discriminative approach

\[ P(s | x) \]

scene, \( s \)

\[ G^{-1} \]

input image, \( x \)

dim(\( s \)) \ll \text{dim}(\( x \))

\( s = \text{scene representation.} \)
Generative Approach (unsupervised)

Find $s$ that makes $X \approx X$. More generally, find $P(s|X)$. The closer $X$ is to $X$, the more like $P(s|X)$ is high iff $X \approx X$. 

Intractable in general. Use variational approximation (later).
Note:
Use Bayes' Rule to compute \( P(s|X) \): 

\[
P(s|X) = \frac{P(X|s) \cdot P(s)}{P(X)}
\]

\[
P(X) = \sum_{s} P(X,s) = \sum_{s} P(X|s) \cdot P(s)
\]

\[
P(X|s) = P(X|X(s))
\]

usually simple (e.g., Gaussian)

\[
P(z|X) = \frac{e^{-\frac{1}{2}z^2 / 2 \sigma^2}}{\sqrt{2\pi} \sigma}
\]
**Variational Autoencoder**

Simplifying assumption:

**Start** Train a NN to compute
(an approximation to) \( p(s|x) \)

\[ \theta(s|x) \rightarrow \text{samples} \rightarrow G \]

Encoder

\( NN \)

Decoder

\( G \)

Train NN to make \( x \approx \tilde{x} \) (for all training points)

Can train \( G \) can be given or can be trained with \( NN \) simultaneously with \( NN \)

Note: must
Note: must perform gradient descent through both $G$ and $NN$ (the sampling operation).

No problem if $G$ is a $NN$.

If $G$ is a graphics program, we must be able to differentiate it. (Hence "differentiable rendering").

Note: unlike discriminative approach, $VAE$ is unsupervised.

No scene labels needed.