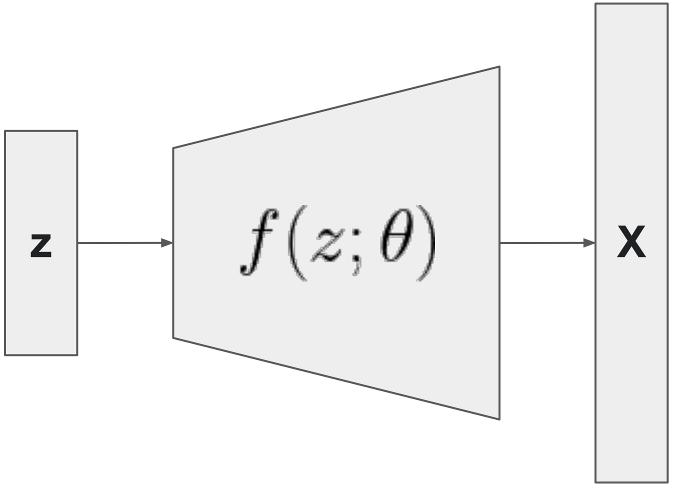
β-Variational Autoencoder

Berk Can Özmen

July 2021



A dataset X, $x \sim p(x)$

A model $p_\theta(x) \approx p(x)$

A latent variable representation $p_\theta(x) = \int p_\theta(x|z)p_\theta(z)dz$

s.t $p_\theta(z) = \mathcal{N}(0, I)$

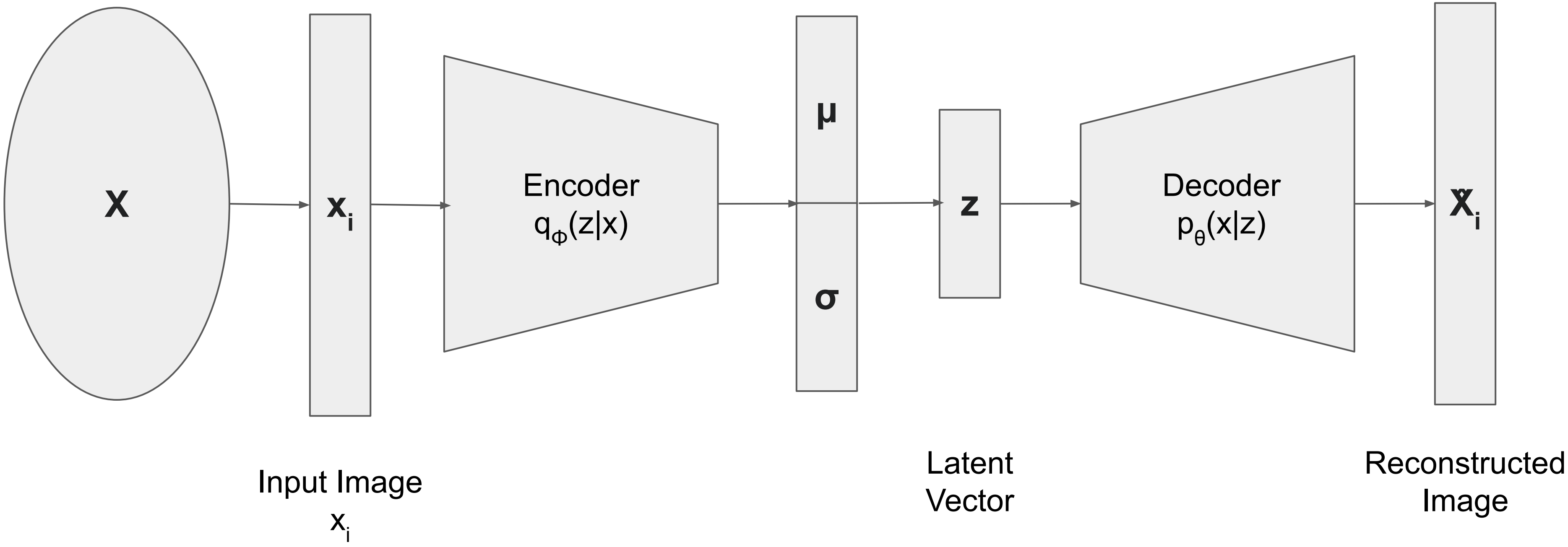
Idea: Construct a tractable $q_\phi(z|x)$ to approximate $p_\theta(z|x)$

$$D_{KL}[q_\phi(z|x)||p_\theta(z|x)] = \mathbb{E}_{z\sim q_\phi(z|x)}[\log q_\phi(z|x) - \log p_\theta(z|x)]$$

$$= \mathbb{E}_{z\sim q_\phi(z|x)}[\log q_\phi(z|x) - \log p_\theta(x|z) - \log p_\theta(z)] + \log p_\theta(x)$$

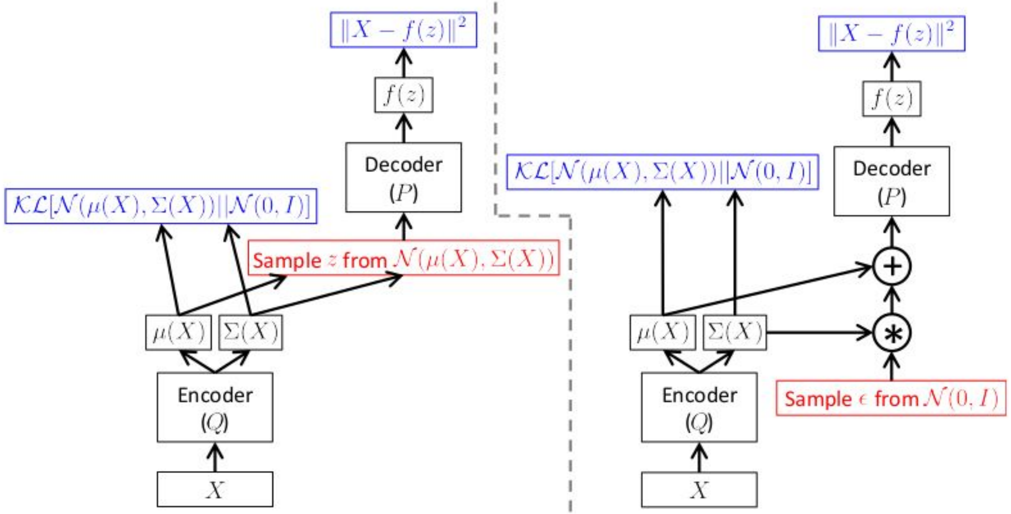
$$\Rightarrow \log p_\theta(x) - D_{KL}[q_\phi(z|x)||p_\theta(z|x)] = \mathbb{E}_{z\sim q_\phi(z|x)}[\log p_\theta(x|z)] - D_{KL}[q_\phi(z|x)||p_\theta(z)]$$

W.r.t $\mathbb{E}_{x\sim p(x)}[\cdot]$



$$\mathcal{L}(\theta, \phi; \mathbf{x}, \mathbf{z}, \beta) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[\log p_\theta(\mathbf{x}|\mathbf{z})] - \beta\ D_{KL}(q_\phi(\mathbf{z}|\mathbf{x})||p(\mathbf{z}))$$

The Reparameterization Trick



(Doersch,2016)

$$\mathbb{E}_{x\sim p(x)}\left[\mathbb{E}_{\epsilon\sim\mathcal{N}(0,I)}[\log p_\theta(x|z = \mu(x) + \Sigma(x)\cdot\epsilon)] - \beta D_{KL}[q_\phi(z|x)||p_\theta(z)]\right]$$

Implementation - Loss function

Assumption: pixels are independent given z and distributed normally

$$\log p_\theta(x|z) = \sum_n \log p_\theta(x_n|z) \propto -\sum_n \|x_n - \hat{x}_n\|^2$$

Reconstruction - MSE

$$\mathrm{D}_{KL}(p||q) = \frac{1}{2}\left[\mu_p^T\mu_p + tr(\Sigma_p) - d - \log|\Sigma_p|\right] \;, where\ q \sim \mathcal{N}(0, I)\ and\ p \sim \mathcal{N}(\mu_p, \Sigma_p)$$

Information Theoretic Perspective

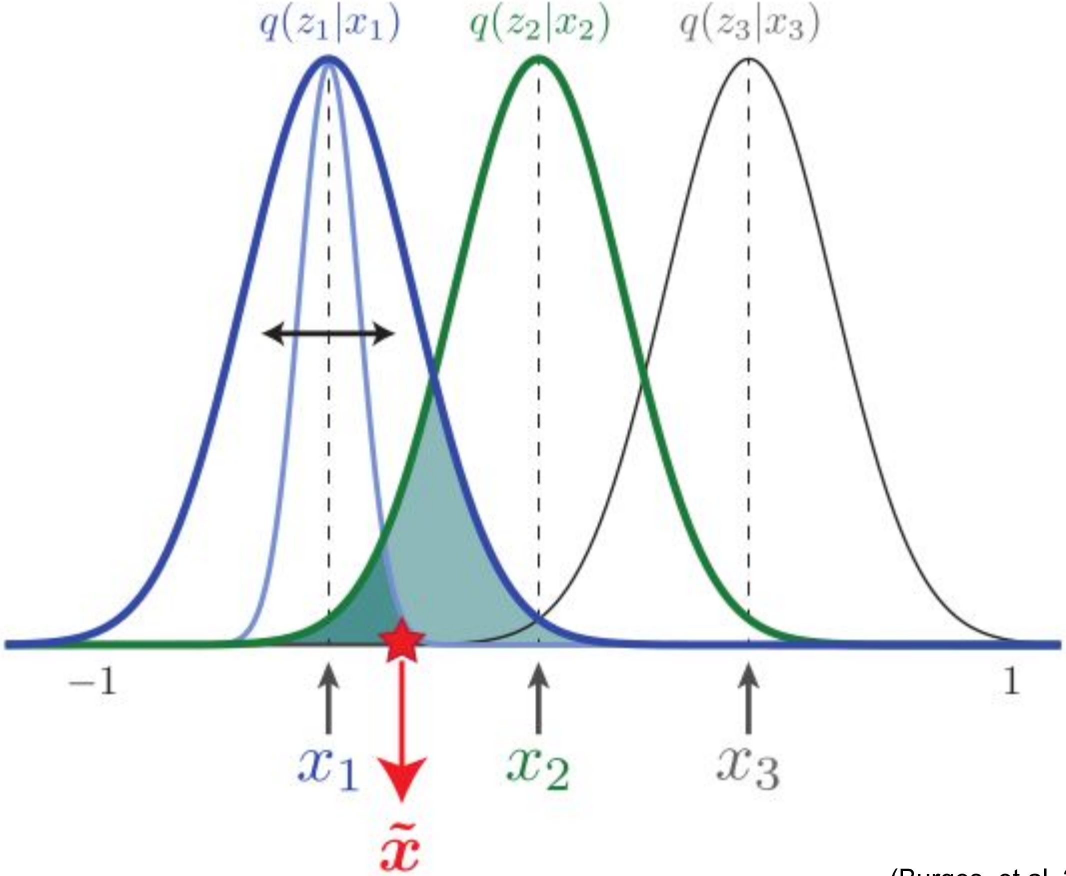
$$D_{KL}[p(x|z)||q(x|z)] \geq 0 \Rightarrow \int dx\ p(x|z) \log p(x|z) \geq \int dx\ p(x|z) \log q(x|z)$$

$$\begin{aligned}
I_{gen}(Z;X) &= \int \int dxdz\ p(x,z) \log \frac{p(x|z)}{p(x)} \\
&= \int dz\ p(z) \left[\int dx\ p(x|z) \log p(x|z) - \int dx\ p(x|z) \log p(x) \right] \\
&\geq \int dz\ p(z) \left[\int dx\ p(x|z) \log p_\theta(x|z) - \int dx\ p(x|z) \log p(x) \right] \\
&= \int \int dxdz\ p(x,z) \log \frac{p_\theta(x|z)}{p(x)} \\
&= \int dx\ p(x) \int dz\ q_\phi(z|x) \log \frac{p_\theta(x|z)}{p(x)} \\
&= \left(- \int dx\ p(x) \log p(x) \right) + \left(\int dx\ p(x) \int dz\ q_\phi(z|x) \log p_\theta(x|z) \right) \\
&= H(x) + \mathbb{E}_{x \sim p(x)} \left[\mathbb{E}_{z \sim q_\phi(z|x)} [\log p_\theta(x|z)] \right]
\end{aligned}$$

Information Theoretic Perspective

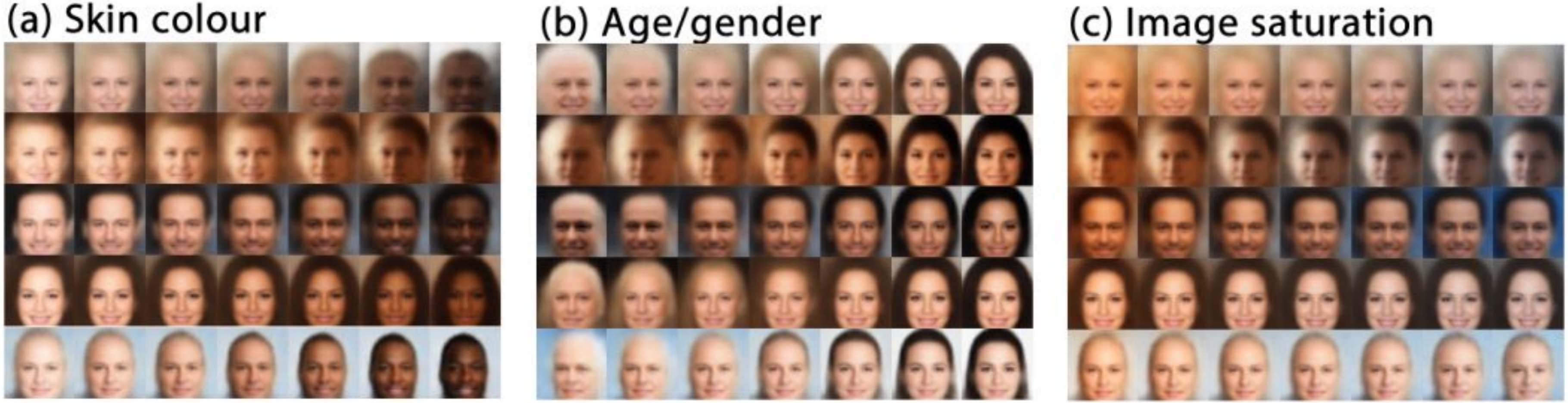
$$\begin{aligned}
I_{rep}(Z;X) &= \int\int dx dz\ p(x,z)\log\frac{q_\phi(z|x)}{p(z)} \\
&\leq \int\int dx dz\ p(x,z)\log\frac{q_\phi(z|x)}{p_\theta(z)} \\
&= \int dx\ p(x)\int dz\ q_\theta(z|x)\log\frac{q_\phi(z|x)}{p_\theta(z)} \\
&= \mathbb{E}_{x\sim p(x)}\left[D_{KL}[q_\phi(z|x)||p_\theta(z)]\right]
\end{aligned}$$

$$I_{gen}(Z;X) - \beta I_{rep}(Z;X) \geq \underbrace{\mathbb{E}_{x\sim p(x)}\left[\mathbb{E}_{z\sim q_\phi(z|x)}[\log p_\theta(x|z)]\right]}_{\text{Information to reconstruct}} - \underbrace{\beta\mathbb{E}_{x\sim p(x)}\left[D_{KL}[q_\phi(z|x)||p_\theta(z)]\right]}_{\text{Extra Information}},$$



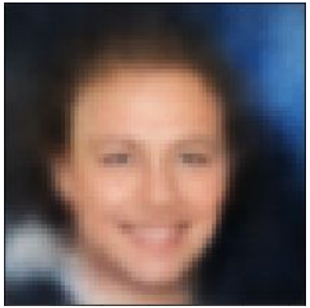
(Burges, et al. 2018)

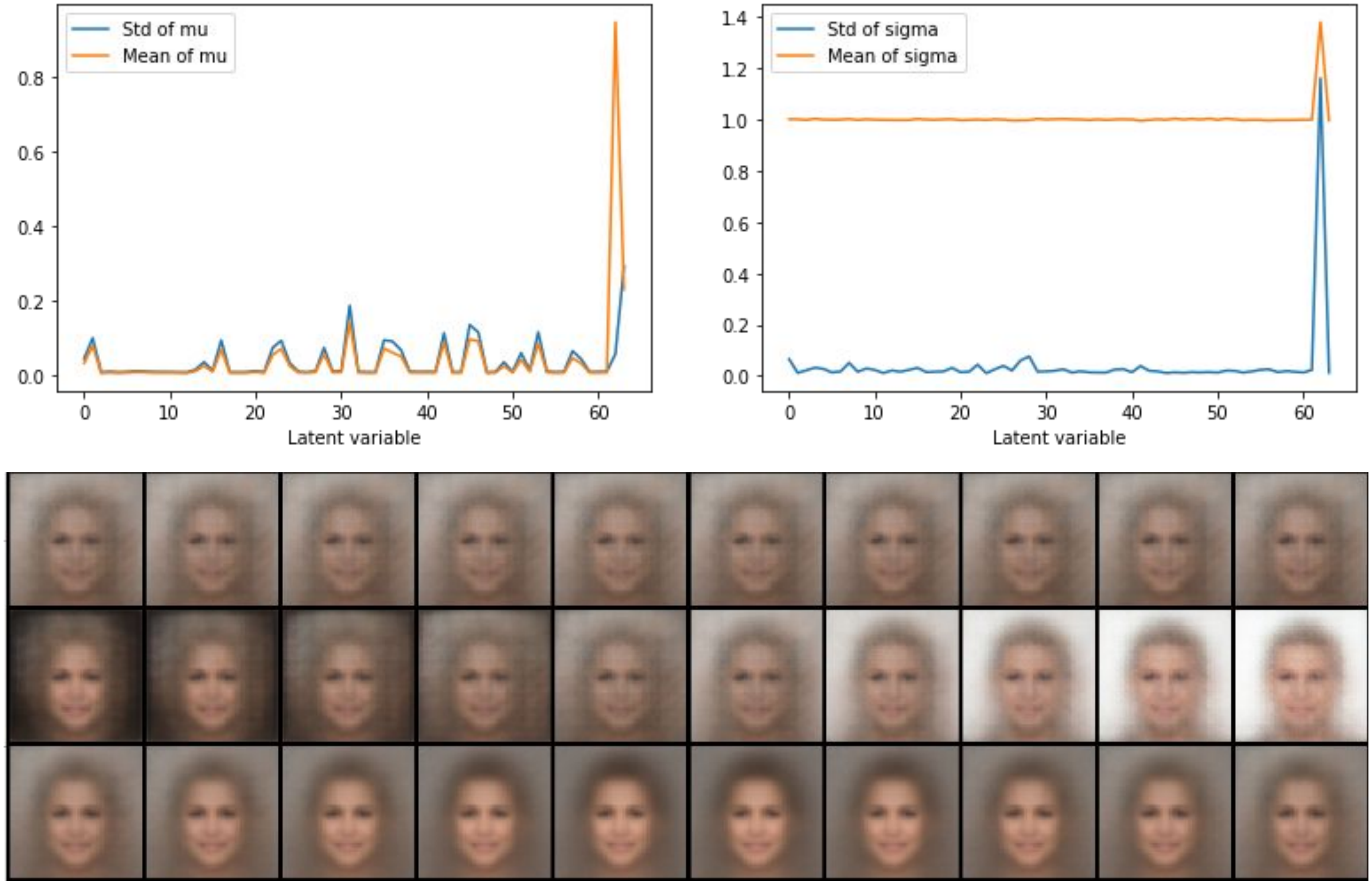
Results (Paper)



(Botvinick, et al. 2017)

Results (Implementation)




Std of mu
Mean of mu
0.0
0.2
0.4
0.6
0.8
0
10
20
30
40
50
60
Latent variable
Std of sigma
Mean of sigma
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0
10
20
30
40
50
60
Latent variable

References

Alemi, A. A., Poole, B., Fischer, I., Dillon, J. V., Saurous, R. A., & Murphy, K. (2018). Fixing a Broken ELBO. https://doi.org/https://arxiv.org/abs/1711.00464

Botvinick, M., Burgess, C., Glorot, X., Higgins, I., Lerchner, A., Matthey, L., Mohamed, S., & Pal, A. (2017). beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework.

Burgess, C., Desjardins, G., Higgins, I., Lerchner, A., Matthey, L., Pal, A., & Watters, N. (2018). Understanding disentangling in beta-VAE. https://doi.org/https://arxiv.org/abs/1804.03599

Doersch, C. (2016). Tutorial on Variational Autoencoders. https://doi.org/https://arxiv.org/abs/1606.05908