40(6):1–19, 2021.

A Appendix

A.1 Implementation Details

Implementation of Proxyboros. We run all experiments on a single NVIDIA RTX 5880 Ada GPU and optimize Proxyboros for 100 rounds, each comprising 5 mesh-space updates followed by 5 rig-space updates. Proxy exploration uses gradient descent with a learning rate of 10^{-3}, with $p=1$, $\alpha=0.99$, and $\lambda_{\text{curv}}=10^6$. For pose adaptation, we use a control interval of 8 frames and set $\lambda_{\text{sag}}=1.0$ and $\lambda_{\text{pose}}=1\times10^{-2}$. The B-spline controls are optimized using Adam (Kingma & Ba, 2014), with a learning rate of 10^{-2} for the first 20 rounds followed by cosine decay to 10^{-3}.

Repair for disconnected non-SMPL meshes. Compared with the complete and consistently connected SMPL-H body model, Objaverse and Truebone Zoo characters often consist of multiple disconnected components, in which geometrically valid assembly contacts can already produce triangle intersections in the rest pose. Directly applying intersection repair to such meshes can therefore incorrectly separate these components or cause unstable deformation around their junctions. We employ a simple heuristic adaptation based on two observations. First, cross-component intersections already present in the rest pose typically reflect persistent assembly contacts rather than motion-induced self-intersections and should therefore be preserved during repair. Second, although these components are geometrically coupled at such contacts, their disconnected topology prevents differential deformation from propagating across the junctions. Our adaptation accordingly preserves the former while compensating for the latter.

Specifically, we identify cross-component triangle pairs that intersect at the rest pose and exclude these pairs, along with their one-ring neighborhoods, from subsequent self-intersection detection. We further construct virtual cross-component connections around these junctions using a robust point-cloud Laplacian (Sharp & Crane, 2020), which is combined with the original cotangent surface Laplacian to propagate local deformation consistently across otherwise disconnected components. This adaptation applies to both Proxyboros and ISIR, as both methods perform local repair using the same collision-pair formulation and Large Steps parameterization. Mesh-Utg instead relies on its own forward/reverse-flow untangling formulation, which does not expose the corresponding collision-pair filtering or Laplacian-based deformation operator. Incorporating the same mechanism would therefore require modifying its core optimization procedure; we retain its original implementation and apply it independently to each frame.

Skeleton Fitting for Shape-space Baselines. ISIR and Mesh-Utg operate directly in shape space and therefore do not provide skeletal motion after repair. In Sec. 4, for evaluation involving skeletal poses, we first apply each method independently to every input frame and then fit the original character rig to the resulting repaired meshes.

Specifically, we retain the original rest-pose geometry, skeleton hierarchy, and skinning weights, and optimize a local axis-angle rotation for each non-root joint and each frame. The fitted vertices are generated using linear blend skinning (LBS), with SMPL-H pose-corrective blend shapes enabled for the SMPL-H sequences. The fitting objective combines vertex-wise correspondence to the repaired mesh, an As-Rigid-As-Possible (ASAP) (Igarashi et al., 2005) term for preserving its local geometry, and a weak ℓ_2 regularization on the optimized joint rotations. The root transformation is fixed according to the corresponding input frame to preserve its global alignment. All frames are fitted independently, without temporal regularization, ensuring that the fitting procedure does not introduce additional temporal priors that are unavailable to the original shape-space methods.

A.2 Evaluation Metrics

We provide detailed definitions of the evaluation metrics used in our experiments. Let $\boldsymbol{V}_{1:T} = \{\boldsymbol{V}_t\}_{t=1}^{T}$ and $\hat{\boldsymbol{V}}_{1:T} = \{\hat{\boldsymbol{V}}_t\}_{t=1}^{T}$ denote the input and repaired mesh sequences, respectively, which share the same mesh connectivity. For metrics related to pose preservation and motion fidelity, global root motion is removed from both sequences before evaluation, such that the measurements reflect changes in relative skeletal articulation and surface motion rather than global character displacement. Unless otherwise specified, each metric is computed per sequence and then averaged over all evaluated sequences.

A.2.1 Self-Intersection Repair Quality

Intersecting Face Resolution (IFR). Following Jang et al. (2025), we detect self-intersections at each frame using BVH-based triangle collision detection (Karras, 2012; Tzionas et al., 2016). Let $\mathcal{C}_t$ and $\hat{\mathcal{C}}_t$ denote the sets of faces involved in at least one self-intersection in the input and repaired meshes at frame t, respectively. For meshes with disconnected components, cross-component face pairs that already intersect in the rest pose are treated as structural assembly contacts and excluded from evaluation. We define the resolved and newly introduced intersecting faces as

$$\mathcal{R}_t = \mathcal{C}_t \setminus \hat{\mathcal{C}}_t, \qquad \mathcal{A}_t = \hat{\mathcal{C}}_t \setminus \mathcal{C}_t, \tag{9}$$

where $\mathcal{R}_t$ contains initially intersecting faces resolved after repair, and $\mathcal{A}_t$ contains newly intersecting faces. The IFR is defined as

$$\mathrm{IFR} = \frac{\sum_{t=1}^{T} (|\mathcal{R}_t| - |\mathcal{A}_t|)}{\sum_{t=1}^{T} |\mathcal{C}_t|}. \tag{10}$$

A larger IFR indicates a greater net reduction in the number of intersecting faces.

Penetration Volume Reduction (PVR). IFR measures self-intersection at the surface level, but the number of intersecting faces does not necessarily reflect the geometric severity of penetration. For example, shallow surface contacts and deep inter-part penetrations may involve similar numbers of intersecting faces. We therefore introduce *Penetration Volume Reduction*, which measures the reduction in volumetric overlap between semantic body parts.

Semantic part construction. To distinguish meaningful inter-part penetration from the internal structure of an articulated mesh, we partition the rest-pose surface into semantic body parts according to its skinning weights. Let $\boldsymbol{W}_{\mathrm{skin}}(i, j)$ denote the skinning weight between vertex i and joint j. We first organize the skeleton joints into semantic groups $\{\mathcal{G}_g\}_{g=1}^{N_p}$ and define the weight of vertex i with respect to semantic part g as

$$\boldsymbol{w}_{i,g} = \sum_{j \in \mathcal{G}_g} \boldsymbol{W}_{\mathrm{skin}}(i, j). \tag{11}$$

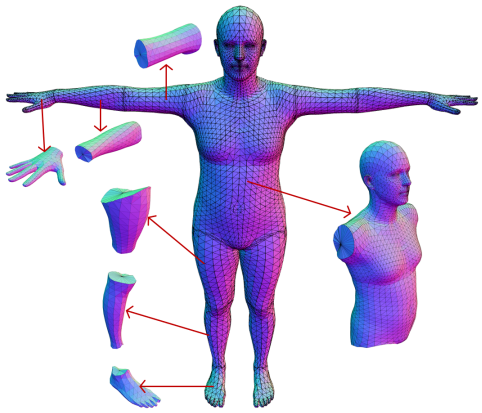


Figure 8: Semantic body-part decomposition for penetration-volume evaluation. The SMPL-H mesh is partitioned into 13 parts from grouped skinning weights, with the torso, neck, and head merged into a single central region and articulated limb segments retained separately.

For SMPL-H, the pelvis, spine, neck, collar, and head joints are grouped into a single torso-head part. These regions form the central body of the mesh and typically do not produce meaningful inter-part penetrations among themselves during articulation. The wrist and finger joints are grouped into the left and right hand parts, while the ankle and foot joints form the left and right foot parts. The left and right hips, knees, shoulders, and elbows are retained as separate semantic parts. This results in 13 parts: { *torso-head, left/right hip, left/right knee, left/right foot, left/right shoulder, left/right elbow, left/right hand* }. The vertex-level part weights are linearly interpolated over each triangle, yielding continuous weight fields over the surface. Semantic boundaries are then defined by the equal-weight contours between neighboring parts, producing geometrically coherent cut surfaces that avoid discrete vertex-level partition artifacts. This continuous partition enables inter-part penetrations induced by arbitrary poses to be consistently localized and faithfully measured during subsequent volumetric evaluation. Accordingly, the surface region associated with part g is defined as

$$\mathcal{P}_g = \{x \in \mathcal{M}_{\mathrm{rest}} \mid w_g(x) \geq w_h(x),\ \forall h \neq g\}. \tag{12}$$

Triangles crossing semantic boundaries are clipped along the corresponding equal-weight contours, yielding a continuous surface partition. Each resulting part is closed by triangulating its boundary loops, enabling volumetric overlap to be measured between different semantic regions. The cut vertices are stored by their barycentric coordinates on the original surface, so that the same partition can be reconstructed consistently throughout the animation. Fig. 8 illustrates the resulting decomposition and several representative closed part meshes.

Pairwise penetration volume. Let $\mathcal{P}_{t,g}$ denote the closed geometry of semantic part g at frame t. For each pair of semantic parts (g, h) with $g < h$, we first compute the intersection of their axis-aligned bounding boxes and sample a regular voxel grid within the overlapping region. The penetration volume between the two parts is approximated as

$$\mathcal{V}_t^{g,h} = s_{t,g,h}^3 \sum_{x \in \mathcal{X}_t^{g,h}} 1\left[x \in \mathcal{P}_{t,g}\right] 1\left[x \in \mathcal{P}_{t,h}\right], \tag{13}$$

where $\mathcal{X}_t^{g,h}$ is the set of sampled voxel centers and $s_{t,g,h}$ is the corresponding voxel size. We set the default voxel size according to the diagonal length d_{bbox} of the rest-pose bounding box:

$$s_0 = \frac{d_{\mathrm{bbox}}}{160}. \tag{14}$$

For large overlapping regions, the voxel size is adaptively increased to limit the number of sampled points. The total penetration volume of a sequence is then calculated by

$$\mathcal{V}(\boldsymbol{V}_{1:T}) = \sum_{t=1}^{T} \sum_{g<h} \mathcal{V}_t^{g,h}. \tag{15}$$

Finally, PVR is defined as

$$\mathrm{PVR} = 1 - \frac{\mathcal{V}(\hat{\boldsymbol{V}}_{1:T})}{\mathcal{V}(\boldsymbol{V}_{1:T})}. \tag{16}$$

A larger PVR indicates a greater reduction in penetration volume, with $\mathrm{PVR} = 1$ corresponding to complete removal of the measured inter-part overlap. Sequences with zero initial penetration volume are excluded from this metric.

A.2.2 Pose Preservation

Skeleton Rotation Similarity (SRS). We measure skeletal pose preservation using the geodesic difference between the global joint rotations of the input and repaired sequences. Let $\boldsymbol{R}_{t,j}, \hat{\boldsymbol{R}}_{t,j} \in \mathrm{SO}(3)$ denote the corresponding global rotations of joint j at frame t. Their geodesic rotation difference is

$$\theta_{t,j} = \arccos\left(\frac{\mathrm{Tr}\left(\hat{\boldsymbol{R}}_{t,j}\boldsymbol{R}_{t,j}^{\top}\right) - 1}{2}\right). \tag{17}$$

We convert this angular difference into a normalized similarity and average it over all frames and joints:

$$\mathrm{SRS} = \frac{1}{TJ} \sum_{t=1}^{T} \sum_{j=1}^{J} \left(1 - \frac{\theta_{t,j}}{\pi}\right). \tag{18}$$

SRS lies in $[0, 1]$, with larger values indicating better preservation of the input skeletal pose.

Vertex Discrepancy (VD). We complement the skeletal measure with the mean vertex displacement between the input and repaired sequences:

$$\mathrm{VD} = \frac{1}{TN_v} \sum_{t=1}^{T} \sum_{i=1}^{N_v} \left\|\hat{\boldsymbol{v}}_{t,i} - \boldsymbol{v}_{t,i}\right\|_2. \tag{19}$$

Lower VD indicates a smaller geometric deviation from the input animation.

A.2.3 Motion Fidelity

Pose preservation alone does not capture whether the temporal characteristics of the input animation are retained after repair. Small frame-wise deviations may still introduce noticeable temporal jitter, while excessive smoothing can suppress meaningful motion variations. We therefore characterize motion fidelity using second-order differences of the root-motion-normalized vertex trajectories, which capture changes in local temporal dynamics. We report two complementary measures: VAD

quantifies how closely the repaired acceleration pattern follows the input motion, whereas VAM measures the absolute temporal variation of the repaired sequence. Together, they distinguish motion fidelity from smoothness and help identify both temporally inconsistent corrections and over-smoothed motion. For each interior frame $t = 2, \dots, T-1$, the discrete vertex accelerations of the input and repaired sequences are defined as

$$\boldsymbol{a}_{t,i} = \boldsymbol{v}_{t+1,i} - 2\boldsymbol{v}_{t,i} + \boldsymbol{v}_{t-1,i}, \qquad \hat{\boldsymbol{a}}_{t,i} = \hat{\boldsymbol{v}}_{t+1,i} - 2\hat{\boldsymbol{v}}_{t,i} + \hat{\boldsymbol{v}}_{t-1,i}. \tag{20}$$

Vertex Acceleration Difference (VAD). VAD measures the deviation of the repaired motion from the temporal dynamics of the input sequence:

$$\mathrm{VAD} = \frac{1}{(T-2)N_v} \sum_{t=2}^{T-1} \sum_{i=1}^{N_v} \left\| \hat{\boldsymbol{a}}_{t,i} - \boldsymbol{a}_{t,i} \right\|_2 . \tag{21}$$

Lower VAD indicates better preservation of the input motion dynamics.

Vertex Acceleration Magnitude (VAM). VAM measures the absolute second-order temporal variation of the repaired sequence:

$$\mathrm{VAM} = \frac{1}{(T-2)N_v} \sum_{t=2}^{T-1} \sum_{i=1}^{N_v} \left\| \hat{\boldsymbol{a}}_{t,i} \right\|_2 . \tag{22}$$

Lower VAM indicates smoother motion with smaller frame-to-frame acceleration variation. Together, VAD and VAM characterize complementary aspects of temporal behavior: VAD measures fidelity to the input motion, whereas VAM measures the temporal smoothness of the repaired sequence.

A.3 Details of Mesh-Space Proxy Exploration

We instantiate the mesh-space proxy exploration with the static self-intersection repair formulation of ISIR (Jang et al., 2025), in which the signed TPE provides directional gradients for separating intersecting surfaces, while curvature preservation limits unintended local distortion. The mesh is optimized under the Large-Steps differential parameterization (Nicolet et al., 2021). Below, we detail how this formulation is incorporated into each proxy-exploration step of Proxyboros.

Signed TPE. For a vertex configuration $\boldsymbol{X}$ with connectivity $\boldsymbol{F}$, let $\mathcal{I}(\boldsymbol{X}, \boldsymbol{F})$ denote the set of pairs of intersecting triangles. For a triangle f_s, we denote its vertices, centroid, unit normal, and area by $\{\boldsymbol{v}_s^k\}_{k=1}^3$, $\boldsymbol{c}_s$, $\boldsymbol{n}_s$, and A_s, respectively. For each $(f_i, f_j) \in \mathcal{I}(\boldsymbol{X}, \boldsymbol{F})$, the bidirectional vertex-based signed TPE (Jang et al., 2025) is

$$\phi_{\mathrm{sTPE}}(f_i, f_j) = A_j \sum_{k=1}^{3} \frac{\langle \boldsymbol{c}_j - \boldsymbol{v}_i^k, \boldsymbol{n}_j \rangle^p}{\|\boldsymbol{c}_j - \boldsymbol{v}_i^k\|_2^{2p}} + A_i \sum_{k=1}^{3} \frac{\langle \boldsymbol{c}_i - \boldsymbol{v}_j^k, \boldsymbol{n}_i \rangle^p}{\|\boldsymbol{c}_i - \boldsymbol{v}_j^k\|_2^{2p}}, \tag{23}$$

where p is a positive odd number. The signed normal projections preserve the direction of the interaction across the opposing surface, yielding gradients that drive intersecting triangles toward separation.

Differential mesh optimization. For each frame t, we compute the cotangent Laplacian $\boldsymbol{L}_t$ from the original posed mesh $\boldsymbol{V}_t$ and construct the Large-Steps transformation

$$\boldsymbol{M}_t = (1-\alpha)\boldsymbol{I} + \alpha \boldsymbol{L}_t, \qquad \boldsymbol{H}_t = \boldsymbol{L}_t \boldsymbol{V}_t, \tag{24}$$

where α controls the spatial coupling of the differential parameterization and $\boldsymbol{H}_t$ serves as a fixed reference for the local differential structure of the input surface. The quantities $\boldsymbol{L}_t$, $\boldsymbol{M}_t$, and $\boldsymbol{H}_t$ are fixed throughout the recurrent optimization.

At round r, the exploration is reinitialized from the latest rigged geometry $\boldsymbol{V}_t^{(r)}$:

$$\boldsymbol{u}_t^{(r,0)} = \boldsymbol{M}_t \boldsymbol{V}_t^{(r)}, \qquad \widetilde{\boldsymbol{V}}_t^{(r,\ell)} = \boldsymbol{M}_t^{-1} \boldsymbol{u}_t^{(r,\ell)}, \tag{25}$$

Table 4: Quantitative ablation for progressive mesh-rig optimization.

Methods	Repair Quality		Pose Preservation		Motion Fidelity	
	PVR (%)↑	IFR (%)↑	SRS (%)↑	VD↓ ($\times 10^2$)	VAD↓ ($\times 10^3$)	VAM↓ ($\times 10^3$)
One-shot PMRO	**95.17**	**91.39**	89.73	6.268	0.863	2.311
Full	94.64	89.73	**93.46**	**5.630**	**0.607**	**2.202**

Table 5: Quantitative ablation for the role of $\mathcal{L}_{sag}$.

Methods	Repair Quality		Pose Preservation		Motion Fidelity	
	PVR (%)↑	IFR (%)↑	SRS (%)↑	VD↓ ($\times 10^2$)	VAD↓ ($\times 10^3$)	VAM↓ ($\times 10^3$)
Point2Point	90.83	86.25	93.12	3.579	**0.434**	**2.144**
Point2Plane	90.39	85.70	88.58	**3.546**	0.498	2.166
L_{sag} **w/o direction**	92.30	87.87	**94.15**	4.065	0.457	2.153
L_{sag} **w/o magnitude**	90.40	85.68	88.54	3.560	0.501	2.167
Full	**94.64**	**89.73**	93.46	5.630	0.607	2.202

where $\boldsymbol{u}_t^{(r,\ell)}$ denotes the differential variables at exploration step ℓ. We optimize

$$\mathcal{L}_{\text{proxy},t}^{(r,\ell)} = \sum_{(f_i,f_j)\in\mathcal{I}(\widetilde{\boldsymbol{V}}_t^{(r,\ell)},\boldsymbol{F})} \phi_{\text{sTPE}}(f_i,f_j) + \lambda_{\text{curv}} \left\| \boldsymbol{L}_t \widetilde{\boldsymbol{V}}_t^{(r,\ell)} - \boldsymbol{H}_t \right\|^2. \tag{26}$$

The signed TPE term drives intersecting regions toward separation, whereas the curvature term preserves the local differential structure of the original posed surface. Rather than running the static repair to convergence, we stop the exploration after K_{proxy} updates and retain $\widetilde{\boldsymbol{V}}_t^{(r)} = \widetilde{\boldsymbol{V}}_t^{(r,K_{\text{proxy}})}$ as the intermediate proxy. Applying the same procedure independently to all frames yields $\widetilde{\boldsymbol{V}}_{1:T}^{(r)}$, which provides the geometric guidance for the subsequent Proxy-Guided Pose Adaptation.

A.4 Ablation of Progressive Mesh-Rig Optimization in Proxyboros

To further examine the progressive design of mesh-rig interaction, we construct a one-shot variant that retains the same proxy-guided pose adaptation while collapsing the recurrent optimization into two sequential stages. Specifically, instead of alternating 5 mesh-space and 5 rig-space updates for 100 rounds, the one-shot variant first performs 500 mesh-space updates to obtain fully optimized frame-wise proxies, followed by 500 rig-space updates against these fixed proxies. The total number of mesh- and rig-space optimization steps is kept the same as in the full model.

As shown in Table 4, despite a small increase in repair quality, removing the progressive mesh-rig interaction leads to clear degradation in pose preservation and motion fidelity. This behavior reflects the ambiguity of static self-intersection repair. When each frame is optimized to convergence before rig-space adaptation, the resulting proxies commit to frame-specific repair solutions without feedback from the realizable skeletal motion. The subsequent pose adaptation must then reconcile these fixed targets within the shared rig space, resulting in greater deviations from the original motion. Progressive optimization instead repeatedly reinitializes mesh-space exploration from the updated rigged geometry, allowing the repair guidance to evolve based on the skeletal response accumulated in previous rounds while remaining compatible with the realizable motion. Fig. 9 further corroborates this phenomenon. Under increasingly stringent PVR requirements, the progressive model retains more samples within restrictive VD ranges, indicating that the recurrent mesh-rig interaction enables strong geometric repair while maintaining smaller deviations from the original pose.

A.5 Ablation of Surface-Aware Proxy Guidance

We further investigate how the two geometric factors in $\mathcal{L}_{\text{sag}}$ contribute to proxy-guided pose adaptation. Its per-vertex residual can be written as $s_{t,i}^2[\overline{\boldsymbol{n}}_{t,i}^\top(\boldsymbol{v}_{t,i}^{\text{rig}} - \tilde{\boldsymbol{v}}_{t,i}^{(r)})]^2$. Here, $\overline{\boldsymbol{n}}_{t,i}$ specifies the local surface direction along which the rig is encouraged to follow the proxy, whereas $s_{t,i}^2$ controls the strength of this guidance according to the agreement between their surface normals. We isolate these

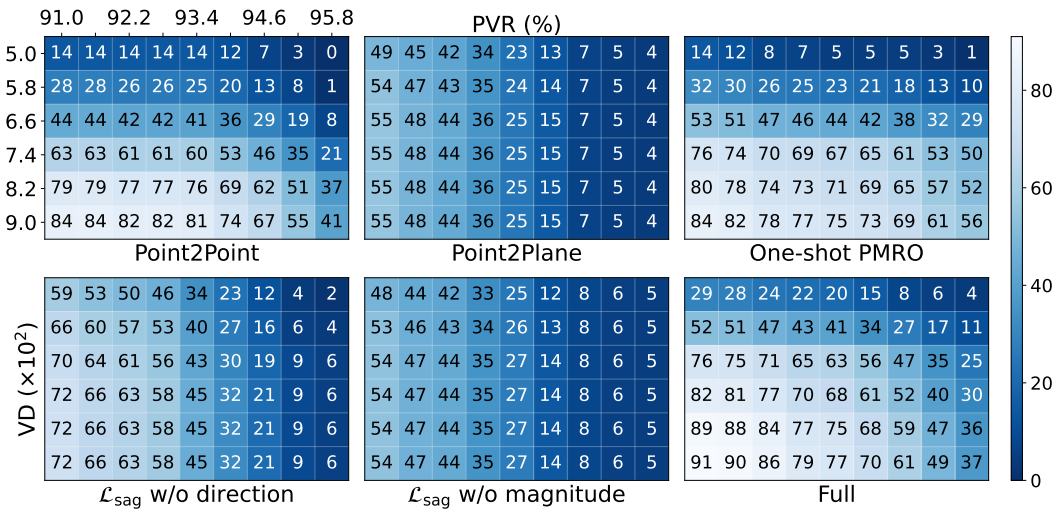


Figure 9: Ablation comparison under joint PVR and VD thresholds. Each cell reports the percentage of samples with PVR above and VD below their respective thresholds, with lighter colors indicating a larger proportion of samples that satisfy both criteria.

two effects by comparing the full objective against point-to-point and conventional point-to-plane alignments (Park & Subbarao, 2003), as well as variants that remove either the directional projection or the orientation-dependent magnitude.

Table 5 reveals distinct roles for the two factors. Removing the orientation-dependent magnitude leaves only the normal-direction projection, yielding results almost identical to those of conventional point-to-plane alignment. This suggests that constraining displacement along the local normal direction alone provides geometric flexibility over point-to-point alignment, but offers little indication of how strongly different proxy corrections should influence the rig. By contrast, removing the directional projection while retaining $s_{t,i}^2$ improves PVR and IFR substantially and achieves the highest SRS. The result indicates that local normal agreement provides an informative signal for modulating proxy guidance: corrections accompanied by small orientation changes receive greater influence, while those involving substantial local reorientation are down-weighted. The full formulation combines this adaptive modulation with normal-direction projection, achieving the highest PVR and IFR. The two factors thus serve complementary purposes, with the magnitude regulating which local corrections are emphasized and the directional term controlling how they are geometrically transferred into rig space.

The improvement in repair quality comes with increased VD, VAD, and VAM, as the full model permits larger geometric and temporal changes to resolve intersections that the more conservative variants leave unresolved. This trade-off is examined more directly in Fig. 9, which reports the proportion of samples satisfying joint PVR and VD thresholds. At loose PVR requirements and very restrictive VD thresholds, conservative alignment variants retain more samples because they introduce smaller deformations. As the required repair quality increases, this advantage diminishes, and the full model preserves a substantially larger proportion of well-repaired samples once moderate pose adjustment is allowed. The heatmap also clarifies why VD should not be interpreted independently: a small deviation may simply reflect insufficient geometric correction. Taken together, these results show that surface-aware proxy guidance improves the rig's ability to absorb effective repair cues while maintaining a controlled balance between penetration removal and the preservation of the original motion.

A.6 Additional Visualizations of Repair Results

We provide additional repair results across diverse geometries, rig structures, and motion sequences on our project page. The interactive viewer supports 360-degree rotation, allowing detailed in-

spection of the repaired self-intersection regions and motion consistency from arbitrary viewpoints. Please move to our project page for more qualitative results.

A.7 Runtime Analysis

Table 6 reports the average optimization time per frame on a 90-frame test sequence. All experiments are conducted on a server equipped with two Intel Xeon Gold 5418Y CPUs and eight NVIDIA RTX 5880 Ada GPUs. To maintain a consistent computational budget across methods, each experiment is restricted to a single GPU and eight CPU workers. Under this setting, ISIR, PoseShield, and Proxyboros use their GPU-parallelized implementations, whereas Mesh-Utg follows its released serial CPU implementation. As shown in Table 6, Proxyboros achieves a practical per-frame runtime while jointly performing self-intersection repair, rig-space pose adaptation, and temporal optimization.

Table 6: Runtime comparison in seconds per frame.

Methods	Runtime (s/frame)
ISIR	**1.53**
Mesh-Utg	34.18
PoseShield	3.04
Ours	<u>2.09</u>

A.8 User Study Details

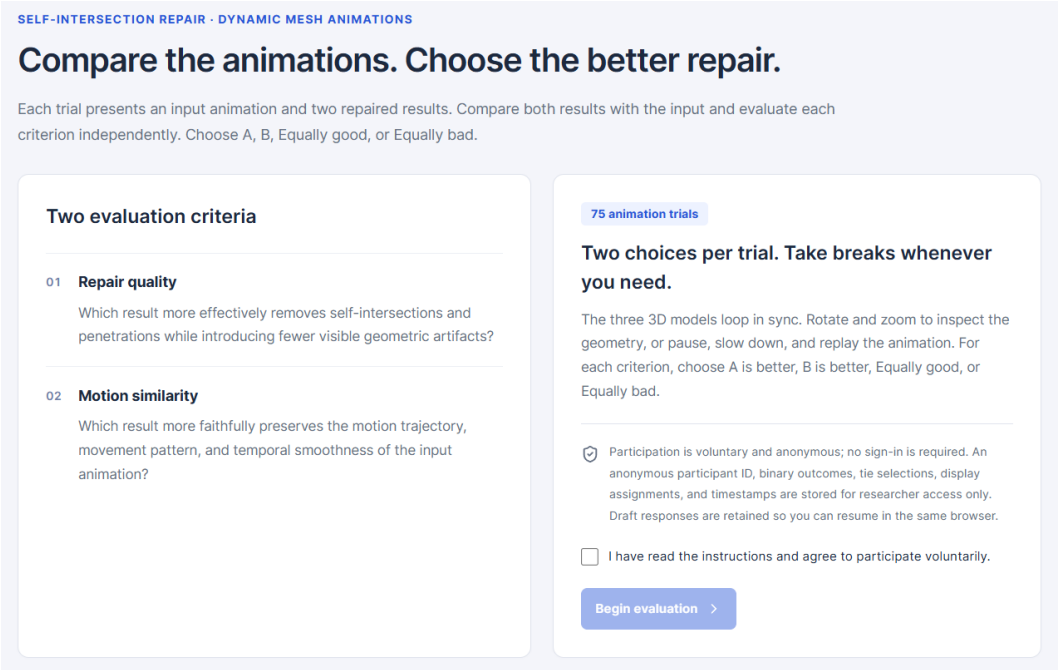
Questionnaire Design. Fig. 10 shows the interface of our user study. Each trial presents the input animation as a reference, together with two anonymous repaired results. All three animations are played synchronously and rendered from aligned viewpoints. Participants can pause, replay, rotate, and zoom the animations to inspect both spatial geometry and temporal behavior. The identities of the compared methods are hidden, while the trial order and the left-right assignment of the two results are randomized to reduce ordering bias.

Evaluation Criteria. Participants evaluate each pair independently under two criteria. *Repair quality* assesses how effectively visible self-intersections and penetrations are removed while avoiding additional geometric artifacts. *Motion similarity* evaluates how well the repaired animation preserves the motion of the input, including its trajectory, movement pattern, and temporal smoothness. For both criteria, participants are instructed to compare the repaired results against the input animation rather than judging their visual appearance alone.

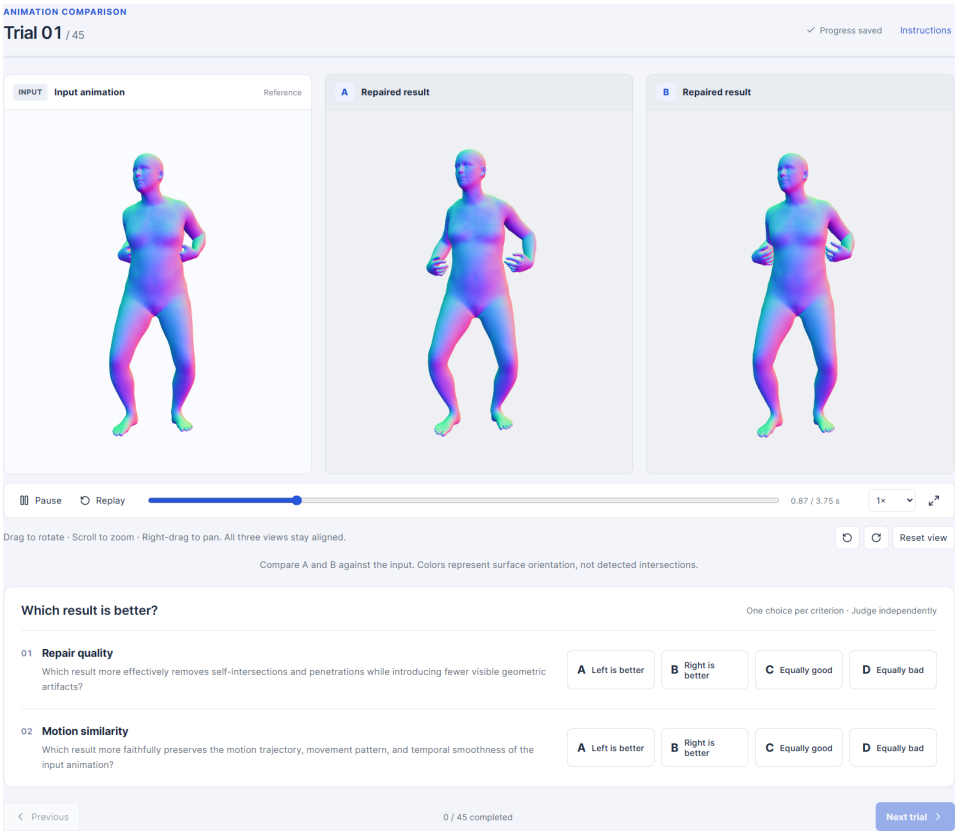
Response Format. For each criterion, participants choose among four options: the left result is better, the right result is better, both are equally good, or both are equally bad. The latter two choices allow participants to distinguish between similarly successful and similarly unsuccessful repairs. In Table 3 of the main paper, the reported tie rate includes only "both are equally good" and thus serves as a positive indicator of cases where our method performs comparably well to the baseline. The win and tie rates are computed from these pairwise responses for each applicable baseline.

A.9 Limitations and Discussions

Proxyboros is constrained by the expressiveness of the original rig, as some severe penetrations may require local deformations that cannot be represented by the available joints and skinning weights. In such cases, complete repair may demand larger pose deviations or remain unattainable within the original deformation space, motivating richer deformation models or more adaptive rig representations. Beyond this representational limitation, Proxyboros is geometry-driven and does not explicitly encode high-level motion semantics. When multiple geometrically feasible repair trajectories exist, geometric cues alone may be insufficient to determine which trajectory best preserves the intended motion, for example, how a limb should pass around the torso or how recurring motion patterns should be preserved across repeated cycles after separating self-intersection regions. A promising direction is to incorporate more generalizable motion priors that capture long-range temporal structure and semantic regularities. Learning such priors robustly remains challenging, in part because large-scale collections of diverse and geometrically valid motion data are limited. By enabling scalable correction of self-intersecting animations, Proxyboros may help facilitate the construction of cleaner motion data for future motion representations.



(a) Starting page with evaluation instructions and criteria.



(b) Example pairwise comparison for an SMPL-H sequence.

Figure 10: Snapshots of user study interface.