Observation

"An illustration of a baby daikon radish in a tutu walking a dog"

$\mathbf{x}^{(i)}$

Text Encoder

$\mathbf{z}^{(i)}$

Image Decoder

Prediction

$\hat{\mathbf{y}}^{(i)}$