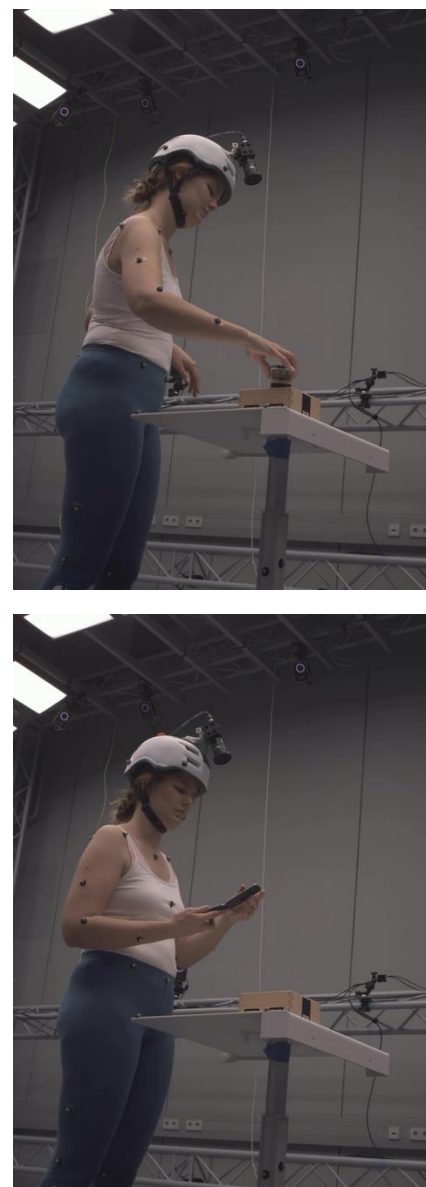


### Input video



Body observation



Body Encoder

Left hand observation



Hand Encoder

Right hand observation



Hand Encoder

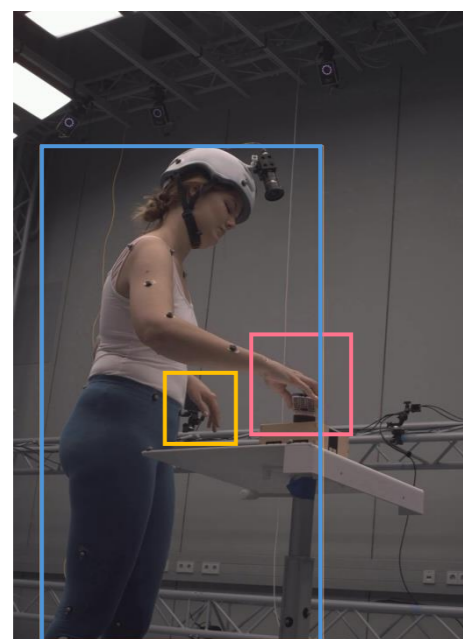
Hand Fusion



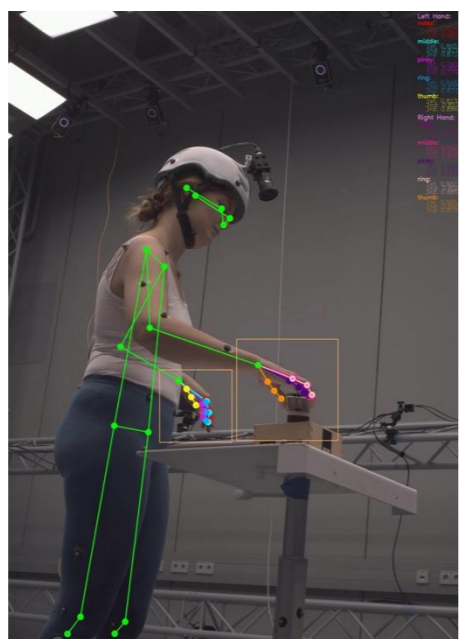
Temporal Motion Encoder

SMPL-X Decoder

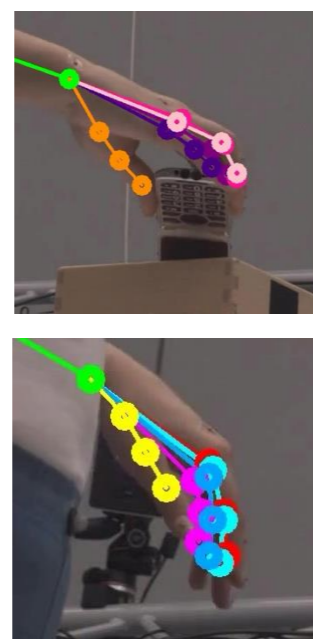
### Part-specific observations



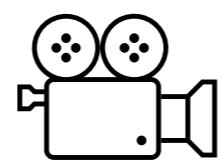
Box



2D keypoints



Vision feature



Camera



### Close-up aware augmentation



Full-body



Upper-body



Hand truncation

### Output SMPL-X sequence

