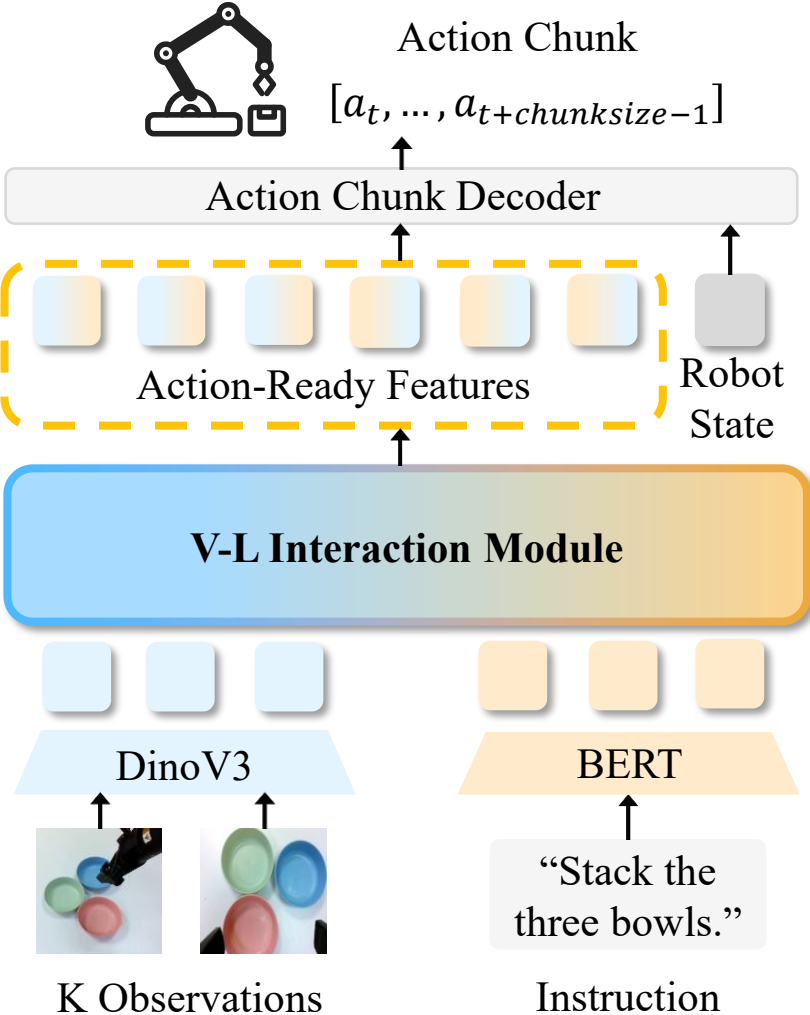
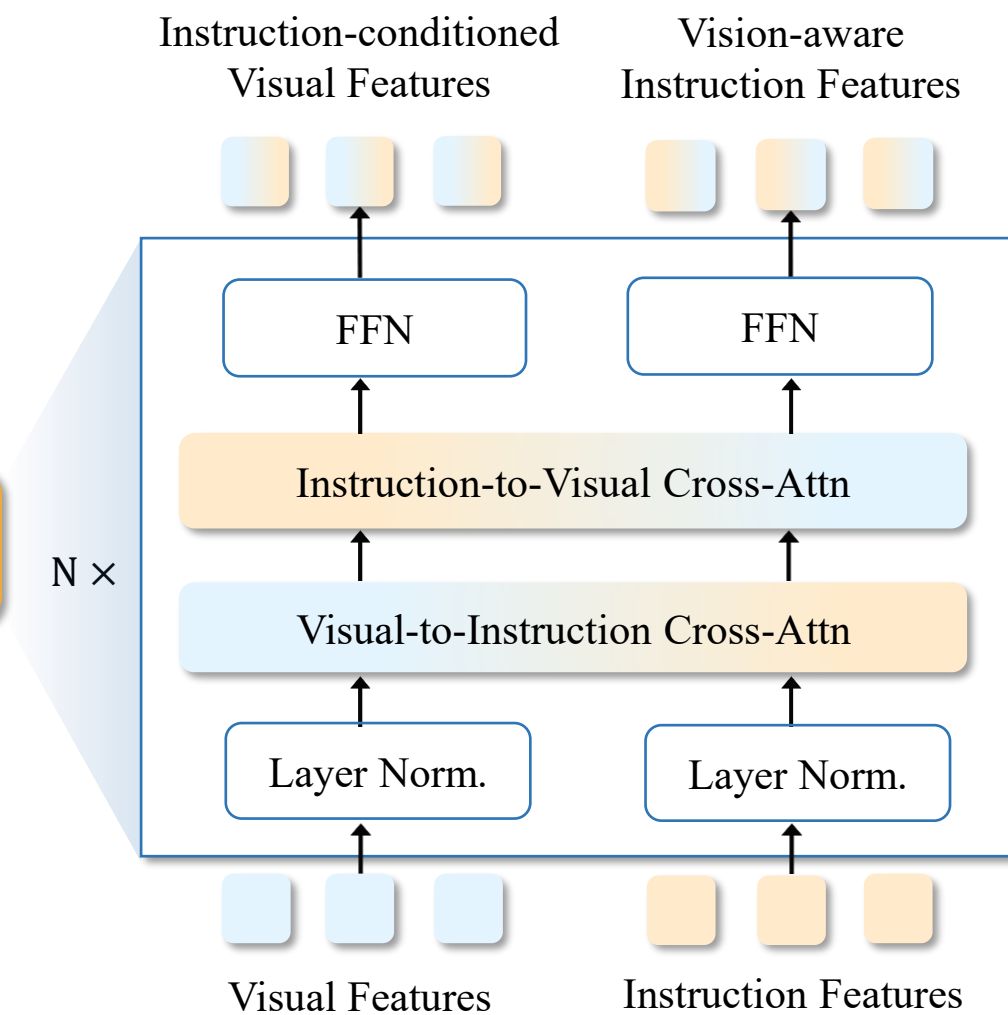




(a) Overall Architecture



(b) Vision-Language Interaction Module