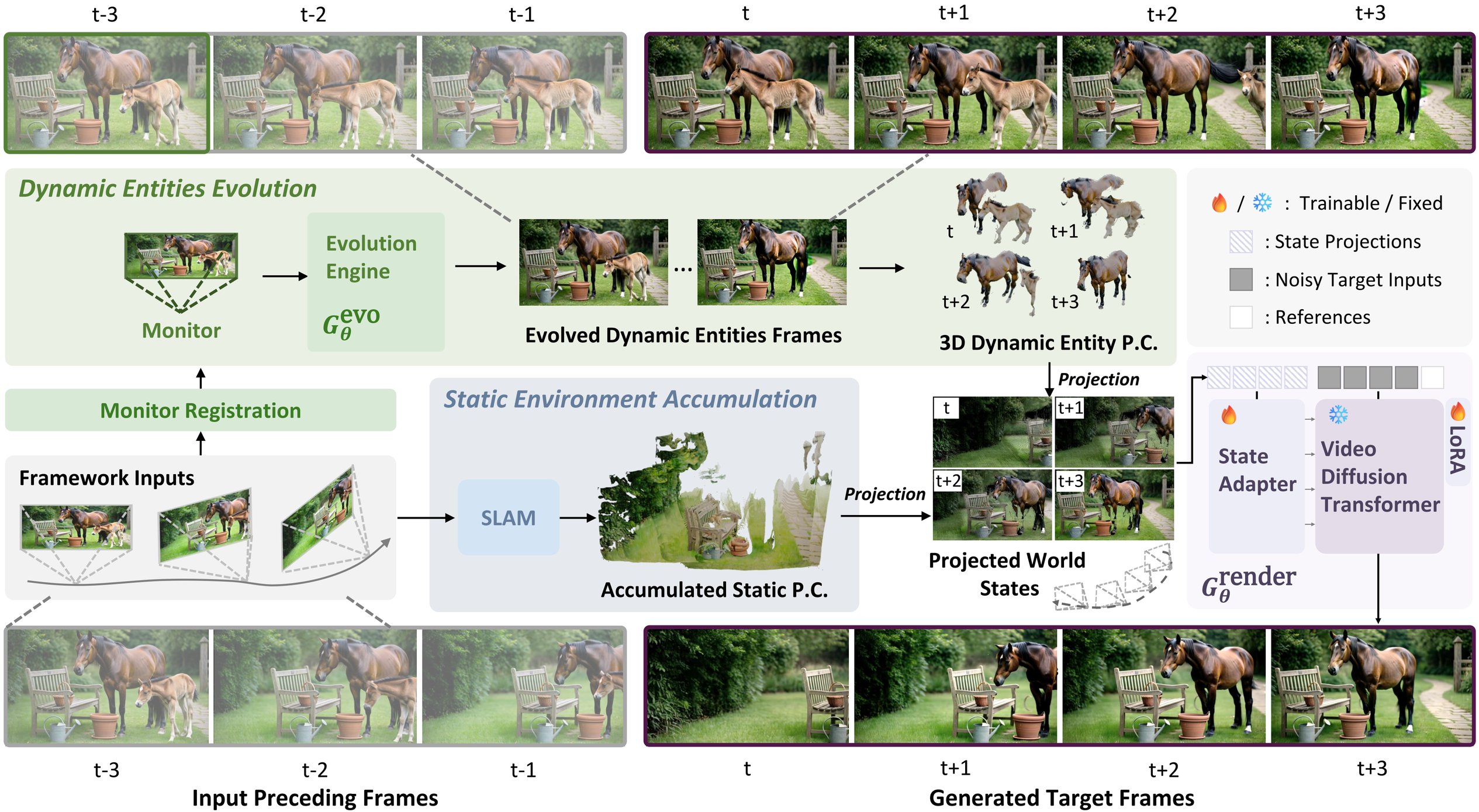
t-3
t-2
t-1
t
t+1
t+2
t+3
Dynamic Entities Evolution
Monitor
Evolution Engine
G_θ^{evo}
Evolved Dynamic Entities Frames
t
t+1
t+2
t+3
3D Dynamic Entity P.C.
/ : Trainable / Fixed
: State Projections
: Noisy Target Inputs
: References
Monitor Registration
Static Environment Accumulation
Projection
t
t+1
t+2
t+3
State Adapter
Video Diffusion Transformer
LoRA
Framework Inputs
SLAM
Accumulated Static P.C.
Projection
Projected World States
G_θ^{render}
t-3
t-2
t-1
t
t+1
t+2
t+3
Input Preceding Frames
Generated Target Frames