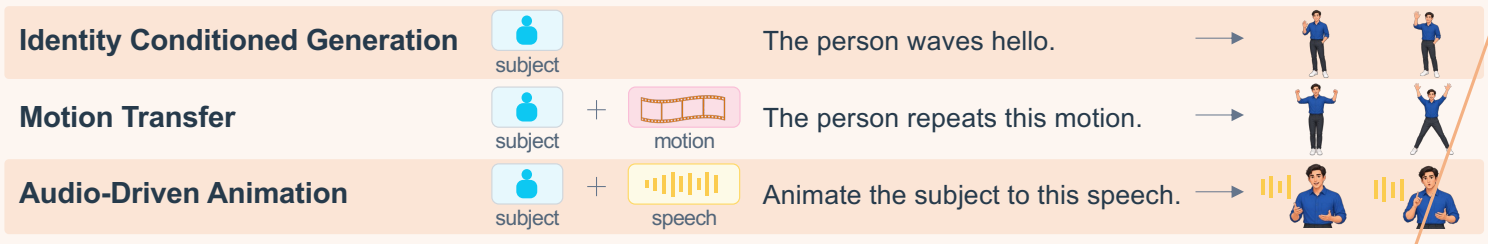
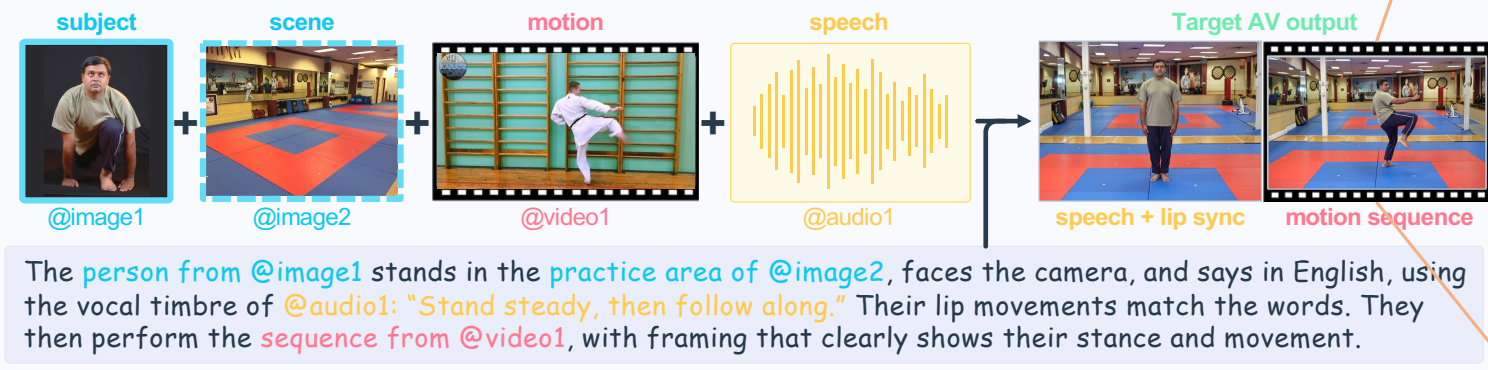


(a) Existing Reference-Conditioned Video Generation



(b) Omni-Reference Audio-Video Generation



(c) Multimodal Context Understanding

