MetaRL for Non-stationary and -parametric Environments [Bing et al. 2023]

- often task changes over time
- especially if a task is comprised of subtasks

MetaRL for Non-stationary and -parametric Environments [Bing et al. 2023]

- often task changes over time
- especially if a task is comprised of subtasks

$\rightarrow$ generate task-embedding via VAE

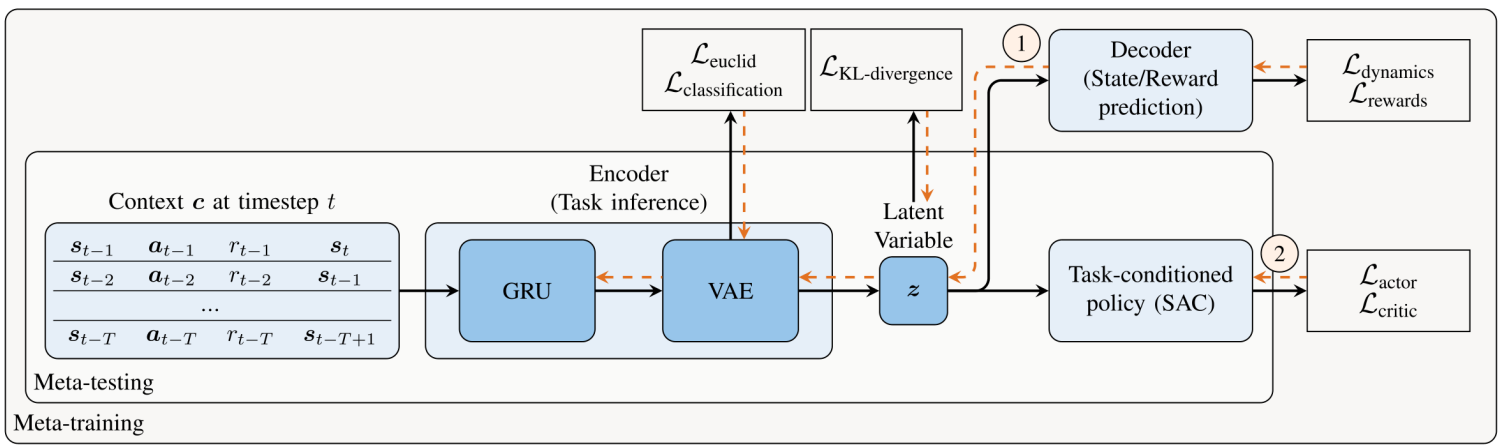

Fig. 2. Meta-training and meta-testing procedure. The encoder learns a task encoding z from the recent context with gradients from the decoder and provides z for the task-conditioned policy trained via SAC during meta-training. In meta-testing, the latent task description z is fed into the task-conditioned policy to perform the actions. Orange arrows outline the gradient flow.

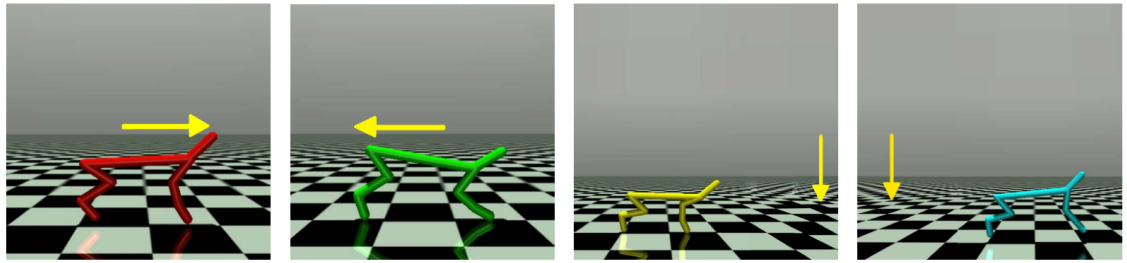

Fig. 1. Visualization of eight base environment tasks (yellow arrows show the movement direction). These eight tasks share similar dynamic models, but are qualitatively distinct. Each task contains parametric variations, e.g., different goal velocities in the run forward/backward task. (a) Run forward. (b) Run backward. (c) Reach front goal. (d) Reach back goal. (e) Front stand. (f) Back stand. (g) Jump. (h) Front flip.

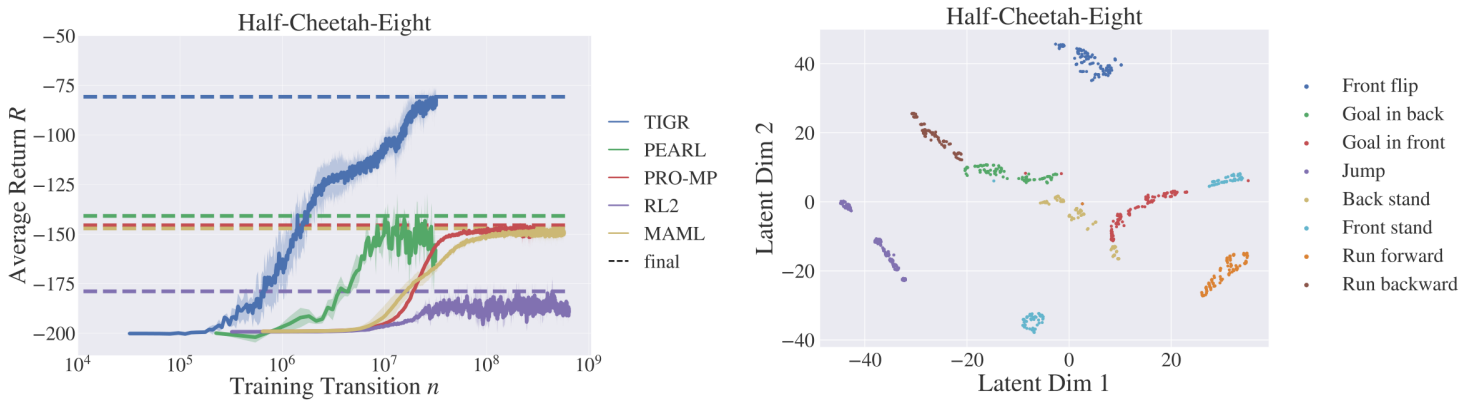

Fig. 7. (a) Meta-testing performance over environment interactions evaluated periodically during the meta-training phase. We show the mean performance from three independent runs. (b) Final encoding of the eight tasks visualized using t-SNE [31] in two dimensions. Note that the x-axis is in **log scale**.