

Diverse Human Motion Prediction via Gumbel-Softmax Sampling from an Auxiliary Space

Lingwei Dang¹, Yongwei Nie^{1*}, Chengjiang Long², Qing Zhang³, Guiqing Li¹

¹ South China University of Technology, ² Meta Reality Lab, ³ Sun Yat-sen University

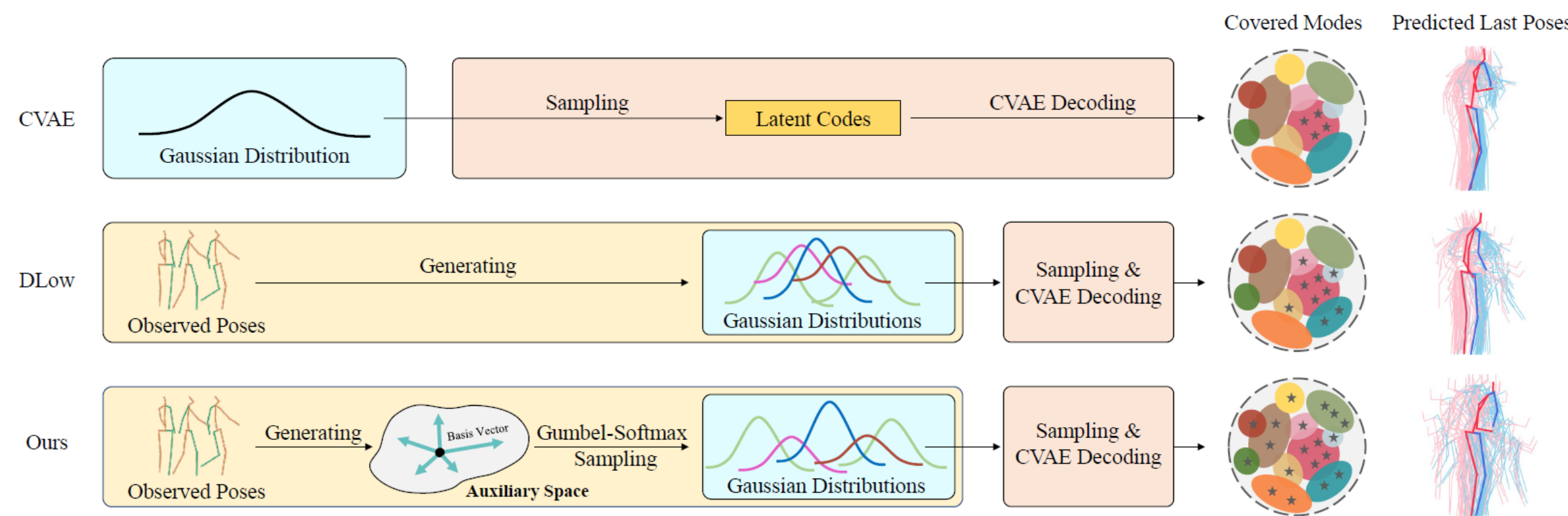
csdanglw@mail.scut.edu.cn, {nieyongwei, ligq}@scut.edu.cn, clong1@fb.com, zhangqing.whu.cs@gmail.com



ACM
Multimedia 2022
Lisbon, Portugal | 10-14 October

Background & Motivation

- Diverse human motion prediction aims at predicting **multiple** possible future sequences from the observed one.
- Poses obtained by previous simple deep generative networks are **not diverse enough**.
- Recent work attempted to model the conditional distribution of data but it can only cover **a fixed number of modes**.
- We propose a novel **sampling strategy** for sampling very diverse results from an imbalanced **multimodal distribution**.



Contributions

- A novel sampling method converts the sampling of the distribution into randomly sampling of points from an **auxiliary space** for **diverse and accurate** sampling.
- A **Gumbel-Softmax** sampling method and a **hinge-diversity loss** both improve the performance of our method.
- Extensive experimental results demonstrate the **effectiveness** of our approach.

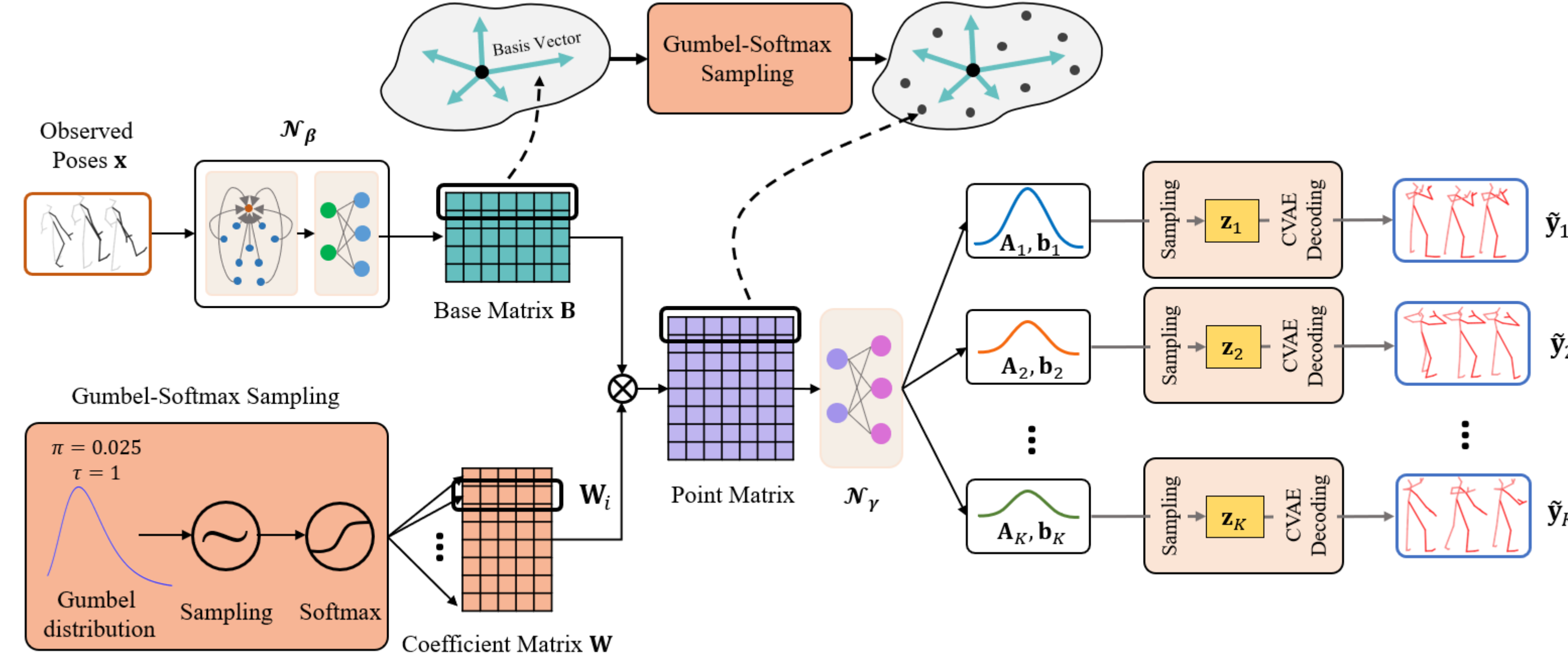
Datasets & Metric & Project

- Datasets: Human3.6M, HumanEva-I Dataset
- Metrics: APD, ADE, FDE, MMAD, MMFDE
- Link: https://github.com/Droliven/diverse_sampling

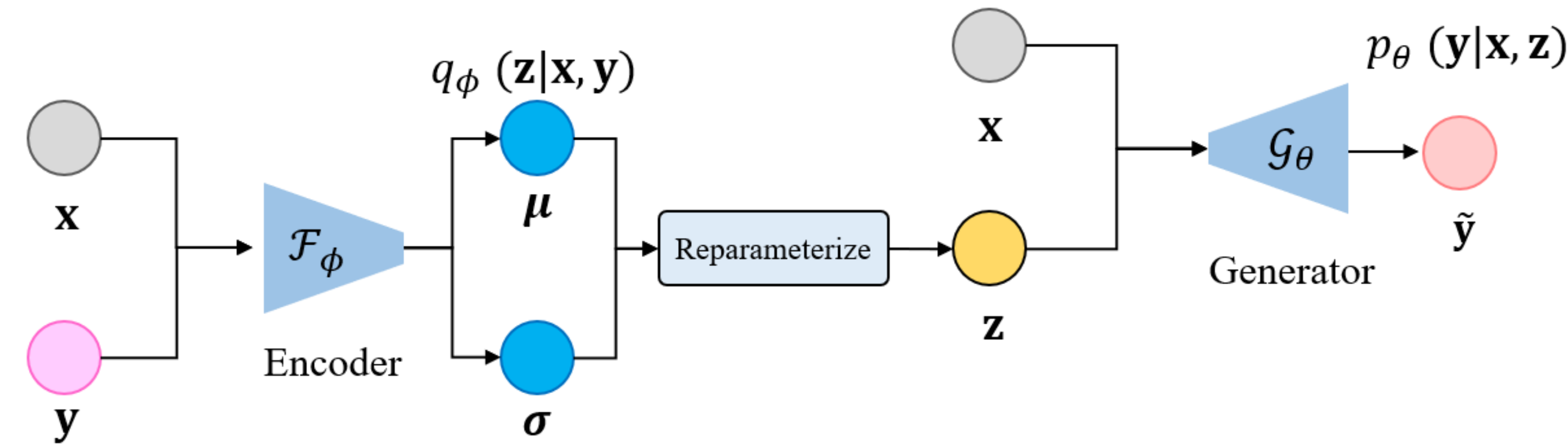


Proposed Approach

Approach Overview



Architecture of the pretrained CVAE



Pseudocode for the overall approach and the Gumbel-Softmax sampling strategy

Algorithm 1 Diverse sampling from a complex distribution by random Gumbel-Softmax sampling from an auxiliary space
Input: Observed pose sequence x , number of samples K , auxiliary space generation network N_β , Gaussian distribution generation network N_γ , CVAE decoder network G_θ that models the target complex distribution
Output: A set of samples $\{\tilde{y}_k\}_{k=1}^K$
1: $B = N_\beta(x)$ // generate an auxiliary space given input poses
2: $W \leftarrow$ Gumbel-Softmax sampling // generate a coefficient matrix by Algorithm 2
3: $P = WB$ // Multiply W and B to obtain a point matrix P
4: $\{A_k, b_k\}_{k=1}^K = N_\gamma(P)$ // convert points into means and variances
5: $\epsilon \sim N(0, 1)$ // sampling an ϵ from the normal distribution
6: **for** $k = 1$ to K **do**
7: $z_k = A_k \epsilon + b_k$ // sample z_k from Gaussian distribution of variance A_k and mean b_k by reparameterization trick
8: $\tilde{y}_k = G_\theta(x, z_k)$ // decode z_k and x into a result \tilde{y}_k
9: **end for**

Loss functions

$$\mathcal{L}_{div} = \frac{1}{K(K-1)} \sum_{i=1}^K \sum_{j \neq i}^K \max(0, \eta - \|\tilde{y}_i - \tilde{y}_j\|_2) \quad \mathcal{L}_{acc} = \min_k \|\tilde{y} - \tilde{y}_k\|_2, k \in [1, K] \quad \mathcal{L}'_{KL} = \mathcal{KL}(r_{\beta, \gamma}(z_k | x) \| p(z)), k \in [1, K]$$

Algorithm 2 Gumbel-Softmax coefficient matrix generation

Input: Number of coefficient vectors K , dimension size M of a coefficient vector, Gumbel distribution parameters π and τ
Output: A coefficient matrix $W \in \mathbb{R}^{K \times M}$
1: Declare a matrix $W \in \mathbb{R}^{K \times M}$
2: **for** $i = 1$ to K **do**
3: **for** $j = 1$ to M **do**
4: $u \sim U(0, 1)$ // sample a value from uniform distribution
5: $g = -\log(-\log(u))$
6: $W_{ij} = \frac{\pi + g}{\tau}$
7: **end for**
8: $W_i = \text{Softmax}(W_i)$ // normalize the i^{th} row of W
9: **end for**

Quantitative Results

Comparisons with baselines

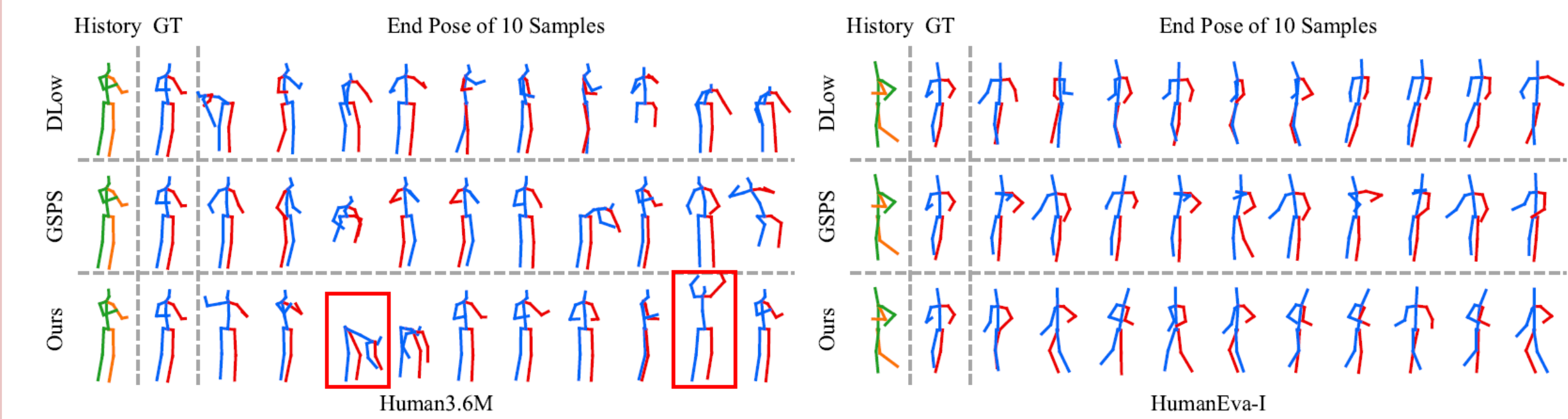
	Method	Human3.6M [18]					HumanEva-I [36]				
		APD \uparrow	ADE \downarrow	FDE \downarrow	MMAD \downarrow	MMFDE \downarrow	APD \uparrow	ADE \downarrow	FDE \downarrow	MMAD \downarrow	MMFDE \downarrow
deterministic	LTD [31]	0.000	0.516	0.756	0.627	0.795	0.000	0.415	0.555	0.509	0.613
	MSR [13]	0.000	0.508	0.742	0.621	0.791	0.000	0.371	0.493	0.472	0.548
	Pose-Knows [41]	6.723	0.461	0.560	0.522	0.569	2.308	0.269	0.296	0.384	0.375
	MT-VAE [42]	0.403	0.457	0.595	0.716	0.883	0.021	0.345	0.403	0.518	0.577
	HP-GAN [6]	7.214	0.858	0.867	0.847	0.858	1.139	0.772	0.749	0.776	0.769
stochastic	BoM [7]	6.265	0.448	0.533	0.514	0.544	2.846	0.271	0.279	0.373	0.351
	GMVAE [14]	6.769	0.461	0.555	0.524	0.566	2.443	0.305	0.345	0.408	0.410
	DeLiGAN [17]	6.509	0.483	0.534	0.520	0.545	2.177	0.306	0.322	0.385	0.371
	DSF [43]	9.330	0.493	0.592	0.550	0.599	4.538	0.273	0.290	0.364	0.340
	DLow [44]	11.741	0.425	0.518	0.495	0.531	4.855	0.251	0.268	0.362	0.339
	GSPS [30]	14.757	0.389	0.496	0.476	0.525	5.825	0.233	0.244	0.343	0.331
	Ours	15.310	0.370	0.485	0.475	0.516	6.109	0.220	0.234	0.342	0.316

Ablation study results

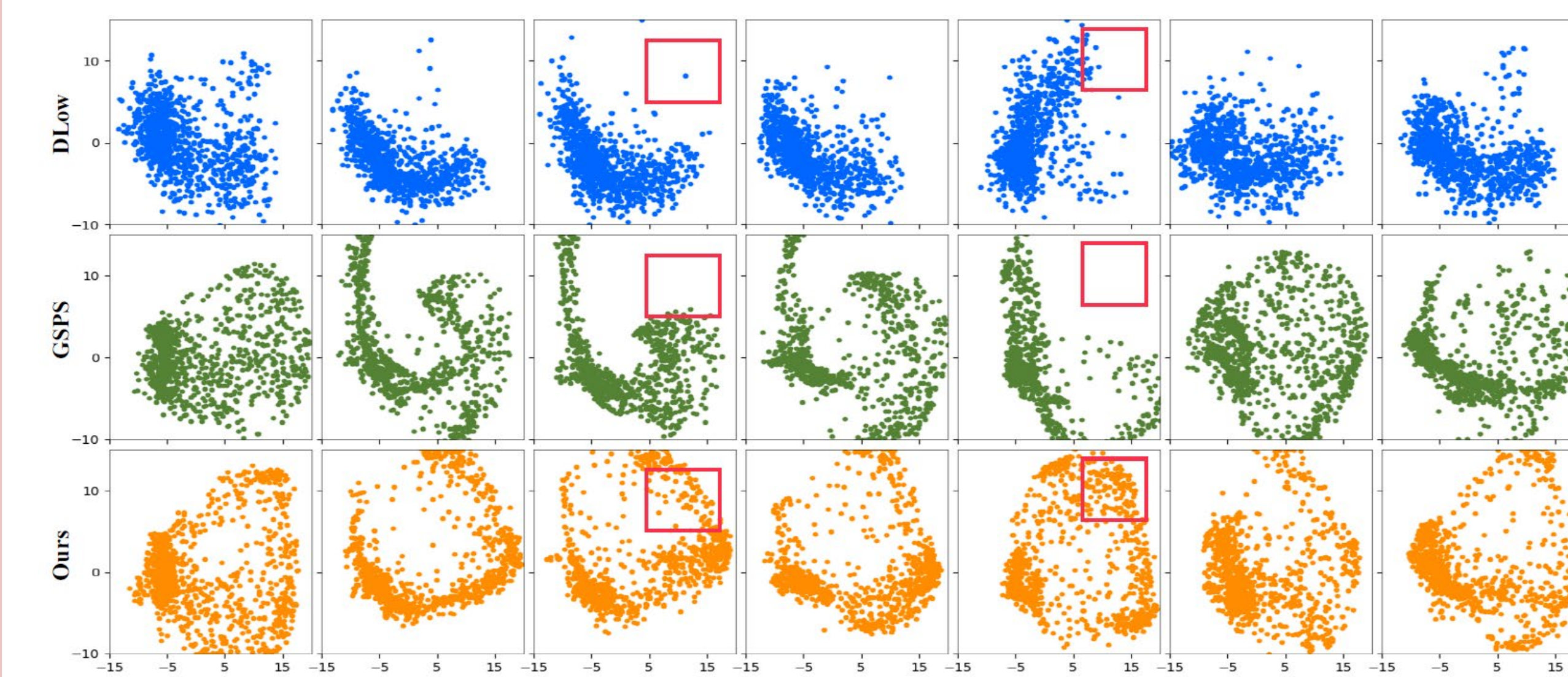
	① Number of basis vectors					② Dimension of auxiliary space					③ Sampling method			④ N_γ		⑤ \mathcal{L}_{div} vs. \mathcal{L}_{div}				
	20	30	40*	50	60	32	64	128*	256	512	Gumbel*	Gaussian	Uniform	w/ N_γ	w/o N_γ	$\mathcal{L}_{div}(25)$	$\mathcal{L}_{div}(25)$	$\mathcal{L}_{div}(25)$	$\mathcal{L}_{div}(25)$	$\mathcal{L}_{div}(1300)$
APD \uparrow	5.929	5.946	5.993	5.969	5.951	5.885	5.931	5.993	5.963	5.957	5.993	5.847	5.730	5.993	5.182	5.993	5.843	5.993	5.993	5.993
ADE \downarrow	0.234	0.234	0.231	0.229	0.233	0.236	0.233	0.231	0.233	0.236	0.231	0.240	0.233	0.231	0.229	0.231	0.211	0.235	0.235	0.235
FDE \downarrow	0.240	0.243	0.240	0.239	0.241	0.245	0.243	0.240	0.241	0.243	0.240	0.246	0.241	0.240	0.236	0.240	0.236	0.240	0.218	0.242
MMAD \downarrow	0.345	0.345	0.340	0.339	0.338	0.337	0.342	0.340	0.336	0.342	0.340	0.343	0.344	0.340	0.322	0.340	0.309	0.340	0.309	0.343
MMFDE \downarrow	0.321	0.319	0.313	0.315	0.312	0.312	0.318	0.313	0.310	0.316	0.313	0.320	0.323	0.313	0.298	0.313	0.287	0.313	0.287	0.321

Qualitative Results

Comparisons with baselines



Holistic views of results after the dimension reduction by PCA



[MSRGCN] Dang L, Nie Y, Long C, *et al.* MSR-GCN: Multi-Scale Residual Graph Convolution Networks for Human Motion Prediction. ICCV, 2021.

[DLow] Yuan Y, Kitani K. Dlow: Diversifying latent flows for diverse human motion prediction. ECCV, 2020.

[GSPS] Mao W, Liu M, Salzman M. Generating smooth pose sequences for diverse human motion prediction. ICCV, 2021.

Acknowledgement

This work is sponsored by Prof. Yongwei Nie's and Prof. Guiqing Li's Natural Science Foundation of China projects (62072191, 61972160), and their Natural Science Foundation of Guangdong Province projects (2019A151010860, 2021A151012301).

Key References