ShotStream

- □□□□□ShotStream: Streaming Multi-Shot Video Generation for Interactive Storytelling
- □□□□□https://arxiv.org/abs/2603.25746
- arxiv 2603

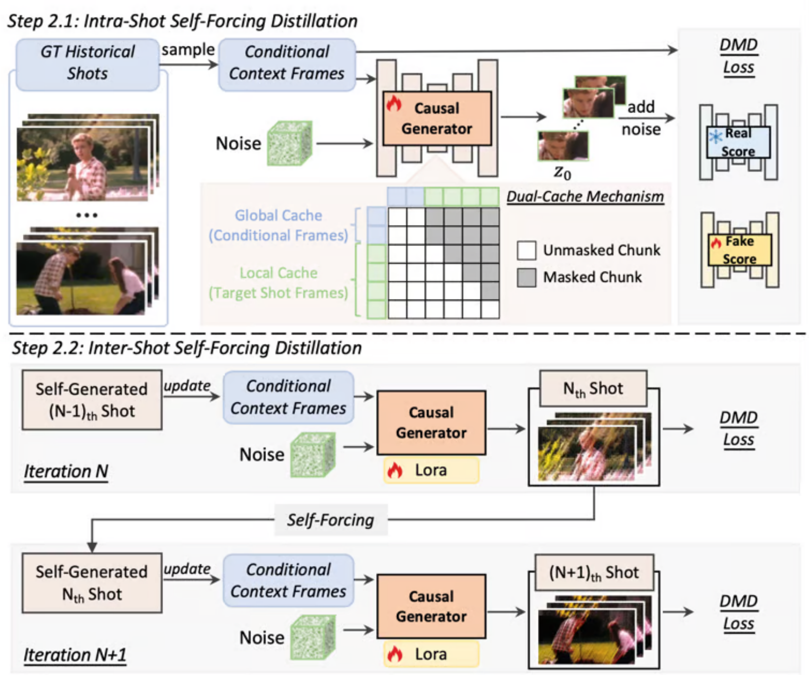

Figure 4. Causal Architecture and Two-Stage Distillation Pipeline. We distill a slow, multi-step bidirectional teacher into an efficient, few-step causal generator. To maintain visual coherence, we propose a novel dual-cache memory mechanism: a global context cache stores conditional frames to ensure inter-shot consistency, while a local context cache retains generated frames within the target shot to guarantee intra-shot consistency. To prevent error accumulation, we employ a progressive two-stage distillation strategy. In the first stage, intra-shot self-forcing distillation (Step 2.1), the model is conditioned on ground-truth historical shots to causally generate the current shot chunk-by-chunk. In the second stage, inter-shot self-forcing distillation (Step 2.2), the model is conditioned on its own previously generated shots, rolling out the video shot-by-shot while iteratively generating the frames of each individual shot chunk-by-chunk.

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□ShotStream□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□next-shot□□□□□□□□□□□

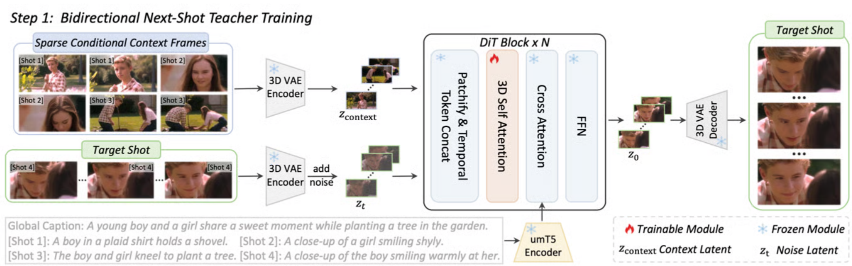

Figure 3. Architecture of the Bidirectional Next-Shot Teacher Model. To realize ShotStream, we first fine-tune a text-to-video model into a bidirectional next-shot model, which generates subsequent shots conditioned on sparse context frames from preceding shots. These conditional context frames are encoded into latents via a 3D VAE and injected by concatenating them with noise latents along the temporal dimension. Notably, only the 3D spatial-temporal attention layers within the DiT Blocks are optimized during fine-tuning. A 4-shot example is shown here for illustration.

□□□□DMD□□□□□□□□student□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□cache□□□□□□□□□□□cache□□global context□□□□□□□cache□□local context□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□self forcing□□□□□□□□□self forcing□

- □□□□□□□□□□□
- □□□32 H800
- □□□https://luo0207.github.io/ShotStream/