DeepCap:
Monocular Human Performance Capture Using Weak Supervision

Marc Habermann, **Weipeng Xu**, Michael Zollhoefer,

Gerard Pons-Moll, and Christian Theobalt

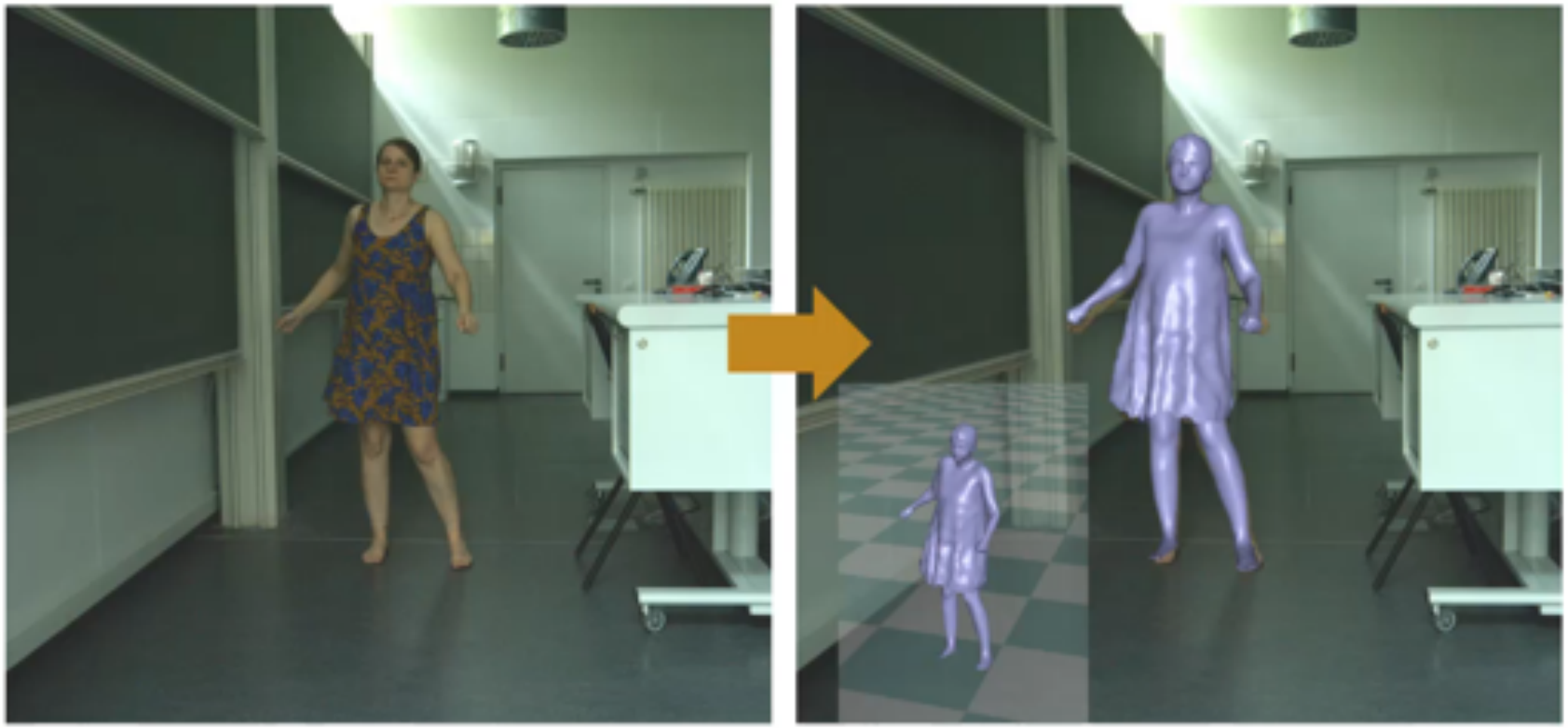
Graphics Vision & Video

max planck institut informatik

Stanford University

SIC Saarland Informatics Campus

DeepCap



Human performance capture from a monocular camera

Challenges

- **Monocular** setting is inherently ambiguous

- High-dimensional problem
  - Pose **and surface deformation**

Source: https://www.fiylo.de/

Related Work

- Capture using parametric models

Kanazawa et al. 2018

Xiang et al. 2018

Metaxas et al. 1993, Plaenkers et al. 2001, Sminchisescu et al. 2003, Sigal et al. 2004, Joo et al. 2018, Pavlakos et al. 2018, Kanazawa et al. 2019, Pavlakos et al. 2019, …

Related Work

- Monocular template-free capture

Zheng et al. 2019

Saito et al. 2019

Huang et al. 2018, Varol et al. 2018, Natsume et al. 2019, …

Related Work

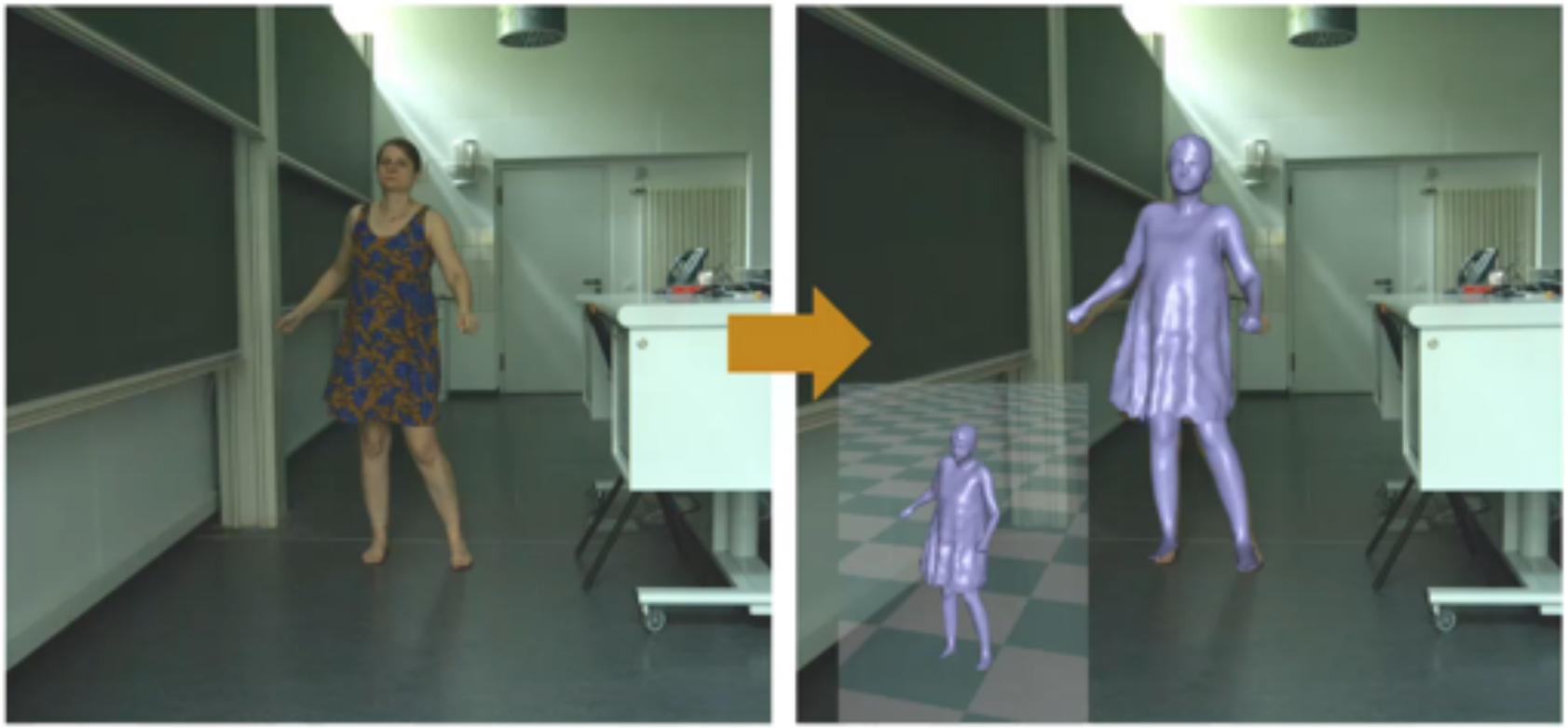
- Template-based capture

Habermann et al. 2019

Xu et al. 2018

Carranza et al. 2003, Bray et al. 2006, Starck et al. 2007, De Aguiar et al. 2008, Brox et al. 2010, Cagniart et al. 2010, …

DeepCap



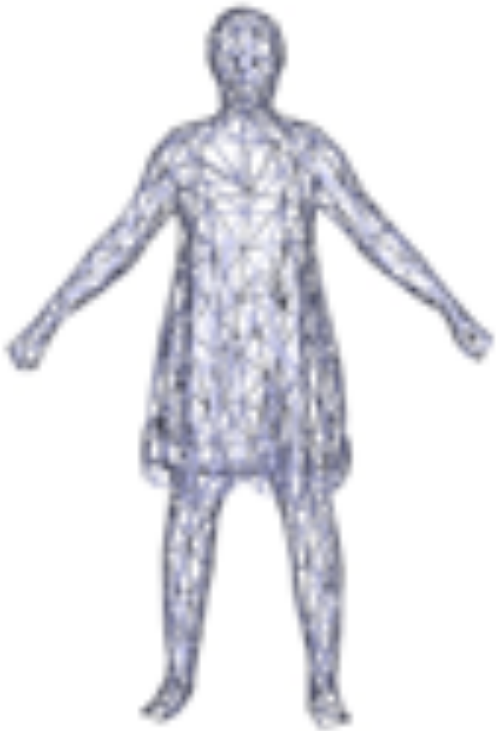
Learning based approach

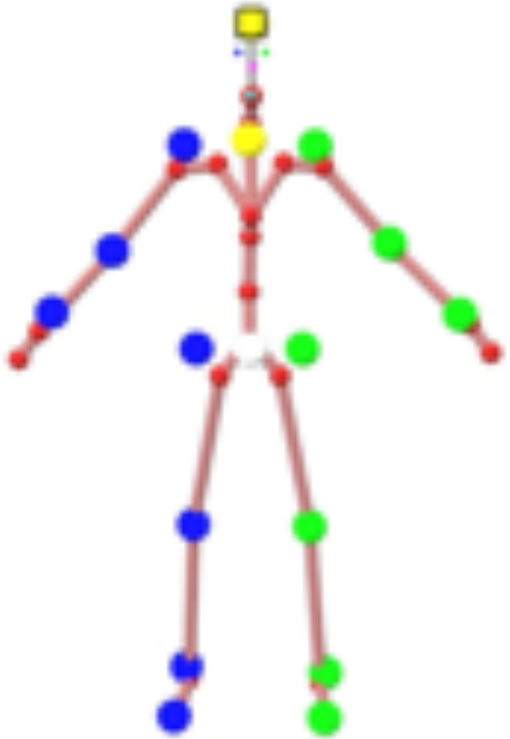
Pose + **surface deformation**

Weak multi-view supervision

Personalized Character Model

Fully automatic



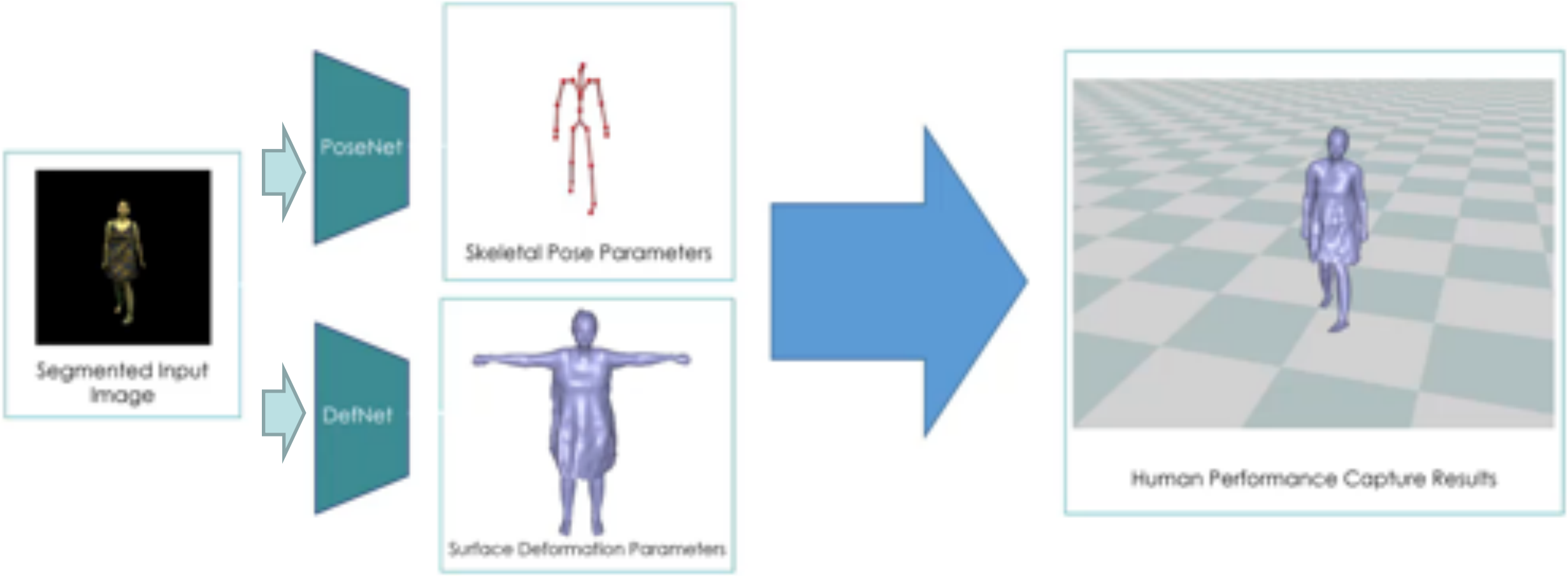


Template mesh

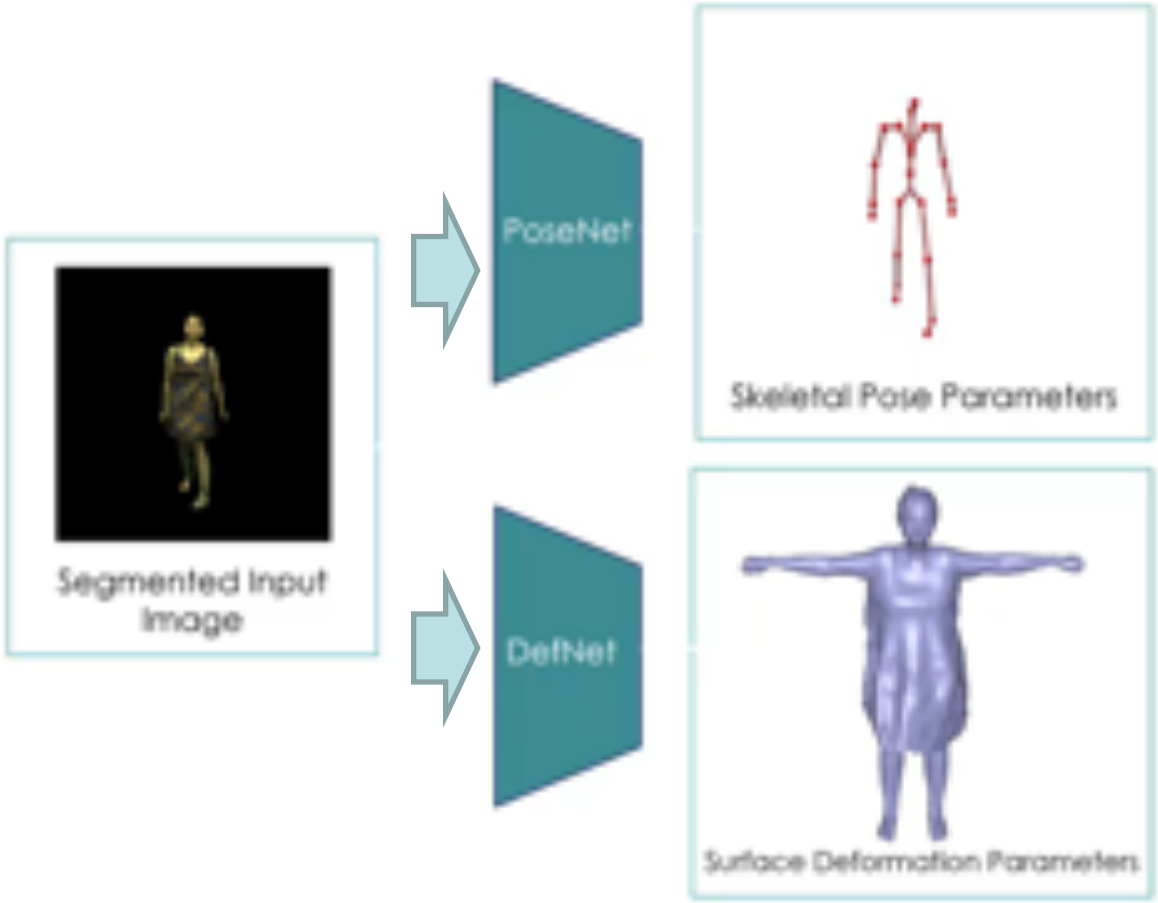
Embedded graph

Skeleton

Inference Time



Direct Supervision?

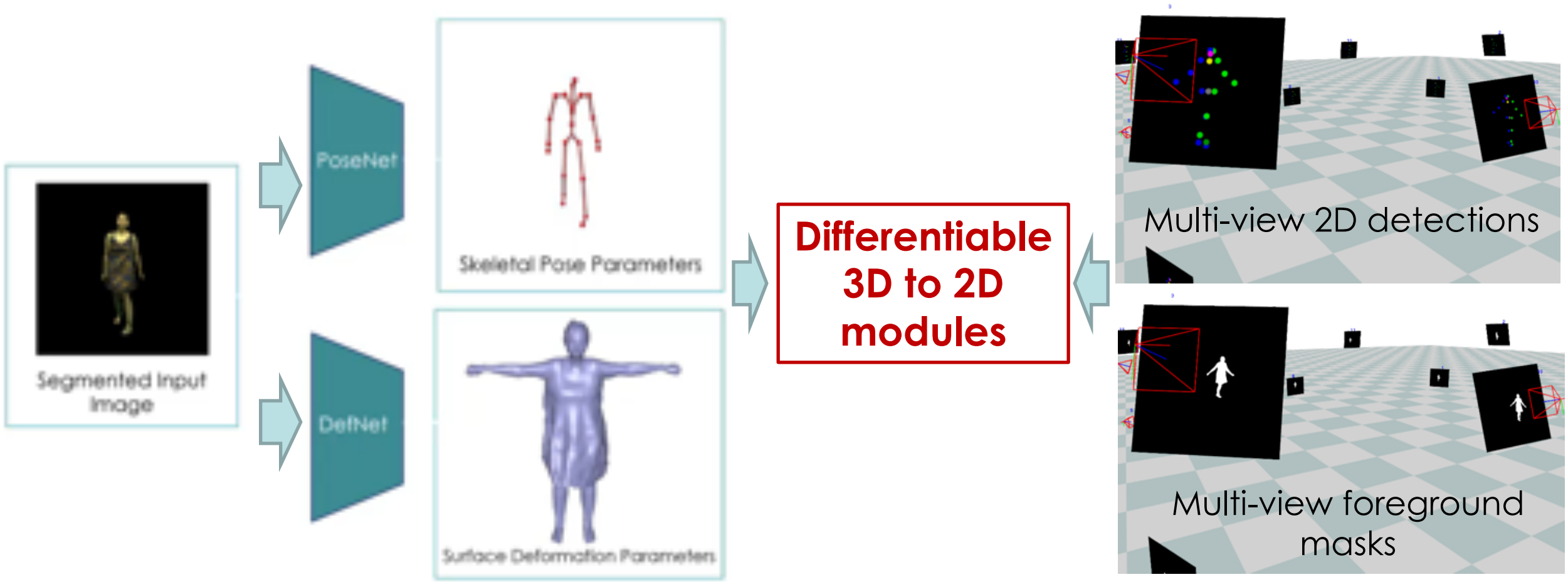


Ground truth 3D pose

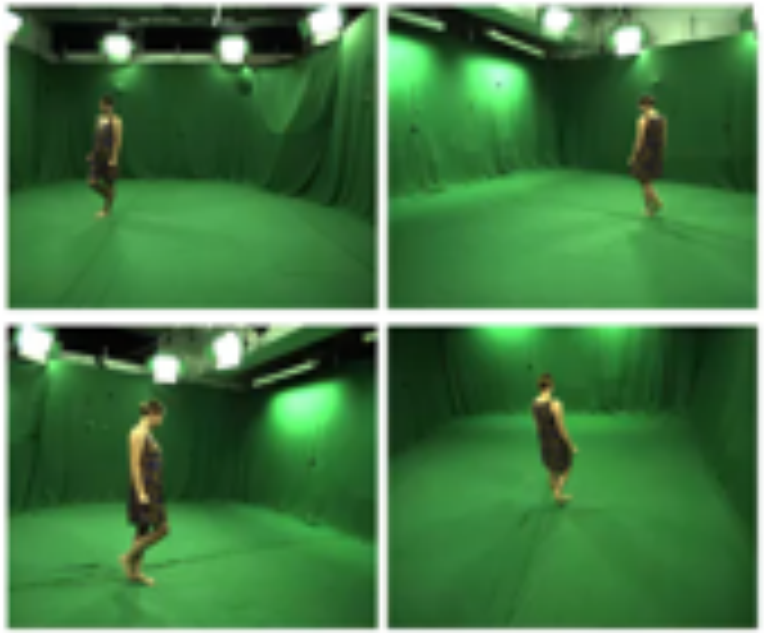
Difficult to obtain

Ground truth 3D surface

Weak Supervision

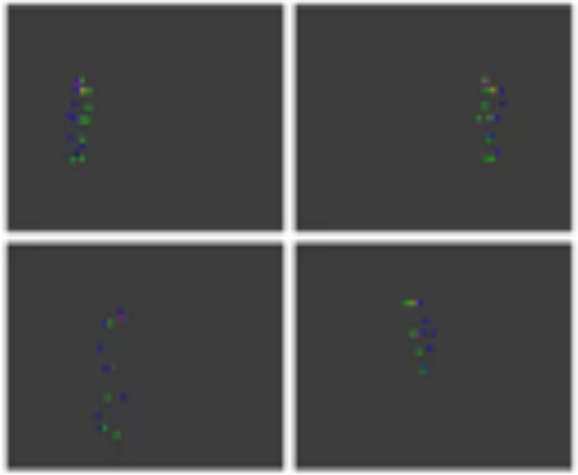


Training Data – Weak Multi View



Calibrated multi-view images

OpenPose
(Cao et al. 2019)

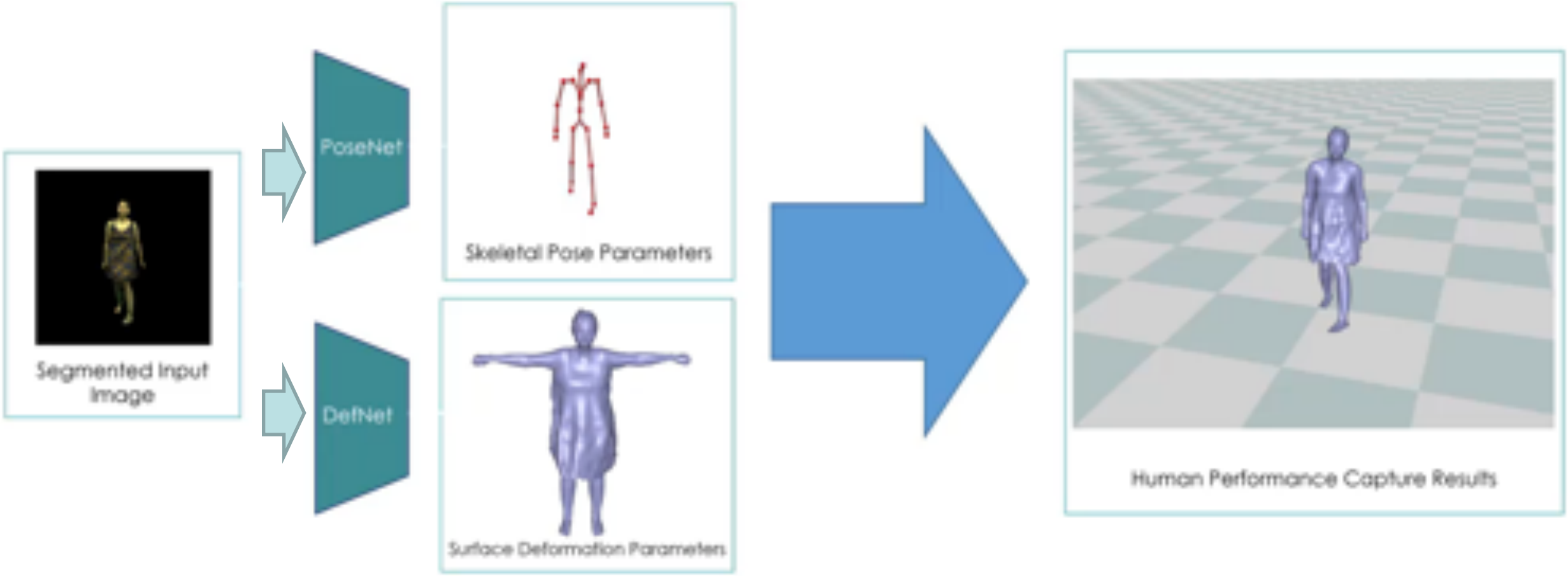


2D keypoints

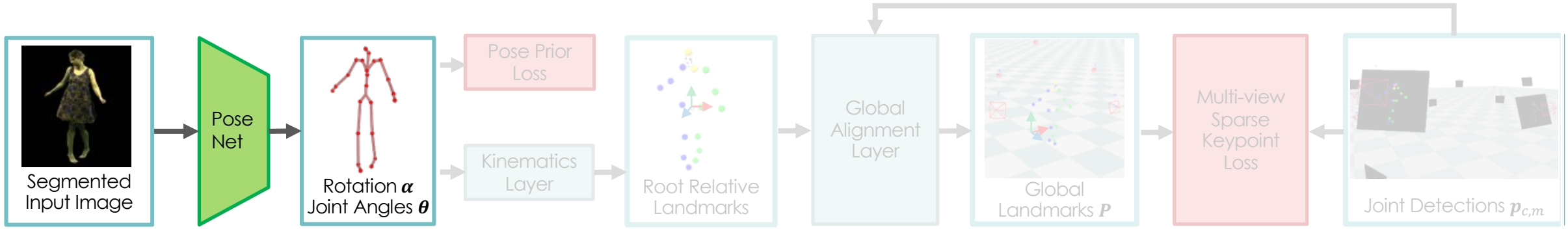
Color keying

Foreground mask

Pipeline


PoseNet
DefNet
Segmented Input Image
Skeletal Pose Parameters
Surface Deformation Parameters
Human Performance Capture Results

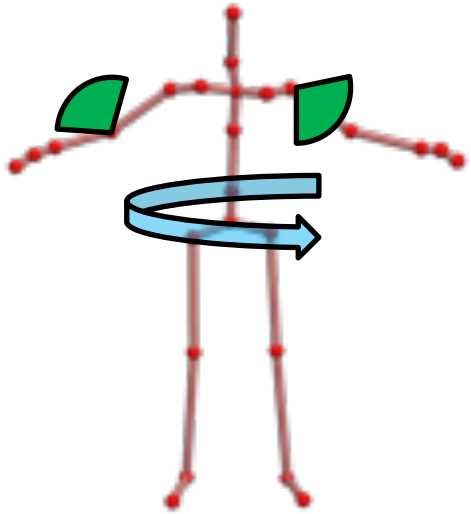
PoseNet



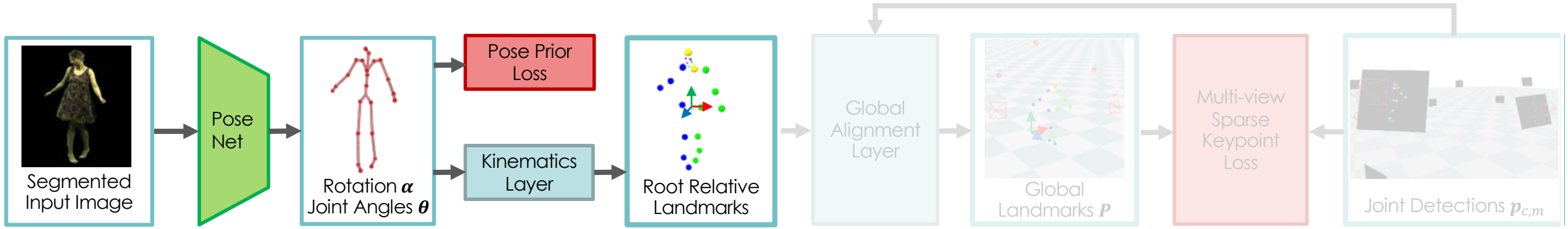
PoseNet

Root rotation $\boldsymbol{\alpha} \in \mathbb{R}^3$

Joint angles $\boldsymbol{\theta} \in \mathbb{R}^3$



PoseNet

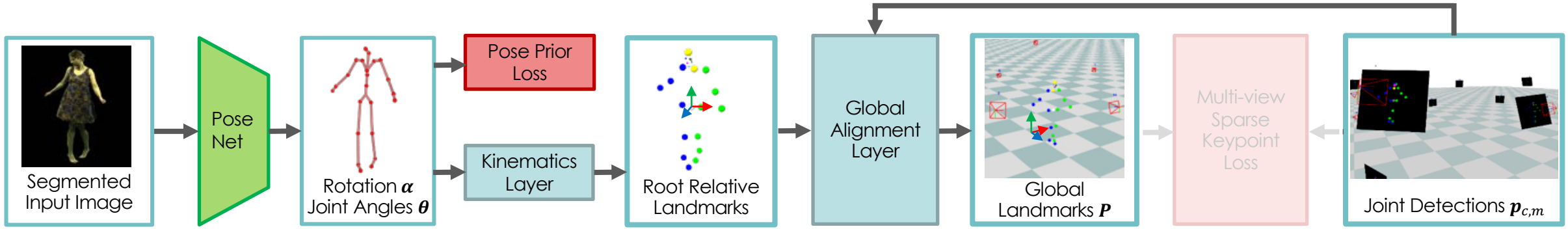


Kinematics Layer

Function $f_m(\boldsymbol{\alpha}, \boldsymbol{\theta})$: $\mathbb{R}^{\mathbf{30}} \rightarrow \mathbb{R}^{\mathbf{3}}$ per landmark m

Skeletool pose → **Camera** and **root relative** 3D landmark positions $\boldsymbol{P}_{c',m}$

PoseNet



Rigid transform for landmark $P_{c',m}$

Camera and **root relative**

3D space

→

Global

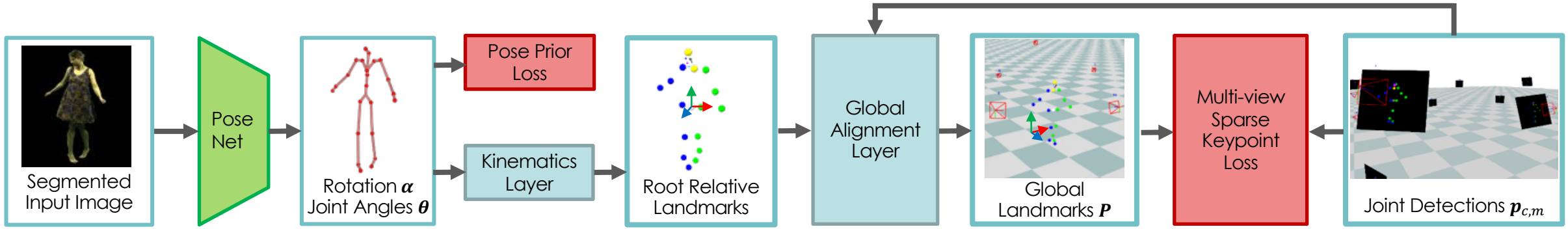
3D space

$$\boldsymbol{P}_m = \boldsymbol{R}_{c'}^T \boldsymbol{P}_{c',m} + \boldsymbol{t}$$

Inverse extrinsic rotation of the input camera c'

Global translation

PoseNet



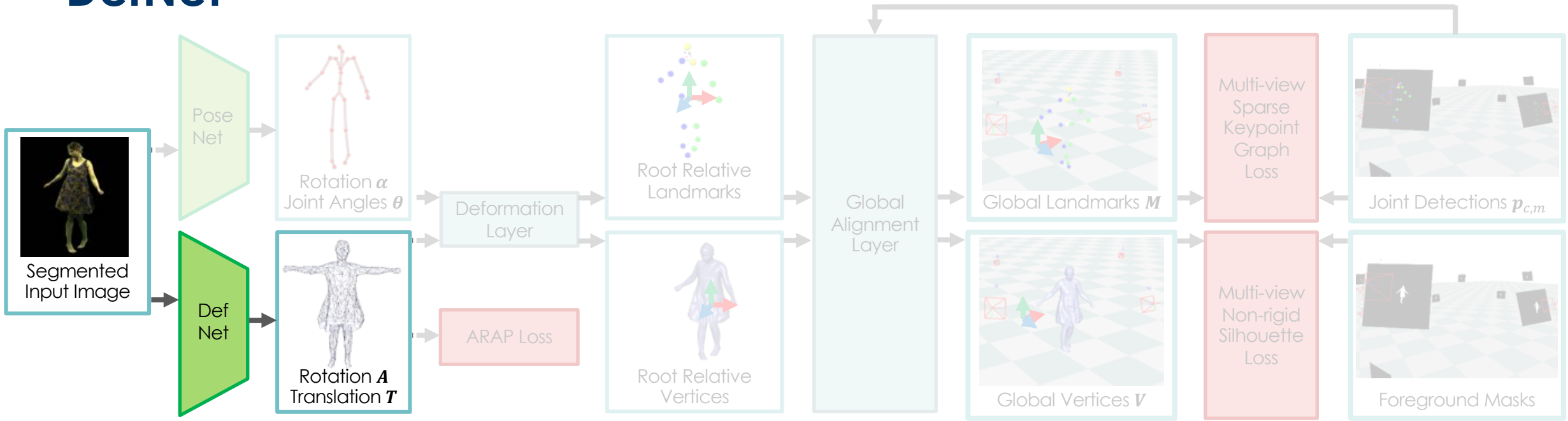
Multi-view Sparse Keypoint Loss

$$L_{kp}(\boldsymbol{P}) = \sum_{c}\sum_{m}\left\|\pi_c(\boldsymbol{P}_m) - \boldsymbol{p}_{c,m}\right\|_2^2$$

Projecting (π) 3D landmark $\boldsymbol{P}_m$ into camera view c

Comparing to 2D joint detection $\boldsymbol{p}_{c,m}$

DefNet



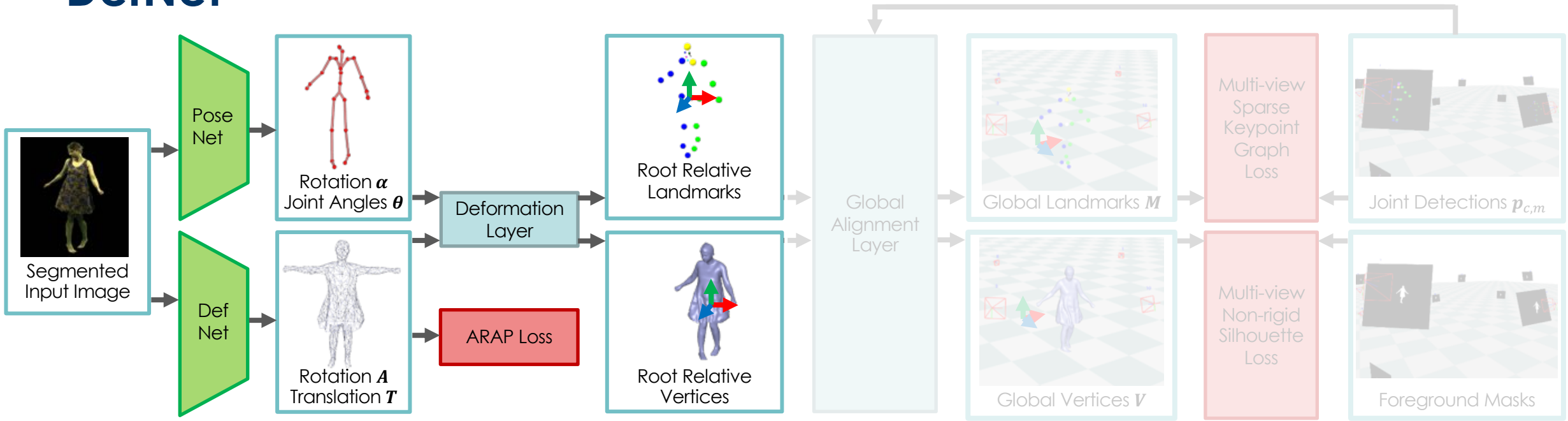
DefNet

Regresses **embedded deformation*** in **canonical pose**

Per node k rotation angles $\boldsymbol{A}_k$ and translation $\boldsymbol{T}_k$

*(Sumner et al. 2007, Sorkine et al. 2007)

DefNet



Pose

Deformation

Deformation Layer

Embedded deformation

Dual Quaternion Skinning

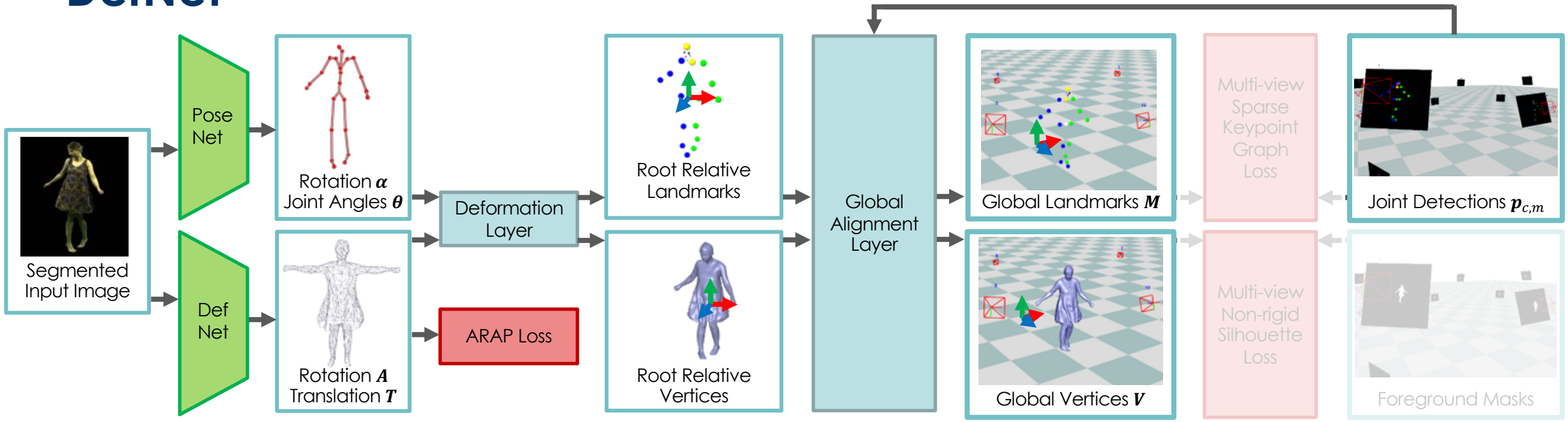
(Kavan et al. 2007)

Posed and **deformed**

Landmarks $\boldsymbol{M}_{c',m}$

Vertices $\boldsymbol{V}_{c',i}$

DefNet



Rigid transform for landmark m and vertex i

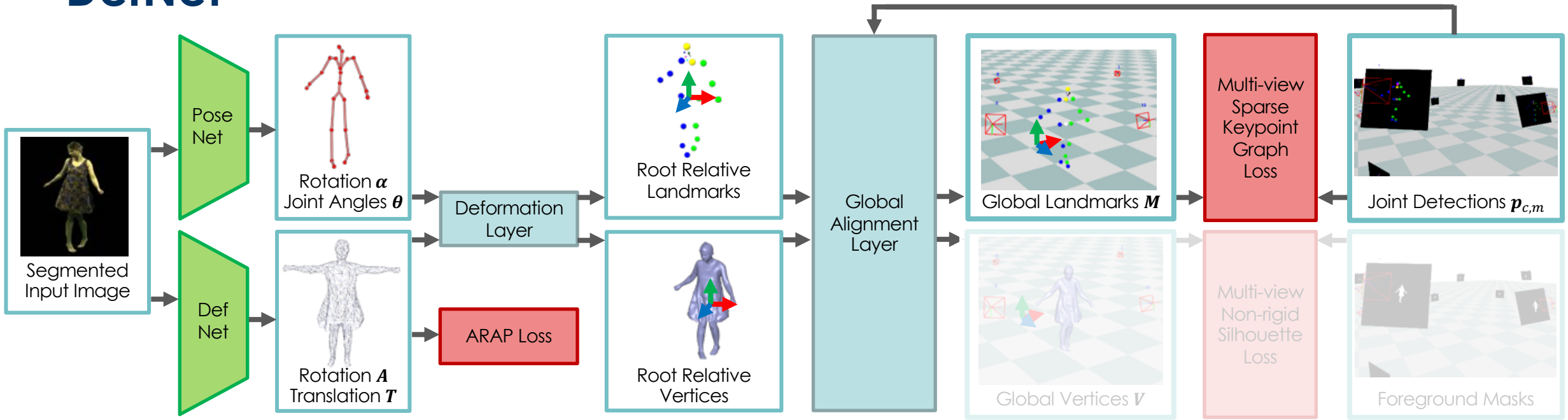
Camera and **root relative**

3D landmark $M_{c',m}$ and vertex $V_{c',i}$

Global

3D landmark M_m and vertex V_i

DefNet

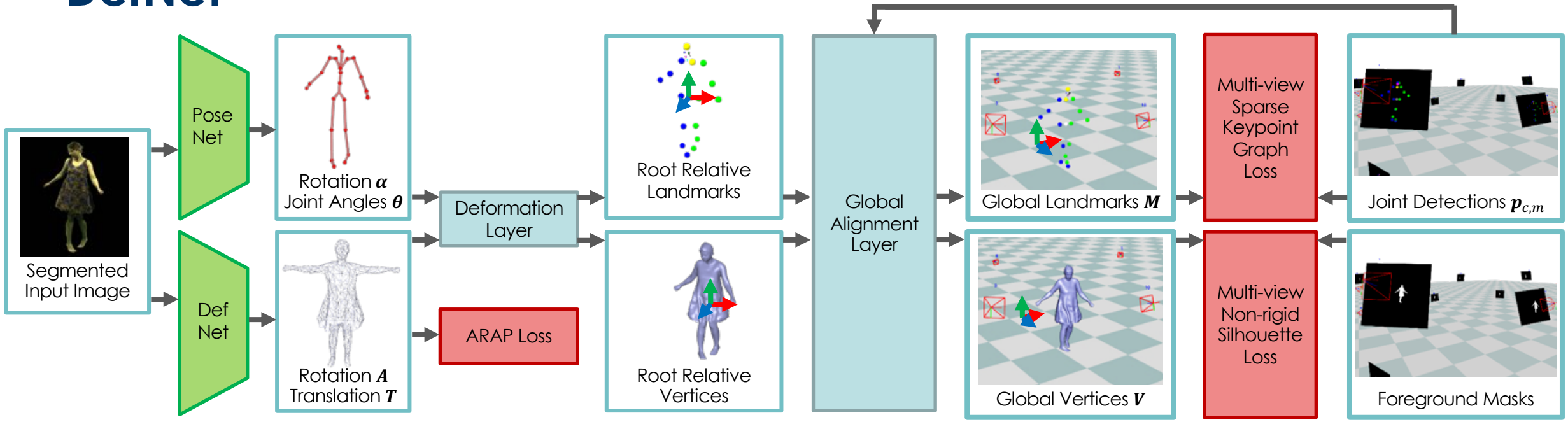


Multi-view Sparse Keypoint Graph Loss

$$L_{kpg}(\boldsymbol{P}) = \sum_{c}\sum_{m}\left\|\pi_c(\boldsymbol{M}_m) - \boldsymbol{p}_{c,m}\right\|_2^2$$

Global 3D landmark $\boldsymbol{M}_m$

DefNet



Non-rigid Silhouette Loss

$$L_{sil}(\boldsymbol{V}) = \sum_{c} \sum_{i \in B_c} \left\| D_c\left(\pi_c(\mathbf{V}_i)\right) \right\|_2^2$$

Set of boundary vertices for camera c

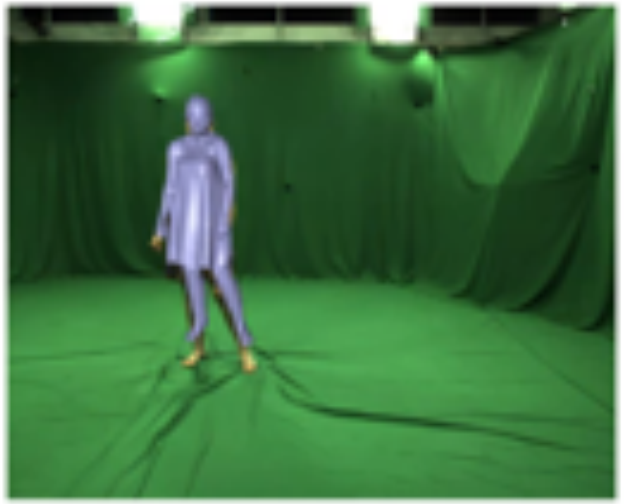
Distance transform image

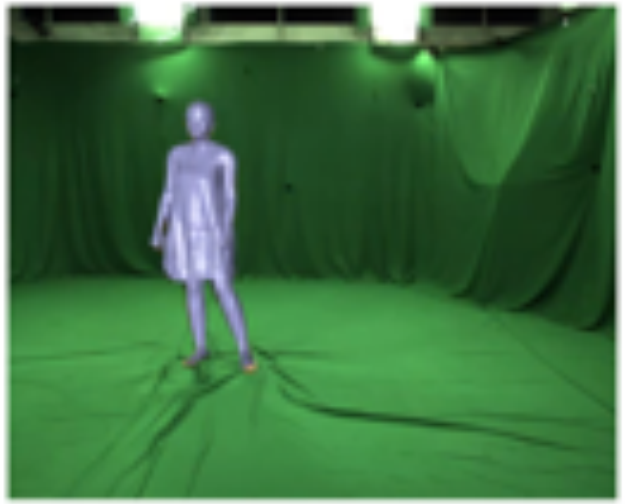
Qualitative Evaluation

Habermann et al. 2019

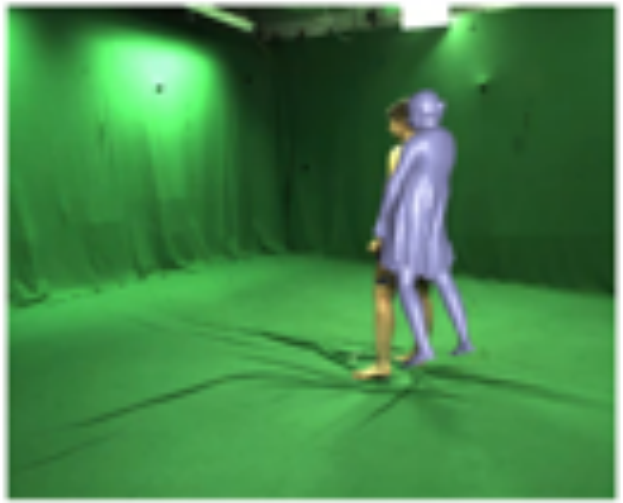
Ours

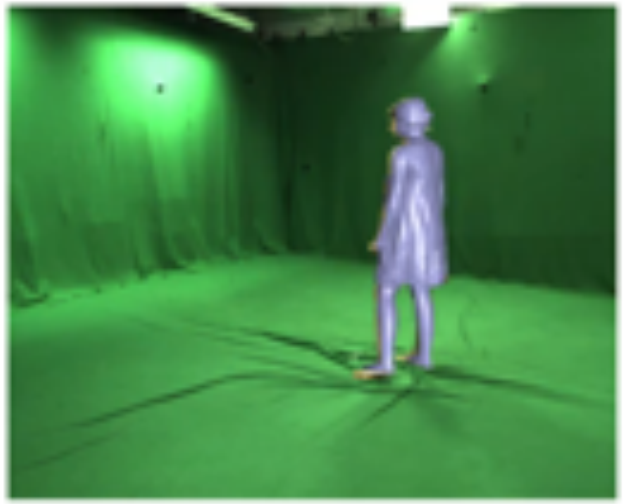
Overlay on **input** image





Overlay on **reference** view





Qualitative Evaluation

3D view

Overlay on input image

Saito et al. 2019 | Zheng et al. 2019 | **Ours**

Quantitative Evaluation

Surface reconstruction accuracy

Method (on S4)	Multi-view IoU* (in %)	
HMR (Kanazawa et al. 2018)	65.1	Person-unspecific
HMMR(Kanazawa et al. 2019)	63.79	
LiveCap (Habermann et al. 2019)	59.96	Person-specific
Ours	**82.53**	

*IoU = Intersection over Union

More results

Thank you!

Marc Habermann

Weipeng Xu

Michael Zollhoefer

Gerard Pons-Moll

Christian Theobalt