Variational Autoencoders- II
Conditional, Joint, Semi-Supervised

Sergei V. Kalinin

PURPOSE

- (Super-brief) introduction into Neural Networks
- What are (Variational) autoencoders?
- Key notions:
    - Encoding and decoding
    - Latent distribution
    - Latent representations
- Why invariances: rotational, translational, and scale
- Other colors of VAEs:
    - Semi-supervised
    - Conditional
    - Joint
- From VAEs to encoder-decoders (VED)
- Further opportunities:
    - Physics constraints
    - Representation learning
- Active learning: DKL

Autoencoders

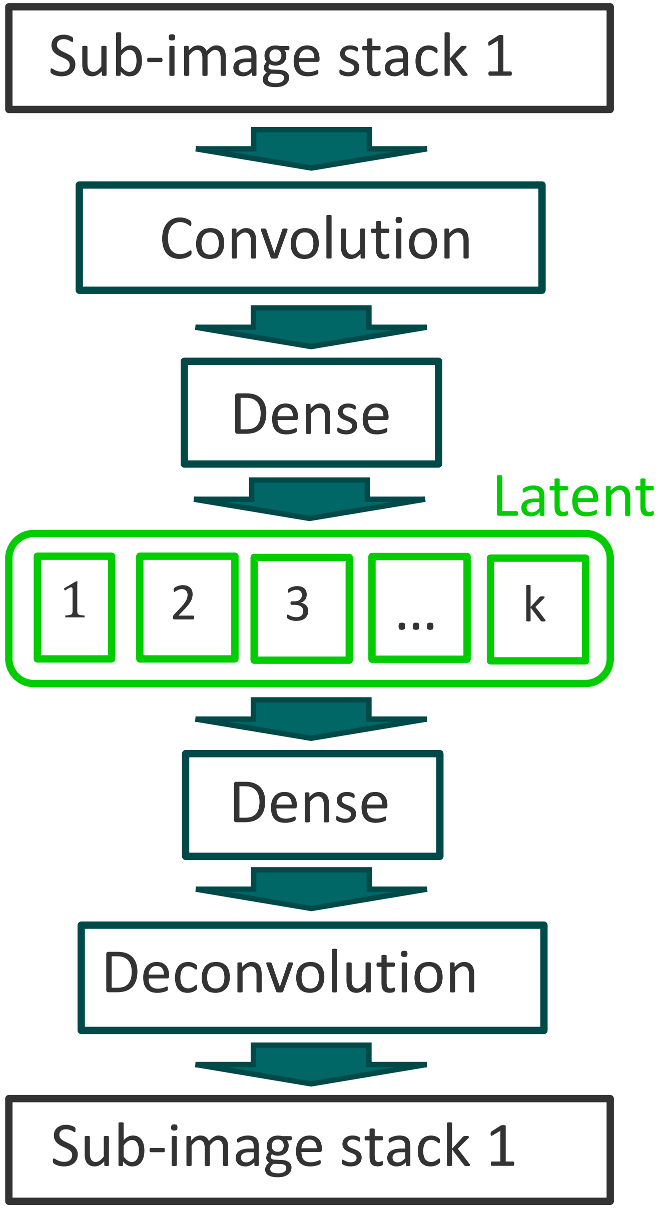
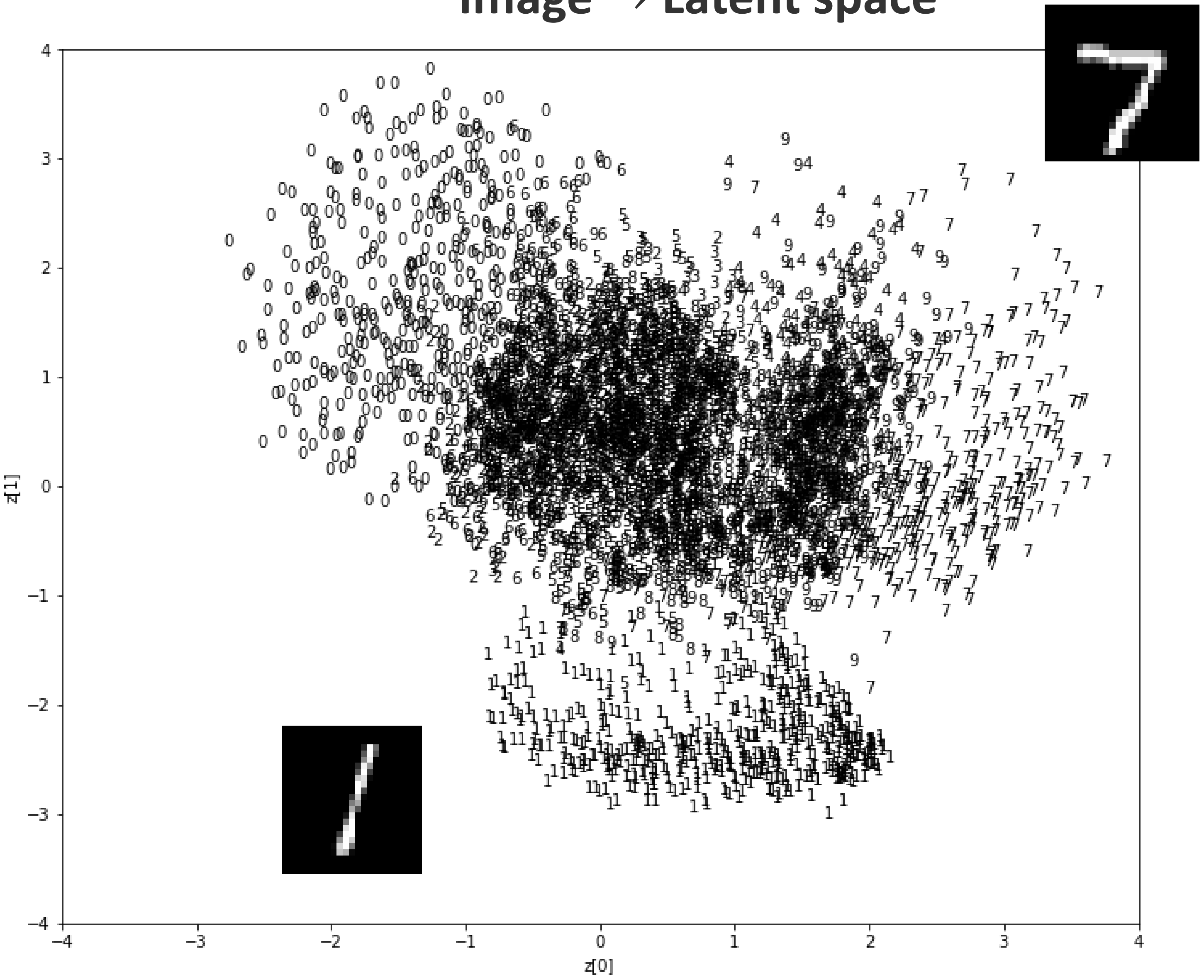
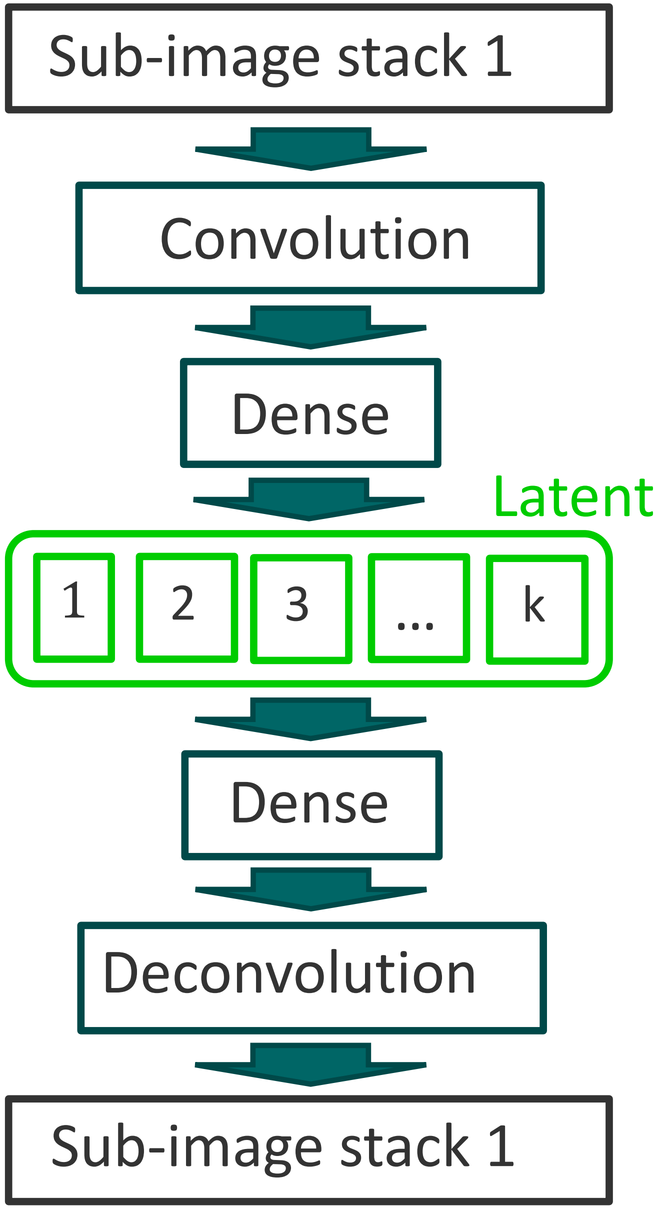


Image → Latent space

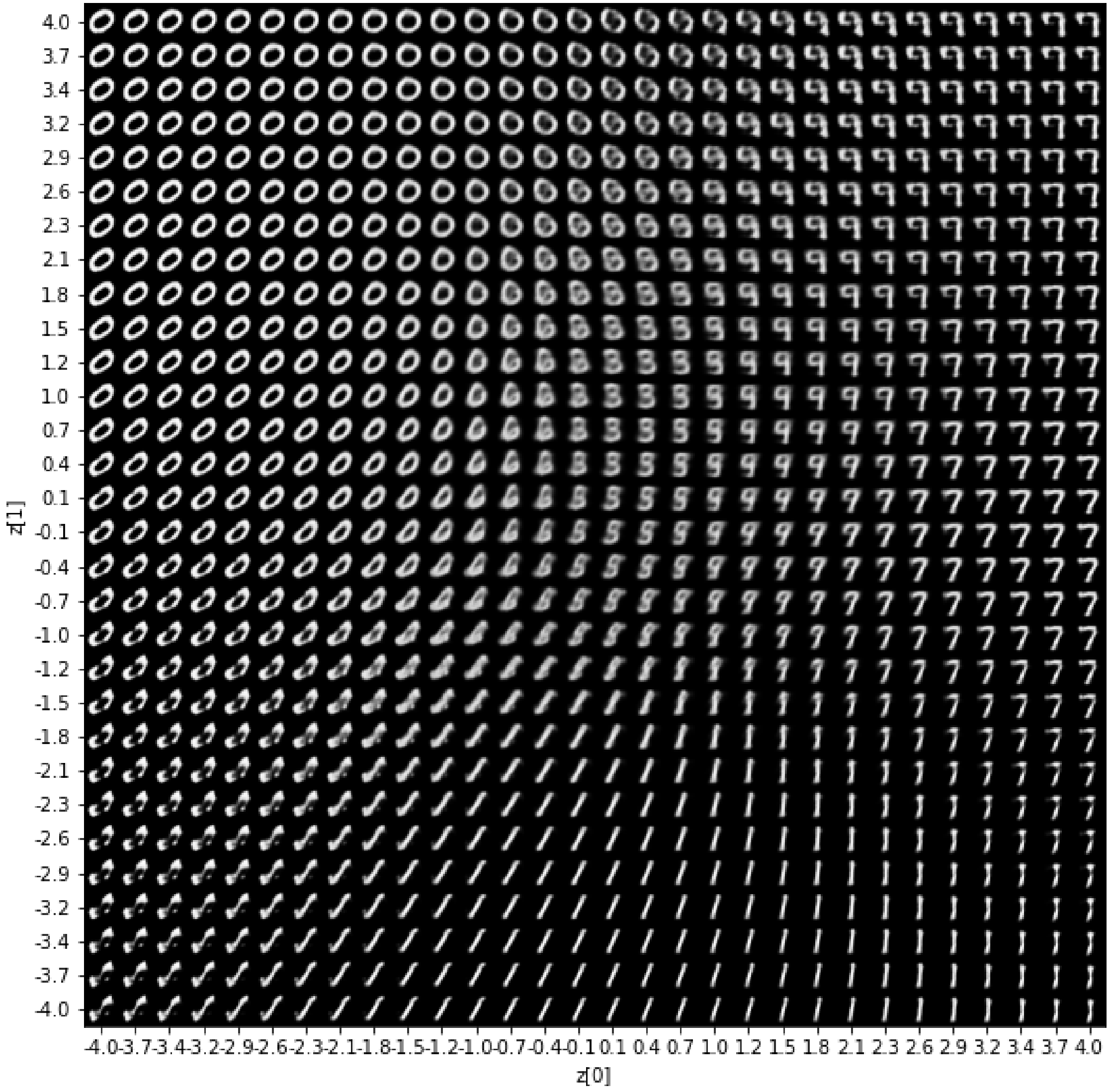


Encoding the data via low dimensional vector

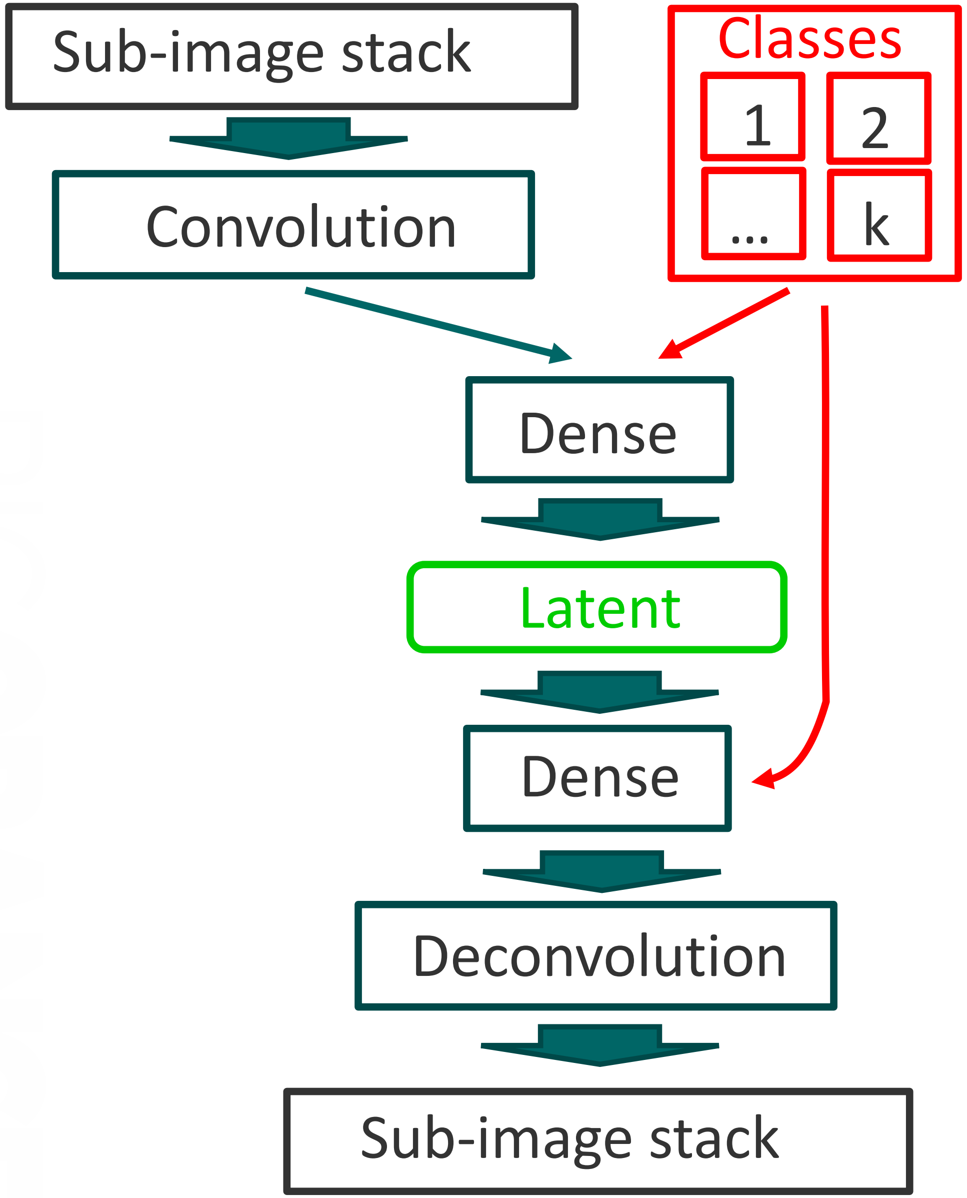
Autoencoders

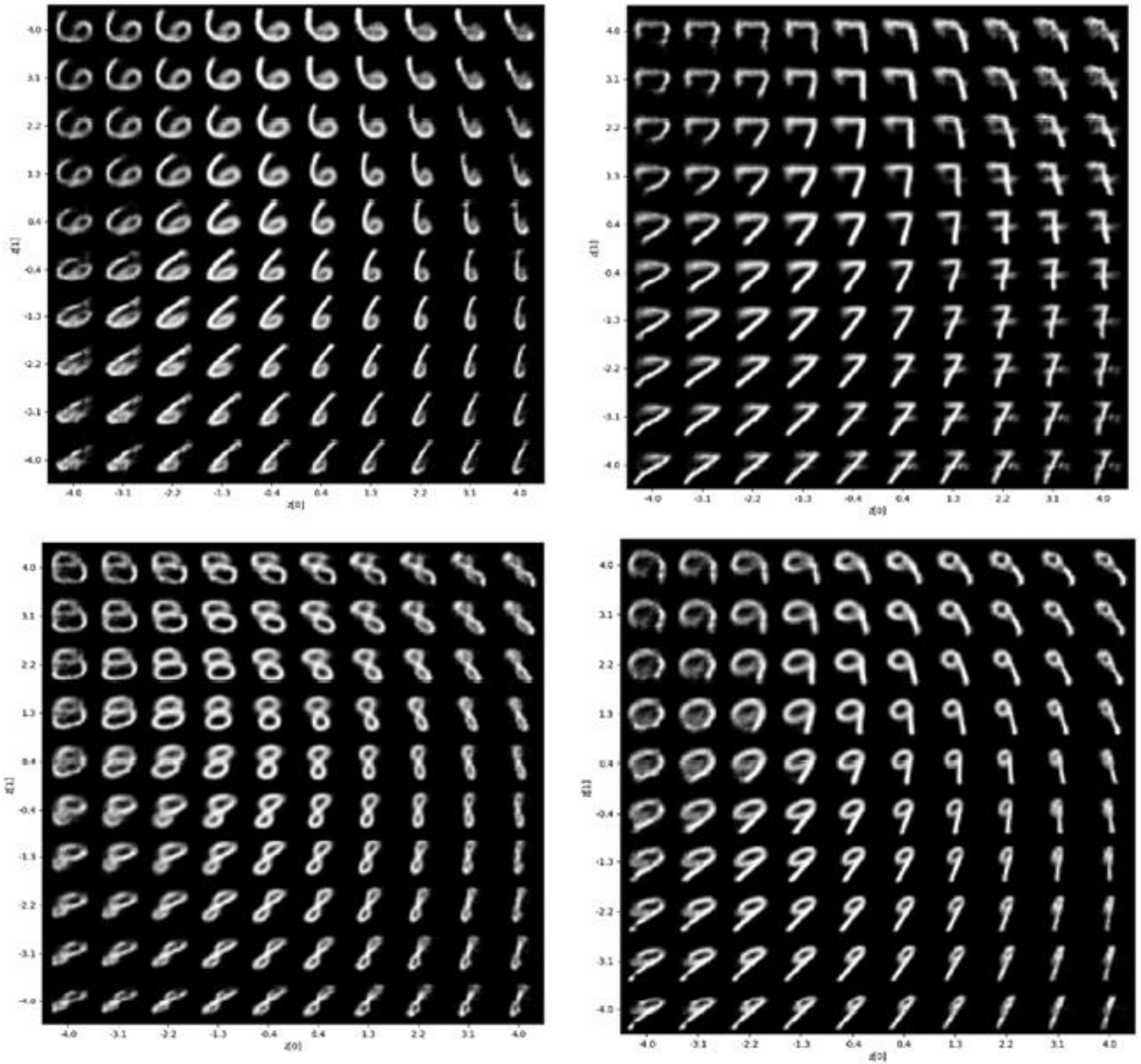


Latent space → Image

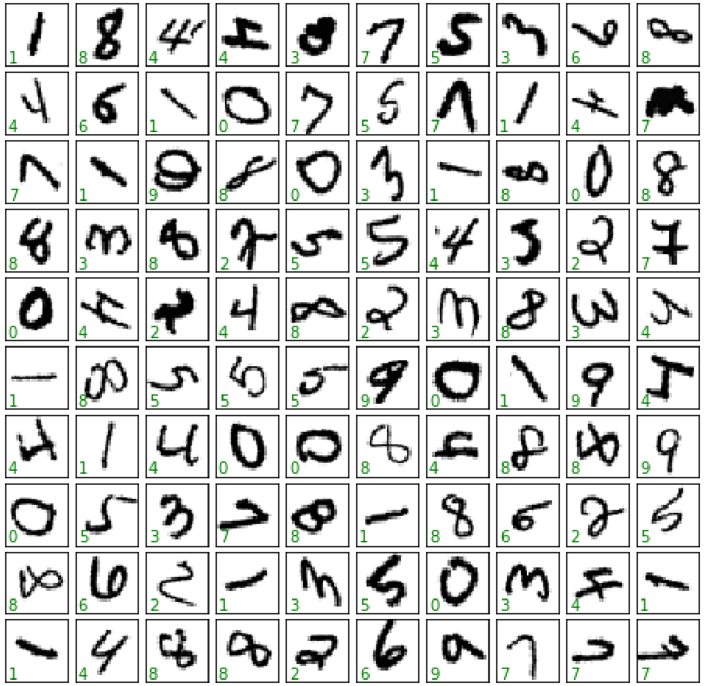


Conditional VAE

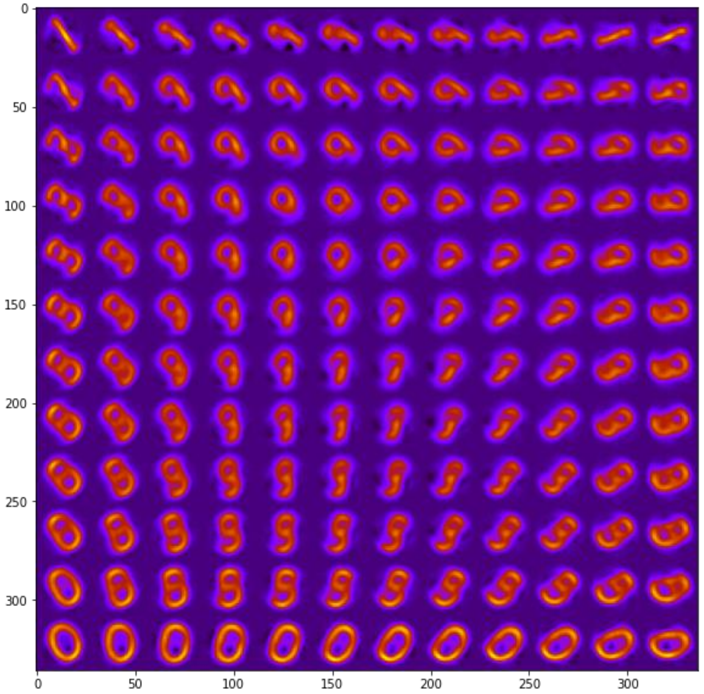




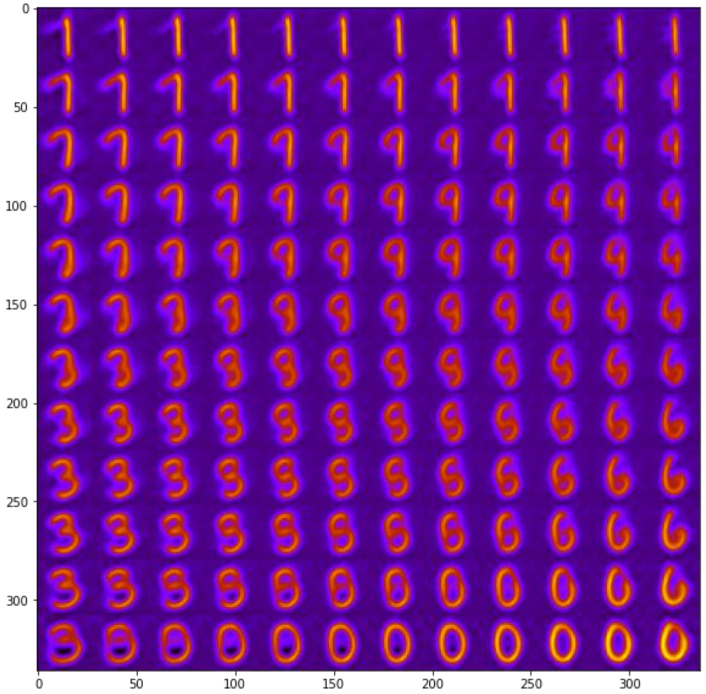
Start with MNIST



VAE



rVAE



Problem 1: Invariances

Problem 2: We share the latent space between multiple factors of variability

- digits
- traits

But first: how do we control latent spaces?

Hyperspherical Variational Auto-Encoders

Tim R. Davidson* Luca Falorsi* Nicola De Cao* Thomas Kipf Jakub M. Tomczak

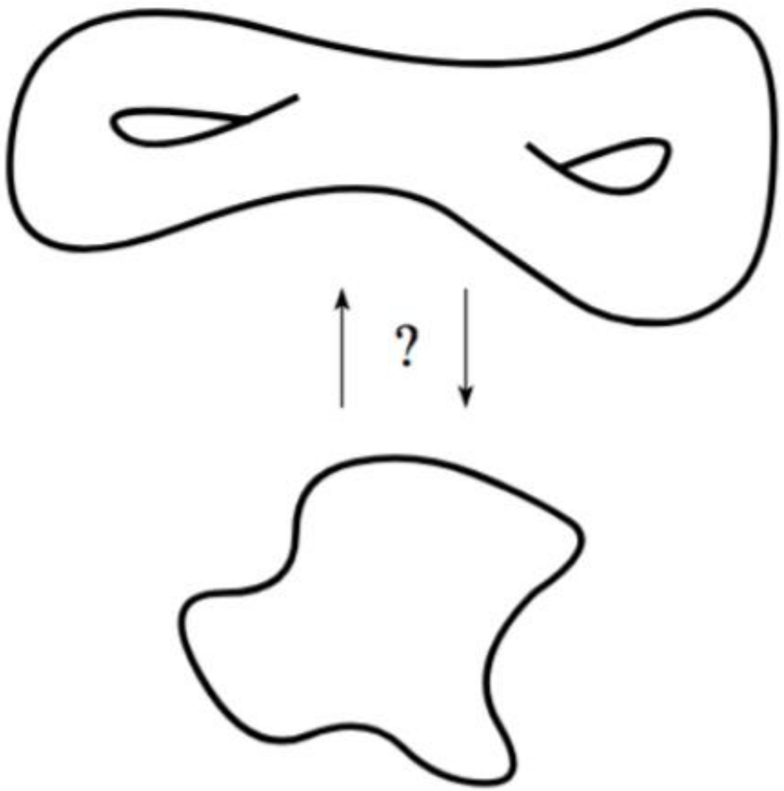
University of Amsterdam

(a) Original (b) Autoencoder (c) $\mathcal{N}$-VAE (d) $\mathcal{N}$-VAE, $\beta = 0.1$ (e) $\mathcal{S}$-VAE

Figure 1: Plots of the original latent space (a) and learned latent space representations in different settings, where β is a re-scaling factor for weighting the KL divergence. (Best viewed in color)

Explorations in Homeomorphic Variational Auto-Encoding

Luca Falorsi [*1] Pim de Haan [*1] Tim R. Davidson [*1] Nicola De Cao [1] Maurice Weiler [1]
Patrick Forré [1] Taco S. Cohen [1 2]



The VAE is a latent variable model, in which $\mathbf{x}$ denotes a set of observed variables, $\mathbf{z}$ stochastic latent variables, and $p(\mathbf{x}, \mathbf{z}) = p(\mathbf{x}|\mathbf{z})p(\mathbf{z})$ a parameterized model of the joint distribution called the *generative model*. Given a dataset $\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_N\}$, we typically wish to maximize the average marginal log-likelihood $\frac{1}{N}\log p(\mathbf{X}) = \frac{1}{N}\sum_{i=1}^{N}\log\int p(\mathbf{x}_i, \mathbf{z}_i)d\mathbf{z}$, w.r.t. the parameters. However when the model is parameterized by neural networks, the marginalization of this expression is generally intractable. One solution to overcome this issue is applying variational inference in order to maximize the *Evidence Lower Bound* (ELBO) for each observation:

$$\begin{aligned}\log p(\mathbf{x}) &= \log\int p(\mathbf{x}, \mathbf{z})d\mathbf{z} \\ &\geq \mathbb{E}_{q(\mathbf{z})}[\log p(\mathbf{x}|\mathbf{z})] - KL(q(\mathbf{z})||p(\mathbf{z})), \quad (1)\end{aligned}$$

where the approximate posterior $q(\mathbf{z})$ belongs to the variational family $\mathcal{Q}$. To make inference scalable an *inference network* $q(\mathbf{z}|\mathbf{x})$ is introduced that outputs a probability distribution for each data point $\mathbf{x}$, leading to the final objective

$$\mathcal{L}(\mathbf{x};\theta) = \mathbb{E}_{q(\mathbf{z}|\mathbf{x})}[\log p(\mathbf{x}|\mathbf{z})] - KL(q(\mathbf{z}|\mathbf{x})||p(\mathbf{z})), \quad (2)$$

with θ representing the parameters of p and q. The ELBO can be efficiently approximated for continuous latent variable $\mathbf{z}$ by Monte Carlo estimates using the *reparameterization trick* of $q(\mathbf{z}|\mathbf{x})$ (Kingma & Welling, 2013; Rezende et al., 2014).

2.2. Lie Groups and Lie Algebras

Lie Group A group is a set equipped with a product that follows the four group axioms: the product is closed and

What if we have multiple classes?

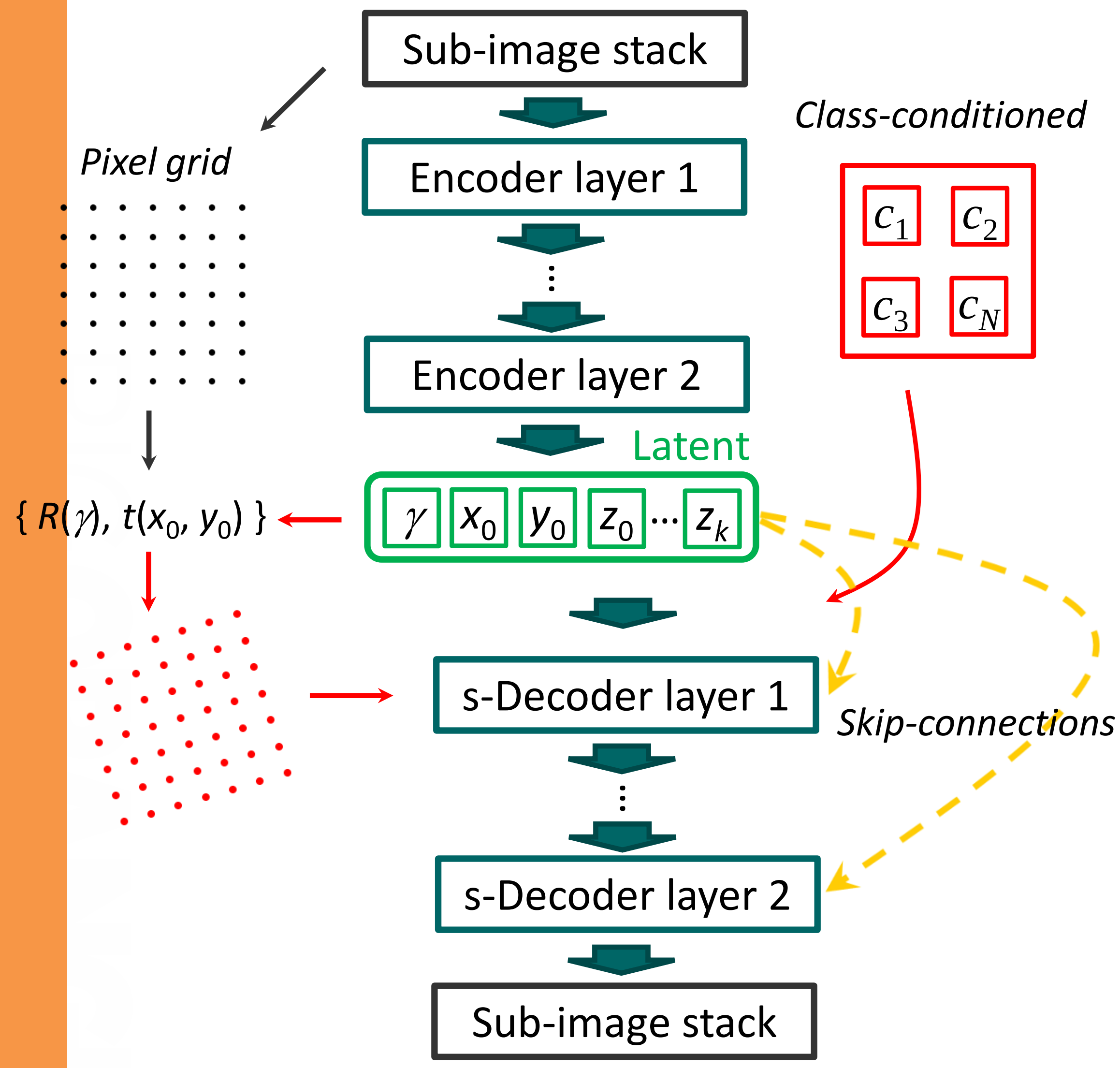
1. Classes are known: conditional (discrete) VAE

2. Factors of variability are known: conditional (continuous) VAE

3. Some classes are known: semi-supervised VAE

4. Number of classes are known: joint VAE

Conditional VAE



- Generative model is a function of spatial coordinate
- 3 additional latent variables to absorb rotations and shifts
- Disentangles rotations and translations from image content
- Ideal for analyzing microscopy sub-images on atomic level

$$\left.\begin{aligned}ELBO &= Reconstruction\ Loss \\ &- D_{KL}\big(q(z|x)\|\mathcal{N}(0, I)\big)\end{aligned}\right\}\ \text{Regular VAE}$$

$$- D_{KL}\left(q(\gamma|x)\|\mathcal{N}(0, s_\gamma^2)\right) \quad \text{Rotation}$$

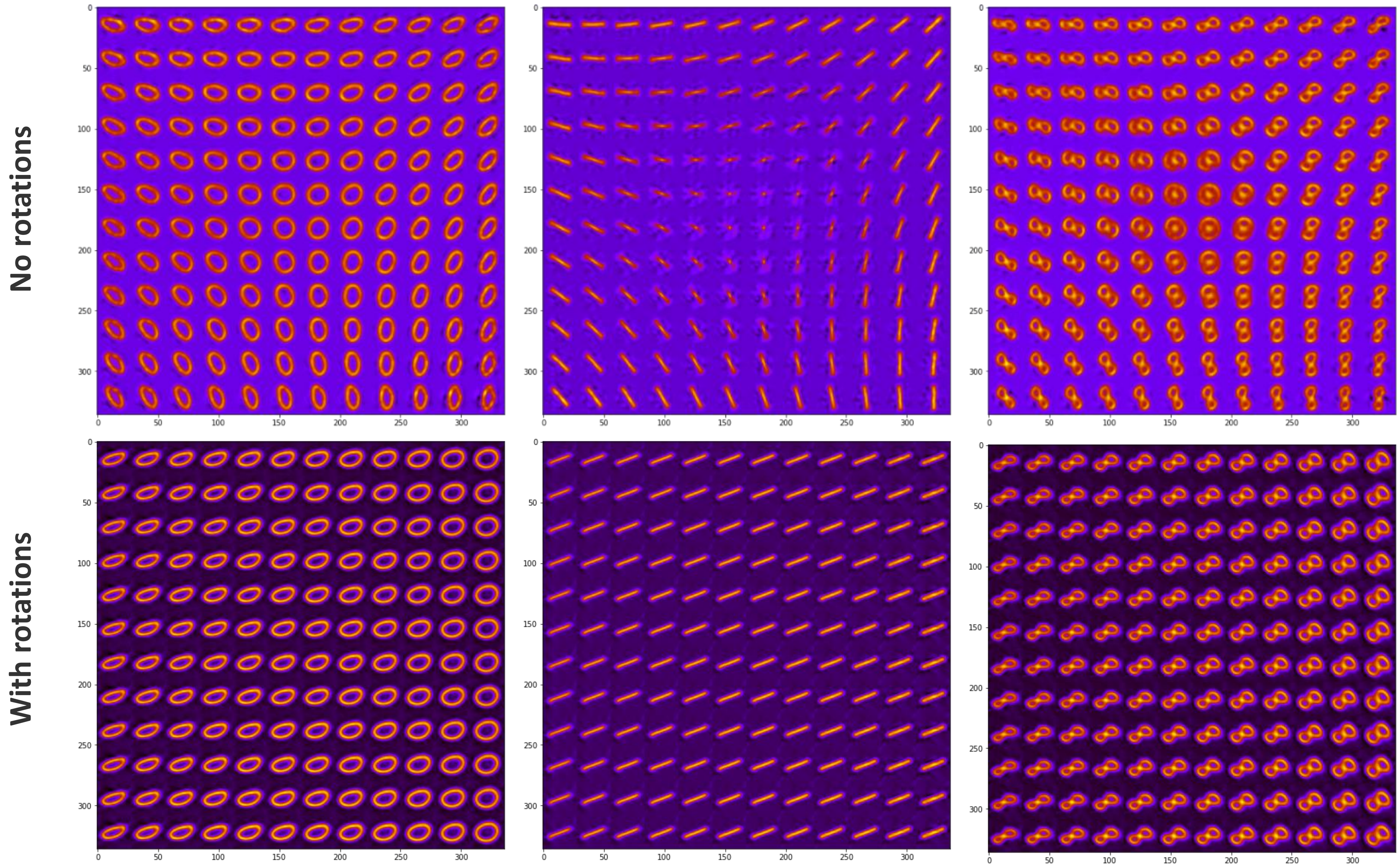
$$- D_{KL}\left(q(\Delta r|x)\|\mathcal{N}(0, s_{\Delta r}^2)\right) \quad \text{Translation}$$

+ D_{KL}(physics-based “priors”) ?

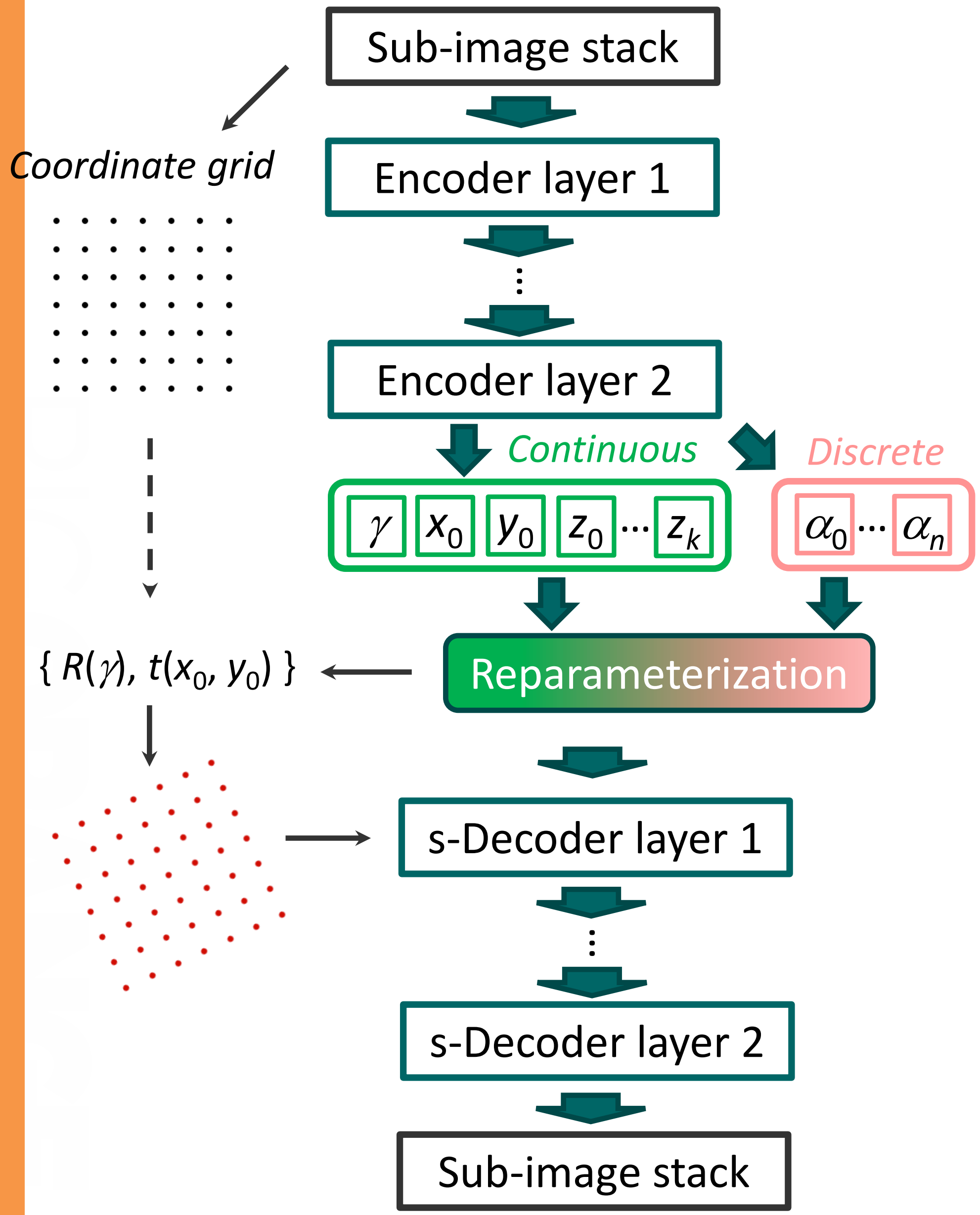
+ D (physics) ?

Bepler *et al.*, NeurIPS 2019

MNIST: cVAE



Joint VAE



- Generative model is a function of spatial coordinate (*e.g.*, via spatial broadcasting)
- 3 additional latent variables to absorb rotations and shifts
- Disentangles rotations and translations from image content
- Learns discrete classes in unsupervised fashion
- Well-suited for analyzing microscopy (sub-)images on atomic and molecular levels

$ELBO =$

$-\ Reconstruction\ Loss$

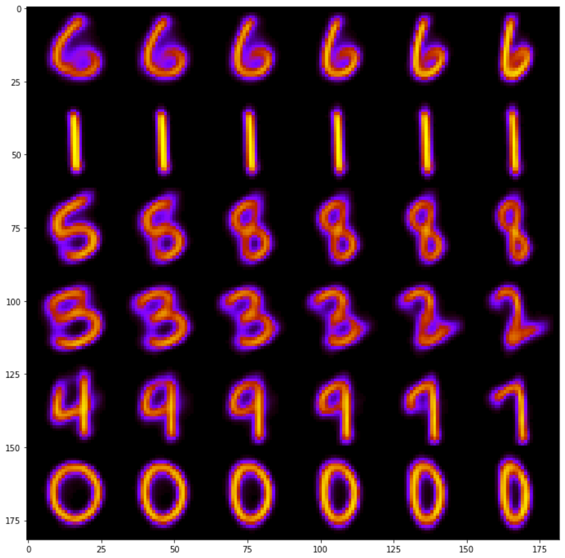
Structure Rotation

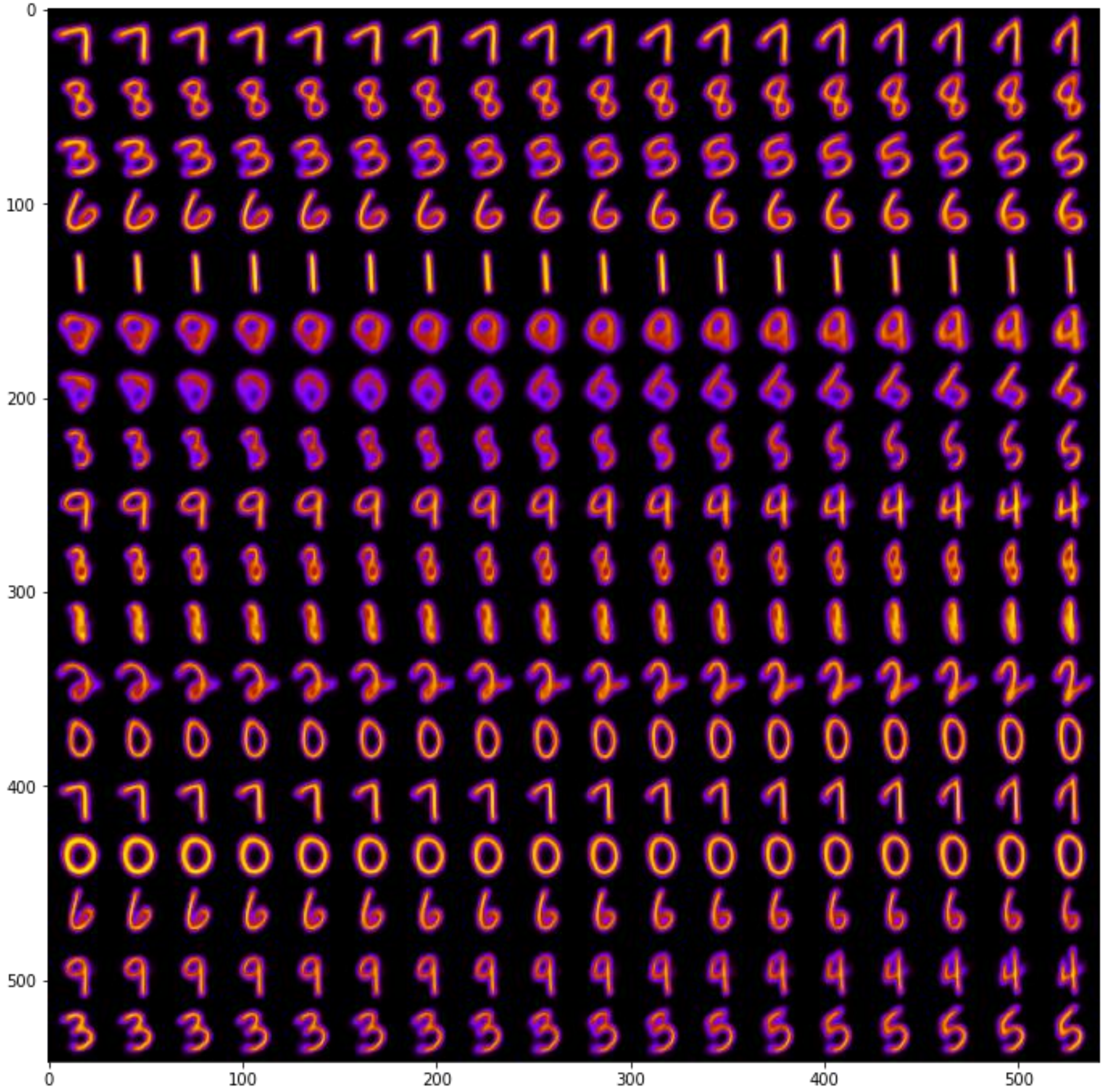
$-\beta_c(t)\left|\left(D_{KL}(q(z|x)\|p(z)) + D_{KL}(q(\gamma|x)\|p(\gamma)\right) - C_z\right|$ Continuous

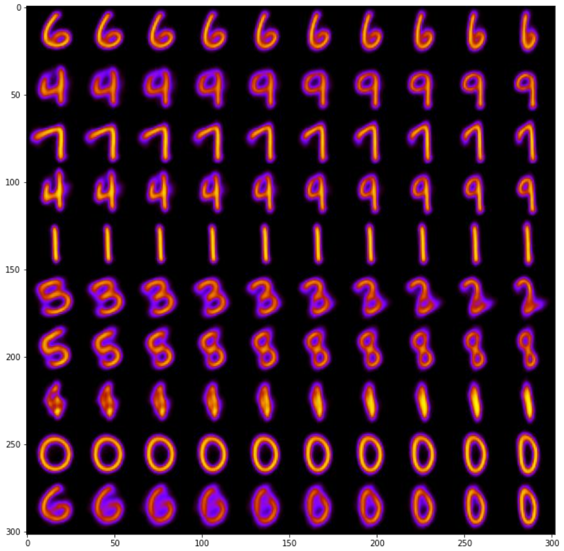
$-\beta_d(t)\left|D_{KL}\left(q(\alpha|x)\|p(\alpha)\right) - C_\alpha\right|$ Discrete

+ physics-based “loss” ?

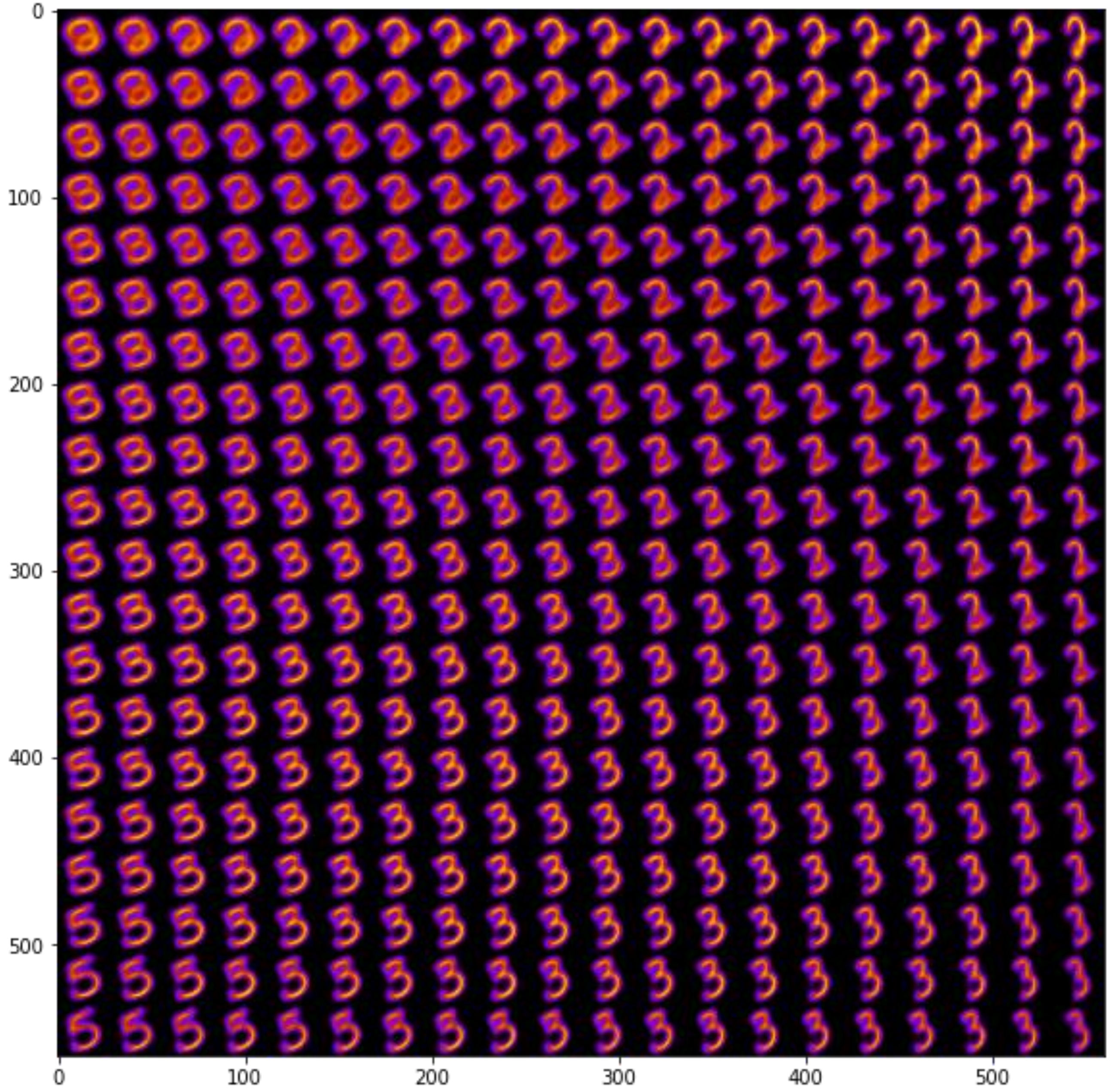
jVAE of MNIST







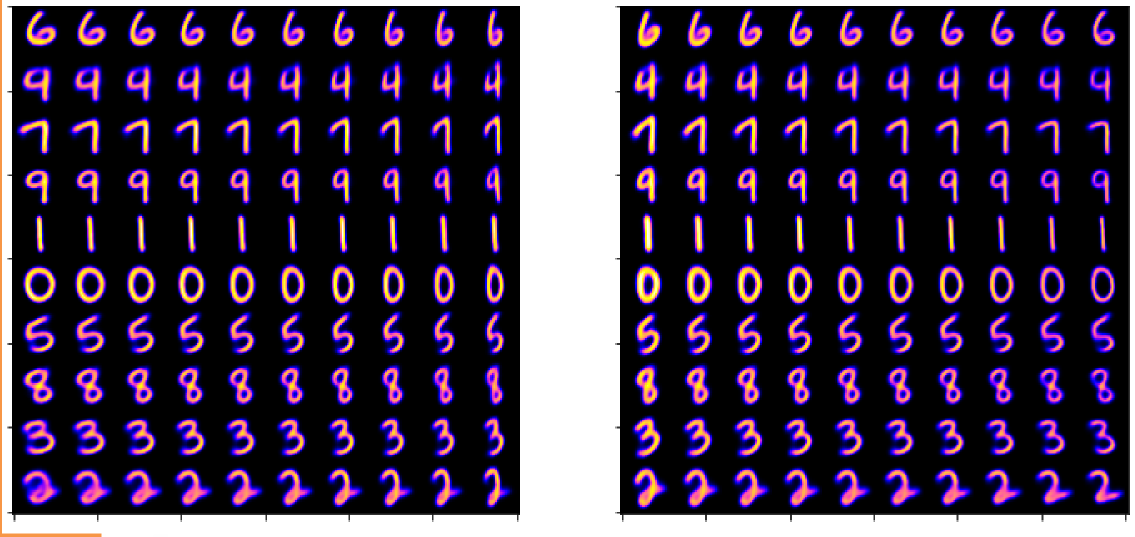
Latent representation



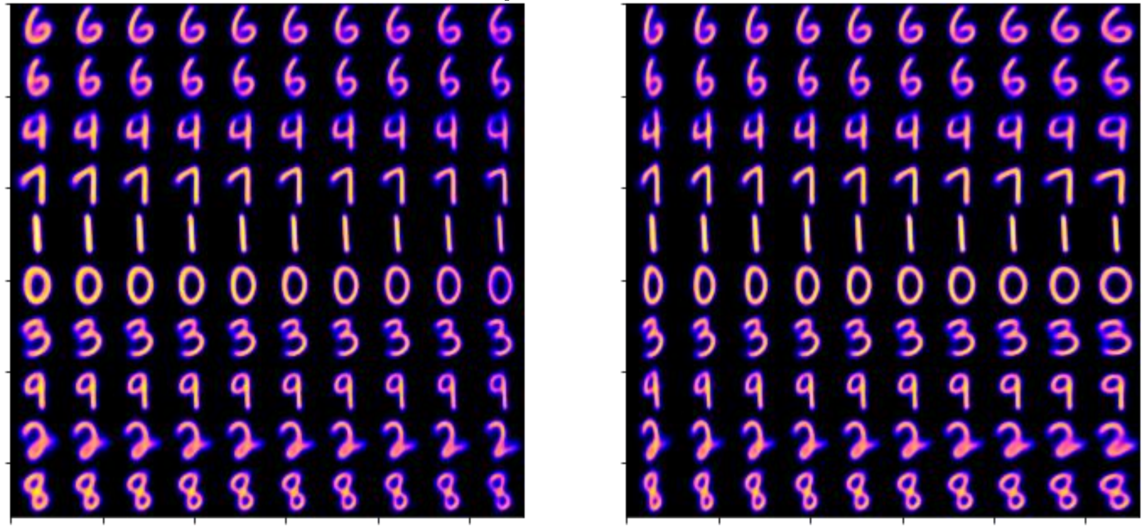
Ensemble jVAE

THE UNIVERSITY of TENNESSEE KNOXVILLE

Predictions from different ensemble models



Mean prediction



Baseline: 10 epochs
Ensemble models: 8

- The unstable classes show the largest "uncertainty"
- Indication of the quality of separation and/or a guide for selection of the number of classes

Dispersion in predictions ('uncertainty')