□

home
papers
study
life

About me

MagicComp

- □□□□□MagicComp: Training-free Dual-Phase Refinement for Compositional Video Generation
- □□□□□https://arxiv.org/abs/2503.14428
- arxiv

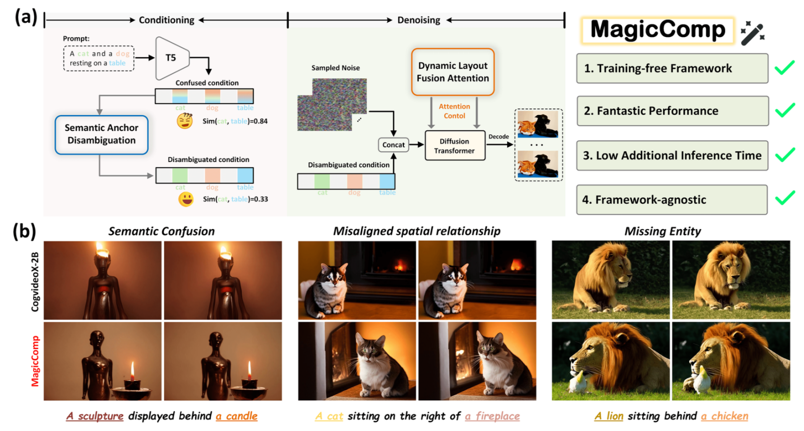

Figure 1. **Overall pipeline for MagicComp.** (a) Our MagicComp comprises two core modules: Semantic Anchor Disambiguation (SAD) for resolving inter-subject ambiguity during conditioning, and Dynamic Layout Fusion Attention (DLFA) for spatial-attribute binding via fused layout masks in denoising. (b) MagicComp is a training-free framework, which effectively address the challenges (e.g., semantic confusion, misaligned spatial relationship, missing entities) in compositional video generation with minimal additional inference overhead.

□□training-free□□□□□□□□□□□□□□□□□□□1□□□□□□□□□□□□token□embedding□□□□□□prompt□□□□□□□□□□token□embedding□□□□□□□□□□2□□□□□□□□□□□token□attention□□mask□□□□□□□□□□□□□□

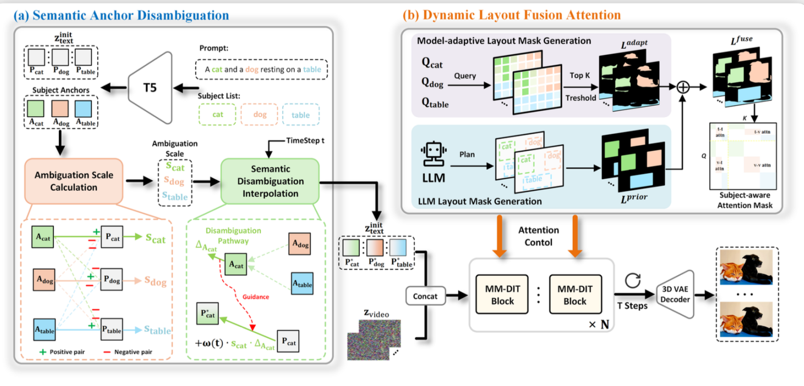

Figure 2. **Detailed architecture of MagicComp.** The dual-phase refinement strategy of MagicComp contains two core steps: (a) Semantic Anchor Disambiguation (SAD) module for inter-subject disambiguation during the conditioning stage. We only display disambiguation process of subject "cat" for simplicity, other subjects follow the similar way. (b) Dynamic Layout Fusion Attention (DLFA) module for precise attribute–location binding of each subject during the denoising stage.

- □□□□□□□□□□
- □□□T2V-CompBench; VBench
- □□□1 A100
- □□□https://hong-yu-zhang.github.io/MagicComp-Page/

Newer Older

2025-06-03 DiTCtrl	2025-06-03 Video-MSG

Archive RSS feed GitHub Email QR Code

Made with Montaigne and bigmission