

Motivation



Input Image



SOTA Method



Our Method

Hard to estimate 3D human pose and shape for people with clothing as clothes vary a lot in shape and size

Previous Methods

Key points

Silhouette / Part Segmentation

Do not supervise body shape

Do not match projected minimally-clothed SMPL

High-Level Semantics
Solution

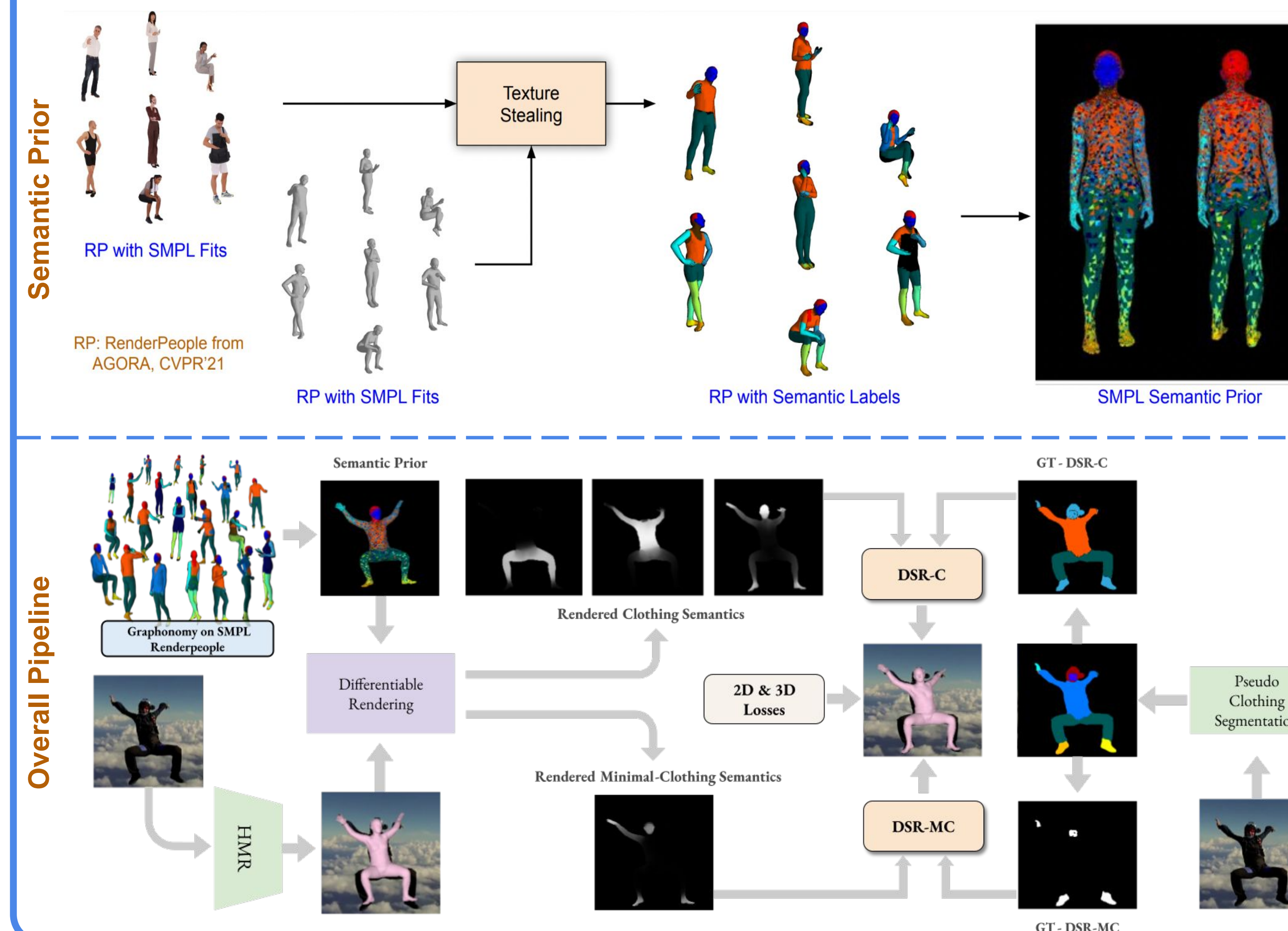
Clothing

Supervise SMPL to lie inside Clothing

Minimal Clothing

Supervise SMPL to exactly match MC

Differentiable Semantic Rendering



Results

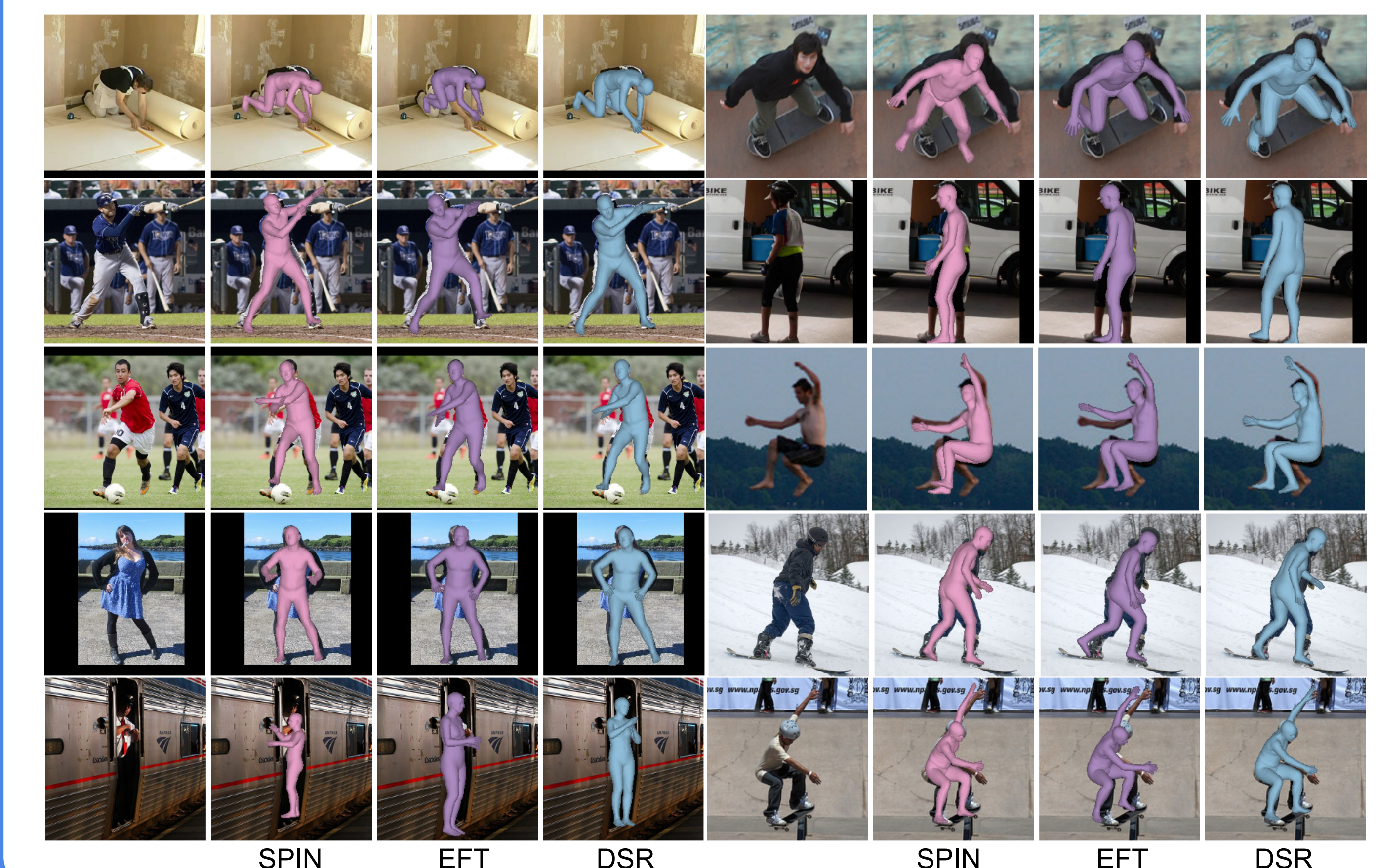
Ablation study

C-EFT: baseline method
DSR-FB: full-body silhouette supervision
DSR-MC: minimal clothing term
DSR-C: clothing term
DSR-MVP: manual labelling of clothing and minimal-clothing.

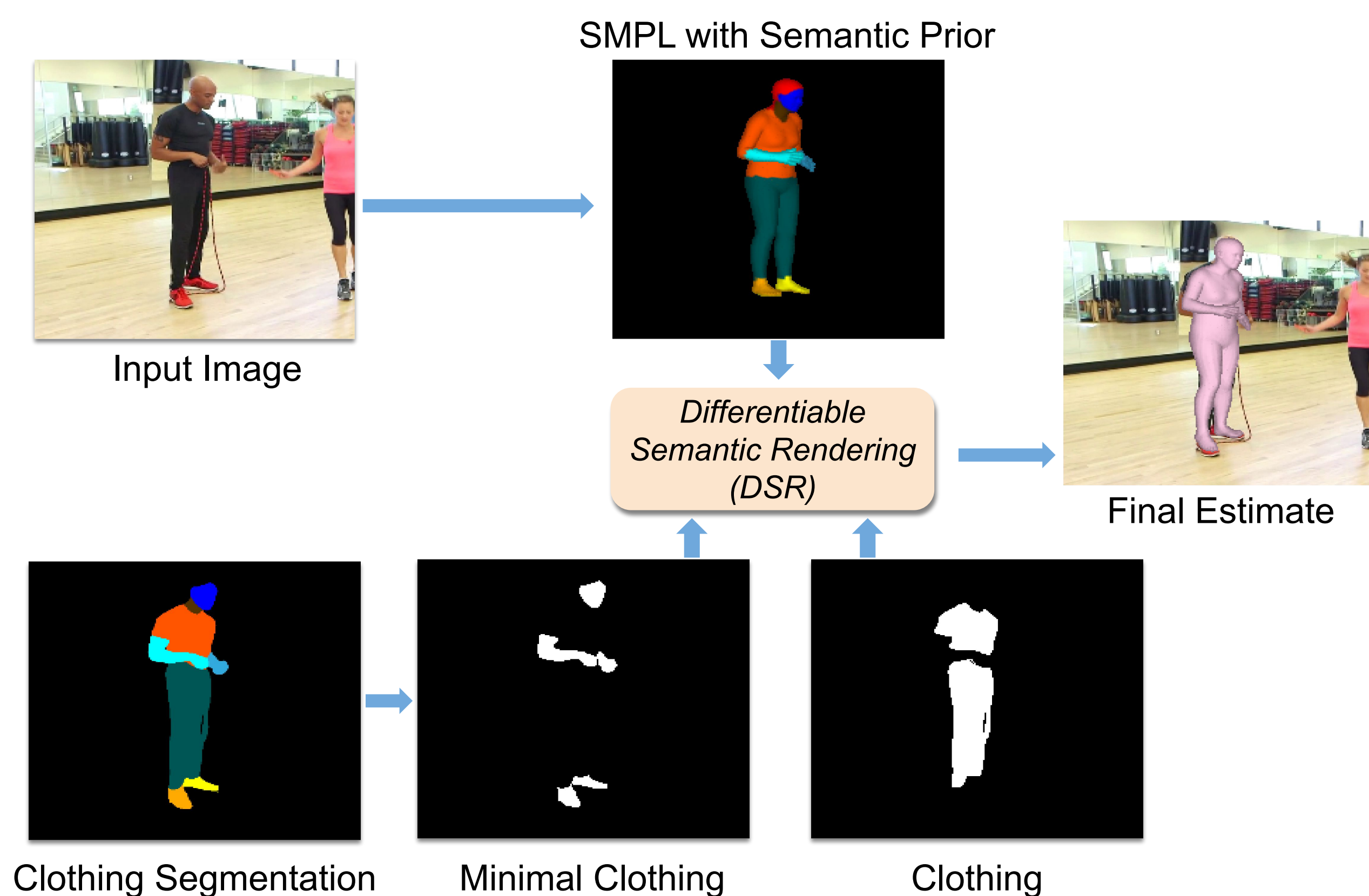
Method	PAMPJPE ↓	MPJPE ↓	PVE ↓
C-EFT	58.5	101.0	119.3
+ DSR-FB	59.8	102.1	120.3
+ DSR-FB (s-DistM)	58.0	100.2	117.8
+ DSR-MC (s-DistM)	58.2	100.6	118.5
+ DSR-MC (s-IoU)	58.0	100.3	118.1
+ DSR-C	57.6	99.8	117.6
+ DSR-MVP	58.1	100.3	117.8
+ DSR-C + DSR-MC (Ours)	57.2	99.2	116.3

SOTA Performance

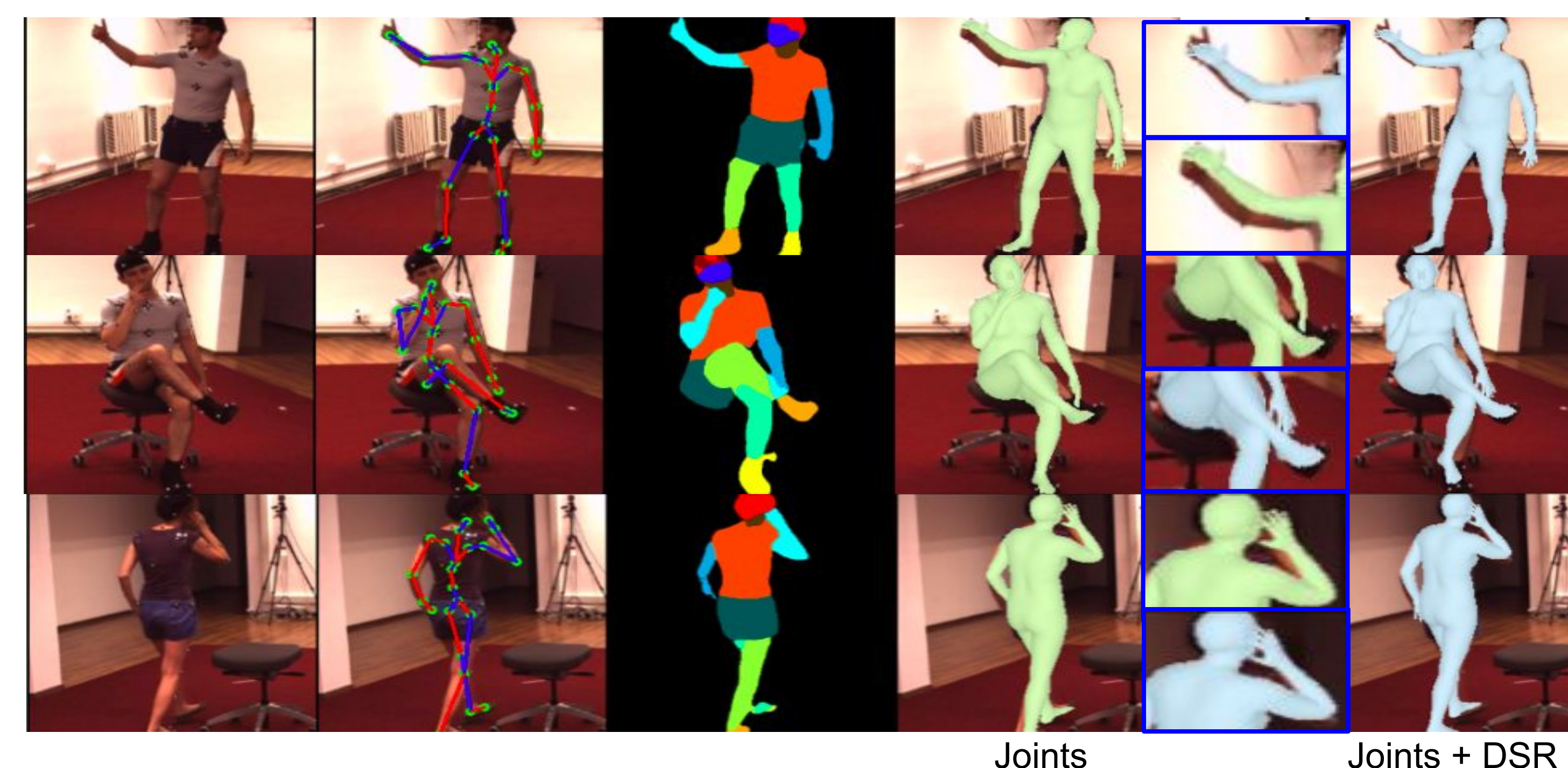
Models	3DPW			Human3.6M	
	PA-MPJPE ↓	MPJPE ↓	PVE ↓	PA-MPJPE ↓	MPJPE ↓
HMR [15]	76.7	130.0	-	56.8	88
CMR [19]	70.2	-	-	50.1	-
SPIN [18]	59.2	96.9	116.4	41.1	62.5
EFT [14]	54.2	-	-	43.7	-
Zanfir et. al [52] (w/ 3DPW train)	57.1	90.0	-	-	-
EFT [14] (w/ 3DPW train)	52.2	-	-	43.8	-
DSR	54.1	91.7	105.8	40.3	60.9
DSR (w/ 3DPW train)	51.7	85.7	99.5	41.4	62.0



Overall Idea



Are 3D Joints Enough?



We over-fit a batch of H36M samples on ground-truth (GT) joints (green) and joints with DSR (blue). The weak supervision with semantic information improves accuracy.

Resources



Project Page

<https://dsr.is.tue.mpg.de/>