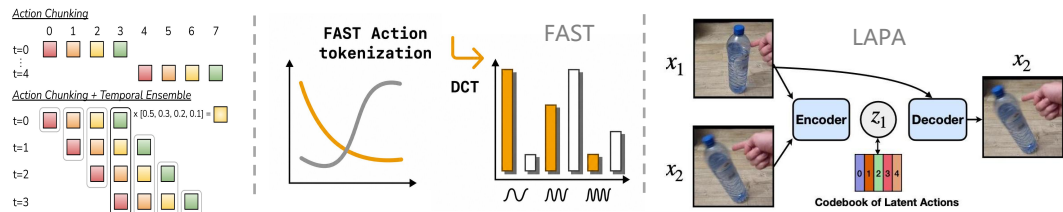


## (a) Pre-Training

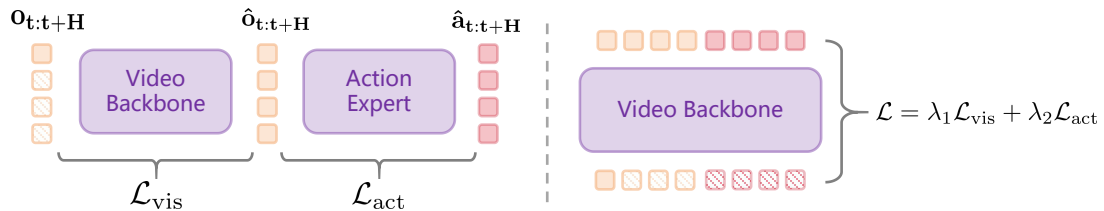
### 1) Spatiotemporal Representation



### 2) Action Representation

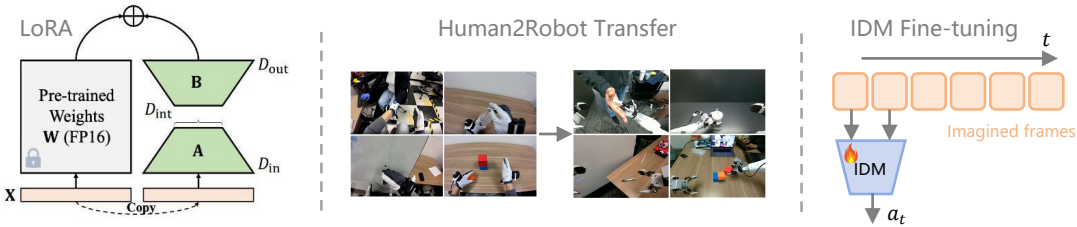


### 3) Video-Action Alignment

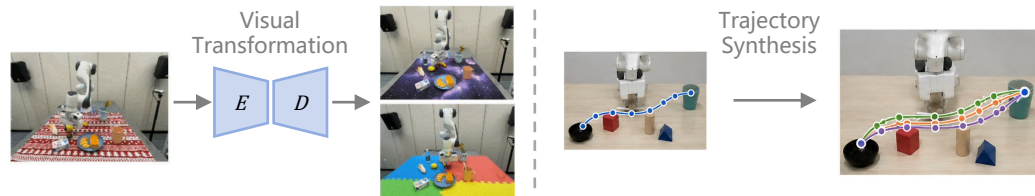


## (b) Post-Training

### 1) Policy Fine-tuning



### 2) Data Augmentation



### 3) Reinforcement Learning

