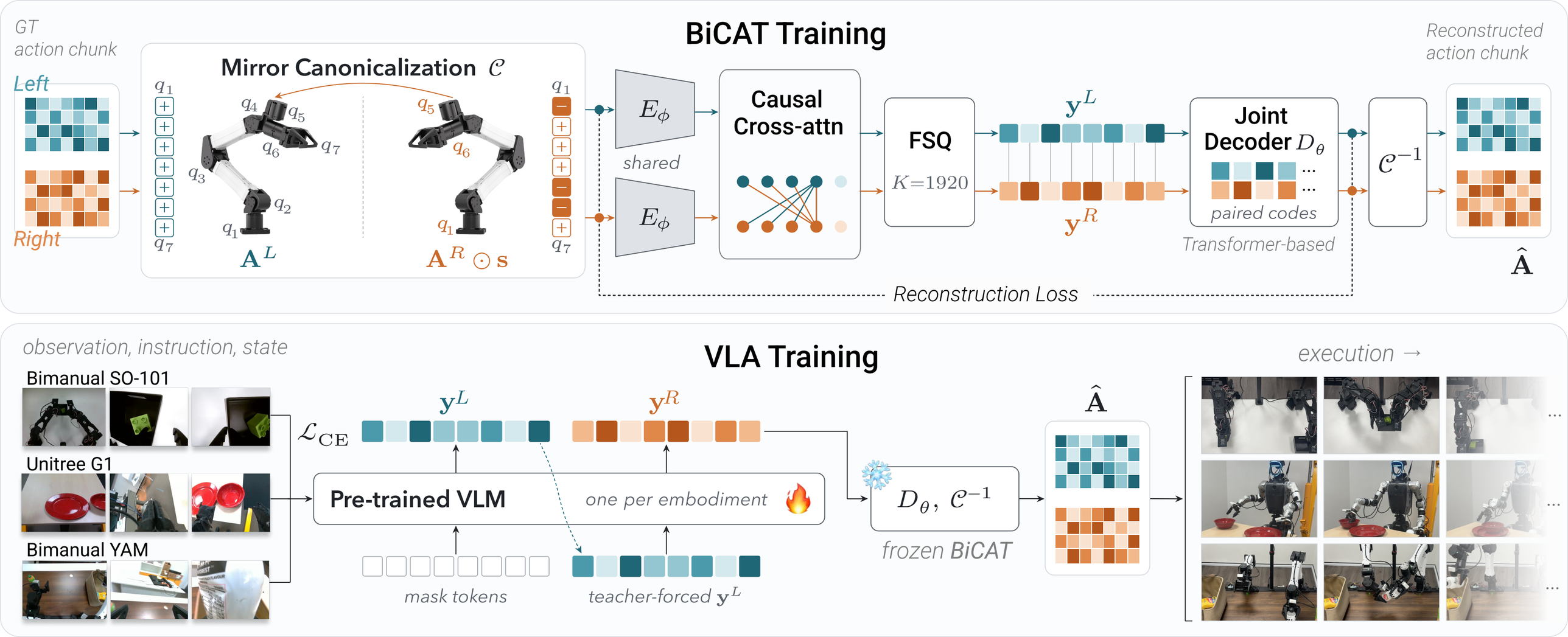

GT
action chunk
Left
Right
BiCAT Training
Mirror Canonicalization $\mathcal{C}$
$\mathbf{A}^L$
$\mathbf{A}^R \odot \mathbf{s}$
E_ϕ
shared
E_ϕ
Causal
Cross-attn
FSQ
$K=1920$
$\mathbf{y}^L$
$\mathbf{y}^R$
Joint
Decoder D_θ
paired codes
Transformer-based
$\mathcal{C}^{-1}$
Reconstructed
action chunk
$\hat{\mathbf{A}}$
Reconstruction Loss
observation, instruction, state
Bimanual SO-101
Unitree G1
Bimanual YAM
VLA Training
$\mathcal{L}_{\text{CE}}$
$\mathbf{y}^L$
$\mathbf{y}^R$
Pre-trained VLM
one per embodiment
mask tokens
teacher-forced $\mathbf{y}^L$
D_θ, $\mathcal{C}^{-1}$
frozen BiCAT
$\hat{\mathbf{A}}$
execution →