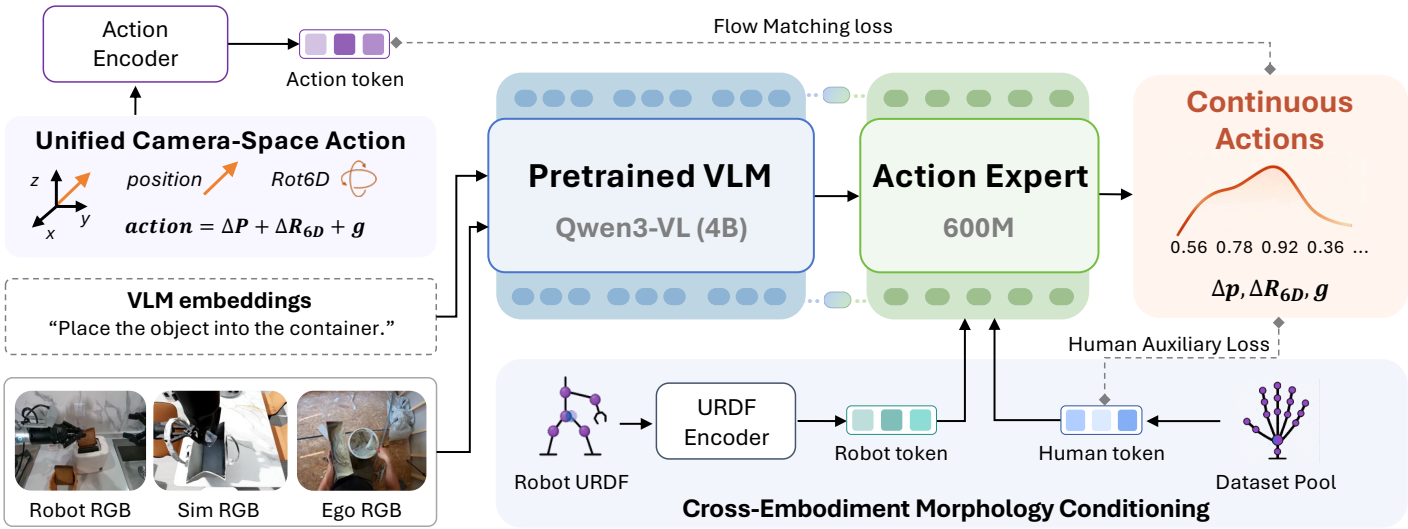
Action Encoder
Action token
Flow Matching loss
Unified Camera-Space Action
z
y
x
position
Rot6D
$action = \Delta P + \Delta R_{6D} + g$
VLM embeddings
"Place the object into the container."
Robot RGB
Sim RGB
Ego RGB
Pretrained VLM
Qwen3-VL (4B)
Action Expert
600M
Continuous Actions
0.56 0.78 0.92 0.36 ...
$\Delta p, \Delta R_{6D}, g$
Human Auxiliary Loss
Robot URDF
URDF Encoder
Robot token
Human token
Dataset Pool
Cross-Embodiment Morphology Conditioning