DSNet: Dynamic skin deformation prediction by Recurrent Neural Network

Hyewon SEO

Seminar@ MLMS

March 2021

cnrs

iCUBE

Université de Strasbourg

Introduction

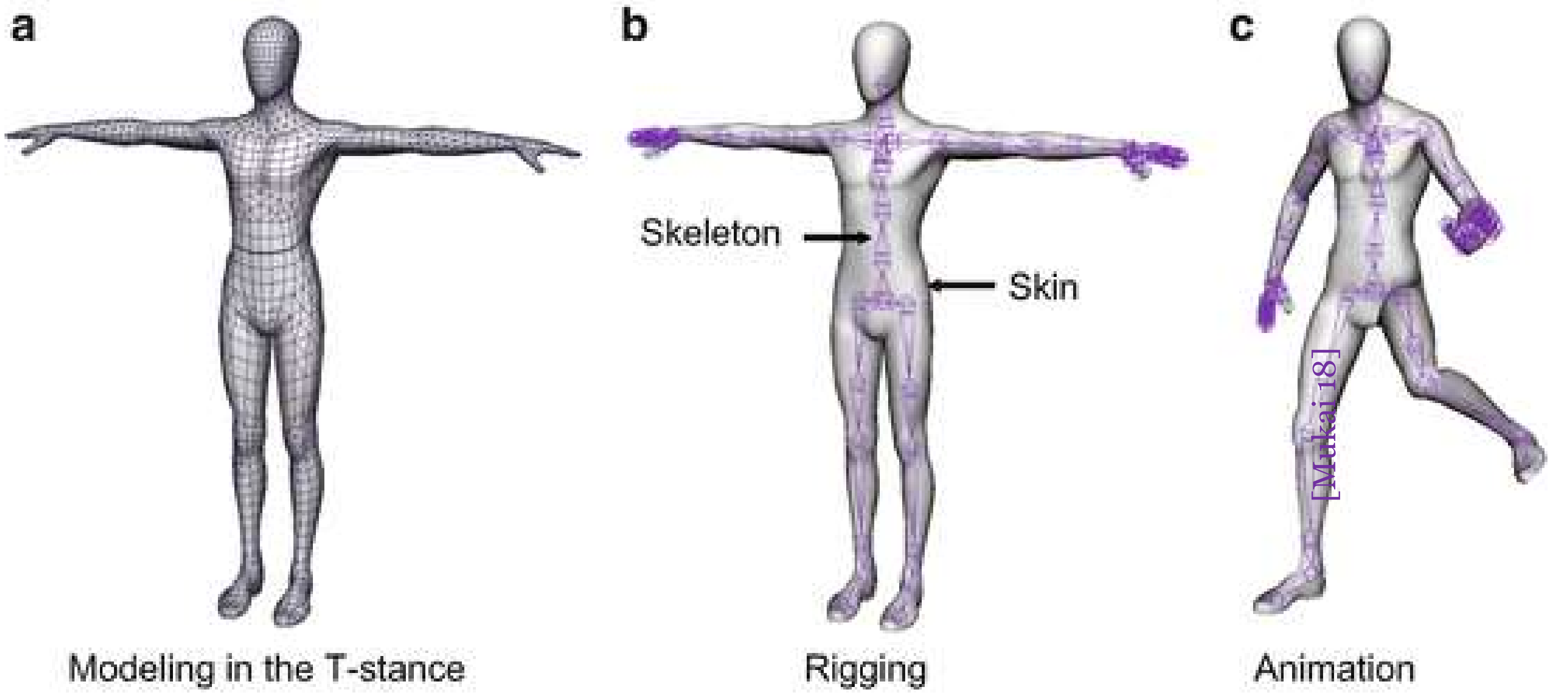
Dynamic skin deformation contributes to the enriched realism of character models in rendered scenes.

https://www.youtube.com/watch?v=Jf6CmxeEpw4&t=15s

It has a long tradition in CG and CA…

Introduction

Linear blend skinning: [MTT91] $\mathcal{W}(\boldsymbol{\theta}; M, \boldsymbol{J}, \boldsymbol{W})$



Get the skin surface M.

Define the skeleton $\boldsymbol{J}$.
Map vertices to the skeleton: $\boldsymbol{W}$

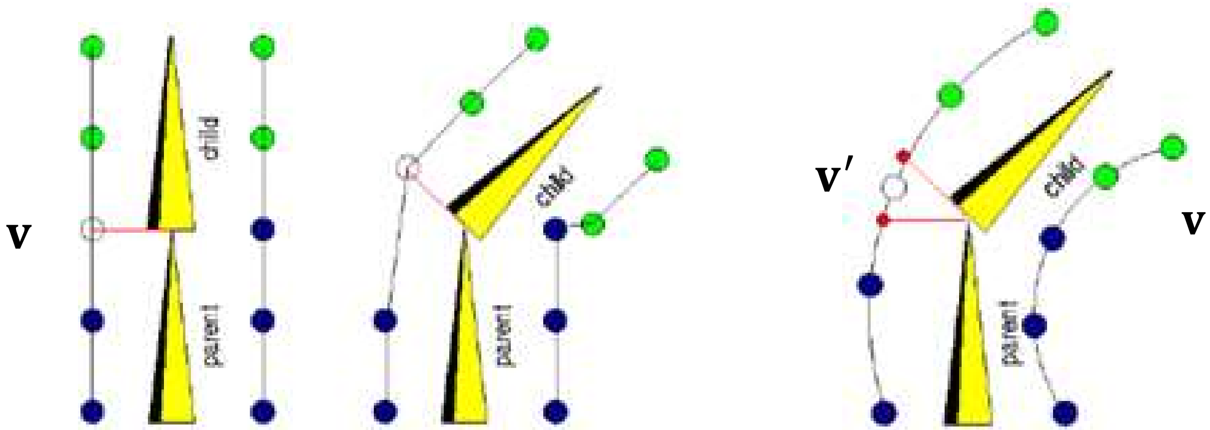
Apply rotations $\boldsymbol{\theta}$ to the skeleton.
Reposition vertices ($\mathcal{W}$).

[MTT91] Magnenat-Thalmann N., Thalmann D., "Human Body Deformations Using Joint-dependent Local Operators and Finite-Element Theory", Making Them Move, N.Badler, B.A.Barsky, D.Zeltzer, eds, Morgan Kaufmann, San Mateo, California, pp.243-262, 1991.

[Mukai18] Tomohiko Mukai, Example-Based Skinning Animation, pp 2093-2112, Handbook of Human Motion, Springer, 2018.

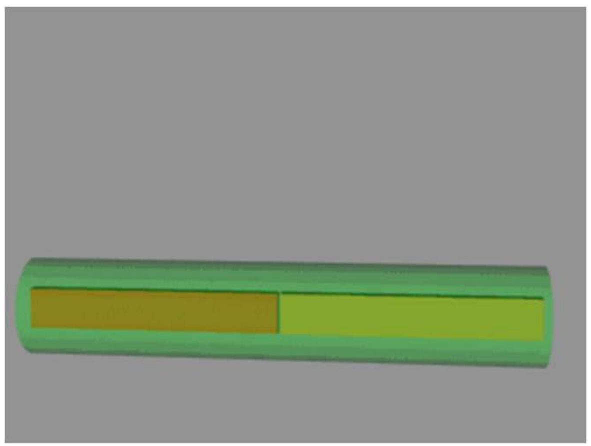
Introduction

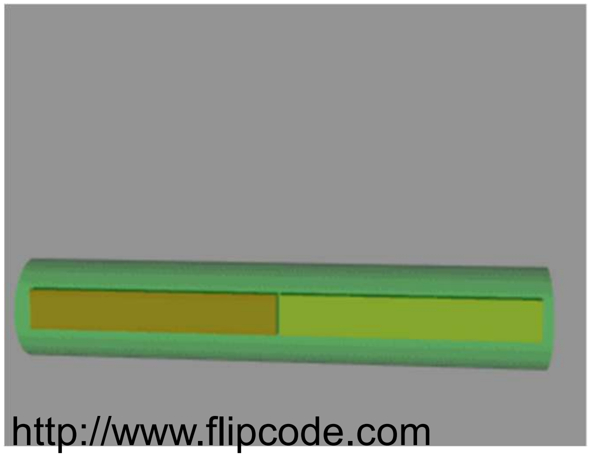
Linear blend skinning $\mathcal{W}(\boldsymbol{\theta}; M, \boldsymbol{J}, \boldsymbol{W})$



$$\mathbf{v}' = \sum w_i(\mathbf{M}_i(\theta_i) \cdot \mathbf{v})$$

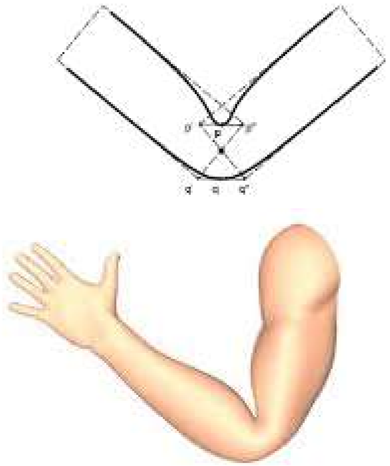
$$\sum w_i = 1$$





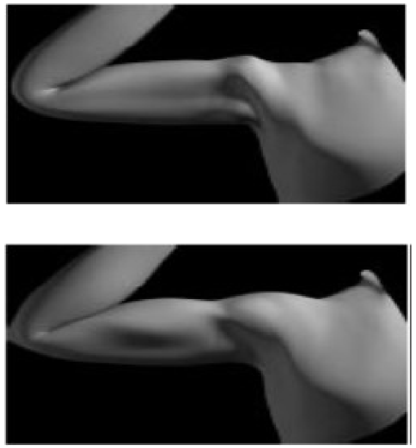
Introduction

Limitations of LBS



[Yang & Zhang 06]

Unnatural deformations
at certain poses



[Lewis et al 06]

Impossible to express
nonlinear deformation
i.e. muscle bulging

[Romero et al 20]

Impossible to simulate
skin dynamics
i.e. jiggle effect

[Yang & Zhang 06] Xiaosong Yang and J. J. Zhang, "Stretch It - Realistic Smooth Skinning," International Conference on Computer Graphics, Imaging and Visualisation (CGIV'06), Sydney, Qld., 2006, pp. 323-328.

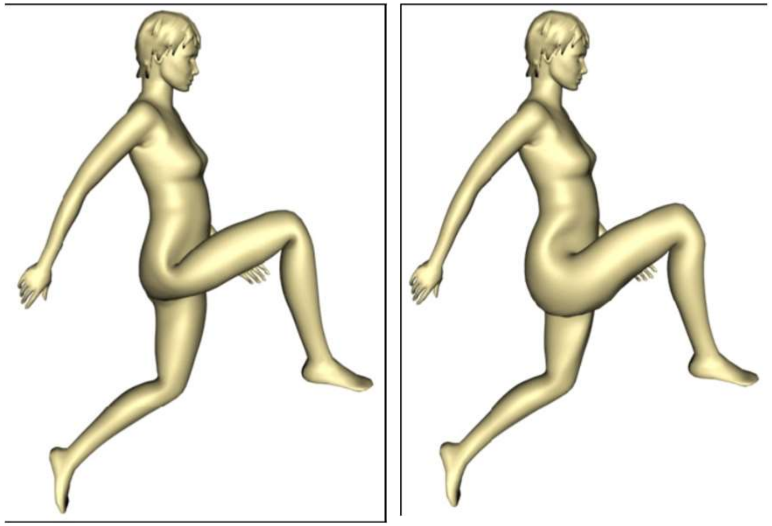
[Lewis et al 06] J. P. Lewis, Matt Cordner, and Nickson Fong. 2000. Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation. Proc Computer graphics and interactive techniques SIGGRAPH '00.

[Romero et al 20] Romero, Cristian & Otaduy, Miguel & Casas, Dan & Perez, Jesus. (2020). Modeling and Estimation of Nonlinear Skin Mechanics for Animated Avatars. Computer Graphics Forum. 39. pp. 77-88.

Previous work

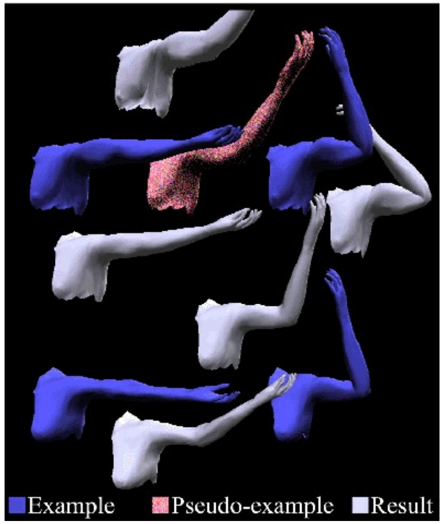
Solutions: Previous work

Geometric



[Magnenat-Thalmann et al. 04]

Example-based



[Sloan et al. 01]

Physics-based

[Ziva Dynamics]

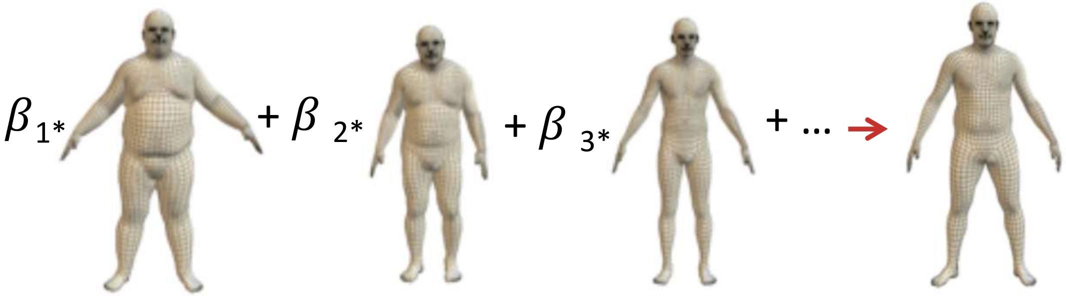
[Magnenat-Thalmann et al. 04] N Magnenat-Thalmann, F Cordier, H Seo, G Papagianakis, Modeling of bodies and clothes for virtual environments, 2004 International Conference on Cyberworlds, 201-208

[Sloan et al. 01] P. P. Sloan, C. Rose and M. Cohen, "Shape by Example", ACM SIGGRAPH Symposium on Interactive 3D Graphics, NC, USA, pp. 135–143, 2001.

Previous work

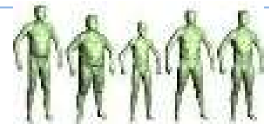
Data-driven body shape modelers

[SMT03, ASK+05]

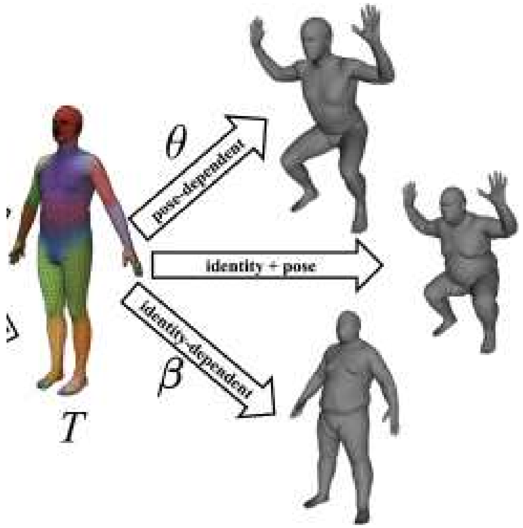


$$M(\vec{\boldsymbol{\beta}}) = \overline{\boldsymbol{T}} + \sum \beta_n \boldsymbol{S_n}$$

Basis shape vectors



A unifying framework for subject- & pose-dependent shapes [HLRB12,LMRP+15]



[SMT03] Seo H., and Magnenat-Thalmann N., "An Automatic Modeling of Human Bodies from Sizing Parameters", ACM SIGGRAPH 2003 Symposium on Interactive 3D Graphics (April), pp.19-26, Monterey, USA, 2003.

[ASK+05] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis J., SCAPE: Shape Completion and Animation of People. ACM Trans. Graph. (Proc. SIGGRAPH 24, 3, 408–416) 2005.

[HLRB12] D. Hirshberg, M. Loper, E. Rachlin, and M. Black, Coregistration: Simultaneous alignment and modeling of articulated 3D shape. In European Conf. on Computer Vision (ECCV), LNCS 7577, Part IV, 242–255, 2012.

[LMRP+15] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black. SMPL: A Skinned Multi-Person Linear Model. ACM Trans. Graphics (Proc. SIGGRAPH Asia), 2015.

Previous work

Data-driven dynamic human shape modelers [PMR+15, CO18]

$$\boldsymbol{M}_{\mathrm{t}} = \boldsymbol{SDD}_{\mathrm{t}} + \Delta_{\mathrm{t}}$$

$$\boldsymbol{SDD}_{\mathrm{t}} = f_{linear}(\boldsymbol{\theta_t}; \bar{M}(\boldsymbol{\beta}), J(\boldsymbol{\beta}), \boldsymbol{W})$$

$$\Delta_{\mathrm{t}} = g_{linear}(\boldsymbol{v_t}, \boldsymbol{a_t}, \dot{\boldsymbol{\theta}}_{\boldsymbol{t}}, \ddot{\boldsymbol{\theta}}_{\boldsymbol{t}}, \Delta_{\mathrm{t-1}}, \Delta_{\mathrm{t-2}}; \bar{M}(\boldsymbol{\beta}))$$

[PMR+15] Pons-Moll G., Romero J., Mahmood N., and Black M. J.: Dyna: a model of dynamic human shape in motion. *ACM Trans. Graph. 34, 4,* Article 120 (July 2015).

[CO18] Casas, D. & Otaduy, M. (2018). Learning Nonlinear Soft-Tissue Dynamics for Interactive Avatars. Proc. ACM Computer Graphics and Interactive Techniques. 1. 1-15.

[BODO18] Bailey S. W., Otte D., Dilorenzo P., and O'Brien J. F.: Fast and Deep Deformation Approximations. *ACM Trans. Graph.*, 37(4):119:1–12, August 2018.

DS-Net : Overview

Our goal is to learn a function $f(\{\boldsymbol{\theta}_t\}) = \{\Delta_t\}, t = 1, \ldots T$

c.f. $\boldsymbol{\varphi}_t = \{\boldsymbol{v}_t, \boldsymbol{a}_t, \dot{\boldsymbol{\theta}}_t, \ddot{\boldsymbol{\theta}}_t\}$

⚠ Both input and outputs are **sequences!!**

- The results of frame *t* depend on the results of previous frames *t-1, t-2, …*
- We should also consider subject specificity i.e. β.

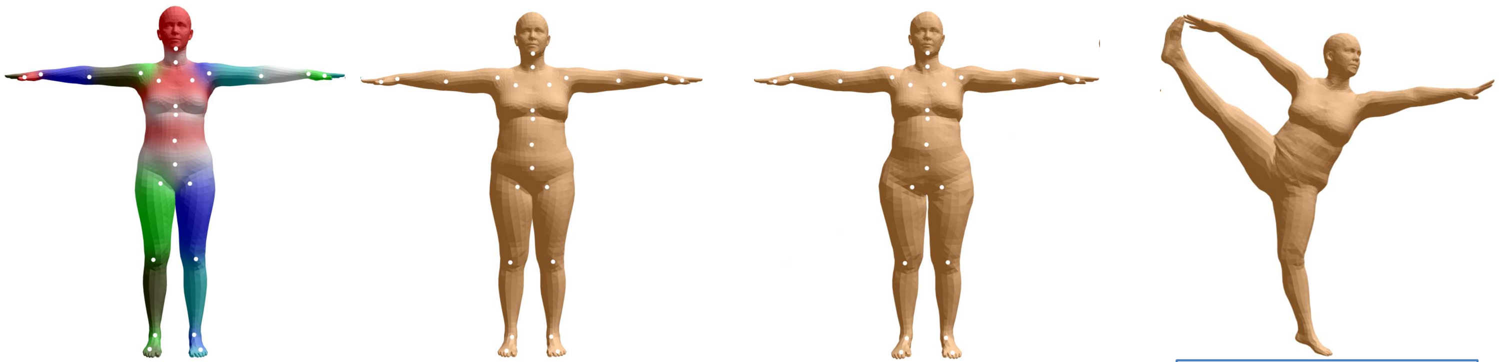
=> $\boxed{\Delta_t = f(\boldsymbol{\theta}_t, f(\boldsymbol{\theta}_{t-1}), \boldsymbol{\beta})}$

- We deploy LSTM network to learn our function.

A common shape space is required: SMPL! (A Skinned Multi-Person Linear Model)

Dynamic skin : model

SMPL: A Skinned Multi-Person Linear Model [LMRP+15]



(a) $\bar{\mathbf{T}}$, $\mathcal{W}$ (b) $\bar{\mathbf{T}} + B_S(\vec{\beta})$, $J(\vec{\beta})$ (c) $T_P(\vec{\beta},\vec{\theta}) = \bar{\mathbf{T}} + B_S(\vec{\beta}) + B_P(\vec{\theta})$ (d) $W(T_P(\vec{\beta},\vec{\theta}), J(\vec{\beta}), \vec{\theta}, \mathcal{W})$

$$M(\boldsymbol{\beta},\boldsymbol{\theta}) = \mathcal{W}(\boldsymbol{\theta}, \boxed{\bar{M}(\boldsymbol{\beta},\boldsymbol{\theta})}, J(\boldsymbol{\beta}), \boldsymbol{W})$$

SDD: linear blend skinning

$$\boxed{\bar{M}(\boldsymbol{\beta},\boldsymbol{\theta})} = \bar{\boldsymbol{T}} + M_S(\boldsymbol{\beta}) + M_P(\boldsymbol{\theta})$$

Template model Shape blend shape Pose blend shape

$$M_S(\boldsymbol{\beta}) = \boldsymbol{\mu}_S + \sum_{n=1}^{|\boldsymbol{\beta}|} \beta_n \boldsymbol{s}_n$$

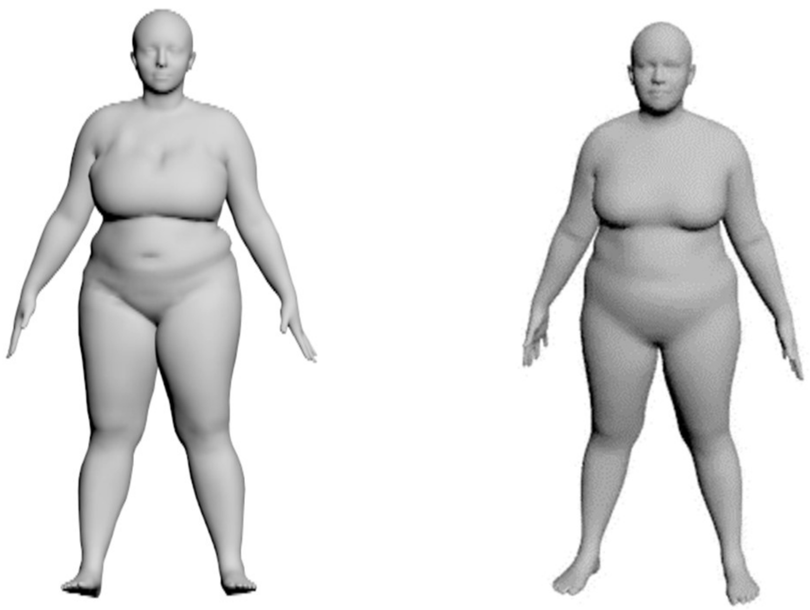
$$M_P(\boldsymbol{\theta}) = \sum_{n=1}^{9K} (R_n(\boldsymbol{\theta}) - R_n(\boldsymbol{\theta}^0))\, \boldsymbol{P}_n$$

[LMRP+15] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black. SMPL: A Skinned Multi-Person Linear Model. ACM Trans. Graphics (Proc. SIGGRAPH Asia), 2015.

Dynamic skin: dataset

Dyna dataset [PRMB15]

- Captured shapes exhibiting dynamic skin deformation
- 5 (female) subjects, 10~14 motions each
- Inter-, intra-subject correspondence with N=6890 vertices, 13776 triangles
- The duration of each sequence varies: 2 ~15 sec.



[PRMB15] Pons-Moll G., Romero J., Mahmood N., and Black M. J.: Dyna: a model of dynamic human shape in motion. ACM Trans. Graph. 34, 4, Article 120 (July 2015), 14 pages.

Dynamic skin: dataset

Dyna [PRMB15] ***: training & validation***

Mosh [LMB14] ***: test***

dataset	subjects	motions	fps	No. sequences (men/women)
Dyna	5 men, 5 women	10~14 motions for each subject: one-leg jumping, light hoping, jumping jacks, shake hips, running in place, etc.	60	66 / 67
Mosh	Same subjects as above	Includes some skin-dynamics inducing motions (side-to-side hoping, basketball, kicking) that are not included Dyna.	100	24 / 30

[PRMB15] Pons-Moll G., Romero J., Mahmood N., and Black M. J.: Dyna: a model of dynamic human shape in motion. ACM Trans. Graph. 34, 4, Article 120 (July 2015), 14 pages.
[LMB14] M. Loper, N. Mahmood, and M. J. Black. MoSh: Motion and Shape Capture from Sparse Markers. ACM Trans. Graph., 33(6):220:1–220:13, Nov. 2014.

Generation of training data

Extraction of SMPL parameters + redisuals Δ, from each mesh.

For each motion sequence m:

1. Compute the best matching SMPL parameters $(\boldsymbol{\beta}, \boldsymbol{\theta}_1)$ at frame 1.

$$\min_{\boldsymbol{\beta},\boldsymbol{\theta}_1} \left\| \mathcal{W}\left(\overline{\boldsymbol{T}} + M_S(\boldsymbol{\beta}) + M_P(\boldsymbol{\theta}_1)\right) - S_1 \right\|_2 .$$

2. Compute the best matching SMPL parameters $\boldsymbol{\theta}_t$ for each frame > 1.

$$\min_{\boldsymbol{\theta}_t} \left\| \mathcal{W}\left(\overline{\boldsymbol{T}} + M_S(\boldsymbol{\beta}^*) + M_P(\boldsymbol{\theta}_t)\right) - S_t \right\|_2 .$$

Fixed throughout all frames > 1.

3. The displacement vector is considered as the dynamic skin component.

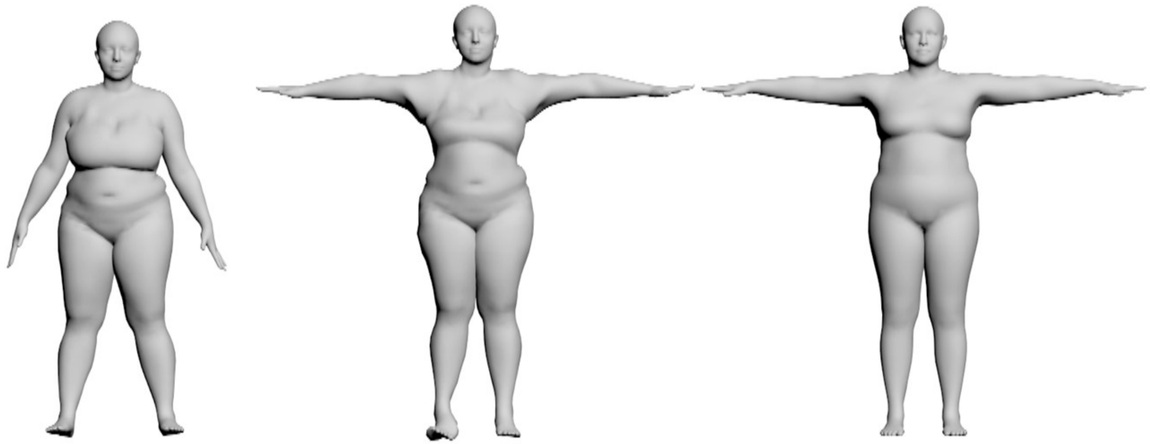
$$\Delta_t = \mathcal{W}^{-1}(S_t) - \left(\overline{\boldsymbol{T}} + M_S(\boldsymbol{\beta}^*) + M_P(\boldsymbol{\theta}_t^*)\right)$$

Unposing operation: transforms a body mesh to its rest pose.

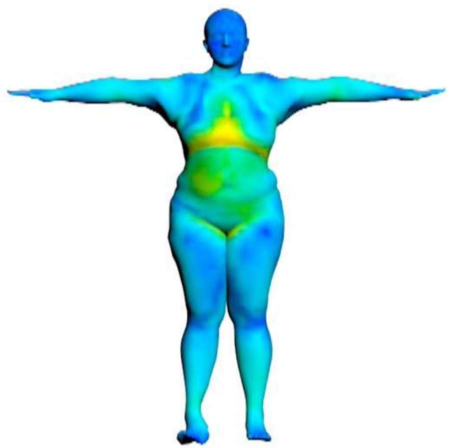
The training data is a set of input and output pairs : $\{(\boldsymbol{\beta}^m, \boldsymbol{\theta}_t^m, \Delta_t^m)\}$, *m*=1…65.

Generation of training data

Mesh alignment results



(a) $\boldsymbol{S}$ (b) $\mathcal{W}^{-1}(\boldsymbol{S}, \boldsymbol{\theta})$ (c) $M_S(\boldsymbol{\beta})$



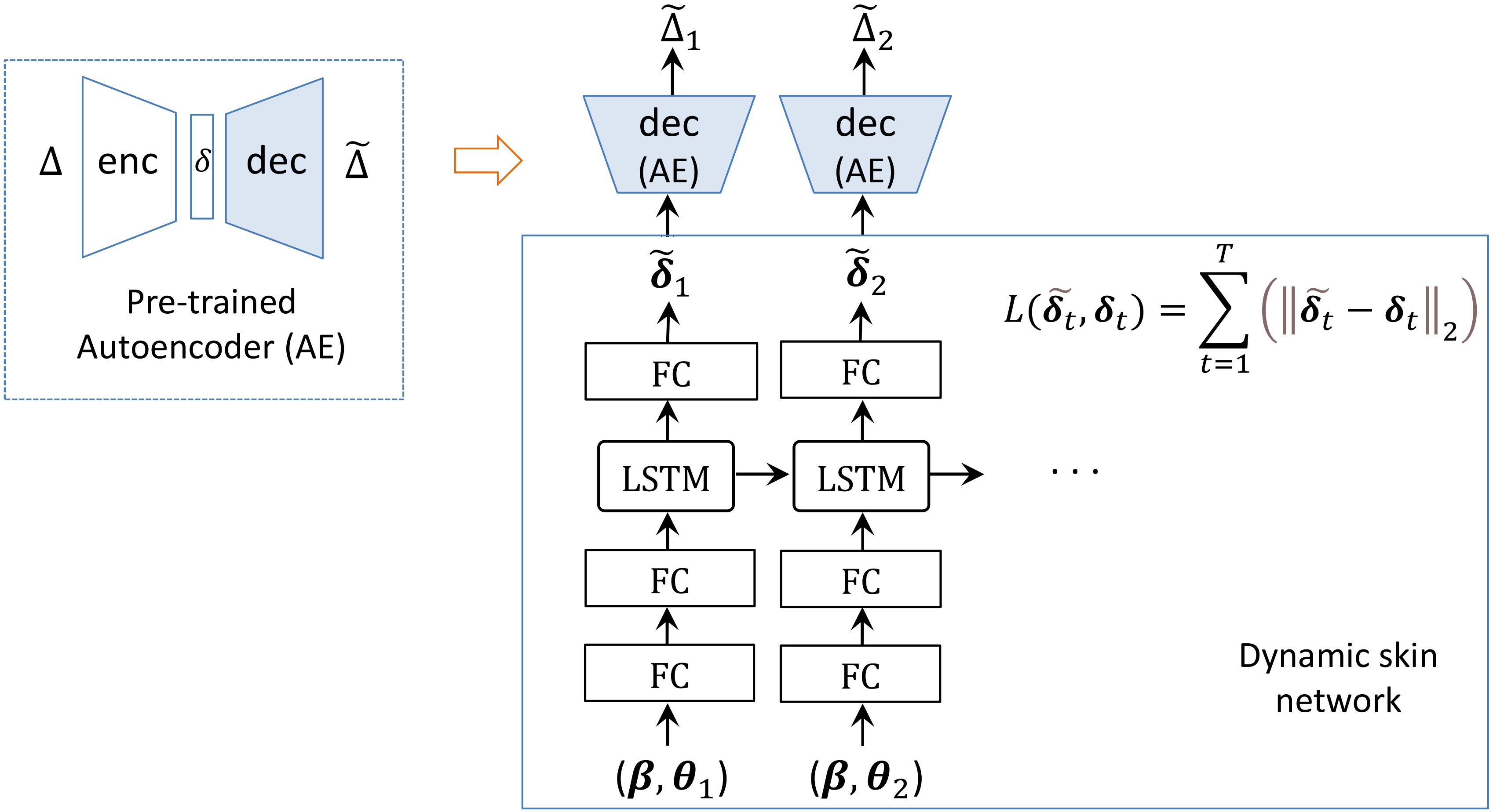
(d) $M_S(\boldsymbol{\beta}) + \Delta$

Skin offsets contributed by the dynamic skin deformation are recorded at a canonical pose $\boldsymbol{\theta}^0$.

DS-Net : network architecture

DSNet: Dynamic skin prediction

- The original data space resides in a high dimensional space: $\Delta_t \in R^{N\times 3}$ ($>$ 20 000)
- We represent them in a latent space by using an autoencoder: $\delta_t \in R^{100}$
- The DSNet LSTM is trained on the latent space



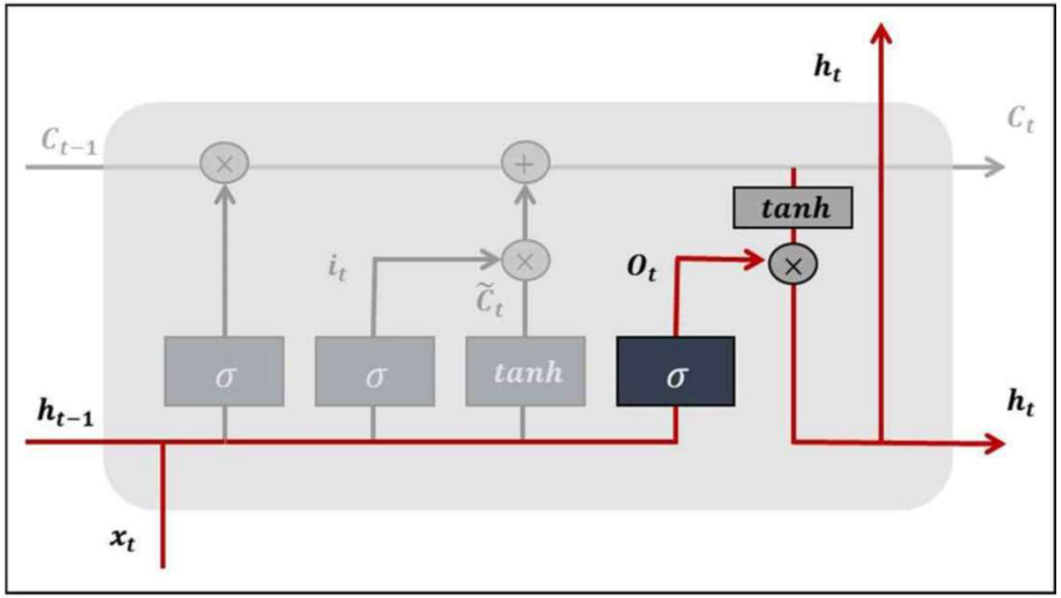
DS-Net : LSTM

Long Short Term Memory network [HS97]

- **It's an RNN, network with recurrent edges**
- **One or more layer is connected to itself**
  - Self connections allow the network to build an **internal representation** of past inputs
  - In effect they serve as network **memory**

Our function

$$\Delta_t = f(\boldsymbol{x}_t, f(\boldsymbol{x}_{t-1}))$$



$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right)$$

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

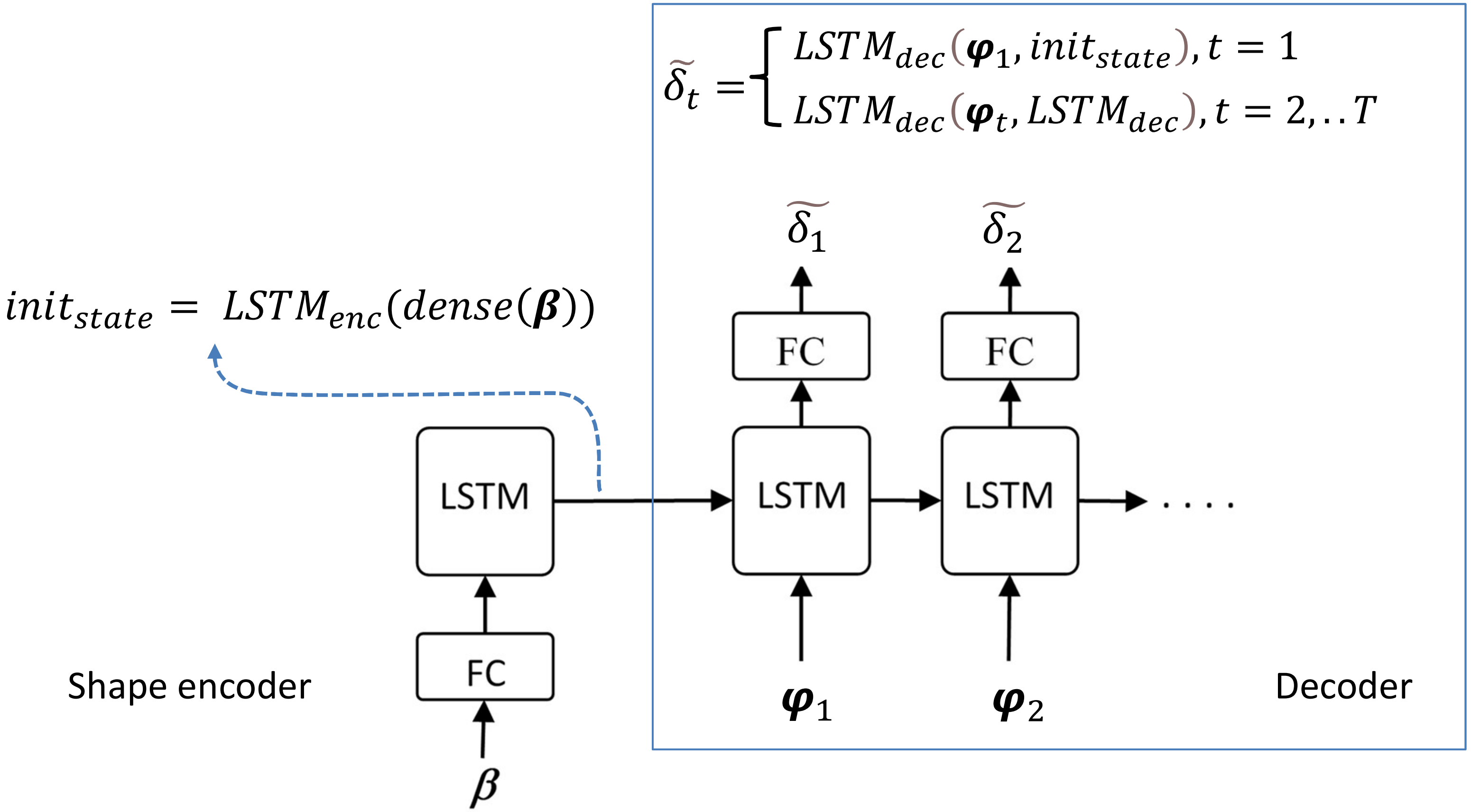
$$o_t = \sigma\left(W_o \left[h_{t-1}, x_t\right] + b_o\right)$$

$$h_t = o_t * \tanh\left(C_t\right)$$

[HS97] Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

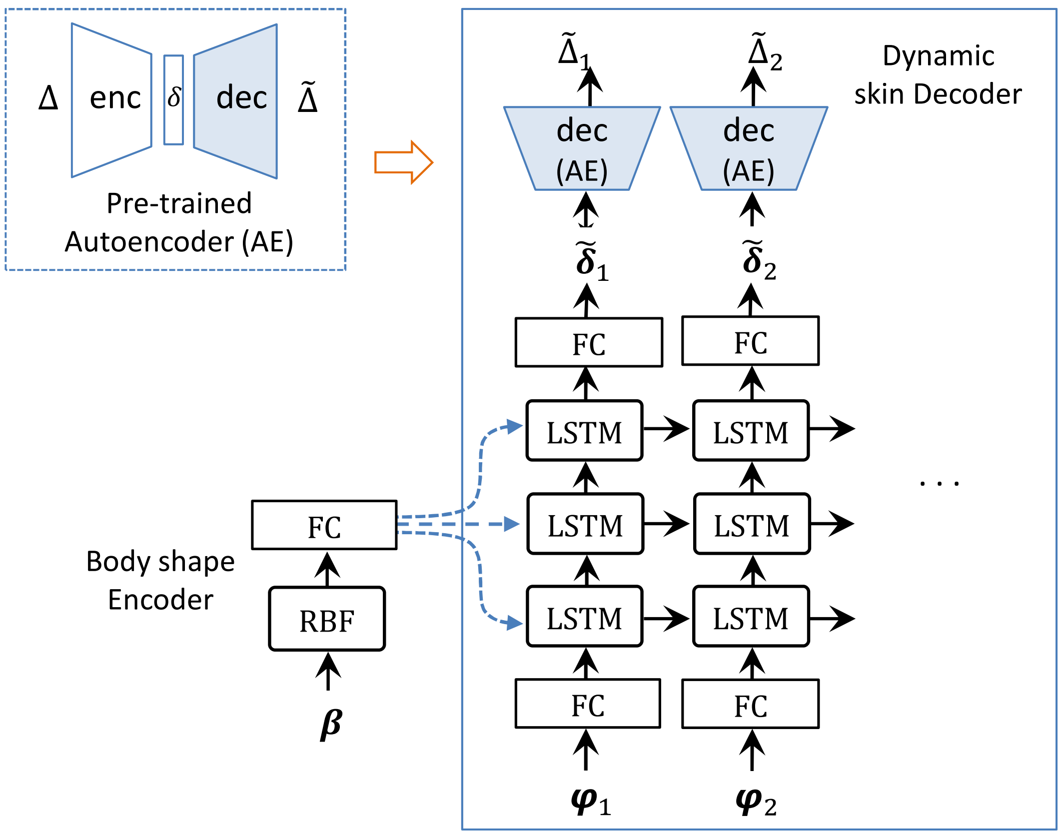
DS-Net : network architecture

DSNet: Earlier versions II



DS-Net : network architecture

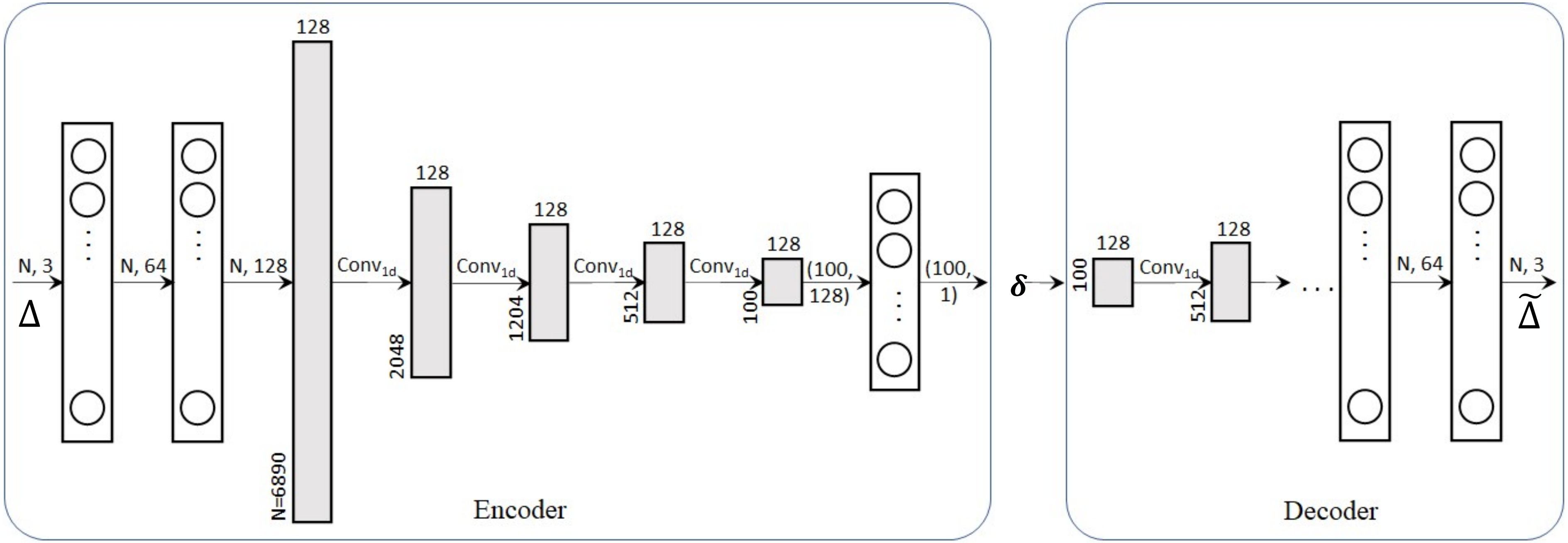
DSNet: Earlier versions I



$$\boldsymbol{\varphi}_t = (\boldsymbol{v}_t, \boldsymbol{a}_t, \dot{\boldsymbol{\theta}}_t, \ddot{\boldsymbol{\theta}}_t)$$

DS-Net : data dimension reduction

Mesh autoencoder (AE):



: dense layer

: data

It reduces the dimension of the original mesh 3N (3×6890= 20,670) to **100!!**

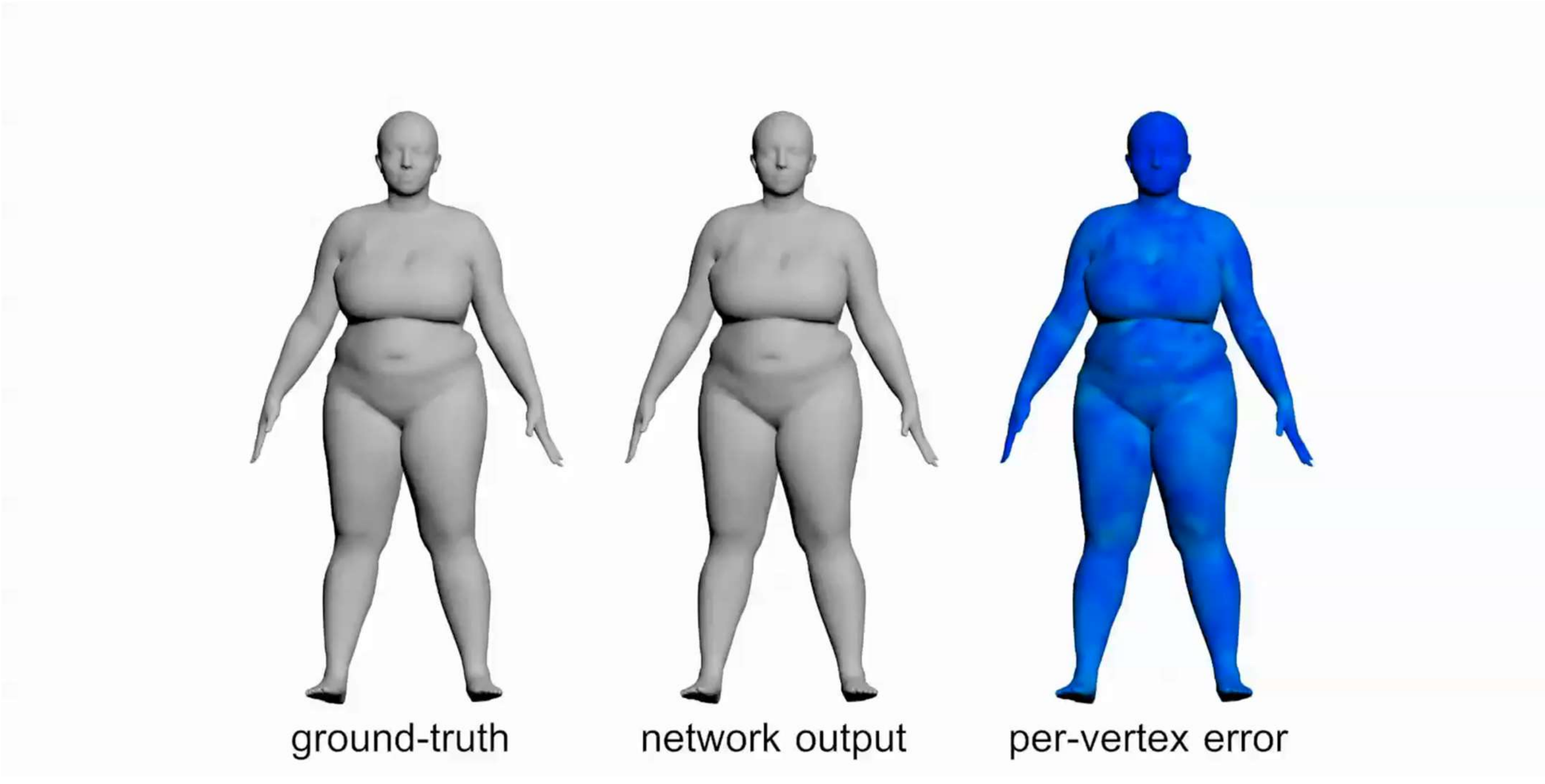
DS-Net : AE details

Mesh autoencoder (AE):

- The input data Δ has been normalized to [-1,1].
- Pytorch implementation of Adam optimizer.
- Batch size 64, learning rate 0,0001.
- 11,8% of network parameters, compared to the other AE.
  => much more efficient to train!!

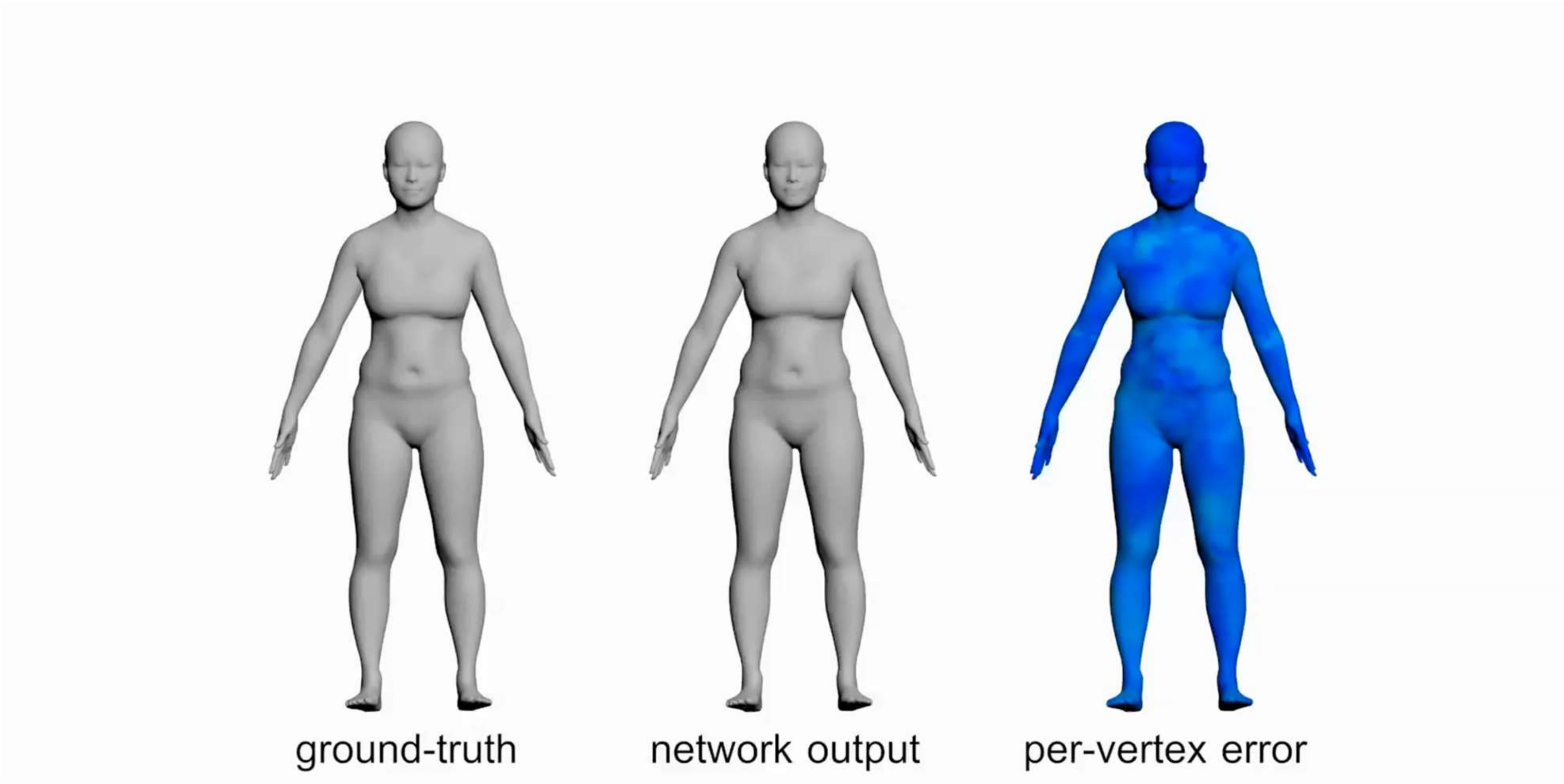
DS-Net : AE results

Reconstruction results: min 0 cm, max 1.033 cm



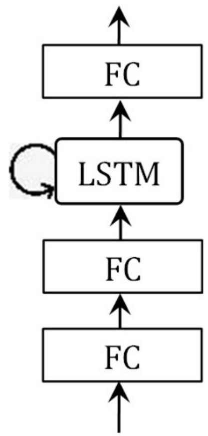
DS-Net : AE results

Reconstruction results: min 0 cm, max 1.000 cm



DS-Net : details

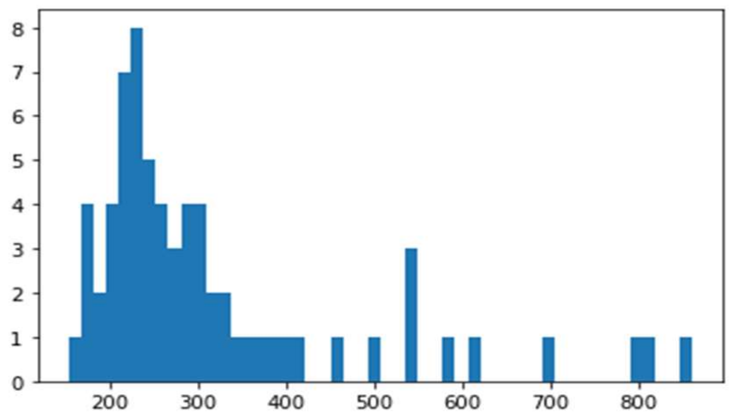
Implementation details



- **Tensorflow 2.0 implementation of Adam optimizer**
- **3rd dimensions of output vectors: 64, 128, 60, 100**
- **Activation functions: linear, tanh, (bath normalization), linear**
- **Batch size=16, lr= 0.0001.**
- **0.05 sec/epoch on a Ubuntu machine with Nvidia GeForce RTX 2080 Super**

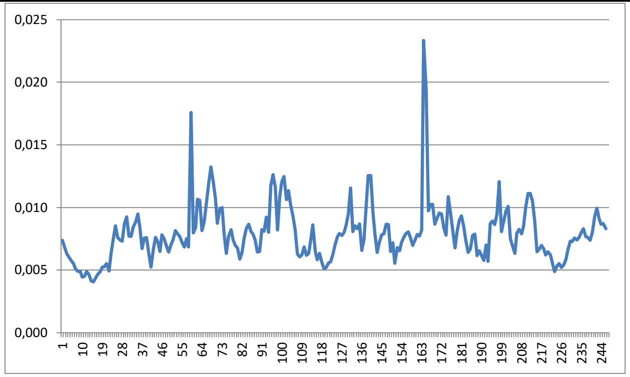
Data preprocessing

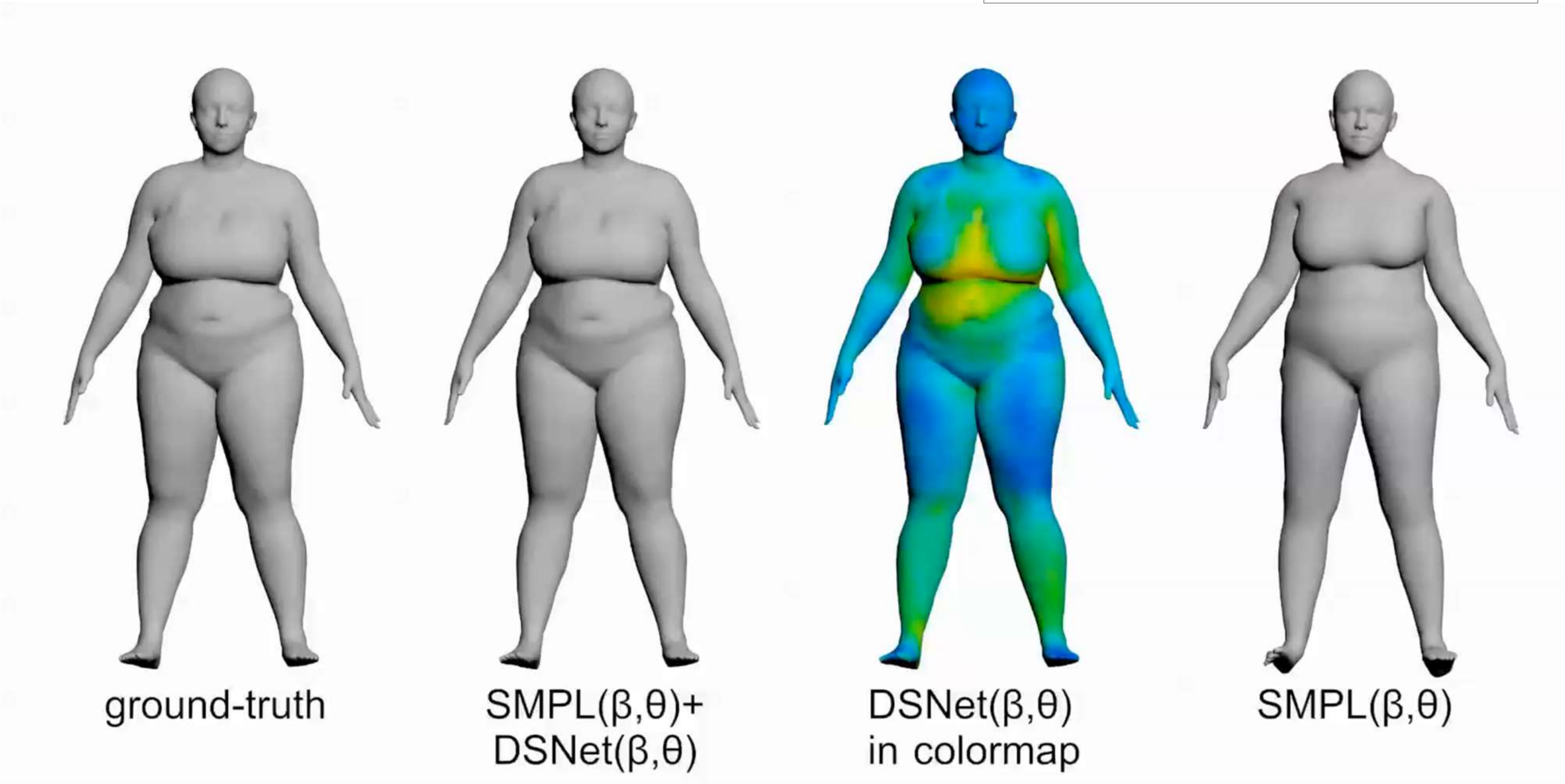
- **Uniformize the frame lengths (to 300) by zero padding or tail clipping.**



DS-Net : Results

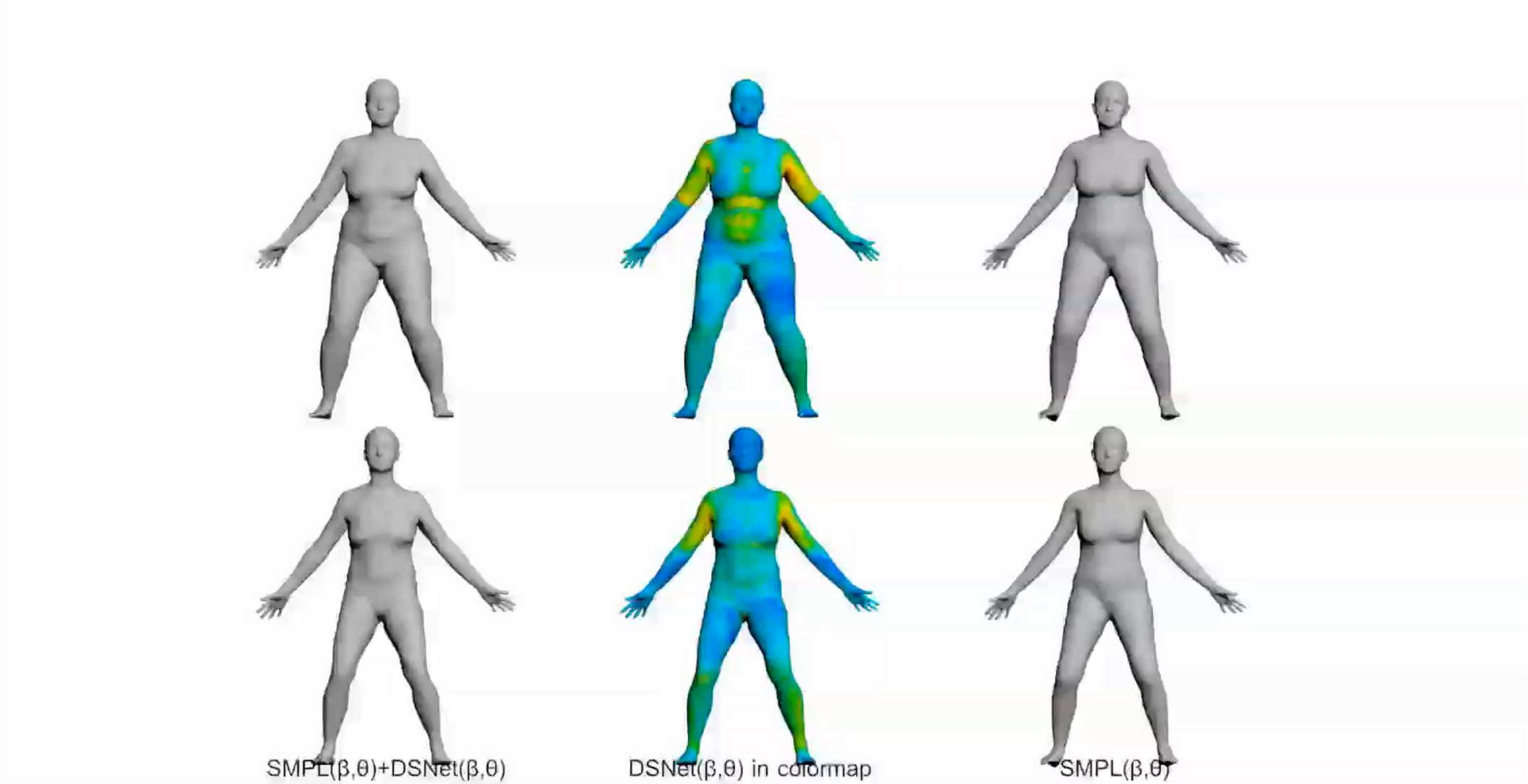
On validation data:





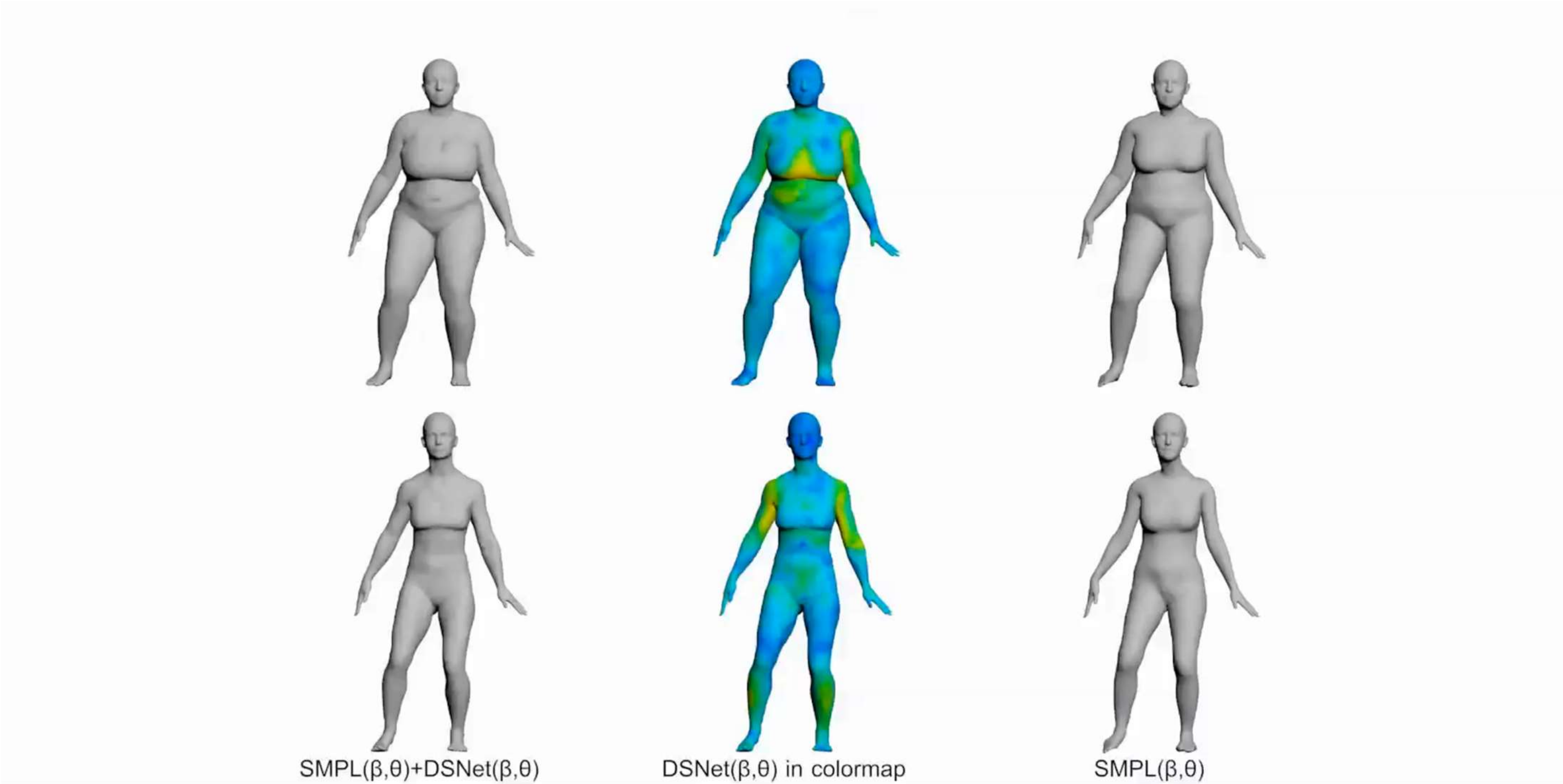
DS-Net : Results

On validation data:



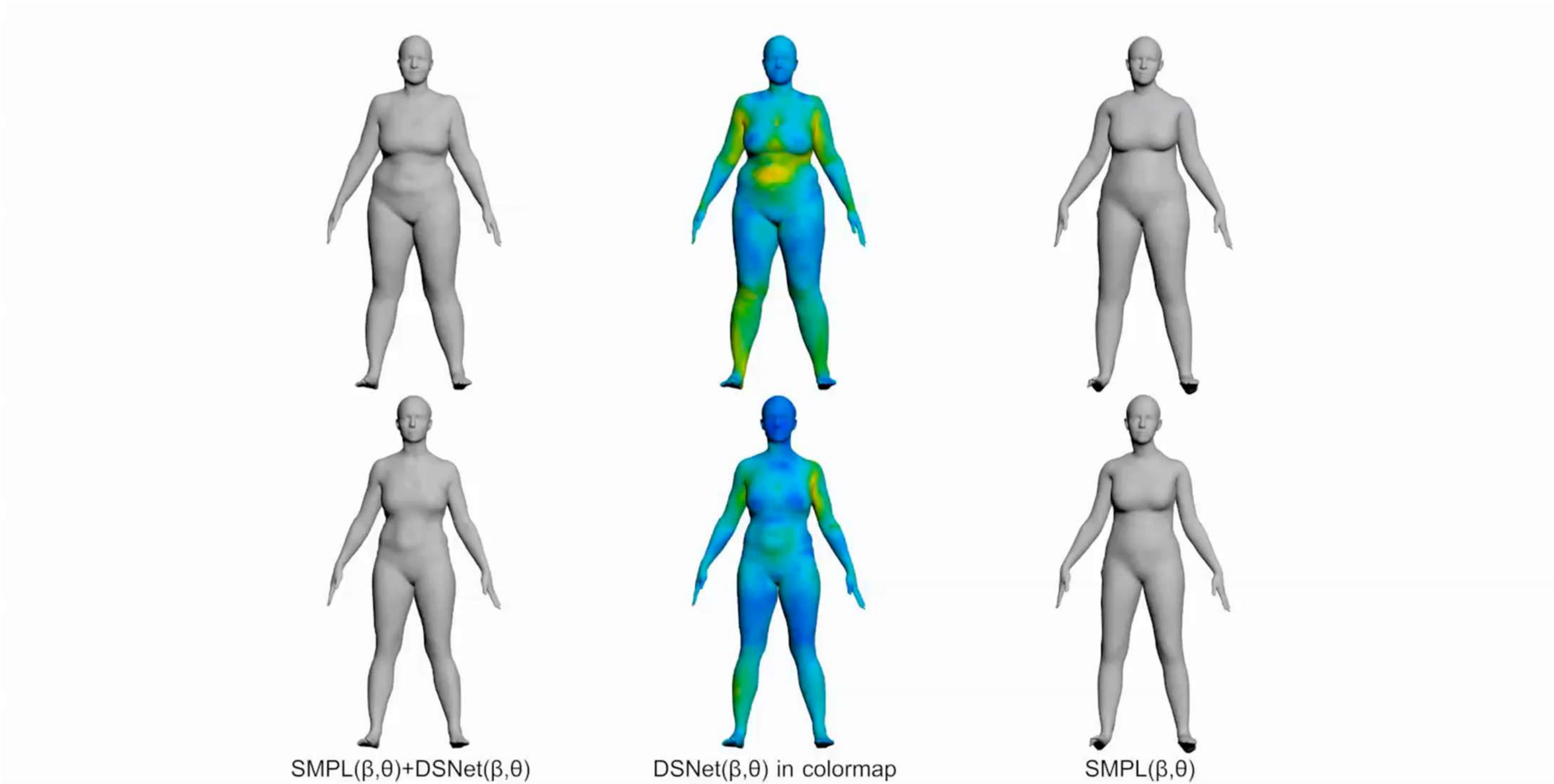
DS-Net : Results

On validation data:



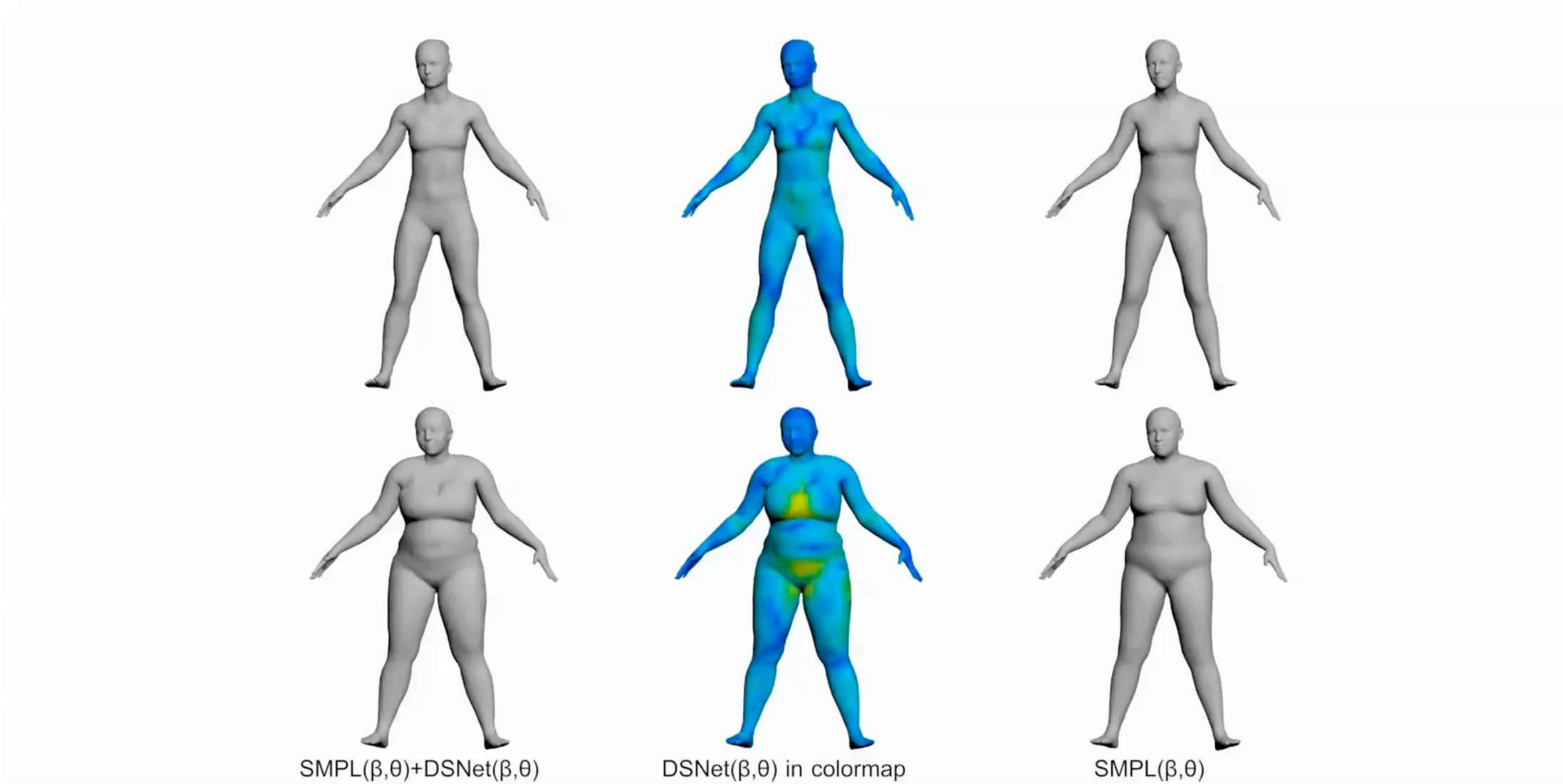
DS-Net : Results

On validation data:



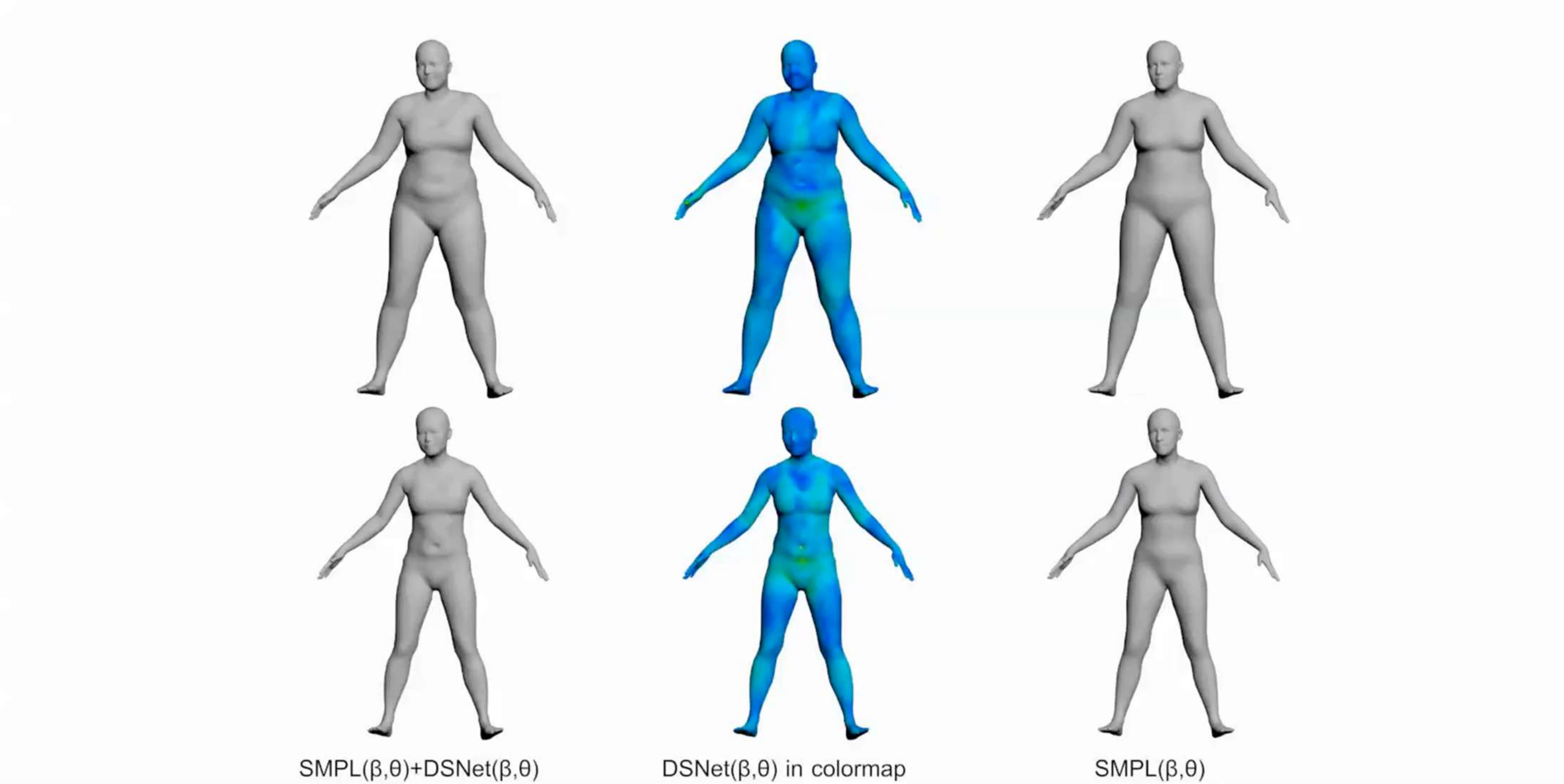
DS-Net : Results

On unseen motions :



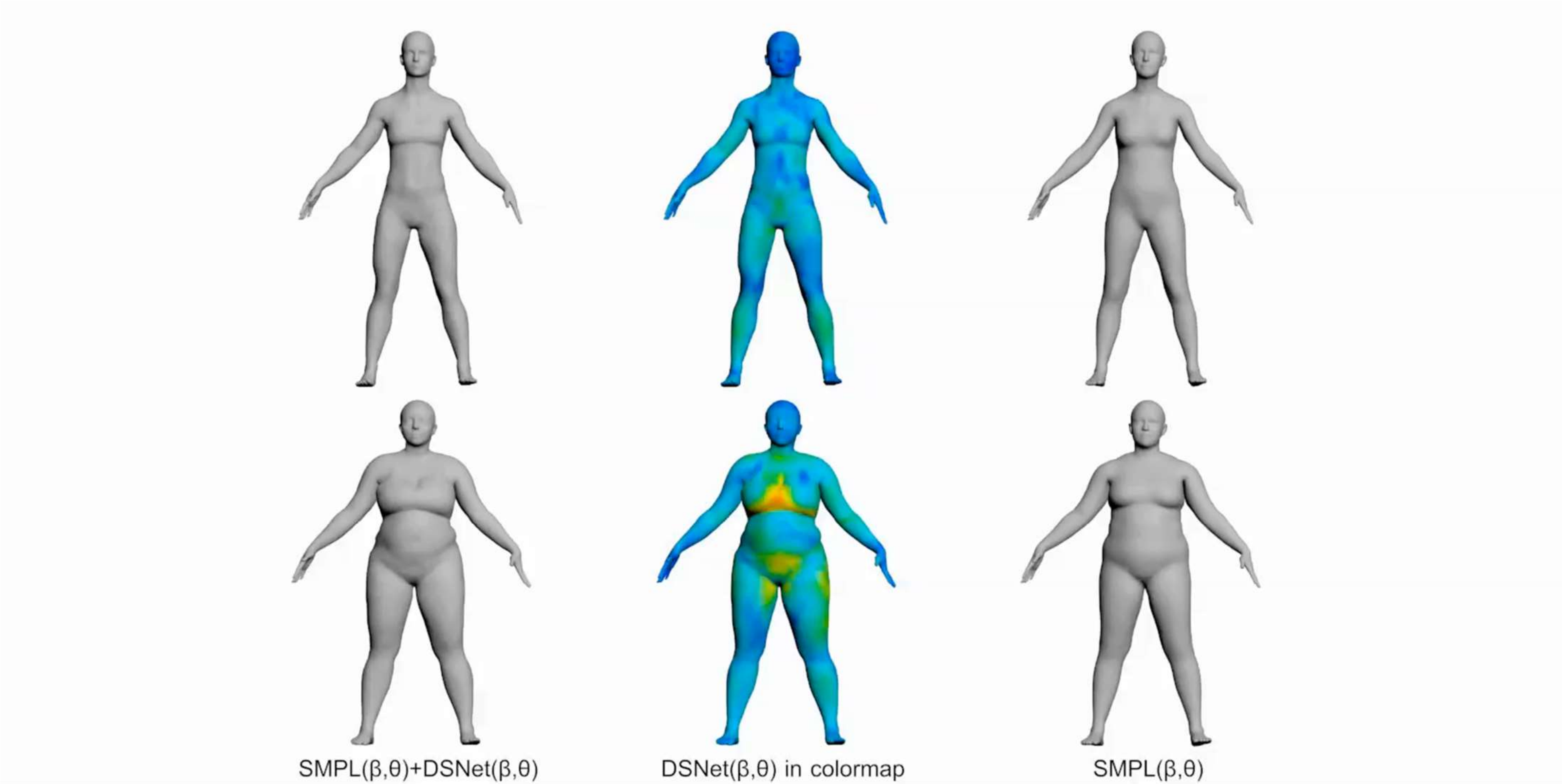
DS-Net : Results

On unseen motions :



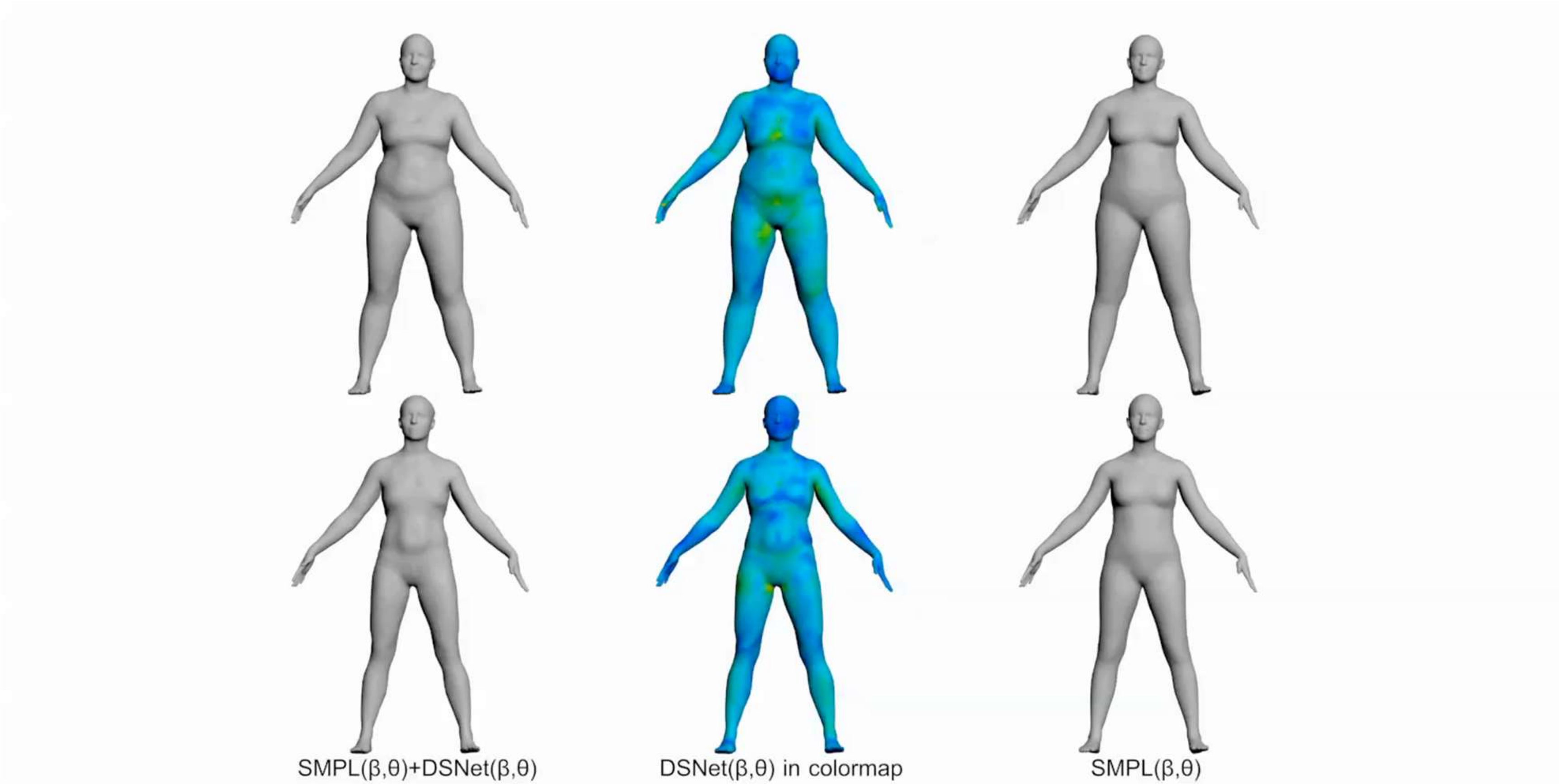
DS-Net : Results

On unseen motions :



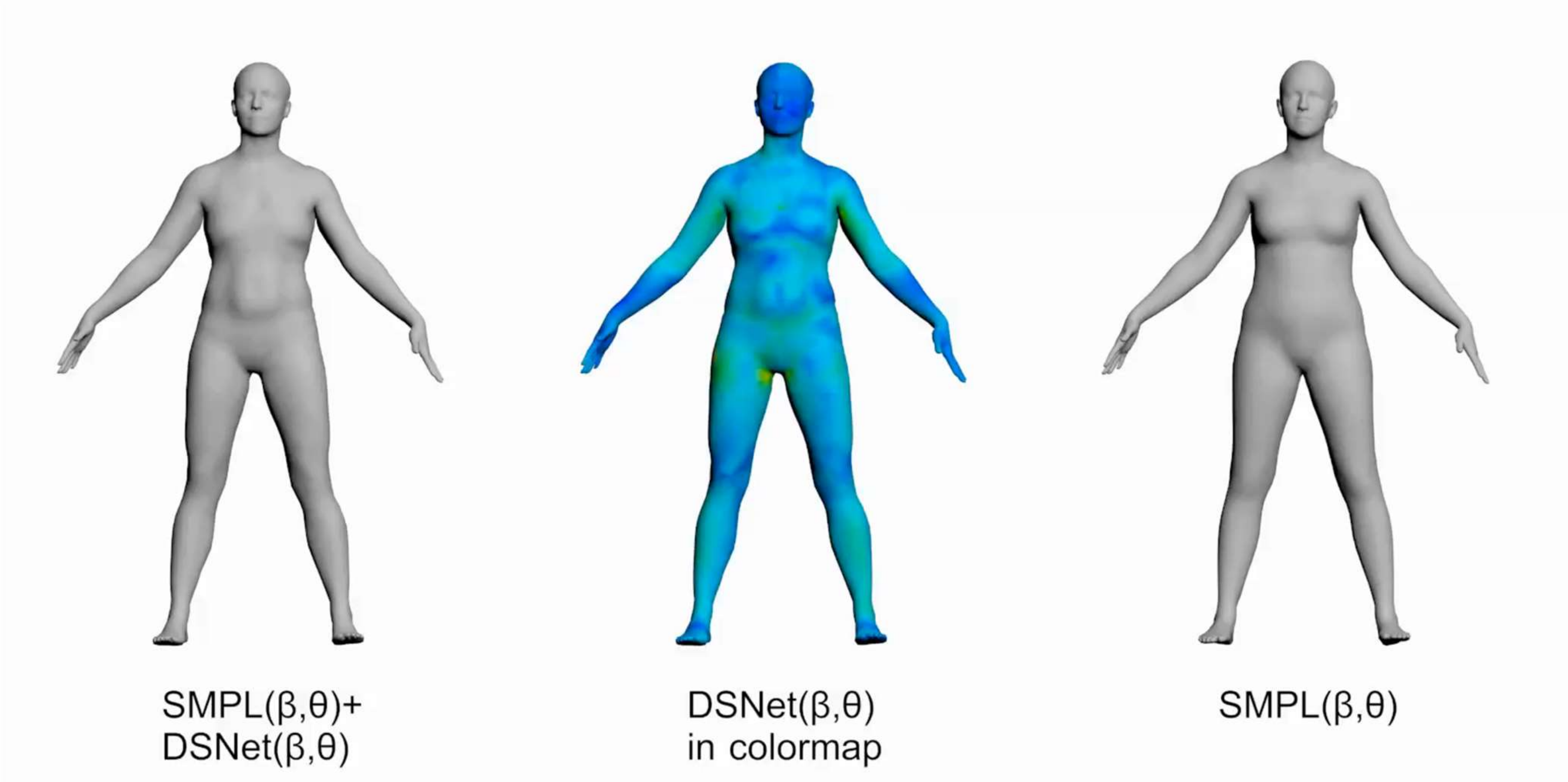
DS-Net : Results

On unseen motions :



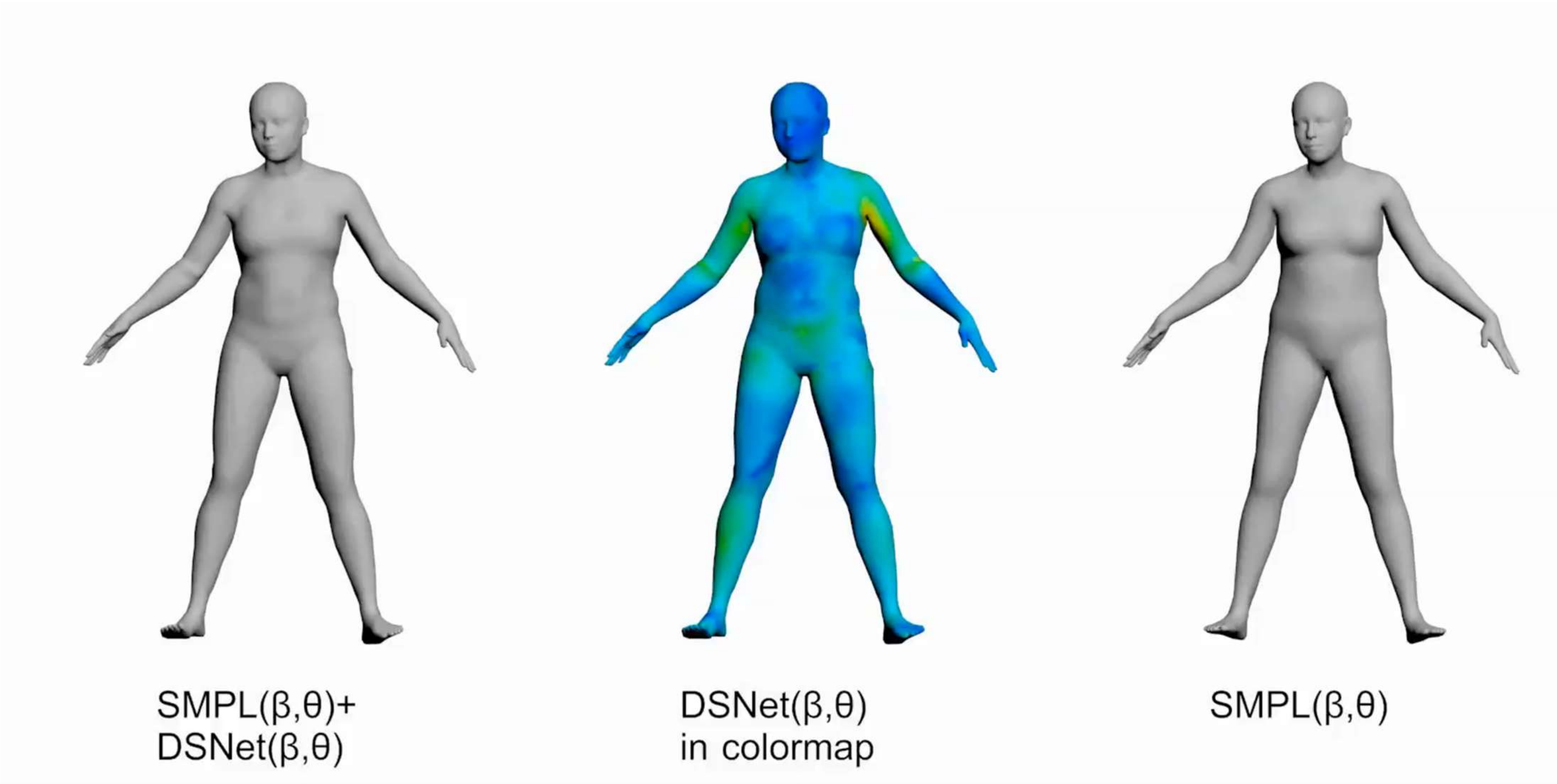
DS-Net : Results

On unseen motions & unseen subjects:



DS-Net : Results

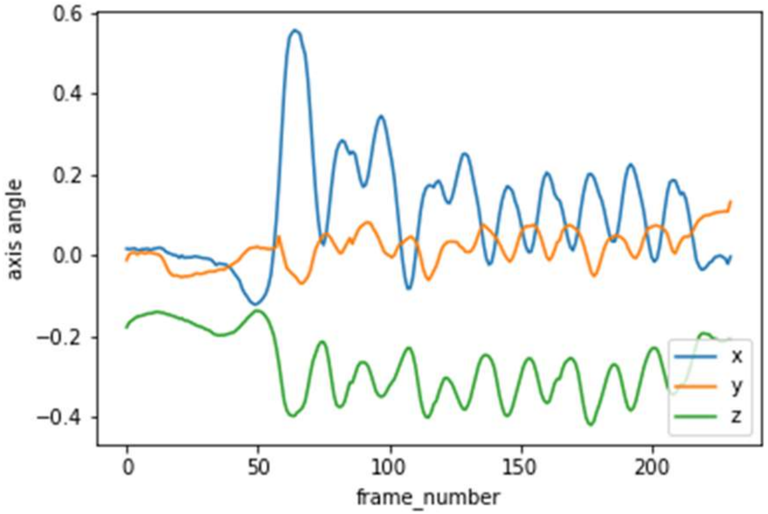
On unseen motions & unseen subjects:



Conlusion

A note on the training data

- **We observed that the dynamics dependent shapes had been partly absorbed by the pose-dependent shape..!!**



- 'spine 2' joint angles during 'Jiggling on toes' motion

- **This means that our training data do not fully capture the observed dynamics…**

Conlusion

- A learning based method to the estimation of quality dynamic skin deformation.

- The dynamic skin deformation has been modeled as a time series data, as a function of pose, body shape, and the results of previous time steps.

  => An LSTM based NN has been developed, trained on sequences of triangular meshes captured from real people.

- Also developed has been an AE, which builds a compact space for the intrinisic representation of DS offset, allowing a very efficient operation of the DSNet.

Many thanks to…

Kaifeng ZOU, Frederic CORDIER