Learning Parallel Dense Correspondence from Spatio-Temporal Descriptors for Efficient and Robust 4D Reconstruction

Jiapeng Tang[1,4] Dan Xu[2] Kui Jia [1] Lei Zhang[3,4]

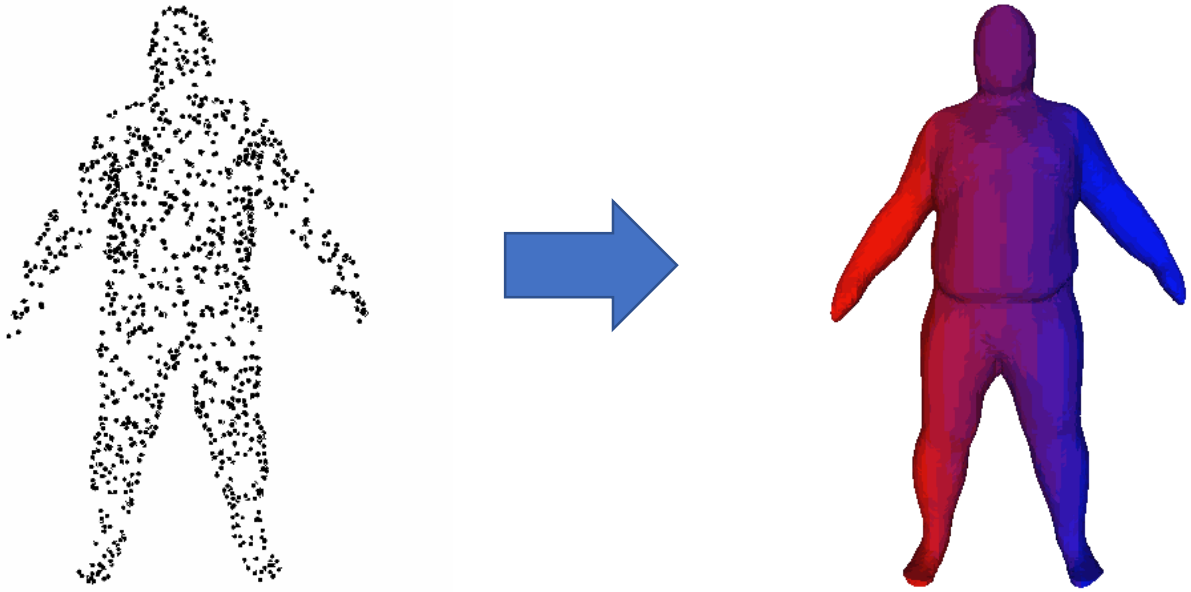
[1] South China University of Technology

[2] The Hong Kong University of Sciences and Technology

[3] The Hong Kong Polytechnic University

[4] DAMO Academy, Alibaba Group

Task: Dynamic 4D Surface Reconstruction

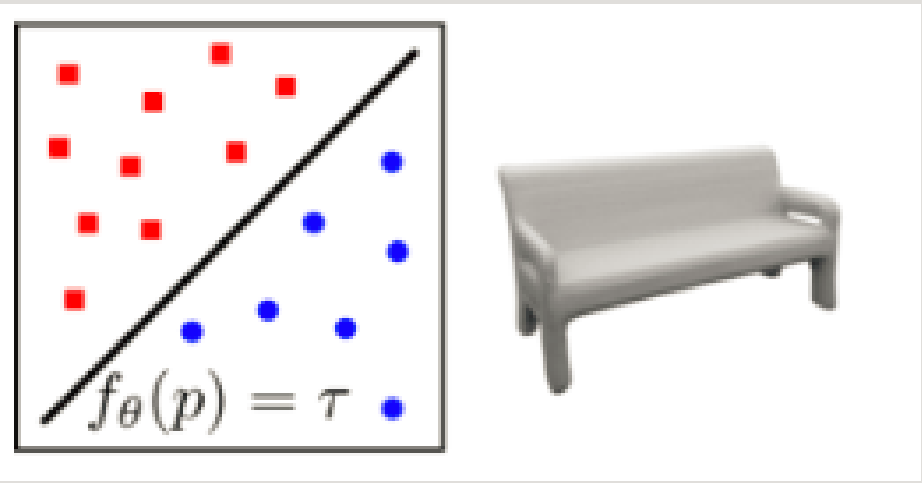


- Goals:

  Accurate Geometry & Temporal Coherence & Dense Correspondence & Faster Inference

Related works: Deep Implicit Representation

- *Occupancy Networks for 3D recon:*
  - represent 3D shape as the continuous decision boundary of a binary classifier.



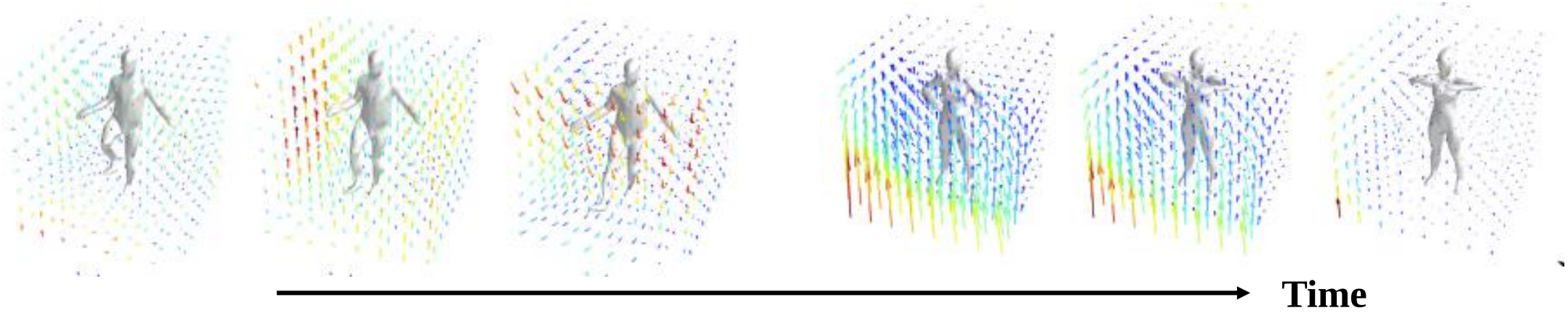
Related works: Extending Deep Implicit Representation into 4D Space

represent **the 3D surface at the first frame** by an occupancy field.

$$f_\theta : \mathbb{R}^3 \rightarrow [0, 1]$$

represent **the temporal evolution of 3D shapes** by a velocity field defined in 4D space.

$$\mathbf{v}_{\hat{\theta}} : \mathbb{R}^3 \times [0, T] \rightarrow \mathbb{R}^3$$



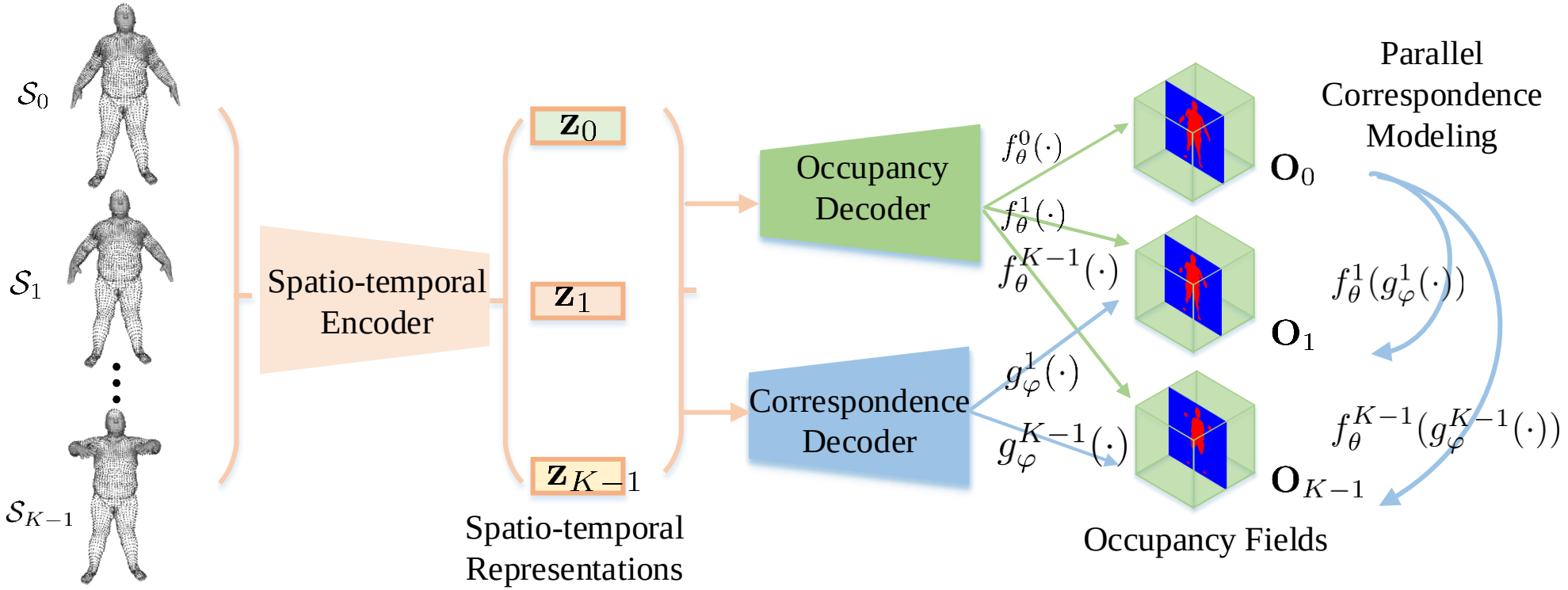
integrating occupancy flow to recover shape motions

Challenges:

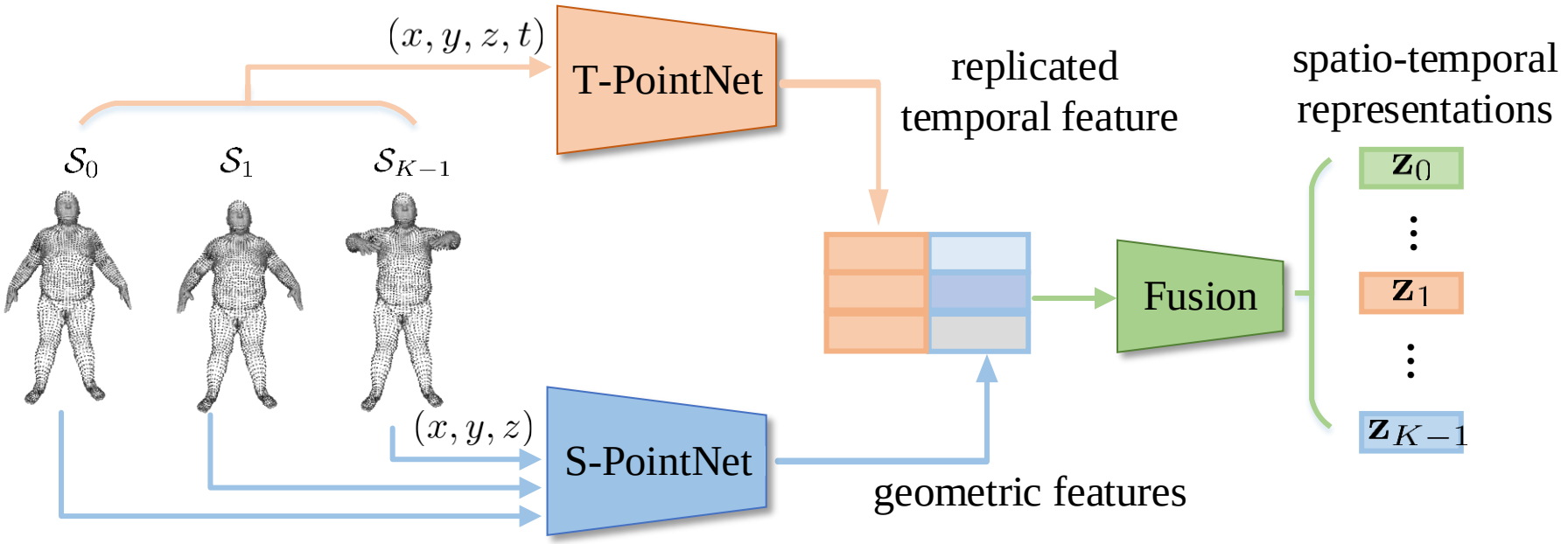
- 4D Point Cloud Encoder
  - Ignore the aggregation of shape properties from multiple frames.
  - Inefficiently capture temporal dynamics.
- Dense Correspondence Modeling
  - Low computational efficiency (duo to solving neural ordinary differential equations).
  - Non-coherent human motions (caused by the accumulated prediction errors).

Overall Pipeline

Key idea: *Parallelly establish the dense correspondence* between predicted occupancy fields at different time steps via explicitly learning continuous displacement vector fields from *robust spatio-temporal shape representations*.



Spatio-temporal Representations Learning



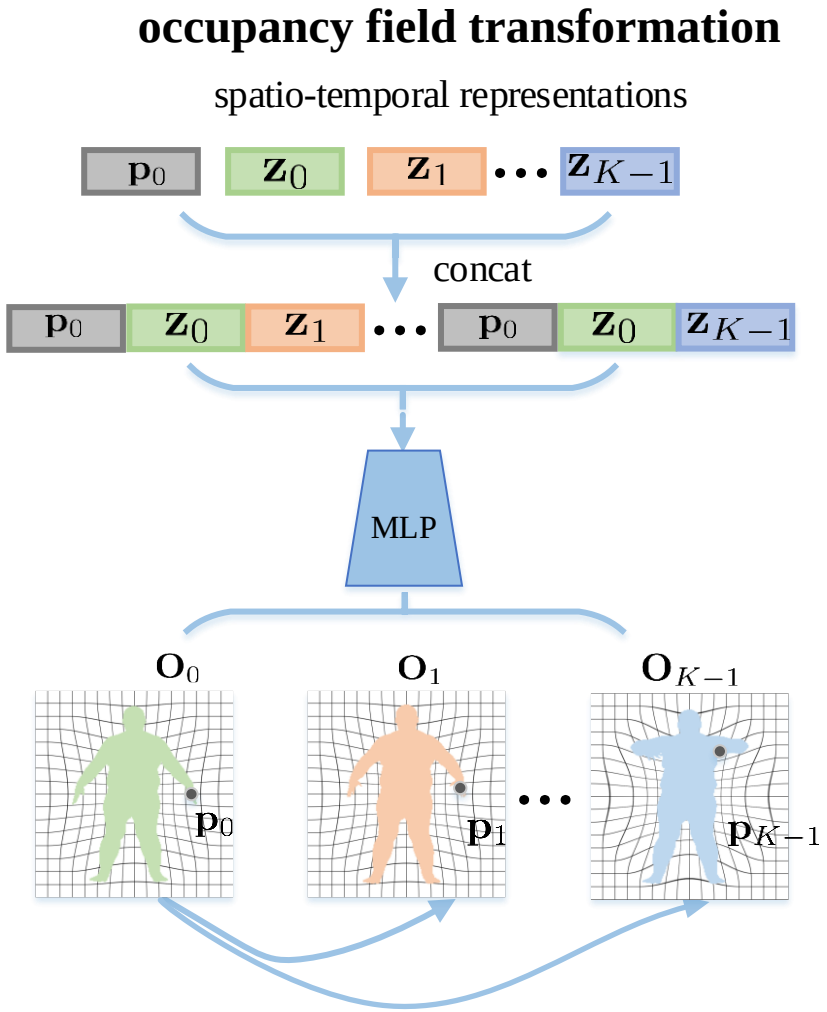
aggregate shape properties and explore dynamics variations

Cross-time Correspondence Modeling

- Predict spatially continuous displacement vector fields in parallel paths through a shared MLP.
- The occupancy field transformation between $\mathbf{O}_0$ and $\mathbf{O}_k$:

$$\mathbf{p}_k - \mathbf{p}_0 = g_\varphi(\mathbf{p}_0 \oplus \mathbf{z}_0 \oplus \mathbf{z}_k)$$

- Bypass the expensive computation of Neural ODE.
- Support parallel surface mesh deformations during inference.



4D Shape Reconstruction

	Method	IoU	Chamfer	Correspond.
S1	PSGN 4D [11]	-	0.101	0.102
	ONet 4D [27]	77.9%	0.084	-
	OFlow [32]	81.5%	0.065	0.094
	Ours	**84.9%**	**0.055**	**0.080**
S2	PSGN 4D [11]	-	0.119	0.131
	ONet 4D [27]	66.6%	0.140	-
	OFlow [32]	72.3%	0.084	0.117
	Ours	**76.2%**	**0.071**	**0.098**

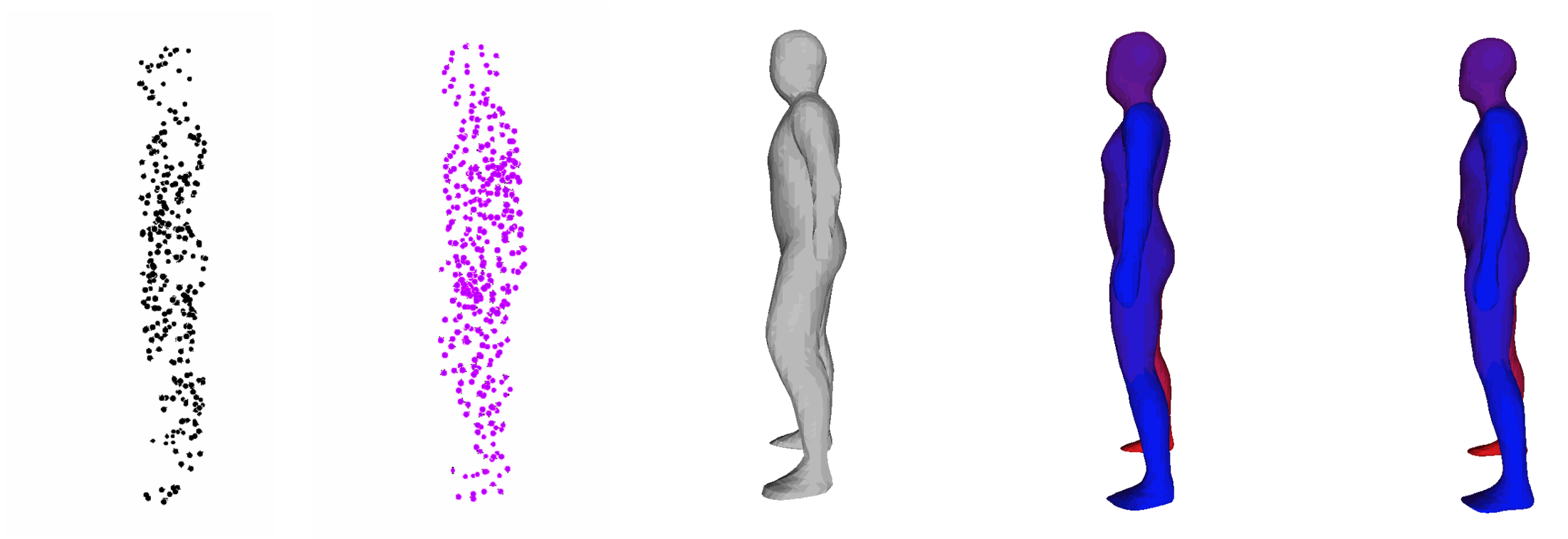
time-evenly sampled point cloud sequence

	Method	IoU	Chamfer	Correspond.
S1	PSGN 4D [11]	-	0.148	0.121
	ONet 4D [27]	71.9%	0.114	-
	OFlow [32]	76.9%	0.090	0.134
	Ours	**83.8%**	**0.059**	**0.090**
S2	PSGN 4D [11]	-	0.155	0.140
	ONet 4D [27]	62.8%	0.130	-
	OFlow [32]	67.2%	0.112	0.178
	Ours	**74.8%**	**0.076**	**0.118**

time-unevenly sampled point cloud sequence with large variations

higher robustness on non-uniform sequences with large variations

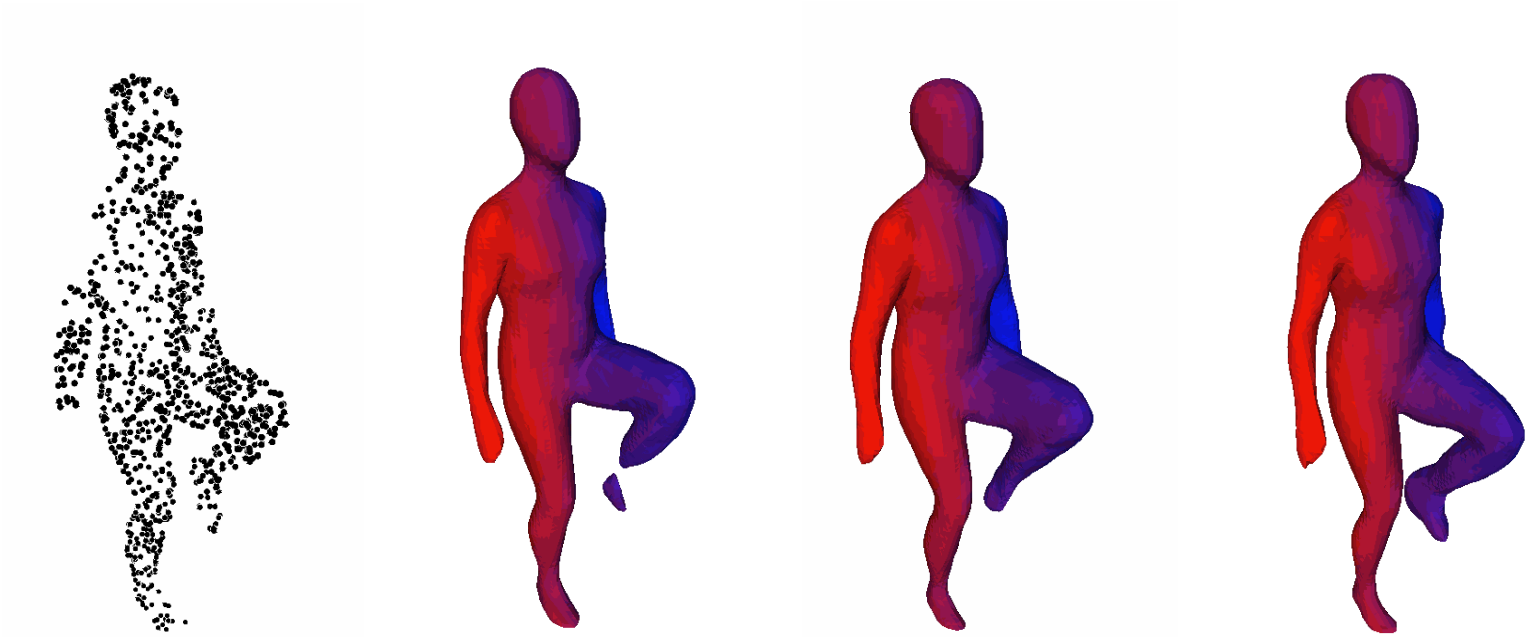
4D Shape Reconstruction



Input | PSGN 4D | ONet 4D | OFlow | Ours

more robust geometries and coherent human motions

Ablation Studies



Input Ours (C1) Ours (C2) Ours

- Ours(C1): without spatio-temporal encoder.
- Ours(C2): without parallel correspondence modeling.

4D Shape Completion

Input

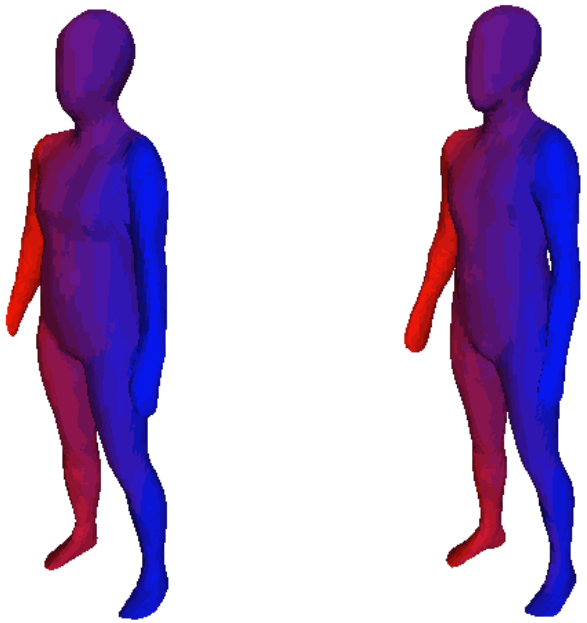
OFlow

Ours

Space and Time Complexity Comparison

Method	Mem. (GB)	Train (day)	Inference (s)
OFlow [32]	3.53	42	1.84
Ours	10.8	10	0.23
	Forward (s).	Backward (s).	Train (s)
OFlow [32]	0.33	4.02	4.35
Ours	0.45	0.49	0.94

about 4 times faster in training and 8 times in inference



OFlow Ours

THANK YOU!

The dataset and code are available at

https://github.com/tangjiapeng/LPDC-Net