(a) Input: Contact Interaction Multi-View Videos

Camera 1

Camera 2

Camera 3

Camera 4

(b) Camera Calibration

(c) 3D Pose + Instance Segmentation

(d) Output: Human Meshes