



# AutoAvatar: Autoregressive Neural Fields for Dynamic Avatar Modeling Ziqian Bai<sup>1,2</sup> Timur Bagautdinov<sup>2</sup> Javier Romero<sup>2</sup> Michael Zollhöfer<sup>2</sup> Ping Tan<sup>1</sup> Shunsuke Saito<sup>2</sup>

<sup>1</sup> Simon Fraser University <sup>2</sup> Reality Labs Research

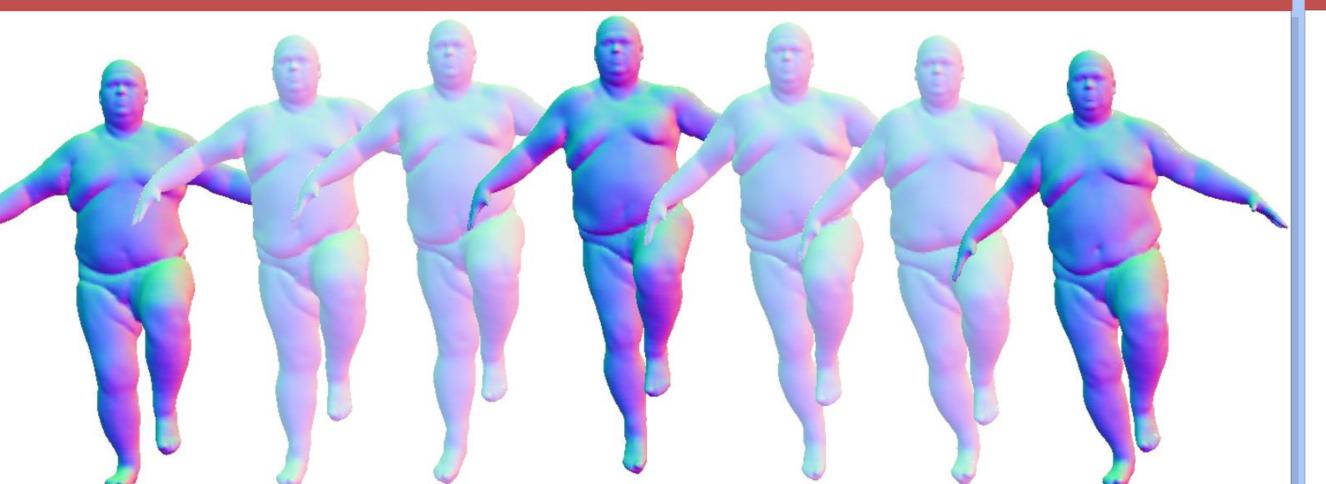




Introduction

Motivation: Dynamic effects are important for realistic body avatars, such as history-dependent soft-tissue deformations (e.g. belly).

Goal: Build body avatars driven by SMPL poses with dynamic effects from raw scans without precise surface registration.



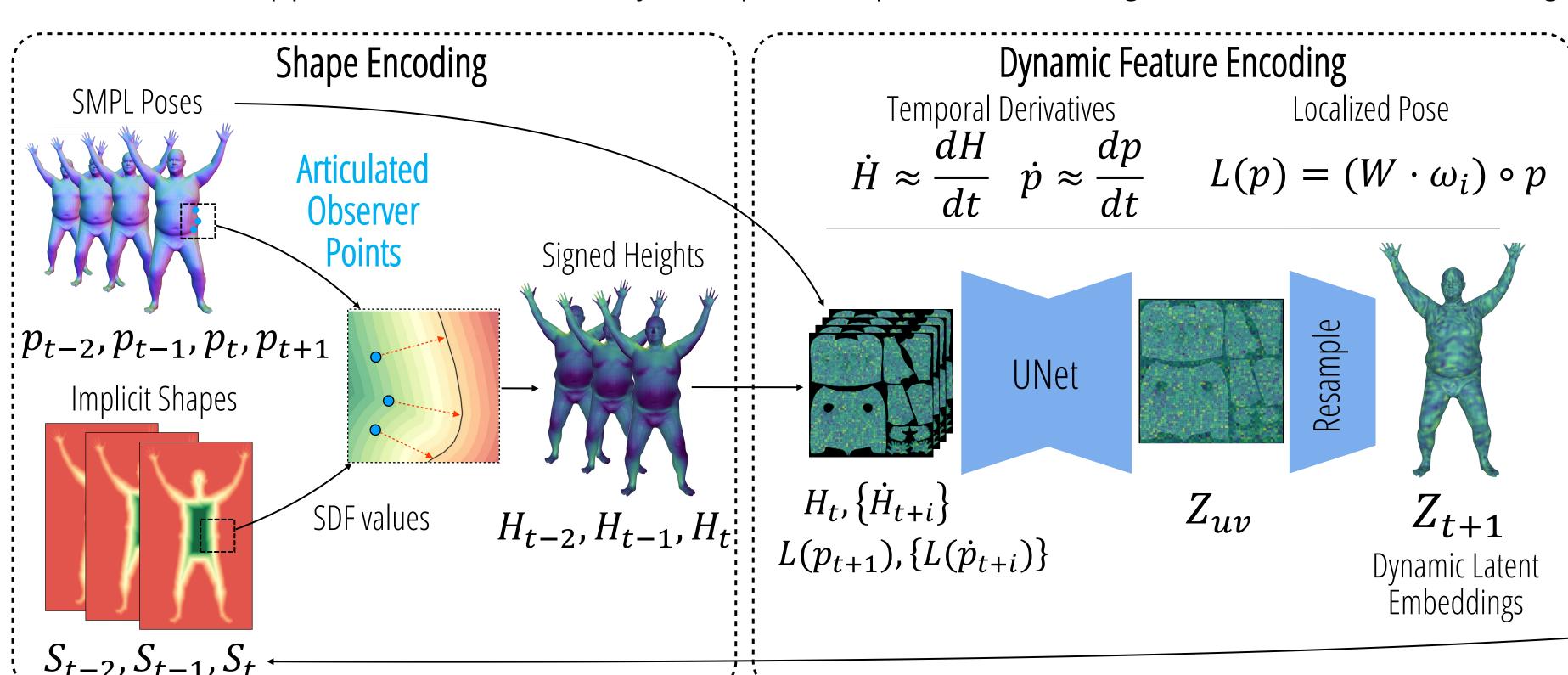
#### Prior works:

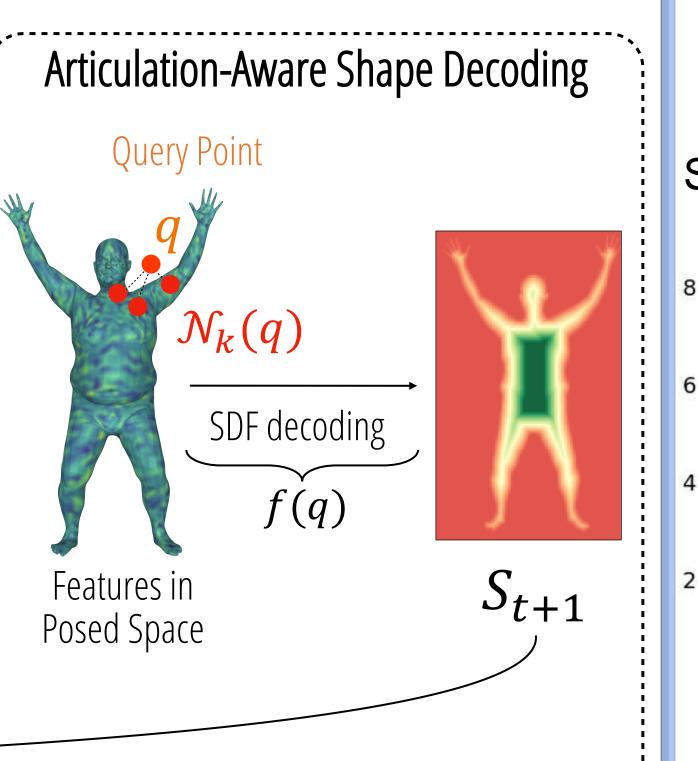
- Some approaches [1] use implicit surfaces to remove the need of precise surface registration, but do not capture temporal dynamics.
- Others use autoregressive frameworks [2] or physics-based simulation [3] to model temporal dynamics, but require precise surface registration or templates that are either hard to obtain or impose undesired topology constraints.

Contributions: The first autoregressive approach for modeling history-dependent implicit surfaces of human bodies.

### Our Approach

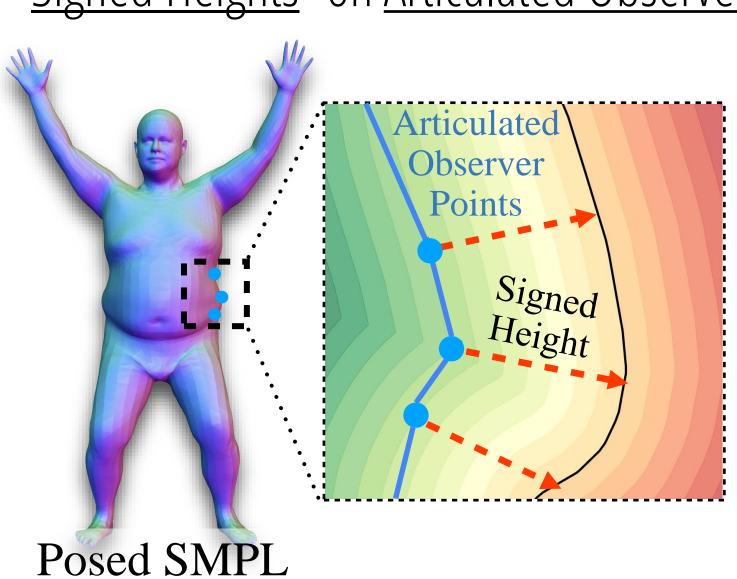
**Overview**: Our approach takes the <u>history of implicit shapes</u> in an <u>autoregressive</u> manner for learning dynamics.





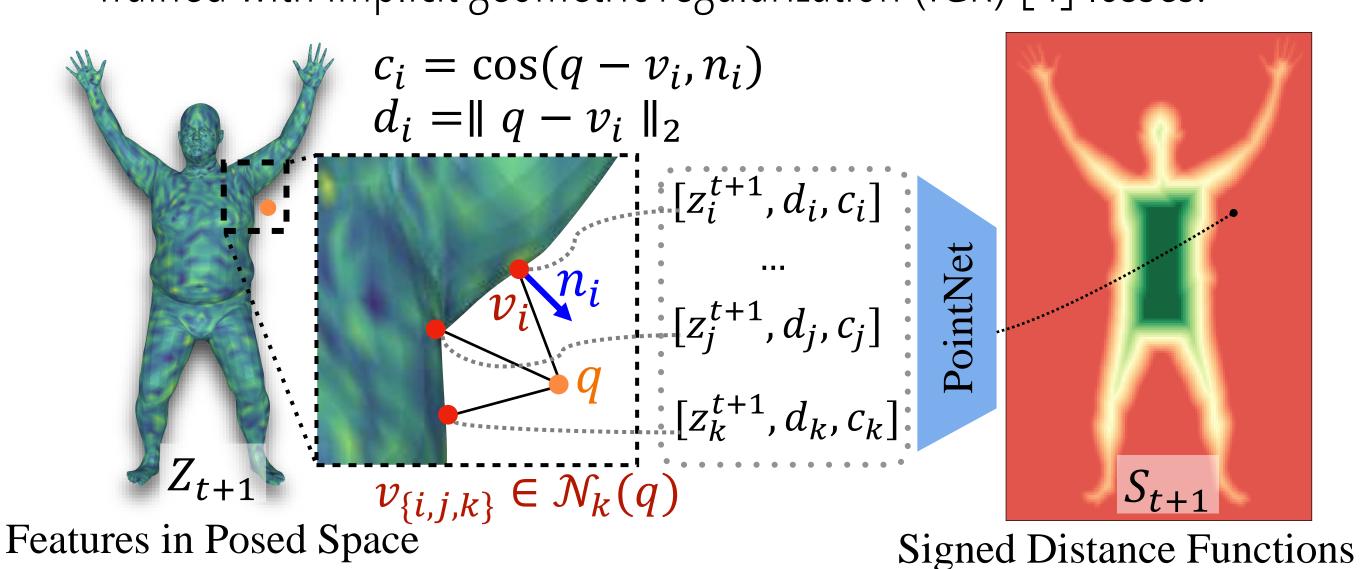
## Shape Encoding:

- How to efficiently encode and feed back implicit neural surfaces (SDF) into the model.
- → "Signed Heights" on Articulated Observer Points.



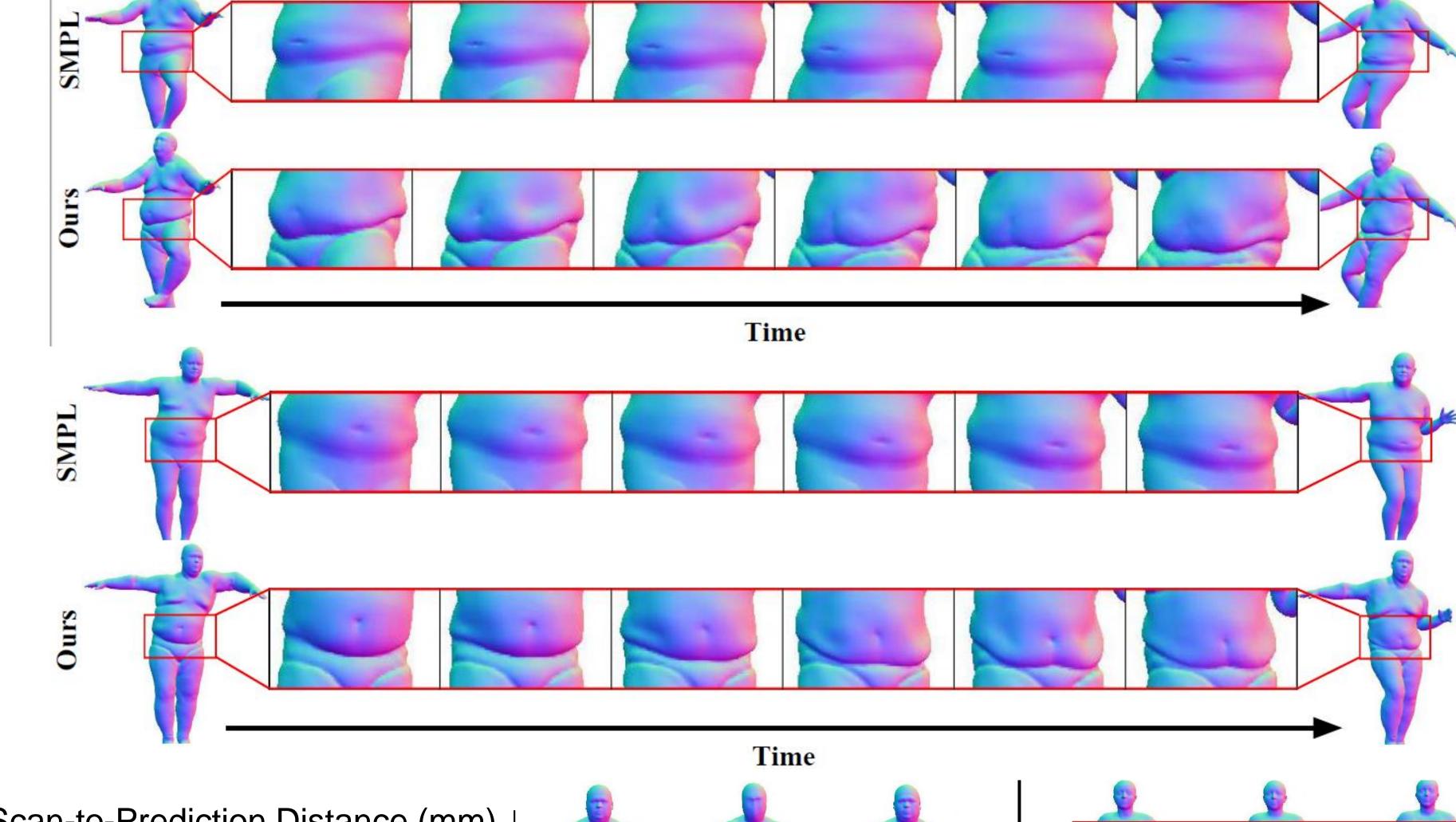
#### Articulation-Aware Shape Decoding:

- f(q): K-nearest vertices  $\mathcal{N}_k(q)$  + PointNet.
- Pose-agnostic spatial representation  $c_i$  and  $d_i$ .
- Trained with implicit geometric regularization (IGR) [4] losses.



### Experiments





Scan-to-Prediction Distance (mm) ↓						9
Ours SoftSMPL				Jump to top		
6-				nf.		J
2-				to bottom		
0 50 100 150 200 250 # of frames	Scan	SNARF	Ours	Scan	SoftSMPL	Our
(a) Mean Scan-to-Prediction Distance (mm) $\downarrow$ on I	DFaust.		(b) Mean Square	d Error of Volume Cha	nge ↓ on DFaust.	

Rollout (# of				(# of frames)					Rollout (# of frames)					
		1	2	4	8	16	30			2	4	8	16	30
Interpolation Set					Interpolation Set									
Non-Autoregressive	SNARF [5]	7.428	7.372	7.337	7.476	7.530	7.656	Non-Autoregressive	SNARF [5]	0.01582	0.01552	0.01610	0.01658	0.01682
	Pose	4.218	4.202	4.075	4.240	4.409	4.426		Pose	0.01355	0.01305	0.01341	0.01367	0.01387
	PoseTCN	4.068	4.118	4.086	4.228	4.405	4.411		PoseTCN	0.01364	0.01323	0.01350	0.01399	0.01416
	Pose + dPose	3.852	3.841	3.764	3.972	4.164	4.156		Pose + dPose	0.01288	0.01247	0.01273	0.01311	0.01321
Autoregressive	G-embed	2.932	3.006	3.131	3.462	3.756	3.793	5.916 Autoregressive	G-embed	0.01179	0.01168	0.01199	0.01248	0.01265
	L-embed	1.784	2.138	2.863	4.250	5.448	5.916		L-embed	0.01003	0.01180	0.01466	0.01716	0.01844
	Ours	1.569	1.914	2.587	3.627	4.736	5.255		Ours	0.00902	0.01053	0.01258	0.01456	0.01565
Extrapolation Set					Extrapolation Set									
Non-Autoregressive P	SNARF [5]	7.264	7.287	7.321	7.387	7.308	7.251	Non-Autoregressive	SNARF [5]	0.01178	0.01194	0.01251	0.01228	0.01206
	Pose	4.303	4.306	4.308	4.299	4.385	4.398		Pose	0.01027	0.01039	0.01074	0.01052	0.01039
	PoseTCN	4.090	4.091	4.105	4.119	4.233	4.257		PoseTCN	0.01020	0.01038	0.01064	0.01040	0.01029
	Pose + dPose	3.984	3.991	4.017	4.063	4.162	4.190		Pose + dPose	0.00992	0.01014	0.01048	0.01029	0.01013
Autoregressive	G-embed	2.884	2.926	3.043	3.258	3.577	3.787	2 Autoregressive	G-embed	0.00936	0.00959	0.00995	0.00996	0.00998
	L-embed	1.329	1.539	2.079	3.326	4.578	5.192		L-embed	0.00648	0.00821	0.01100	0.01308	0.01402
	Ours	1.150	1.361	1.834	2.689	3.789	4.526		Ours	0.00567	0.00715	0.00915	0.01039	0.01107

#### References

- [1] SCANimate: Weakly supervised learning of skinned clothed avatar networks, Saito et al., CVPR 2021.
- [2] Softsmpl: Data-driven modeling of nonlinear soft-tissue dynamics for parametric humans, Santesteban et al., Computer Graphics Forum 2020.
- [3] Fem simulation of 3d deformable solids: A practitioner's guide to theory, discretization and model reduction, Sifakis et al., SIGGRAPH Courses 2012.
- [4] Implicit geometric regularization for learning shapes, Gropp et al., ICML 2020.
- [5] Snarf: Differentiable forward skinning for animating non-rigid neural implicit shapes, Chen et al., ICCV 2021.