

StoryMem

- StoryMem: Multi-shot Long Video Storytelling with Memory
- <https://arxiv.org/abs/2512.19539>
- arxiv

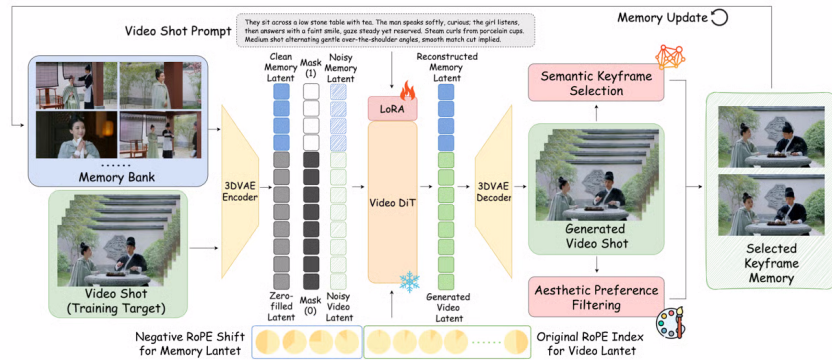


Figure 2 Overview of StoryMem. *StoryMem* generates each shot conditioned on a memory bank that stores keyframes from previously generated shots. During generation, the selected memory frames are encoded by a 3D VAE, fused with noisy video latents and binary masks, and fed into a LoRA-finetuned memory-conditioned Video DiT to synthesize the current shot. After generating each shot, semantic keyframe selection and aesthetic preference filtering are applied to obtain informative and reliable memory frames, enabling long-range cross-shot consistency and natural narrative progression. By iteratively generating shots with memory updates, *StoryMem* produces coherent minute-long, multi-shot story videos.

[illegible]

StoryMem
 Wan2.2I2V

CLIP

- 0000000000000000
- 000<https://kevin-thu.github.io/StoryMem/>

Newer

Older

2026-1-8

ConceptAttention

2026-1-6
VideoLCT