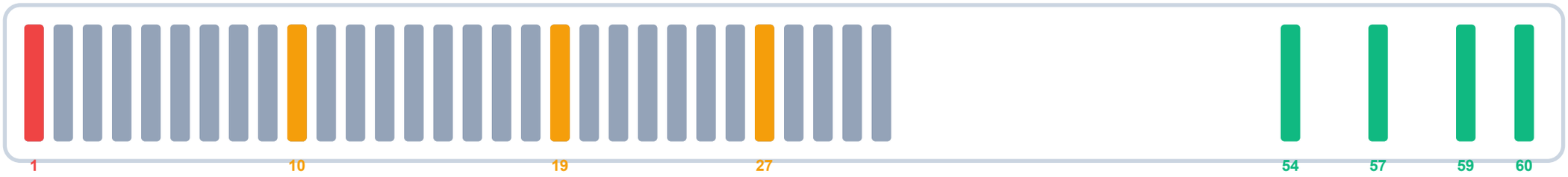
Learned Memory Pruning: Which Frames Does SAM2-Lite Keep?

Example 60-frame sequence (2 seconds @ 30fps)



Frame 1 (Current frame) Frame 60

Learned Memory Selection Strategy

1. Anchor Frames (Always Keep)

- First 1-2 frames serve as reference templates
- Critical for long-term identity tracking

~2 frames
(3-5% budget)

2. Appearance Change Points (High Priority)

- Frames with object rotations, occlusions, or lighting shifts
- High motion magnitude and prediction uncertainty

~5-8 frames
(15-25% budget)

3. Recent Context Window (Temporal Smoothness)

- Last 2-4 frames for smooth tracking transitions
- Ensures temporal consistency between predictions

~3-4 frames
(8-12% budget)

Result: 256-512 informative tokens vs. 60 frames × 256 tokens = 15,360 total tokens
30-60× memory reduction with minimal accuracy loss through intelligent frame selection

- Static object appearance with minimal changes

(75% removed)