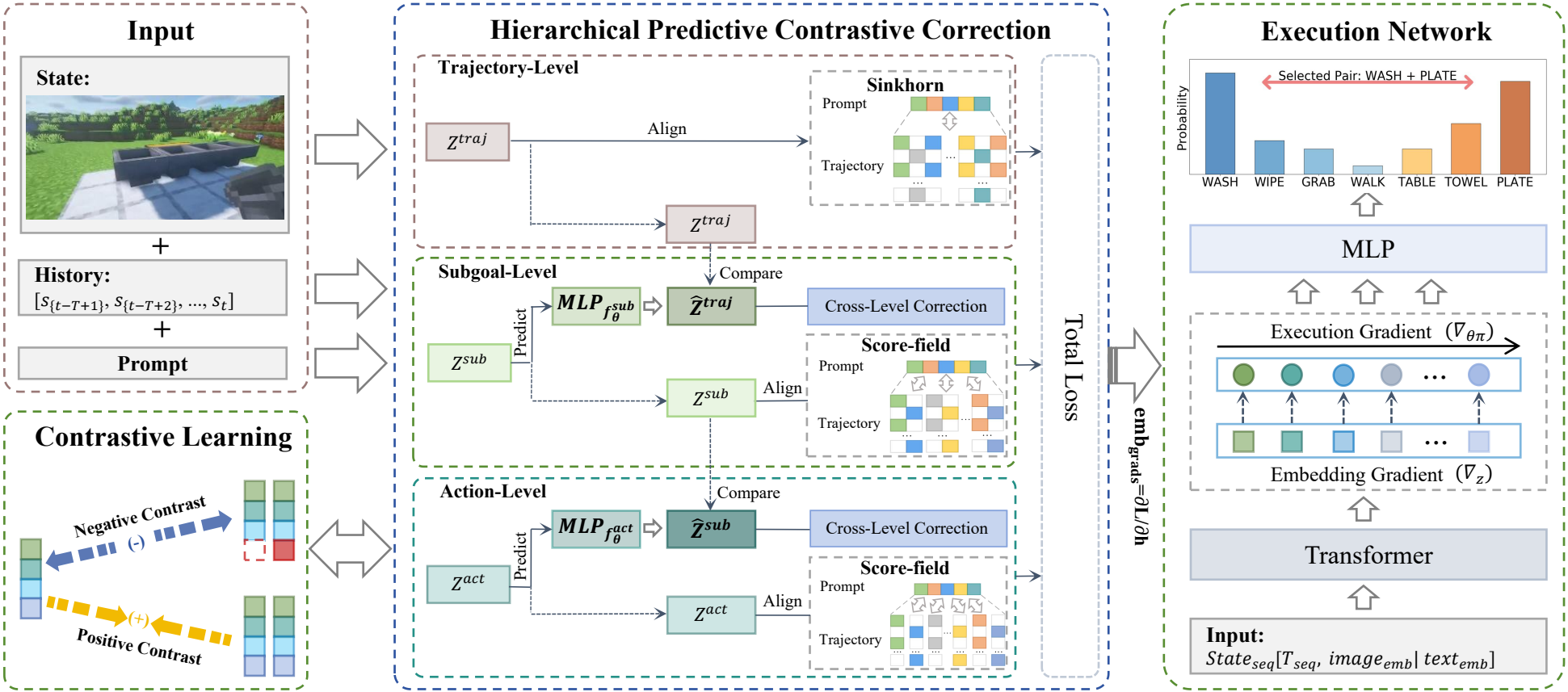
Input
State:
+
History:
$[s_{\{t-T+1\}}, s_{\{t-T+2\}}, \ldots, s_t]$
+
Prompt
Contrastive Learning
Negative Contrast
(-)
(+)
Positive Contrast
Hierarchical Predictive Contrastive Correction
Trajectory-Level
Z^{traj}
Align
Sinkhorn
Prompt
Trajectory
Z^{traj}
Compare
Subgoal-Level
Predict
$MLP_{f_\theta^{sub}}$
$\hat{Z}^{traj}$
Cross-Level Correction
Z^{sub}
Score-field
Prompt
Trajectory
Z^{sub}
Align
Action-Level
Compare
Predict
$MLP_{f_\theta^{act}}$
$\hat{Z}^{sub}$
Cross-Level Correction
Z^{act}
Score-field
Prompt
Trajectory
Z^{act}
Align
Total Loss
$emb_{grads}=\partial L/\partial h$
Execution Network
Probability
Selected Pair: WASH + PLATE
WASH
WIPE
GRAB
WALK
TABLE
TOWEL
PLATE
MLP
Execution Gradient $(\nabla_{\theta\pi})$
Embedding Gradient (∇_z)
Transformer
Input:
$State_{seq}[T_{seq}, image_{emb}|\ text_{emb}]$