

Input
Video



ASR Model

PySceneDetect

Video segments



Transcription



Qwen-VL-8B

Scene Descriptions
with Timestamp



Fused contextual block



Gemini-3-flash preview

Inference-Time Contextual Guidance

Output
Timestamp



Start

End