Dauphine | PSL
UNIVERSITÉ PARIS

Variational autoencoder for image generation

Bencheikh Elhocine Mohammed Amine
Benmaiza Mohammed
Chikh Yanis

Data science project
IASD

1. Autoencoder

01

1. Autoencoder (1/3)

A deep neural network architecture, composed of an encoder network and decoder, aims to find optimal compression for input data.

- The encoder component learns to map any input vector X to a low latent space vector Y.
- The decoder component learns to map back the compressed vector Y from latent space to the original space (reconstructing original data vector X) and generate X'.
- The loss function to optimize is the reconstruction error (minimize the difference between X' and X).

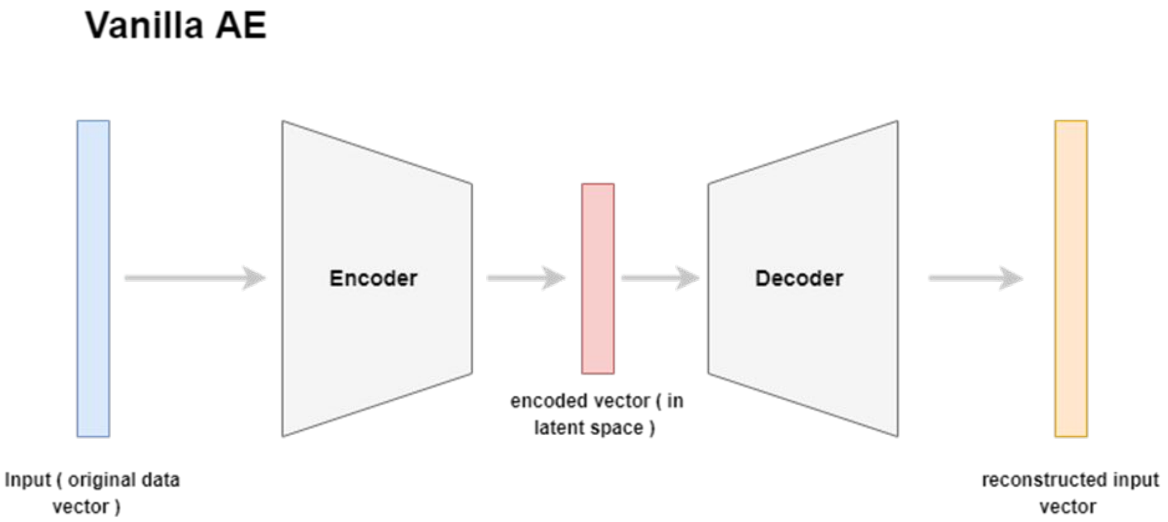


Figure : Autoencoder architecture (authors).

1. Autoencoder (2/3)

Is it possible to use Autoencoders architecture to generate new data samples?

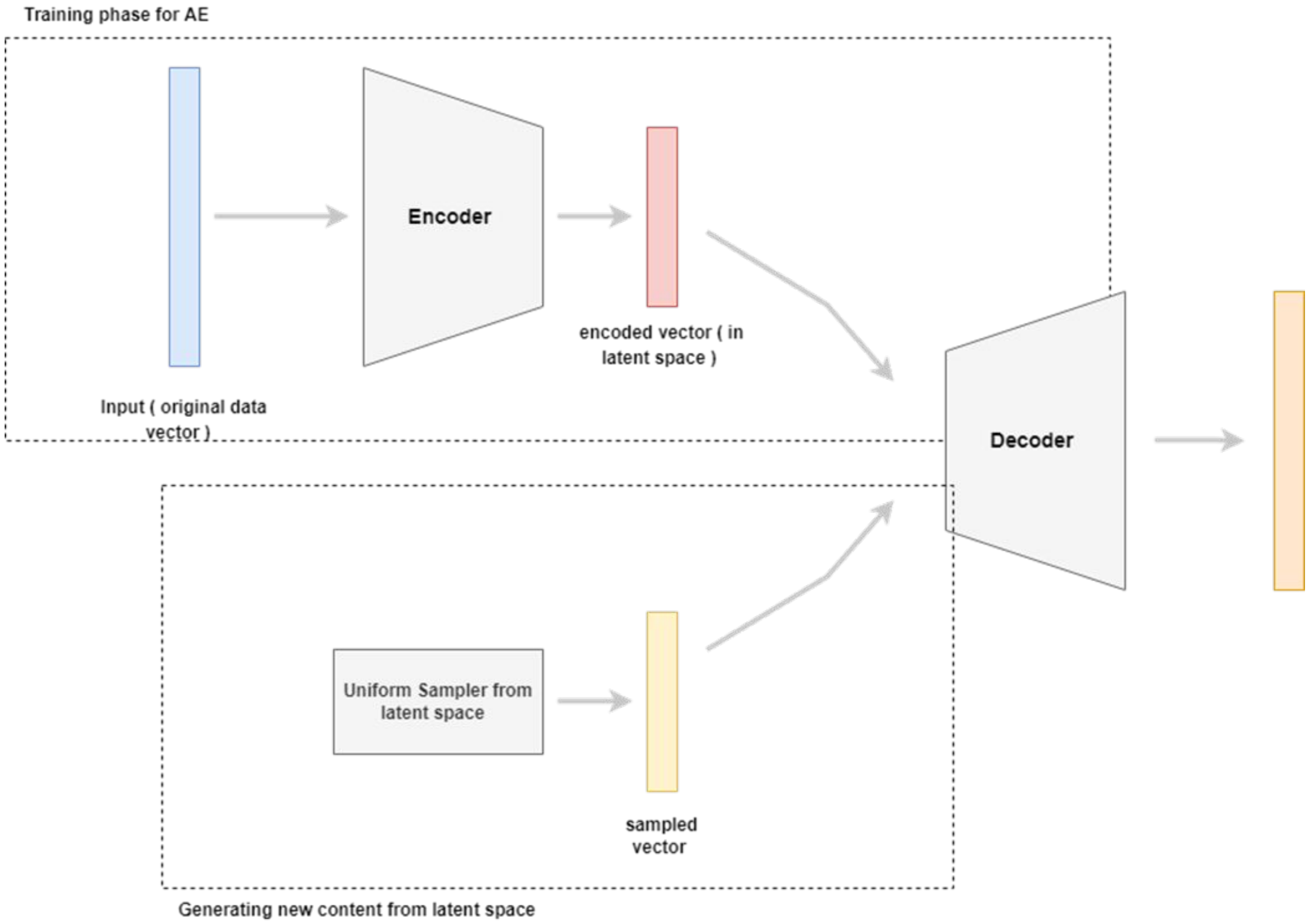


Figure : Is generating new data possible using AE? (authors).

1. Autoencoder (3/3)

No, the generated latent space is not regular; it is possible to sample a point that is meaningless to the decoder!

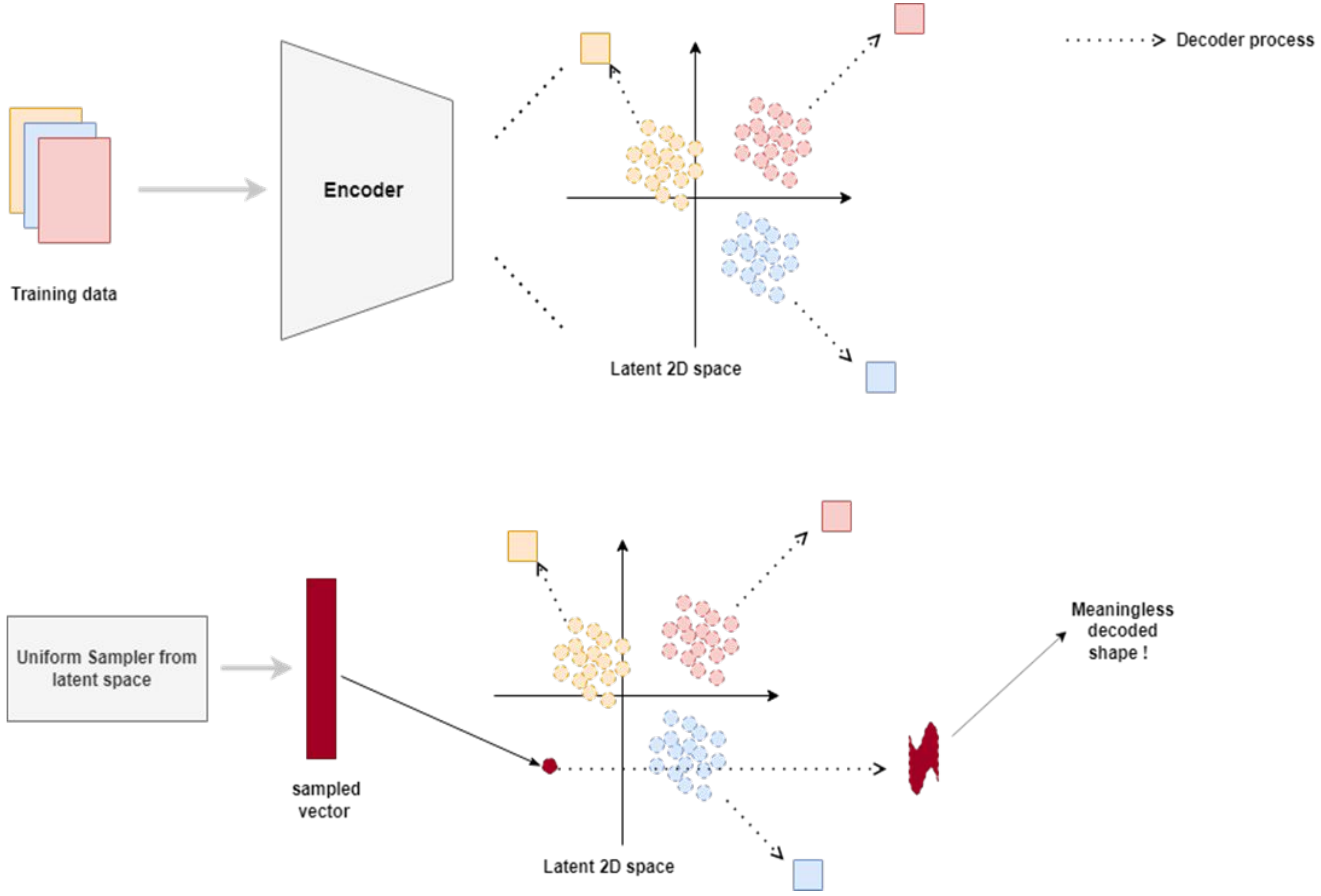


Figure : problem with AE’s latent space*.

*Inspired from illustrations in: Towards data science

2. Variational autoencoder

02

2. Variational Autoencoder (1/6)

Variational AutoEncoders (VAE) are **AutoEncoders** adapted to the task of content generation.

- Generates new data by decoding points that are randomly sampled from the **latent space**. The quality and relevance of generated data depend on the regularity of the latent space.
- Overcomes the problem of the latent space **irregularity** by making the encoder return parameters **of a distribution** over the latent space variables instead of a single point.
- Adds to the loss function a regularization term.

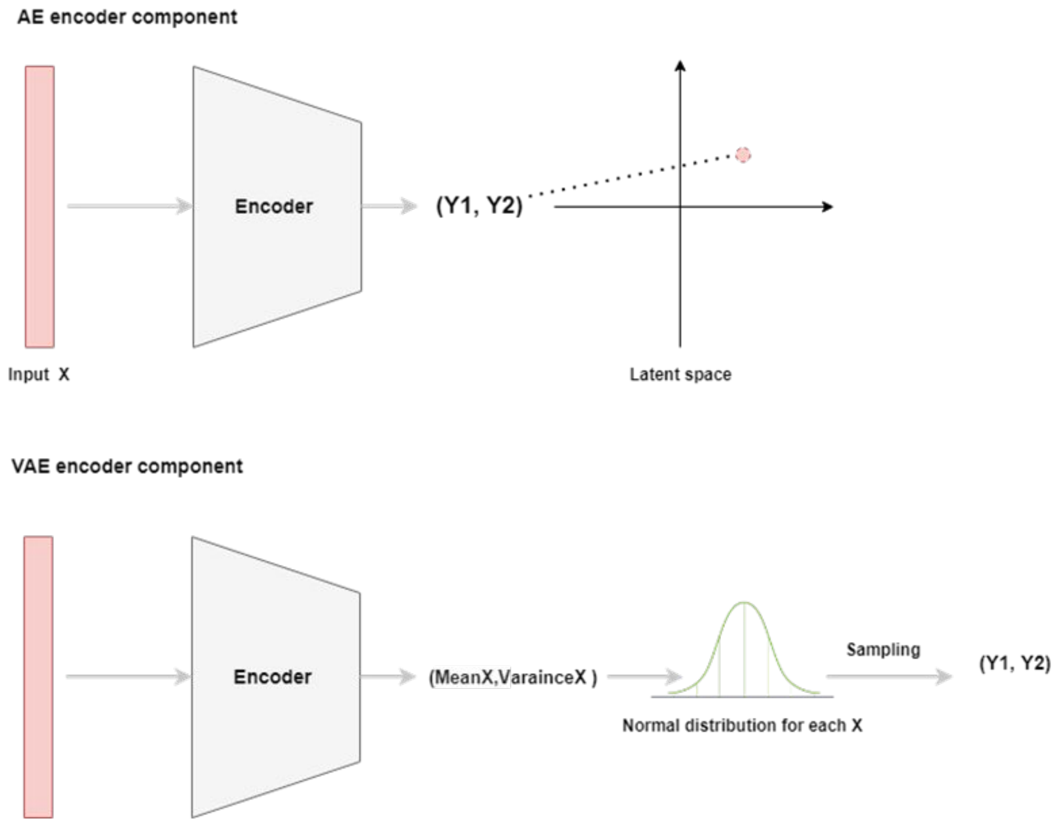


Figure : encoder from AE to VAE (authors).

*Lei et al, Oct 2019. Modeling tabular data using conditional GAN. https://arxiv.org/abs/1907.00503

2. Variational Autoencoder (2/6)

Variational AutoEncoders (VAE) are **AutoEncoders** adapted to the task of content generation.

- Generates new data by decoding points that are randomly sampled from the **latent space**. The quality and relevance of generated data depend on the regularity of the latent space.
- Overcomes the problem of the latent space **irregularity** by making the encoder return parameters **of a distribution** over the latent space variables instead of a single point.
- Adds to the loss function a regularization term.

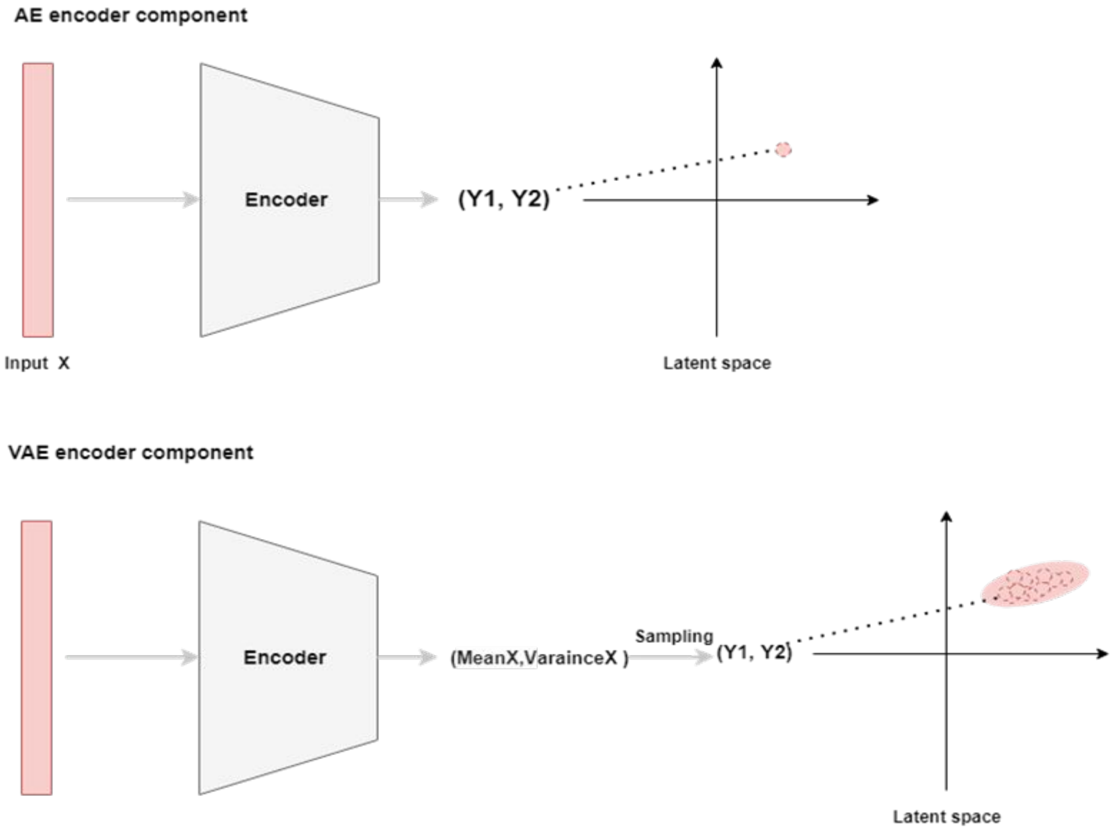
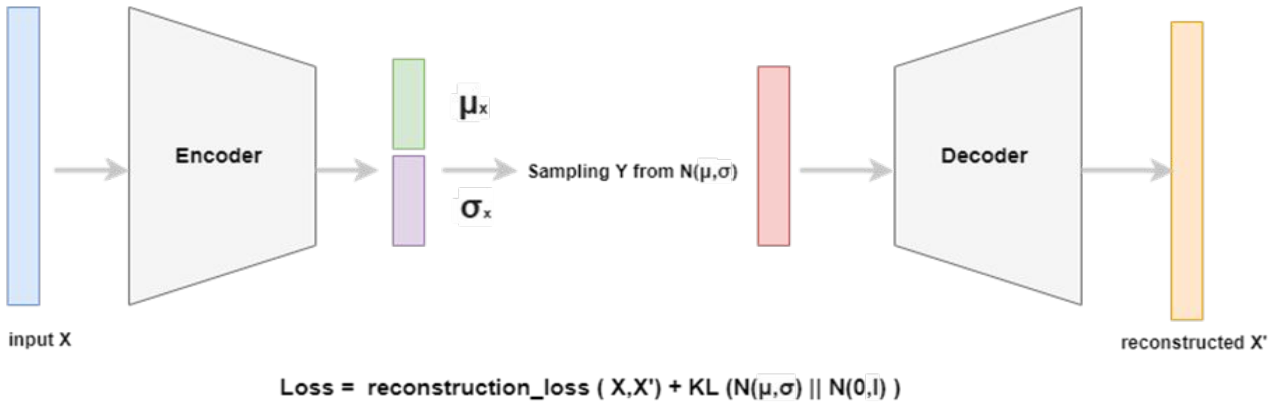


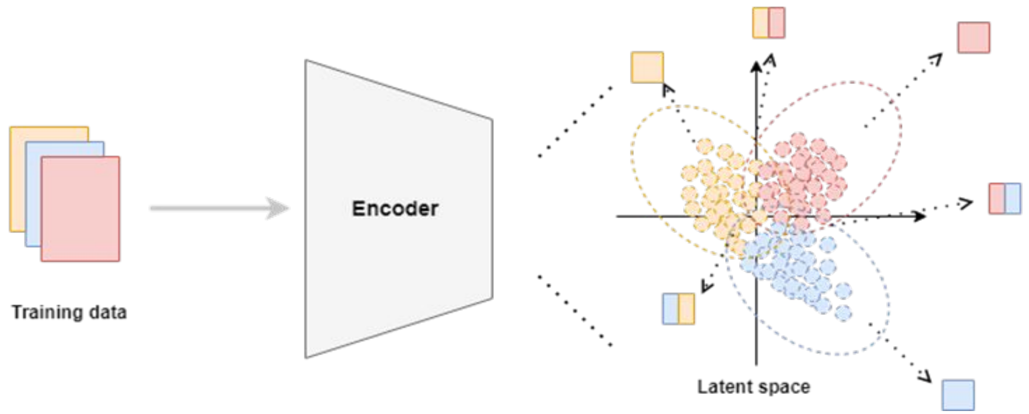
Figure : encoder from AE to VAE (authors).

*Lei et al, Oct 2019. Modeling tabular data using conditional GAN.
https://arxiv.org/abs/1907.00503

2. VARIATIONAL AUTOENCODER (3/6)

- VAE adds a regulation term : The KL-divergence distance between sample distribution and a normal N(0,I) distribution.
- The idea is to construct for each input X a distribution and then make all these distributions near to each other (since we try to make them near the N(0,I) distribution).
- Sampling is then done from the N(0,I) distribution.





2. RESULTS (4/6)

VAE with latent dimension = 2

VAE with latent dimension = 8

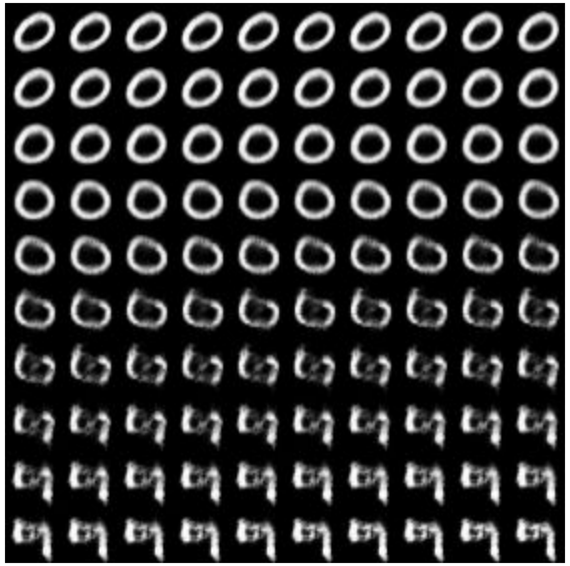
VAE with latent dimension = 32

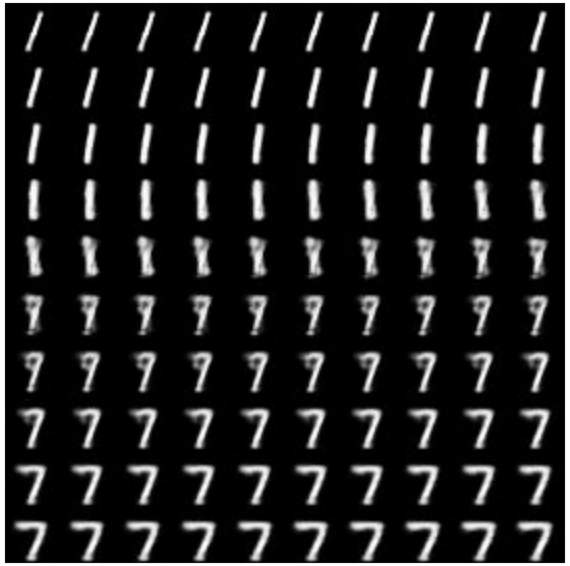
2. RESULTS (5/6)

Latent projection using t-SNE

2. VAE PROBLEM (6/6)

There is no variation in the generated images



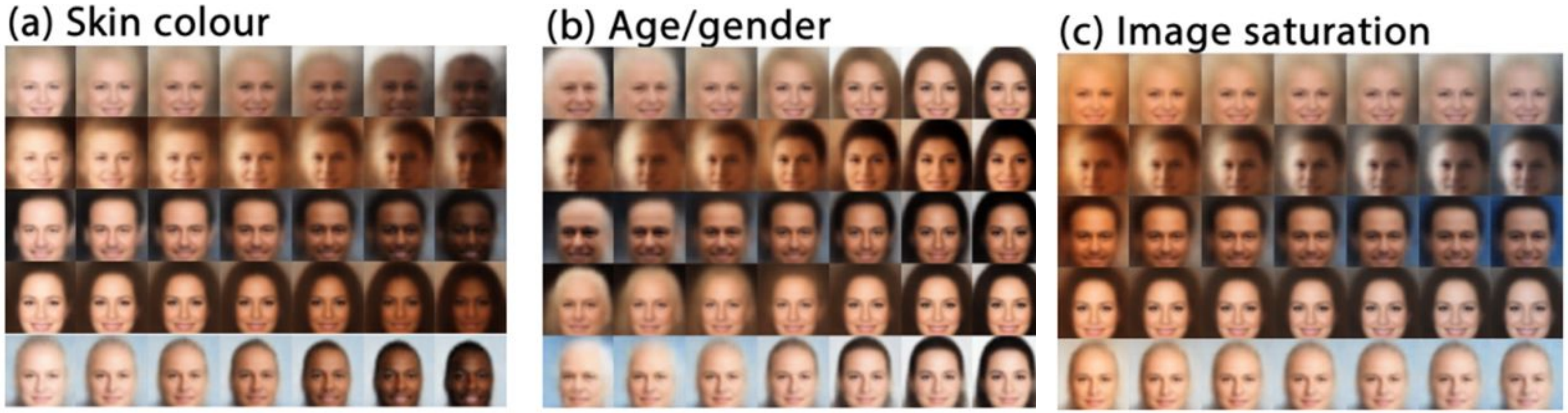


3. BETA-VARIATIONAL AUTOENCODERS

03

3.BETA-VARIATIONAL AUTOENCODERS (1/6)

- The latent space generated by the VAE is very entangled*.
- If we assume that the real images are created from some independent generative factors then we can learn a disentangled latent representation where each latent unit is sensitive to changes in a single generative factor while being relatively invariant to changes in other factors.



*Higgins et al, β-VAE: LEARNING BASIC VISUAL CONCEPTS WITH A CONSTRAINED VARIATIONAL FRAMEWORK

beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework | OpenReview

3.BETA-VARIATIONAL AUTOENCODERS (2/6)

- The beta VAE is able to learn a disentangled representation by introducing a beta factor in the total loss.

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

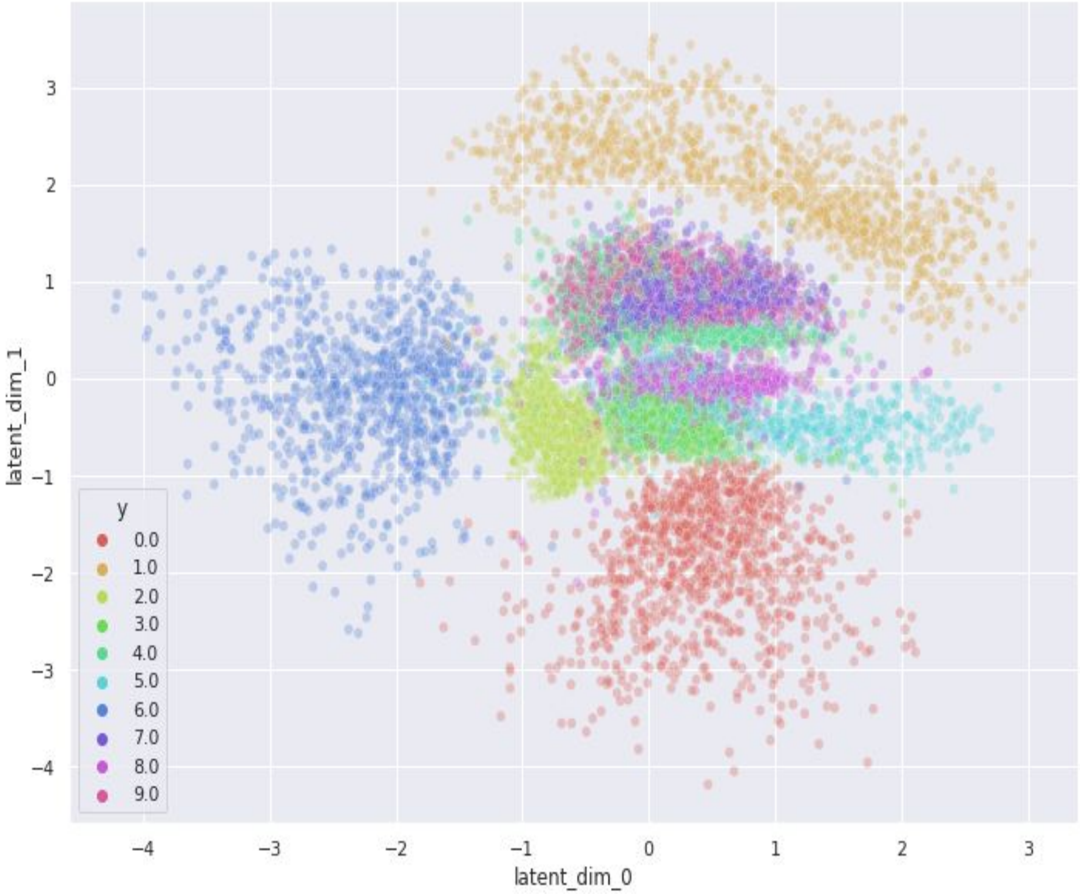
Loss for the VAE *

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

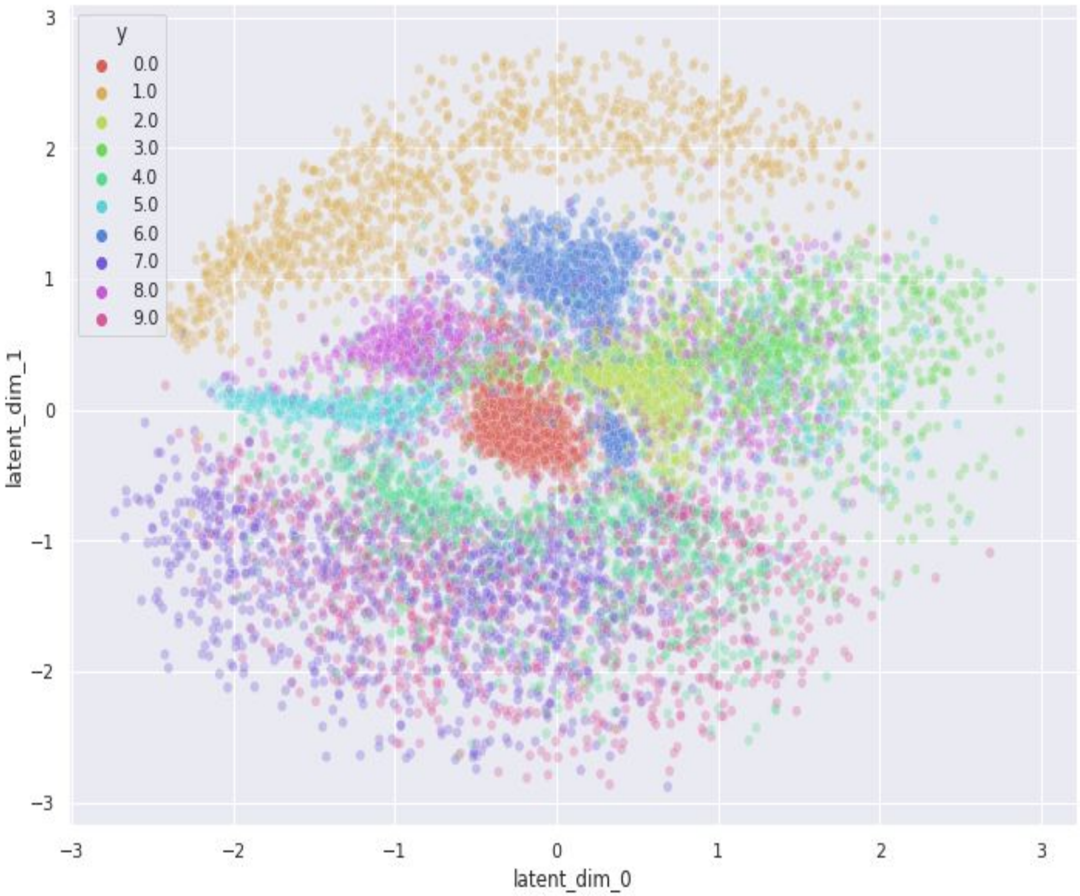
Loss for the Beta-VAE *

*Higgins et al, β-VAE: LEARNING BASIC VISUAL CONCEPTS WITH A CONSTRAINED VARIATIONAL FRAMEWORK
[beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework | OpenReview]

3.RESULTS (3/6)

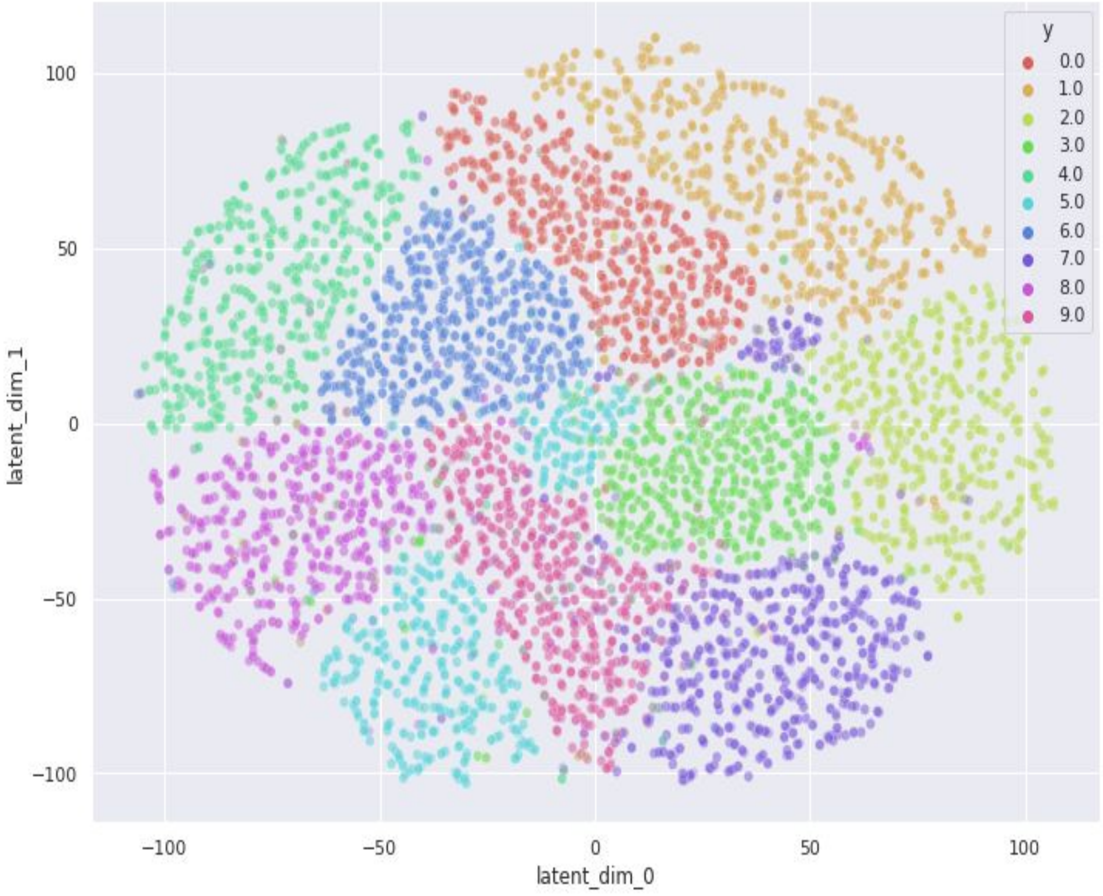


Latent projection using t-SNE
beta=2, latent dimension=2



Latent projection using t-SNE
beta=10, latent dimension=2

3.RESULTS (5/6)



Latent projection using t-SNE
beta=2, latent dimension=8

3.RESULTS (6/6)

beta-VAE with latent dimension = 2 , beta=10

beta-VAE with latent dimension = 8 , beta=2

4. Vector Quantized Variational Autoencoder (VQ-VAE)

04

4. Vector Qauntized-VAE (1/4)

The main idea behind **vector quantized*** VAE is to learn a dictionary-based discrete latent space where the output of the encoder is mapped to an embedding dictionary then passed to the decoder.

- The posterior distribution q(Z|X) is a one-hot distribution*.
- If we define a uniform categorical prior over Z, the KL-divergence term will be constant.
- Then, No need for KL divergence in the loss function.

Given that :

$$Z_e = Encoder(X)$$

The input of the decoder will be:

$$Z_q = e_i, \ where \ i = argmin_{j \epsilon [1..k]}(||Z_e - e_j||)$$

The posterior will become :

$$q(Z_q = e_i|X) = \begin{cases} 1 & \text{if } i = argmin(||Z_e - e_j||) \\ 0 & \text{otherwise} \end{cases}$$

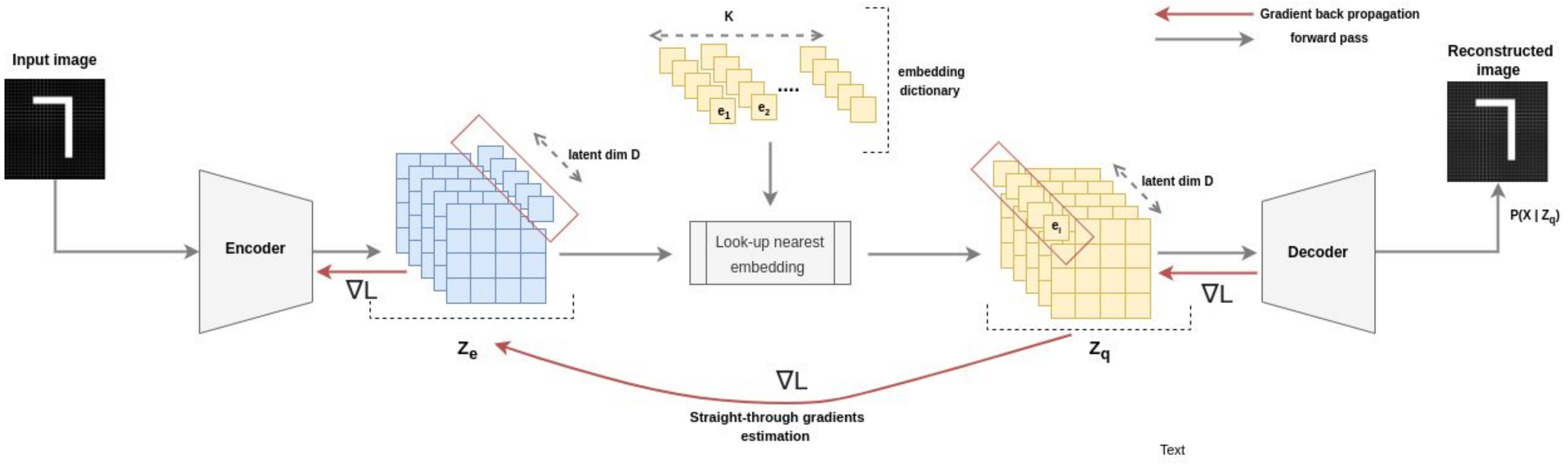
*van den Oord et al, May 2018. Neural Discrete Representation Learning https://arxiv.org/pdf/1711.00937.pdf

4.Vector Qauntized-VAE (2/4)

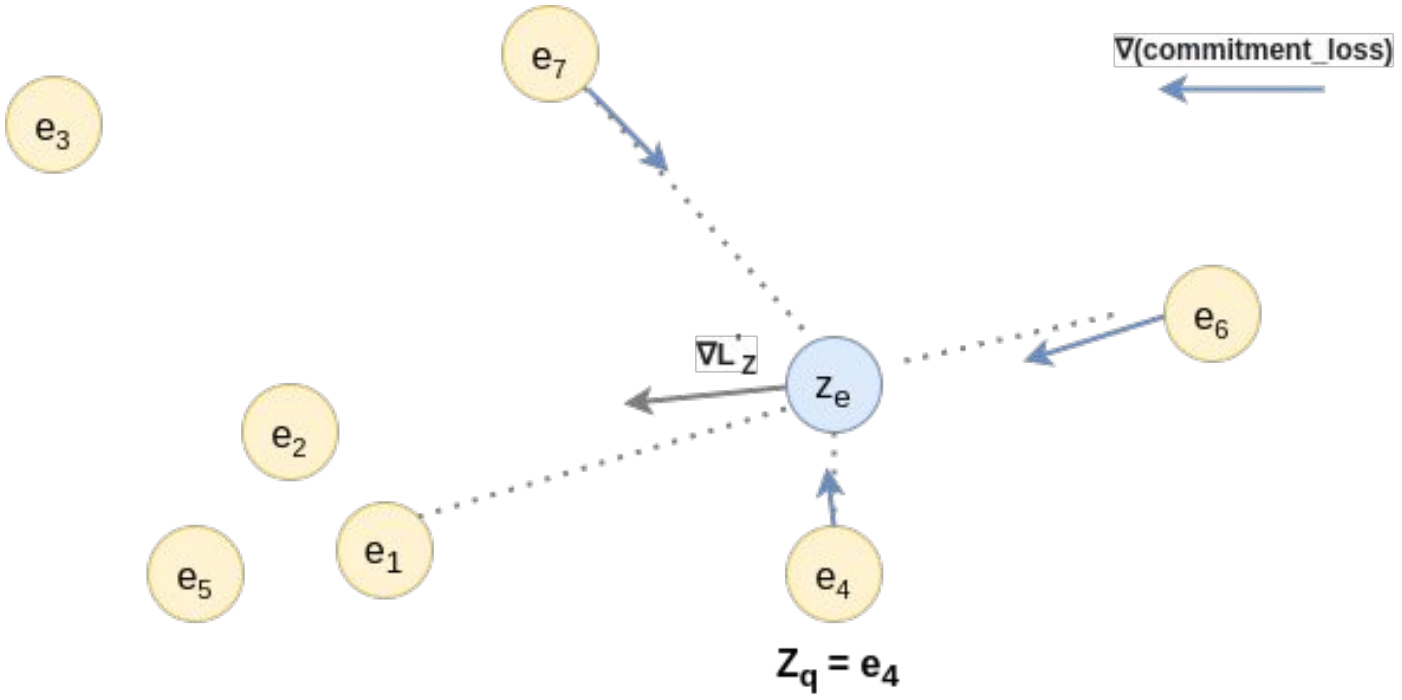
$$L = \underbrace{log[p(X|Z_q)]}_{\text{Reconstruction loss}} + \underbrace{\|sg(Z_q) - e\|_2^2 + \beta\|Z_q - sg(e)\|_2^2}_{\text{quantization loss}}$$

- **Reconstruction loss** : how reconstructed image is similar to the input.
- **Quantization loss =** embedding_loss + commitment loss.
- embedding loss pushes embeddings vectors towards encoder output, the commitment loss forces the encoder to commit to the embedding dictionary.

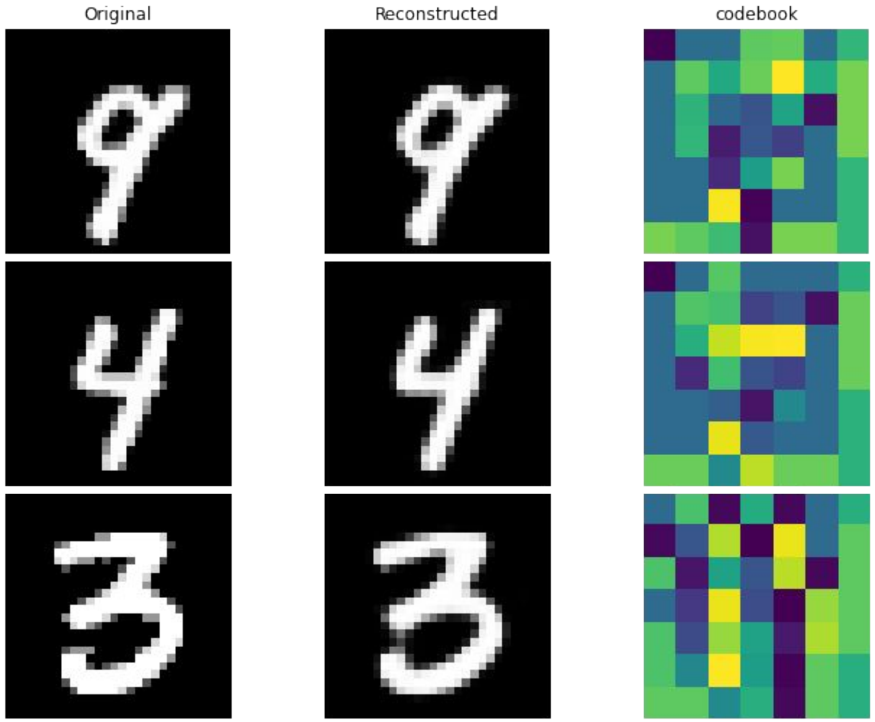
* sg means stop gradients, we don't compute gradients for what is inside.


Input image
Encoder
∇L
latent dim D
K
embedding dictionary
e1
e2
....
Look-up nearest embedding
Gradient back propagation
forward pass
latent dim D
ei
∇L
Decoder
Reconstructed image
P(X | Zq)
Ze
Zq
∇L
Straight-through gradients estimation
Text

4.Vector Qauntized-VAE (4/4)



3.RESULTS



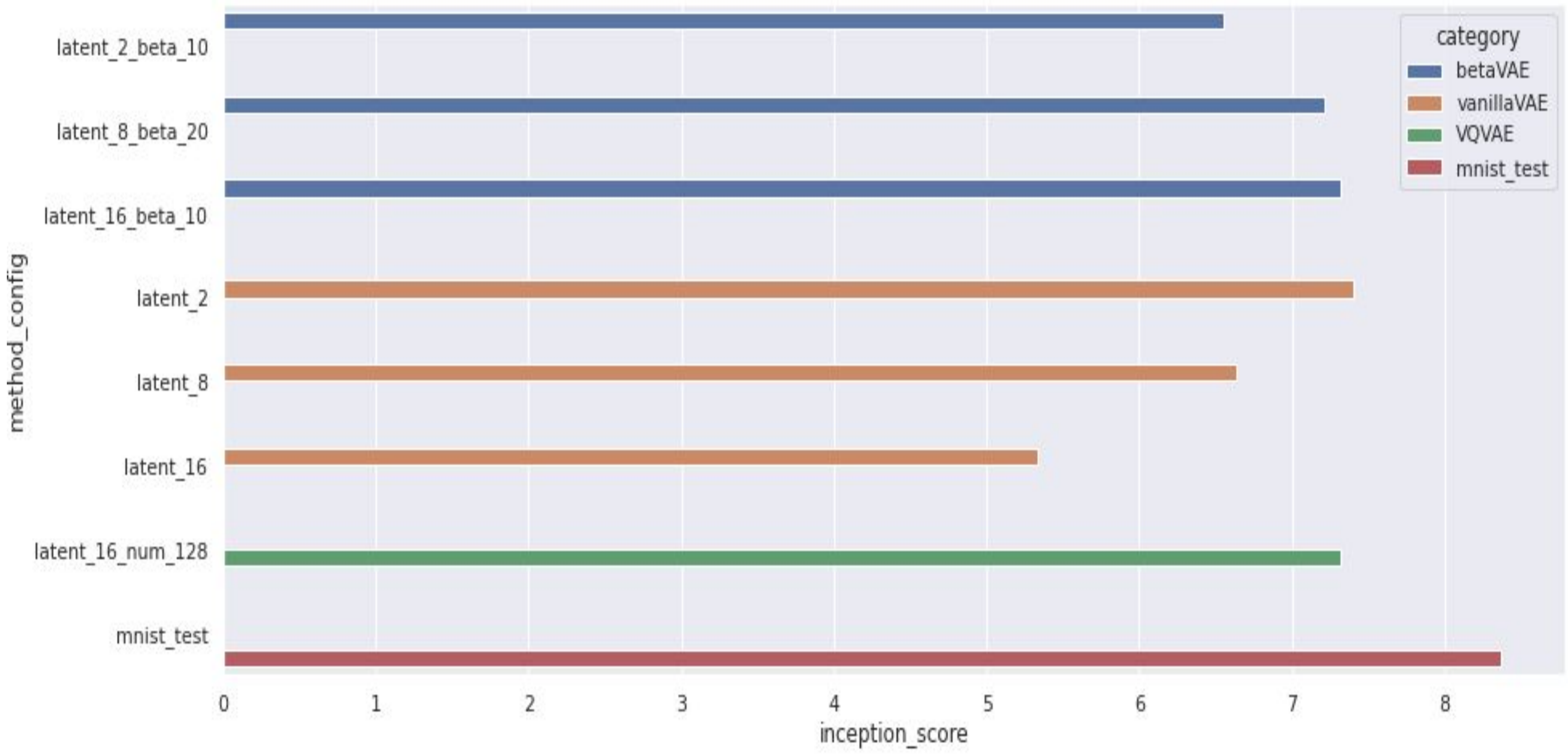
Comparison between original and reconstructed images with corresponding codebook.

Samples generated during training of VQ-VAE with 16 latent dimension

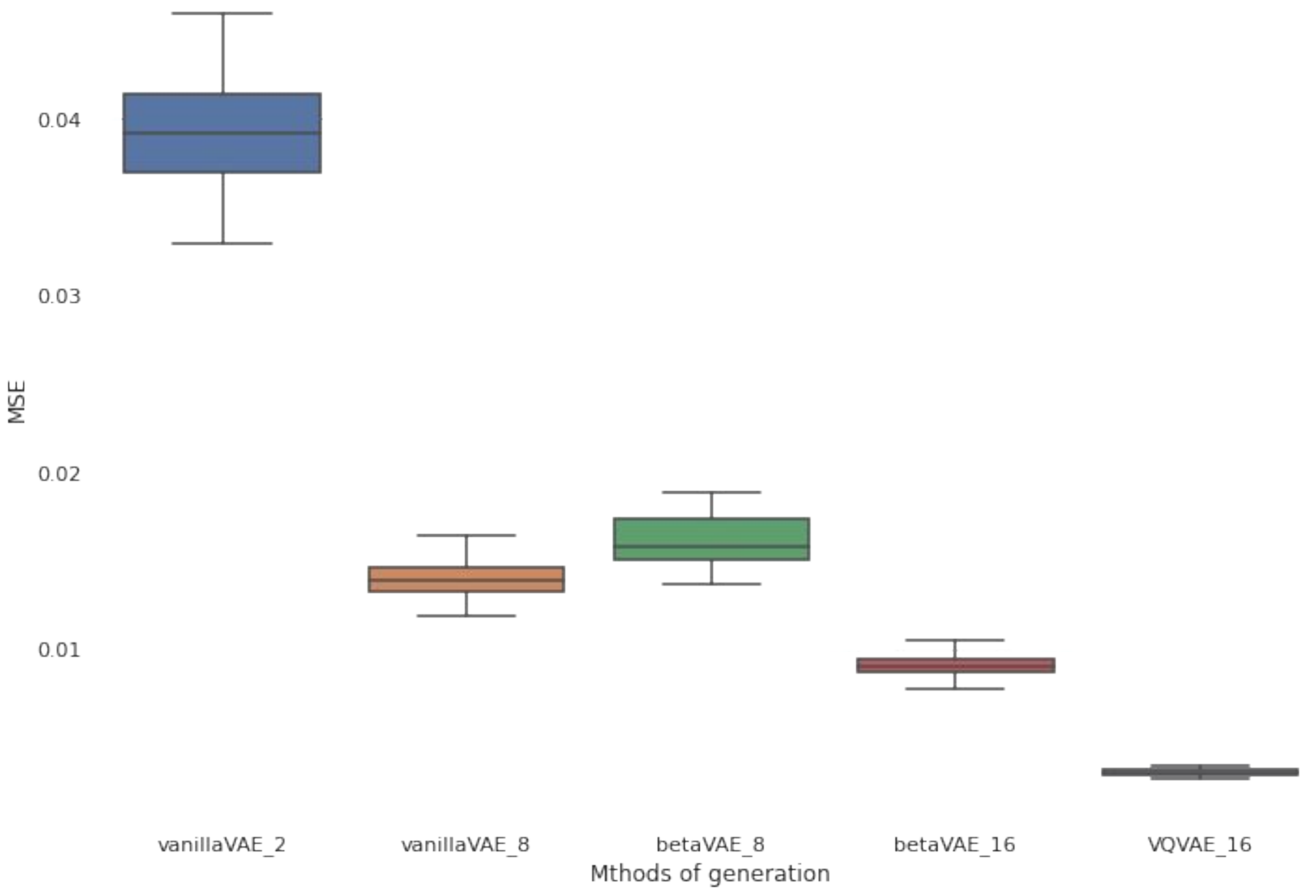
5. EVALUATION

05

5.EVALUATION INCEPTION SCORE (1/2)



5.EVALUATION RECONSTRUCTION LOSS (2/2)

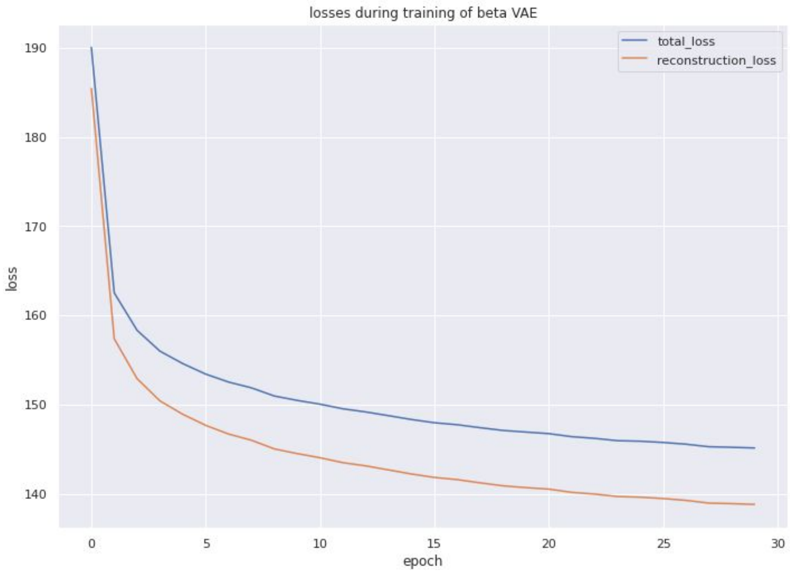


REFERENCES

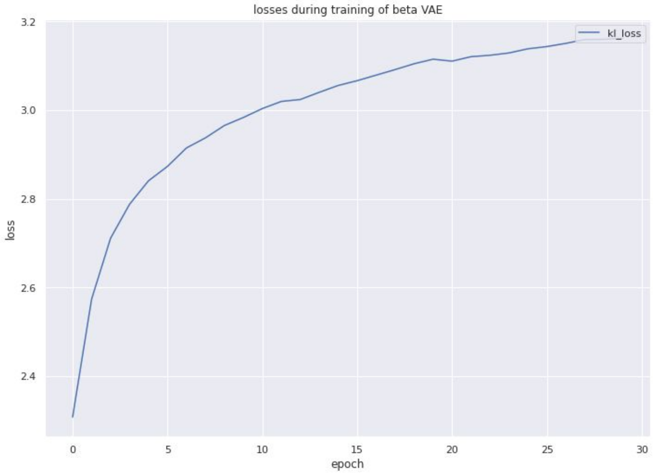
- Van den Oord et al, May 2018. Neural Discrete Representation Learning https://arxiv.org/pdf/1711.00937.pdf
- Higgins et al, β-VAE: LEARNING BASIC VISUAL CONCEPTS WITH A CONSTRAINED VARIATIONAL FRAMEWORK beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework | OpenReview
- Oord, Aaron V., Kalchbrenner, Nal, Vinyals, Oriol, Espeholt, Lasse, Graves, Alex, and Koray Kavukcuoglu. "Conditional Image Generation with PixelCNN Decoders." *arXiv*, (2016). https://doi.org/10.48550/arXiv.1606.05328.
- Kingma, Diederik P., and Max Welling. "Auto-Encoding Variational Bayes." *arXiv*, (2013). https://doi.org/10.48550/arXiv.1312.6114.

,

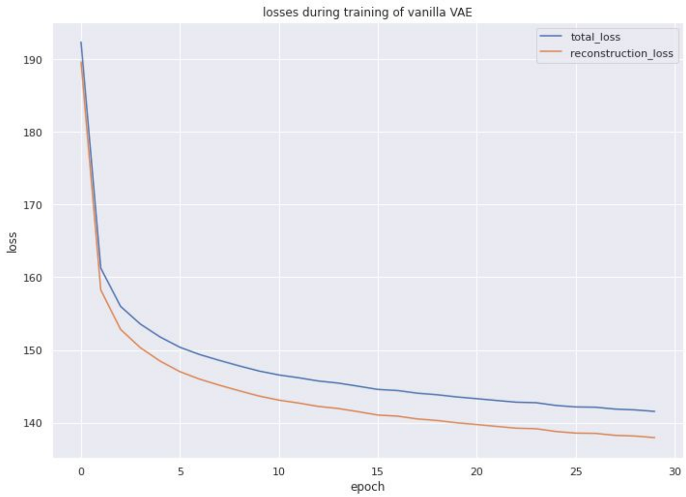
THANK YOU FOR YOUR ATTENTION !



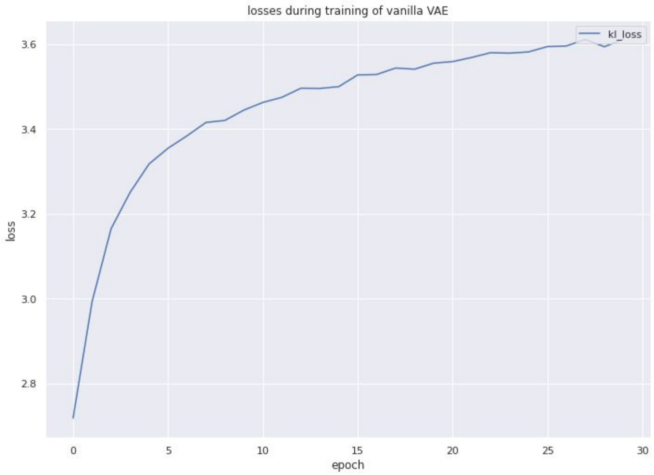
losses during training of beta VAE
total_loss
reconstruction_loss
loss
epoch



losses during training of beta VAE
kl_loss
loss
epoch



losses during training of vanilla VAE
total_loss
reconstruction_loss
loss
epoch



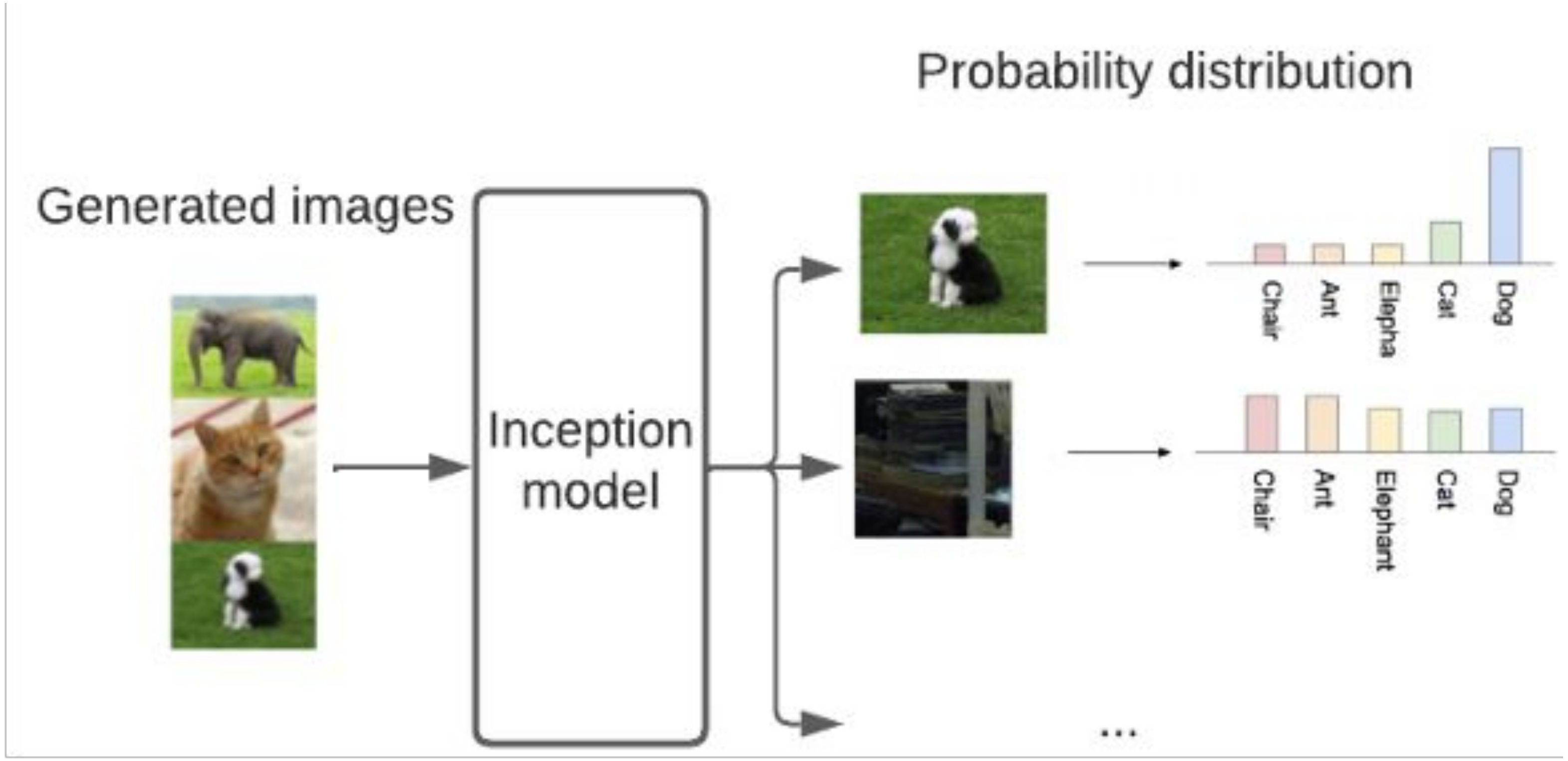
losses during training of vanilla VAE
kl_loss
loss
epoch

REFERENCES

- https://research.aimultiple.com/synthetic-data/ (a blog article about synthetic data from AI multiple)
- https://www.statice.ai/post/ai-data-agility-synthetic-data-insurance (statice blog article about synthetic data in insurance)
- https://www.gartner.com/en/information-technology/glossary/synthetic-data (Gartner glossary definition for synthetic data)
- https://www.nvidia.com/en-us/deep-learning-ai/resources/accelerating-ai-with-synthetic-data-ebook/ (NVidia eBook about synthetic data in AI)
- https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73 (medium TDS blog about VAE)
- https://towardsdatascience.com/art-of-generative-adversarial-networks-gan-62e96a21bc35 (medium TDS blog about GANs and implementation)
- https://github.com/Baukebrenninkmeijer/table-evaluator (synthetic data evaluation library on GitHub)
- https://blogs.nvidia.com/blog/2021/06/08/what-is-synthetic-data/ (the future of IA and synthetic data, nvidia blog article)
- tps://towardsdatascience.com/synthetic-data-vault-sdv-a-python-library-for-dataset-modeling-b48c406e7398 (blog article from medium about synthetic data and SDV)
- https://towardsdatascience.com/understanding-generative-adversarial-networks-gans-cd6e4651a29 (Understanding Generative adversarial nets)
- https://github.com/Team-TUD/CTAB-GAN (CTAB-GAN GitHub repository)
- https://medium.com/octavian-ai/a-simple-explanation-of-the-inception-score-372dff6a8c7a

Images variety

INCEPTION SCORE
(IS)*

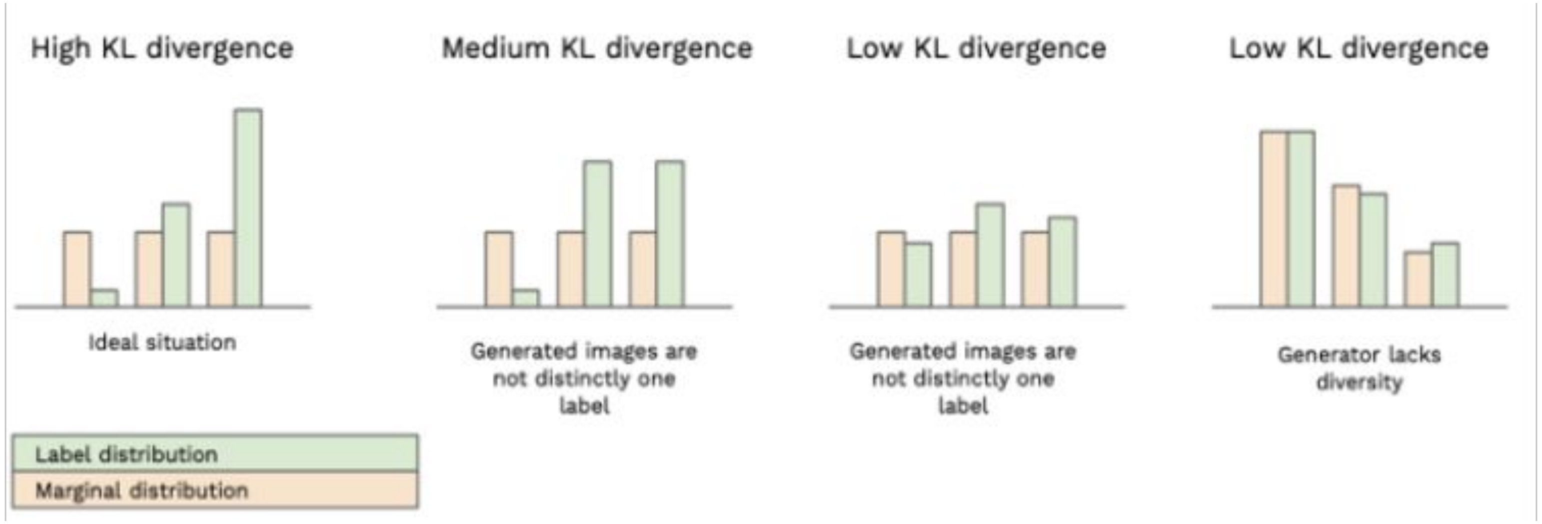


Each image distinctly looks like something

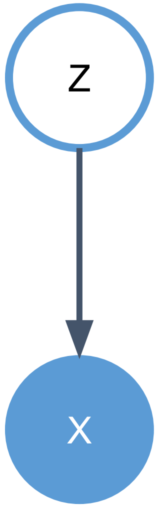
*[A simple explanation of the Inception Score | by David Mack | Octavian | Medium]

5.METRICS (2/2)

- We calculate this score using a statistics formula called the Kullback-Leibler (KL) divergence, which is a measure of how similar/different two probability distributions are.



2. VARIATIONAL AUTOENCODER (4/7)



Latent variables

Observed variables

The marginal log-likelihood *

$$\max_{p\in\mathcal{P}_{\mathbf{x},\mathbf{z}}} \sum_{\mathbf{x}\in D} \log p(\mathbf{x}) = \sum_{\mathbf{x}\in D} \log \int p(\mathbf{x},\mathbf{z})\, \mathrm{d}\mathbf{z}.$$

intractable for high-dimensional

The Evidence Lower Bound *

$$\mathrm{ELBO}(\mathbf{x};\theta,\phi) = \mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}\left[\log \frac{p_\theta(\mathbf{x},\mathbf{z})}{q_\phi(\mathbf{z}\mid\mathbf{x})}\right]$$

*REf

3.BETA-VARIATIONAL AUTOENCODERS (2/6)

- The beta VAE is able to learn a disentangled representation by introducing a beta factor in the total loss.
- These constraint impose a limit on the capacity of the latent information channel and control the emphasis on learning statistically independent latent factors. With $\beta > 1$ the model is pushed to learn a more efficient latent representation of the data, which is disentangled if the data contains at least some underlying factors of variation that are independent.

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

Loss for the VAE *

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

Loss for the Beta-VAE *

*Higgins et al, β-VAE: LEARNING BASIC VISUAL CONCEPTS WITH A CONSTRAINED VARIATIONAL FRAMEWORK
beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework | OpenReview