Gradients

$$\nabla_\phi f(w, \phi) = \sum_i \nabla_\phi \mathbb{E}_{q(t_i \mid x_i, \phi)} \log p(x_i \mid t_i, w)$$

Gradients

$$\nabla_\phi f(w,\phi) = \sum_i \nabla_\phi \mathbb{E}_{q(t_i \mid x_i, \phi)} \log p(x_i \mid t_i, w)$$

$$t_i \sim q(t_i \mid x_i, \phi) = \mathcal{N}(m_i, \operatorname{diag}(s_i^2))$$

Gradients

$$\nabla_\phi f(w, \phi) = \sum_i \nabla_\phi \mathbb{E}_{q(t_i | x_i, \phi)} \log p(x_i \mid t_i, w)$$

$$t_i \sim q(t_i \mid x_i, \phi) = \mathcal{N}(m_i, \operatorname{diag}(s_i^2))$$

$$t_i = \varepsilon_i \odot s_i + m_i$$

$$\varepsilon_i \sim p(\varepsilon_i) = \mathcal{N}(0, I)$$

Gradients

$$\nabla_\phi f(w,\phi) = \sum_i \nabla_\phi \mathbb{E}_{q(t_i \mid x_i,\phi)} \log p(x_i \mid t_i, w)$$

$$t_i \sim q(t_i \mid x_i, \phi) = \mathcal{N}(m_i, \operatorname{diag}(s_i^2))$$

$$t_i = \varepsilon_i \odot s_i + m_i = g(\varepsilon_i, x_i, \phi)$$

$$\varepsilon_i \sim p(\varepsilon_i) = \mathcal{N}(0, I)$$

Gradients

$$\nabla_\phi f(w,\phi) = \sum_i \nabla_\phi \mathbb{E}_{q(t_i|x_i,\phi)} \log p(x_i \mid t_i, w)$$

$$= \sum_i \nabla_\phi \mathbb{E}_{p(\varepsilon_i)} \log p(x_i \mid g(\varepsilon_i, x_i, \phi), w)$$

$$t_i \sim q(t_i \mid x_i, \phi) = \mathcal{N}(m_i, \operatorname{diag}(s_i^2))$$

$$t_i = \varepsilon_i \odot s_i + m_i = g(\varepsilon_i, x_i, \phi)$$

$$\varepsilon_i \sim p(\varepsilon_i) = \mathcal{N}(0, I)$$

Gradients

$$\nabla_\phi f(w,\phi) = \sum_i \nabla_\phi \mathbb{E}_{q(t_i|x_i,\phi)} \log p(x_i \mid t_i, w)$$

$$= \sum_i \int p(\varepsilon_i) \nabla_\phi \log p(x_i \mid g(\varepsilon_i, x_i, \phi), w) d\varepsilon_i$$

$$t_i \sim q(t_i \mid x_i, \phi) = \mathcal{N}(m_i, \mathrm{diag}(s_i^2))$$

$$t_i = \varepsilon_i \odot s_i + m_i = g(\varepsilon_i, x_i, \phi)$$

$$\varepsilon_i \sim p(\varepsilon_i) = \mathcal{N}(0, I)$$

Gradients

$$\nabla_\phi f(w,\phi) = \sum_i \nabla_\phi \mathbb{E}_{q(t_i|x_i,\phi)} \log p(x_i \mid t_i, w)$$

$$= \sum_i \int p(\varepsilon_i) \nabla_\phi \log p(x_i \mid g(\varepsilon_i, x_i, \phi), w) d\varepsilon_i$$

$$= \sum_i \mathbb{E}_{p(\varepsilon_i)} \nabla_\phi \log p(x_i \mid g(\varepsilon_i, x_i, \phi), w)$$

$$t_i \sim q(t_i \mid x_i, \phi) = \mathcal{N}(m_i, \operatorname{diag}(s_i^2))$$

$$t_i = \varepsilon_i \odot s_i + m_i = g(\varepsilon_i, x_i, \phi)$$

$$\varepsilon_i \sim p(\varepsilon_i) = \mathcal{N}(0, I)$$

Gradients

$$\nabla_\phi f(w,\phi) = \sum_i \nabla_\phi \mathbb{E}_{q(t_i|x_i,\phi)} \log p(x_i \mid t_i, w)$$

$$= \sum_i \int p(\varepsilon_i) \nabla_\phi \log p(x_i \mid g(\varepsilon_i, x_i, \phi), w) d\varepsilon_i$$

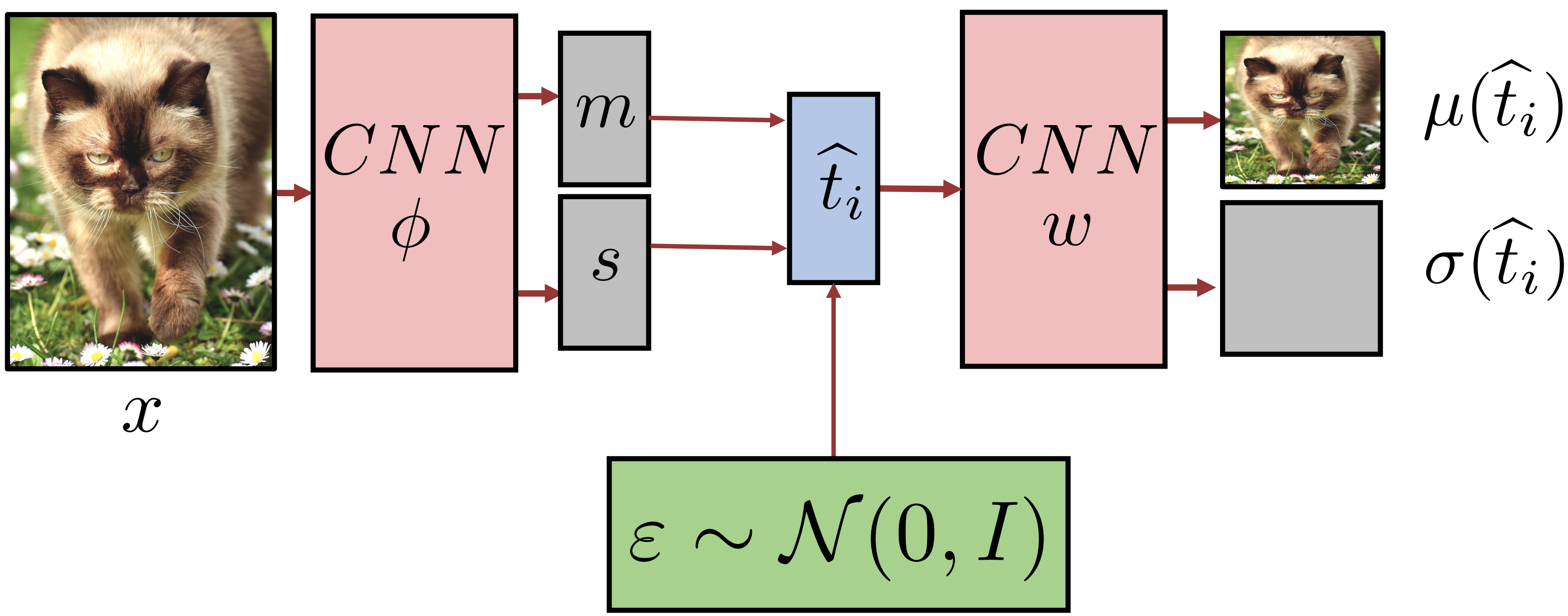
$$= \sum_i \mathbb{E}_{p(\varepsilon_i)} \nabla_\phi \log p(x_i \mid g(\varepsilon_i, x_i, \phi), w)$$

Gradients

$$\nabla_\phi f(w,\phi) = \sum_i \mathbb{E}_{p(\varepsilon_i)} \nabla_\phi \log p(x_i \mid g(\varepsilon_i, x_i, \phi), w)$$

$$p(\varepsilon_i) = \mathcal{N}(0, I)$$

$$t_i = \varepsilon_i \odot s_i + m_i = g(\varepsilon_i, x_i, \phi)$$



Variational Autoencoder summary

- Infinite mixture of Gaussians
- To learn: EM + approximate q with Gaussians + stochastic variational inference
- Like plain autoencoder, but with noise and KL regularization
- Generates nice images