Interpolation in Auto Encoders with bridge processes

Carl Ringqvist, Judith Butepage, Hedvig Kjellström and Henrik Hult

Introduction

Our suggested method for interpolation can be used in many latent variable model frameworks. We concentrate on the *Variational Auto Encoder* (VAE) model for image reconstruction. The VAE model consists of:

- An encoder $q_\phi(z|x)$, transforming images x to a representation (of lower dimension) z, through a neural network function.
- A prior $p(z)$, enforcing a structure on the latent distribution of data
- A decoder, $p_\theta(x|z)$, transforming a z-representation to an image representation x.

Introduction

Most common interpolation method is *linear interpolation*:

1. Encode $x^{(i)}, x^{(j)}$ through sampling $z^{(i)} \sim q_\phi(z|x^{(i)}), z^{(j)} \sim q_\phi(z|x^{(j)})$
2. Pick points of suitable distance along the line between decoded data points: $[z^{(i)}, z, ..., z^{(j)}]$
3. Decode the latent path by sampling

$$[x^{(i)} \sim p_\theta(x|z^{(i)}), x \sim p_\theta(x|z), ..., x^{(j)} \sim p_\theta(x|z^{(j)})]$$

Problem and idea

- Good samples are generally produced close to data latent representation.
- Many prior structures, especially in higher dimension, and especially the commonly used normal prior, enforces latent data representations with "holes".
- Lines between data points hence often traverse through "empty" areas of the latent space, creating images of low fidelty.
- Many methods have been developed to remedy this problem, notably *spherical interpolation* for higher dimensional normal priors.
- We suggest a novel *stochastic* interpolation scheme, that also address the above problem. It is to our knowledge the first stochastic interpolation method presented. We argue for that stochasticity is preferable and interesting in its own right for some applications.

Diffusion approach

- Our approach starts from the following obsdervation. Given a probability p of the form $p(\mathrm{x}) \propto e^{-E(\mathrm{x})}$ a Langevin diffusion of the form

$$d\mathrm{X} = -\nabla E(\mathrm{X})dt + \sqrt{2}d\mathrm{W} \tag{1}$$

has p as its stationary distribution (under suitable conditions)
- In the context of VAE with $p(x)$ as the prior, the process will reside "close" to latent data representations.

Diffusion approach

- In order to create an interpolation scheme from the stochastic process, we note that the corresponding bridge process, $\mathrm{X}^{x_0 x_T}$ from x_0 to x_T is given by

$$dX^{x_0 x_T} = \Big(-\nabla E(X^{x_0 x_T}) + \sigma\sigma^T \nabla \log p(x_T, T | X^{x_0 x_T}, t)\Big) dt + \sigma dW, \tag{2}$$

- This gives a stochastic interpolation scheme for a completely general prior. However, the term p is hard to calculate for most priors. This can be solved with numerical methods. For some priors, notably the normal distribution, p can be solved for explicitly.

Diffusion approach

If the prior $p(\mathrm{z})$ is an n-dimensional standard normal distribution (e.g the prior in the VAE setting),

$$p(\mathrm{z}) = (2\pi)^{-n/2} e^{-2^{-1}\mathrm{z}\mathrm{z}^T}, \tag{3}$$

it follows that the bridge for the corresponding diffusion process reads

$$\begin{aligned} dZ = \Big[-Z + \frac{2e^{-(T-t)}}{1-e^{-2(T-t)}} \big(z_T - Ze^{-(T-t)} \big) \Big] dt \\ + \sqrt{2} dW, \end{aligned} \tag{4}$$

We use this bridge process for interpolation between two latent data representations, when the VAE prior p is a normal distribution.

Gaussian Process approach

The approach outlined has the advantage of being very general. However, for normal priors, Gaussian processes can be deployed as well. The kernel parameterization of Gaussian processes allows for greater control over the properties of the bridge. For our examples, we use two kernels

$$k(h) = \exp\left\{ -\beta |h|^{\alpha} \right\} \tag{5}$$

$$k(h) = \exp\left\{ -\frac{2}{\ell^2} \sin^2\left(\pi \frac{|h|^2}{p} \right) \right\} \tag{6}$$

Kernel (5) is suitable for strong control over smoothness in image transitions. Kernel (6) allows for a periodic behavior of the interpolation path.

Gaussian Process approach

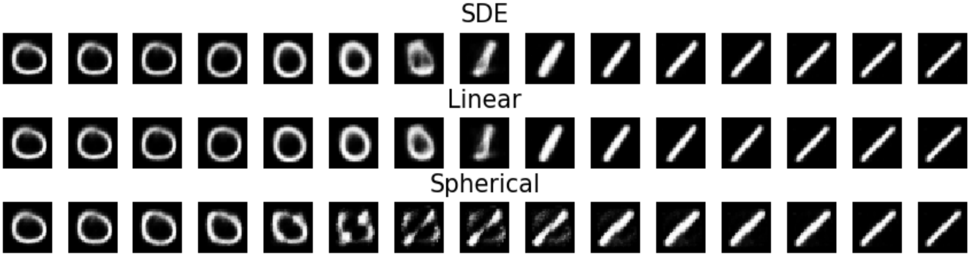
In order to construct an interpolation path with a Gaussian Processes of kernel k, we consider the joint Gaussian distribution of $(Z(0), Z(t_1), \dots, Z(t_m), Z(T))$, conditioning on $(Z(0), Z(T))$. Using the properties of conditional Gaussian distributions we obtain the mean $\hat{\mu}_{z_0,z_T}(t)$ and covariance $\hat{k}(t,s)$ for the bridge process

$$\hat{\mu}_{z_0,z_T}(t) = \frac{z_0[k(t)-k(T-t)k(T)]+z_T[k(T-t)-k(t)k(T)]}{1-k(T)^2} \tag{7}$$

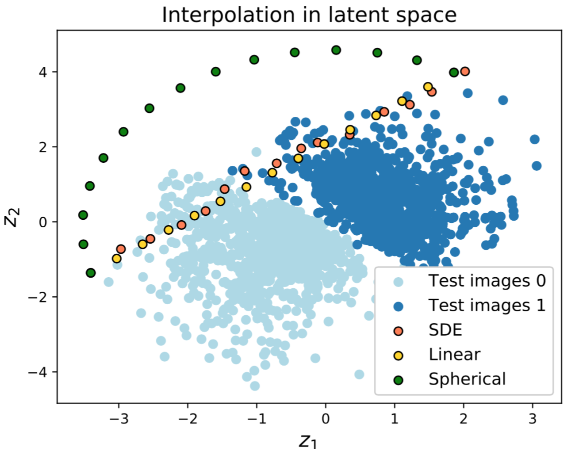
and similar holds for the bridge covariance.

MNIST

For lower dimension, the normal distribution is concentrated around the origin, and linear interpolation generally works well. Here, the commonly used (in higher dimensions) *spherical interpolation* has unsatisfactory results. The picture shows the benefit of our stochastic method in that it can be adjusted so as to reproduce linear interpolation, through shortening the interpolation time T.

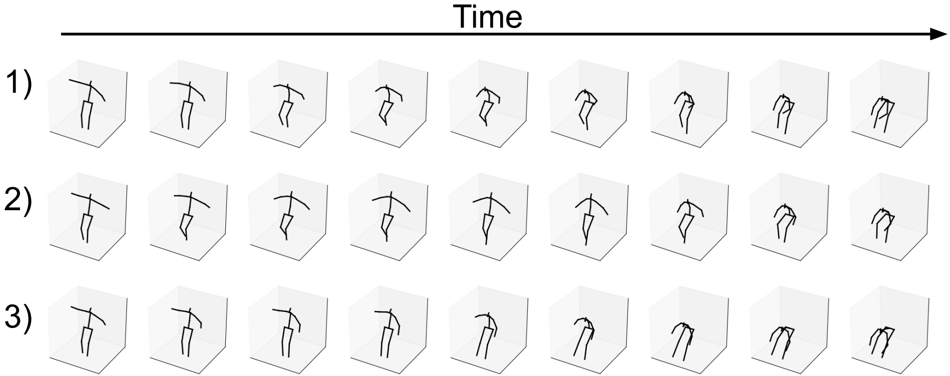


MNIST


Interpolation in latent space
4
2
0
−2
−4
−3
−2
−1
0
1
2
3
z_2
z_1
Test images 0
Test images 1
SDE
Linear
Spherical

HUMAN POSES

In order to test the method over higher dimensions and in order to demonstrate the interesting stochastic property of the method, we deploy the interpolation scheme over the data set *Human Poses*.



HUMAN POSES

The following samples where generated with a Gaussian process bridge between two human pose pictures. The kernel is periodic. Note that for every sample, the start and end point is the same, and that the method hence produces an interesting and plausible variability.

- Sample ”struggling while walking”: `https://youtu.be/BKErJVI2bSA=`
- Sample `"walking with confidence":` `https://youtu.be/94-L5idkwJY=`
- Sample `"avoiding projectile while walking":` `https://youtu.be/oPTB_GrJzPQ=`

U. Cetin and A. Danilova.
Markov Bridges: SDE Representation
Stochastic Processes and their Applications, 126(3):651-679, 2016.

P. Hall, J.S. Marron, and A. Neeman.
Geometric representation of high dimension, low sample size data.
Journal of the Royal Statistical Society: Series B (Stastistical Methodology), 67(3):427-444, 2005.

D. Kingma and M. Welling.
Auto-encoding variational Bayes.
Preprint, arXiv:1312.6114, 2013.

T. White.
Sampling generative network.
Preprint, arXiv:1609.04468, 2016.