Nonisotropic Gaussian Diffusion for Realistic 3D Human Motion Prediction

Cecilia Curreli, Dominik Muhle, Abhishek Saroha, Zhenzhang Ye, Riccardo Marin, Daniel Cremers

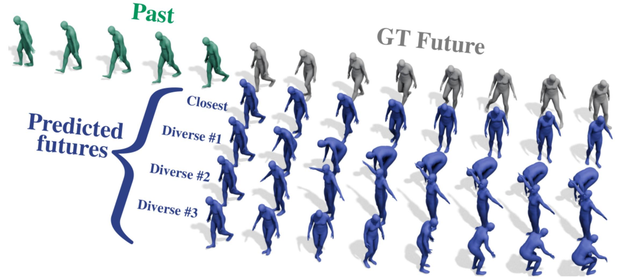
Munich Center for Machine Learning · Technical University of Munich

CVPR Nashville JUNE 11-15, 2025

Code available! GitHub

Task: Human Motion Prediction

For a given past motion, generate multiple diverse futures



Limitations

FAIL Baselines generates *diverse* but not ***realistic*** futures

FAIL Conventional Denoising Diffusion Probabilistic Models [3][4] (**isotropic** Gaussian diffusion) disregard joint relationships.

Forward noise transitions $\mathcal{N}(\boldsymbol{x}_t; \sqrt{\alpha_t}\boldsymbol{x}_{t-1}, \boldsymbol{\Sigma}_t)$

Isotropic covariance $\boldsymbol{\Sigma}_t = (1-\alpha_t)\mathbb{I}$

$$\boldsymbol{x}_t = \sqrt{\alpha_t}\boldsymbol{x}_{t-1} + (1-\alpha_t)\epsilon \qquad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbb{I})$$

with $\boldsymbol{x}_0$ the true variable and $\boldsymbol{x}_T$ pure noise

Citations

[1] Yuan et al. *Dlow: Diversifying latent flows for diverse human motion prediction*. ECCV20
[2] Sun et al. *Towards consistent stochastic human motion prediction via motion diffusion*. ECCV24
[3] Ho et al. *Denoising diffusion probabilistic models*. NeurIPS20
[4] Rombach et al. *High-Resolution Image Synthesis with Latent Diffusion Models*. CVPR22

Contributions

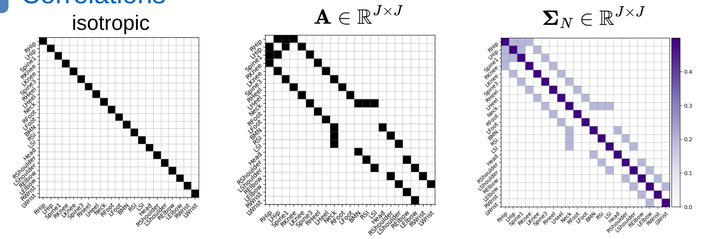
- The first **Nonisotropic Gaussian Diffusion** training and sampling formulation for a structured problem
- **SkeletonDiffusion** achieves **SOTA** performance on several datasets, also in challenge scenarios (zero-shot & noisy data)

Our formulation is general and applicable to other problems!

Nonisotropic Gaussian Diffusion

- A **training** and **sampling** formulation that lifts the i.i.d. assumption
- The noise reflects **correlations** between the J human body joints

1 Correlations

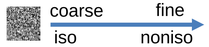


$$\boldsymbol{\Sigma}_N = \frac{\mathbf{A} - \lambda_{\min}(\mathbf{A})\mathbb{I}}{\lambda_{\max}(\mathbf{A}) - \lambda_{\min}(\mathbf{A})}$$

✓ Positive definite
✓ Normalized eigenvalues

2 Noise Blending

$$\boldsymbol{\Sigma}_t = (1-\alpha_t)(\gamma_t\boldsymbol{\Sigma}_N + (1-\gamma_t)\mathbb{I})$$



3 Forward and Reverse Equations

$$\boldsymbol{x}_t = \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \boldsymbol{U}\bar{\boldsymbol{\Lambda}}_t^{1/2}\boldsymbol{\epsilon}$$

$$\boldsymbol{x}_{t-1} = \boldsymbol{\mu}_q + \boldsymbol{U}\boldsymbol{\Lambda}_q\boldsymbol{\epsilon}$$

$$\boldsymbol{\Sigma}_t = \boldsymbol{U}\boldsymbol{\Lambda}_t\boldsymbol{U}^\top$$

Find the details in the paper!

4 Training Objective

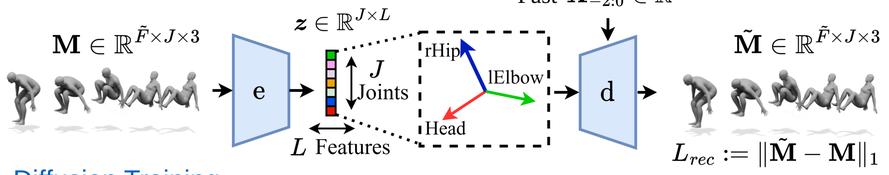
$$L_{diff}(\boldsymbol{x_\theta}, \boldsymbol{x}_0) = \bar{\alpha}_t\|\bar{\boldsymbol{\Lambda}}_t^{-1/2}\boldsymbol{U}^T(\boldsymbol{x_\theta} - \boldsymbol{x}_0)\|^2$$

SkeletonDiffusion

SkeletonDiffusion is a **latent diffusion model** implementing nonisotropic diffusion with a **graph attention architecture** explicitly considering joint types and connectivity

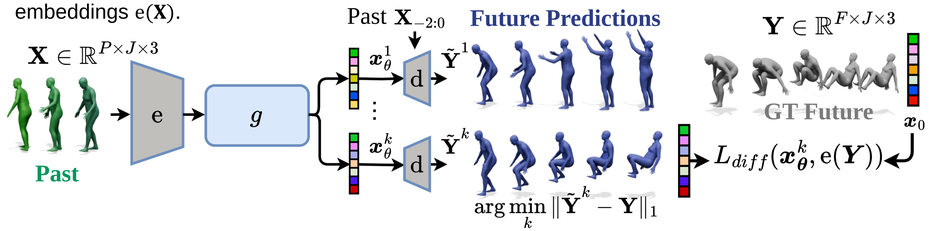
Latent Space Training

Train an autoencoder to reconstruct motions $\boldsymbol{M}$ of random length $\tilde{F}$ via curricular learning to learn a temporally-compressed latent space $\boldsymbol{z}$



Diffusion Training

Denoise $\boldsymbol{x}_t \in \mathbb{R}^{J\times L}$ into latent embeddings $\boldsymbol{z} = \mathrm{e}(\mathbf{Y})$ of the GT future $\mathbf{Y}$ conditioning on past embeddings $\mathrm{e}(\mathbf{X})$.



To increase diversity, relax the loss by denoising k=50 samples. The feature dimension L is diffused isotropically.

Results

Demo online!

- Demonstrate overlooked issues: unrealistic motions may 'cheat' diversity
- Explicit inductive bias in the architecture improves body realism
- **Nonisotropic diffusion improves performance**: precision, diversity and limb stretching

Performance · Parameters · Convergence · Footprint

Method	Precision	Diversity	Body Realism			
			mean ↓		RMSE ↓	
	ADE ↓	APD ↑	str	jit	str	jit
Zero-Velocity	0.755	0.000	0.00	0.00	0.00	0.00
DLow	0.590	13.170	8.41	0.40	11.06	0.58
DivSamp	0.564	**24.724**	11.17	0.82	16.71	1.0
CoMusion	0.494	10.848	4.04	0.25	5.63	0.52
Ours (w/o Graph-Att)	0.502	8.021	3.90	0.20	5.31	0.27
Ours (isotropic)	0.499	8.788	3.72	0.18	4.93	0.24
SkeletonDiffusion	**0.480**	9.456	**3.15**	**0.20**	**4.45**	**0.26**