ARAPDIFFUSION: GEOMETRY–DISTRIBUTION FEEDBACK FOR DEFORMABLE 3D SHAPE GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Deformable 3D shapes can undergo large changes in pose and morphology while still preserving plausible local geometry, yet generative models often produce samples with unnatural stretching or distortion. In latent generative models, such failures can arise from two coupled sources: the decoder determines the geometry produced at a latent code, while the generative distribution determines which latent regions are actually visited. We introduce ARAPDiffusion, a latent generative framework that establishes feedback between deformation geometry and the generative distribution. First, latent codes sampled from the current diffusion model determine where as-rigid-as-possible (ARAP) constraint is applied, concentrating geometric supervision in regions that the generator actually visits. Second, generated shapes are evaluated by their deformation relative to the observed shape family, and these scores modulate the influence of generated samples during diffusion refinement. In this way, the generative distribution determines where geometry is enforced, while deformation quality in turn reshapes what the model generates. ARAPDiffusion applies to both consistently meshed shapes and unordered point clouds, and the learned prior can also be used for conditional reconstruction. Experiments across human, animal, and bone shape families show consistent improvements in generated-shape quality, distribution matching, and reconstruction accuracy.

1 INTRODUCTION

A central challenge in deformable 3D shape generation is to capture the diversity of a shape family without allowing geometrically implausible deformations. Humans, animals, and anatomical structures can differ substantially in pose and morphology, yet not every geometric variation is plausible: a shape with a reasonable global configuration may still contain unnatural local stretching or distortion. A successful generator must therefore model not only the range of possible shapes, but also how valid shapes can vary from one another. This requirement is important for both unconditional generation and applications that use the learned generator as a shape prior, such as reconstructing complete geometry from incomplete observations.

This issue becomes more subtle in latent generative models because shape generation is split between a learned representation and a learned distribution over that representation. Reconstruction losses constrain the decoder at latent codes corresponding to observed shapes, but provide limited guidance on how the decoder should behave at newly generated codes. A latent diffusion model, meanwhile, learns which regions of the encoded shape space are likely to be sampled. As a result, geometric failures can arise in two different ways: the decoder may behave poorly in regions not sufficiently constrained by reconstruction, or the generative model may place probability on such regions. Addressing either component alone can leave the other source of error unresolved.

Deformation geometry provides a natural prior for this problem. Even under substantial global changes, local regions of deformable shapes often move approximately rigidly. ARAP captures this structure by factoring out local rigid motion and penalizing the remaining distortion, making it well suited to articulated shape families where large valid motions should not themselves be suppressed; previous approaches Huang et al. (2021); Yang et al. (2024) use such geometry to regularize learned shape representations. Latent diffusion, however, exposes a mismatch that representation-level regularization alone does not address. Geometric constraints are typically applied at latent locations chosen independently of the learned generative distribution, even though some regions are far more relevant to generation than others. Conversely, improving the decoder at a latent code does not change

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.

how much probability the diffusion model assigns to that region. This leads to two questions: where should deformation geometry be enforced in a generative latent space, and how can deformation quality influence the generative distribution itself?

We introduce ARAPDiffusion as a geometry–distribution feedback framework whose defining feature is a bidirectional dependency between decoder geometry and the latent generative distribution. First, in *generation-aware geometric refinement*, the current diffusion model supplies latent codes at which ARAP constraints are applied to the decoder. Geometric supervision is therefore concentrated in latent regions that the learned generator actually visits, rather than at locations sampled independently of the generative distribution. Second, in *geometry-aware diffusion refinement*, we evaluate generated shapes by their deformation relative to the observed shape family and use this geometric quality to refine the diffusion model. Samples with better deformation quality receive greater influence during this update. Together, these two state-dependent directions define one feedback step: the current generative distribution determines where geometry is refined, while the resulting deformation quality determines the update to the next generative state.

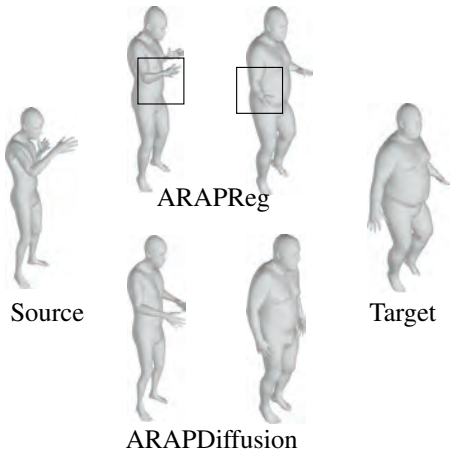


Figure 1: Intermediate shapes between the same source and target. ARAPReg Huang et al. (2021) produces visible geometric distortions in intermediate shapes, while ARAPDiffusion better preserves local geometry throughout the deformation.

The key methodological distinction is the dependency structure between the two refinements rather than either refinement in isolation. Geometric supervision is conditioned on the current generative distribution, whereas the subsequent generative update is conditioned on deformation quality induced by the refined representation. Thus, the supervision received by each component evolves with the state of the other instead of being specified once under a fixed sampling or weighting rule. Alternating these updates realizes a closed state-dependent feedback loop. Fig. 1 illustrates the decoder-side effect by comparing intermediate shapes between the same endpoints, where ARAPDiffusion better preserves local geometry. The same framework applies to consistently meshed shapes and unordered point clouds through mesh and implicit decoders, respectively, and the learned deformation-aware prior can also be used for conditional reconstruction. Across human and animal datasets, with additional results on bone shapes, ARAPDiffusion improves generated-shape quality and distribution matching while also improving reconstruction accuracy. Ablations directly test the two cross-dependencies by retaining the corresponding refinement procedure while replacing its state-dependent signal with a frozen or deformation-independent counterpart.

2 Related Work

Generative models for deformable shapes. For consistently meshed data, COMA Ranjan et al. (2018), Neural3DMM Bouritsas et al. (2019), SP-Disentangle Zhou et al. (2020a), and MeshConv Zhou et al. (2020b) learn low-dimensional representations of non-rigid shape variation, while implicit generative models Atzmon & Lipman (2021); Atzmon et al. (2021) remove the need for shared mesh connectivity. Diffusion-based methods have also been applied to humans and articulated shapes Tevet et al. (2023); Stathopoulos et al. (2024); Zhang et al. (2024); Yi et al. (2024); Lu et al. (2025); Ho et al. (2025), often in parametric spaces such as SMPL Loper et al. (2015) or SMAL Zuffi et al. (2017). Latent diffusion Rombach et al. (2022) performs generation in learned representation spaces and has been widely adopted for 3D generation Zhang et al. (2023); Cheng et al. (2023); Xiang et al. (2025), with joint learning of latent representations and generative models also explored Vahdat et al. (2021). GPLD3D Dong et al. (2024) further incorporates geometric and physical quality priors into latent diffusion for 3D shape generation. Our focus is instead on coupling the geometry induced by a learned shape representation with its generative distribution, so that generation determines where geometric supervision is applied and the resulting geometry in turn influences subsequent generation, for both consistently meshed shapes and unordered point clouds.

Deformation geometry for generative shape models. ARAPReg Huang et al. (2021) regularizes mesh generators using as-rigid-as-possible deformation geometry, while subsequent work develops

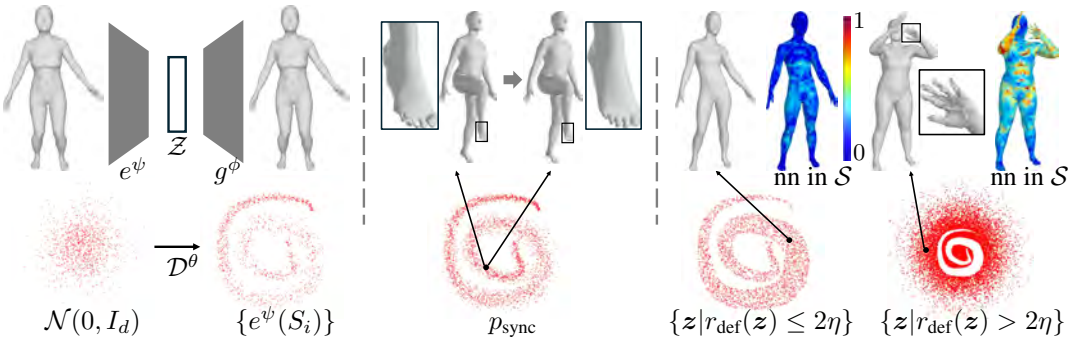


Figure 2: Overview of ARAPDiffusion. Starting from standard latent-diffusion initialization, ARAPDiffusion alternates two feedback directions: samples from the current generative distribution determine where ARAP supervision is applied, while deformation quality of generated shapes modulates their influence during diffusion refinement. Because each direction changes the state used to construct the supervision signal for the other, their alternation forms the proposed geometry–distribution feedback loop.

related geometric constraints, equivariant formulations, and Riemannian descriptions of learned shape spaces Dummer et al. (2024); Atzmon et al. (2022); Muralikrishnan et al. (2022); Foti et al. (2023); Maesumi et al. (2023); Atzmon et al. (2024); GenCorres Yang et al. (2024) extends deformation-based regularization to implicit generators without known correspondence. Related work in computational shape analysis studies geometric metrics and geodesics for deformable shape spaces Klassen et al. (2004); Michor & Mumford (2006); Srivastava et al. (2011); Hartman et al. (2023b); Yang et al. (2023); Hartman et al. (2023a; 2024; 2025); Kilian et al. (2007); Eisenberger et al. (2021); GeoLatent Yang et al. (2023), for example, derives a Riemannian metric from ARAP deformation geometry. These methods establish deformation geometry as an effective prior for learning shape variation, but geometric regularization is typically defined independently of where a learned generative model places probability, and improving the decoder does not by itself alter that distribution. ARAPDiffusion instead couples the two: the current generative distribution determines where geometric supervision is applied, while the resulting deformation quality provides feedback for subsequent generation.

GPLD3D Dong et al. (2024) incorporates geometric and physical priors into latent diffusion, while reward-weighted regression applies quality-weighted denoising to generated samples Black et al. (2024). Our diffusion update follows this general principle, retaining observed-data supervision and restricting generated-sample supervision to low noise levels.

ARAPDiffusion makes geometric representation learning and generative refinement mutually state-dependent: the current generator determines where ARAP regularization is applied, while the refined decoder provides deformation scores for the next generative update. Our ablations test both dependencies using frozen sampling and mean-matched uniform controls.

3 Problem Statement and Approach Overview

Problem statement. Let $\overline{\mathcal{S}}$ denote a deformable shape family and $\mathcal{S} = \{S_1, \ldots, S_N\} \subset \overline{\mathcal{S}}$ a collection of observed shapes. We aim to learn a generative model that produces new, geometrically plausible shapes from the same family, primarily considering unconditional generation and later extending the learned prior to conditional reconstruction. The framework supports both consistently meshed shapes, such as SMPL Loper et al. (2015) and SMAL Zuffi et al. (2017), and unordered point clouds without shared connectivity.

Approach overview. As shown in Fig. 2, ARAPDiffusion builds on latent diffusion Rombach et al. (2022), with an encoder e^ψ mapping observed shapes to a latent space $\mathcal{Z} \cong \mathbb{R}^d$, a decoder $g^\phi : \mathcal{Z} \to \overline{\mathcal{S}}$, and a diffusion model $\mathcal{D}^\theta$ over the encoded shapes. Starting from standard latent-diffusion initialization, ARAPDiffusion alternates two feedback directions: samples from the current generative distribution determine where ARAP supervision is applied, while deformation quality of generated shapes modulates their influence during diffusion refinement. Their alternation forms the proposed geometry–distribution feedback loop. The framework applies to both fixed-topology mesh decoders and implicit decoders for unordered point clouds.

4 METHOD

ARAPDiffusion realizes geometry–distribution feedback through alternating state-dependent updates, where each update changes the sampling or weighting rule used by the other.

4.1 STANDARD LATENT-DIFFUSION INITIALIZATION

We first train an autoencoder on the observed shapes. A shape encoder $e^\psi : \overline{\mathcal{S}} \to \mathcal{Z}$ maps each shape to a latent code, and a decoder $g^\phi : \mathcal{Z} \to \overline{\mathcal{S}}$ reconstructs the shape:

$$\min_{\psi,\phi} \quad \sum_{S\in\mathcal{S}} d^2(S, g^\phi(e^\psi(S))). \tag{1}$$

For consistently meshed shapes, d^2 is the cumulative squared distance between corresponding vertices, using the encoder and decoder architectures of ARAPReg. For unordered point clouds, we use PointNet++ Qi et al. (2017) as the encoder and the GenCorres implicit decoder, with squared algebraic distance as the reconstruction loss.

We then fit a diffusion model to the encoded training shapes. Let p_{data} denote the empirical distribution over $\{e^\psi(S) \mid S \in \mathcal{S}\}$. Following EDM Karras et al. (2022), we train

$$\min_\theta \mathop{\mathbb{E}}_{\boldsymbol{z}\sim p_{\text{data}}} \mathbb{E}\mathbb{E}\lambda(\sigma)\|\mathcal{D}^\theta(\boldsymbol{z}+\boldsymbol{n};\sigma) - \boldsymbol{z}\|^2.$$

The denoising architecture follows DiffuseSDF Chou et al. (2023) and contains attention Vaswani et al. (2017), fully connected, and normalization layers. This provides the initial latent-diffusion model used by the two refinement stages below.

4.2 GENERATION-AWARE GEOMETRIC REFINEMENT

The first refinement improves decoder geometry in latent regions relevant to generation. Standard ARAP regularization samples latent locations from a fixed prior, which need not reflect where the learned generator places probability. We instead sample latent codes from the current diffusion model itself. Let p_{sync} denote this sampling distribution. Holding the diffusion model fixed, we refine the encoder and decoder using

$$\min_{\psi,\phi} \ \frac{1}{|\mathcal{S}|}\sum_{S\in\mathcal{S}} d^2(S, g^\phi(e^\psi(S))) + \mathop{\mathbb{E}}_{\boldsymbol{z}\sim p_{\text{sync}}} l_{\text{reg}}(\boldsymbol{m}^\phi(\boldsymbol{z})) \tag{2}$$

The reconstruction term keeps the representation anchored to the observed shapes, while the second term realizes distribution-to-geometry feedback: drawing $\boldsymbol{z}$ from p_{sync} rather than a fixed Gaussian concentrates ARAP supervision in latent regions emphasized by the current generator.

ARAP should measure geometric distortion rather than deformation magnitude. Following ARAPReg Huang et al. (2021) and GeoLatent Yang et al. (2023), let $H(\boldsymbol{m}^\phi(\boldsymbol{z}))$ denote the ARAP pull-back metric in latent space, which measures the local geometric distortion induced by latent perturbations; its construction is reviewed in Sec. D.1. For example, a limb may undergo a large displacement while remaining locally close to rigid. We therefore normalize the ARAP energy by the magnitude of the induced deformation. For a tangent direction $\boldsymbol{v}$ at $\boldsymbol{z}$, we define

$$r_{\text{reg}}(\boldsymbol{z},\boldsymbol{v}) := \frac{\boldsymbol{v}^T H(\boldsymbol{m}^\phi(\boldsymbol{z}))\boldsymbol{v}}{\boldsymbol{v}^T E(\boldsymbol{m}^\phi(\boldsymbol{z}))\boldsymbol{v}}, \qquad E(\boldsymbol{m}^\phi(\boldsymbol{z})) = \frac{\partial \boldsymbol{m}^\phi(\boldsymbol{z})}{\partial \boldsymbol{z}}^T \frac{\partial \boldsymbol{m}^\phi(\boldsymbol{z})}{\partial \boldsymbol{z}} + \delta I_d,$$

where $\delta = 10^{-10}$ prevents numerical instability for infinitesimal displacements. This yields the normalized metric

$$H_n(\boldsymbol{m}^\phi(\boldsymbol{z})) = E(\boldsymbol{m}^\phi(\boldsymbol{z}))^{-\frac{1}{2}} H(\boldsymbol{m}^\phi(\boldsymbol{z})) E(\boldsymbol{m}^\phi(\boldsymbol{z}))^{-\frac{1}{2}}. \tag{3}$$

We use

$$l_{\text{reg}}(\boldsymbol{m}^\phi(\boldsymbol{z})) = \|H_n(\boldsymbol{m}^\phi(\boldsymbol{z}))\|_\star,$$

where $\|\cdot\|_\star$ denotes the trace norm. This normalization makes the regularizer sensitive to distortion relative to deformation magnitude, rather than penalizing large but geometrically valid shape changes.

For implicit decoders, $\boldsymbol{m}^\phi(\boldsymbol{z})$ denotes a mesh discretization of the implicit surface, and we evaluate the same pull-back metric using the local deformation construction of GenCorres Yang et al. (2024); details are provided in Sec. D.2.

4.3 GEOMETRY-AWARE DIFFUSION REFINEMENT

Improving the decoder does not determine how probability should be distributed across latent space. A latent region may decode smoothly yet still produce shapes that are poor members of the observed family. The reverse geometry-to-distribution direction closes the feedback loop by using deformation quality to refine the diffusion model on generated samples.

For a generated latent code $\boldsymbol{z}$, we define a family-relative deformation score $r_{\text{def}}(\boldsymbol{z})$. A small value indicates that the decoded shape can be related to the observed shape family through relatively low-distortion deformation. We convert this score into

$$w(\boldsymbol{z}) = \exp\left(- \frac{r_{\text{def}}^2(\boldsymbol{z})}{2\eta^2} \right),$$

where η is the median leave-one-out deformation score of the encoded training shapes; when calibrating this scale on a training shape, the query shape itself is excluded from the nearest-neighbor search. At each alternation, p_{data} is formed from the training shapes encoded by the current encoder after the distribution-to-geometry update. We then fix the encoder and decoder and refine $\mathcal{D}^\theta$ by augmenting the original EDM objective with weighted generated samples:

$$\min_\theta \quad \mathop{\mathbb{E}}_{\boldsymbol{z}\sim p_{\text{data}}} \mathop{\mathbb{E}}_{\sigma}\mathop{\mathbb{E}}_{\boldsymbol{n}} \lambda(\sigma)\|\mathcal{D}^\theta(\boldsymbol{z}+\boldsymbol{n};\sigma) - \boldsymbol{z}\|^2 + \mathop{\mathbb{E}}_{\boldsymbol{z}\sim p_{\text{sync}}} \mathop{\mathbb{E}}_{\sigma\le\sigma_0}\mathop{\mathbb{E}}_{\boldsymbol{n}} \lambda(\sigma) w(\boldsymbol{z}) \|\mathcal{D}^\theta(\boldsymbol{z}+\boldsymbol{n};\sigma) - \boldsymbol{z}\|^2. \quad (4)$$

The first term anchors diffusion to the encoded observations. The second uses deformation quality as a confidence signal for generated samples: geometrically better generated samples receive greater influence, while poorly deformed samples are suppressed. Viewed in isolation, this is quality-weighted refinement on generated samples. Within the alternating procedure, however, these weights are produced by decoder geometry refined at locations sampled from the current generator; updating $\mathcal{D}^\theta$ then changes that generator, and hence the latent locations used by the next geometry update. Because the weights are not normalized across generated samples, they affect both relative contributions among generated samples and the aggregate strength of generated-sample refinement with respect to the data term. We intentionally retain both effects, treating $w(\boldsymbol{z})$ as confidence rather than as a normalized importance ratio, and therefore do not interpret Eq. 4 as defining an explicit normalized target distribution. In Sec. 5.4, a mean-matched uniform control separately tests the sample-dependent component while preserving aggregate generated-sample weight. We restrict this update to small noise levels $\sigma \le \sigma_0$, so the feedback remains local around generated codes while the large-noise regime remains anchored by p_{data}. We set σ_0 to half of the average distance from each training latent code to its nearest other training code.

Deformation score. For consistently meshed shapes, we evaluate deformation quality relative to the closest observed shape in latent space. Because mesh connectivity is shared, vertex correspondence is known. At each vertex, we factor out the best local rotation and measure the residual deformation; small residuals indicate locally near-rigid changes, while large residuals indicate stronger distortion.

Formally, for two meshes with vertex vectors $\boldsymbol{p}, \boldsymbol{q} \in \mathbb{R}^{3n}$ and vertex neighborhood $\mathcal{N}(i)$, we define

$$r_{\text{def}}^i(\boldsymbol{p}, \boldsymbol{q}) = \left(\min_{R\in \mathrm{SO}(3)} \sum_{j\in\mathcal{N}(i)} \|R(\boldsymbol{p}_i - \boldsymbol{p}_j) - (\boldsymbol{q}_i - \boldsymbol{q}_j)\|^2 \right)^{\frac{1}{2}}. \quad (5)$$

The optimal rotation is obtained by singular-value decomposition of $M_i = \sum_{j\in\mathcal{N}(i)} (\boldsymbol{p}_i - \boldsymbol{p}_j)(\boldsymbol{q}_i - \boldsymbol{q}_j)^T$ Arun et al. (1987). Let $S_{i_{\boldsymbol{z}}}$ denote the closest training shape to $\boldsymbol{m}^\phi(\boldsymbol{z})$. We compute $r_{\text{def}}^i(\boldsymbol{m}^\phi(\boldsymbol{z}), S_{i_{\boldsymbol{z}}})$ over all vertices and use their p_c-percentile as $r_{\text{def}}(\boldsymbol{z})$, with $p_c = 80$ in our experiments.

We use a percentile rather than the maximum because articulated joints can naturally exhibit larger deformation than most of the surface, causing a small number of vertices to dominate as p_c approaches 100. The score correlates with SMPL/SMAL reconstruction error and remains stable over a broad range of p_c; we provide the corresponding analysis in Appendix B.

Implicit generative model. The same principle applies to implicit generators, but explicit vertex correspondence is unavailable. We therefore use GenCorres Yang et al. (2024) to deform the closest

training shape toward the generated implicit surface, providing a temporary correspondence through which the same local deformation error can be evaluated.

Let $S_{i_{\boldsymbol{z}}}(\boldsymbol{z})$ denote the training shape after deformation toward $g^{\phi}(\boldsymbol{z})$. We define

$$r_{\text{def}}(\boldsymbol{z}) = \text{prctile}\Big(r^{j}_{\text{def}}\big(S_{i_{\boldsymbol{z}}}, S_{i_{\boldsymbol{z}}}(\boldsymbol{z})\big), p_c\Big) + d_{\text{Haus}}\big(S_{i_{\boldsymbol{z}}}(\boldsymbol{z}), g^{\phi}(\boldsymbol{z})\big).$$

The first term measures the deformation required from the observed shape, while the second measures how well the deformed shape reaches the generated surface. Validation of the deformation score for implicit generators is provided in Appendix B. For the implicit datasets considered here, we additionally set $w(\boldsymbol{z}) = 0$ for topologically invalid generated meshes whose Euler characteristic $|\mathcal{V}| - |\mathcal{E}| + |\mathcal{F}| \neq 2$. Additional score variants are discussed in the supplementary material.

4.4 Extension to Conditional Reconstruction

The deformation-aware prior learned above can also be used for reconstruction from partial observations. We first learn the decoder using the unconditional procedure in Secs. 4.1–4.3, then fix it and train a conditional diffusion model $\mathcal{D}^{\theta}(\boldsymbol{z} + \boldsymbol{n}, \sigma; c)$, where c denotes the observation.

A valid conditional sample must satisfy both the observed condition and the deformation prior. We therefore draw latent codes from the current conditional diffusion model, denoting the resulting distribution by p^{c}_{sync}, and weight their contribution using the same deformation score:

$$\min_{\theta} \quad \mathop{\mathbb{E}}_{\boldsymbol{z}\sim p_{\text{data}}} \mathop{\mathbb{E}}_{\sigma} \mathop{\mathbb{E}}_{\boldsymbol{n}} \lambda(\sigma)\|\mathcal{D}^{\theta}(\boldsymbol{z}+\boldsymbol{n},\sigma,c)-\boldsymbol{z}\|^2 + \mu \mathop{\mathbb{E}}_{\boldsymbol{z}\sim p^{c}_{\text{sync}}} \mathop{\mathbb{E}}_{\sigma\le\sigma_0} \mathop{\mathbb{E}}_{\boldsymbol{n}} \lambda(\sigma) w(\boldsymbol{z}) \|\mathcal{D}^{\theta}(\boldsymbol{z}+\boldsymbol{n},\sigma,c)-\boldsymbol{z}\|^2. \quad (6)$$

5 Experimental Results

5.1 Experimental Setup

We evaluate ARAPDiffusion on the Human and Animal datasets used by ARAPReg Huang et al. (2021) and GenCorres Yang et al. (2024), which are generated from SMPL Loper et al. (2015) and SMAL Zuffi et al. (2017), respectively. For mesh generators, we follow the training/test splits of ARAPReg; for implicit generators, we follow those of GenCorres. Additional experiments on the Bone datasets from ARAPReg are provided in the supplementary material.

We evaluate four complementary aspects of the learned model. The reconstruction error e_r, following ARAPReg and GenCorres, measures reconstruction on held-out shapes. For generated-shape quality, e_t is the Hausdorff distance to the corresponding SMPL/SMAL fit (Sec. E); because fitting is initialized and regularized using the nearest training shape, we treat e_t as a fitting-based proxy rather than an exact distance to the parametric manifold. The distribution metric d_{W} is the empirical Wasserstein distance between 2000 generated and training samples in ambient space, using Chamfer distance as the pairwise ground cost. We therefore use e_t as a generated-shape quality proxy and rely on d_W as a complementary distribution-level measure, rather than interpreting e_t alone as a direct estimate of distributional fidelity. Finally, d_n is the Hausdorff distance to the nearest training shape, serving as a sanity check on training-set proximity rather than a standalone coverage metric. For implicit generators, we extract triangular meshes before evaluation. Additional evaluation details and timing are provided in Sec. E.

5.2 Unconditional Mesh Generators

We first evaluate unconditional mesh generation against SP-Disent. Zhou et al. (2020a), COMA Ranjan et al. (2018), Neural3DMM Bouritsas et al. (2019), MeshConv Zhou et al. (2020b), ARAPReg Huang et al. (2021), FrameAve Atzmon et al. (2022), GeoLatent Yang et al. (2023), and BRESA Hartman et al. (2025). Tab. 1 reports the quantitative results, while Fig. 3 provides a spatial comparison of reconstruction errors on held-out test shapes.

ARAPDiffusion outperforms all prior baselines in e_r, e_t, and d_W on both Human and Animal. The largest gains occur in e_t and d_W, the metrics most directly associated with the two goals of our framework: improving the geometry of individual generated shapes and improving distribution-level matching. The d_n values remain comparable to those of the baselines, providing no indication that these gains are obtained simply by moving generated samples closer to the training set.

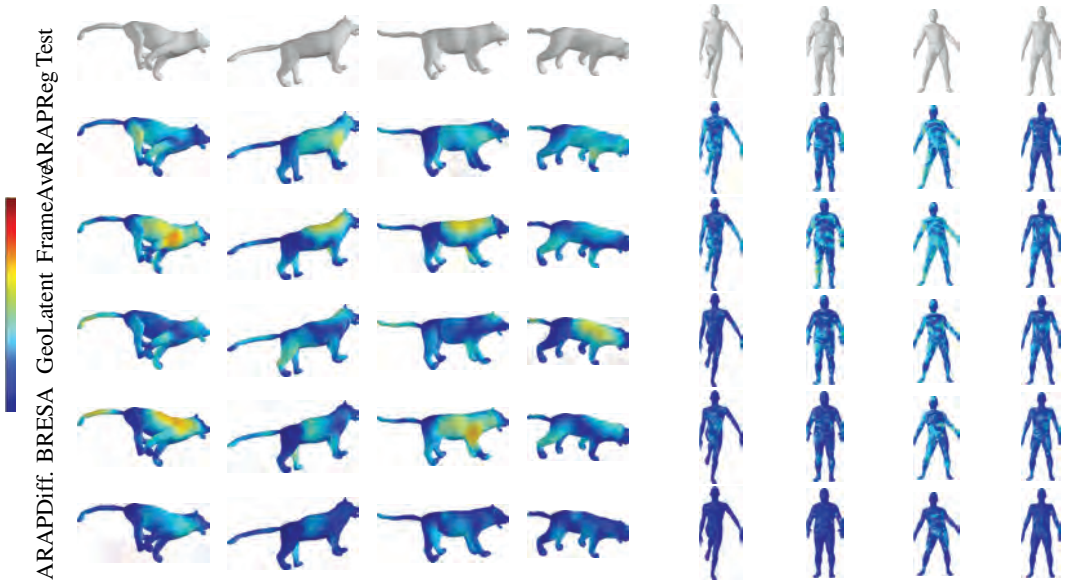


Figure 3: Color-coded reconstruction errors on held-out test shapes. We compare ARAPDiffusion with the strongest deformation-aware baselines.

NoReg is already a strong latent-diffusion baseline and substantially outperforms prior methods in generated-shape quality e_t and distribution matching d_W. Its advantage does not extend uniformly to reconstruction, however: its e_r is worse than several deformation-aware baselines on both datasets. ARAPDiffusion further reduces $e_r/e_t/d_W$ over NoReg by 19.3%/32.0%/17.9% on Human and 15.6%/16.1%/20.4% on Animal, showing that deformation geometry provides substantial complementary gains on top of the strong latent-diffusion backbone. Additional random samples from all methods are provided in the supplementary material.

5.3 Unconditional Implicit Generators

We next test whether the same geometry–generation coupling extends beyond meshes with known correspondence. Fig. 4 and Tab. 2 compare ARAPDiffusion with SALD Atzmon & Lipman (2021) and GenCorres Yang et al. (2024). We additionally include HY-LoRA, a strong reference obtained by fine-tuning the pretrained Hunyuan3D 2.5 model Lai et al. (2025) on the training data using LoRA Hu et al. (2022).

	Human				Animal			
	$e_r\downarrow$	$e_t\downarrow$	$d_W\downarrow$	d_n	$e_r\downarrow$	$e_t\downarrow$	$d_W\downarrow$	d_n
SP-Disent.	10.0	24.1	23.1	14.2	15.2	14.3	26.1	15.3
COMA	8.80	22.3	20.2	10.7	14.5	13.2	23.2	12.9
3DMM	7.39	19.1	21.3	9.81	17.3	13.6	24.7	11.1
MeshConv	5.43	16.8	19.3	9.76	8.01	11.9	20.1	10.8
ARAPReg	4.52	19.7	18.9	6.81	9.68	12.5	22.5	10.9
FrameAve	4.94	20.6	16.7	9.54	12.5	10.6	20.2	11.1
GeoLatent	4.31	17.2	19.1	9.81	5.98	11.2	21.2	11.2
BRESA	4.41	22.4	20.1	9.71	6.12	12.6	23.4	10.7
ARAPDiff.	**4.18**	**7.21**	**12.4**	9.85	**5.75**	**7.35**	**14.4**	10.9
NoReg	5.18	10.6	15.1	9.86	6.81	8.76	18.1	10.9
NoDeReg	4.74	8.05	13.9	9.77	6.16	7.99	16.1	11.0
NoLaReg	4.34	7.27	13.4	9.87	6.03	7.71	15.3	11.1
FrozenInit	4.32	7.89	13.6	9.88	6.01	7.84	15.8	11.0
MeanMatch	4.21	7.26	13.2	9.86	5.82	7.61	15.2	10.8
AlterDiff	4.16	7.19	12.3	9.86	5.77	7.34	14.3	10.8

Table 1: Comparison with prior mesh-generation methods and variants of ARAPDiffusion. We report e_r and e_t in mm and d_W and d_n in cm. ARAPDiffusion is highlighted in bold; ablations and the alternative diffusion backbone are reported below the dashed line.

ARAPDiffusion achieves the lowest e_r, e_t, and d_W on both Human and Animal. More importantly, the improvement pattern mirrors the mesh setting: the gains are substantially larger in generated-shape quality and distribution matching than in reconstruction. Compared with GenCorres, e_r decreases by 2.06% and 5.95% on Human and Animal, respectively, while the larger reductions in e_t and d_W show that incorporating deformation geometry into generation benefits both decoded samples and the learned distribution. This consistency indicates that the proposed coupling does not rely on explicit mesh correspondence.

ARAPDiffusion also improves e_r, e_t, and d_W relative to HY-LoRA, despite the latter starting from a large-scale pretrained 3D model. HY-LoRA has a smaller d_n, indicating that its samples lie closer to the training set. Fig. 4 provides qualitative comparisons of the reconstructed shapes.

	Human				Animal			
	$e_r\downarrow$	$e_t\downarrow$	$d_W\downarrow$	d_n	$e_r\downarrow$	$e_t\downarrow$	$d_W\downarrow$	d_n
SALD	7.21	16.9	21.8	10.2	8.12	18.9	20.5	11.2
GenCorres	6.31	16.4	22.1	9.89	7.56	19.1	21.2	11.3
HY-LoRA	6.23	8.84	15.9	9.8	7.23	9.25	16.9	8.5
ARAPDiff.	**6.18**	**8.39**	**14.8**	10.1	**7.11**	**8.91**	**16.5**	11.1

Table 2: Comparison with implicit-generation baselines under e_r, e_t, d_W, and d_n. The best result for e_r, e_t, and d_W is bold-faced.

5.4 Ablation Study

We ablate the feedback loop at two levels: NoDeReg and NoLaReg remove one refinement direction entirely, while FrozenInit and MeanMatchedUniform retain the corresponding refinement procedure but replace its state-dependent signal with a static or deformation-independent counterpart, respectively. FrozenInit tests the distribution-to-geometry dependency by sampling geometric-refinement locations throughout training from a frozen copy of the diffusion model obtained after standard initialization, rather than from the evolving generator. MeanMatchedUniform isolates the sample-dependent component of geometry-to-distribution feedback. For each generated minibatch $\mathcal{B}$, we first compute the deformation-dependent weights $w(\boldsymbol{z}_i)$ used by the full model and their mean $\bar{w}_{\mathcal{B}} = \frac{1}{|\mathcal{B}|}\sum_{\boldsymbol{z}_i\in\mathcal{B}} w(\boldsymbol{z}_i)$. We then assign the same weight $\bar{w}_{\mathcal{B}}$ to every generated sample in that minibatch. Consequently, the aggregate generated-sample weight is preserved, while the deformation-dependent relative weighting among samples is removed. All other training and evaluation settings are kept unchanged across each controlled comparison, so that each variant alters only the feedback component under study. Results for mesh generators are reported in Tab. 1.

NoReg (standard latent diffusion). Removing both geometry-based refinements increases e_r, e_t, and d_W by 23.9%/18.4%, 47.0%/19.1%, and 21.7%/25.7% on Human/Animal, respectively. Thus, while latent diffusion already provides strong generation, deformation-aware refinement contributes substantial additional gains across representation and generation metrics.

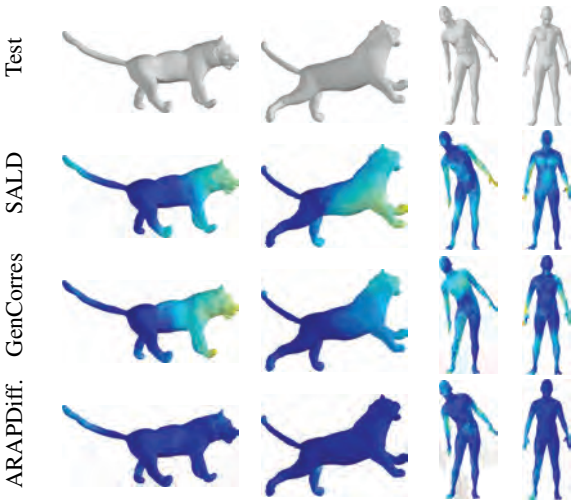


Figure 4: Qualitative reconstruction comparison for implicit generators. ARAPDiffusion more closely reproduces the held-out test shapes.

NoDeReg (without geometric decoder regularization). Removing generation-aware geometric regularization, while retaining the reconstruction update and geometry-aware diffusion refinement, increases e_r, e_t, and d_W by 13.4%/7.13%, 11.6%/8.71%, and 12.1%/11.8%, respectively. The degradation relative to the full model shows that geometric supervision of the decoder in regions sampled by the current generator provides a complementary contribution.

NoLaReg (geometric refinement only). Removing geometry-aware diffusion refinement increases e_r, e_t, and d_W by 3.83%/4.87%, 0.83%/4.90%, and 8.06%/6.25% on Human/Animal, respectively. The largest and most consistent degradation occurs in d_W, which is consistent with geometry-to-distribution feedback acting primarily on the generative model rather than directly on decoder geometry. These results show that geometry-aware diffusion refinement provides a complementary benefit beyond decoder-side geometric refinement. Together with NoDeReg, they support the contribution of both directions to the full geometry–distribution feedback loop.

FrozenInit (frozen-initial-generator refinement). To assess whether distribution-to-geometry feedback benefits from tracking the evolving generative distribution, we freeze a copy of the diffusion model immediately after standard latent-diffusion initialization. At every subsequent geometry-refinement update, we draw fresh latent samples from this fixed initial generator rather than from the current diffusion model. The trainable diffusion model, geometry-aware diffusion refinement,

optimization schedule, and all other settings remain unchanged. Compared with FrozenInit, the full model reduces e_t from 7.89 to 7.21 on Human and from 7.84 to 7.35 on Animal, with corresponding reductions in d_W from 13.6 to 12.4 and from 15.8 to 14.4. Since both variants use latent locations sampled from a learned diffusion model and differ in whether this sampling distribution follows the evolving generator, this comparison tests the benefit of updating geometric supervision with the current generative state rather than fixing it at initialization.

MeanMatchedUniform (mean-matched uniform refinement). To isolate deformation-dependent sample modulation from the aggregate strength of geometry-aware diffusion refinement, for each generated minibatch $\mathcal{B}$ we replace the individual weights $w(\boldsymbol{z}_i)$ with their batch mean $\bar{w}_{\mathcal{B}}$. This preserves the total generated-sample weight within the minibatch while removing all deformation-dependent relative weighting among its samples. Compared with MeanMatchedUniform, the full model reduces d_W from 13.2 to 12.4 on Human and from 15.2 to 14.4 on Animal, while e_t changes from 7.26 to 7.21 and from 7.61 to 7.35. The larger effect on d_W is consistent with the role of this intervention: decoder-side geometric refinement is unchanged, while deformation-dependent relative weighting is removed specifically from the generative-model update. Because the two weighting rules match the aggregate weight on the same generated minibatch, the comparison separates sample-dependent geometric feedback from changes in overall generated-sample weight. Together with FrozenInit, these controls support the two state dependencies underlying the feedback loop: geometric supervision tracks the evolving generative distribution, while generated-sample influence depends on deformation quality.

AlterDiff (alternative generative backbone). Finally, we replace EDM with the rectified flow model Liu et al. (2023). Performance remains nearly unchanged, and in several entries differs only marginally from the EDM version. This suggests that the observed improvements are not specific to EDM and persist under an alternative generative formulation.

5.5 Conditional Reconstruction

Fig. 5 qualitatively evaluates the extension to reconstruction from partial observations against GeoLatent and BRESA. Each method uses a conditional diffusion model in its own latent space to infer a complete shape from the observed points. Across the shown examples, ARAPDiffusion recovers plausible unobserved geometry while remaining consistent with the partial input, illustrating that the learned deformation-aware prior can also be used for conditional reconstruction.

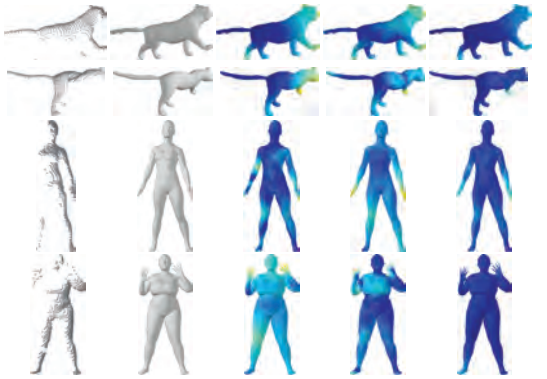


Figure 5: Conditional reconstruction from partial observations. ARAPDiffusion uses the learned shape prior together with the observed points to recover the complete shape.

6 Conclusions and Future Work

We presented ARAPDiffusion, which couples decoder geometry and latent generation through a closed state-dependent loop. The current generator determines where decoder geometry is refined, while the refined decoder determines how generated samples update the next generator. Because each step changes the state conditioning the other, the contribution lies in this bidirectional dependency rather than either refinement or alternation alone. Our ablations test both directions and their state dependence against static or deformation-independent controls. The resulting coupling improves generated-shape quality and distribution matching, applies to mesh and implicit representations, and yields a deformation-aware prior for conditional reconstruction from partial observations.

Future work includes integrating deformation quality more directly into diffusion training or sampling, extending the framework to broader shape representations and families, and exploring dynamic settings such as 4D shape and motion generation. More broadly, our results suggest that geometric structure can provide useful feedback not only for learning shape representations, but also for shaping the generative distributions built on top of them.

Ethics Statement

This work does not involve human-subject studies, animal experiments, or the collection of personally identifiable information. The datasets used in our experiments are derived from established shape models or adopted from prior work. We are not aware of any specific ethical concerns beyond the general considerations associated with generative modeling.

Reproducibility Statement

We provide the model formulation, training objectives, evaluation protocols, implementation details, and hyperparameters in the main paper and appendices. Dataset splits and experimental settings follow the corresponding prior work where applicable. Ablation variants are evaluated under the same experimental protocol, with each control modifying only the component specified in Sec. 5.4.

AI Use Statement

Generative AI tools were used to assist with language editing, manuscript preparation, and to assist with coding for experimental implementation. The clarity of presentation all technical claims, mathematical formulations, experimental results, figures, tables, and conclusions were verified, and interpreted by the authors. The authors reviewed all AI-assisted edits and take full responsibility for the final content of the paper.

References

K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(1):67–71, 1987.

Matan Atzmon and Yaron Lipman. SALD: sign agnostic learning with derivatives. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=7EDgLu9reQD.

Matan Atzmon, David Novotný, Andrea Vedaldi, and Yaron Lipman. Augmenting implicit neural shape representations with explicit deformation fields. *CoRR*, abs/2108.08931, 2021. URL https://arxiv.org/abs/2108.08931.

Matan Atzmon, Koki Nagano, Sanja Fidler, Sameh Khamis, and Yaron Lipman. Frame averaging for equivariant shape space learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 631–641, June 2022.

Matan Atzmon, Jiahui Huang, Francis Williams, and Or Litany. Approximately piecewise E(3) equivariant point networks. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net, 2024. URL https://openreview.net/forum?id=aKJEHWmBEf.

Kevin Black, Michael Janner, Yilun Du, Ilya Kostrikov, and Sergey Levine. Training diffusion models with reinforcement learning. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=nMkdvj7BF8.

Giorgos Bouritsas, Sergiy Bokhnyak, Stylianos Ploumpis, Stefanos Zafeiriou, and Michael M. Bronstein. Neural 3d morphable models: Spiral convolutional networks for 3d shape representation learning and generation. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pp. 7212–7221. IEEE, 2019. doi: 10.1109/ICCV.2019.00731. URL https://doi.org/10.1109/ICCV.2019.00731.

Yen-Chi Cheng, Hsin-Ying Lee, Sergey Tulyakov, Alexander G Schwing, and Liang-Yan Gui. Sdfusion: Multimodal 3d shape completion, reconstruction, and generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4456–4465, 2023.

Gene Chou, Yuval Bahat, and Felix Heide. Diffusion-sdf: Conditional generative modeling of signed distance functions. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2023.

Yuan Dong, Qi Zuo, Xiaodong Gu, Weihao Yuan, Zhengyi Zhao, Zilong Dong, Liefeng Bo, and Qixing Huang. Gpld3d: Latent diffusion of 3d shape generative models by enforcing geometric and physical priors. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 56–66. IEEE, 2024.

Sven Dummer, Nicola Strisciuglio, and Christoph Brune. Rda-inr: Riemannian diffeomorphic autoencoding via implicit neural representations. *SIAM Journal on Imaging Sciences*, 17(4):2302–2330, 2024. doi: 10.1137/24M1644730. URL https://doi.org/10.1137/24M1644730.

Marvin Eisenberger, David Novotný, Gael Kerchenbaum, Patrick Labatut, Natalia Neverova, Daniel Cremers, and Andrea Vedaldi. Neuromorph: Unsupervised shape interpolation and correspondence in one go. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, pp. 7473–7483. Computer Vision Foundation / IEEE, 2021. doi: 10.1109/CVPR46437.2021.00739. URL https://openaccess.thecvf.com/content/CVPR2021/html/Eisenberger_NeuroMorph_Unsupervised_Shape_Interpolation_and_Correspondence_in_One_Go_CVPR_2021_paper.html.

Simone Foti, Bongjin Koo, Danail Stoyanov, and Matthew J. Clarkson. 3d generative model latent disentanglement via local eigenprojection. *Comput. Graph. Forum*, 42(6), 2023. doi: 10.1111/CGF.14793. URL https://doi.org/10.1111/cgf.14793.

Emmanuel Hartman, Emery Pierson, Martin Bauer, Nicolas Charon, and Mohamed Daoudi. Bareesa: A riemannian framework for unregistered human body shapes. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*, pp. 14135–14145. IEEE, 2023a. doi: 10.1109/ICCV51070.2023.01304. URL https://doi.org/10.1109/ICCV51070.2023.01304.

Emmanuel Hartman, Yashil Sukurdeep, Eric Klassen, Nicolas Charon, and Martin Bauer. Elastic shape analysis of surfaces with second-order sobolev metrics: A comprehensive numerical framework. *Int. J. Comput. Vis.*, 131(5):1183–1209, 2023b. doi: 10.1007/S11263-022-01743-0. URL https://doi.org/10.1007/s11263-022-01743-0.

Emmanuel Hartman, Nicolas Charon, and Martin Bauer. Self supervised networks for learning latent space representations of human body scans and motions, 2024. URL https://arxiv.org/abs/2411.03475.

Emmanuel Hartman, Emery Pierson, Martin Bauer, Mohamed Daoudi, and Nicolas Charon. Basis restricted elastic shape analysis on the space of unregistered surfaces. *Int. J. Comput. Vis.*, 133(4):1999–2024, 2025. doi: 10.1007/S11263-024-02269-3. URL https://doi.org/10.1007/s11263-024-02269-3.

Hsuan-I Ho, Chen Guo, Po-Chen Wu, Ivan Shugurov, Chengcheng Tang, Abhay Mittal, Sizhe An, Manuel Kaufmann, and Linguang Zhang. Phd: Personalized 3d human body fitting with point diffusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7526–7537, 2025.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

Qixing Huang, Xiangru Huang, Bo Sun, Zaiwei Zhang, Junfeng Jiang, and Chandrajit Bajaj. Arapreg: An as-rigid-as possible regularization loss for learning deformable shape generators. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5815–5825, October 2021.

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022. URL http://papers.nips.cc/paper_files/paper/2022/hash/a98846e9d9cc01cfb87eb694d946ce6b-Abstract-Conference.html.

Martin Kilian, Niloy J. Mitra, and Helmut Pottmann. Geometric modeling in shape space. *ACM Trans. Graph.*, 26(3):64–es, July 2007. ISSN 0730-0301. doi: 10.1145/1276377.1276457. URL https://doi.org/10.1145/1276377.1276457.

Eric Klassen, Anuj Srivastava, Washington Mio, and Shantanu H. Joshi. Analysis of planar shapes using geodesic paths on shape spaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(3):372–383, 2004. doi: 10.1109/TPAMI.2004.1262333. URL https://doi.org/10.1109/TPAMI.2004.1262333.

Zeqiang Lai, Yunfei Zhao, Haolin Liu, Zibo Zhao, Qingxiang Lin, Huiwen Shi, Xianghui Yang, Mingxin Yang, Shuhui Yang, Yifei Feng, Sheng Zhang, Xin Huang, Di Luo, Fan Yang, Fang Yang, Lifu Wang, Sicong Liu, Yixuan Tang, Yulin Cai, Zebin He, Tian Liu, Yuhong Liu, Jie Jiang, Linus, Jingwei Huang, and Chunchao Guo. Hunyuan3d 2.5: Towards high-fidelity 3d assets generation with ultimate details. *CoRR*, abs/2506.16504, 2025. doi: 10.48550/ARXIV.2506.16504. URL https://doi.org/10.48550/arXiv.2506.16504.

Xingchao Liu, Chengyue Gong, and Qiang Liu. Flow straight and fast: Learning to generate and transfer data with rectified flow. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/forum?id=XVjTT1nw5z.

Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. SMPL: A skinned multi-person linear model. *ACM Transactions on Graphics*, 34(6), 2015.

Junzhe Lu, Jing Lin, Hongkun Dou, Ailing Zeng, Yue Deng, Xian Liu, Zhongang Cai, Lei Yang, Yulun Zhang, Haoqian Wang, et al. Dposer-x: Diffusion model as robust 3d whole-body human pose prior. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9988–9997, 2025.

Arman Maesumi, Paul Guerrero, Noam Aigerman, Vladimir Kim, Matthew Fisher, Siddhartha Chaudhuri, and Daniel Ritchie. Explorable mesh deformation subspaces from unstructured 3d generative models. In *SIGGRAPH Asia 2023 Conference Papers*, SA '23, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400703157. doi: 10.1145/3610548.3618192. URL https://doi.org/10.1145/3610548.3618192.

Peter W. Michor and David B. Mumford. Riemannian geometries on spaces of plane curves. *Journal of the European Mathematical Society*, 8(1):1–48, 2006. doi: 10.4171/jems/37. URL https://geodesic-test.mathdoc.fr/articles/10.4171/jems/37/.

Sanjeev Muralikrishnan, Siddhartha Chaudhuri, Noam Aigerman, Vladimir G. Kim, Matthew Fisher, and Niloy J. Mitra. GLASS: geometric latent augmentation for shape spaces. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pp. 470–479. IEEE, 2022. doi: 10.1109/CVPR52688.2022.01800. URL https://doi.org/10.1109/CVPR52688.2022.01800.

Charles R. Qi, Li Yi, Hao Su, and Leonidas J. Guibas. Pointnet++: deep hierarchical feature learning on point sets in a metric space. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 5105–5114, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Anurag Ranjan, Timo Bolkart, Soubhik Sanyal, and Michael J. Black. Generating 3d faces using convolutional mesh autoencoders. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (eds.), *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part III*, volume 11207 of *Lecture Notes in Computer Science*, pp. 725–741. Springer, 2018. doi: 10.1007/978-3-030-01219-9_43. URL https://doi.org/10.1007/978-3-030-01219-9_43.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, pp. 10674–10685. IEEE, 2022. doi: 10.1109/CVPR52688.2022.01042. URL https://doi.org/10.1109/CVPR52688.2022.01042.

Anuj Srivastava, Eric Klassen, Shantanu H. Joshi, and Ian H. Jermyn. Shape analysis of elastic curves in euclidean spaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(7):1415–1428, 2011. doi: 10.1109/TPAMI.2010.184. URL https://doi.org/10.1109/TPAMI.2010.184.

Anastasis Stathopoulos, Ligong Han, and Dimitris N. Metaxas. Score-guided diffusion for 3d human recovery. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16-22, 2024*, pp. 906–915. IEEE, 2024. doi: 10.1109/CVPR52733.2024.00092. URL https://doi.org/10.1109/CVPR52733.2024.00092.

Guy Tevet, Sigal Raab, Brian Gordon, Yonatan Shafir, Daniel Cohen-Or, and Amit Haim Bermano. Human motion diffusion model. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/forum?id=SJ1kSyO2jwu.

Arash Vahdat, Karsten Kreis, and Jan Kautz. Score-based generative modeling in latent space. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 11287–11302, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/5dca4c6b9e244d24a30b4c45601d9720-Abstract.html.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 6000–6010, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

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 *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pp. 21469–21480, June 2025.

Haitao Yang, Bo Sun, Liyan Chen, Amy Pavel, and Qixing Huang. Geolatent: A geometric approach to latent space design for deformable shape generators. *ACM Trans. Graph.*, 42(6), December 2023. ISSN 0730-0301. doi: 10.1145/3618371. URL https://doi.org/10.1145/3618371.

Haitao Yang, Xiangru Huang, Bo Sun, Chandrajit L. Bajaj, and Qixing Huang. Gencorres: Consistent shape matching via coupled implicit-explicit shape generative models. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net, 2024. URL https://openreview.net/forum?id=dGH4kHFKFj.

Brent Yi, Vickie Ye, Maya Zheng, Lea Müller, Georgios Pavlakos, Yi Ma, Jitendra Malik, and Angjoo Kanazawa. Estimating body and hand motion in an ego-sensed world. *CoRR*, abs/2410.03665, 2024. doi: 10.48550/ARXIV.2410.03665. URL https://doi.org/10.48550/arXiv.2410.03665.

Biao Zhang, Jiapeng Tang, Matthias Nießner, and Peter Wonka. 3dshape2vecset: A 3d shape representation for neural fields and generative diffusion models. *ACM Trans. Graph.*, 42(4), July 2023. ISSN 0730-0301. doi: 10.1145/3592442. URL https://doi.org/10.1145/3592442.

Siwei Zhang, Bharat Lal Bhatnagar, Yuanlu Xu, Alexander Winkler, Petr Kadlecek, Siyu Tang, and Federica Bogo. Rohm: Robust human motion reconstruction via diffusion. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16-22, 2024*, pp. 14606–14617. IEEE, 2024. doi: 10.1109/CVPR52733.2024.01384. URL https://doi.org/10.1109/CVPR52733.2024.01384.

Keyang Zhou, Bharat Lal Bhatnagar, and Gerard Pons-Moll. Unsupervised shape and pose disentanglement for 3d meshes. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (eds.), *ECCV (22)*, volume 12367 of *Lecture Notes in Computer Science*, pp. 341–357. Springer, 2020a. ISBN 978-3-030-58542-6. URL http://dblp.uni-trier.de/db/conf/eccv/eccv2020-22.html#ZhouBP20.

Yi Zhou, Chenglei Wu, Zimo Li, Chen Cao, Yuting Ye, Jason M. Saragih, Hao Li, and Yaser Sheikh. Fully convolutional mesh autoencoder using efficient spatially varying kernels. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020b. URL https://proceedings.neurips.cc/paper/2020/hash/68dd09b9ff11f0df5624a690fe0f6729-Abstract.html.

Silvia Zuffi, Angjoo Kanazawa, David Jacobs, and Michael J. Black. 3D menagerie: Modeling the 3D shape and pose of animals. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

A NORMALIZED ARAP REG

Our goal is to compute the derivative of

$$r^\theta(\boldsymbol{z}) = \mathrm{Tr}\big(E^\theta(z)^{-\frac{1}{2}} H^\theta(\boldsymbol{z}) E^\theta(z)^{-\frac{1}{2}}\big) = \mathrm{Tr}\big(H^\theta(\boldsymbol{z}) E^\theta(z)^{-1}\big)$$

where

$$E^\theta(\boldsymbol{z}) = J^\theta(\boldsymbol{z})^T J^\theta(\boldsymbol{z}), \quad H^\theta(\boldsymbol{z}) = J^\theta(\boldsymbol{z})^T \overline{H}^\theta(\boldsymbol{z}) J^\theta(\boldsymbol{z})$$

where

$$J^\theta(\boldsymbol{z}) = \frac{\partial \boldsymbol{g}^\theta(\boldsymbol{z})}{\partial \boldsymbol{z}}.$$

Using the chain rule, we have

$$\begin{aligned}\frac{\partial r^\theta(\boldsymbol{z})}{\partial \theta} &= \mathrm{Tr}\big(E^\theta(z)^{-\frac{1}{2}} \frac{\partial H^\theta(\boldsymbol{z})}{\partial \theta} E^\theta(z)^{-\frac{1}{2}}\big) \\ &\quad - \mathrm{Tr}\big(H^\theta(\boldsymbol{z})^{\frac{1}{2}} E^\theta(z)^{-1} \frac{\partial E^\theta(\boldsymbol{z})}{\partial \theta} E^\theta(z)^{-1} H^\theta(\boldsymbol{z})^{\frac{1}{2}}\big)\end{aligned} \tag{7}$$

Let eigenvalue decomposition of E^θ and singular value decomposition of $H^\theta(\boldsymbol{z})^{\frac{1}{2}} E^\theta(z)^{-1}$ be

$$E^\theta(\boldsymbol{z}) = U^\theta(\boldsymbol{z}) \Lambda^\theta(\boldsymbol{z}) U^\theta(\boldsymbol{z})^T$$

and

$$H^\theta(\boldsymbol{z})^{\frac{1}{2}} E^\theta(z)^{-1} = V^\theta(\boldsymbol{z}) \Sigma^\theta(\boldsymbol{z}) W^\theta(\boldsymbol{z})^T.$$

We have

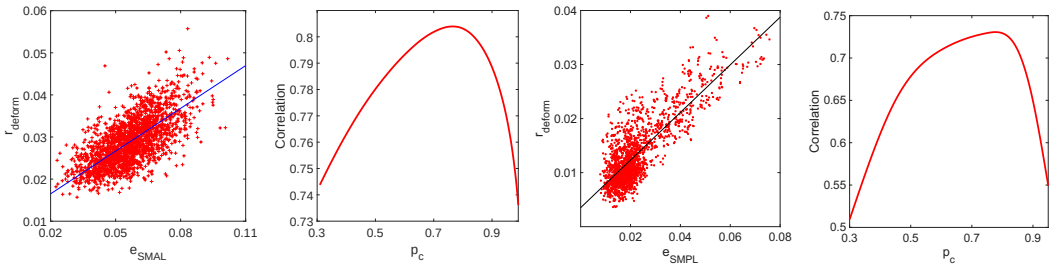
$$\begin{aligned}&\mathrm{Tr}\big(E^\theta(z)^{-\frac{1}{2}} \frac{\partial H^\theta(\boldsymbol{z})}{\partial \theta} E^\theta(z)^{-\frac{1}{2}}\big) \\ =&\frac{2}{\epsilon^2} \sum_{i=1}^{d} \frac{1}{\lambda_i^\theta(\boldsymbol{z})} \big(\frac{\partial \boldsymbol{g}^\theta(\boldsymbol{z} + \epsilon \boldsymbol{u}_i^\theta(\boldsymbol{z}))}{\partial \theta} - \frac{\partial \boldsymbol{g}^\theta(\boldsymbol{z})}{\partial \theta}\big)^T \overline{H}^\theta(\boldsymbol{z}) \\ &\big(\boldsymbol{g}^\theta(\boldsymbol{z} + \epsilon \boldsymbol{u}_i^\theta(\boldsymbol{z})) - \boldsymbol{g}^\theta(\boldsymbol{z})\big) \\ +&\frac{1}{\epsilon^2} \sum_{i=1}^{d} \frac{1}{\lambda_i^\theta(\boldsymbol{z})} \big(\boldsymbol{g}^\theta(\boldsymbol{z} + \epsilon \boldsymbol{u}_i^\theta(\boldsymbol{z})) - \boldsymbol{g}^\theta(\boldsymbol{z})\big)^T \frac{\partial \overline{H}^\theta(\boldsymbol{z})}{\partial \theta} \\ &\big(\boldsymbol{g}^\theta(\boldsymbol{z} + \epsilon \boldsymbol{u}_i^\theta(\boldsymbol{z})) - \boldsymbol{g}^\theta(\boldsymbol{z})\big)\end{aligned}$$

Similarly, we have

$$\begin{aligned}&\mathrm{Tr}\big(H^\theta(\boldsymbol{z})^{\frac{1}{2}} E^\theta(z)^{-1} \frac{\partial E^\theta(\boldsymbol{z})}{\partial \theta} E^\theta(z)^{-1} H^\theta(\boldsymbol{z})^{\frac{1}{2}}\big) \\ =&\frac{2}{\epsilon^2} \sum_{i=1}^{d} \sigma_i^\theta(\boldsymbol{z})^2 \big(\frac{\partial \boldsymbol{g}^\theta(\boldsymbol{z} + \epsilon \boldsymbol{w}_i^\theta(\boldsymbol{z}))}{\partial \theta} - \frac{\partial \boldsymbol{g}^\theta(\boldsymbol{z})}{\partial \theta}\big)^T \\ &\big(\boldsymbol{g}^\theta(\boldsymbol{z} + \epsilon \boldsymbol{w}_i^\theta(\boldsymbol{z})) - \boldsymbol{g}^\theta(\boldsymbol{z})\big).\end{aligned}$$

B VALIDATION OF THE DEFORMATION SCORE

We further validate the deformation score r_{def} used for geometry-to-distribution feedback in the main paper. For mesh generators, we examine its correlation with SMPL/SMAL fitting error and its sensitivity to the percentile parameter p_c. For implicit generators, we perform the corresponding analysis using the deformation score defined through GenCorres-based correspondence.



(a) (b) (c) (d)

Figure 6: Validation of the deformation score r_{def}. Panels (a,b) show the relationship between r_{def} and SMPL/SMAL fitting error and its sensitivity to p_c for mesh generators; panels (c,d) show the corresponding behavior for implicit generators. The correlation remains stable over a broad range of p_c.

Across both mesh and implicit generators, r_{def} remains positively associated with the corresponding SMPL/SMAL fitting error over a broad range of p_c, supporting its use as a family-relative geometric quality signal rather than relying on a narrowly tuned percentile choice.

C ADDITIONAL DEFORMATION-SCORE VARIANTS

For completeness, we describe additional deformation-score variants considered in our study. All results reported in the main paper use the deformation score defined in Sec. 4.3; the variants below are not used to produce the main quantitative results. Deformation energy between two meshes of the same topology. Suppose the vertex vectors are $\boldsymbol{p} \in \mathbb{R}^{3n}$ and $\boldsymbol{q} \in \mathbb{R}^{3n}$ where the i-th vertices are given by $\boldsymbol{p}_i$ and $\boldsymbol{q}_i$. Let $\mathcal{N}(i)$ collect 1-ring neighbors of the i-th vertex. We measure the generalized deformation energy between them as

$$f(\boldsymbol{p}, \boldsymbol{q}) = \sum_{i=1}^{n} \Big(\lambda^{\text{rigid}} f_i^{\text{rigid}}(\boldsymbol{p}, \boldsymbol{q}) + \lambda^{\text{sim}} f_i^{\text{sim}}(\boldsymbol{p}, \boldsymbol{q}) + \lambda^{\text{aff}} f_i^{\text{aff}}(\boldsymbol{p}, \boldsymbol{q}) \Big) \tag{8}$$

Here

$$f_i^{\text{rigid}}(\boldsymbol{p}, \boldsymbol{q}) = \min_{R \in \text{SO}(3)} \sum_{j \in \mathcal{N}(i)} \|R(\boldsymbol{p}_i - \boldsymbol{p}_j) - (\boldsymbol{q}_i - \boldsymbol{q}_j)\|^2.$$

Denote $P_i = (\boldsymbol{p}_i - \boldsymbol{p}_j) \in \mathbb{R}^{3 \times |\mathcal{N}(i)|}$ and $Q_i = (\boldsymbol{q}_i - \boldsymbol{q}_j) \in \mathbb{R}^{3 \times |\mathcal{N}(i)|}$. Then

$$\begin{aligned} f_i^{\text{rigid}}(\boldsymbol{p}, \boldsymbol{q}) &= \min_{R \in \text{SO}(3)} \|RP_i - Q_i\|_{\mathcal{F}}^2 \\ &= \|P_i\|_{\mathcal{F}}^2 + \|Q_i\|_{\mathcal{F}}^2 - 2 \max_{R \in \text{SO}(3)} \langle RP_i, Q_i \rangle \\ &= \|P_i\|_{\mathcal{F}}^2 + \|Q_i\|_{\mathcal{F}}^2 - 2 \max_{R \in \text{SO}(3)} \text{Trace}(RP_i Q_i^T). \end{aligned} \tag{9}$$

Denote $S = P_i Q_i^T$. Let its singular-value decomposition (SVD) be $S = U \Sigma V^T$ where the singular values in the diagonal of Σ are ordered in the decreasing manner. The optimal R that maximizes $\text{Trace}(RS)$ is given by $R = V \text{diag}(1, 1, \text{sign}(\det(S))) U^T$. Note that $\text{sign}(\det(S))$ enforces that $R \in \text{SO}(3)$.

Similarly,

$$f_i^{\text{sim}}(\boldsymbol{p}, \boldsymbol{q}) = \min_{s > 0, R \in \text{SO}(3)} \|sRP_i - Q_i\|_{\mathcal{F}}^2.$$

It is easy to see that the optimal rotation can be derived from SVD of $S = P_i Q_i^T$ in the same way as above. Given the optimal $R^\star$, the optimal scale is

$$s^\star = \frac{\langle R^\star P_i, Q_i \rangle}{\|P_i\|_{\mathcal{F}}^2}.$$

Finally,

$$f_i^{\text{aff}}(\boldsymbol{p}, \boldsymbol{q}) = \min_A \|AP_i - Q_i\|_{\mathcal{F}}^2$$

and the optimal

$$A^\star = (Q_i P_i^T)(P_i P_i^T)^{-1}.$$

- For each synthetic shape, obtain top-K closest training shapes using L^2-distance between vertex positions.
- For each closest shape, calculate the deformation energy using $f(\boldsymbol{p}, \boldsymbol{q})$ with hyper-parameters λ^{sim} and λ^{aff}.
- Sort the distances and use the k-th smallest value as the score.
- There are three critical hyper-parameters above, i.e., λ^{sim}, λ^{aff}, and k. K is less sensitive.

D Details in Preliminaries

D.1 Details of ARAPReg

ARAPReg Huang et al. (2021) introduces as-rigid-as-possible (ARAP) deformation geometry into a mesh generator $\boldsymbol{m}^\phi : \mathcal{Z} \to \mathbb{R}^{3n}$ with fixed connectivity. Its objective combines data fitting with geometric regularization at latent codes sampled from a fixed Gaussian:

$$\min_\phi l_{\text{data}}(\mathcal{S}, \boldsymbol{m}^\phi) + \lambda \mathop{\mathbb{E}}_{\boldsymbol{z} \sim \mathcal{N}_d(\boldsymbol{0}, I_d)} l_{\text{reg}}(\boldsymbol{m}^\phi(\boldsymbol{z})). \tag{10}$$

The corresponding latent-space deformation metric is obtained by pulling back the mesh-space ARAP metric,

$$H(\boldsymbol{m}^\phi(\boldsymbol{z})) = \frac{\partial \boldsymbol{m}^\phi(\boldsymbol{z})}{\partial \boldsymbol{z}}^T \overline{H}(\boldsymbol{m}^\phi(\boldsymbol{z})) \frac{\partial \boldsymbol{m}^\phi(\boldsymbol{z})}{\partial \boldsymbol{z}}.$$

We next describe the construction of $\overline{H}$.

The sparse matrix $\overline{H}(\boldsymbol{m}^\phi(\boldsymbol{z}))$ is based on an ARAP deformation model that measures the deformation between a mesh whose vertices are given by $\boldsymbol{p} \in \mathbb{R}^{3n}$ and the same mesh topology whose vertices are given by $\boldsymbol{p} + \boldsymbol{d} \in \mathbb{R}^{3n}$. Let $\mathcal{N}(i)$ collect the adjacent vertices of i. The deformation model is given by

$$e(\boldsymbol{p}, \boldsymbol{d}) = \sum_{i=1}^{n} \min_{\boldsymbol{c}_i} \sum_{j \in \mathcal{N}(i)} \|\boldsymbol{c}_i \times (\boldsymbol{p}_i - \boldsymbol{p}_j) - (\boldsymbol{d}_i - \boldsymbol{d}_j)\|^2.$$

Introduce $L \in \mathbb{R}^{n \times n}$, $B(\boldsymbol{p}) \in \mathbb{R}^{3n \times 3n}$, and $C(\boldsymbol{p}) \in \mathbb{R}^{3n \times 3n}$:

$$L_{ij} = \begin{cases} |\mathcal{N}(i)| & i = j \\ -1 & j \in \mathcal{N}(i) \\ 0 & \text{otherwise} \end{cases} \qquad B_{ij} = \begin{cases} \sum_{j \in \mathcal{N}(i)} (\boldsymbol{p}_i - \boldsymbol{p}_j)\times & i = j \\ -(\boldsymbol{p}_i - \boldsymbol{p}_j)\times & j \in \mathcal{N}(i) \\ 0 & \text{otherwise} \end{cases}$$

$$C_{ij} = \begin{cases} \sum_{j \in \mathcal{N}(i)} ((\boldsymbol{p}_i - \boldsymbol{p}_j)\times)((\boldsymbol{p}_i - \boldsymbol{p}_j)\times)^T & i = j \\ 0 & \text{otherwise} \end{cases}$$

Let $\boldsymbol{c} \in \mathbb{R}^{3n}$ collect $\boldsymbol{c}_i$ in its corresponding blocks. Since the neighborhood relation is symmetric, the directed sum over i and $j \in \mathcal{N}(i)$ counts each mesh edge twice. With the definitions above, the deformation energy can therefore be written as

$$\begin{aligned} e(\boldsymbol{p}, \boldsymbol{d}) &= \min_{\boldsymbol{c}} \Big(2\boldsymbol{d}^T (L \otimes I_3)\boldsymbol{d} - 2\boldsymbol{d}^T B(\boldsymbol{p})^T \boldsymbol{c} + \boldsymbol{c}^T C(\boldsymbol{p}) \boldsymbol{c} \Big) \\ &= \boldsymbol{d}^T \Big(2L \otimes I_3 - B(\boldsymbol{p})^T C(\boldsymbol{p})^{-1} B(\boldsymbol{p}) \Big) \boldsymbol{d}. \end{aligned}$$

Hence, the mesh-space ARAP metric is

$$\overline{H}(\boldsymbol{m}^{\phi}(\boldsymbol{z})) = 2L \otimes I_3 - B(\boldsymbol{m}^{\phi}(\boldsymbol{z}))^T C(\boldsymbol{m}^{\phi}(\boldsymbol{z}))^{-1} B(\boldsymbol{m}^{\phi}(\boldsymbol{z})).$$

By construction, $\overline{H}$ is positive semidefinite and vanishes on infinitesimal rigid motions, consistent with the original squared-residual ARAP energy.

D.2 Details of GenCorres

Consider a point $\boldsymbol{m}_i^{\phi}$ on the implicit surface $g^{\phi}(\boldsymbol{x}, \boldsymbol{z}) = 0$. Let $\epsilon \boldsymbol{d}$ be its perturbation so that $\boldsymbol{m}_i^{\phi} + \epsilon \boldsymbol{d}$ lies on $g^{\phi}(\boldsymbol{x}, \boldsymbol{z} + \epsilon \boldsymbol{v}) = 0$. Applying the first-order Taylor expansion of $g^{\phi}(\boldsymbol{x}, \boldsymbol{z})$, we have

$$\frac{\partial g^{\phi}}{\partial \boldsymbol{x}}(\boldsymbol{m}_i^{\phi}(\boldsymbol{z}), \boldsymbol{z}) \cdot \boldsymbol{d} + \frac{\partial g^{\phi}}{\partial \boldsymbol{z}}(\boldsymbol{m}_i^{\phi}(\boldsymbol{z}), \boldsymbol{z}) \boldsymbol{v} = 0. \tag{11}$$

Stacking Eq. (11) over all mesh vertices yields the linear system used by GenCorres to compute the local deformation induced by the latent displacement $\boldsymbol{v}$.

E Details of Experiments

E.1 Hyper Parameters of ARAPDiffusion

The initialization step follows the standard protocol of ARAPReg and GenCorres to train the AE, which go through 200 epochs, where each epoch goes through a training shape once. The latent diffusion model is trained with 200 epochs as well.

We perform 10 alternations between generation-aware geometric refinement of the encoder–decoder and geometry-aware diffusion refinement. See Fig. 7 for illustration of convergence. In each alternation, we train the encoder–decoder for 50 epochs and fine-tune the diffusion model for 100 epochs.

When training the AE with regularization, each epoch picks a subset of 12800 samples for training. When training the latent diffusion model with regularization, we use $4N$ samples p_{sync} to train $\mathcal{D}^{\theta}$ in each epoch.

E.2 Parametric Fitting of SMPL and SMAL Models

For simplicity, we only describe the procedure of fitting the SMPL model described as $\boldsymbol{p}(\theta, \beta)$ to a 3D shape, where θ and β are the pose and shape parameters, respectively. The same fitting procedure, including initialization and parameter regularization, is used for all compared methods; we therefore use e_t as a common fitting-based comparison metric rather than an exact projection distance to the parametric shape manifold.

E.2.1 Mesh Fitting

Suppose S has the same mesh as the SMPL model where mesh vertices are given by $\boldsymbol{p}^0$. For reconstruction, our goal is to solve for θ and β to optimize

$$\min_{\theta,\beta} \|p^0 - p(\theta, \beta)\|_2^2 + \lambda \left(\|\theta - \theta_m\|_2^2 + \|\beta - \beta_m\|_2^2 \right) \tag{12}$$

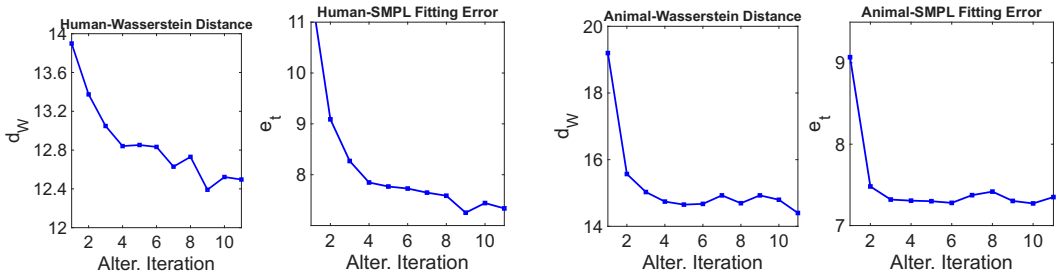
where θ_m and β_m are the codes of the closest training shape to p^0. The value of λ is set so that the fitting term and the regularization term have the same values. Optimization employs the Gauss-Newton method, starting from θ_m and β_m.

E.2.2 Implicit Surface Fitting

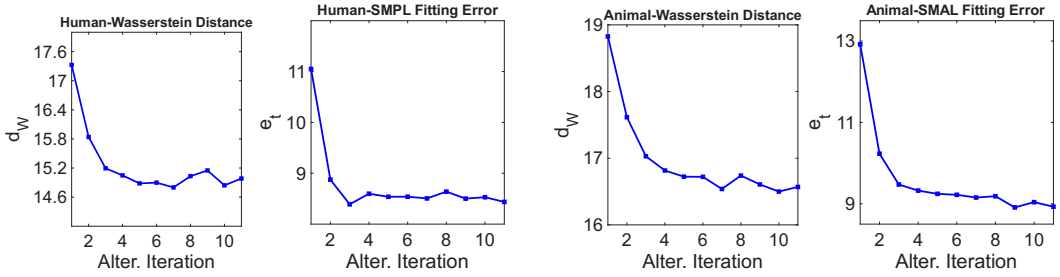
For an implicit surface defined by $g(\boldsymbol{p})$, we solve for

$$\min_{\theta,\beta} \sum_{i=1}^{n} g^2(p_i(\theta, \beta)) + \lambda \left(\|\theta - \theta_m\|_2^2 + \|\beta - \beta_m\|_2^2 \right) \tag{13}$$

where θ_m and β_m are the codes of the closest training shape to the mesh discretization of $g(\boldsymbol{p})$ with respect to the Chamfer distance. We again set value of λ so that the fitting term and the regularization term have the same values. Optimization employs the Gauss-Newton method, starting from θ_m and β_m.



(a) Human (Explicit Mesh) (b) Animal (Explicit Mesh)



(c) Human (Implicit Surface) (d) Animal (Implicit Surface)

Figure 7: Convergence of alternating geometry–distribution feedback under different settings: (a) and (b) show the explicit mesh generator for Human and Animal; (c) and (d) show the implicit surface generator for Human and Animal.

E.3 Running Time

All experiments use a single NVIDIA H200 GPU (140GB). For explicit SMAL/SMPL models, latent-diffusion initialization takes approximately 5/10 hours and alternating feedback takes 3/6 hours, respectively. For implicit models, the corresponding times are 20/50 hours and 10/20 hours.

F Convergence of Alternating Feedback Optimization

We examine convergence using the Wasserstein distance d_W and SMPL/SMAL fitting error e_t.

Across all four settings in Fig. 7, d_W and e_t improve rapidly in the first two alternations and then stabilize within a narrow range.

G Additional Experiments

G.1 Performance on Bone Datasets

We additionally evaluate on the Bone dataset introduced in ARAPReg, which contains 40 shapes per category. ARAPDiffusion achieves the lowest e_r, e_t, and d_W among the compared methods, while d_n remains comparable to the baselines.

Bone	e_r	e_t	d_W	d_n
SP-disent.	5.34	8.01	12.9	6.86
COMA	4.14	7.09	12.1	6.78
3DMM	4.03	7.12	11.9	6.72
MeshConv	4.47	7.34	12.1	6.69
ARAPReg	3.76	6.77	11.6	6.76
FrameAVE	3.92	6.81	11.9	6.91
GeoLatent	3.67	6.37	11.5	6.83
BRESA	3.61	6.45	11.7	6.91
ARAPDiffusion	**3.25**	**5.92**	**10.4**	6.79

G.2 Performance When Scaling the Data

The quantitative results on Human are shown below, comparing ARAPDiffusion with the standard latent-diffusion baseline (NoReg) as the amount of training data varies:

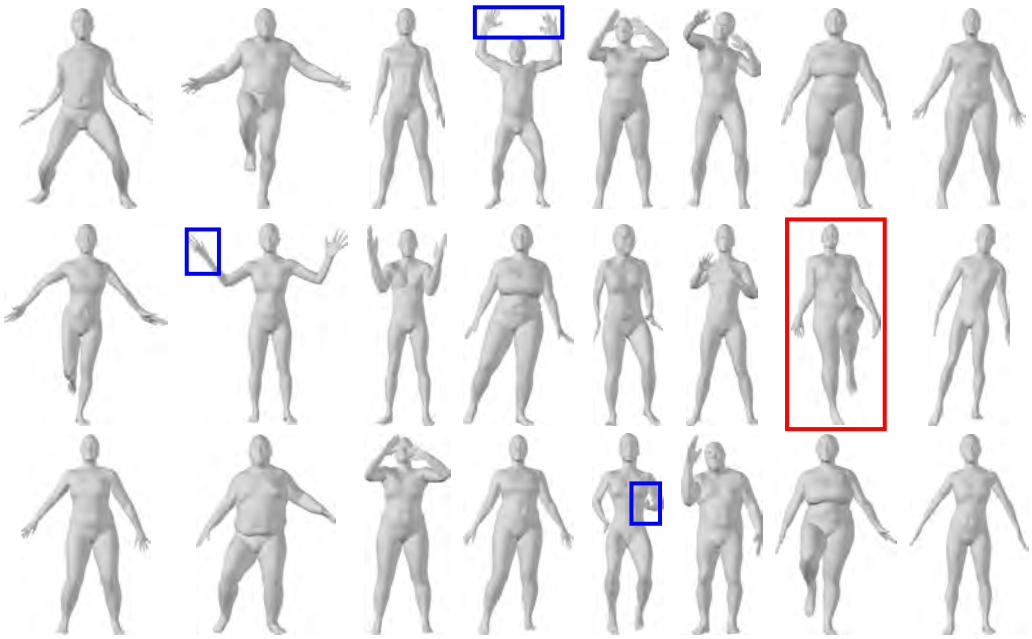


Figure 8: ARAPReg results. Red boxes indicate global failures; blue boxes indicate local artifacts.

Training data (%)	Method	e_t	d_W
50	Standard LD (No Reg)	14.20	19.50
50	ARAPDiffusion	10.85	15.20
100	Standard LD (No Reg)	10.60	15.10
100	ARAPDiffusion	7.21	12.40
200	Standard LD (No Reg)	6.95	12.10
200	ARAPDiffusion	5.8	10.20

Note: 100% data corresponds to the original setting in Table 1 of our main paper. These results suggest that the geometric refinement improves data efficiency: ARAPDiffusion with reduced training data approaches the performance of the unregularized model trained with substantially more data. Furthermore, the refinement remains beneficial as the training set grows, with ARAPDiffusion continuing to outperform the unregularized baseline at 200% data.

H Samples of Random Generated Shapes

H.1 Human

Fig. 8, Fig. 9, Fig. 10, and Fig. 11 show randomly generated shapes of the baseline approaches ARAPReg, FrameAVE, Geolatent, and BRESA, respectively. All of them share the same issue that a subset of the synthetic shapes have low-quality overall shapes. Moreover, their synthetic shapes show a variety of noticeable distortions.

In contrast, the random samples in Fig. 12 show substantially fewer visible distortions and more consistently plausible geometry. The distortion of each generate shape is also visually smaller.

Fig. 13 shows for each generated shape (on the left) its closest training shape (on the right). We can see that the generated shapes do not replicate the training shapes.

H.2 Animal

Fig. 14, Fig. 15, Fig. 16, and Fig. 17 show randomly generated shapes of the baseline approaches ARAPReg, FrameAVE, Geolatent, and BRESA, respectively. All of them share the same issue that a subset of the synthetic shapes have low-quality overall shapes. Moreover, their synthetic shapes show a variety of noticeable distortions.

In contrast, as shown in Fig. 18, ARAPDiffusion does not have low-quality generated shapes. The distortion of each generate shape is also visually smaller.

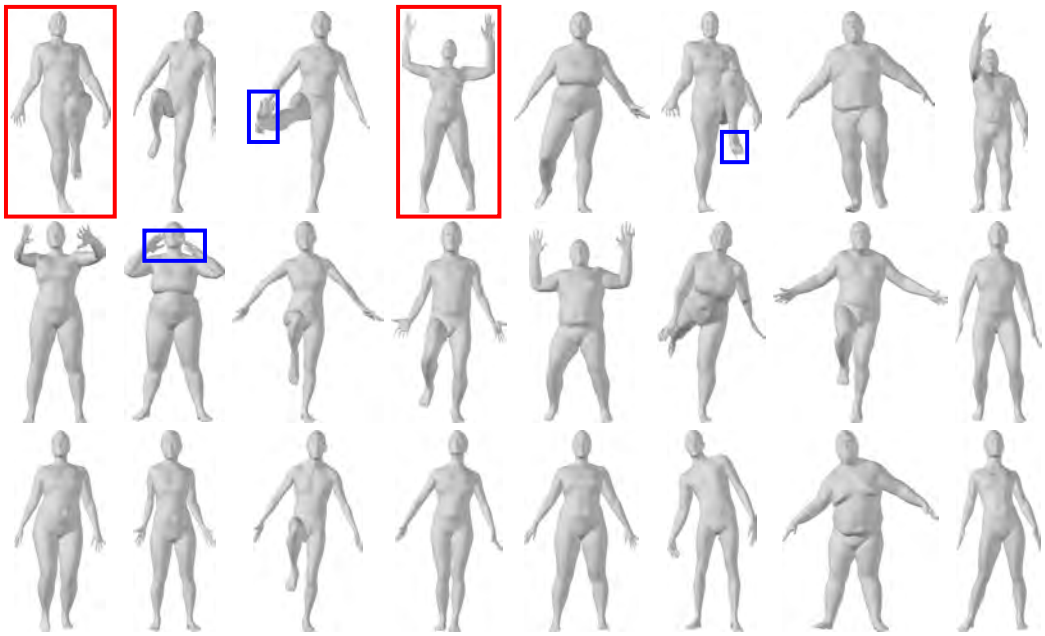


Figure 9: FrameAVE results. They show improved individual shape quality and higher success rate.

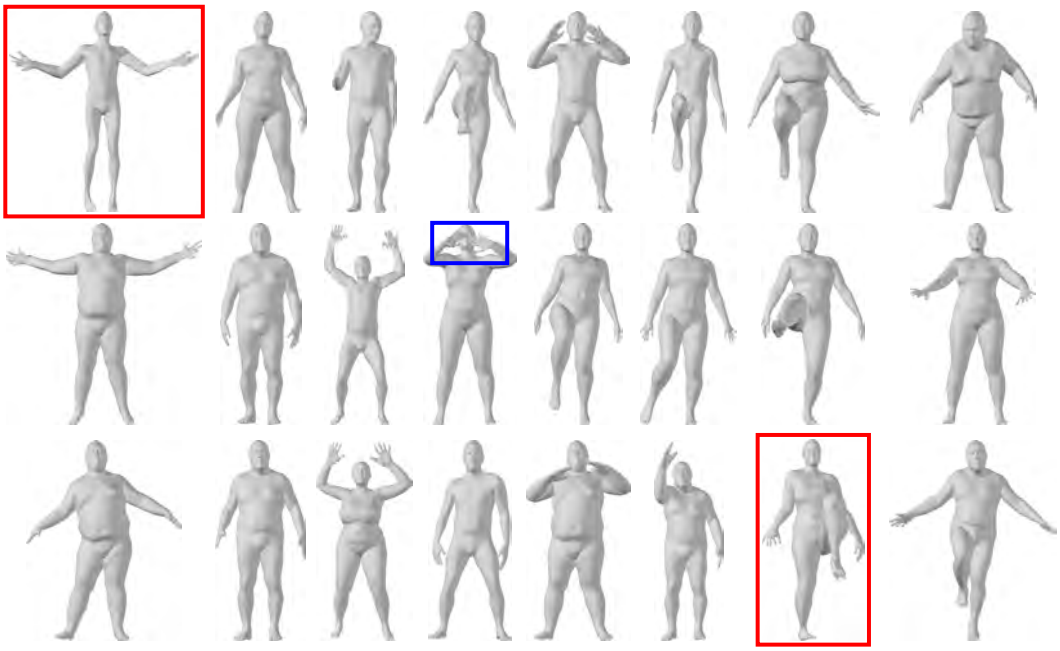


Figure 10: GeoLatent results. They show improved individual shape quality and higher success rate.

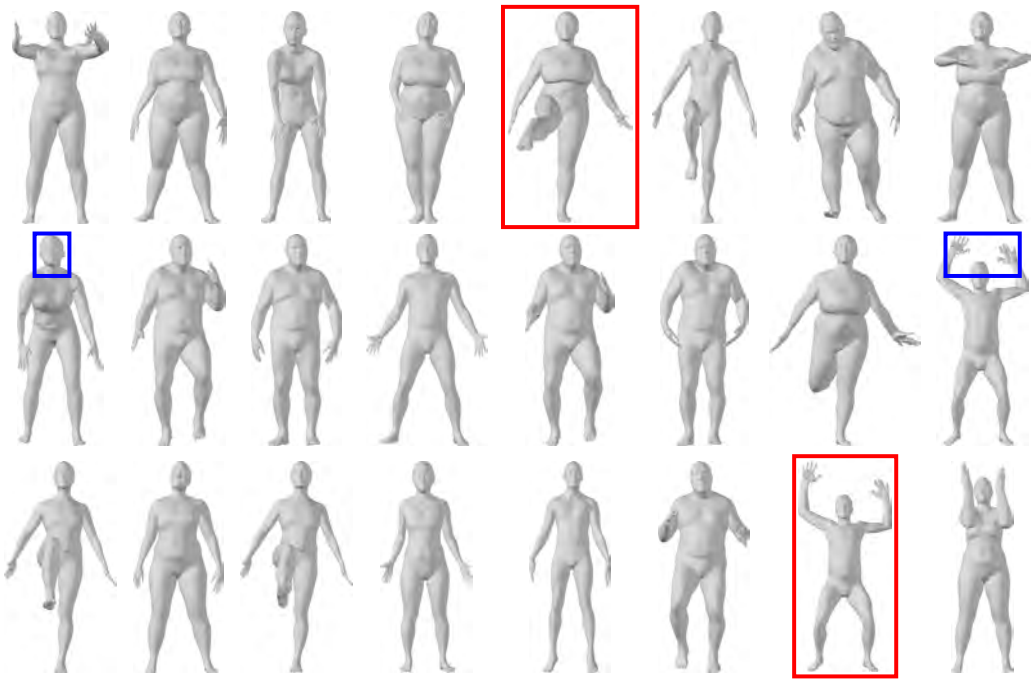


Figure 11: BRESA results. They show improved individual shape quality and higher success rate.

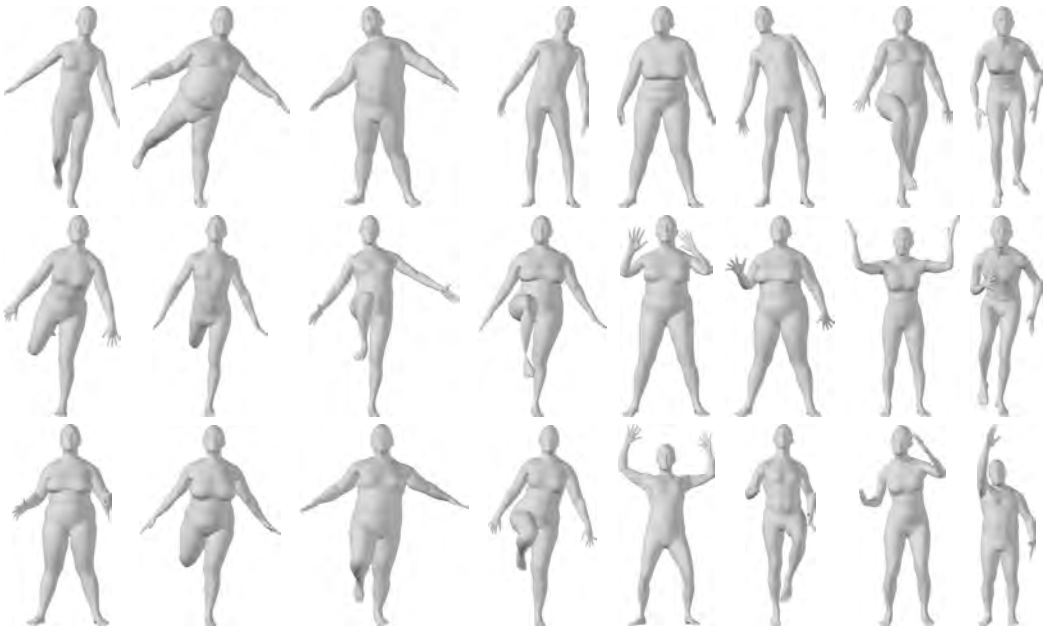


Figure 12: ARAPDiffusion results. They show improved individual shape quality and higher success rate.

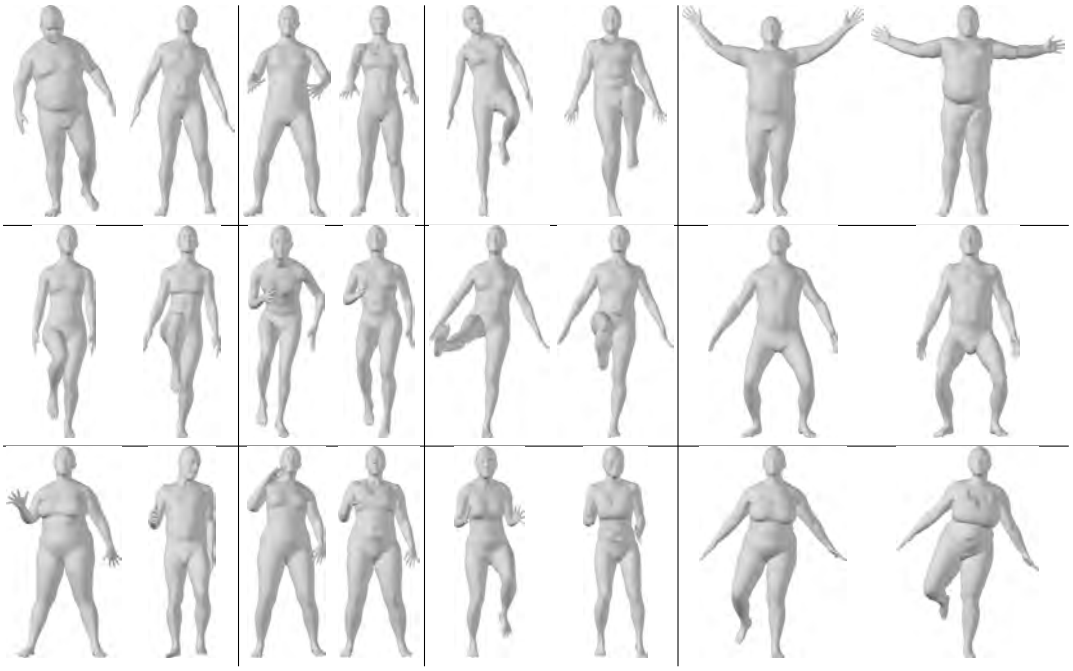


Figure 13: The closest training shape of each generated shape. In each cell, the generated shape is shown on the left, while the closest training shape is shown on the right.

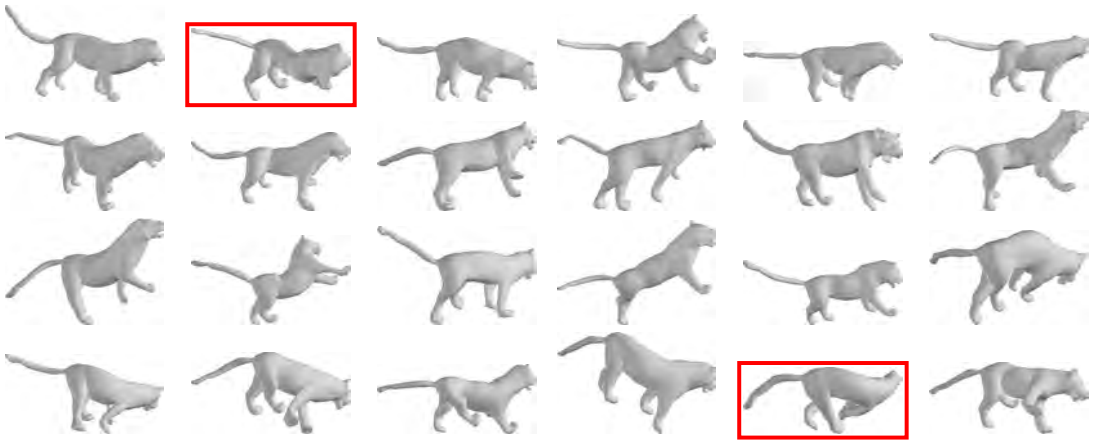


Figure 14: ARAPReg results. Similar to other baseline approaches, some of the generated shapes have significantly distorted geometries. Most of shapes have noticeable local distortions.

Fig. 19 shows for each generated shape (on the left) its closest training shape (on the right). We can see that the generated shapes do not replicate the training shapes.

I Additional Conditional Generation Results

Fig. 20 and Fig. 21 show qualitative conditional generation results. Compared to baseline approaches, our approach leads to results that with significantly improved alignments with the ground-truth. Even if our results do not align with the ground-truth due to limited observations, our results are still plausible.

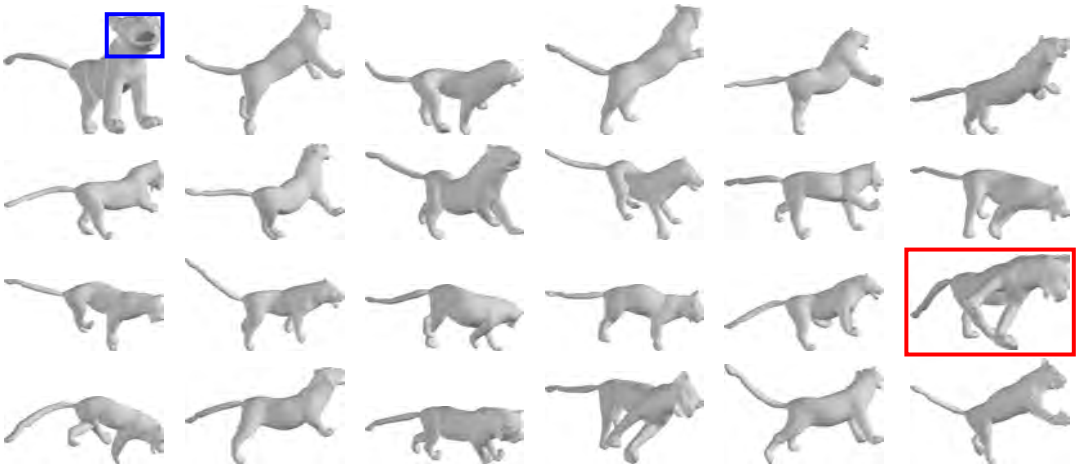


Figure 15: FrameAVE results. Similar to other baseline approaches, some of the generated shapes have significantly distorted geometries. Most of shapes have noticeable local distortions.

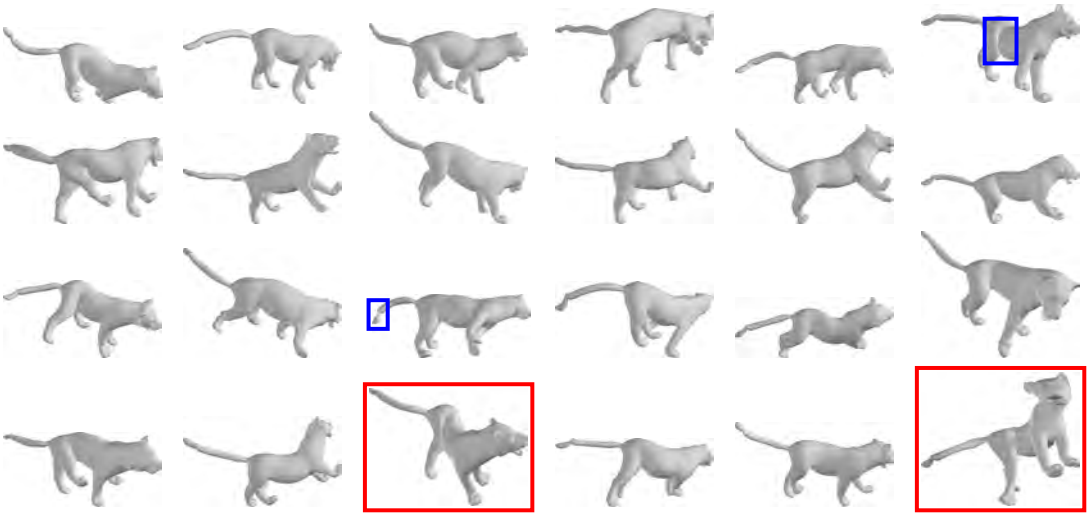


Figure 16: GeoLatent results. Similar to other baseline approaches, some of the generated shapes have significantly distorted geometries. Most of shapes have noticeable local distortions.

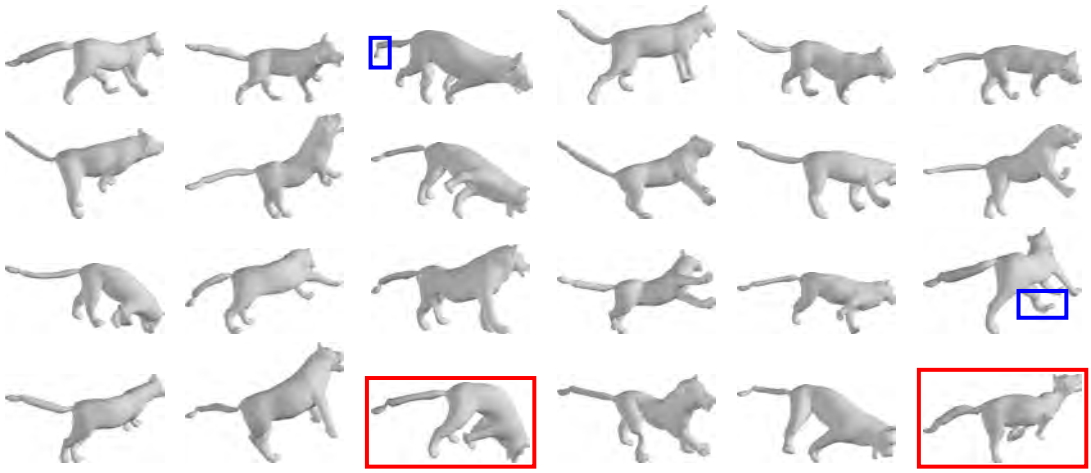


Figure 17: BREAS results. Similar to other baseline approaches, some of the generated shapes have significantly distorted geometries. Most of shapes have noticeable local distortions.

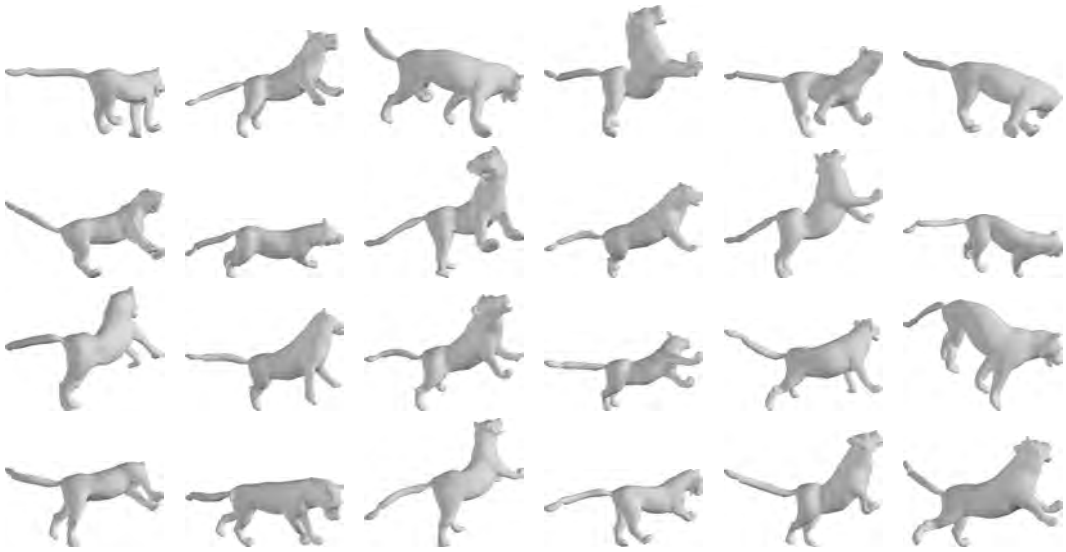


Figure 18: ARAPDiffusion results. They show improved individual shape quality and higher success rate.

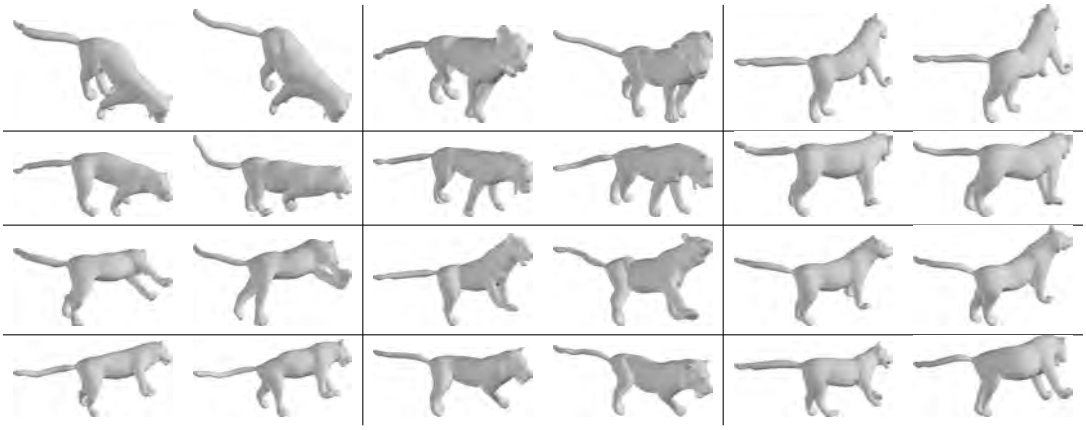


Figure 19: The closest training shape of each generated shape. In each cell, the generated shape is shown on the left, while the closest training shape is shown on the right.

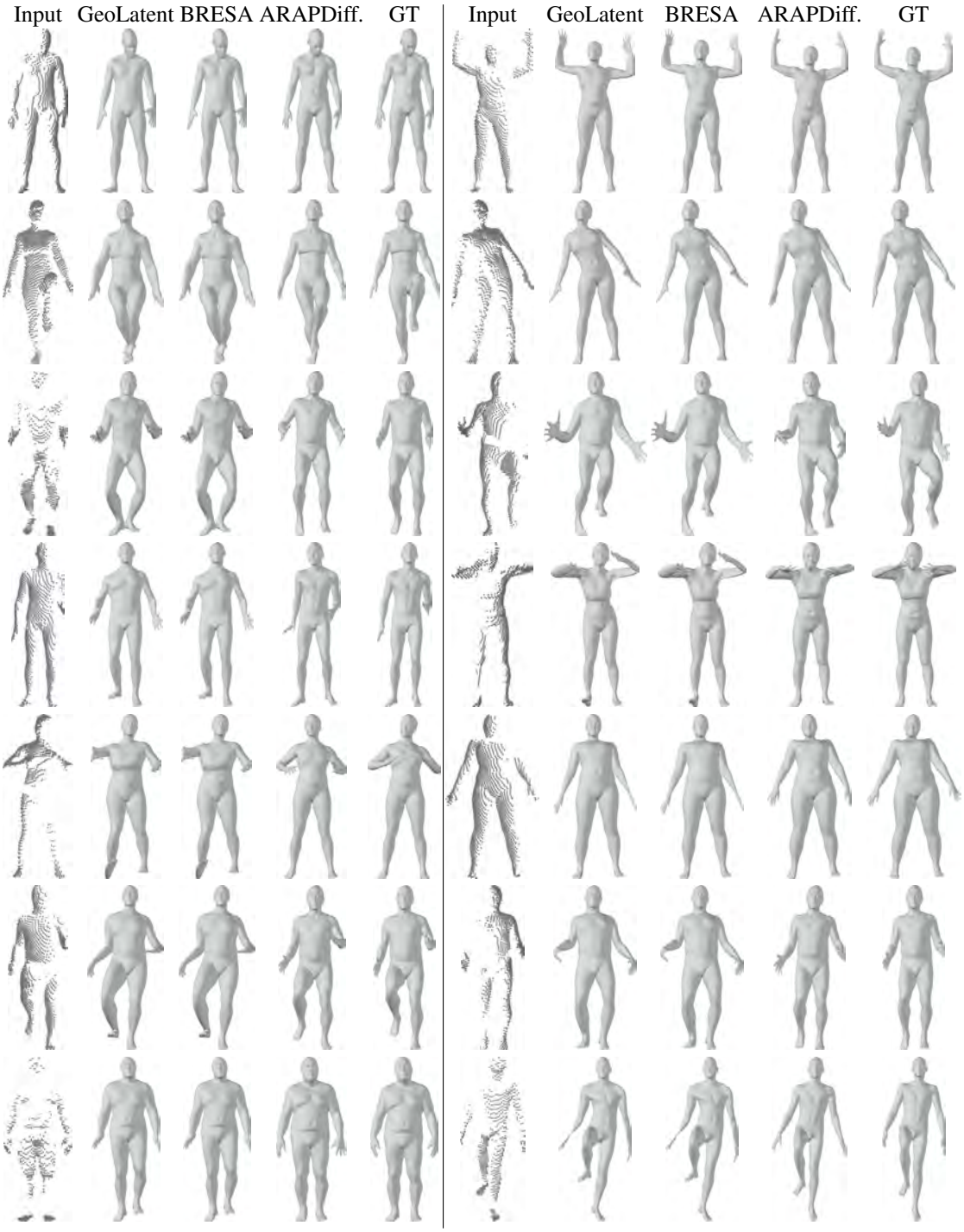


Figure 20: Human conditional generation results. Our results align with the ground-truth better than baseline approaches.



Figure 21: Animal conditional generation results. Our results align with the ground-truth better than baseline approaches.