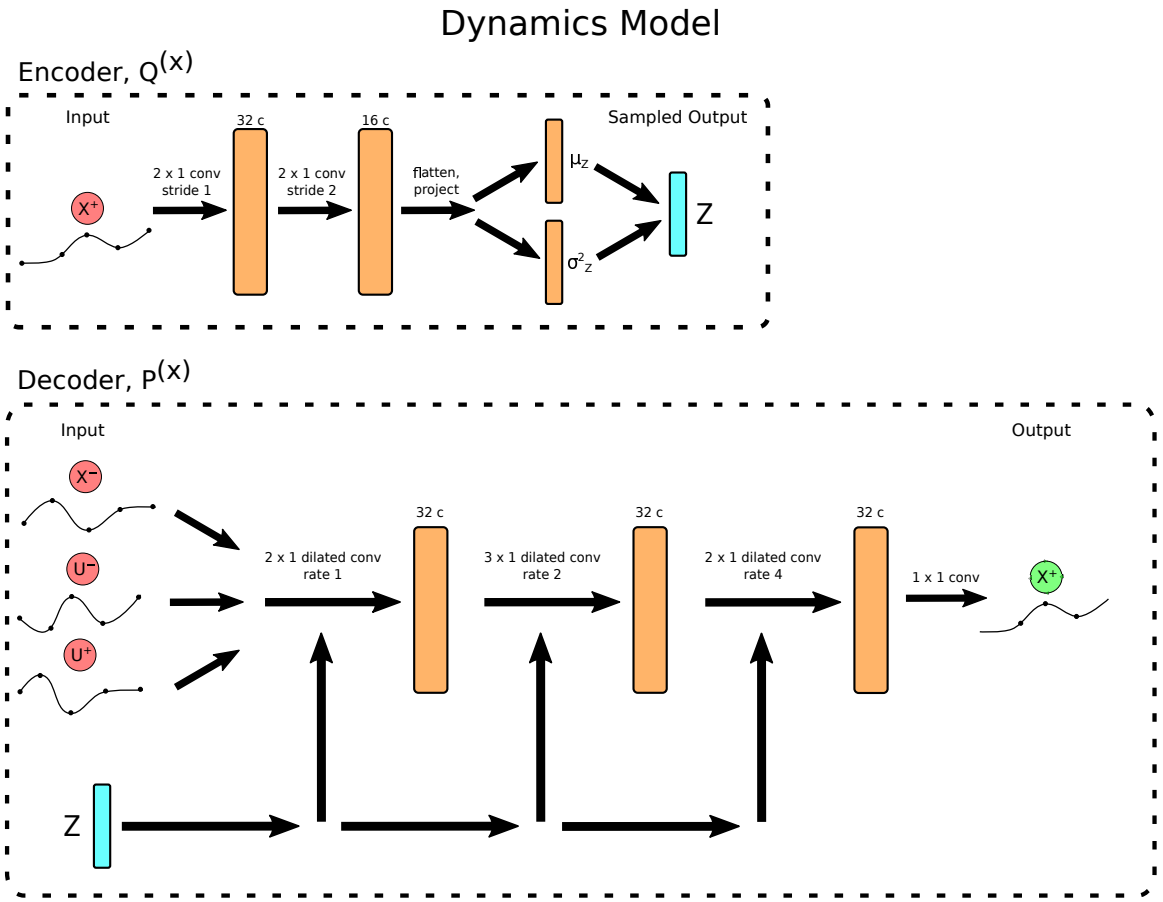
A. Appendix

Here we give a more precise description of the architectures of the models we introduced in the paper. Both the dynamics model and the latent action prior were trained using Adam with the default parameters.

A.1. Dynamics Model

The following figure depicts the detailed encoder and decoder architectures for our dynamics models. The encoder uses 1D-convolutions (across the temporal dimension) and the ReLU activation function. The decoder is autoregressive, using dilated causal 1D-convolutions and the gated activation function described in Section 3.1.

The layer sizes indicated below correspond to the model we trained for the basic Reacher environment. For the obstacle and pushing environments, we used the same encoder architecture. The decoder for those environments had 64 channels in all layers, and had an additional 1×1 convolution with 128 channels before the final layer.



A.2. Latent Action Prior

The architecture for the latent action prior is quite similar to that of our dynamics models as depicted on the previous page. We used the same architecture for all experiments.