Generative Models

Variational Autoencoders-VAE

Marcos C. S. Santana, Prof. João Paulo Papa

December 18, 2019

Universidade Estadual Paulista "Júlio de Mesquita Filho" (UNESP)
Faculdade de Ciências (FC) / Departamento de Computação (DCo)
Bauru, SP - Brasil

Outline

- VAE - Variational Autoencoder
    - Variational Inference
    - Evidence Lower Bound (ELBO)
    - Variational Autoencoder
    - Variations
    - Applications

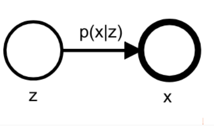
The Problem

How can we perform efficient inference and learning in **directed probabilistic models**, in the presence of **continuous latent variables** with **intractable posterior distributions**, and large datasets? Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.

[Kingma and Welling, 2013]

The Problem

Given a direct graph that models the data **x** related to a continuous latent variable **z**.



The posterior porbability $p(x|z)$ can be defined by Bayesian Statistics:

$$p(z|x) = \frac{p(x,z)}{p(x)} = \frac{p(x|z)p(z)}{p(x)} \tag{1}$$

The Problem

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)} \tag{2}$$

- $p(x|z)$ is the certainty of the observation x, if we are certain about the hidden state z.
- $p(z)$ is the prior probability that can be chosen as a tractable. It corresponds to the beforehand knowledge.
- $p(x)$ is unknown and often untractable. It is a normalizing factor (partition function).

$$p(x) = \int p(x|z)p(z)dz \tag{3}$$

- Not important for some cases, but *essential* for generative models.

RECOGNA LABORATORY

Possible options...

1. Use Monte Carlo-related methods to integrate $p(x) = \int \int \dots \int \dots$.
    - Samples unnormalized.
    - Unbiased.
    - Needs a lot of samples.
    - It is SLOW.
2. **Use Variational Inference Methods**.
3. - Approximates $p(z|x)$.
    - Biased.
    - Fast and scalable.

RECOGNA
LABORATORY

Variational Inference

The method consists to find an approximation of a tractable $q(z)$ distribution that is as close as possible of the untractable $p(z|X)$. But we need:

1. To convert the inference to an optimization problem.
    - Making $q(z)$ parametrized as $q(z;\theta)$
2. To define a criterion for the optimization.
    - Find out a way to measure how close the parametric distribution $q(z;\theta)$ is to $p(z|x)$.

RECOGNA
LABORATORY

Variational Inference - Information difference

Given two probability distribution $q(z)$ and $p(z)$, the information of the distributions is:

$$I_p(z) = -\log p(z) \tag{4}$$

$$I_q(z) = -\log q(z) \tag{5}$$

The information difference between $p(z)$ and $q(z)$ is:

$$\Delta I(z) = -\log p(z) + \log q(z) = \log \frac{q(z)}{p(z)} \tag{6}$$

Taking the expectation w.r.t q(z):

$$\mathbb{E}_{z \sim q(z)}\left[\Delta I(z)\right] = \int q(z) \log \frac{q(z)}{p(z)} dz \tag{7}$$

That is the definition of Kullback–Leibler divergence:

$$KL\big(q(z)||p(z)\big) = \int q(z) \log \frac{q(z)}{p(z)} dz \tag{8}$$

Variational Inference - KL Properties

- If $q(z) = p(z)$ then :

$$KL(p(z)||p(z)) = -\int p(z) \log \frac{p(z)}{p(z)} dz = 0 \tag{9}$$

- DL is not symmetric unless $q(z) = p(z)$.

$$KL(q(z)||p(z)) = \int q(z) \log \frac{q(z)}{p(z)} dz \tag{10}$$

$$KL(p(z)||q(z)) = -\int p(z) \log \frac{q(z)}{p(z)} dz \tag{11}$$

$$KL(p(z)||q(z)) \neq KL(q(z)||p(z)) \tag{12}$$

- The lack of symmetry of KL makes it a divergence and not a metric (distance).

RECOGNA LABORATORY

Variational Inference - KL Properties

- $KL\big(p(z)||p(z)\big) \geq 0$: We know that:

$$\log t \leq t - 1 \tag{13}$$

Thus:

$$-KL\big(p||q\big) = \int q(z) \log \frac{p(z)}{q(z)} dz \leq \int q(z) \left(\frac{p(z)}{q(z)} - 1 \right) dz \tag{14}$$

$$\int q(z) \left(\frac{p(z)}{q(z)} - 1 \right) dz = \int p(z) dz - \int q(z) dz = 1 - 1 = 0 \tag{15}$$

So:

$$-KL\big(p||q\big) \leq 0 \tag{16}$$

Finally:

$$KL\big(p||q\big) \geq 0 \tag{17}$$

RECOGNA LABORATORY

Variational Inference as Optimization Problem

The main idea is to find $q(z;\theta) \sim p(z|X)$:

$$q(z;\theta) = \min_{q(z;\theta)} KL\Big(q(z;\theta)||p(z|X)\Big) \tag{18}$$

But we have two problems:

1. How to compute posterior?
2. How to optimize a distribution?

RECOGNA LABORATORY

Variational Inference: Mathemagics ...

Let's make some transformations on $\log p(x)$

- Multiply $\log p(x)$ by 1:

$$\log p(x) = 1 \times \log p(x) = \log p(x) \int q(z)dz = \int q(z) \log p(x) dz \tag{19}$$

- Substitute by Bayes relation :

$$p(z|x) = \frac{p(x,z)}{p(x)} \rightarrow p(x) = \frac{p(x,z)}{p(z|x)} \tag{20}$$

$$\log p(x) = \int q(z) \log \frac{p(x,z)}{p(z|x)} dz \tag{21}$$

- Multiply by 1:

$$\log p(x) = \int q(z) \log \left(\frac{p(x,z)}{p(z|x)} \frac{q(z)}{q(z)} \right) dz \tag{22}$$

RECOGNA
LABORATORY

Variational Inference: Mathemagics ...

- Split into two parts:

$$\log p(x) = \int q(z) \log \left(\frac{p(x,z)}{q(z)} \right) dz - \int q(z) \log \frac{q(z)}{p(z|x)} \tag{23}$$

- Given that:

$$KL\Big(q(z;\theta) || p(z|X) \Big) = \int q(z) \log \frac{q(z)}{p(z|x)} \tag{24}$$

We have:

$$\log p(x) = \mathbb{E}_{z \sim q(z)} \left[\log \left(\frac{p(x,z)}{q(z)} \right) \right] - KL\Big(q(z) || p(z|X) \Big) \tag{25}$$

$$\mathcal{L}(q(z)) = \mathbb{E}_{z \sim q(z)} \left[\log \left(\frac{p(x,z)}{q(z)} \right) \right] \tag{26}$$

$$\log p(x) = \mathcal{L}(q(z)) - KL\Big(q(z) || p(z|X) \Big) \tag{27}$$

- What is $\mathcal{L}(q(z))$?

RECOGNA
LABORATORY

Variational Inference: Mathemagics ...

- Split into two parts:

$$\log p(x) = \int q(z) \log\left(\frac{p(x,z)}{q(z)}\right) dz - \int q(z) \log \frac{q(z)}{p(z|x)} \tag{28}$$

- Given that:

$$KL\Big(q(z;\theta)||p(z|X)\Big) = \int q(z) \log \frac{q(z)}{p(z|x)} \tag{29}$$

We have:

$$\log p(x) = \mathbb{E}_{z\sim q(z)}\left[\log\left(\frac{p(x,z)}{q(z)}\right)\right] - KL\Big(q(z)||p(z|X)\Big) \tag{30}$$

$$\mathcal{L}(q(z)) = \mathbb{E}_{z\sim q(z)}\left[\log\left(\frac{p(x,z)}{q(z)}\right)\right] \tag{31}$$

$$\log p(x) = \mathcal{L}(q(z)) - KL\Big(q(z)||p(z|X)\Big) \tag{32}$$

- What is $\mathcal{L}(q(z))$?

RECOGNA LABORATORY

Variational Inference: Evidence Lower Bound

- $\log p(X)$ is independent of z and $KL\Big(q(z)||p(z|X)\Big) \geq 0$:

$$\log p(x) \geq \mathcal{L}(q(z)) \tag{33}$$

- $\mathcal{L}(q(z))$ is the lower bound of $\log p(x)$.
- From bayesian nomeclature:

$$posterior = \frac{likelihood \times prior}{evidence} \tag{34}$$

- $\mathcal{L}(q(z))$ is known as Evidence Lower Bound - ELBO [Blei et al., 2017].

RECOGNA LABORATORY

Variational Inference: Evidence Lower Bound

- Maximizing $\mathcal{L}(q(z))$ is equivlent to minimize $KL\Big(q(z)||p(z|X)\Big)$:

$$L = \max_{q(z;\theta)} \mathcal{L}(q(z;\theta)) = \min_{q(z;\theta)} KL\Big(q(z;\theta)||p(z|X)\Big) \tag{35}$$

- The $\mathcal{L}(q(z;\theta))$ to depend on z using the equations:

$$\mathcal{L}(q(z)) = \mathbb{E}_{z\sim q(z)}\left[\log\left(\frac{p(x,z)}{q(z)}\right)\right] = \int q(z)\log\frac{p(x,z)}{q(z)}dz \tag{36}$$

$$\frac{p(x,z)}{q(z)} = \frac{p(x|z)p(z)}{q(z)} \tag{37}$$

RECOGNA LABORATORY

Variational Inference: Evidence Lower Bound

- Manipulating:

$$\mathcal{L}(q(z)) = \int q(z) \log \frac{p(x|z)p(z)}{q(z)} dz \tag{38}$$

$$\mathcal{L}(q(z)) = \int q(z) \log p(x|z) dz + \int q(z) \log \frac{p(z)}{q(z)} dz \tag{39}$$

$$\mathcal{L}(q(z)) = \mathbb{E}_{z \sim q(z)} \Big[\log p(x|z) \Big] - KL\Big(q(z) || p(z) \Big) \tag{40}$$

- The first term is related to data reconstruction.
- The second term plays a regularizing role.
- $p(x|z)$ is the data generated from hidden variable. We can choose a tractable form.
- $p(z)$ and $q(z)$ are prior distributions. They can also be chosen to be tractable.

RECOGNA LABORATORY

Variational Autoencoder - VAE

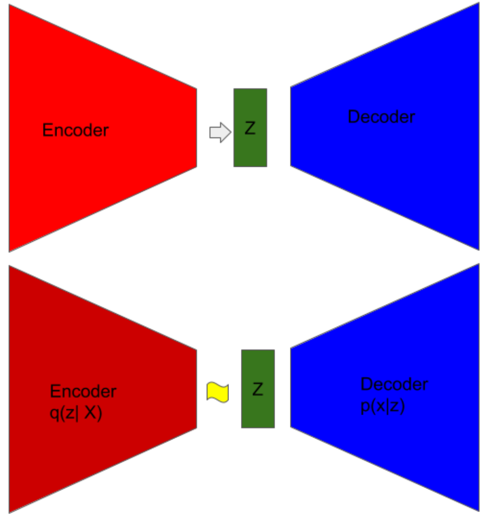


Figura 1: Top: Autoencoder. Bottom: Variational Autoencoder.

RECOGNA
LABORATORY

Variational Autoencoder - Reparametrization trick

- The neural networks are trained using backpropagation algorithm.
- How train a network where the sample layer is not differentiable?
- Reparametrization trick for Gaussian distribution:

$$\mathcal{N}(\mu, \sigma I) = \mu + \sigma \times \mathcal{N}(0, I) \tag{41}$$

- Now, the layer is differentiable. Thus we can use backpropagation to adjust the parameters μ and σ.

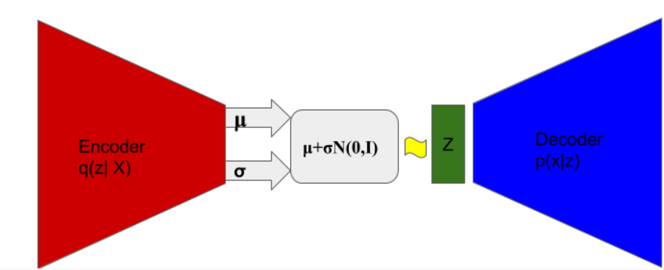


Figura 2: Variational Autoencoder [Kingma and Welling, 2013].

Variational Autoencoder - Loss Function

- From ELBO:

$$\mathcal{L}(q(z)) = \mathbb{E}_{z\sim q(z)}\Big[\log p(x|z)\Big] - KL\Big(q(z;\theta)||p(z)\Big) \tag{42}$$

1. We choose p(x) such as:

$$p(x) \rightarrow \frac{1}{\sqrt{2\pi\sigma_p^2}} \exp\left(-\frac{(x-\mu_p)}{2\sigma_p^2}\right) \tag{43}$$

2. We choose $q(z|x)$ such as:

$$q(z|x) \rightarrow \frac{1}{\sqrt{2\pi\sigma_q^2}} \exp\left(-\frac{(z-\mu_q)}{2\sigma_q^2}\right) \tag{44}$$

[Doersch, 2016]

RECOGNA LABORATORY

Variational Autoencoder - Loss Function

- The close form of the regularization (KL) for Gaussian:

$$KL\left(q(z;\theta)||p(z)\right) = \log\frac{\sigma_q}{\sigma_p} - \frac{\sigma_q^2 + (\mu_q - \mu_p)^2}{2\sigma_q^2} + \frac{1}{2} \quad (45)$$

- We choose $\sigma_p = 1.0$ and $\mu_p = 0$:

$$KL\left(q(z;\theta)||p(z)\right) = \frac{1}{2}\left[1 + \log\sigma_q - \sigma_q^2 - \mu_q^2\right] \quad (46)$$

- The loss function becomes:

$$\mathcal{L}(q(z)) = \mathbb{E}_{z\sim q(z)}\left[\log p(x|z)\right] - \frac{1}{2}\left[1 + \log\sigma_q - \sigma_q^2 - \mu_q^2\right] \quad (47)$$

RECOGNA
LABORATORY

Variational Autoencoder - Loss Function

- If $p(x|z) \approx \exp(x - \mu_p)^2$ the term is related to image reconstruction as found in AE:

$$\log p(x|z) \approx (x - \mu_p)^2 \tag{48}$$

RECOGNA LABORATORY

Variational Autoencoder - Training

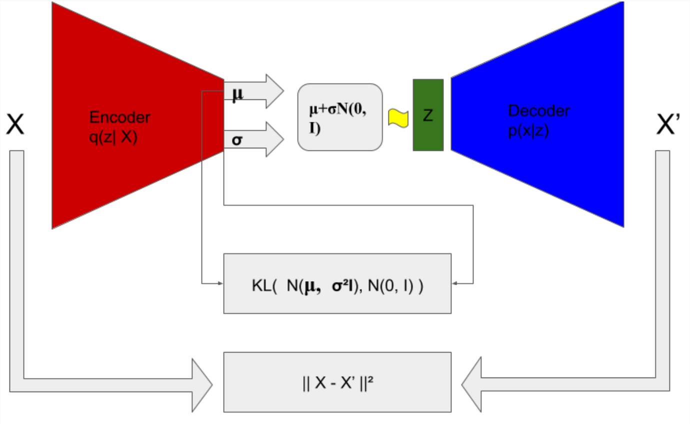


Figura 3: VAE - Training process.

RECOGNA
LABORATORY

Variational Autoencoder - Generating

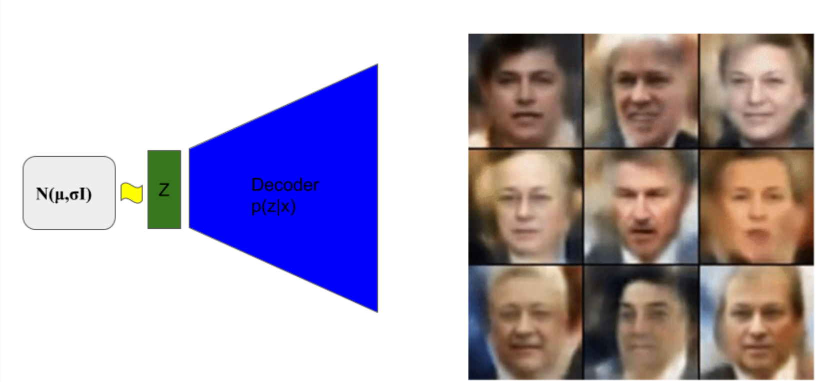


Figura 4: VAE - Generating process [Kingma and Welling, 2013].

Conditional Variational Autoencoder

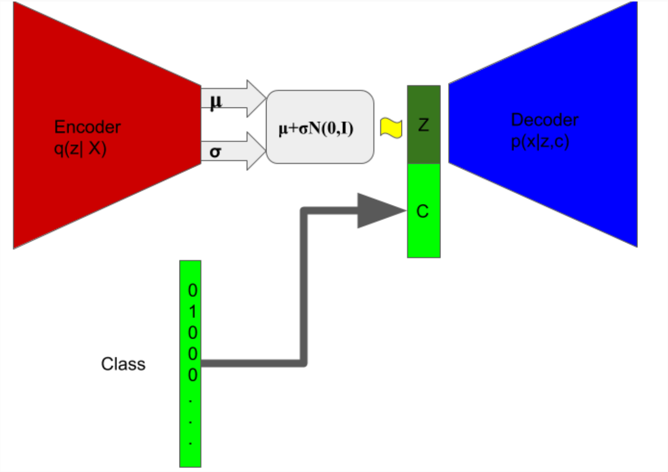


Figura 5: CVAE: [Sohn et al., 2015]

RECOGNA
LABORATORY

Conditional Variational Autoencoder - Generating

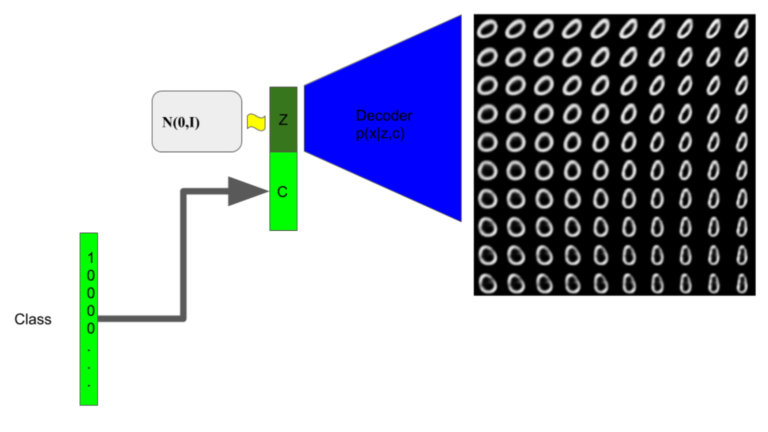


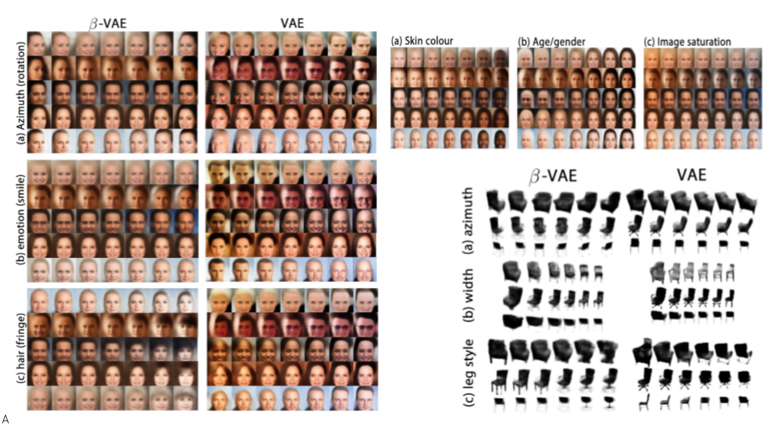
Figura 6: CVAE- Generating digits from classes.

RECOGNA
LABORATORY

β - Variational Autoencoder

Unbalance the loss function terms $\beta > 1$ to promote feature desentanglement [Higgins et al., 2017].

$$\mathcal{L}(q(z)) = \mathbb{E}_{z \sim q(z)}\Big[\log p(x|z)\Big] - \beta \times KL\Big(q(z;\theta)||p(z)\Big) \tag{49}$$



RECOGNA LABORATORY

Vector Quantized Variational Autoencoders (VQ-VAE)

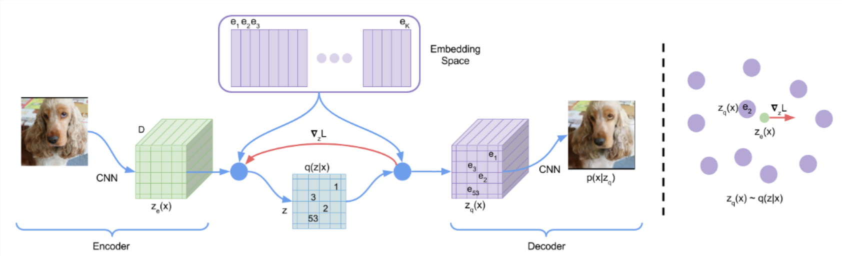


Figura 8: A figure describing the VQ-VAE. [van den Oord et al., 2017]

Vector Quantized Variational Autoencoders (VQ-VAE)

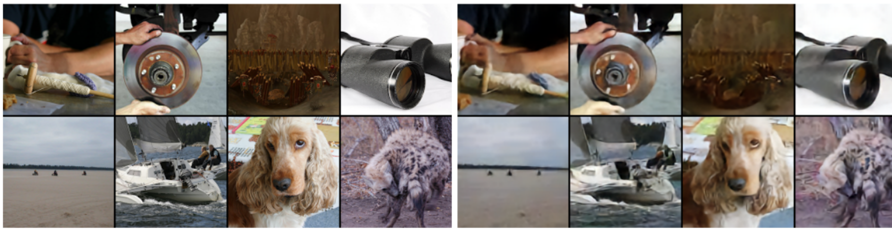


Figura 9: Left: Input images. Right: Reproductions.
[van den Oord et al., 2017]

Variational Autoencoder -Manifold Exploration

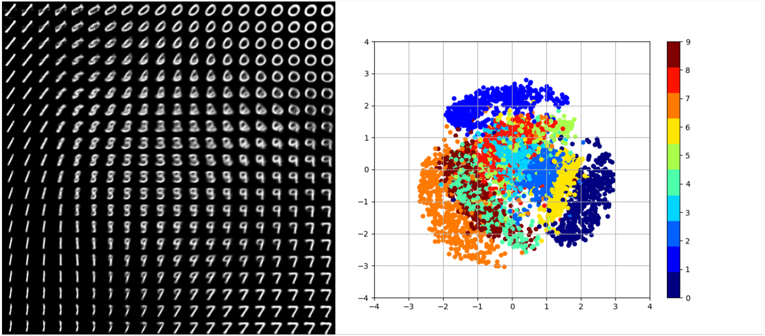


Figura 10: Image produced from learnt manifold. Plot of 2D latent space .

Variational Autoencoder - Annomaly Detection

- Train VAE with Normal class.
- Measure the quality of reconstruction.
- Classify as anomalous if the reconstruction is above some threshold.

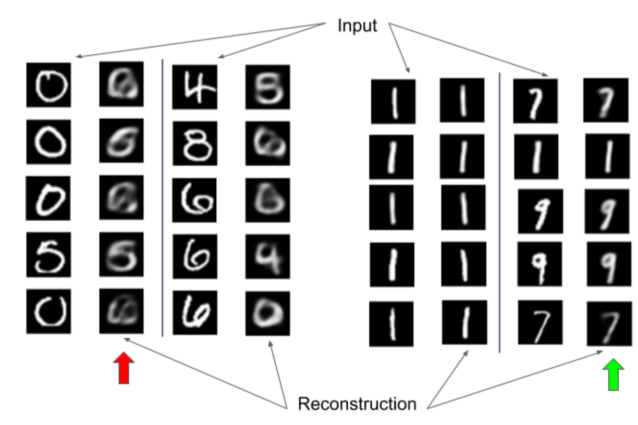


Figura 11: Image from: [An and Cho, 2015]

Variational Autoencoder - Face Editing.

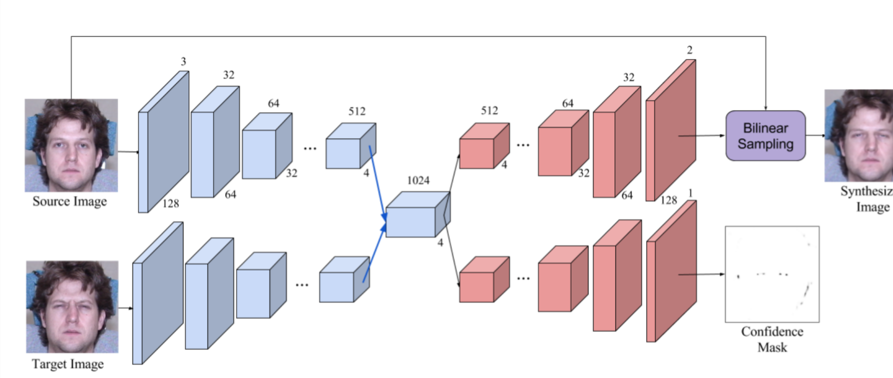


Figura 12: [Yeh et al., 2016].

RECOGNA
LABORATORY

Variational Autoencoder - New Molecule Findings

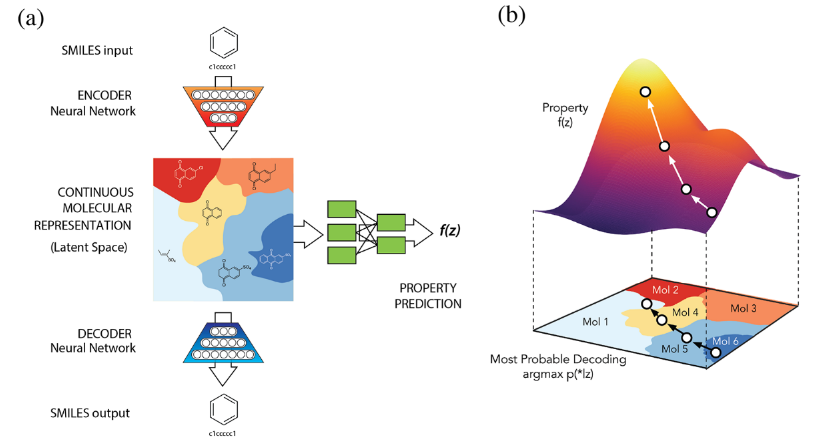


Figura 13: Credits: [Gómez-Bombarelli et al., 2018]

RECOGNA LABORATORY

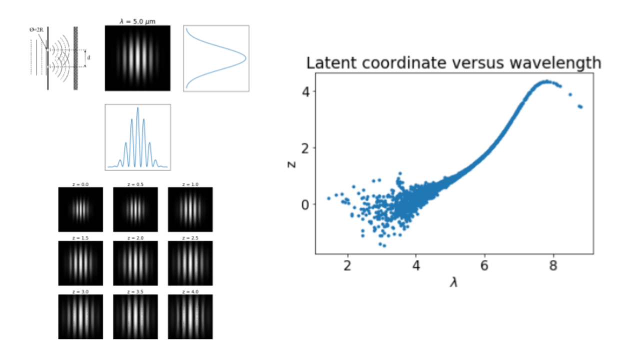


Figura 14: Credits:[Vargas et al., 2018]

Variational Autoencoder - Stellar Clustering

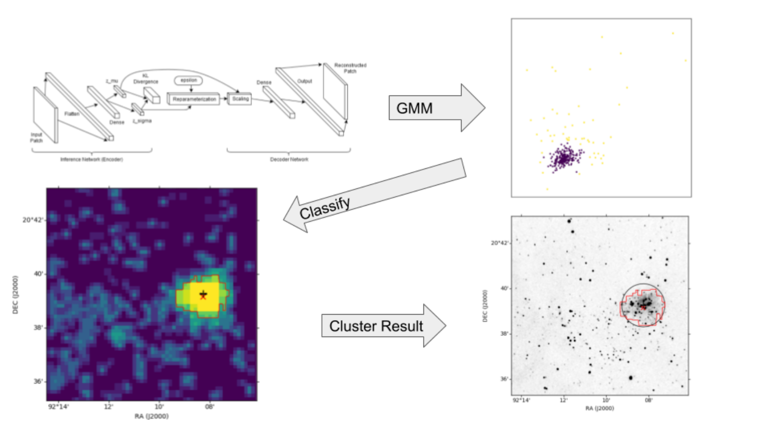


Figura 15: Credits: [Karmakar et al., 2018]

RECOGNA
LABORATORY

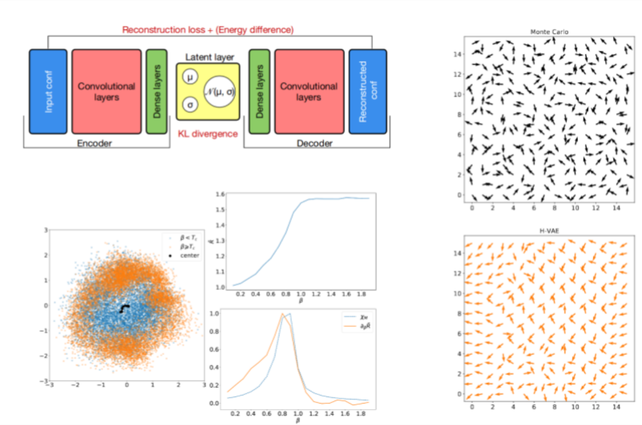


Figura 16: Credits: [Cristoforetti et al., 2017]

Variational Autoencoder - Fluid Simulation

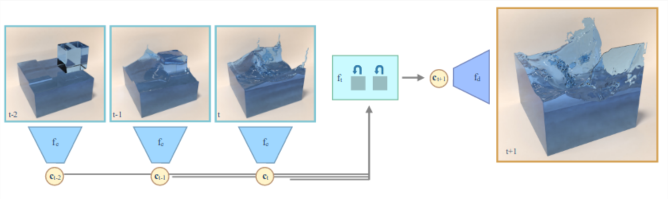


Figura 17: Credits: [Wiewel et al., 2019]

Variational Autoencoder - Fluid Simulation

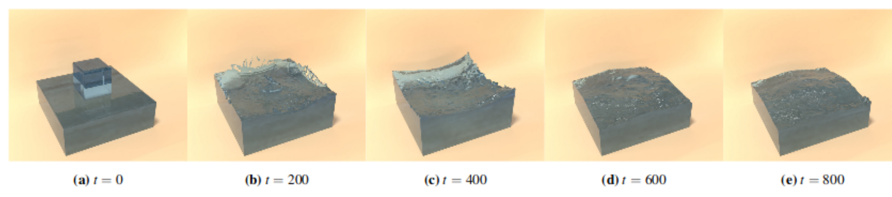


Figura 18: Credits: [Wiewel et al., 2019]

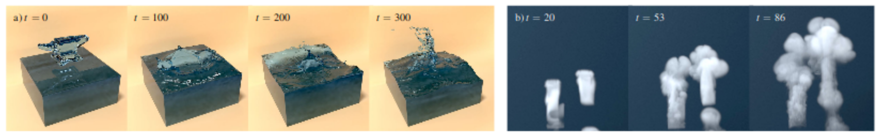


Figura 19: Credits: [Wiewel et al., 2019]

RECOGNA LABORATORY

Variational Autoencoder - Object Position

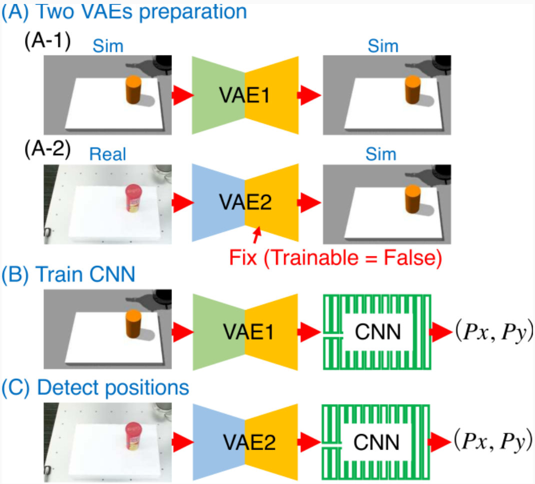


Figura 20: Credits: [Inoue et al., 2017]

RECOGNA LABORATORY

Variational Autoencoder - Conclusions

Pros:

- Well stablished framework.
- Easier to train.
- Allows to infer $q(z|x)$.
- Useful representation learning.

Cons:

- Generated data are blurry.
- Too many assumptions (simplifications) about the priors.

Future:

- VQ-VAE.
- GMM.
- More complex PDF estimators.

RECOGNA LABORATORY

References i

An, J. and Cho, S. (2015).
Variational autoencoder based anomaly detection using reconstruction probability.
Special Lecture on IE, 2(1).

Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017).
Variational inference: A review for statisticians.
Journal of the American statistical Association, 112(518):859–877.

Cristoforetti, M., Jurman, G., Nardelli, A. I., and Furlanello, C. (2017).
Towards meaningful physics from generative models.
arXiv preprint arXiv:1705.09524.

RECOGNA LABORATORY

References ii

Doersch, C. (2016).
Tutorial on variational autoencoders.
arXiv preprint arXiv:1606.05908.

Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P., and Aspuru-Guzik, A. (2018).
Automatic chemical design using a data-driven continuous representation of molecules.
ACS central science, 4(2):268–276.

RECOGNA
LABORATORY

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., and Lerchner, A. (2017).
beta-vae: Learning basic visual concepts with a constrained variational framework.
ICLR, 2(5):6.

Inoue, T., Chaudhury, S., De Magistris, G., and Dasgupta, S. (2017).
Transfer learning from synthetic to real images using variational autoencoders for robotic applications.
arXiv preprint arXiv:1709.06762.

RECOGNA LABORATORY

References iv

Karmakar, A., Mishra, D., and Tej, A. (2018).
Stellar cluster detection using gmm with deep variational autoencoder.
In *2018 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, pages 122–126. IEEE.

Kingma, D. P. and Welling, M. (2013).
Auto-encoding variational bayes.
arXiv preprint arXiv:1312.6114.

Sohn, K., Lee, H., and Yan, X. (2015).
Learning structured output representation using deep conditional generative models.
In *Advances in neural information processing systems*, pages 3483–3491.

RECOGNA LABORATORY

van den Oord, A., Vinyals, O., et al. (2017).
Neural discrete representation learning.
In *Advances in Neural Information Processing Systems*, pages 6306–6315.

Vargas, J. C. R., Novaes, S. F., Cóbe, R., Iope, R., Stanzani, S. L., and Tomei, T. (2018).
Shedding light on variational autoencoders.
In *2018 XLIV Latin American Computer Conference (CLEI)*, pages 294–298. IEEE.

RECOGNA
LABORATORY

References vi

Wiewel, S., Becher, M., and Thuerey, N. (2019).
Latent space physics: Towards learning the temporal evolution of fluid flow.
In *Computer Graphics Forum*, volume 38, pages 71–82. Wiley Online Library.

Yeh, R., Liu, Z., Goldman, D. B., and Agarwala, A. (2016).
Semantic facial expression editing using autoencoded flow.
arXiv preprint arXiv:1611.09961.

RECOGNA LABORATORY