



PUCP

Skoltech

Skolkovo Institute of Science and Technology

Towards Disentangled Representations via Variational Sparse Coding

LatinX in AI Research Workshop - ICML 2019

Robert Aduviri and Alfredo De La Fuente

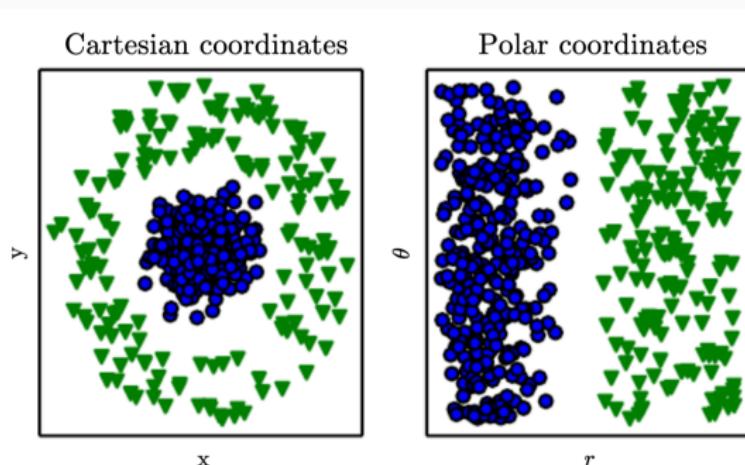
Pontifical Catholic University of Peru, Skolkovo Institute of Science and Technology

Table of contents

1. Motivation
2. Research Problem
3. Technical Contribution
4. Current Results
5. Next Steps

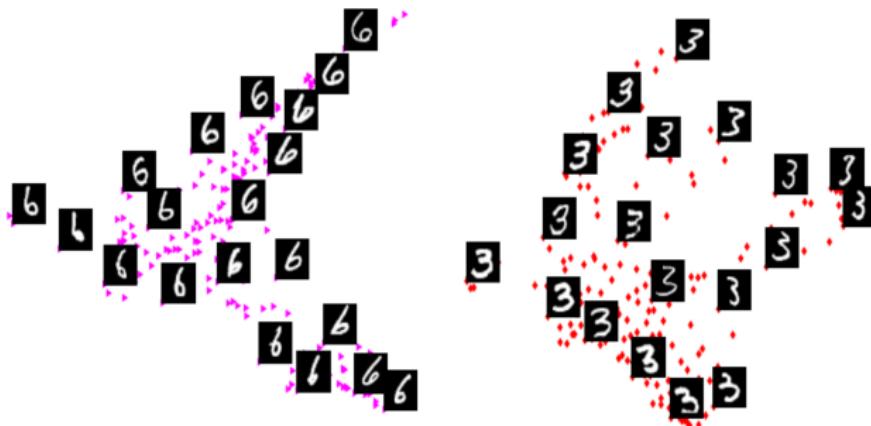
Motivation

Representation Learning



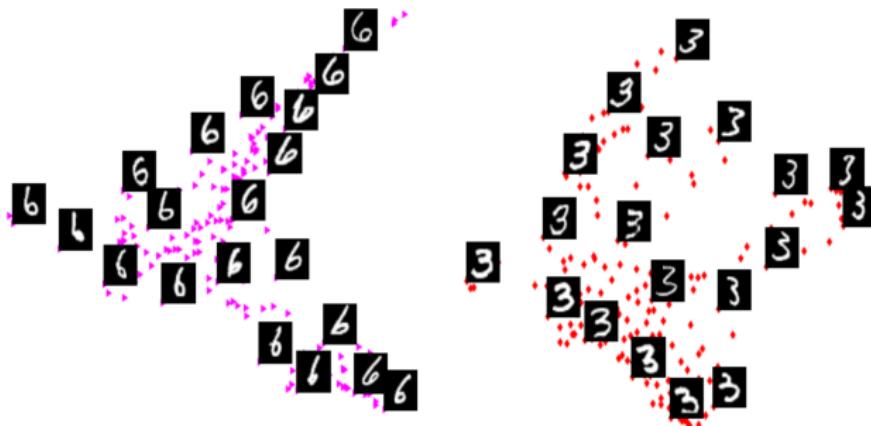
- Simple machine learning algorithms depend *heavily* on the representation of the data they are given.
- The process of designing the right representation for a specific task is commonly known as **feature engineering**.
- An alternative to hand-design these representations is to **learn** them automatically.

Representation Learning



- Lower dimensional representation of raw data (images, text, etc.).
- Efficiently sample from a high-dimensional data distribution.
- Latent space with meaningful properties.

Representation Learning



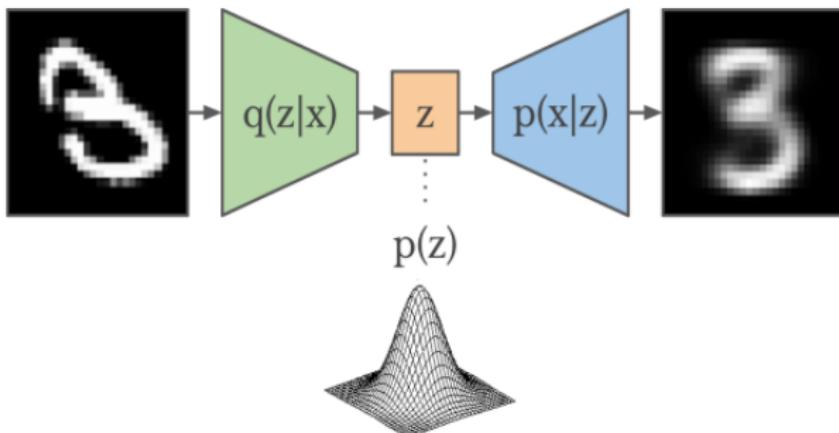
- Lower dimensional representation of raw data (images, text, etc).
- Efficiently sample from a high-dimensional data distribution.
- Latent space with meaningful properties.

→ Generative Models to the rescue!

Variational AutoEncoders (VAE)

Proposed by Kingma & Welling (2013) and Rezende et al. (2014)

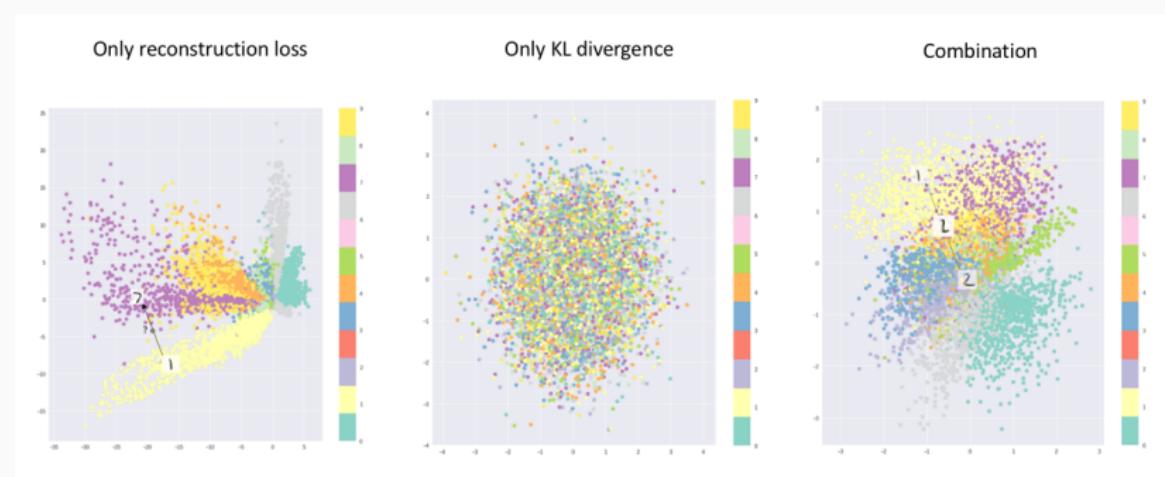
$$\mathcal{L}(\theta, \phi) = \mathbb{E}_{q_\phi(z|x)} [\log p_\theta(x|z)] - \text{KL}(q_\phi(z|x) \| p(z)) \quad (1)$$



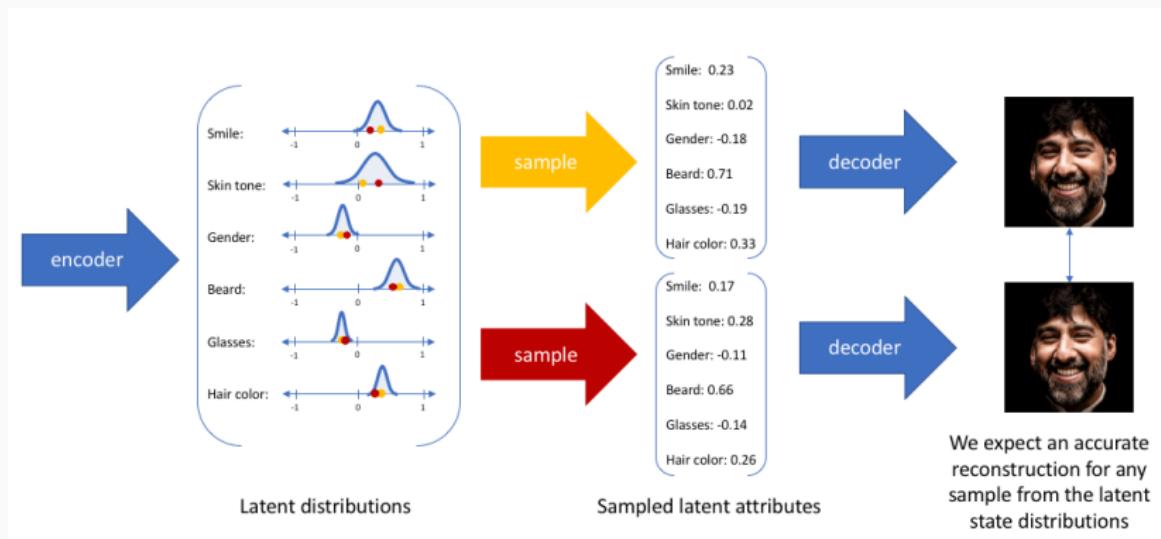
*How expressive can a Gaussian
latent prior distribution be?*

VAE vs AE

Both the reconstruction loss and the KL divergence are necessary to produce a *smooth* latent representation of the data.

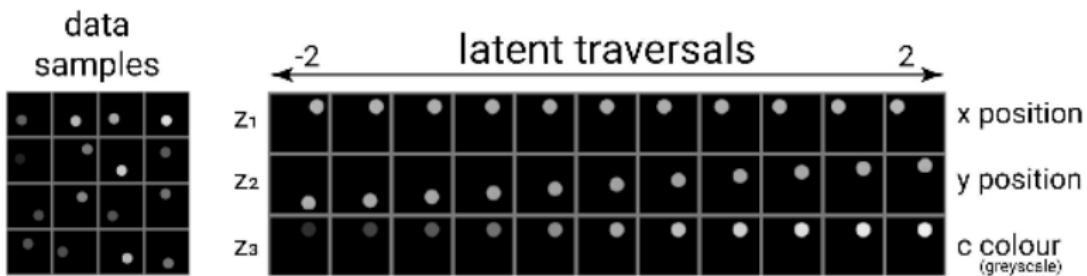


VAE latent codes distribution



Disentanglement

Constitutes the *complex* task of learning representations that separate the **underlying structure** of the world into disjoint parts of its representation.

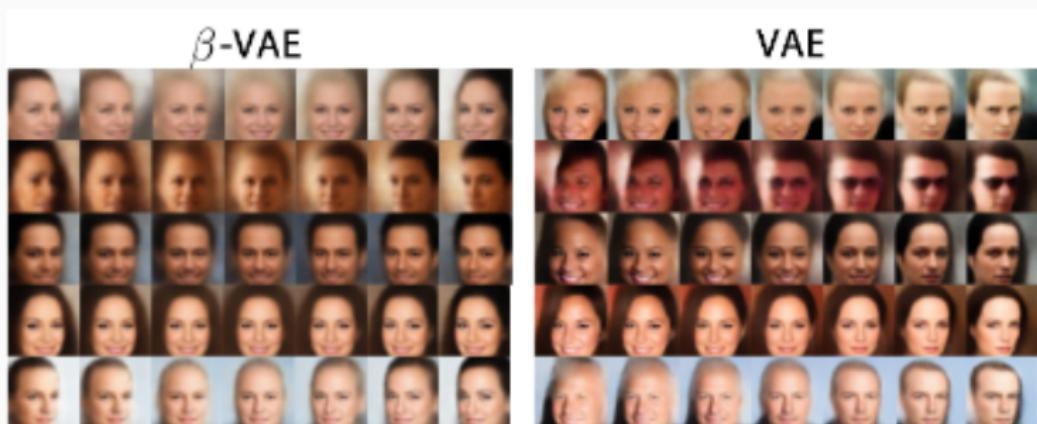


Scheme from the paper “Towards a Definition of Disentangled Representations” by Higgins et al. (2018)

β - VAE

Proposed by Higgins et al. (2017) as a constrained version of VAE to discover disentangled latent factors.

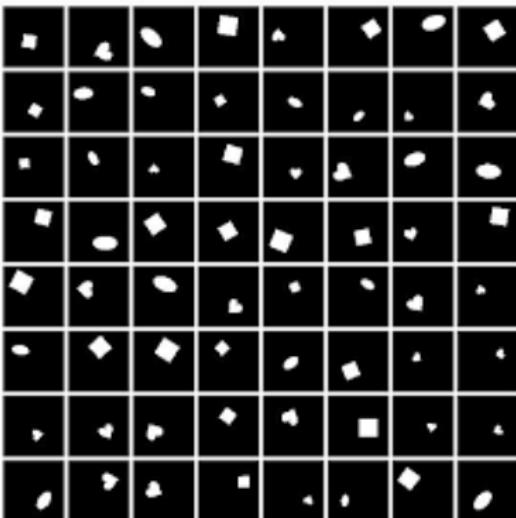
$$\mathcal{L}_{\text{Beta}}(\theta, \phi) = \mathbb{E}_{q_\phi(z|x)} [\log p_\theta(x|z)] - \beta \text{KL}(q_\phi(z|x) \| p(z)) \quad (2)$$



Azimuth(orientation) traversal comparison.

dSprites Dataset

Created by Matthey et al. (2017) as a way to assess the disentanglement properties of unsupervised learning methods. These 2D shapes were procedurally generated from 6 ground truth independent latent factors: color, shape, scale, rotation, x and y positions of a sprite.



Research Problem

Learning Disentangled Representations

We aim to tackle the following challenges:

- Meaningful low-dimensional representations of images

Learning Disentangled Representations

We aim to tackle the following challenges:

- Meaningful low-dimensional representations of images
- Interpretable and disentangled features on latent space.

Learning Disentangled Representations

We aim to tackle the following challenges:

- Meaningful low-dimensional representations of images
- Interpretable and disentangled features on latent space.
- Quantitatively and qualitative evaluation of disentanglement.

Technical Contribution

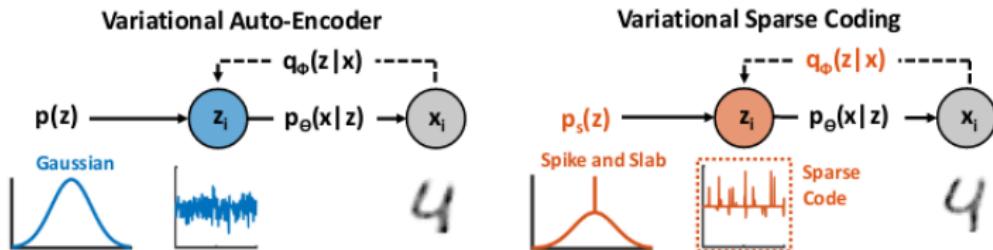
Variational Sparse Coding (VSC)

Tonolini et al. (2019) suggest the use of a **Spike-and-Slab** prior $p(z)$.

$$p_s(z) = \prod_{j=1}^J (\alpha \mathcal{N}(z_j; 0, 1) + (1 - \alpha)\delta(z_j)) \quad (3)$$

which leads to a recognition function as a discrete mixture model,

$$q_\phi(z|x_i) = \prod_{j=1}^J \left(\gamma_{i,j} \mathcal{N}(z_{i,j}; \mu_{z,i,j}, \sigma_{z,i,j}^2) + (1 - \gamma_{i,j})\delta(z_{i,j}) \right) \quad (4)$$



The model captures subjectively **understandable** sources of variation.

Convolutional encoder/decoder

A convolutional architecture was used for the encoder/decoder of the VAE and VSC for comparison, based on the configuration used by Higgins et al. (2017).

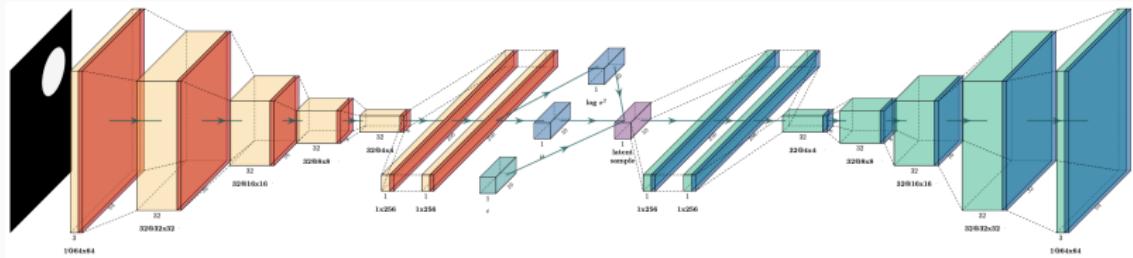


Figure 1: Convolutional architecture used for VAE and VSC

Current Results

Latent Codes Comparison

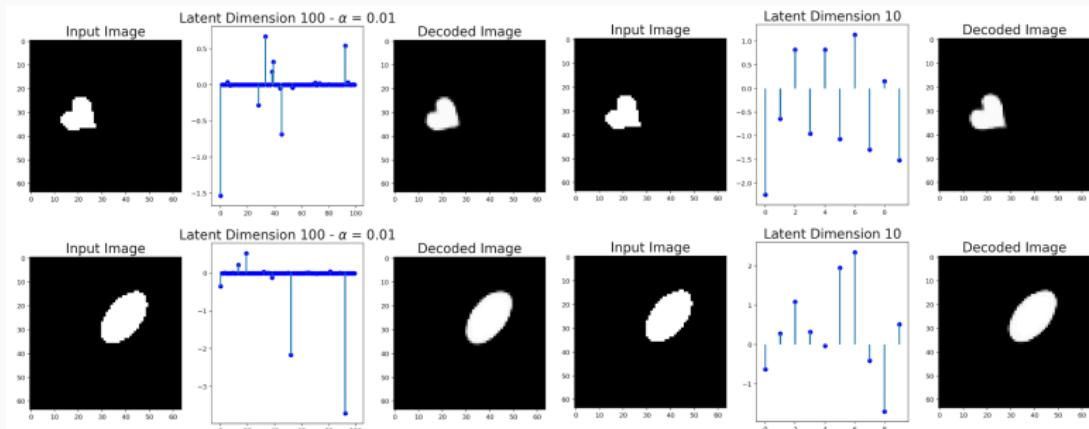


Figure 2: Reconstruction and latent codes of Convolutional VSC (left) ($\alpha = 0.01$, $\beta = 2$) and Convolutional VAE (right) ($\beta = 2$) models with the dSprites dataset.

Latent Space Traversal via VSC

3 3 3 3 3 3 3
7 7 7 7 7 7 7
4 4 4 4 4 4 4
1 1 1 1 1 1 1
9 9 9 9 9 9 9
2 2 2 2 2 2 3
9 9 9 9 9 4 4 4

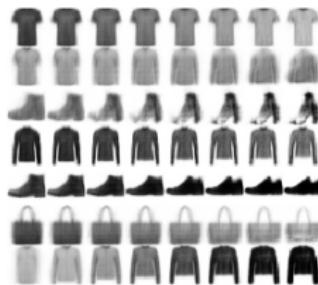


Figure 3: Latent traversals on MNIST (left) and Fashion-MNIST (right).

Latent Space Traversal via VSC

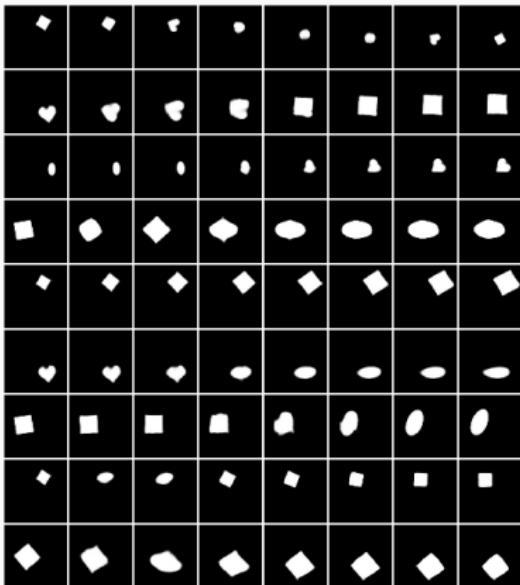
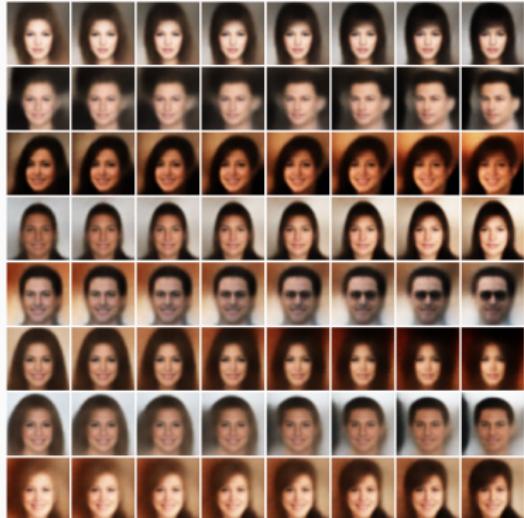


Figure 4: Latent traversals on CelebA (left) and dSprites (right).

Latent Space Traversal Comparison

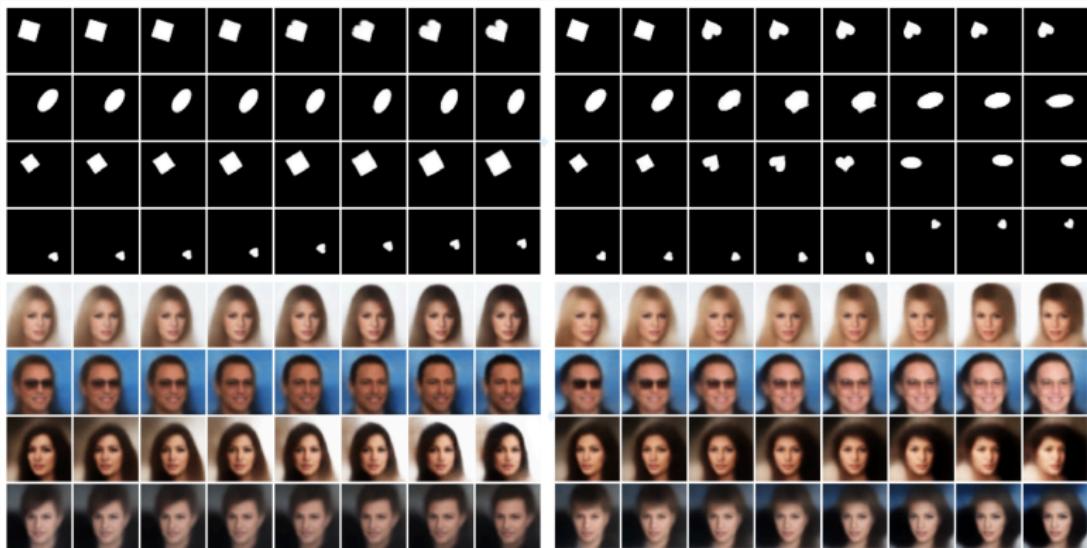


Figure 5: Latent traversals using the Convolutional VSC (left) and Convolutional VAE (right) models with the dSprites and CelebA datasets.

Next Steps

Disentanglement Metrics and Models

The quantitative evaluation of disentanglement is a recent area of research, with several metrics being constantly proposed, in addition to new models and datasets:

- **Metrics:** BetaVAE score, FactorVAE score, Mutual Information Gap, SAP score, DCI, MCE, IRS
- **Models:** BetaVAE, FactorVAE, BetaTCVAE, DIP-VAE, InfoGAN
- **Datasets:** dSprites, Color/Noisy/Scream-dSprites, SmallNORB, Cars3D, Shapes3D

Next Steps

- Perform quantitative disentanglement evaluation with previously proposed metrics.

Next Steps

- Perform quantitative disentanglement evaluation with previously proposed metrics.
- Extend comparison with recent models also proposed for disentanglement, both VAE-based and GAN-based.

Next Steps

- Perform quantitative disentanglement evaluation with previously proposed metrics.
- Extend comparison with recent models also proposed for disentanglement, both VAE-based and GAN-based.
- Perform ablation studies for key features of the model, such as the sparse prior, β -VAE regularization and encoder/decoder used.

Thank you!

Our source code and experiments are available at:

github.com/Alfo5123/Variational-Sparse-Coding

See you at the poster session!

Contact info:

robert.aduviri@pucp.edu.pe

alfredo.delafuente@skoltech.ru