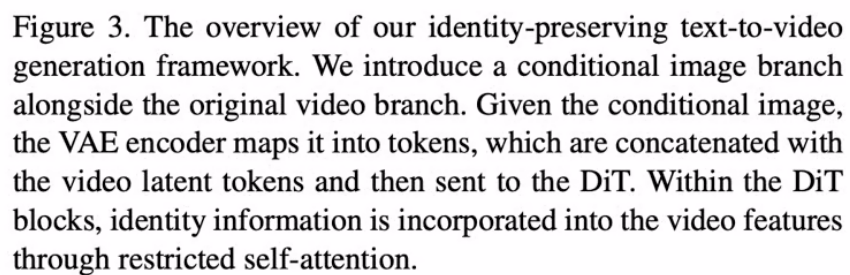


- Stand-In: A Lightweight and Plug-and-Play Identity Control for Video Generation
- <https://arxiv.org/abs/2508.07901>
- arxiv 2025



The diagram illustrates the architecture of the proposed LoRA-based video adapter. It consists of two main parallel paths for temporal (T_I) and spatial (T_V) features. Each path starts with a QKV projection (marked as trainable) and a LoRA module (marked as frozen). The output is then processed by a Conditional Position Mapping block, which uses a 3D volume (height, width, frame) to map spatial coordinates (p_I , p_V) to temporal coordinates (Q'_I , K'_I , Q'_V , K'_V). The final output is the product of the LoRA module and the mapped features (marked as trainable).

LoRA

ConceptAttention