GENERATIVE MODELLING

Kaushik Koneripalli
Geometric Media Lab
MS, Electrical Engineering
Arizona State University

PART I

AN INTRODUCTION TO NEURAL NETWORK BASED GENERATIVE MODELS

BACKGROUND

1. One of the long standing goals of Machine Learning is to address the question: **What is the underlying probability distribution of data?**

2. **Many statistical methods** in literature like GMM, NB, etc. However, they are not generalizable and don't extend to more complex data.

3. In 2013, the paper "**Autoencoding Variational Bayes**"[1] was a breakthrough in generative modelling which combined latent variable models with neural network based methods.

4. This was closely followed by "**Generative Adversarial Networks**"[2] in 2015 which was also a neural network based generative model, but trained with a novel adversarial strategy.

5. Let us briefly overview these methods.

VARIATIONAL AUTOENCODERS

VARIATIONAL INFERENCE:

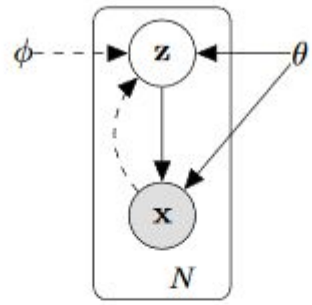


Figure 1: The type of directed graphical model under consideration. Solid lines denote the generative model $p_{\boldsymbol{\theta}}(\mathbf{z})p_{\boldsymbol{\theta}}(\mathbf{x}|\mathbf{z})$, dashed lines denote the variational approximation $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x})$ to the intractable posterior $p_{\boldsymbol{\theta}}(\mathbf{z}|\mathbf{x})$. The variational parameters $\boldsymbol{\phi}$ are learned jointly with the generative model parameters $\boldsymbol{\theta}$.

$$P(\mathbf{Z} \mid \mathbf{X}) = \frac{P(\mathbf{X} \mid \mathbf{Z})P(\mathbf{Z})}{P(\mathbf{X})} = \frac{P(\mathbf{X} \mid \mathbf{Z})P(\mathbf{Z})}{\int_{\mathbf{Z}} P(\mathbf{X}, \mathbf{Z})\, d\mathbf{Z}}$$

However, P(X) is intractable to compute !!!

VARIATIONAL INFERENCE:

Instead, pick an auxillary parametric distribution

$$q_\phi(z|x)$$

and pose the problem as minimizing

$$KL(q_\phi(z|x)||p(z|x))$$

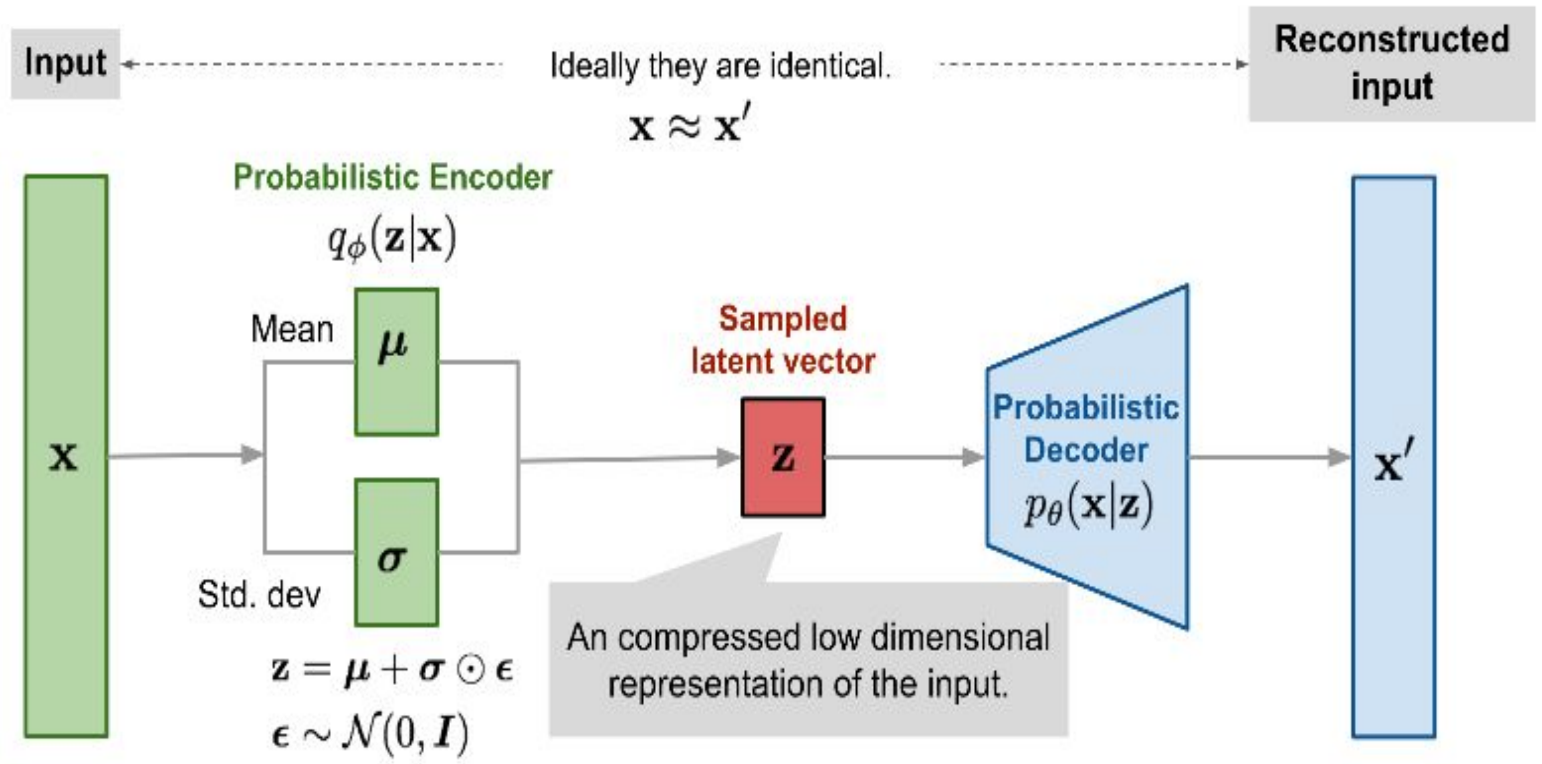
After rearranging terms in the KL divergence, the final equation is as follows:

$$\log p(x) = KL[q_\phi(z|x)||p(z|x)] + ELBO$$

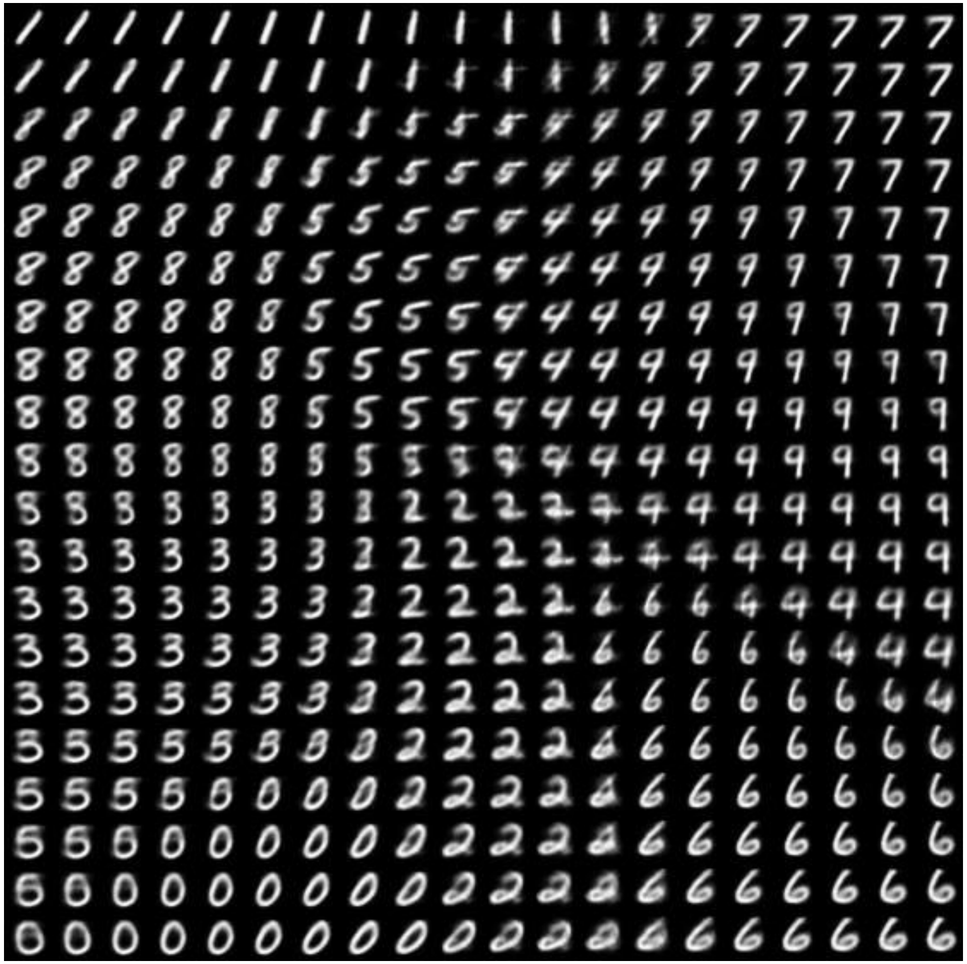
Since for a given x, log p(x) is constant, minimizing the KL divergence is equivalent to maximizing the ELBO.

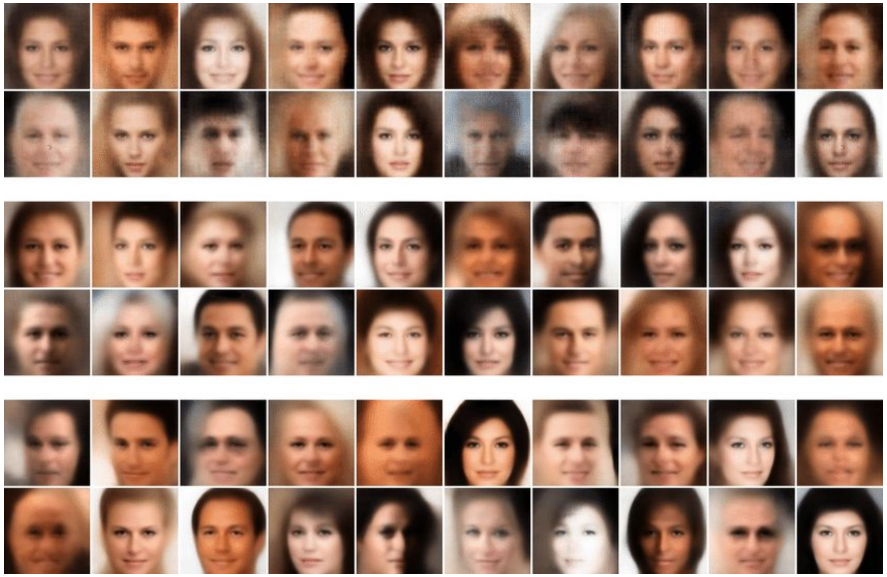
$$ELBO = -KL[q_\phi(z|x)||p(z)] + E_{z\sim q_\phi(z|x)} \log p_\theta(x|z)$$

VARIATIONAL AUTOENCODERS:



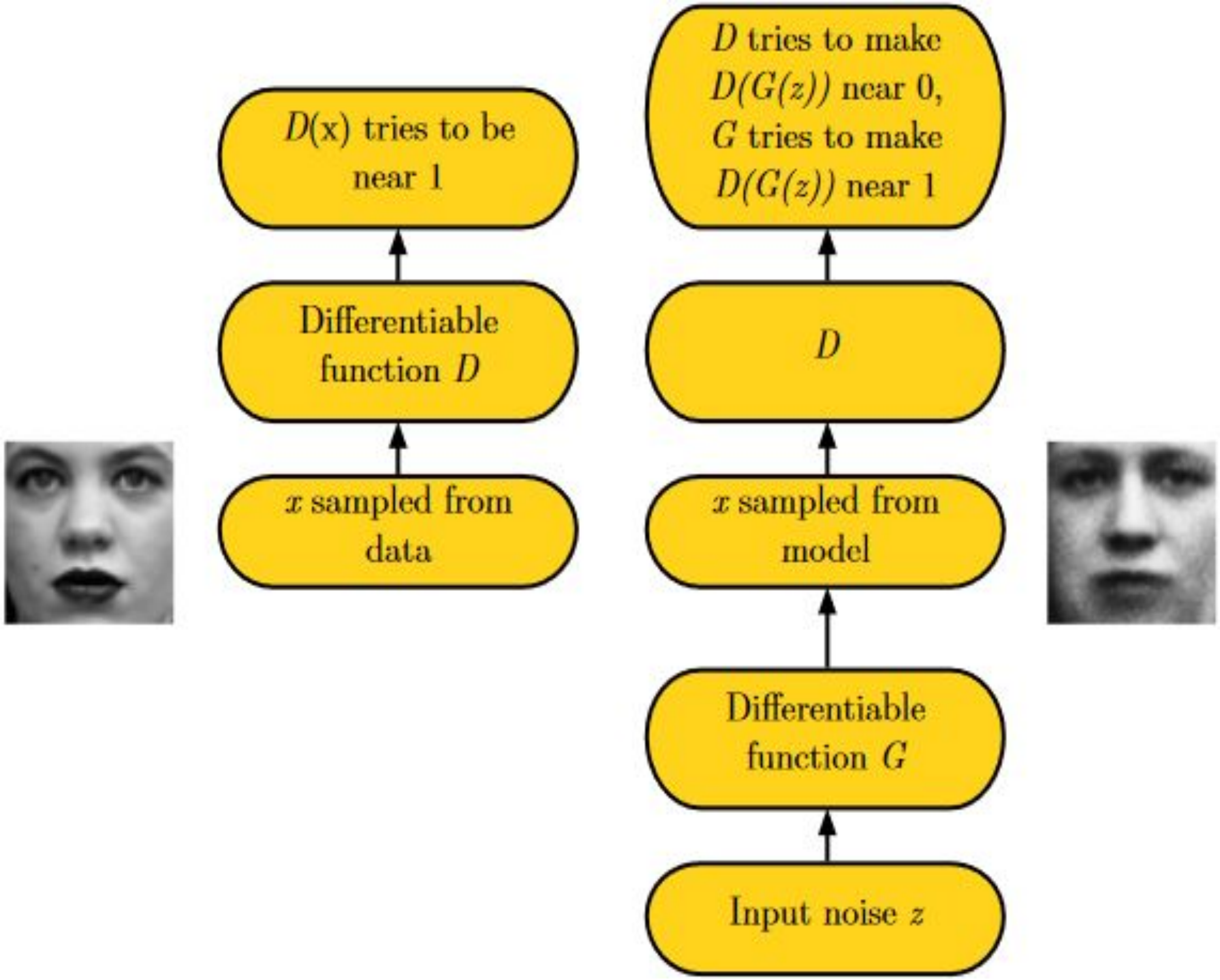
VAE interpolations and reconstructions:



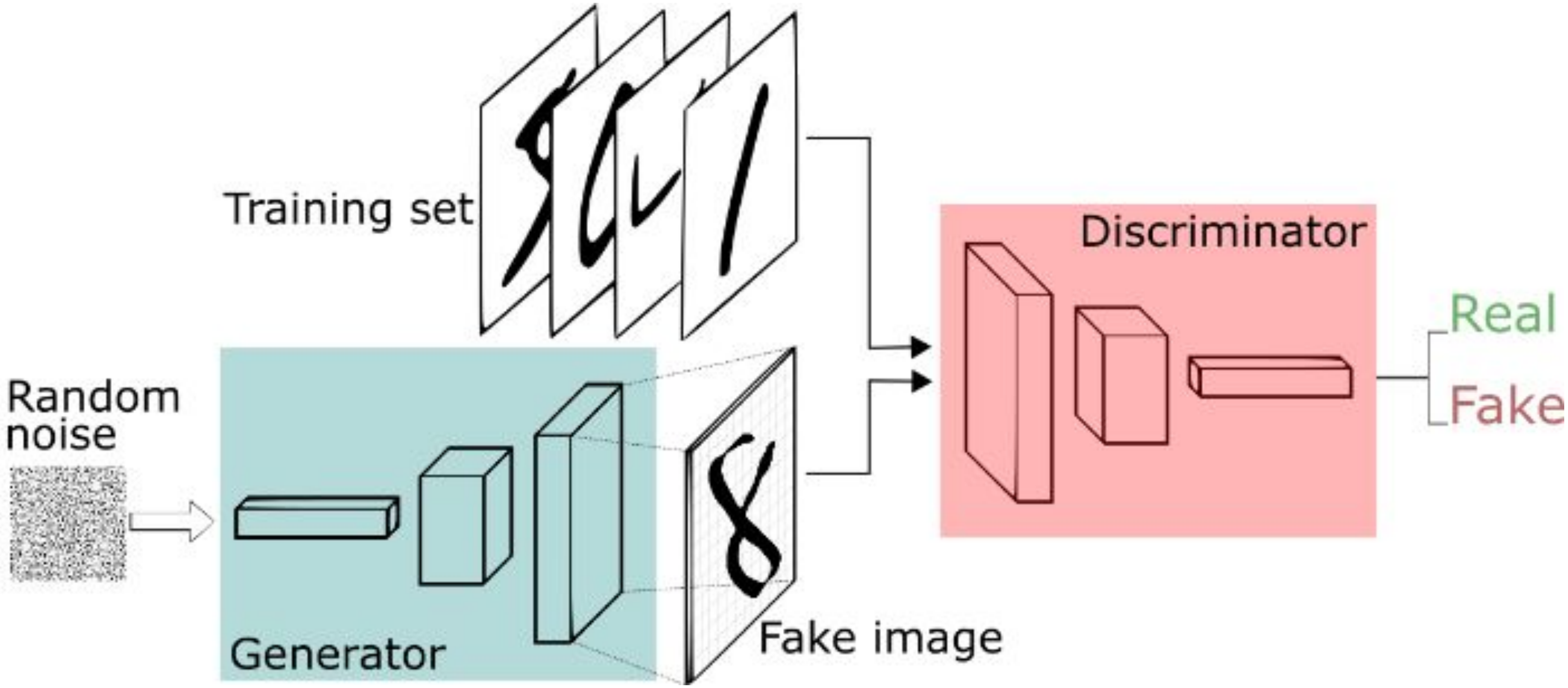


GENERATIVE ADVERSARIAL NETWORKS

GAN MODEL


D(x) tries to be near 1
Differentiable function D
x sampled from data
D tries to make D(G(z)) near 0, G tries to make D(G(z)) near 1
D
x sampled from model
Differentiable function G
Input noise z

GAN MODEL


Training set
Discriminator
Real
Fake
Random noise
Generator
Fake image

GAN Objective dissection

GAN objective function is given by :

$$\min_G \max_D V(G, D) = E_{x \sim p_{data}(x)} \log D(x) + E_{z \sim p_z(z)} \log(1 - D(G(z)))$$

Lets first maximize it w.r.t D. This means, set the derivative to zero:

$$\frac{\partial V(G, D)}{\partial D} = 0$$

We also make an assumption that x = G(z) i.e. in the optimal case, the generated image is from the data distribution. Solving this differential equation yields:

$$D^*(x) = \frac{p_{data}(x)}{p_{data}(x) + p_g(x)}$$

where

$$p_g(x) = p_z(G^{-1}(x))(G^{-1})'(x)$$

On substituting back the optimal D, we can see that the objective function reduces to :

$$KL(p_{data} || \frac{p_g(x) + p_{data}(x)}{2}) + KL(p_g || \frac{p_g(x) + p_{data}(x)}{2}) - \log(4)$$

Minimizing this function basically means that the 2 KL terms go to zero. But this can only happen when data distribution is same as generated image distribution. Thus, GANs provide theoretical guarentees of being able to learn the underlying distribution of data.

GAN generated images

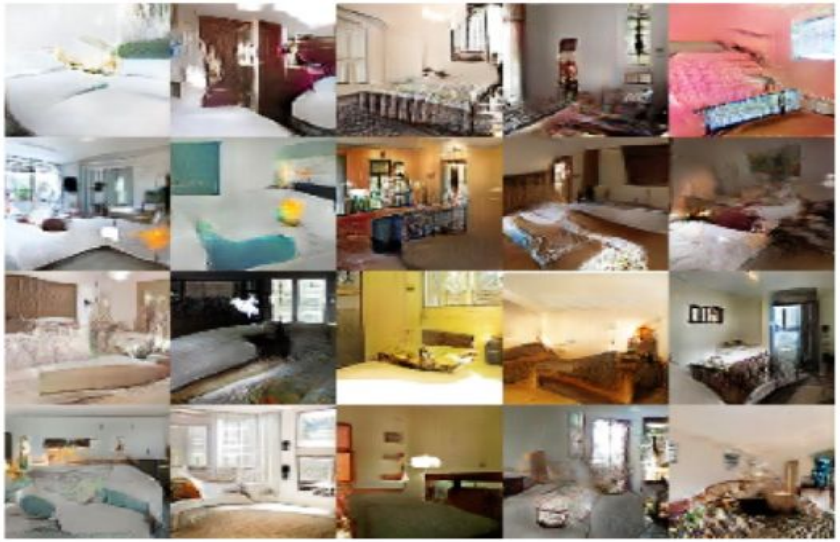


Figure 18: Samples of images of bedrooms generated by a DCGAN trained on the LSUN dataset.

PART II

STRUCTURED DISENTANGLING NETWORKS FOR DEFORMATION INVARIANT LATENT SPACES

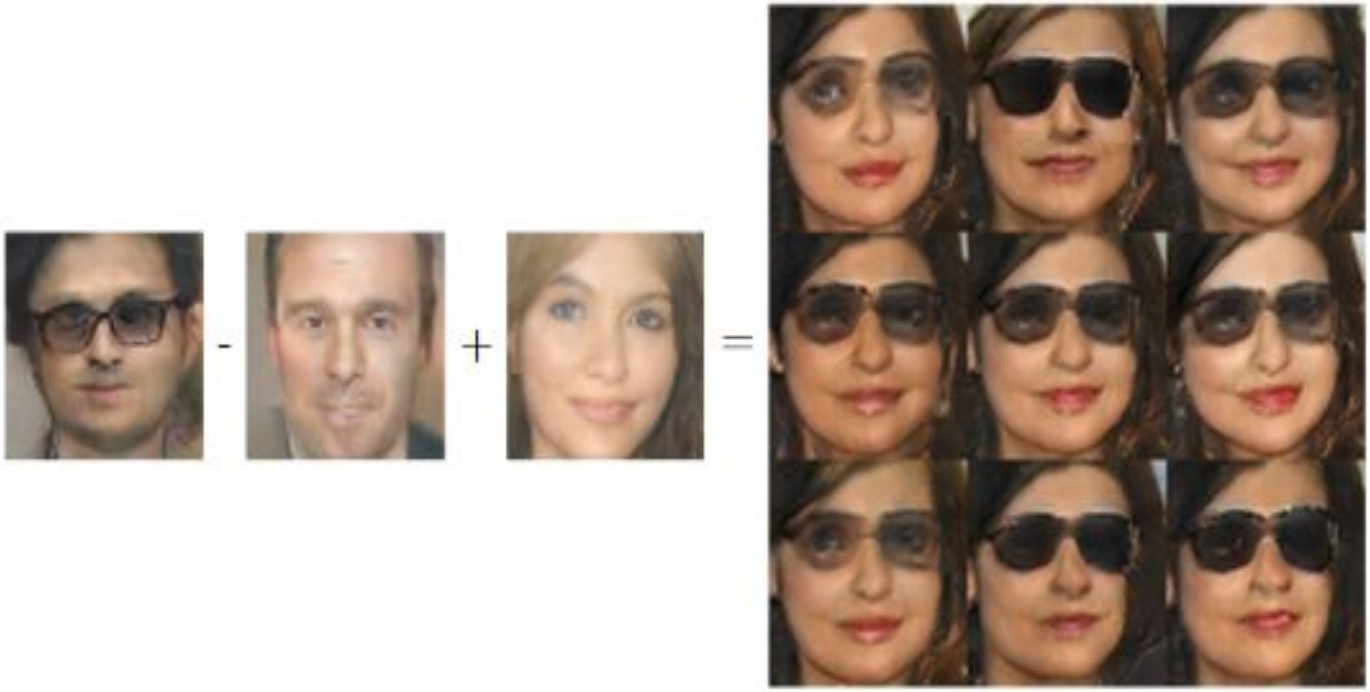
MOTIVATION:

1. With the advent of GANs and VAEs in generative modelling, the next research direction that became very popular is : **How to control the attributes of data when you generate it?**

2. For example, when a face is generated, is it possible to generate/vary only the hair color of the person, keeping all other attributes constant?

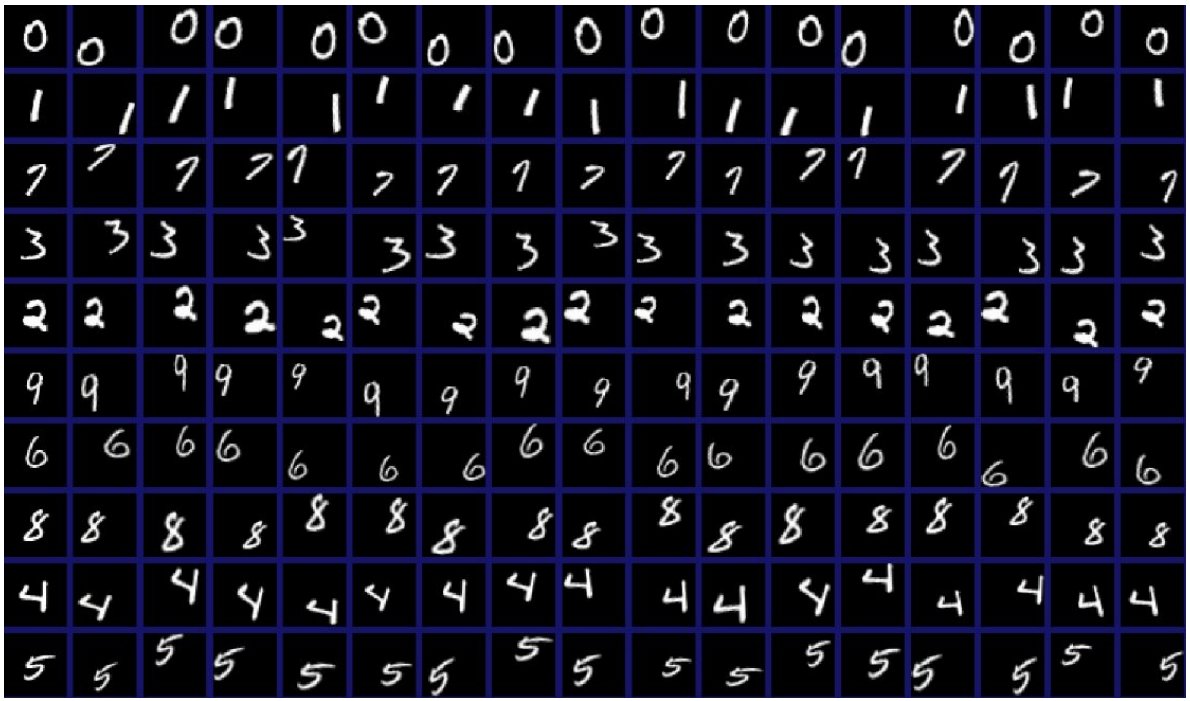
3. This lead to a stream of research popularly known as **“Disentangling the latent spaces”**.

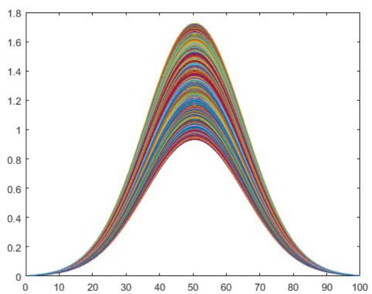
Example of disentangled representations: Doing feature arithmetic

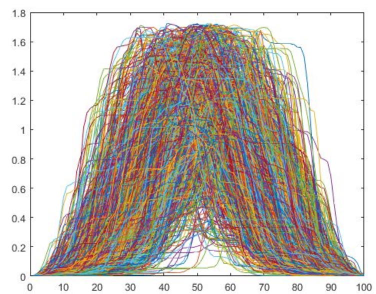


MOTIVATION:

1. Often times, real world data do not come in a "clean" or "unperturbed" format. Nuisance factors typically manifest themselves as affine transforms for images and rate transforms for time series signals.





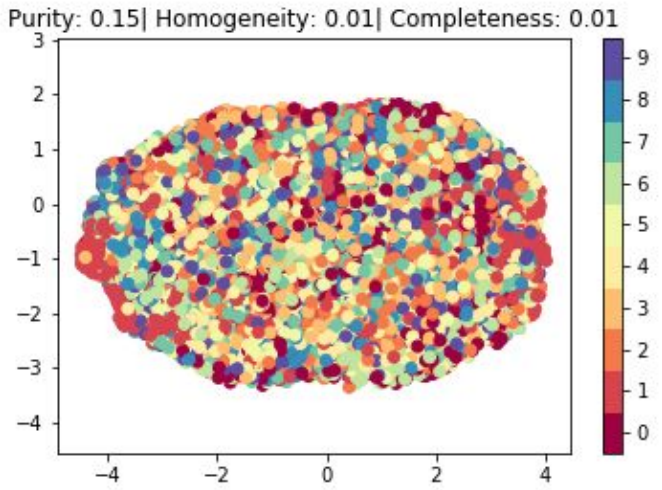


MOTIVATION:

1. These simple transformations can lead to the feature representation being learnt to become indiscriminate w.r.t class information.

2. So, even though the latent space learnt is disentangled, these representations cannot be used for even simple vision tasks like classification.

3. Below, is a visualization of the latent space of a disentangled AE on the Affine MNIST dataset.



PROBLEM STATEMENT:

Is it possible to **disentangle** the latent space of an Autoencoder in an **unsupervised** manner, by **pre-specifying** what attributes the latent variables will represent, whilst maintaining the **class-discriminative** ability?

OUR CONTRIBUTIONS:

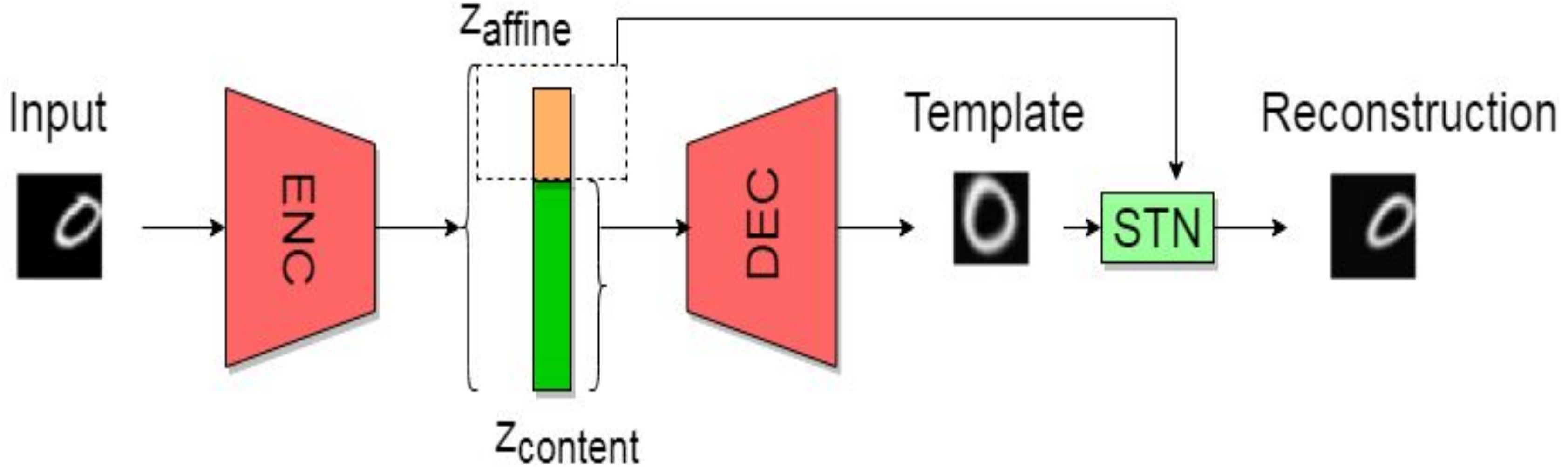
1. We propose a method to ensure **class discriminative-ness** of the latent space for any **disentangling strategy**.

2. The affine/rate transform parameters that are output by the encoder are in the **human-interpretable** transform.

3. **No additional trainable parameters**.

4. By-product of this approach is **template estimation**.

PROPOSED APPROACH



STRATEGIES AND DATASETS CONSIDERED

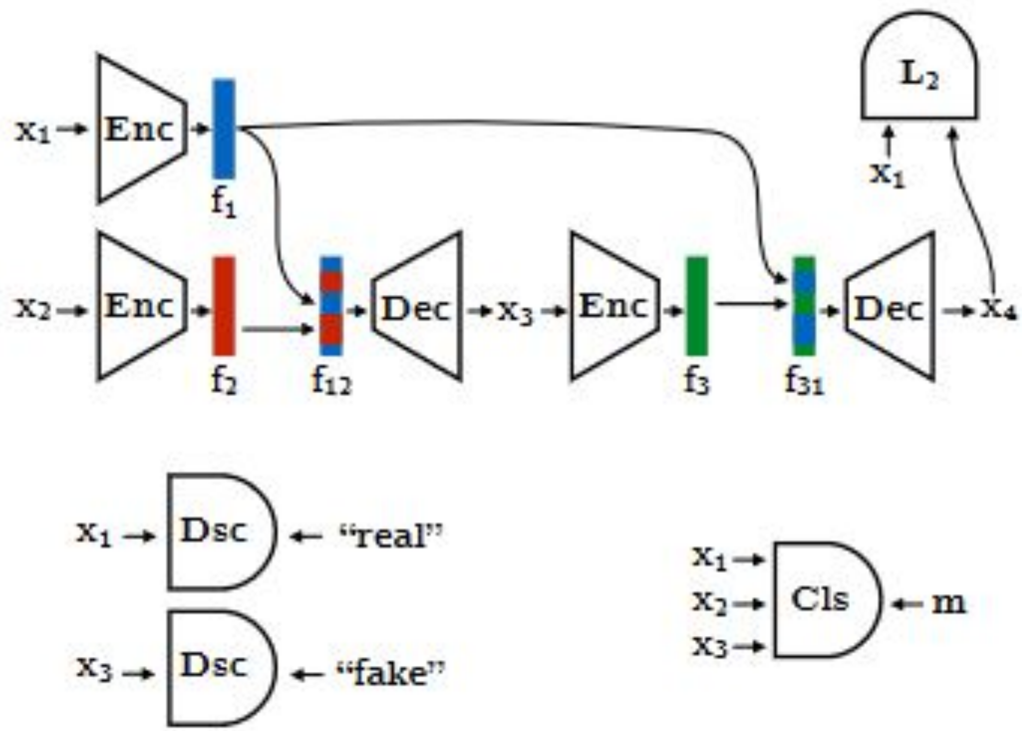
Disentangling strategies considered :

1. Beta-VAE
2. Mixing Autoencoders

Datasets:

1. Affine MNIST
2. Disentangling Sprites (DSprites)

MIXING AUTOENCODERS:



MIXING AUTOENCODERS:

Fig 1. Estimated Affine-Invariant Templates

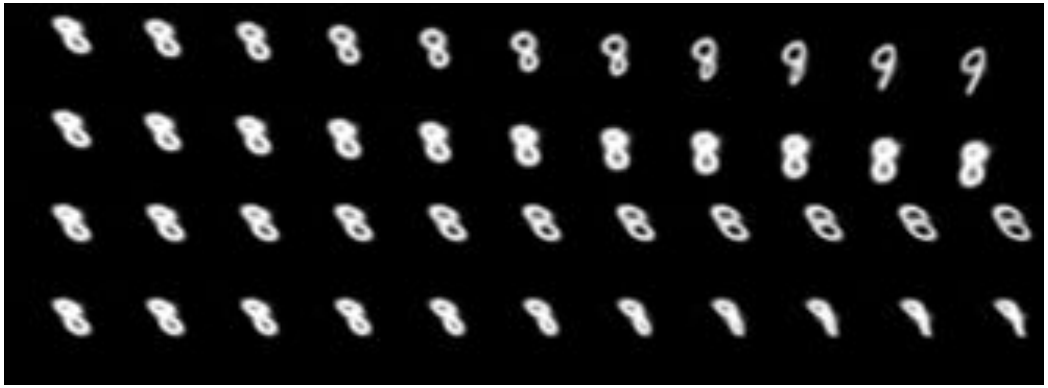


Fig 2. Latent Space Interpolations along chunks

MIXING AUTOENCODERS:

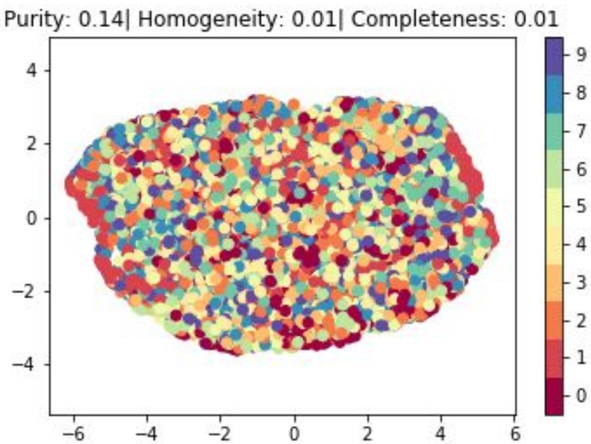


Fig 1. TSNE of latent space of Mixing AE

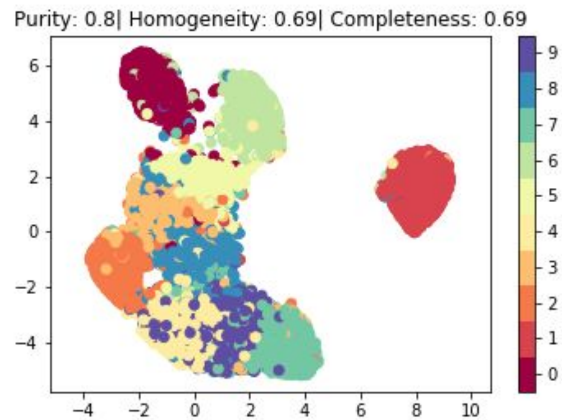
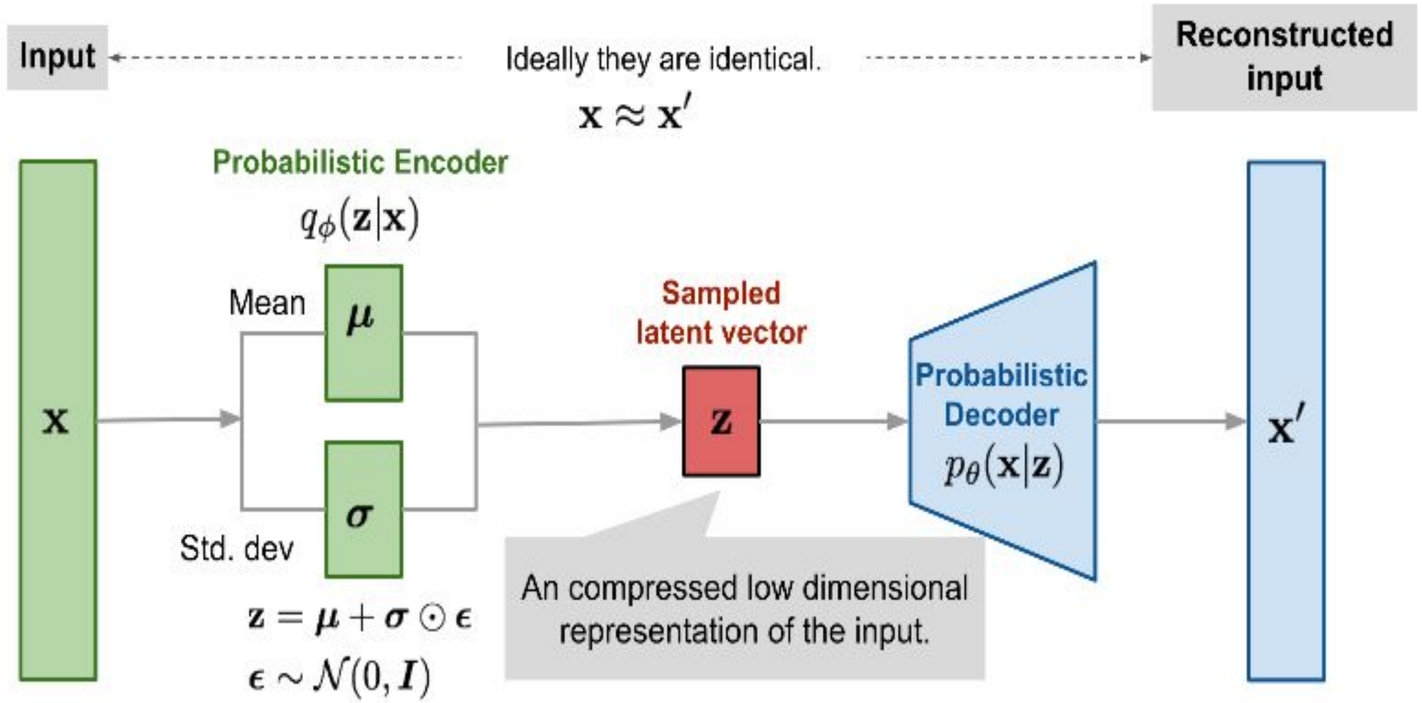


Fig 2. TSNE of latent space of Mixing AE + our training strategy

Beta-VAE:



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

Beta-VAE:

Fig 1. Estimated Affine-Invariant Templates

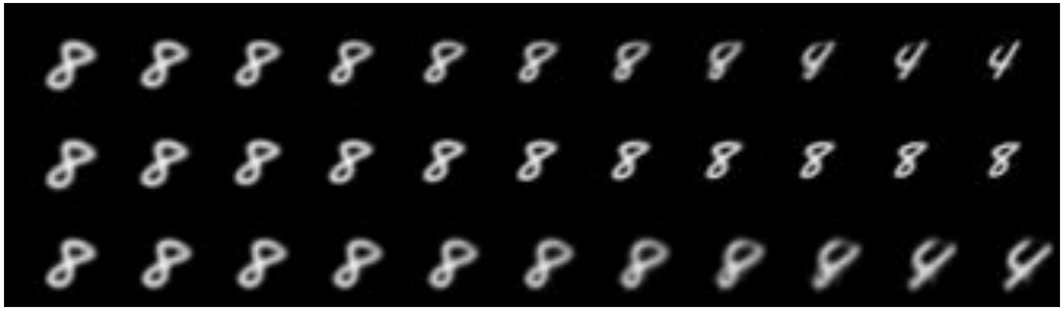


Fig 2. Latent Space Interpolations along chunks

Beta-VAE:

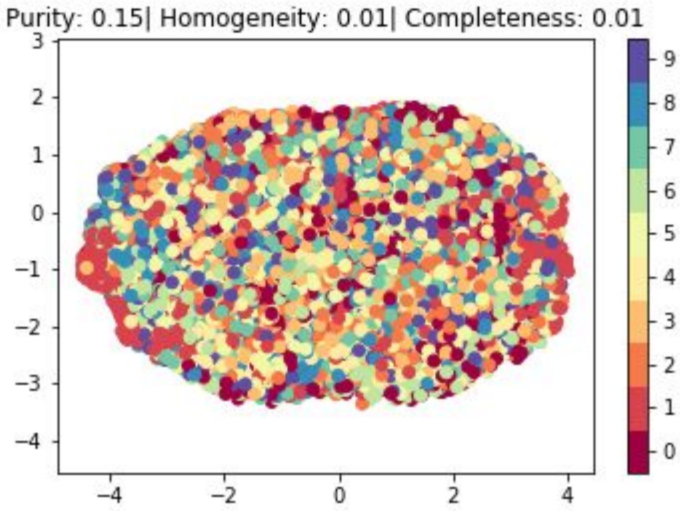


Fig 1. TSNE of latent space of Mixing AE

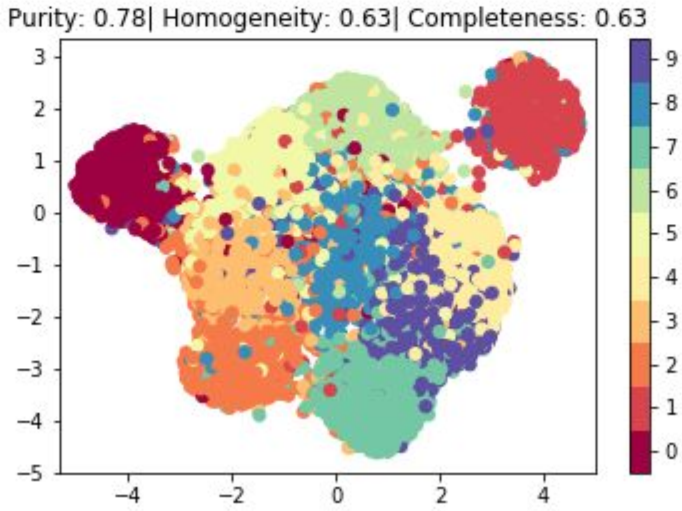


Fig 2. TSNE of latent space of Mixing AE + our training strategy

EXTENSION TO TIME SERIES:

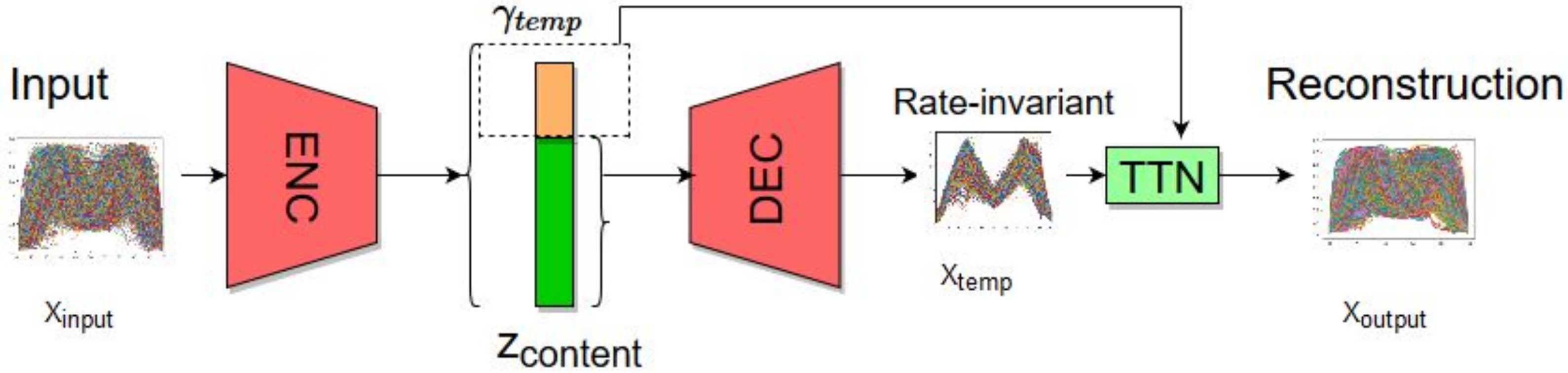


Fig 1. TCN based AE architecture to achieve rate-invariant representations

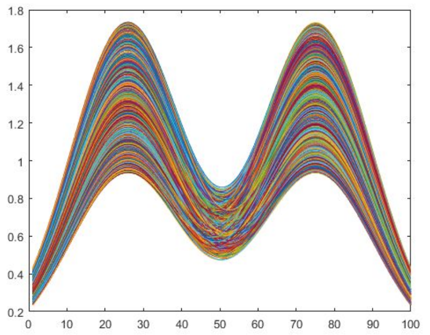


Fig 2. Original Sequence

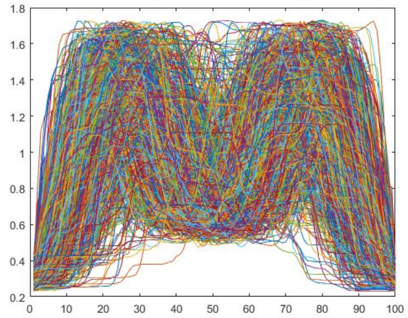


Fig 3. Warped Sequence

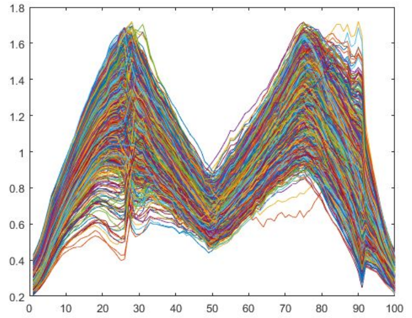


Fig 4. Rate-Invariant sequence predicted by AE

CONCLUSIONS

1. In this work, we have seen how to achieve affine invariant latent spaces in an unsupervised manner and also achieving class-discriminability as a by-product.

2. We also saw that this novel training strategy can be extended to time-series signals to achieve rate invariant representations of the signal.

References:

[1] Kingma, Diederik P., and Max Welling. "**Auto-encoding variational bayes**." *arXiv preprint arXiv:1312.6114* (2013).

[2] Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "**Generative adversarial nets**." In *Advances in neural information processing systems*, pp. 2672-2680. 2014.

[3] Hu, Qiyang, Attila Szabó, Tiziano Portenier, Paolo Favaro, and Matthias Zwicker. "**Disentangling factors of variation by mixing them.**" In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3399-3407. 2018.

[4] Higgins, Irina, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. "**beta-vae: Learning basic visual concepts with a constrained variational framework.**" In *International Conference on Learning Representations*, vol. 3. 2017.

[5] AffNIST dataset - https://www.cs.toronto.edu/~tijmen/affNIST/

[6] DSprites dataset - https://github.com/deepmind/dsprites-dataset.

QUESTIONS?