CAPTURING LABEL CHARACTERISTICS IN VAEs

MATHÉMATIQUES VISION APPRENTISSAGE

Mohamed Benyahia-Tancrede Martinez-Hadrien Levechin

école normale supérieure paris–saclay

Background and Context

Variational Autoencoders (VAEs) are widely used for generative modeling, but their basic design lacks supervised conditional generation capabilities.
A common approach adds a projection of one-hot labels OHE(y) via an MLP to the latent variable z, producing z+MLP(OHE(y)). While effective in some cases, this raises concerns about fully utilizing label information and ensuring that similar latents with slight label differences generate "close" outputs.

Proposed solution

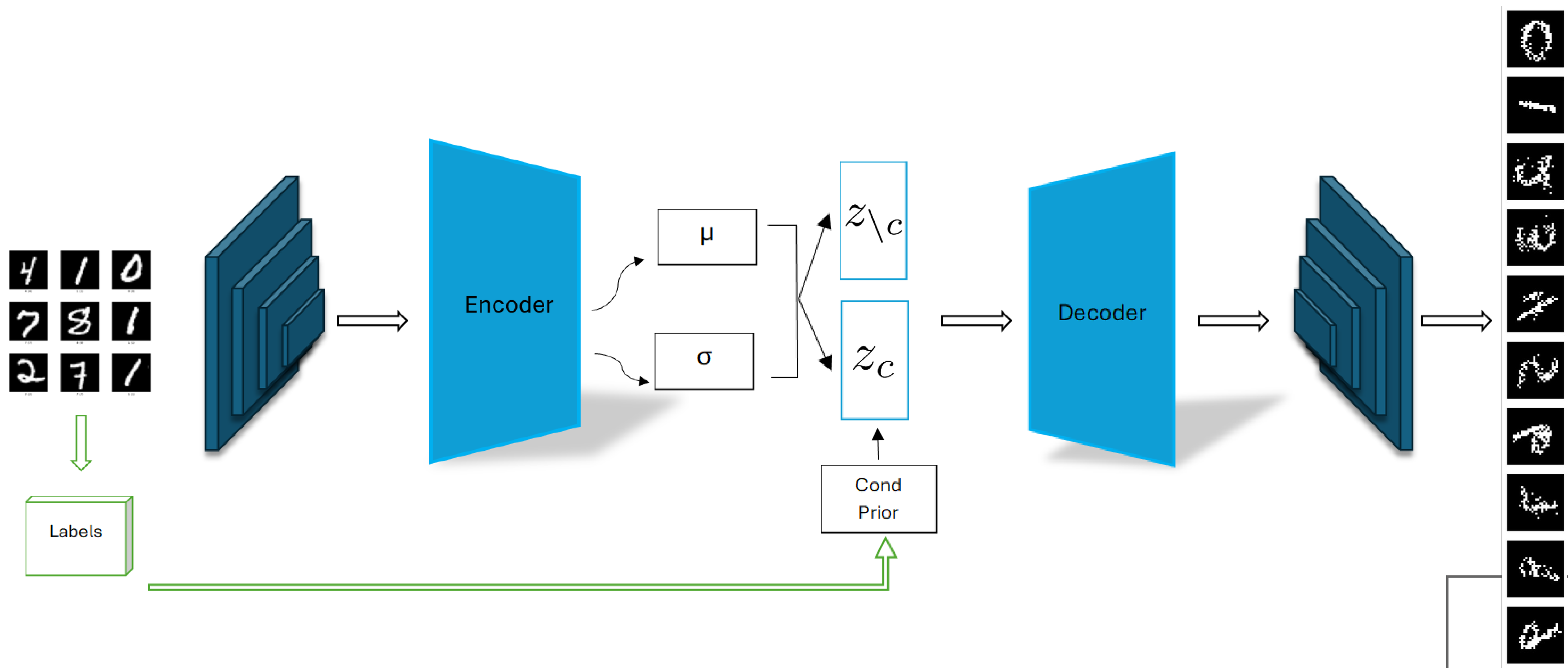
$$z = \{z_c, z_{\backslash c}\}$$

z_c characteristical latent

$p_\psi(z_c|y)$ characteristic latent conditional distribution

$z_{\backslash c}$ style latent

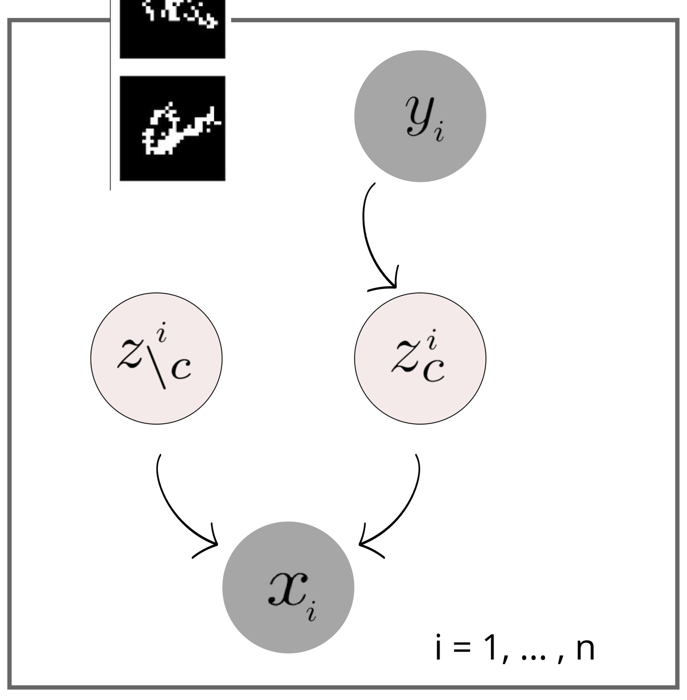
A Characteristic Capturing Variationnal AutoEncoders (CCVAEs)



$$\mathcal{L}(x,y) = -\mathbb{E}_{p(z)}(\log p_\theta(x|z)) + \beta\mathcal{D}_{KL}(q_\phi(z_{\backslash c}|x)|p(z_{\backslash c})) + \beta\mathcal{D}_{KL}(q_\phi(z_c|x)|p(z_c|y)) + CE(y, decoder(z))$$

$$\mathcal{L}(x) = -\mathbb{E}_{p(z)}(\log p_\theta(x|z)) + \beta\mathcal{D}_{KL}(q_\phi(z|x)|p(z)) + \mathbb{E}_{q_\phi(z|x)}\left[\mathrm{H}\left(q_\gamma(y|z_c)\right)\right]$$

$$\mathcal{L}(\mathcal{S},\mathcal{U}) = \sum_{(x,y)\in\mathcal{S}} \mathcal{L}(x,y) + \sum_{x\in\mathcal{U}} \mathcal{L}(x)$$



Our Personal Experiments

Figure 1: Reconstructed images $\beta = 0.05$

Figure 2: Reconstructed images $\beta = 0.4$

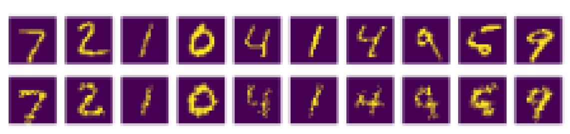


Figure 9: Reconstructed images latent dim $= 16$

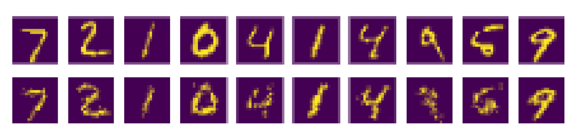


Figure 10: Reconstructed images latent dim $= 128$

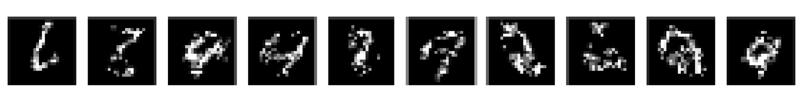


Figure 3: Conditionally generated images $\beta = 0.001$

Figure 4: Conditionally generated images $\beta = 0.05$

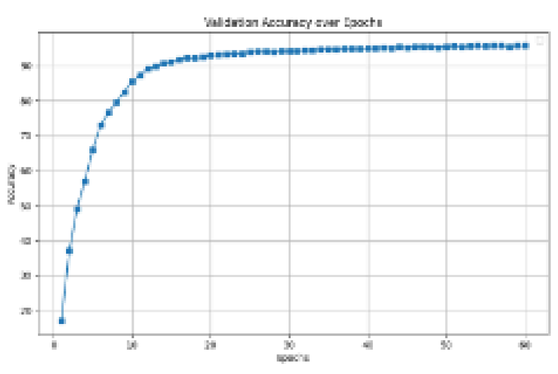


Figure 11: Validation accuracy $\beta = 0.05$

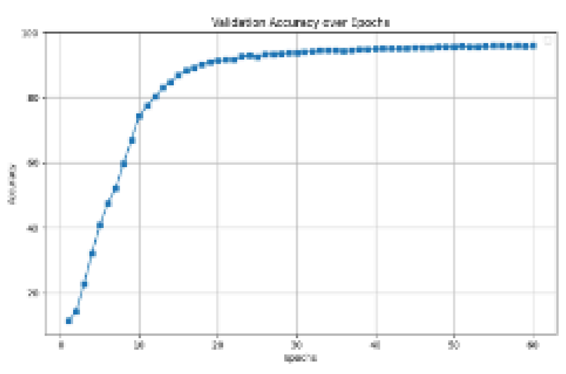


Figure 12: Validation accuracy $\beta = 0.4$

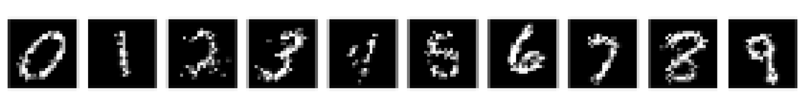


Figure 5: Conditionally generated images $\beta = 0.2$

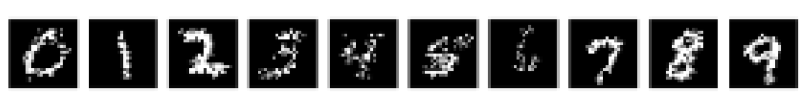


Figure 6: Conditionally generated images $\beta = 0.4$

Figure 7: Interpolated images $\beta = 0.05$

Figure 8: Interpolated images $\beta = 0.4$

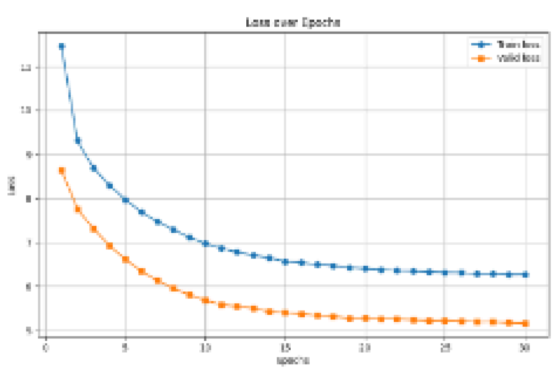


Figure 13: Train and Validation losses latent dim $= 16$

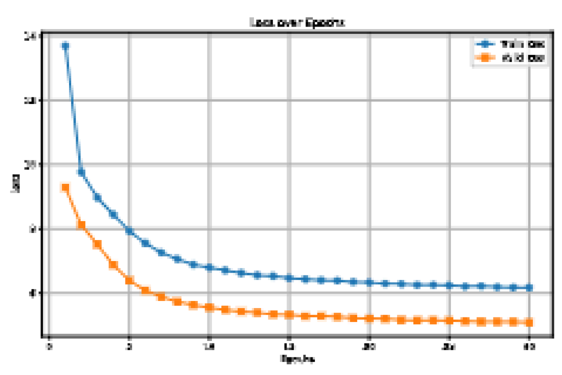


Figure 14: Train and Validation loss latent dim $= 128$