

FCAI

fcai.fi

DWMM: Components

Encoder

$$\mathbf{x}_t = e_{\theta}(\mathbf{s}_t)$$

Latent quantization $\mathbf{c}_t = f(\mathbf{x}_t) \in \mathcal{E}$

Dynamics $\hat{\mathbf{c}}_{t+1} \sim \text{Categorical}(p_1, \dots, p_{|\mathcal{C}|})$ with $p_i = P_\phi(\mathbf{c}_{t+1} = \mathbf{c}^{(i)} \mid \mathbf{c}_t, \mathbf{a}_t)$

Reward

$$\hat{r}_{t+1} = R_{\xi}(\mathbf{c}_t, \mathbf{a}_t)$$

Critic

$$q_t = Q_{\psi}(\mathbf{c}_t, \mathbf{a}_t)$$

Prior Policy $\mathbf{a}_t \sim \pi_{\eta}(\mathbf{a}_t | \mathbf{c}_t)$



DCWM: Components

Encoder $\mathbf{x}_t = e_\theta(\mathbf{s}_t)$

Latent quantization $\mathbf{c}_t = f(\mathbf{x}_t) \in \mathcal{C}$

Dynamics $\hat{\mathbf{c}}_{t+1} \sim \text{Categorical}(p_1, \dots, p_{|\mathcal{C}|})$ with $p_i = P_\phi(\mathbf{c}_{t+1} = \mathbf{c}^{(i)} \mid \mathbf{c}_t, \mathbf{a}_t)$

Reward $\hat{r}_{t+1} = R_\xi(\mathbf{c}_t, \mathbf{a}_t)$

Critic $q_t = Q_\psi(\mathbf{c}_t, \mathbf{a}_t)$

Prior Policy $\mathbf{a}_t \sim \pi_\eta(\mathbf{a}_t \mid \mathbf{c}_t)$

Model-based Reinforcement Learning

