

## Paradigm I

Observation

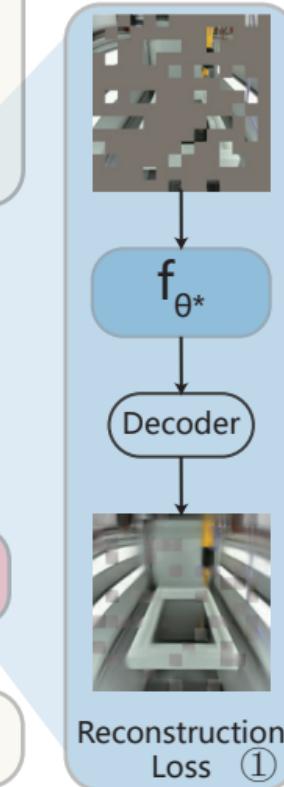


I

Visual Encoder  
 $f_{\theta^*}$

Policy  $\pi$

Action a



(a)

## Paradigm II

Observation

Stage 1?

Yes

Random Motor Control

Action

Success?

Yes

Store

$f_{\theta} + \pi$

Update

(b)