Generative ConvNet

with Continuous Latent Factors

Stu. Jerry Xu **Mentor: Prof. Ying Nian Wu**

Shanghai Jiao Tong University / University of California, Los Angeles

UCLA CSST --- Aug. 31st , 2016

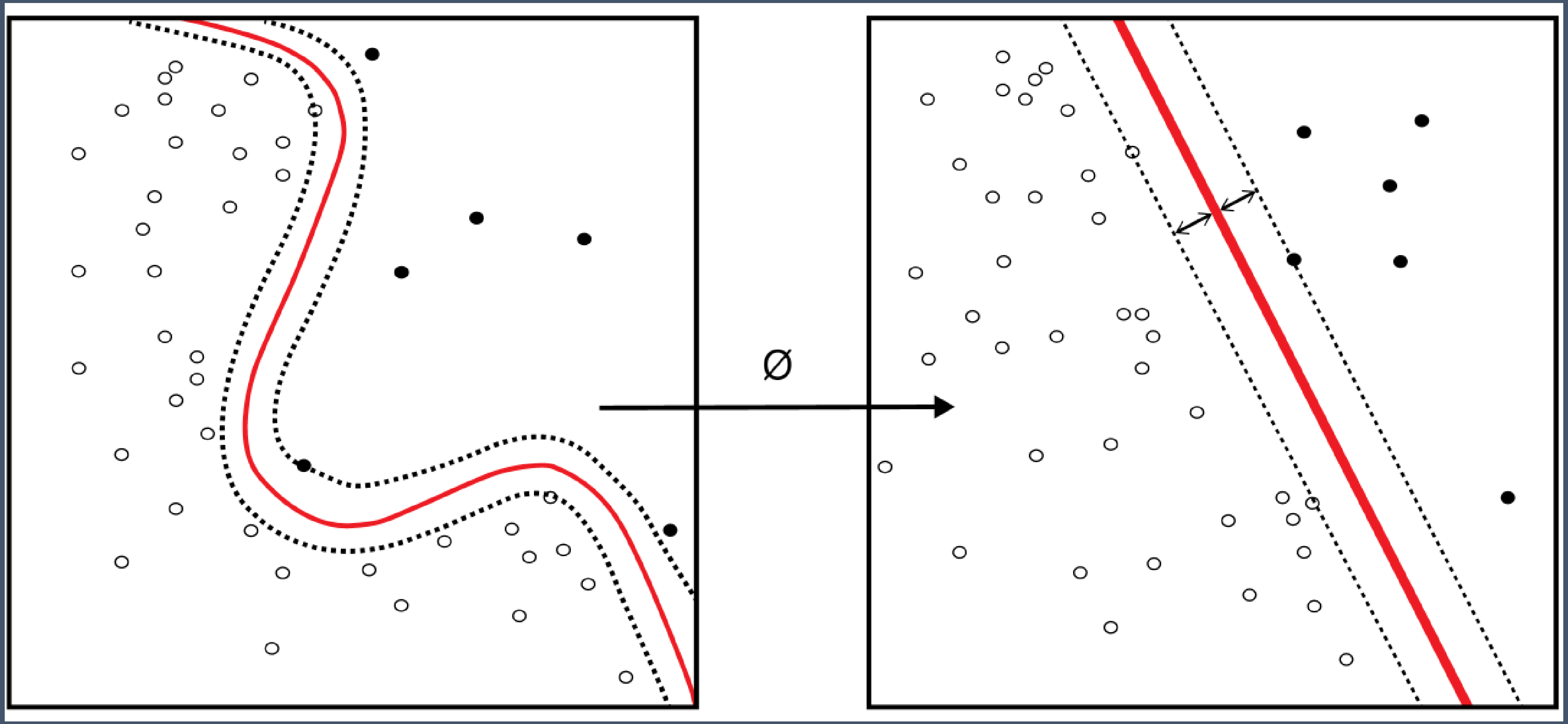
AI & Machine Learning
Introduction

Face Recognition

Anomaly Detection

AI & Machine Learning

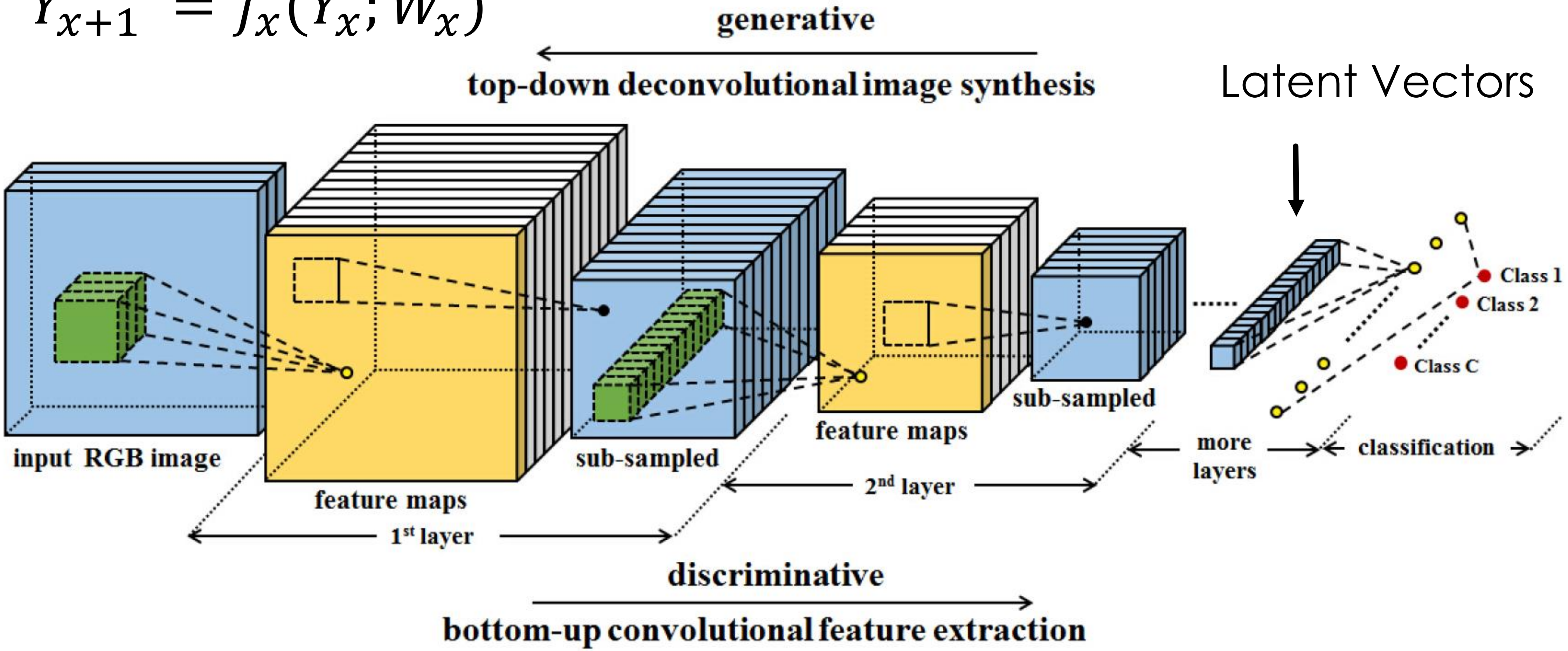
Introduction



Discriminating and generating
Introduction

$$Y_{x+1} = f_x(Y_x; W_x)$$

Latent Vectors

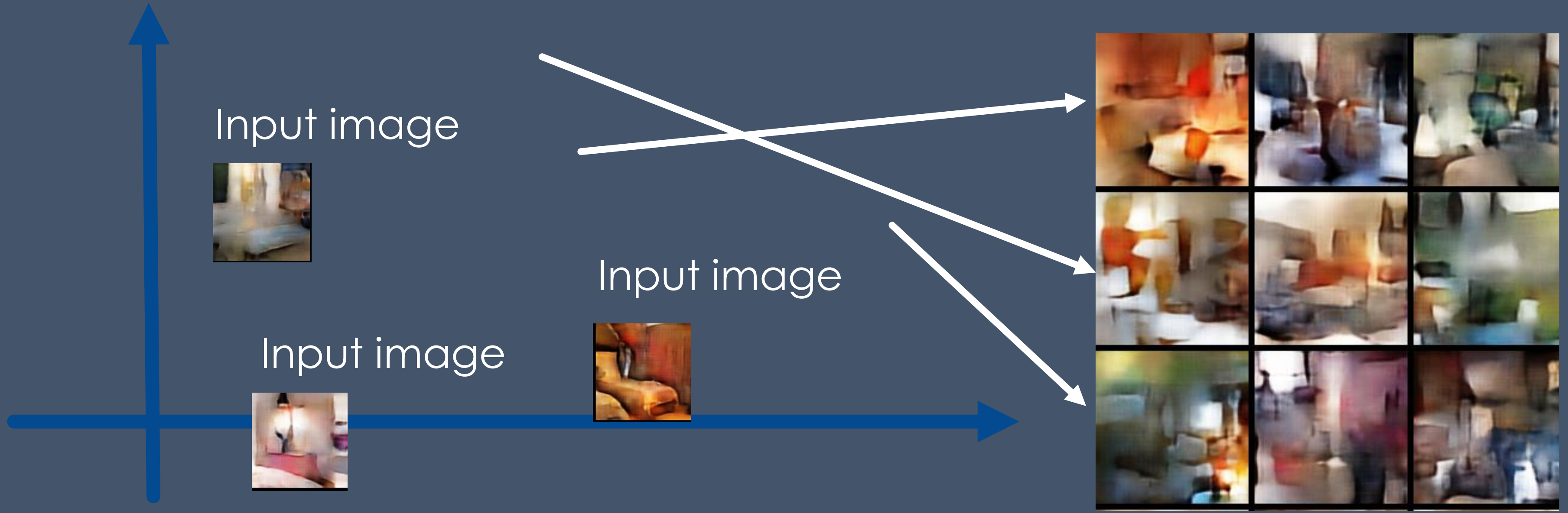


Objective

Introduction

- A Model truly understand the data.
- Find a significative mapping from data to latent vector.
- On the other hand, every latent vector can reconstruct data.

$$I \rightarrow Z$$



1 Technic detail of the model

2 The evaluation

3 Scaling up

Outline

Generative ConvNet
with Continuous Latent Factors

$I \in R^{10000} \quad W \in R^{20\times 10000} \quad Z \in R^{20}$

$$I = WZ + \epsilon$$

Basic Linear Model

Factor analysis / principal component analysis

Map @ Generative ConvNet
Introduction

$I = WZ + \epsilon$ $\rightarrow$ $I = f(Z; W) + \epsilon$

Factor analysis Model
Linear; One-layer

Dimension reduction

Dictionary learning

Latent factor extracting

e.g.
(PCA)Principal component analysis

(NMF)Non-negative matrix factorization

(ICA)Independent component analysis

Generlization

Our Model
Non-linear; Muiti-layer

More explicitly

horizontal unfolding (Convolutional)

hieratical unfolding

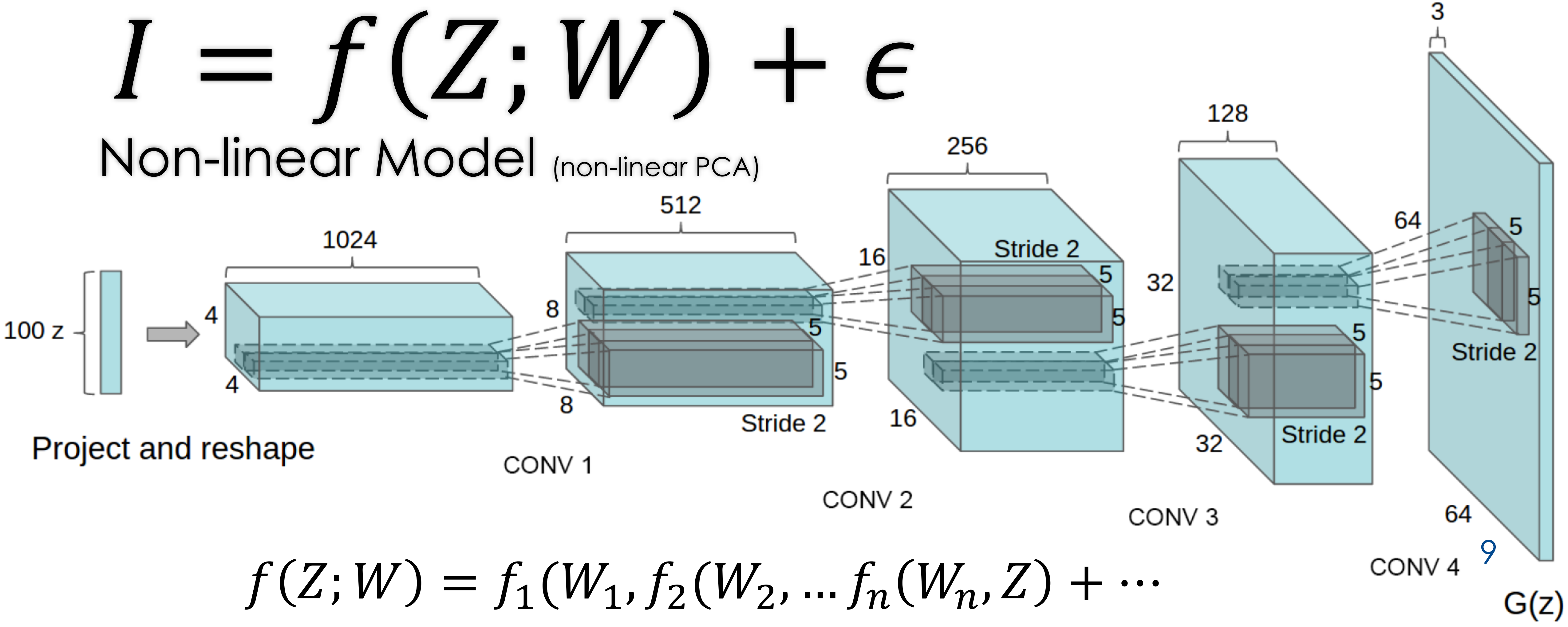
$f(Z; W)$:

Parameterized by ConvNet

Generative ConvNet
with Continuous Latent Factors

$$I = f(Z; W) + \epsilon$$

Non-linear Model (non-linear PCA)



$$f(Z; W) = f_1(W_1, f_2(W_2, \ldots f_n(W_n, Z) + \cdots$$

Generative ConvNet
with Continuous Latent Factors

Reconstruction Error

$$err = \left|\left|Y - f(Z;W)\right|\right|_2^2$$

Likelihood for data {Y}

$$L(W,\{Z_i\}) = \sum_i^n \log p(Y_i, Z_i; W) = -\sum_i^n \left[\frac{\| Y - f(Z_i;W) \|^2}{2\sigma^2} + \frac{\| Z_i \|^2}{2}\right] + constant$$

We need to

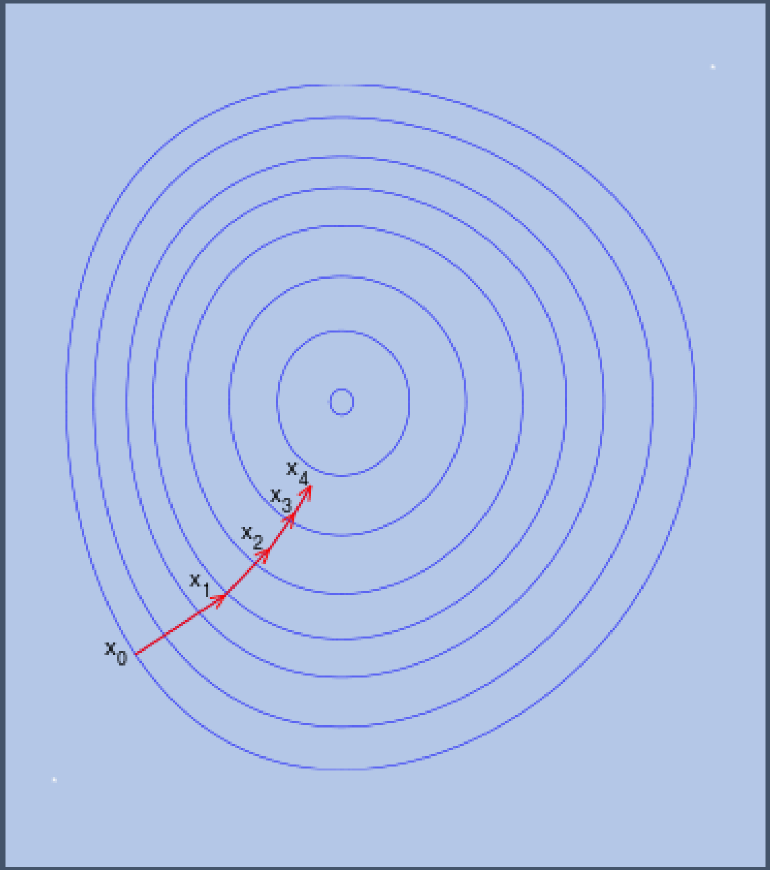
$$\max_W L(W,\{Z_i\})$$

Basic Gradient Descent

On training Generative ConvNet

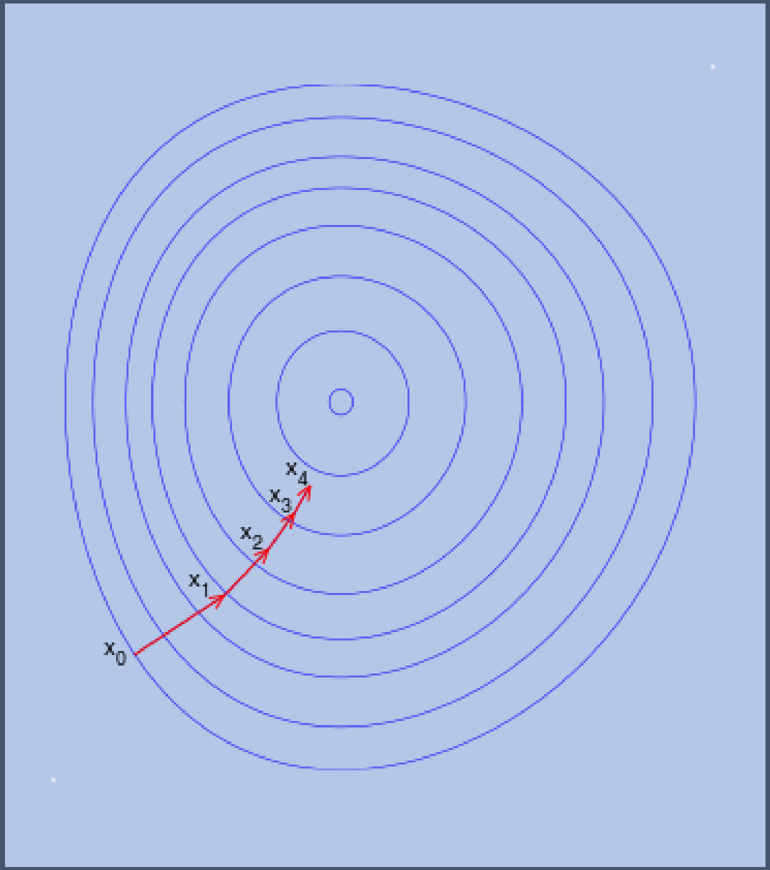
$$\frac{\partial L}{\partial Z_i} = \frac{1}{\sigma^2}\left(Y_i - f(Z_i, W)\right)\frac{\partial f}{\partial Z_i} - Z_i$$

$$\frac{\partial L}{\partial W} = \sum_{i=1}^{n} \frac{1}{\sigma^2}\left(Y_i - f(Z_i, W)\right)\frac{\partial f}{\partial W}$$



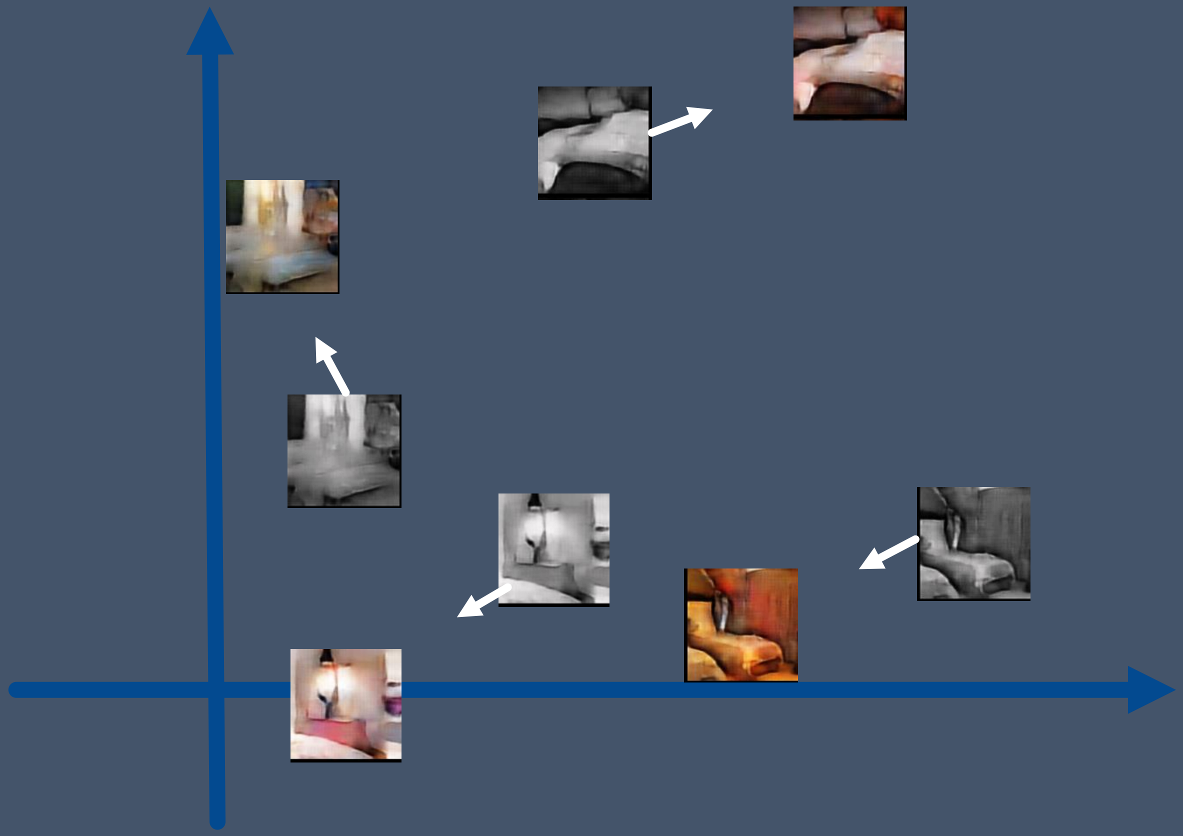
Alternating Gradient Descent
On training Generative ConvNet

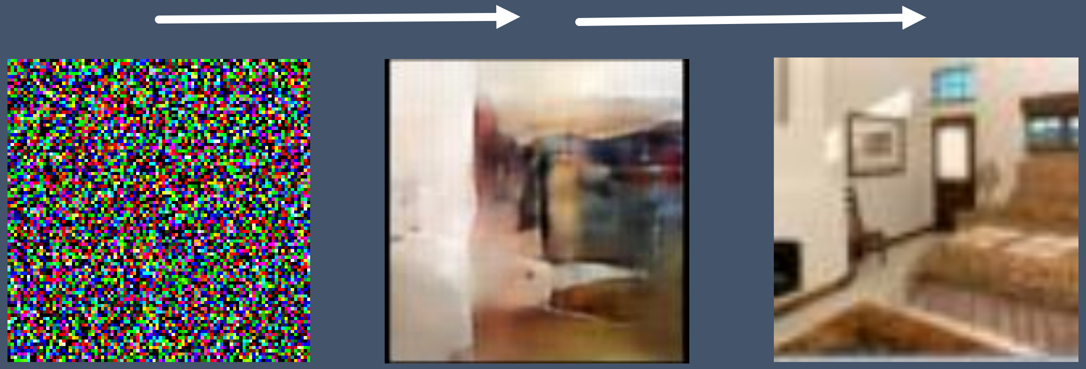
- (1) Inferential back-propagation: For each I, run L steps of gradient descent to update $Z_i \leftarrow Z_i + \eta \partial L / \partial Z_i$
- (2) Learning back-propagation: Update $W \leftarrow W + \eta \partial L / \partial W$



Alternating Gradient Descent

On training Generative ConvNet





Langevin Sampling

On training Generative ConvNet

- maximizing the observed data log-likelihood $L(W) = \sum_{i=1}^{n} \log p(Y_i; W) = \sum_{i=1}^{n} \log \int p(Y_i, Z_i; W) dZ_i$
- $Z_{i+1} = Z_i + \frac{\Delta^2}{2}\left[\frac{1}{\sigma^2}\left(Y - f(Z_i; W)\right)\frac{\partial f}{\partial W} - Z_i\right] + \Delta\epsilon$

Comparison

On training Generative ConvNet

gradient decent

Langevin Sampling

1 by Reconstruction Error

2 by Testing and Negative

3 Future Work

Model Evaluation

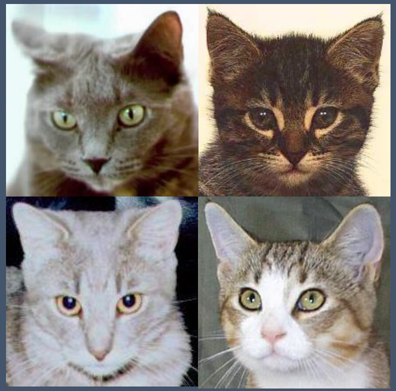
Model Evaluation

By reconstruction error

Reconstruction Error:

$$err = ||Y - f(Z;W)||_2^2$$

Test



Train

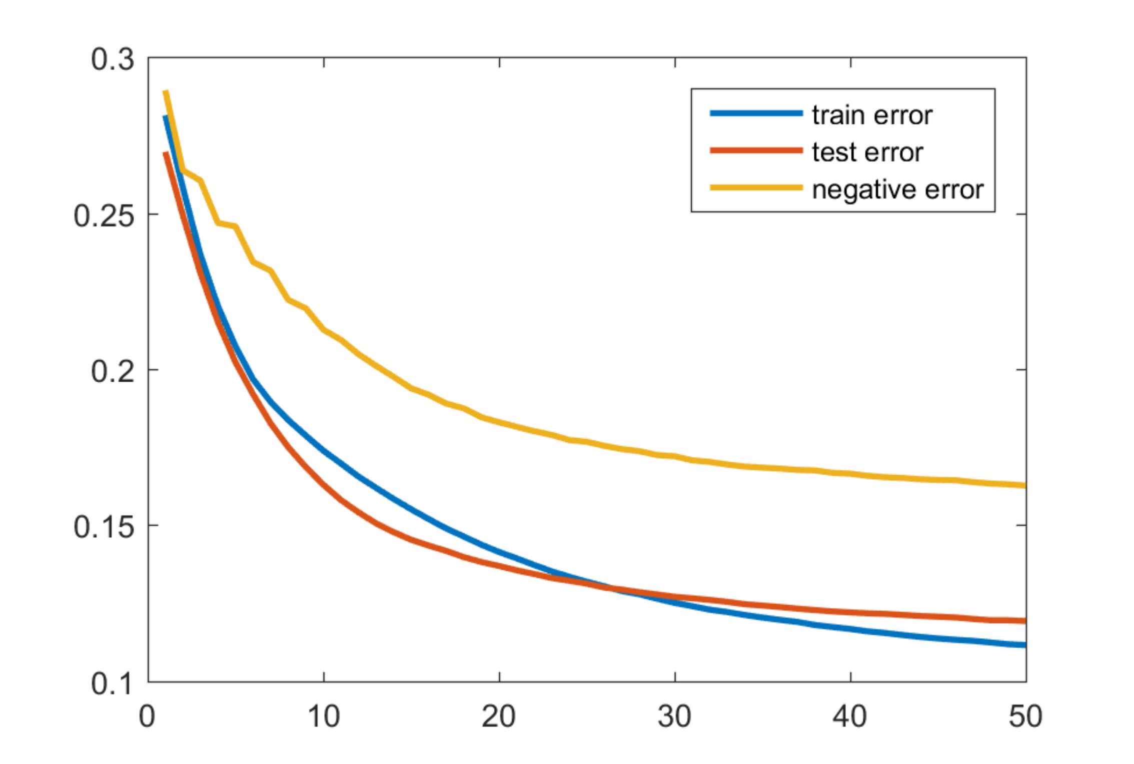
Negative

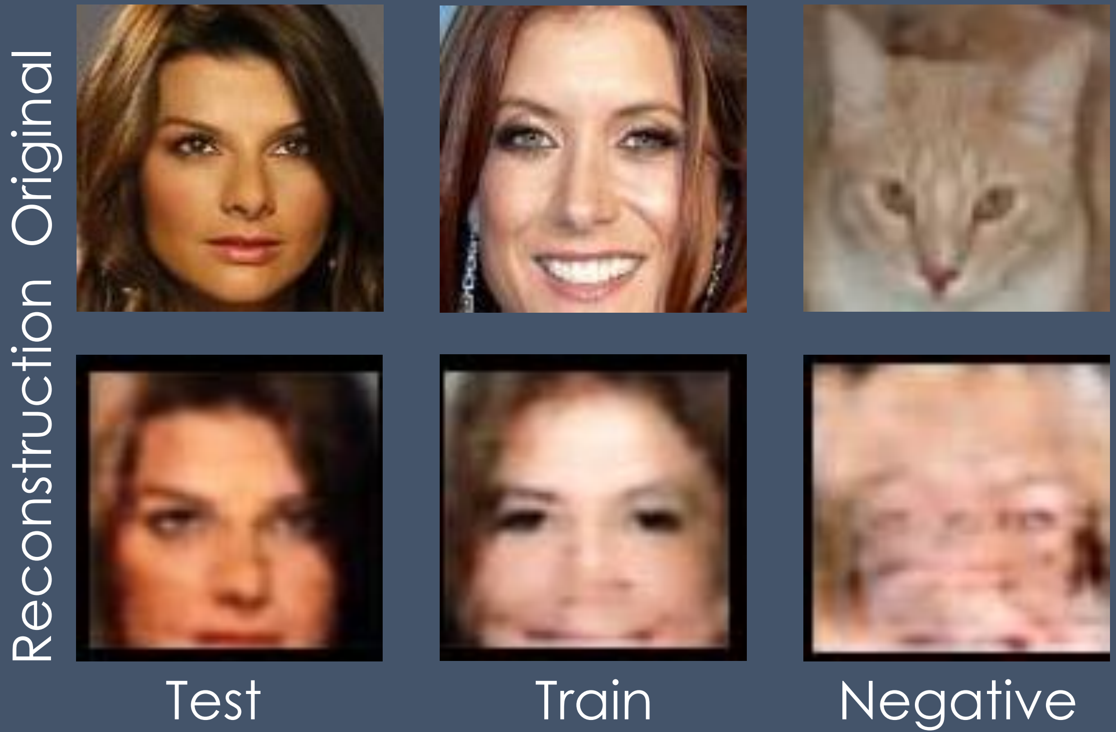
Model Evaluation

By reconstruction error

1. Fix Weights, use Langevin Sampling to Infer the corresponding z.
2. Go through the network, use z to reconstruct the image.
3. Calculate the reconstruct error between reconstructed image and original image.

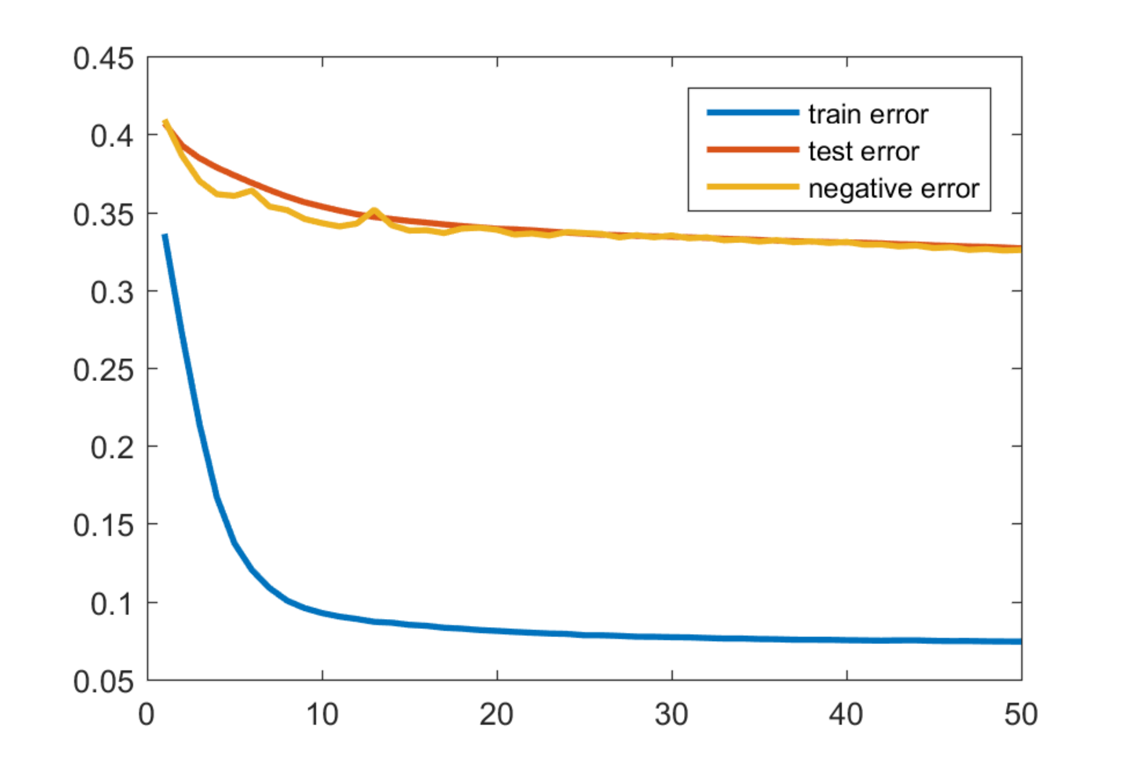
Sample reconstruction Error



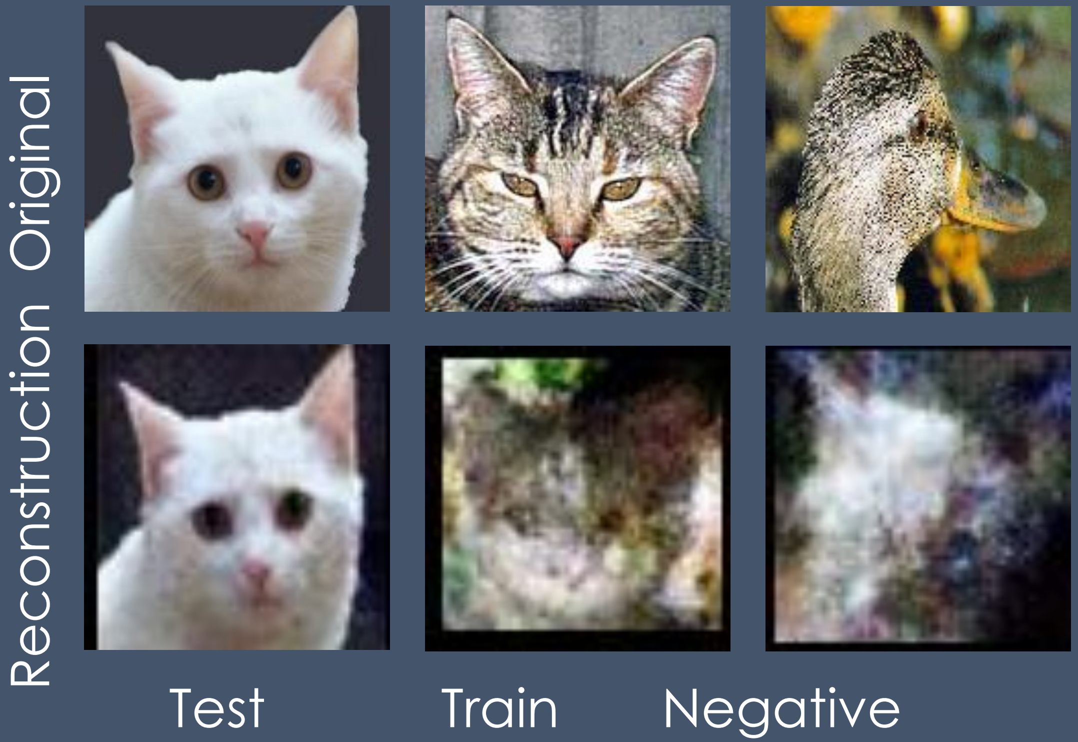


Trainging on 10000 face images --- Good result

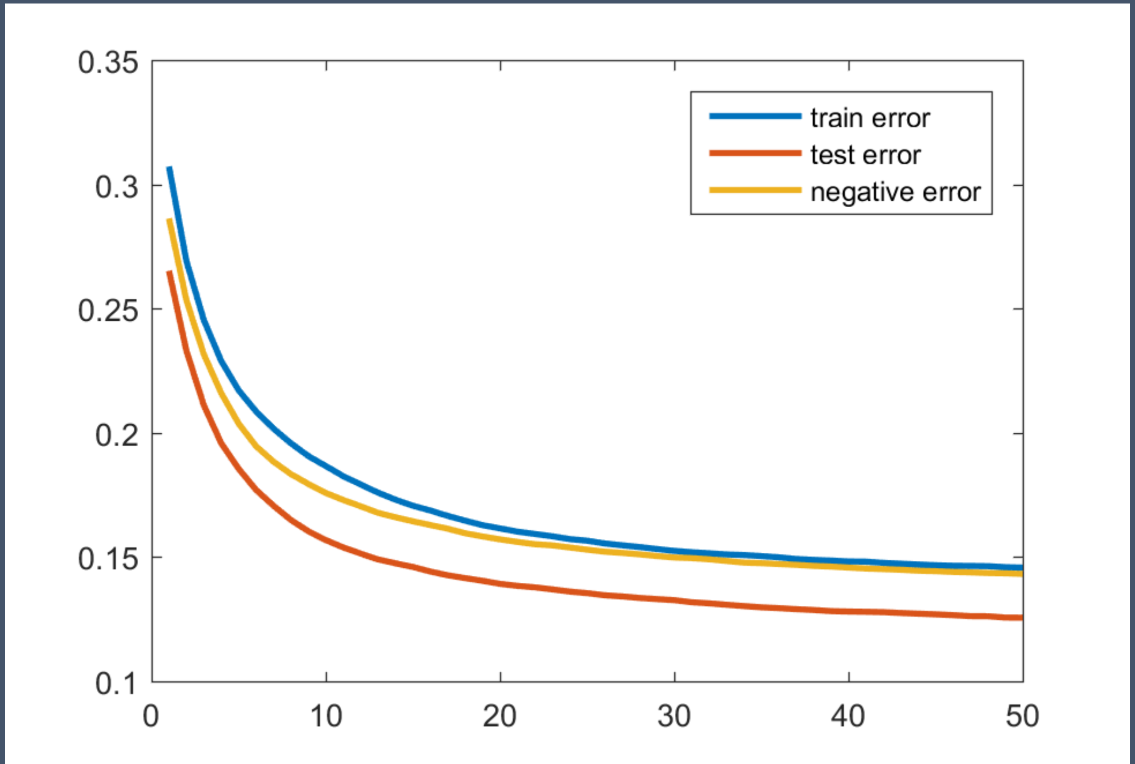
Sample reconstruction Error



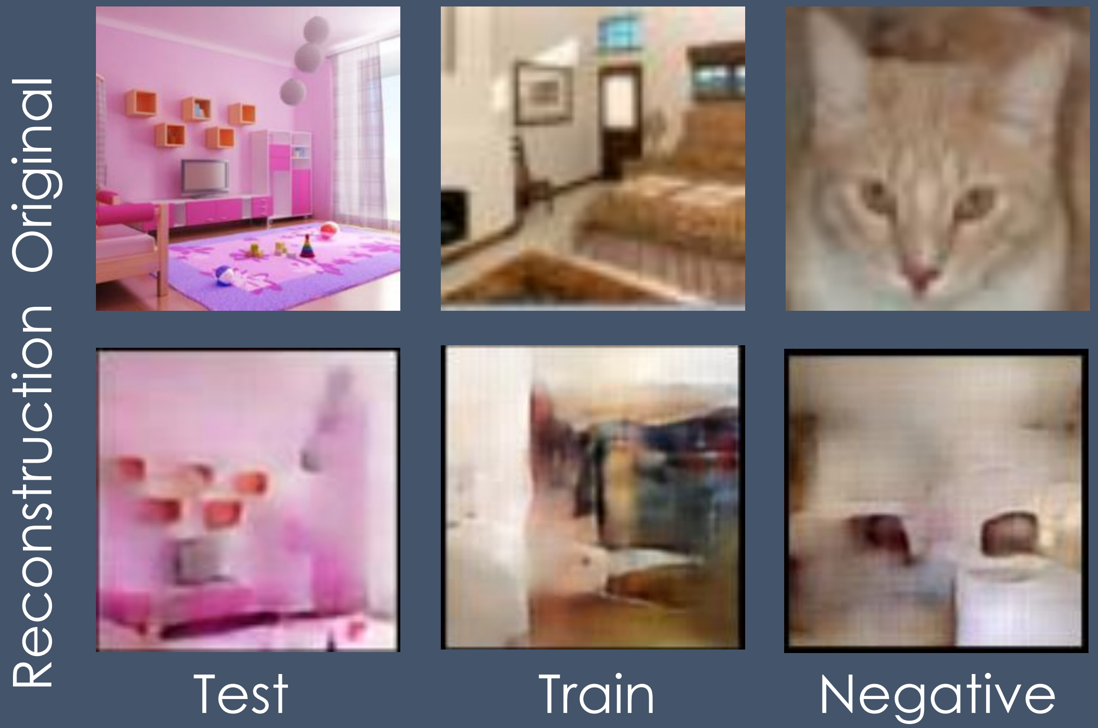
Trainging on 100 cat images --- Overfitting



Sample reconstruction Error



Trainging on 20000 face images --- Underfitting



1 What and Way

2 Way to make it better

3 Results

Scaling up and Discovery

Why we need scaling up

- ‘Not Bad’ result on small data
- ‘Nonsenses’ result on small data

Way to make it better

- Turning Configuration
  - Dimension of Z
  - Learning rate
  - Langavin Step/Size
  - Momentum
- Modifying Net Structure
  - Number of FC layers
  - Add Conv layers
  - Modify # of channels
  - Size of output

Dimension of Z

Result on turing configuration

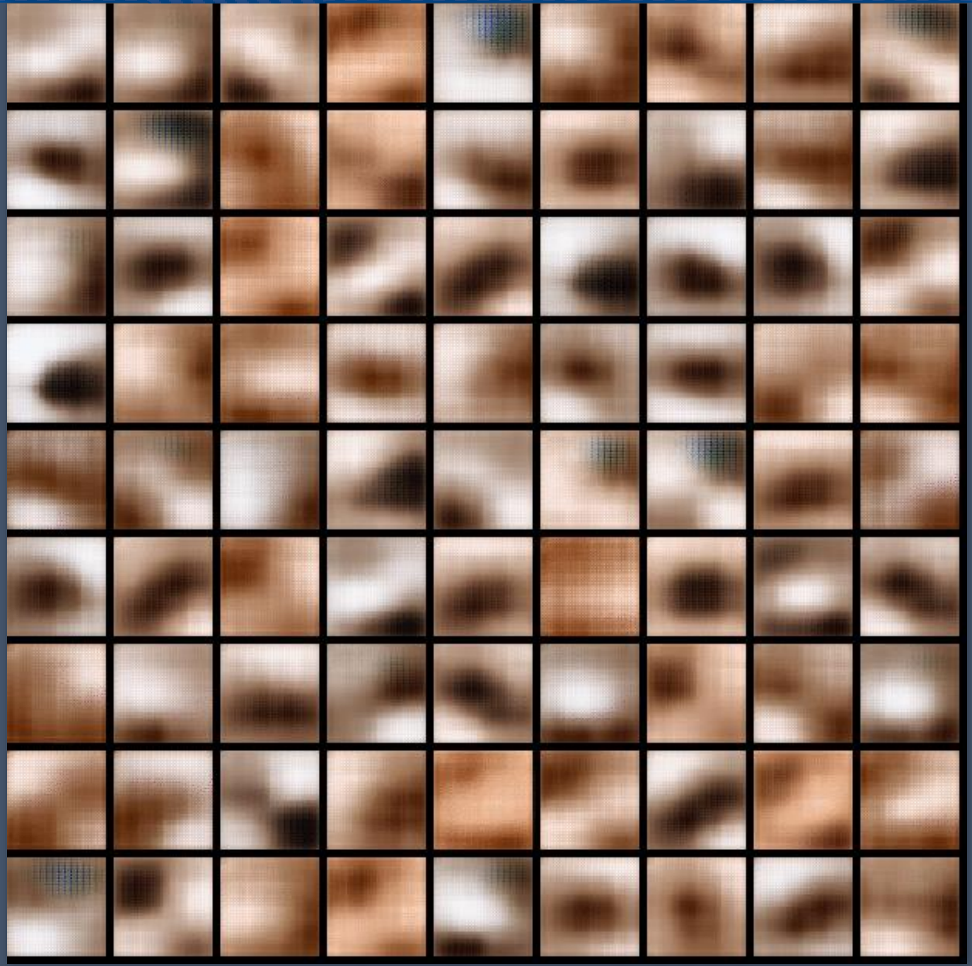
2

25

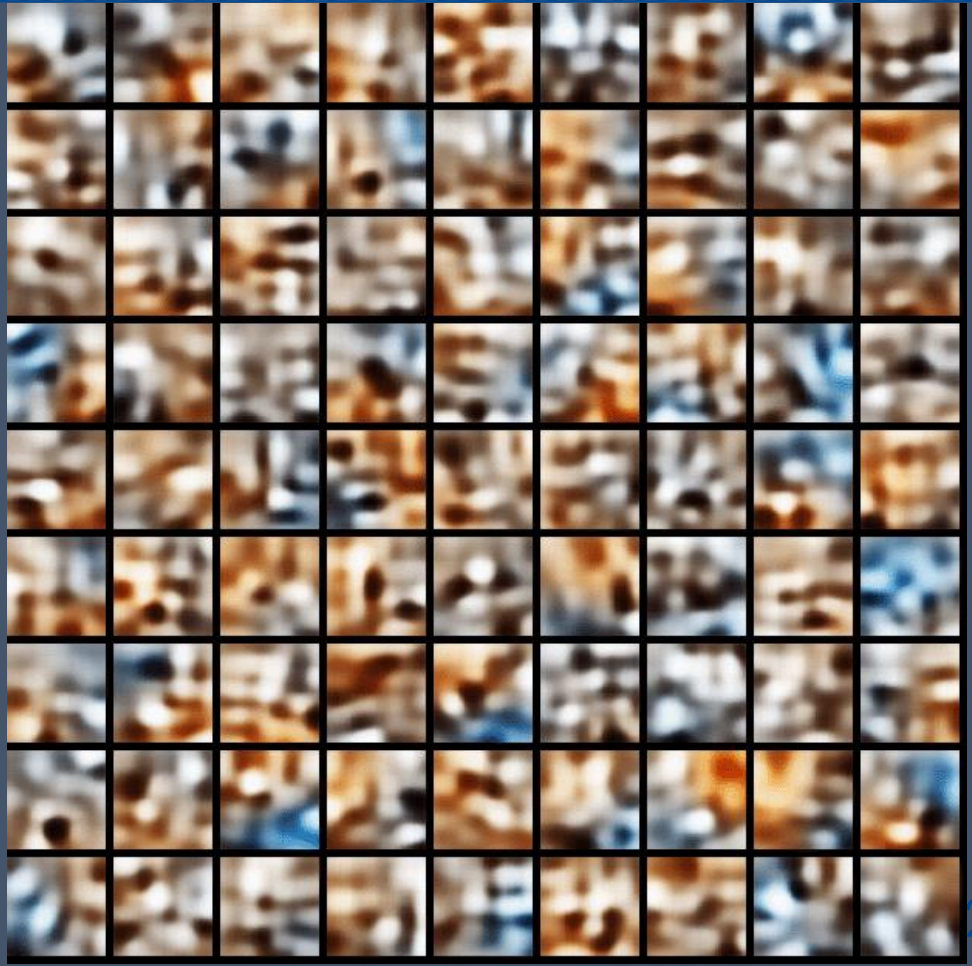
100

Learning Rate

Result on turing configuration



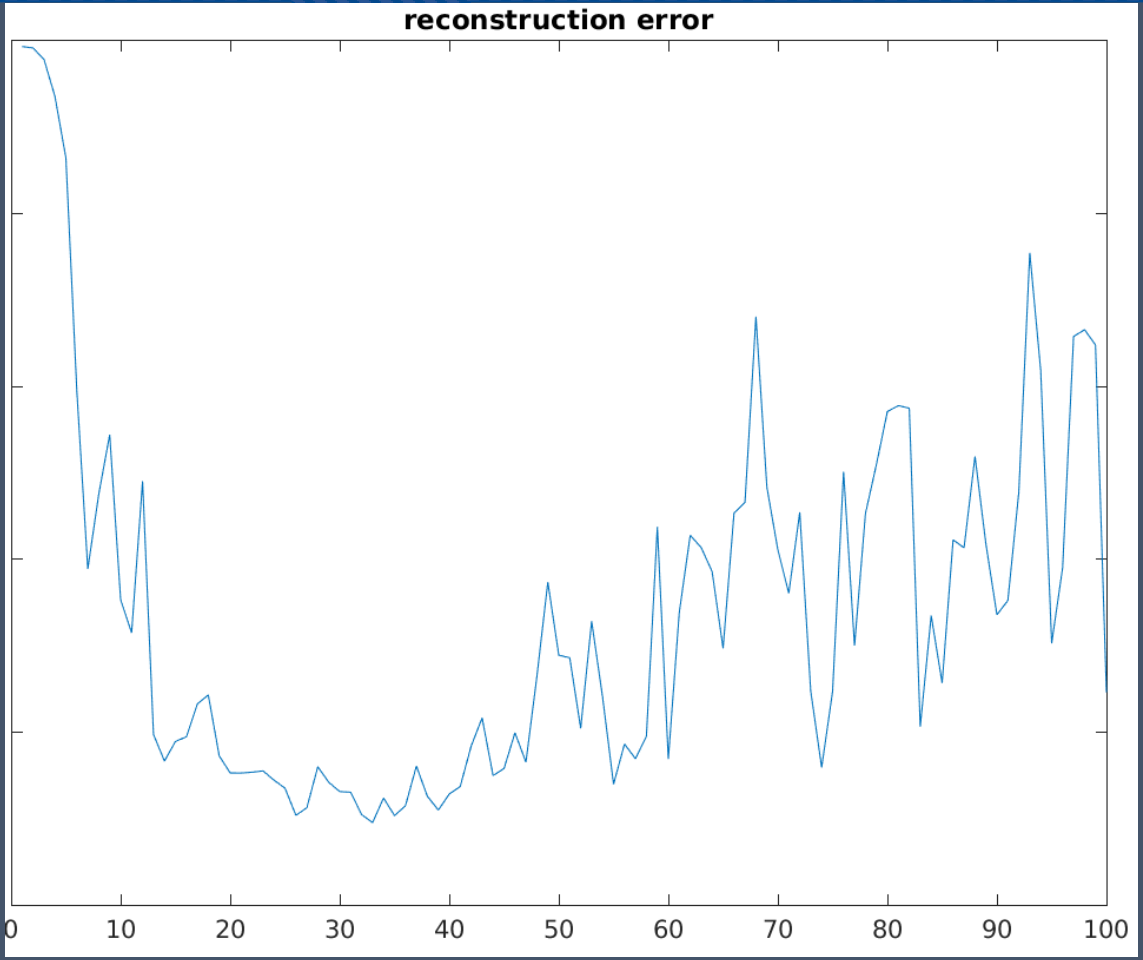
Learning Rate = 0.0003



Learning Rate = 0.00005

Learning Rate

Result on turing configuration



Learning Rate = 0.0003

Learning Rate = 0.00005

Momentum

Result on turing configuration

0

0.2

0.4

0.6

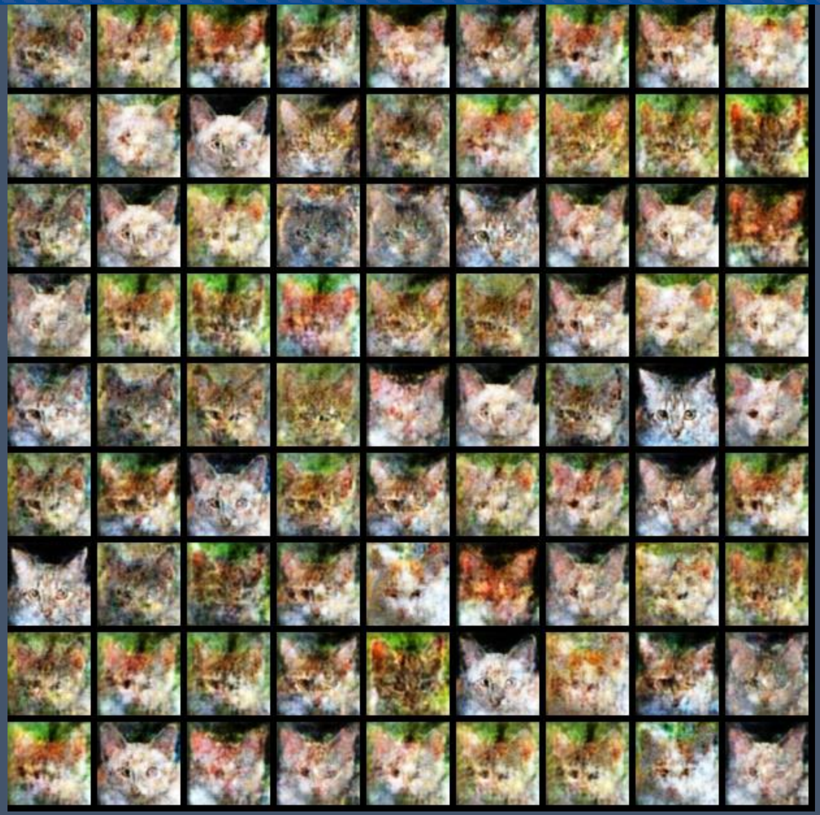
0.8

Number of FC layers

Result on Modifying Net Structure

FC = 1

FC = 2

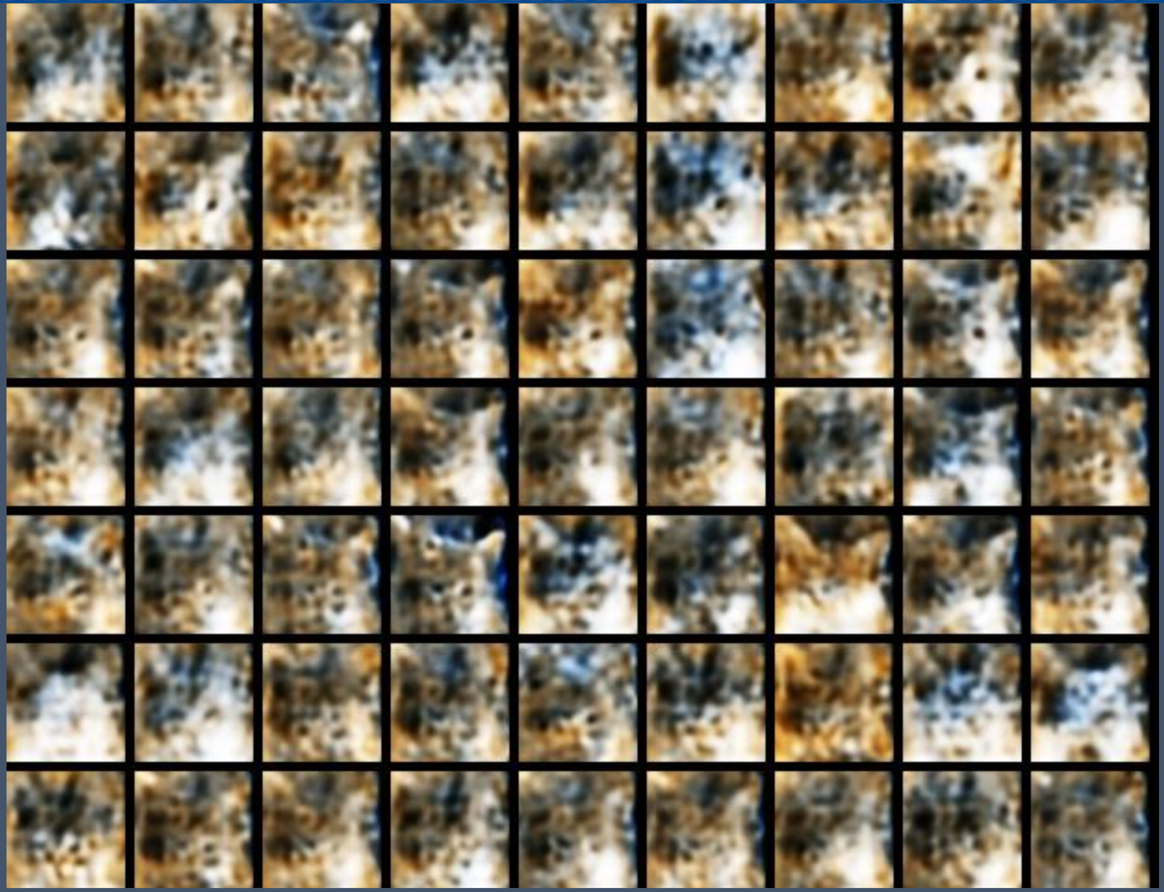


FC = 3

Add Conv layers

Result on Modifying Net Structure

Before Add Conv Layers

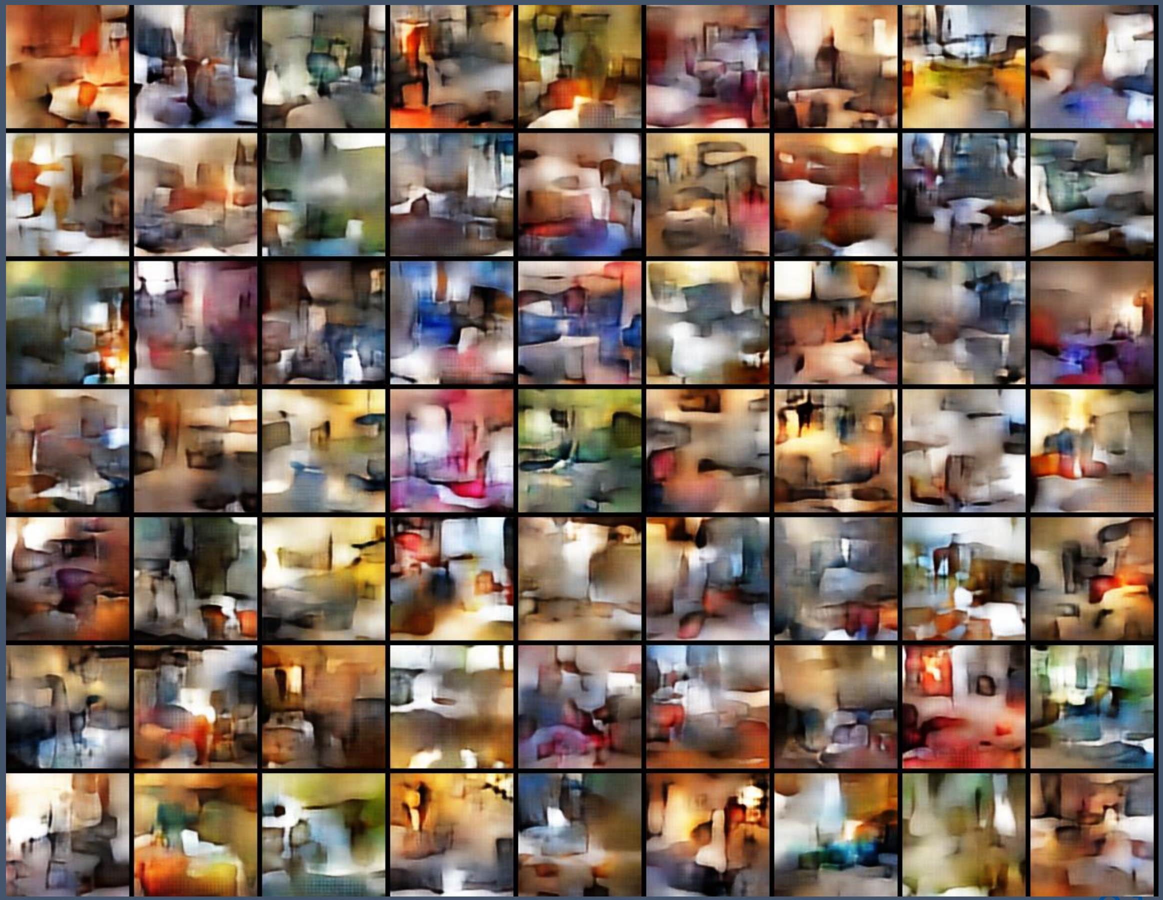


After Add Conv Layers

Modify # of channels

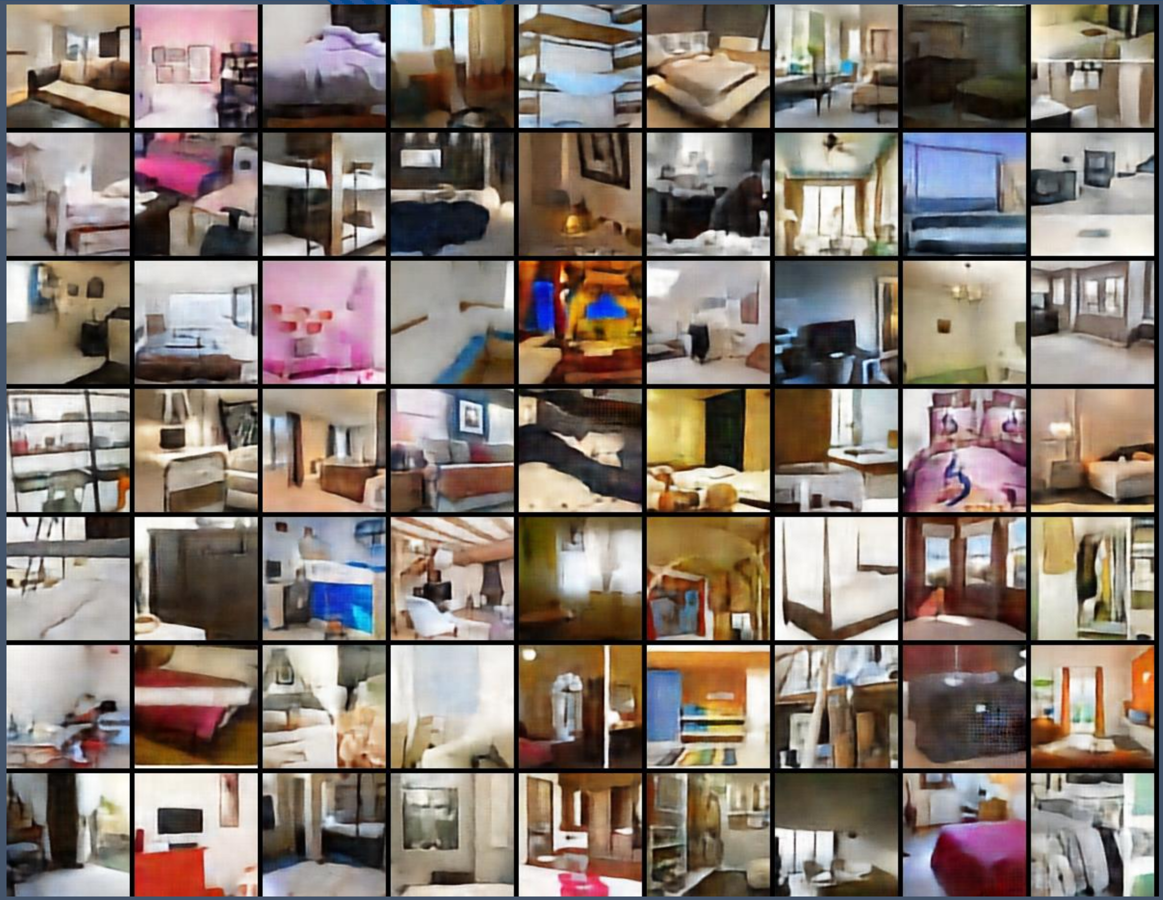
Result on Modifying Net Structure

Before Double Channels

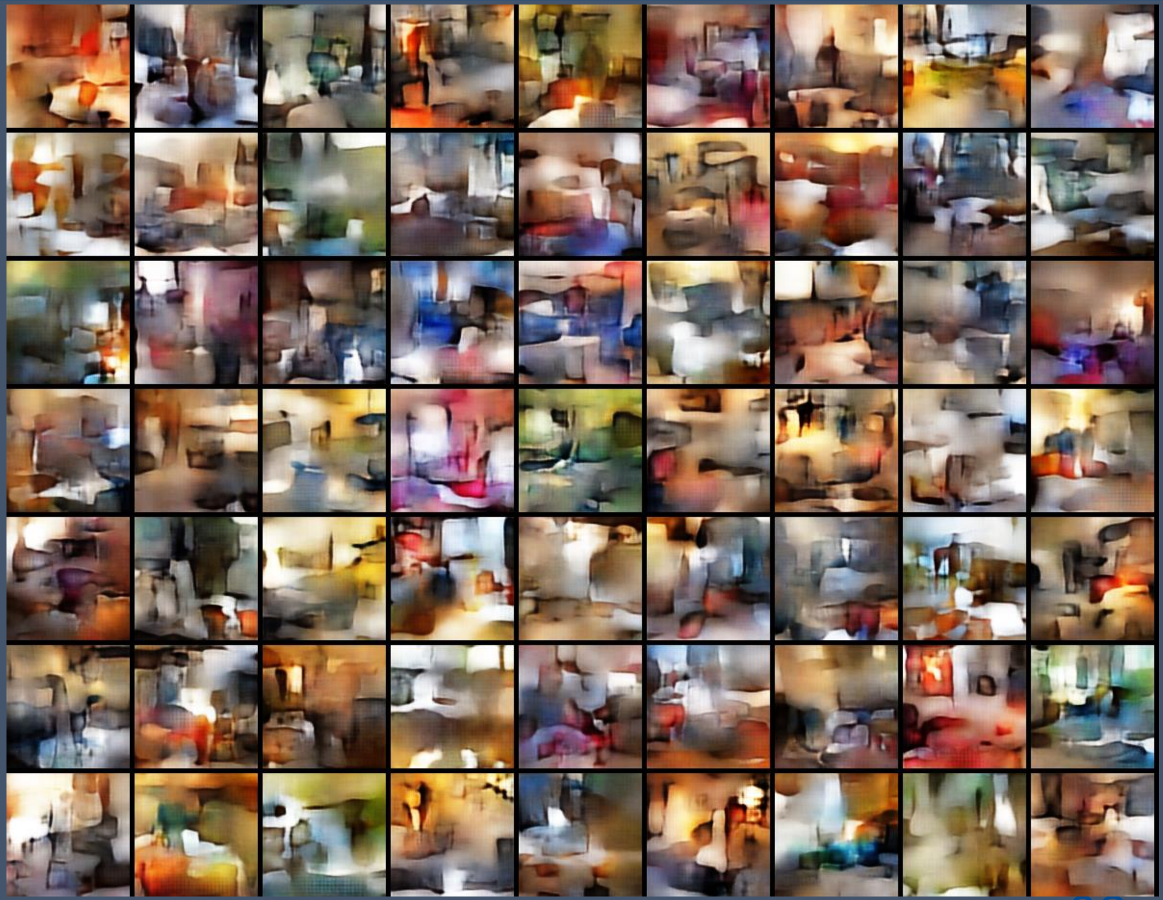


After Double Channels

Results



Reconstruction Results



Synthesis Results

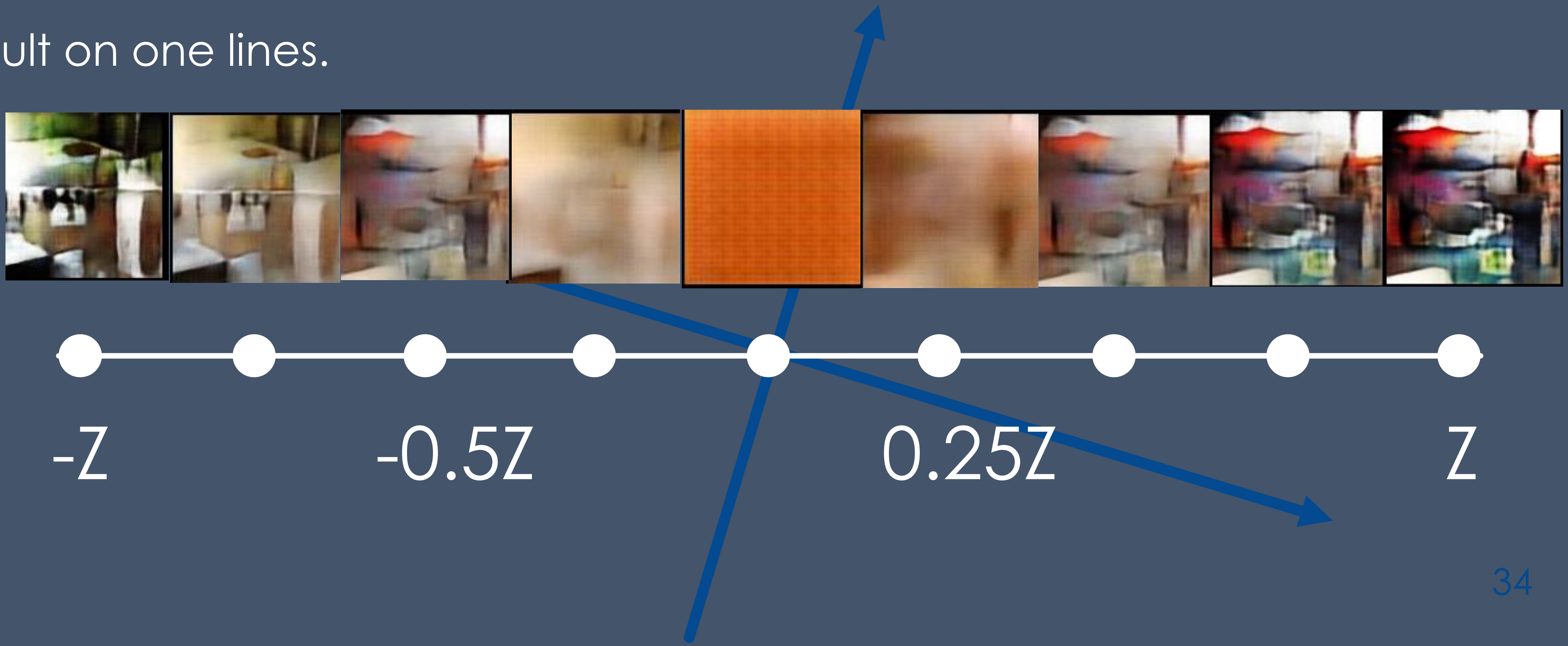
Results

Synthesis Results on Face

Synthesis Results on Clothes

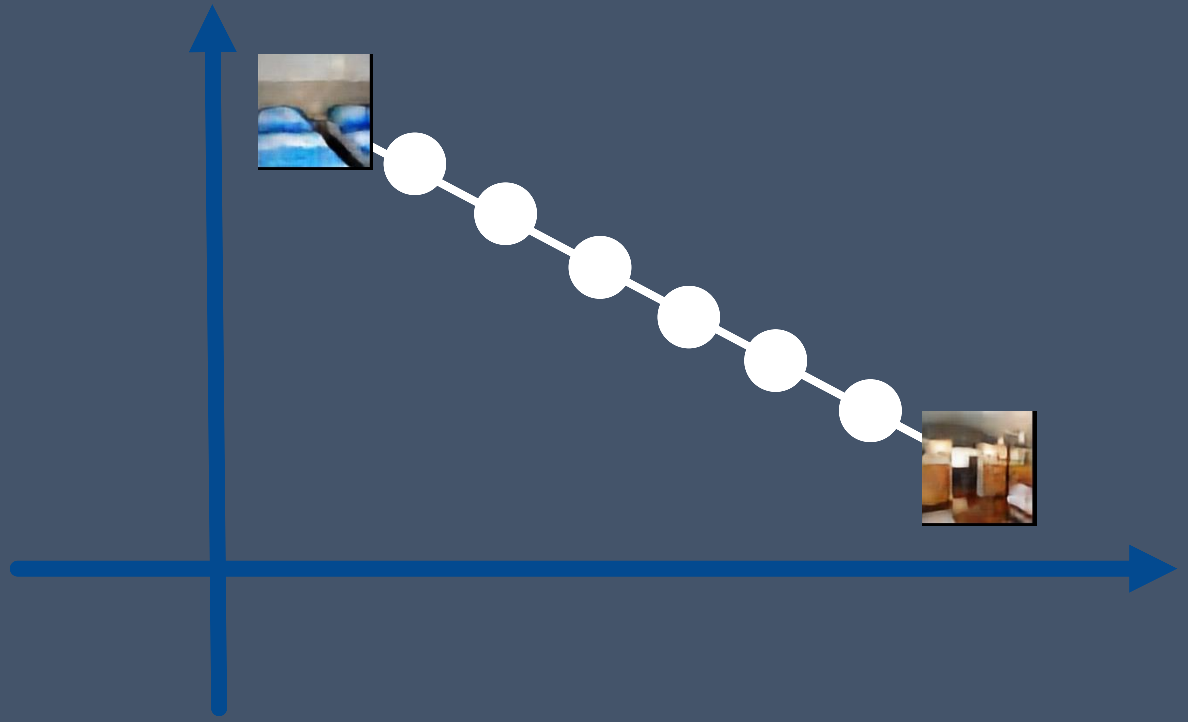
Interpolation Results

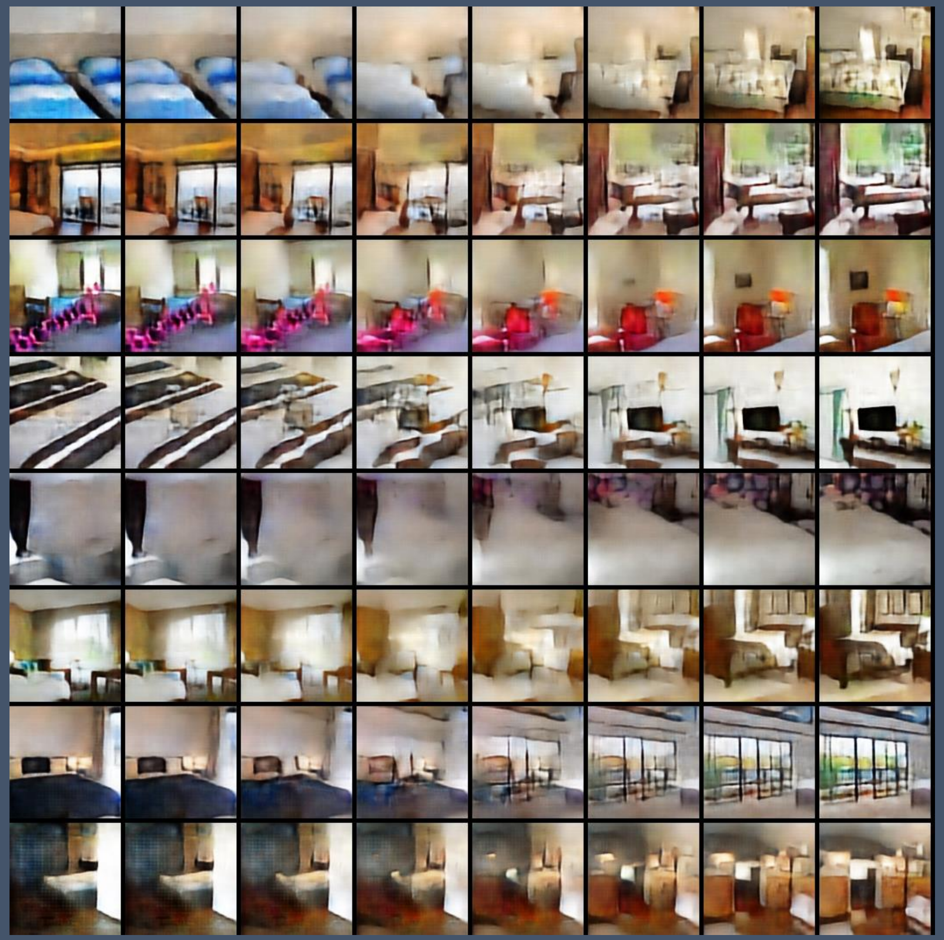
Result on one lines.



Interpolation Results

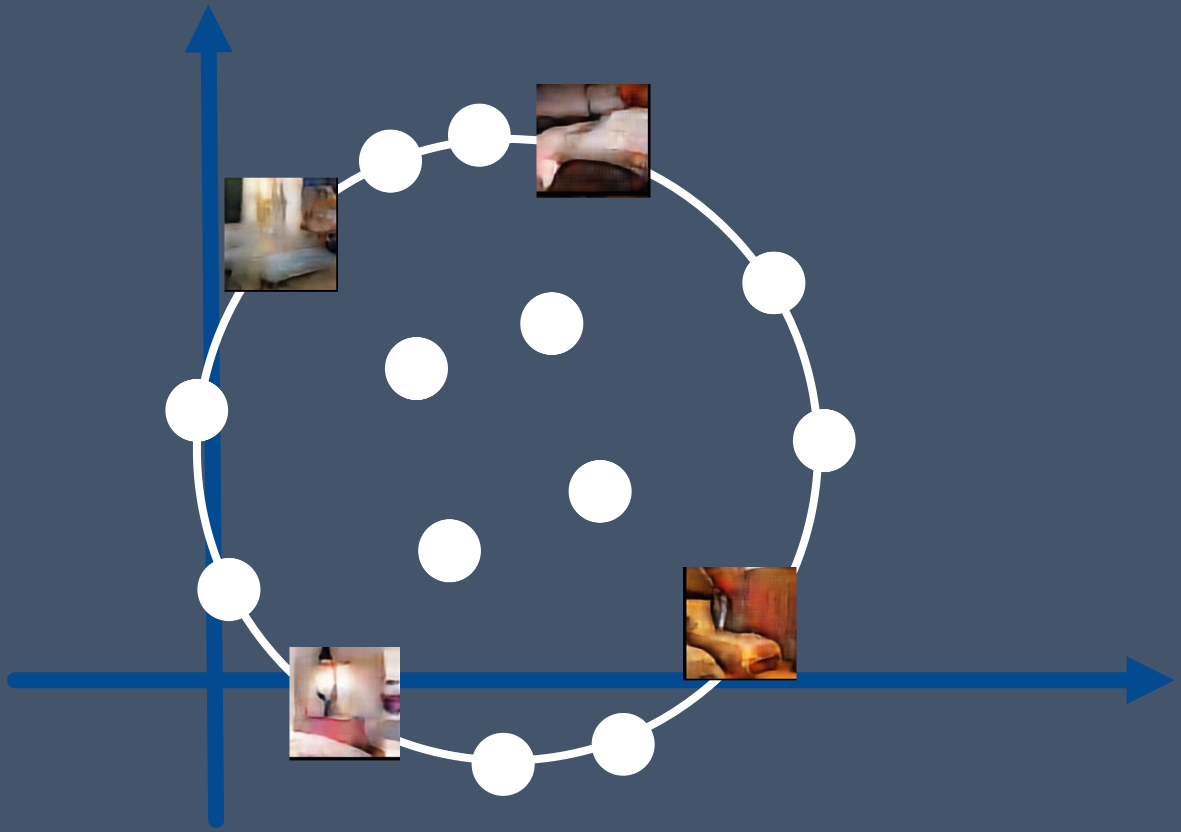
The linear combination of two or more inferred Z (by training image) is interpolation.

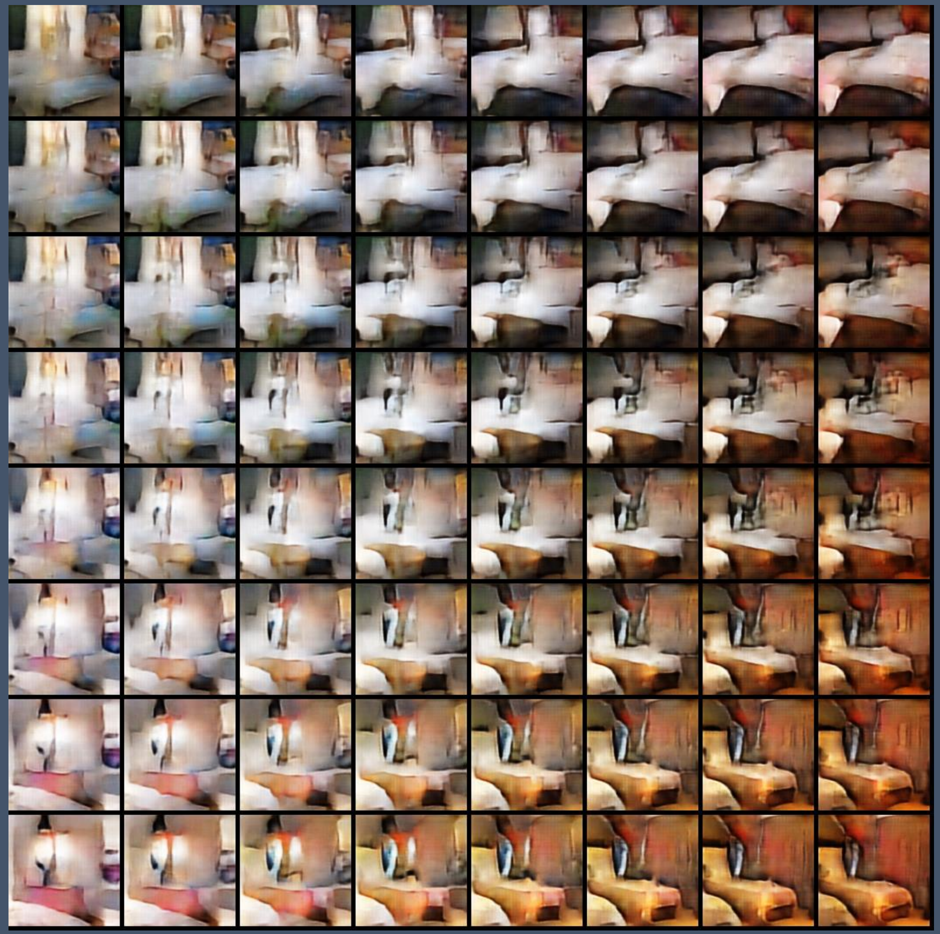




Interpolation Results

The sphere interpolation





Conclusion

Generative ConvNet with Continuous Latent Factors

Synthesis Result is meaningful. As a result, our model can significate explain the data.

- 1. VS **Energy based generative ConvNet**: Both generative. We don't need MCMC which is time consuming. Easy to synthesis.
- 2. VS **Generative Adversely Net** : Need second net, a discriminator to auxiliary training. Our model is simpler.

Reference

- Han, Tian, et al. "**Learning Generative ConvNet with Continuous Latent Factors by Alternating Back-Propagation.**" *arXiv preprint arXiv:1606.08571* (2016).
- Xie, Jianwen, et al. "**A theory of generative convnet**." *arXiv preprint arXiv:1602.03264* (2016).
- Karpathy, Andrej, et al. "**Large-scale video classification with convolutional neural networks**." *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2014.
- Radford, Alec, Luke Metz, and Soumith Chintala. "**Unsupervised representation learning with deep convolutional generative adversarial networks.**" *arXiv preprint arXiv:1511.06434* (2015).

Acknowledgement

- Thanks to:

Prof. Yingnian Wu

Jianwen Xie

Tian Han

Yang Lu

Thanks

to CSST guys

Jerry Xu

Shanghai Jiao Tong University

Q&A

Generative ConvNet Model
by Continuous Latent Factors

- **Generative ConvNet**
  - Non-Linear PCA
  - Alternative Back Propagation
- **Evaluation**
  - Reconstruction Error
  - Train / Test / Negative error
- **Scaling up**
  - Turning Configuration
  - Modifying net structure