x_1 x_2

Patch Emb

Patch Emb

Latent Actions

z_2

Spatial Transformer

Spatial Transformer

Spatial Transformer

Causal Transformer

Causal Transformer

Discretize

Discretize

Patch Emb

z_1 z_2 x_2

Encoder **Decoder**