ACM SIGGRAPH / Eurographics Symposium on Computer Animation (2026)
R. Sugimoto (Editors)



# Latent Space Skinning: Learning Compact Representations for Mesh Animations

G. Drongoulas[1], G. Tsopouridis[1], A. Aristidou[2] and I. Fudos[1]

[1] Department of Computer Science and Engineering, University of Ioannina, Greece,
[2] Department of Computer Science, University of Cyprus & CYENS Centre of Excellence, Cyprus

**Abstract**

*Latent-Space Skinning (LSS) is a neural character animation framework that replaces explicit skinning weights with a shared latent representation of skeletal motion and rest-pose geometry, which is decoded into per-vertex mesh deformations over time. This compact and expressive latent space enables high compression ratios, accurate reconstruction, and intuitive motion manipulation operations such as blending and cross-character transfer.*

**CCS Concepts**

• ***Computing methodologies*** → ***Animation;*** *Machine learning;* ***Procedural animation;***

## 1. Introduction

In this work, we revisit 3D character animation by reformulating the classical skinning pipeline in latent space. Traditional approaches depend on hand-crafted rigs, skeletal animation, and skinning operators such as Linear Blend Skinning (LBS), which, despite their efficiency, introduce artifacts and require manual setup. More advanced deformation methods reduce these artifacts at the cost of higher complexity and continued dependence on mesh topology. Recent learning-based techniques and latent motion representations show that animation data is highly redundant and compressible, yet they typically remain tied to specific rigs or topologies and treat motion, shape, and deformation as separate problems. Consequently, compression, synthesis, and transfer are often addressed independently rather than within a unified framework.

We introduce *Latent-Space Skinning (LSS)*, a neural framework that learns a compact, motion-conditioned latent representation of mesh deformation. Instead of explicitly predicting skinning weights, LSS encodes skeletal motion and rest-pose geometry into a shared latent space and decodes directly to per-vertex deformations over time, maintaining compatibility with skeleton-based pipelines while reducing reliance on explicit skinning. This design exploits temporal redundancy for high compression, corrects common skinning artifacts, and naturally supports interpolation, blending, and motion transfer as a proof-of-concept. Positioned within modern motion modeling (e.g., BiMotion [WYC*26], 4D generative models, deep skinning prediction [MTF23], and latent-space transfer such as SMF [MDM25]), our method provides a simple, topology-dependent baseline that unifies deformation, reconstruction, and motion reuse in a single latent formulation.



## 2. Latent Space Skinning

*Latent-Space Skinning (LSS)* is a neural framework that models mesh animation as a learned mapping from skeletal motion to vertex deformations in a compact latent space. In contrast to classical skinning methods, which explicitly compute vertex positions from bone transformations and skinning weights, LSS directly learns the relationship between skeletal motion, rest-pose geometry, and the resulting animated mesh. Consequently, traditional rigging, skinning weights, and deformation modules are replaced by a latent representation and a decoder that reconstructs the animation sequence. This formulation enables the network to capture both skeletal motion and complex non-linear deformation effects within a unified representation.

Modeling mesh animation requires capturing both temporal motion dynamics and shape-dependent deformations, as character motion evolves continuously over time while remaining strongly influenced by morphology. To address these challenges, we employ an encoder-decoder sequence architecture based on LSTMs with latent-space fusion. An overview of the proposed LSS architecture is shown in Figure 1. The source code is available at: https://github.com/gdrongoulas/LatentSpaceSkinning.

This formulation captures both skeletal motion and non-linear deformation effects within a unified latent representation. The resulting latent space enables compact motion encoding for animation compression, latent-space interpolation for animation synthesis, and the transfer of motion across characters through the recombination of motion and shape representations.

The network is trained to reconstruct the ground-truth animated mesh sequence $\mathbf{V}1^{gt}$ from the rest-pose mesh $\mathbf{V}^{r}$ and the skele-

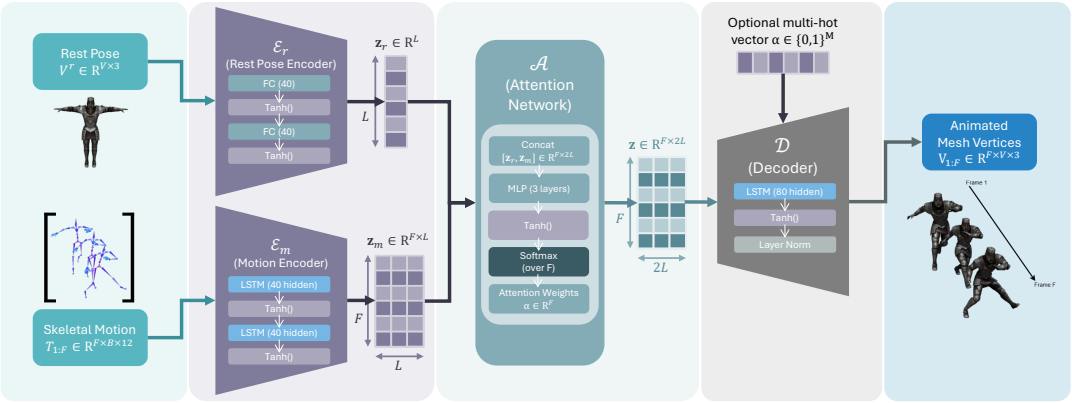


**Figure 1:** ***LSS architecture.*** *The motion sequence* $\mathbf{T}_{1:F}$ *and rest mesh* $\mathbf{V}^r$ *are encoded into latent representations* $\mathbf{z}_m$ *and* $\mathbf{z}_r$*. These are fused via* $\mathcal{A}$ *to produce* $\mathbf{z}$*, which is decoded by* $\mathcal{D}$ *into vertex trajectories* $\mathbf{V}_{1:F}$*. Here the latent dimension* $L = 40$*,* $F$ *is the number of frames,* $B$ *is the number of bones, and* $V$ *is the number of vertices.*

tal motion sequence $\mathbf{T}1$. Optimization is performed using AdamW with a learning rate of $10^{-3}$ and weight decay of $10^{-4}$ for 6,000 epochs on a single GPU. Models are trained independently for each character using between 7 and 20 animation sequences.

**Loss Function** The final training objective combines reconstruction accuracy with geometric, volumetric, and temporal regularization terms. Specifically, $\mathcal{L}$G preserves local surface curvature, $\mathcal{L}$vol enforces segment volume preservation, and $\mathcal{L}$C promotes temporal smoothness by minimizing per-vertex acceleration:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{MSE}} + \lambda_2 \mathcal{L}_{\text{C}} + \lambda_3 \mathcal{L}_{\text{G}} + \lambda_4 \mathcal{L}_{\text{vol}} \tag{1}$$

where $\lambda_1 = 0.7$ and $\lambda_2 = \lambda_3 = \lambda_4 = 0.1$ are empirically chosen weights.

Together, these terms balance reconstruction accuracy, surface detail preservation, volume consistency, and motion smoothness, producing animations that are both geometrically faithful and visually plausible.

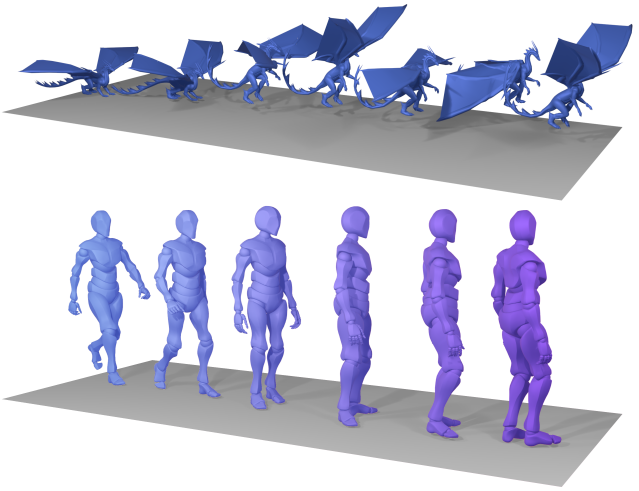

**Figure 2:** ***Animation reconstruction and blending.*** *Top: Dragon animation reconstructed by our model, preserving motion with high fidelity. Bottom: Walk and turn motion bending, generating a smooth transition from walking (light blue) to turning (purple).*

## 3. Results & Evaluation

We evaluate LSS along three dimensions: (i) reconstruction and compression, (ii) motion blending, and (iii) motion transfer. All models are trained per character using 7-13 motion sequences, with additional sequences reserved for evaluation. Unless otherwise stated, all quantitative results are reported on unseen motions.

Representative reconstruction and blending examples are shown in Figure 2, with additional results provided in the supplementary video. Qualitative transfer results are also included in the supplementary material. LSS achieves high-fidelity reconstruction while providing substantially higher compression rates than prior methods, improving compression performance by approximately 15%–40% across the evaluated datasets.

## References

[MDM25] MURALIKRISHNAN S., DUTT N. S., MITRA N. J.: Smf: Template-free and rig-free animation transfer using kinetic codes. *ACM Trans. Graph. 44*, 6 (Dec. 2025). 1

[MTF23] MOUTAFIDOU A., TOULATZIS V., FUDOS I.: Deep fusible skinning of animation sequences. *Vis. Comput. 40*, 8 (Nov. 2023), 5695–5715. 1

[WYC*26] WANG M., YAN Q., CAO Z., LI Y., MAC AODHA O., CORSO J. J., VAXMAN A.: Bimotion: B-spline motion for text-guided dynamic 3d character generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2026), CVPR'26. 1