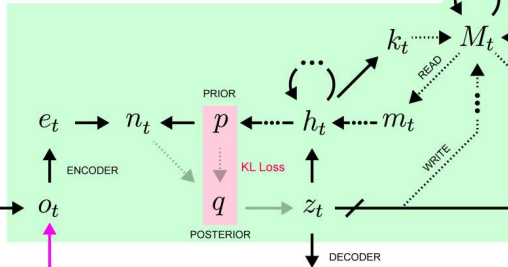


c. MERLIN

ENVIRONMENT

(I_t, v_t, r_{t-1}, T_t)

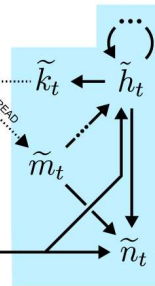
MEMORY-BASED PREDICTOR



$(\hat{I}_t, \hat{R}_t, \hat{v}_t, \hat{a}_{t-1}, \hat{r}_{t-1}, \hat{T}_t)$

Reconstruction Loss

READ-ONLY POLICY



a_t

Policy Loss

