DreamID-Omni

- DreamID-Omni: Unified Framework for Controllable Human-Centric Audio-Video Generation
- https://arxiv.org/abs/2602.12160
- ICML 2026

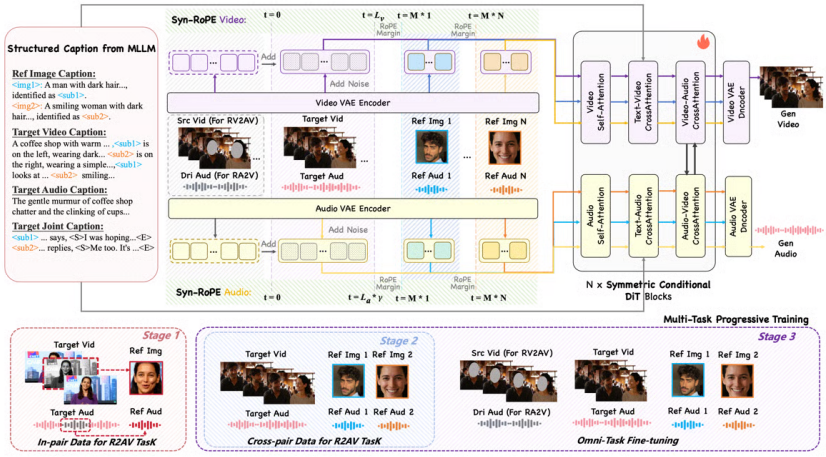
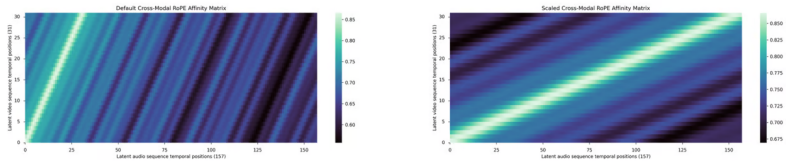


Figure 2 Overview of DreamID-Omni framework. We integrate reference-based generation (R2AV), editing (RV2AV), and animation (RA2V) using a Symmetric Conditional DiT trained via a multi-task progressive training strategy. Structured Caption and Syn-RoPE ensure robust dual-level disentanglement in multi-person scenarios.

R2AV、RV2AV、RA2V id anchor 1 2 3

DreamID-Omni ref image、audio token add token token Ovi



(a) Default (unscaled) RoPE affinity. (b) Aligned RoPE after scaling.

Figure 2: Cross-modal RoPE affinity matrices before and after scaling. Scaling aligns audio and video temporal positions, improving synchronization.

reference token reference ref 1 in-pair R2AV 2 cross-pair R2AV 3

pipeline

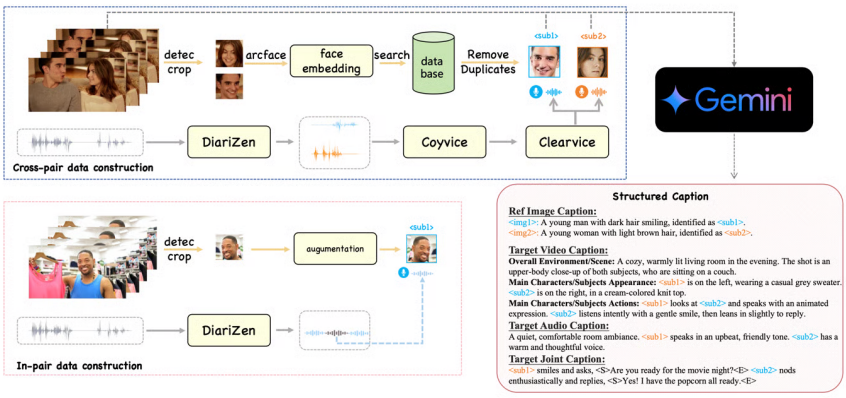


Figure 7 Data construction pipeline.

Newer Older

2026-5-6 2026-3-9

ShotStream

Flow-DPO

leicheng © 2022-2025