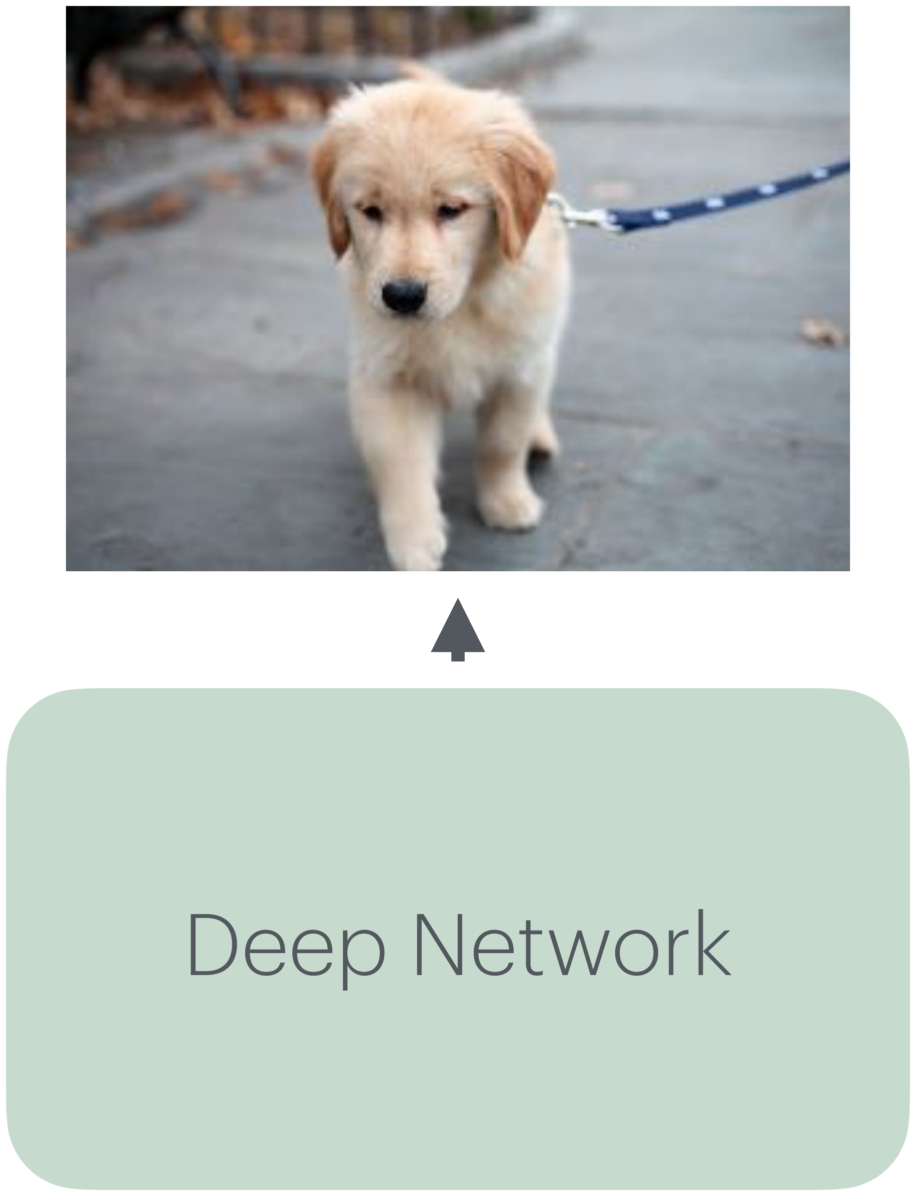
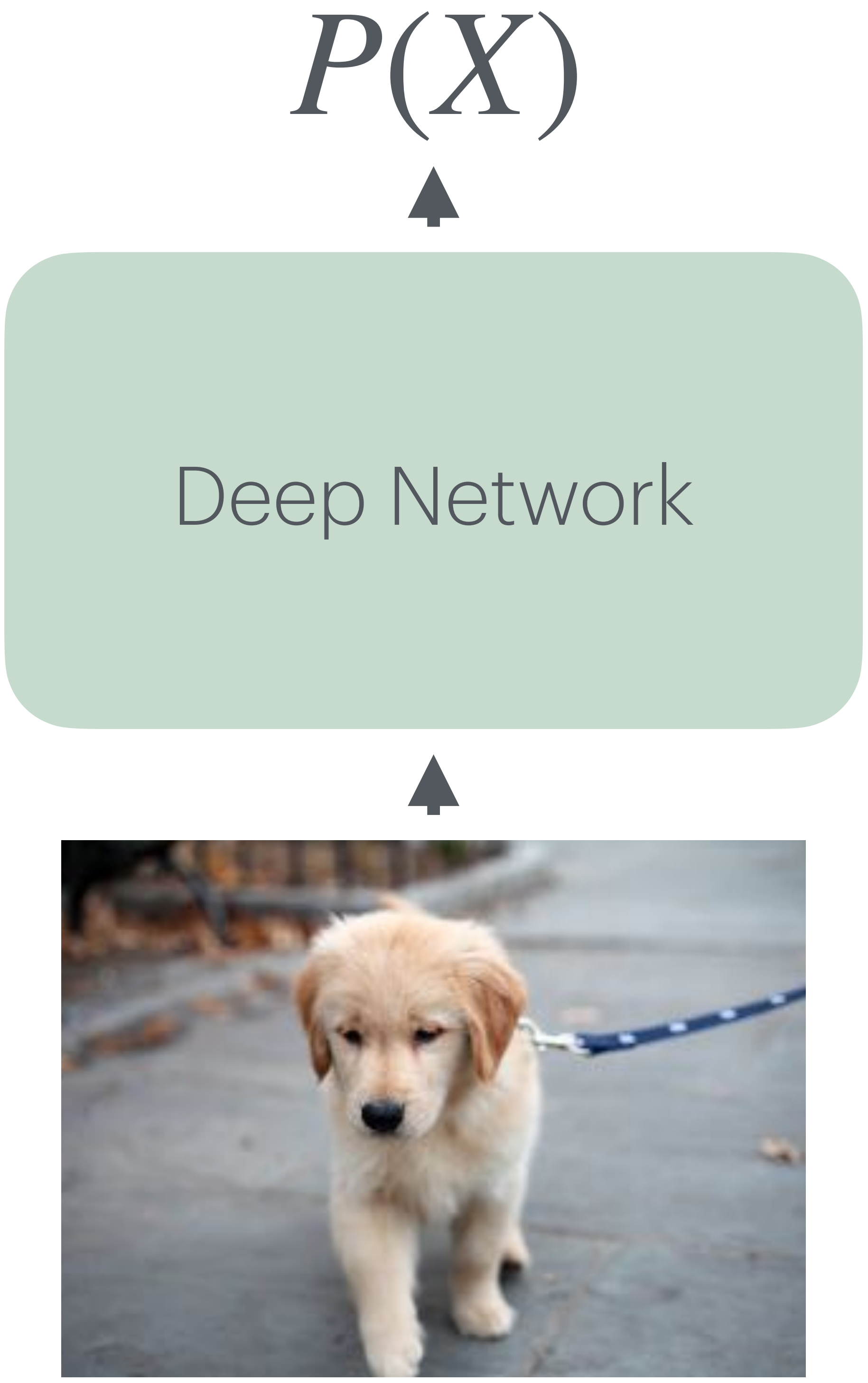
Variational Auto Encoders

Philipp Krähenbühl, UT Austin

Generative models

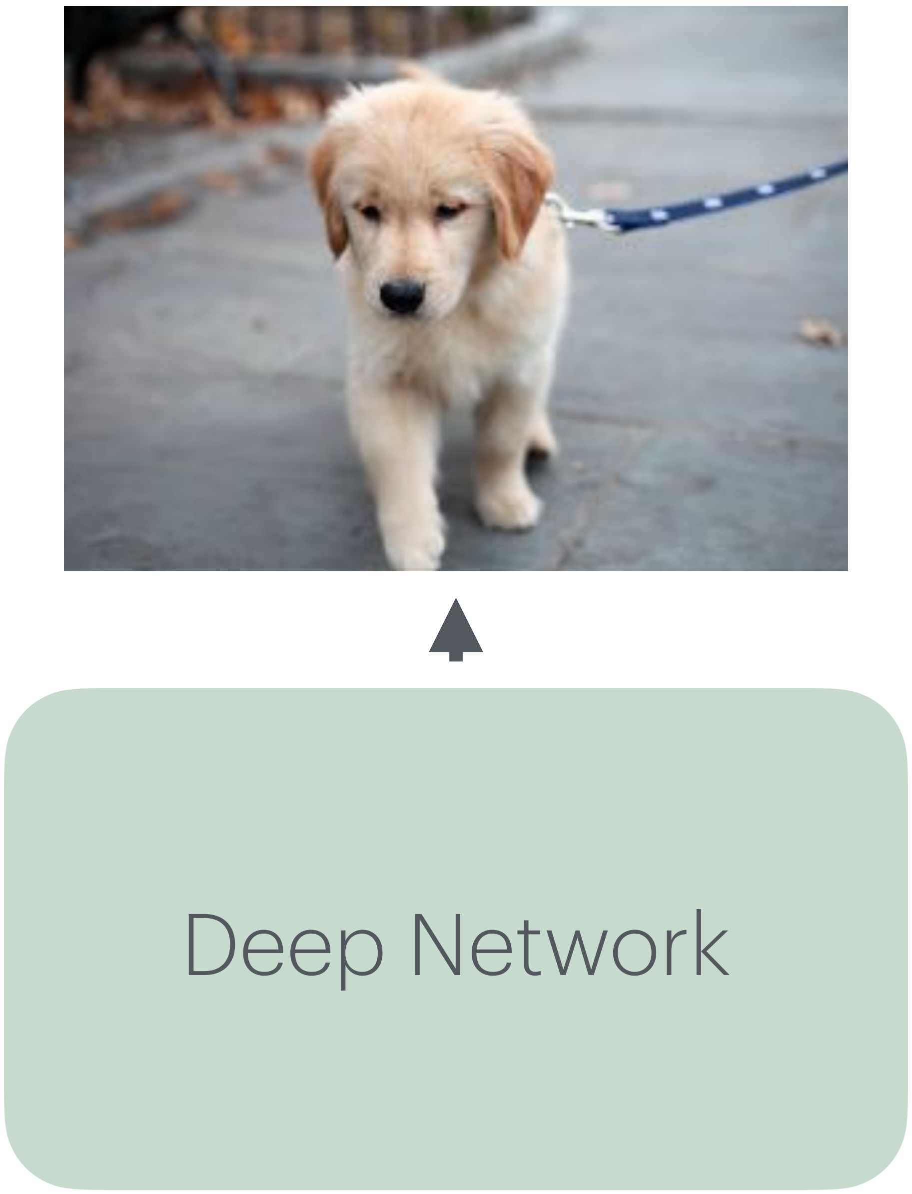
- Two tasks of a generative model $P(X)$
  - Sampling: $x \sim P(X)$
  - Density estimation: $P(X = x)$

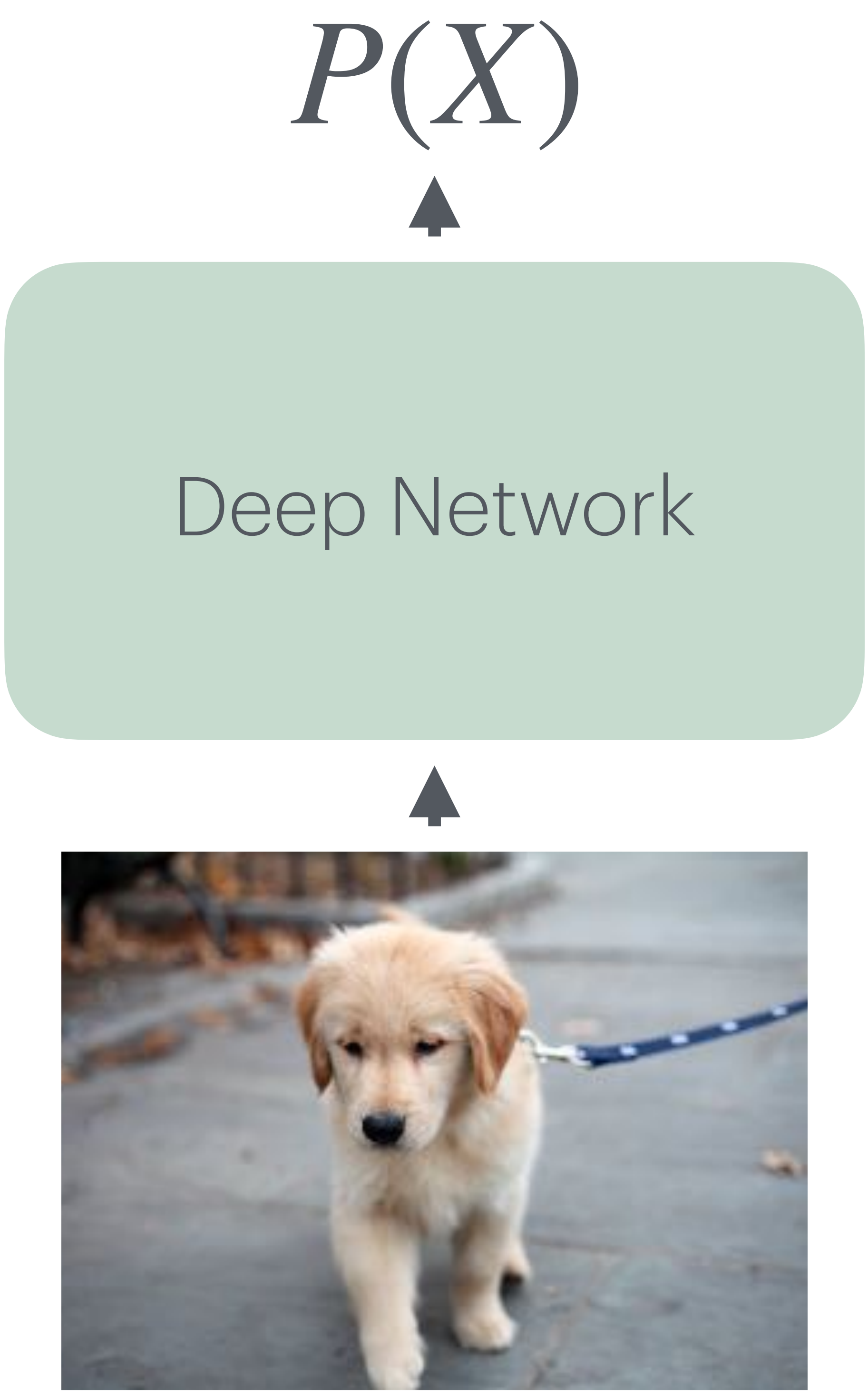




Generative modeling is hard

- Density estimation $P(X = x)$
  - How to ensure $\sum_x P(x) = 1$ for all x
  - Impossible to compute (in general)
- Sampling $x \sim P(X)$
  - What is the input to the network?





Generative models

Two kinds of models

Sampling based $x \sim P(X)$

- Sample $z \sim P(Z)$
- Learn transformation
  - $P(x \mid z)$ or $f: z \rightarrow x$

z Deep Network

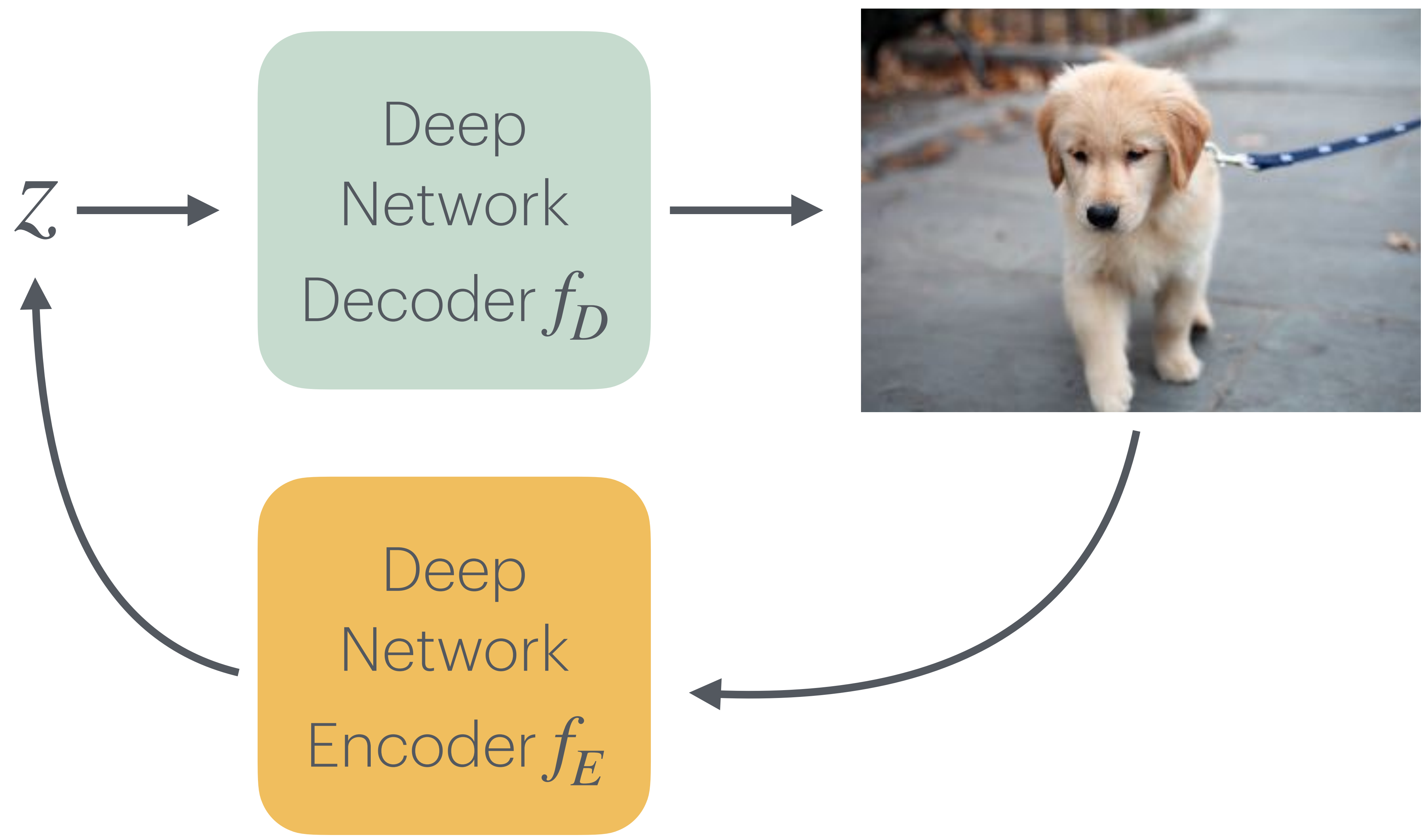
Density estimation based $P(X)$

- Learn special form of $P(X)$
- Model specific sampling / generation

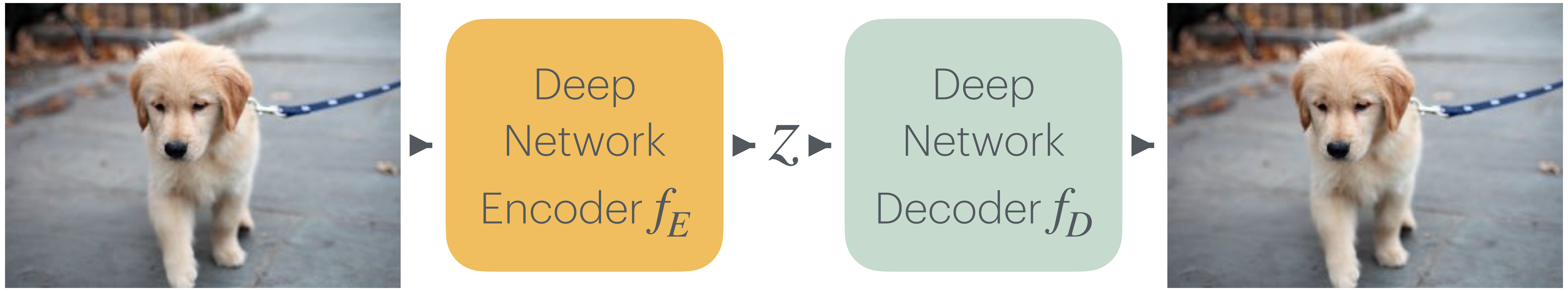
Deep Network $P(X)$

Generative models

- Goal: Learn decoder $f_D : z \to x$
- What should z be?
  - Let a deep network decide
    - Encoder $f_E : x \to z$



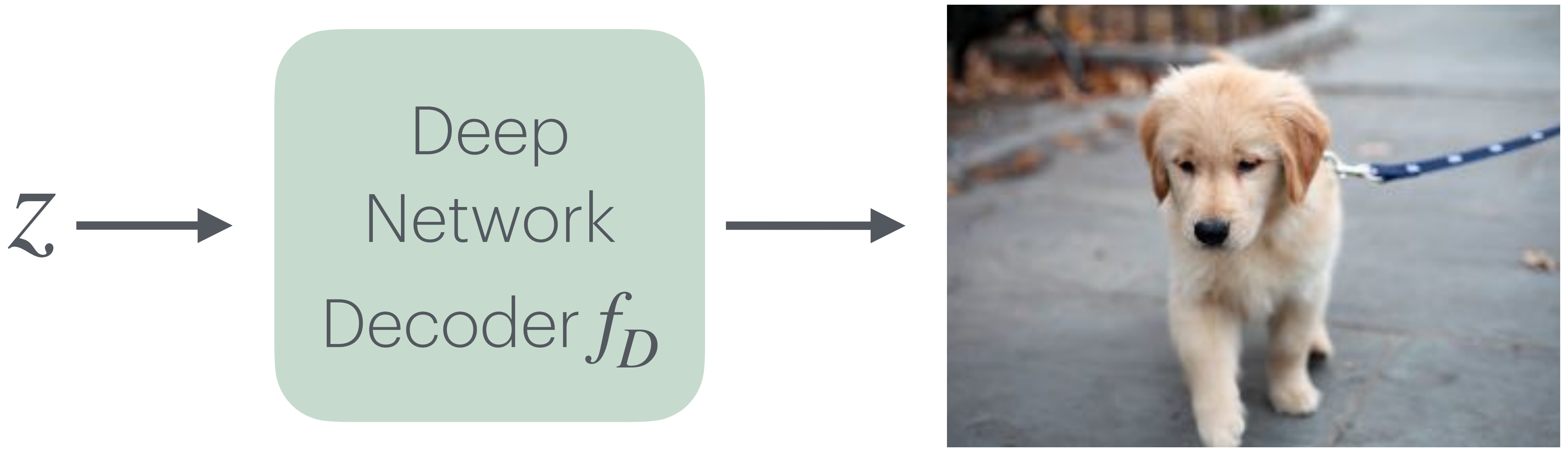
Auto-encoder



- encoder $z = f_E(x)$
- decoder $\hat{x} = f_D(z)$
- Training
    - Supervised learning on large dataset
    - $\ell = E_x\left[\,|f_D(f_E(x)) - x|\,\right]$

Auto-encoder

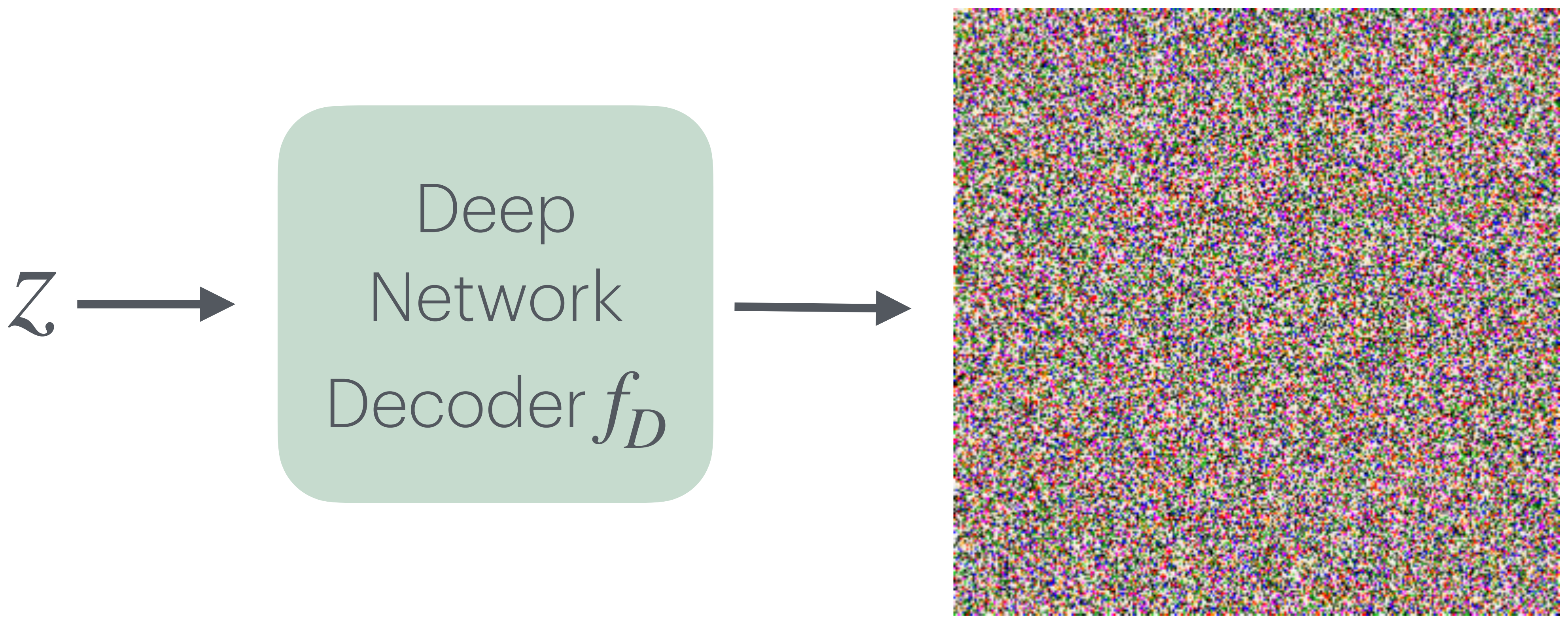
as a Generative model



- Decoder $f_D : z \to x$
- Inference / Sampling
  - What is z at test time?

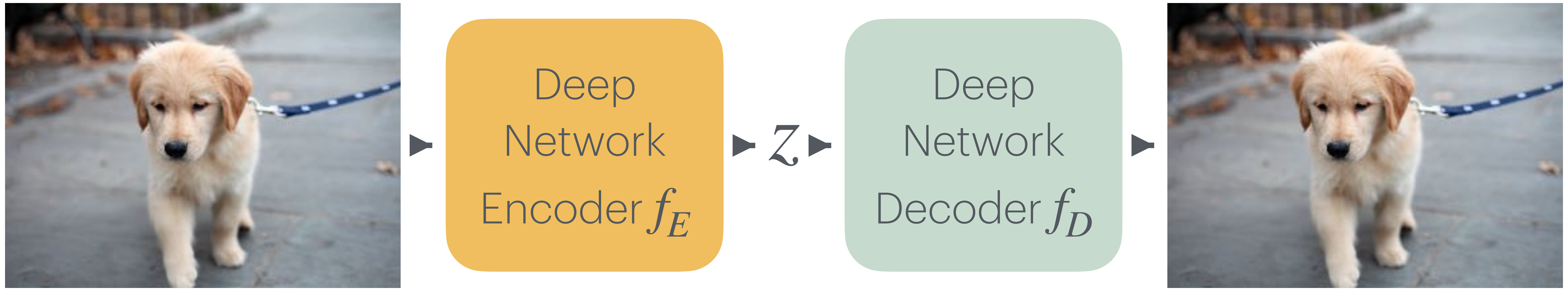
Auto-encoder

Generation



- Decoder $f_D : z \to x$
- Inference / Sampling
  - What is z at test time?
    - Network output -> no new image
    - Random input -> Garbage
    - Interpolation -> Garbage

What does an auto-encoder learn?

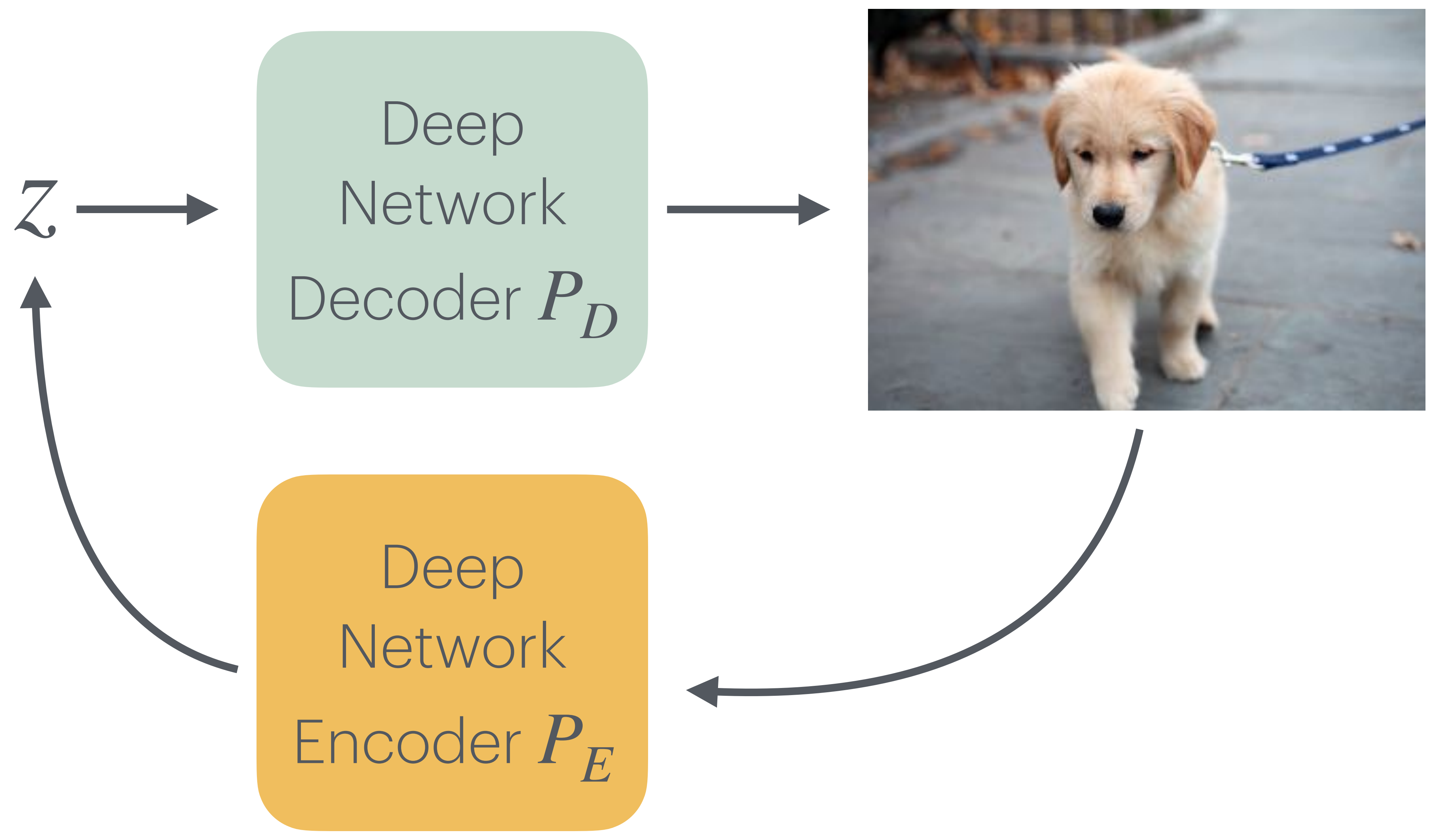


- Compression
- “Invertible” mapping
- Does it learn the structure of images?
  - Only in the limit
  - Perfect compression = understanding
- Poor generation

Variational auto-encoder

A "probabilistic" auto-encoder

- Goal: Learn decoder $P_D(x \mid z)$
- What should z be?
  - Let a deep network decide
    - Encoder $P_E(z \mid x)$

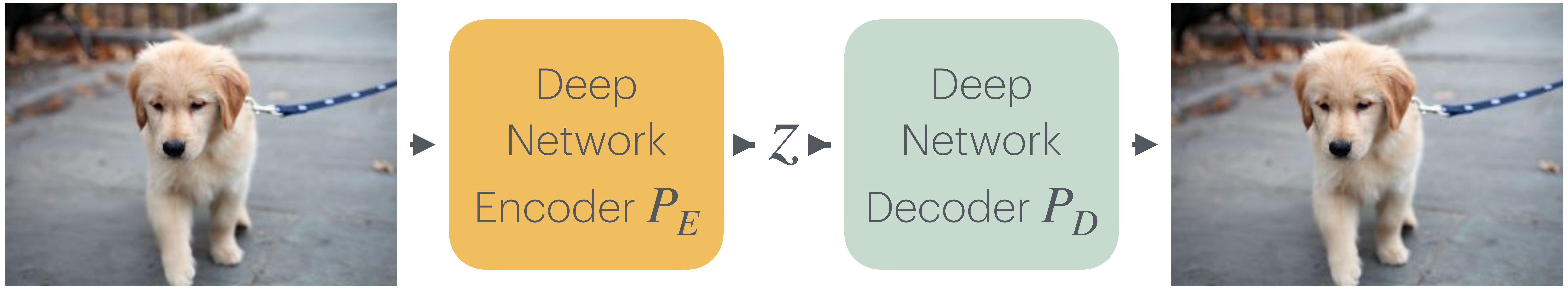


[1] Auto-Encoding Variational Bayes. Kingma et al. 2014.

Variational auto-encoder

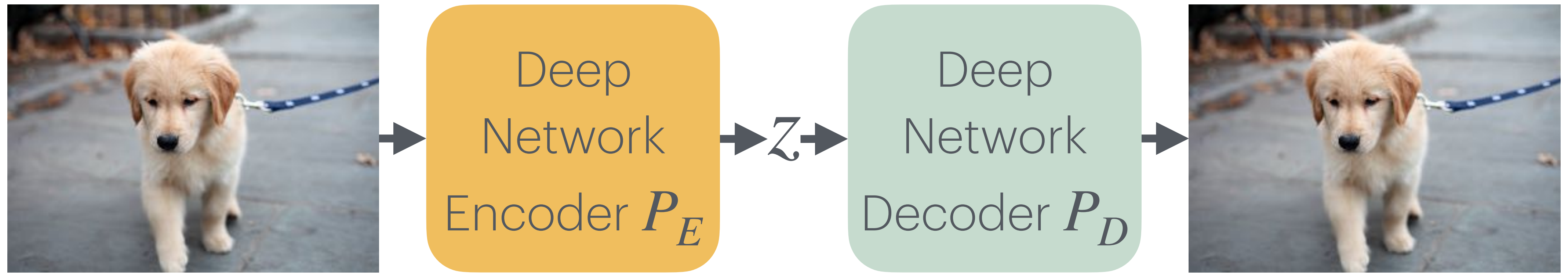
A "probabilistic" auto-encoder

- Decoder $P_D(x \mid z)$ (similar to discriminative model)
- Encoder $P_E(z \mid x)$ (similar to discriminative model)
- Assume $P(Z) = \mathcal{N}(0,1)$
- $$P(x) = \sum_z P_D(x \mid z) P(z)$$
  - $z \sim P(X)$ is equivalent to $z \sim P(Z)$ and $x \sim P(x \mid z)$



Variational auto-encoder

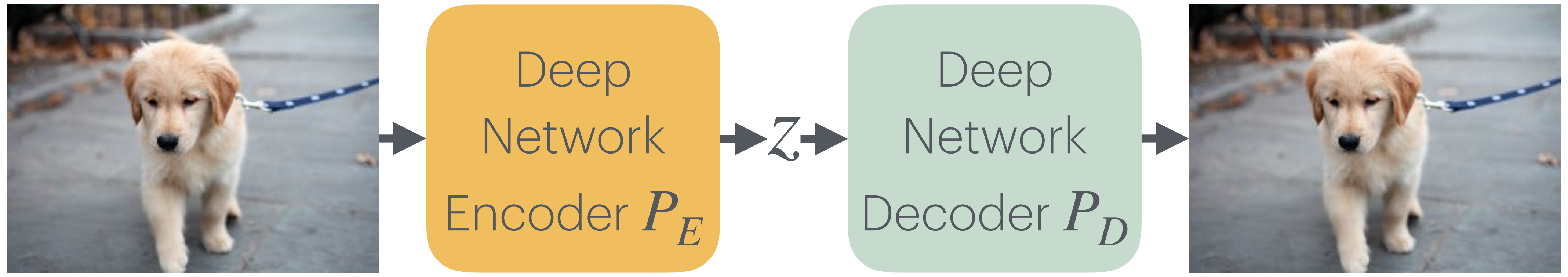
A "probabilistic" auto-encoder



- Decoder $P_D(x \mid z)$ (similar to discriminative model)
- Encoder $P_E(z \mid x)$ (similar to discriminative model)
- Assume $P(Z) = \mathcal{N}(0,1)$
- Bayes rule $P_E(z \mid x) = \dfrac{P_D(x \mid z)P(z)}{P(x)}$ ←intractable

Variational auto-encoder

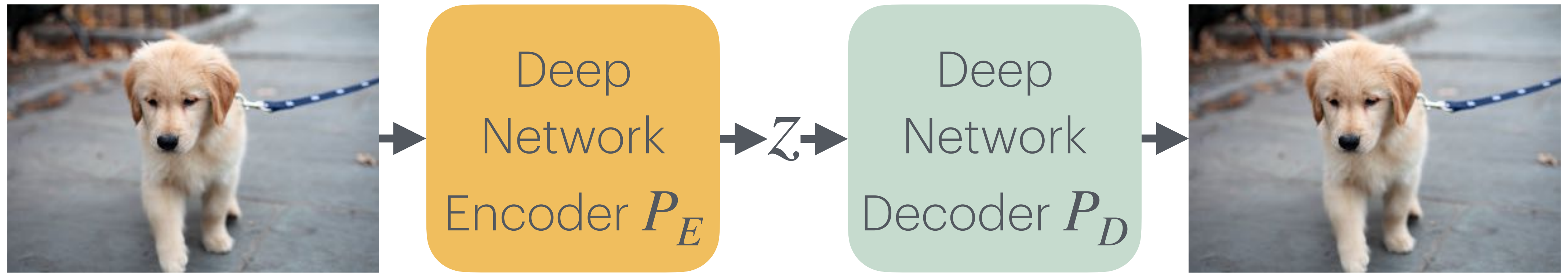
A "probabilistic" auto-encoder



- Decoder $P_D(x \mid z)$ (similar to discriminative model)
- Encoder $Q(z \mid x)$ (similar to discriminative model)
- Assume $P(Z) = \mathcal{N}(0,1)$
  - Bayes rule $P_E(z \mid x) = \dfrac{P_D(x \mid z)P(z)}{P(x)}$ ←intractable
  - Learn $Q \approx P_E$ that minimizes $D_{KL}(Q \mid P_E)$

Variational auto-encoder

A "probabilistic" auto-encoder



- Learn $Q \approx P_E$ that minimizes

$$D_{KL}(Q(z \mid x) \| P_E(z \mid x)) = \log P(x) + E_{z \sim Q}\left[\log \frac{P(z)P_D(x \mid z)}{Q(z \mid x)}\right]$$

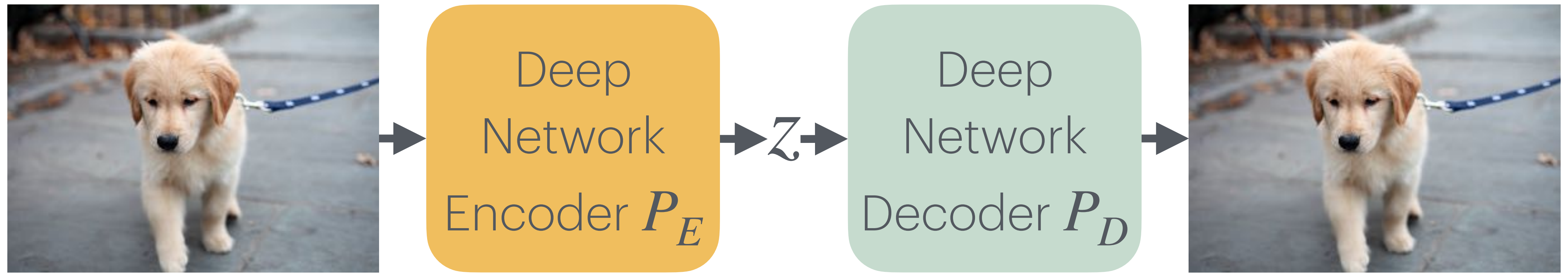
- Maximize $\log P(x)$ of real data, minimize D_{KL}

$$\log P(x) - D_{KL}(Q(z \mid x) \| P_E(z \mid x)) = E_{z \sim Q}\left[\log \frac{Q(z \mid x)}{P(z)P_D(x \mid z)}\right]$$

  - Known as ELBO (Evidence Lower BOund)

Variational auto-encoder

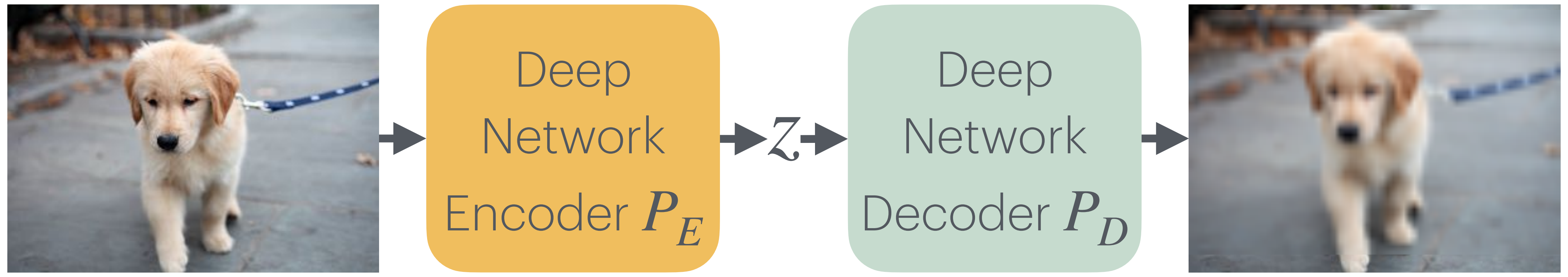
A "probabilistic" auto-encoder



- ELBO $E_{z\sim Q}\left[\log \frac{Q(z\,|\,x)}{P(z)P_D(x\,|\,z)}\right]$ for Gaussians
  - $-\frac{1}{2}\mathbb{E}_{z\sim Q}\left[\|x-\mu_D(z)\|_2^2\right]-\frac{1}{2}\left(N\sigma_Q(x)^2+\|\mu_Q(x)\|_2^2-2N\log\sigma_Q(x)\right)+Const$
- Reparametrization trick
  - For $Q(z\,|\,x)=\mathcal{N}(z;\mu_Q(x),\sigma_Q^2(x))$
  - $\mathbb{E}_{z\sim Q}\left[\|x-\mu_D(z)\|_2^2\right]=\mathbb{E}_{\varepsilon\sim\mathcal{N}(0,1)}\left[\|x-\mu_D(\mu_Q(x)+\varepsilon\sigma_Q(x))\|_2^2\right]$

Variational auto-encoder

A "probabilistic" auto-encoder

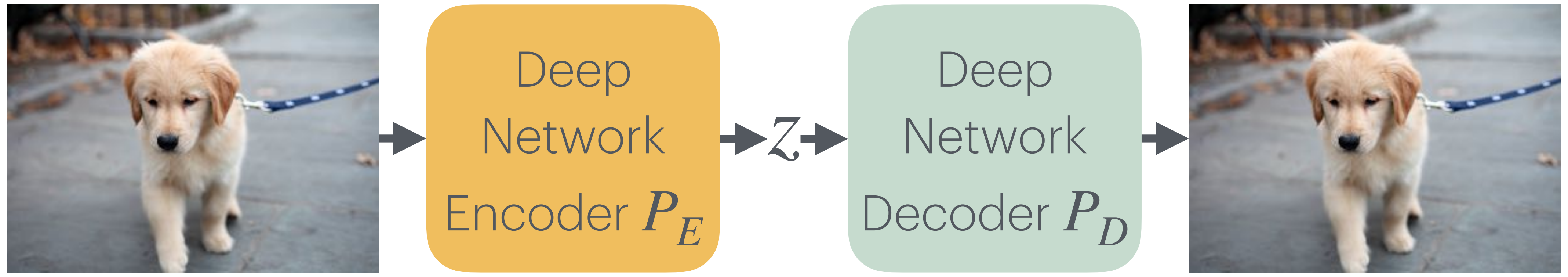


- Can learn $P(X)$ and sampling function $x \sim P$
- Issues
  - Reconstruction loss: Pixel-level l2 loss
    - Blurry outputs
  - Approximation Q: Gaussian assumption
    - Sphere packing in higher dimensions
    - Lots of empty space

[1] Auto-Encoding Variational Bayes. Kingma et al. 2014.

Variational auto-encoder

A "probabilistic" auto-encoder



- Learn a model of $P(x) = P_D(x \mid z)P(z)$ with $P(z) = \mathcal{N}(z; 0,1)$
  - Training: Maximize $P(x)$ of data
  - Approximate $Q \approx P_E$

[1] Auto-Encoding Variational Bayes. Kingma et al. 2014.

References

- [1] Auto-Encoding Variational Bayes. Kingma et al. 2014.