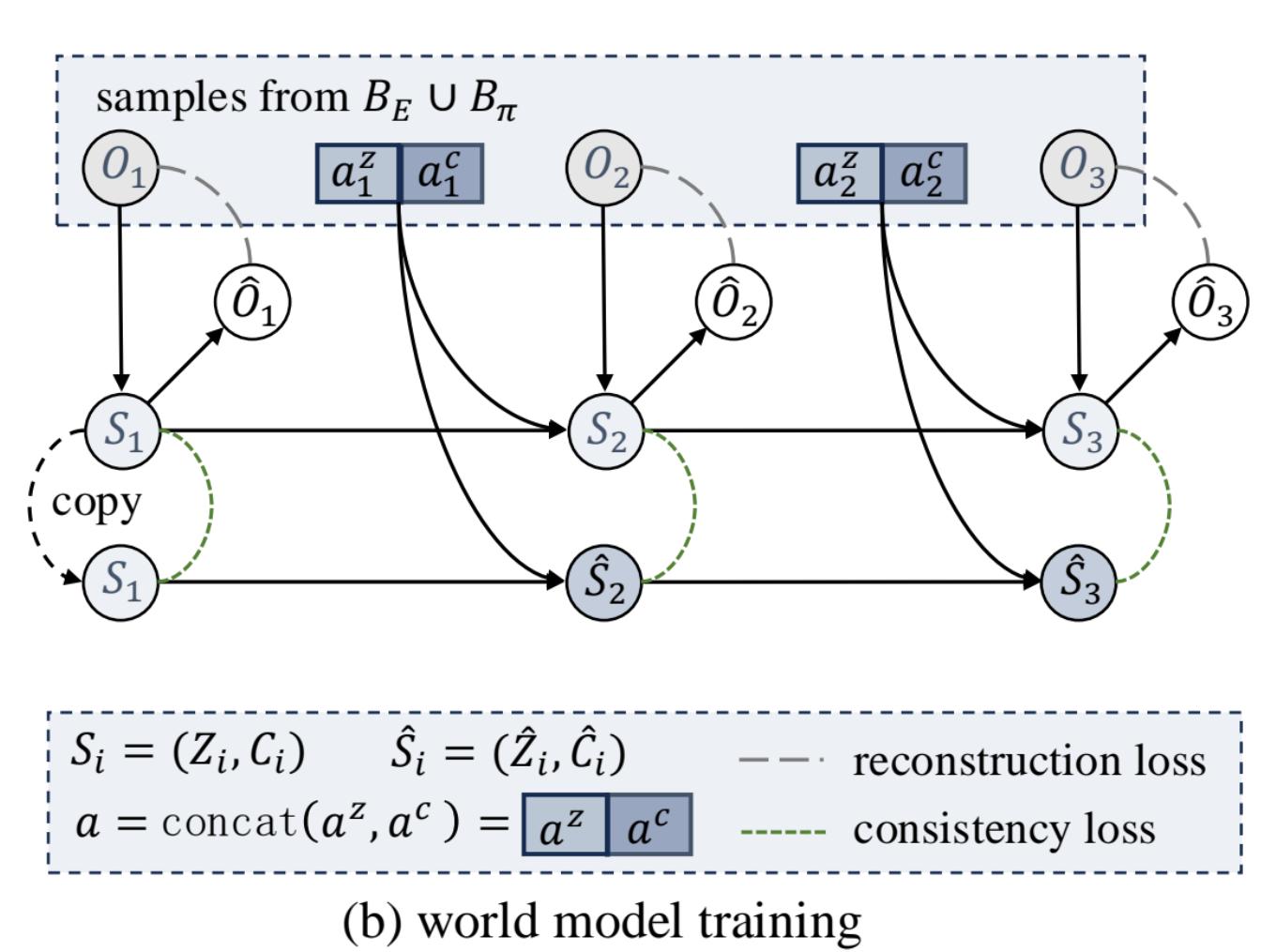


(a) policy training



(b) world model training

$$\begin{aligned}
 S_i &= (Z_i, C_i) & \hat{S}_i &= (\hat{Z}_i, \hat{C}_i) \\
 a &= \text{concat}(a^z, a^c) = [a^z \ a^c] & \text{---} & \text{reconstruction loss} \\
 & & \text{---} & \text{consistency loss}
 \end{aligned}$$