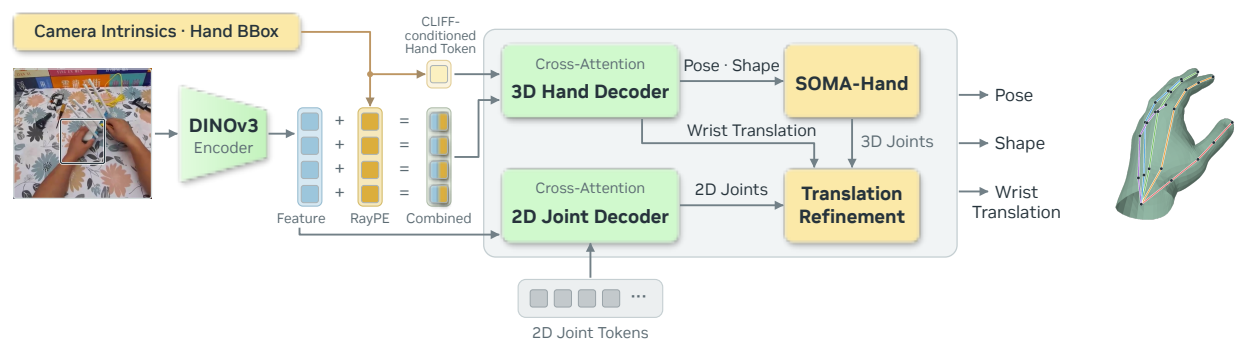

Camera Intrinsics · Hand BBox
CLIFF-conditioned Hand Token
DINOv3 Encoder
Feature
RayPE
Combined
Cross-Attention
3D Hand Decoder
Pose · Shape
SOMA-Hand
Pose
Shape
Wrist Translation
3D Joints
Cross-Attention
2D Joint Decoder
2D Joints
Translation Refinement
Wrist Translation
2D Joint Tokens