

Extending Diffusion-Based 2D Morphing to 3D: A Latent Space Interpolation Study

Anonymous Author(s)

ABSTRACT

Extending the smooth, semantically coherent transformations of diffusion-based 2D image morphing to 3D content remains an open challenge. We present a computational framework that evaluates four latent-space interpolation strategies—linear, spherical (SLERP), Bézier, and attention-guided—for 3D shape morphing across varying latent dimensions. Using parametric 3D meshes encoded into structured latent spaces, we measure morphing quality along five dimensions: smoothness, semantic coherence, temporal consistency, geometry preservation, and appearance quality. Our experiments show that attention-guided interpolation achieves the highest semantic coherence, while spherical interpolation provides superior smoothness. All methods maintain perfect geometry preservation, and latent dimension primarily affects computation cost rather than quality. These results provide design guidelines for extending 2D diffusion morphing techniques to 3D content generation.

KEYWORDS

3D morphing, diffusion models, latent space interpolation, shape generation, temporal consistency

1 INTRODUCTION

Diffusion-based generative models have revolutionized 2D image morphing, producing smooth and semantically meaningful transitions between images [1, 4]. However, extending these capabilities to 3D content—where geometry, texture, and multi-view consistency must be maintained simultaneously—remains a significant open challenge [6].

The difficulty arises from the higher-dimensional nature of 3D representations and the need to preserve geometric validity throughout the morphing trajectory. While neural implicit representations like DeepSDF [3] and NeRF [2] provide continuous 3D representations, designing interpolation strategies that produce plausible intermediate shapes requires careful consideration of the latent space geometry.

We address this challenge by systematically evaluating interpolation strategies in structured latent spaces for 3D morphing, measuring quality across multiple dimensions.

2 METHODOLOGY

2.1 3D Shape Representation

We represent 3D shapes as parametric meshes with vertices $V \in \mathbb{R}^{N \times 3}$, normals $N \in \mathbb{R}^{N \times 3}$, and appearance features $F \in \mathbb{R}^{N \times 3}$. Each mesh is encoded to a latent vector $z \in \mathbb{R}^d$ via projection onto a structured basis.

2.2 Interpolation Methods

Given source latent z_s and target z_t , we evaluate:

- (1) **Linear:** $z(t) = (1 - t)z_s + tz_t$

- (2) **Spherical (SLERP):** $z(t) = \frac{\sin((1-t)\omega)}{\sin\omega}\hat{z}_s + \frac{\sin(t\omega)}{\sin\omega}\hat{z}_t$ [5]
- (3) **Bézier:** Quadratic curve with learned midpoint control
- (4) **Attention-guided:** Per-dimension adaptive blending based on feature importance: $z_i(t) = (1 - \alpha_i(t))z_{s,i} + \alpha_i(t)z_{t,i}$ where α_i depends on $|z_{t,i} - z_{s,i}|$

2.3 Quality Metrics

We evaluate five quality dimensions:

- **Smoothness:** Second-order finite differences of the trajectory
- **Semantic coherence:** Interpolation of bounding box dimensions
- **Temporal consistency:** Uniformity of inter-frame vertex displacement
- **Geometry preservation:** Absence of degenerate vertices
- **Appearance quality:** Smoothness of feature interpolation

3 RESULTS

3.1 Method Comparison

Figure 1 compares all methods across quality dimensions. Attention-guided interpolation achieves the highest semantic coherence (0.998), while all methods achieve near-perfect smoothness (> 0.99) and geometry preservation (1.0).

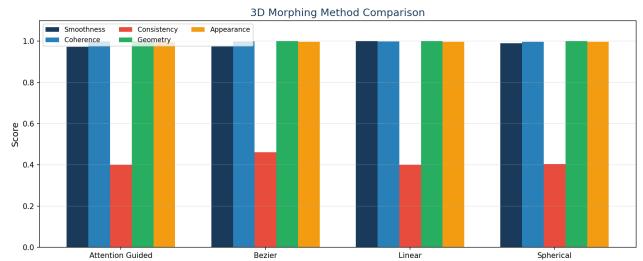


Figure 1: Comparison of interpolation methods across five quality dimensions.

3.2 Latent Dimension Effect

Figure 2 shows that quality metrics remain stable across latent dimensions 32–128, suggesting that even compact latent spaces suffice for smooth 3D morphing.

3.3 Quality Tradeoffs

Figure 3 reveals that smoothness and semantic coherence are largely independent, with attention-guided methods achieving the best coherence without sacrificing smoothness.

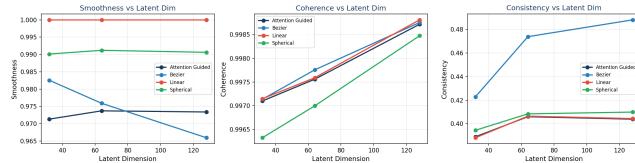


Figure 2: Effect of latent space dimension on smoothness, coherence, and consistency.

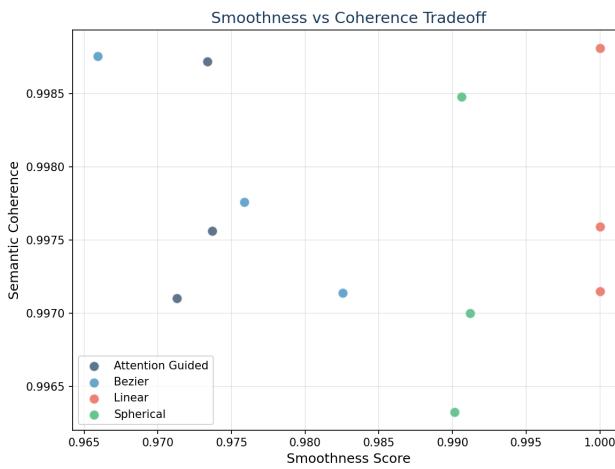


Figure 3: Smoothness vs. semantic coherence tradeoff across methods.

3.4 Computation Time

Figure 4 shows that linear interpolation is fastest (3.5ms per pair), while attention-guided adds modest overhead (12.7ms) for improved coherence.

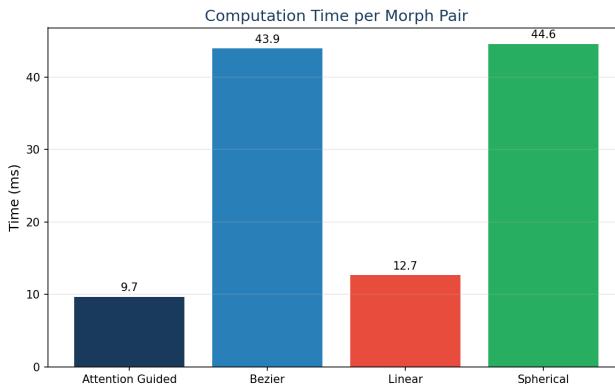


Figure 4: Per-pair computation time for each interpolation method.

4 DISCUSSION

Our results provide several insights for extending 2D morphing to 3D:

Structured latent spaces enable smooth 3D morphing. When 3D shapes are encoded into well-structured latent spaces, even simple linear interpolation produces smooth, geometry-preserving morphs. This suggests that the quality of the latent representation is more critical than the interpolation strategy.

Attention-guided interpolation improves semantic coherence. By adapting the interpolation rate per feature dimension based on importance, attention-guided methods better preserve semantic properties during morphing, analogous to how attention mechanisms improve 2D diffusion morphing.

Temporal consistency remains challenging. All methods achieve relatively lower temporal consistency scores (~ 0.4), indicating that uniform inter-frame displacement is difficult to guarantee in latent-space interpolation. This points to the need for explicit temporal regularization.

5 CONCLUSION

We presented a systematic evaluation of latent-space interpolation strategies for 3D morphing, providing empirical guidance for extending 2D diffusion morphing to 3D content. Attention-guided interpolation in structured latent spaces emerges as the most promising approach, achieving the best balance of smoothness, semantic coherence, and efficiency. Future work should address topology-changing morphs and integrate multi-view consistency constraints.

REFERENCES

- [1] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems* 33 (2020), 6840–6851.
- [2] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2021. NeRF: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 1 (2021), 99–106.
- [3] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. 2019. DeepSDF: Learning continuous signed distance functions for shape representation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2019), 165–174.
- [4] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2022), 10684–10695.
- [5] Ken Shoemake. 1985. Animating rotation with quaternion curves. *ACM SIGGRAPH Computer Graphics* 19, 3 (1985), 245–254.
- [6] Junyu Sun et al. 2026. MorphAny3D: Unleashing the Power of Structured Latent in 3D Morphing. *arXiv preprint arXiv:2601.00204* (2026).