



home
papers
study
life

About me

OmniCustomizer

- [Omni-Customizer: End-to-End MultiModal Customization for Joint Audio-Video Generation](#)
- <https://arxiv.org/abs/2605.17488>
- arxiv 2026

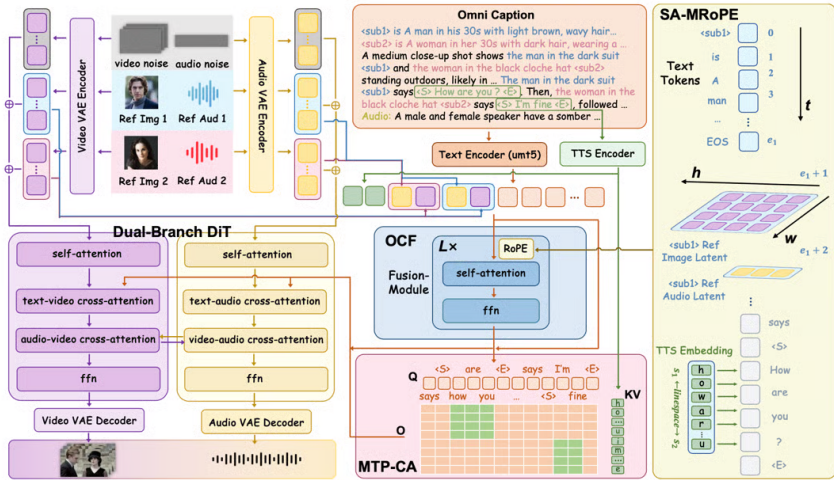


Figure 2: Framework of Omni-Customizer: The text prompt, TTS embeddings, reference images, and audios are integrated by the OCF module, which employs SA-MRoPE for precise multimodal binding. Additionally, MTP-CA ensures exclusive pronunciation enhancement for spoken texts. The enhanced context is injected into the dual-stream backbone for cohesive multi-subject identity preservation.

•
 In-Context

Newer

Older

2026-5-26
Just-Dub-It

2026-5-20
DreamID-Omni