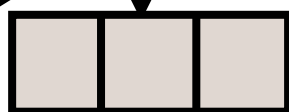
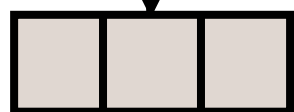


x_1 x_2 

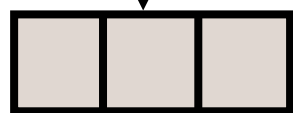
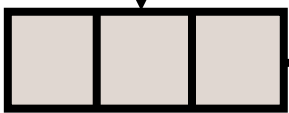
Patch Emb

Patch Emb

Spatial
TransformerSpatial
TransformerCausal
TransformerCausal
Transformer

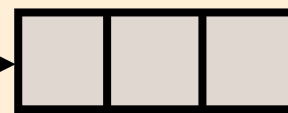
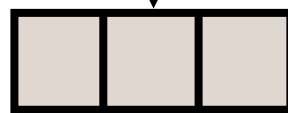
Discretize

Discretize

 z_1  z_2

Encoder

Latent Actions

 z_2 Spatial
Transformer

Patch Emb

 x_2

Decoder