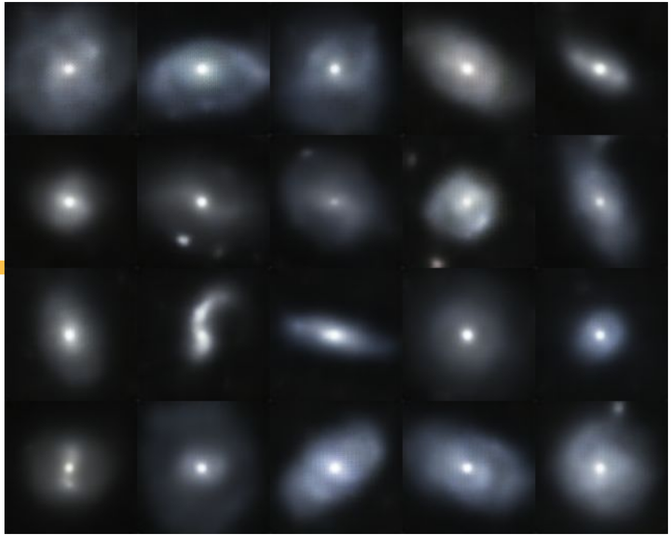
This galaxy does not exist



Matias Carrasco Kind, NCSA

Astronomy AI Seminar, Feb 20th 2019

This person does not exist

https://thispersondoesnotexist.com/

https://arxiv.org/abs/1812.04948

Are we ever going to get this good?

.

.

.

No, and probably don’t need to

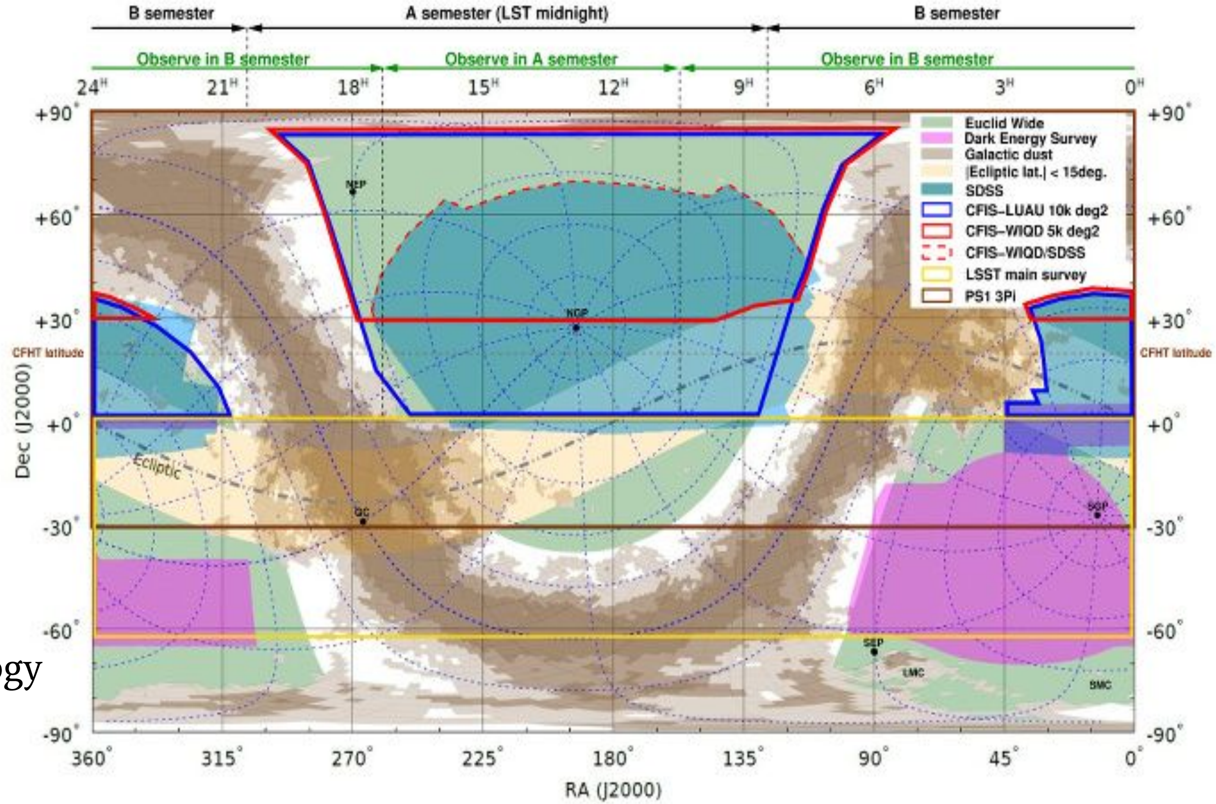
Introduction

- DES DR1 ~ 300M galaxies r ~ 24
- LSST DR1 ~ 9B galaxies r ~ 25
- LSST DR11 ~ 18B galaxies r ~ 27.5

- For galaxies with i < 17,
  DR1 ~ 60K galaxies and
  LSST DR1 ~ 250K

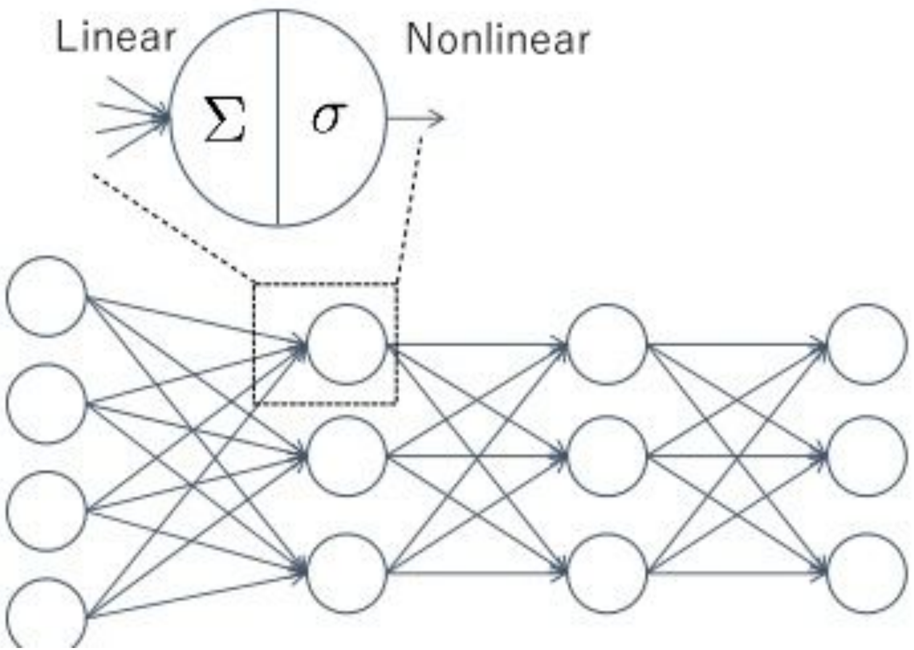
- Single visits from LSST will be same depth as DES final coadds ~25

- LSST overlaps DES completely
- Photo-z still only option for cosmology
- Big data problem → compresion techniques



Deep Learning brief Introduction

There are many different architectures, but this is the basic NN component

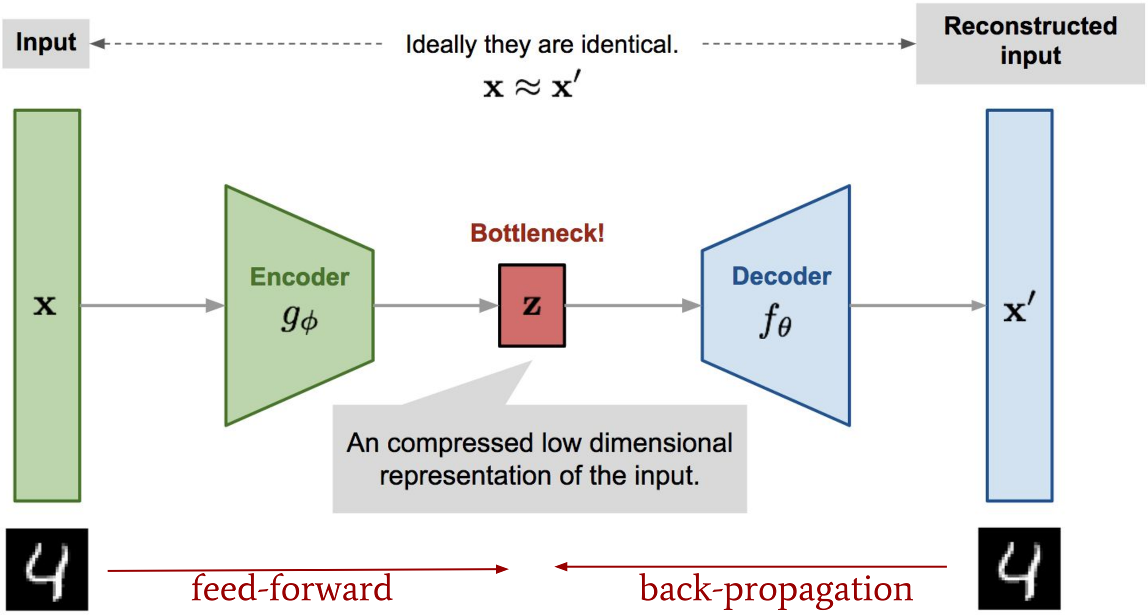


NN: DAG (Directed acyclic graph) of non-linear (sigmoid, relu) operations which are deterministic and differentiable.

Different kind of layers:
Fully connected, CNN, Pooling, etc...

Use feedforward/backpropagation with SGD

Autoencoders



- Around since the 80's
- Data compression
- Anomaly detection
- Denoising
- Regular Machine Learning
- PCA

$$z = g_\phi(x)$$

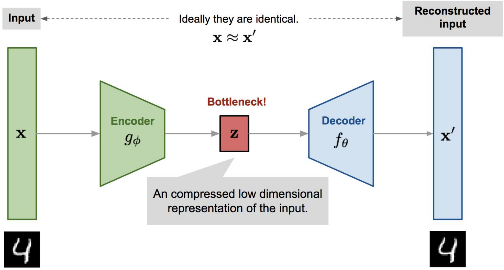
$$x' = f_\theta(g_\phi(x))$$

$$\mathcal{L}(x, x') + reg.$$

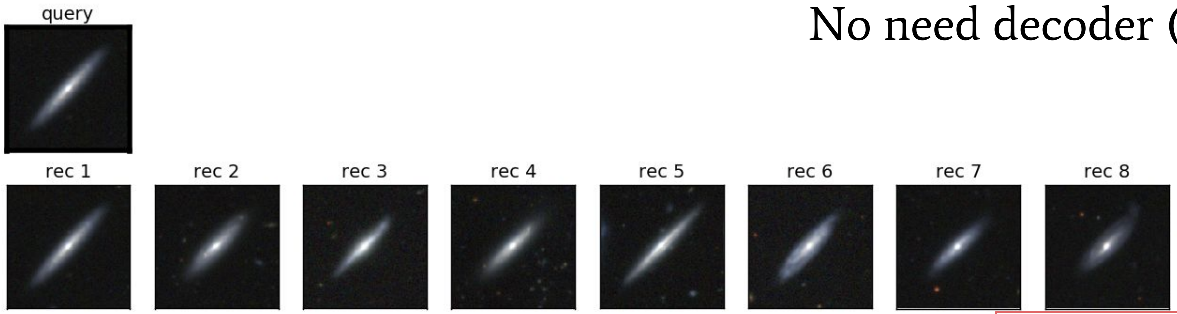
$$\ell_1 = \lambda \sum_i |a_i^{(h)}|$$

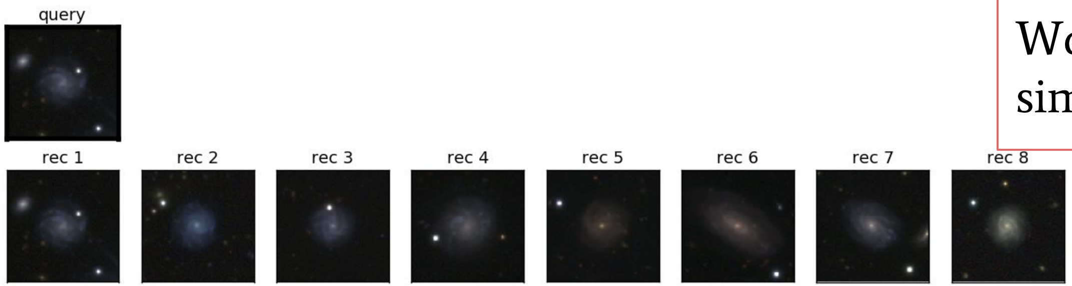
$$\mathcal{L}(x, x') = ||x - x'||^2$$

Autoencoders Applications: Image compression



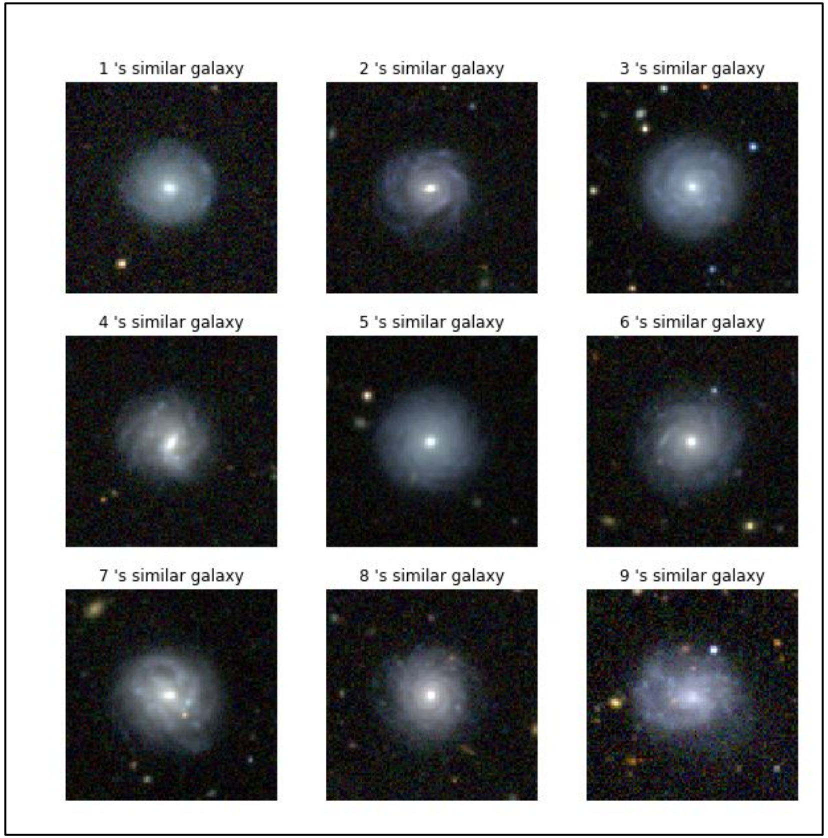
Compress images from 220x220x3 pixels to 50-vector (2000x), for fast similarity search, anomaly detection, etc...
No need decoder (only for Loss)

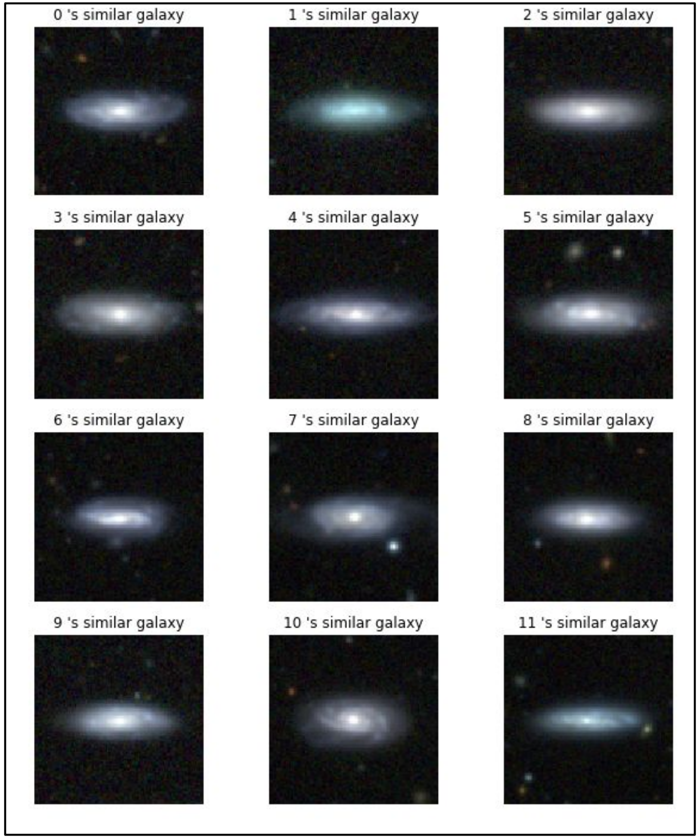




Working on a service for image similarity ranking for DES images

Galaxy selection and similarity search





Bowen, Carrasco-Kind, et al. in prep.

Autoencoders Applications: Data compression

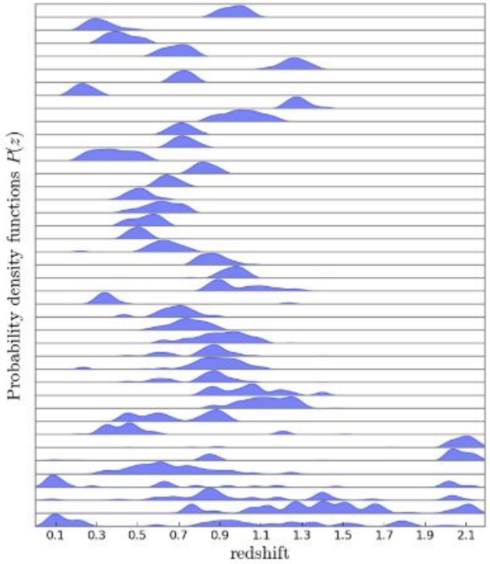
Mon. Not. R. Astron. Soc. **000**, 1–12 (2014) Printed 11 June 2018 (MN LaTeX style file v2.2)

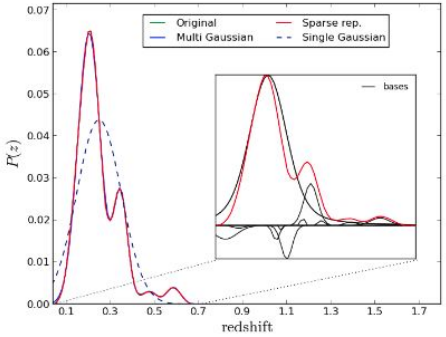
Sparse Representation of Photometric Redshift PDFs: Preparing for Petascale Astronomy

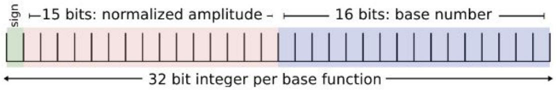
Matias Carrasco Kind* and Robert J. Brunner
Department of Astronomy, University of Illinois, Urbana, IL 61820 USA

https://arxiv.org/pdf/1404.6442.pdf

How to compress photo-z PDF information for millions of galaxies and multiple methods? With Sparse with can achieve 99% accuracy and 95% compression. Can we do better?







Autoencoders Applications: Data compression

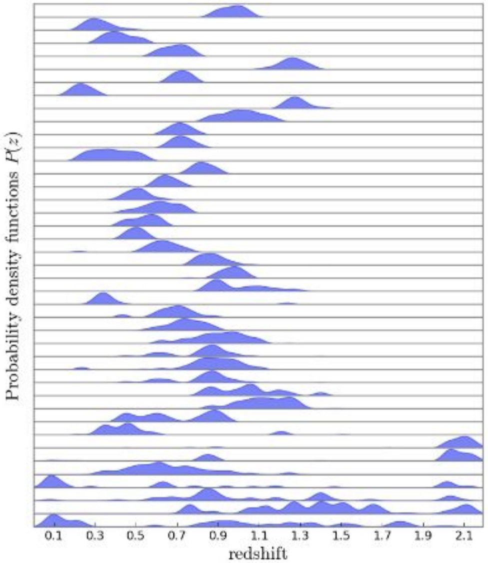
Mon. Not. R. Astron. Soc. 000, 1–12 (2014) Printed 11 June 2018 (MN LaTeX style file v2.2)

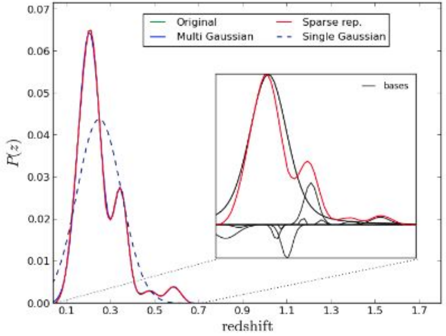
Sparse Representation of Photometric Redshift PDFs: Preparing for Petascale Astronomy

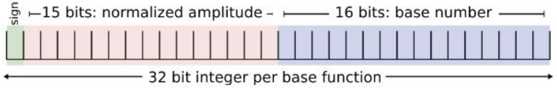
Matias Carrasco Kind* and Robert J. Brunner
Department of Astronomy, University of Illinois, Urbana, IL 61820 USA

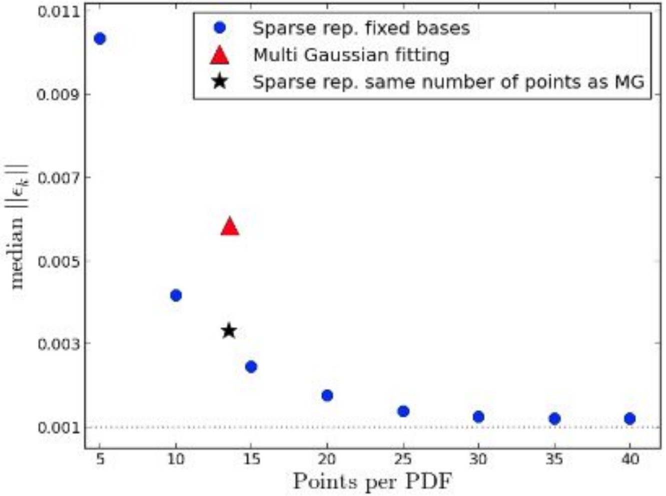
https://arxiv.org/pdf/1404.6442.pdf

How to compress photo-z PDF information for millions of galaxies and multiple methods? With Sparse with can achieve 99% accuracy and 95% compression. Can we do better?









Autoencoders Applications: Data compression

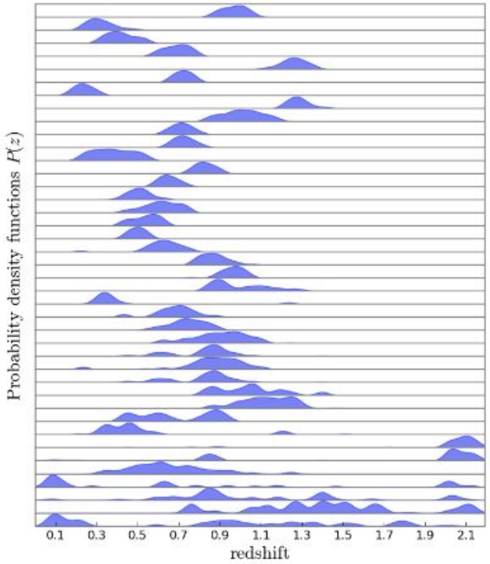
Mon. Not. R. Astron. Soc. **000**, 1–12 (2014) Printed 11 June 2018 (MN LaTeX style file v2.2)

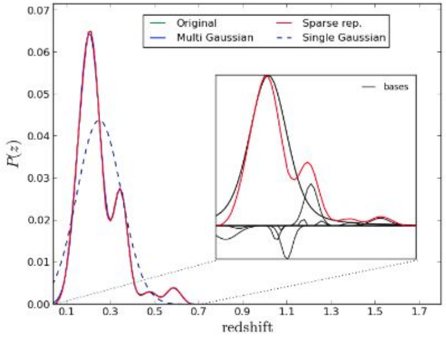
Sparse Representation of Photometric Redshift PDFs: Preparing for Petascale Astronomy

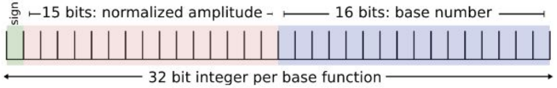
Matias Carrasco Kind* and Robert J. Brunner
Department of Astronomy, University of Illinois, Urbana, IL 61820 USA

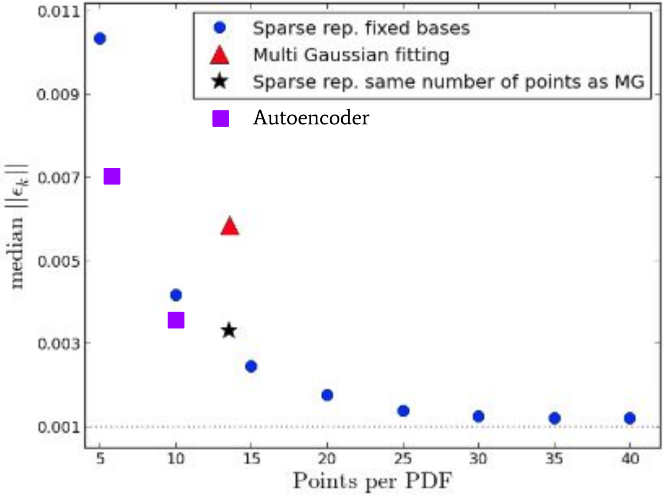
https://arxiv.org/pdf/1404.6442.pdf

How to compress photo-z PDF information for millions of galaxies and multiple methods? With Sparse with can achieve 99% accuracy and 95% compression. Can we do better?



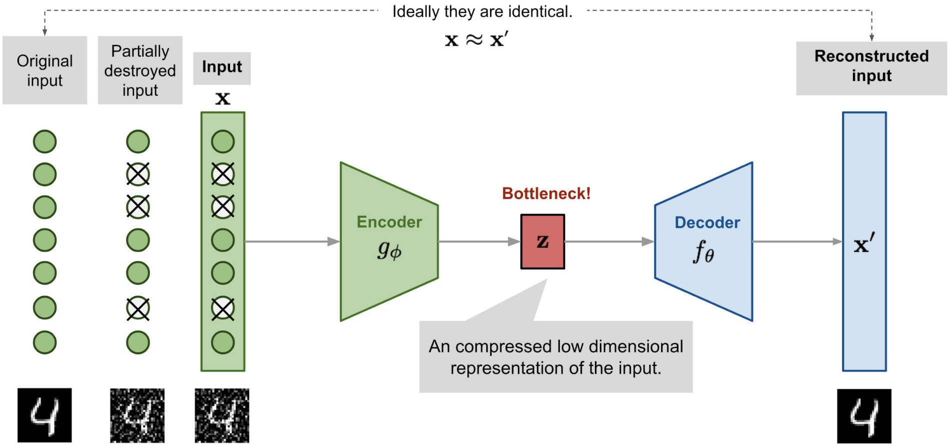


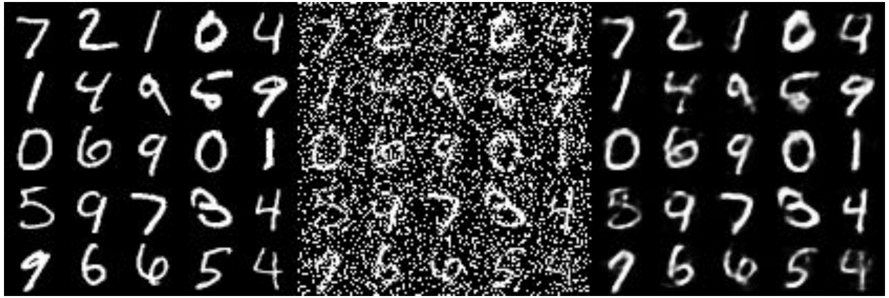




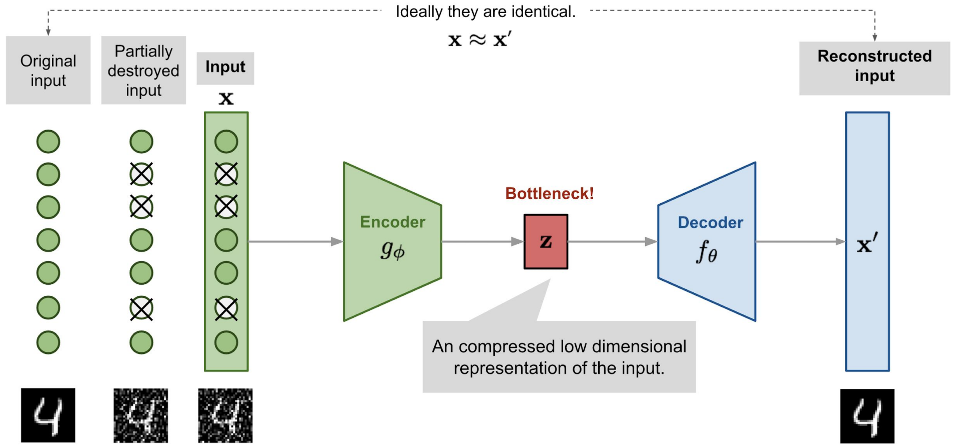
Need powerful decoder, small latent space, no regularization … ongoing/open project

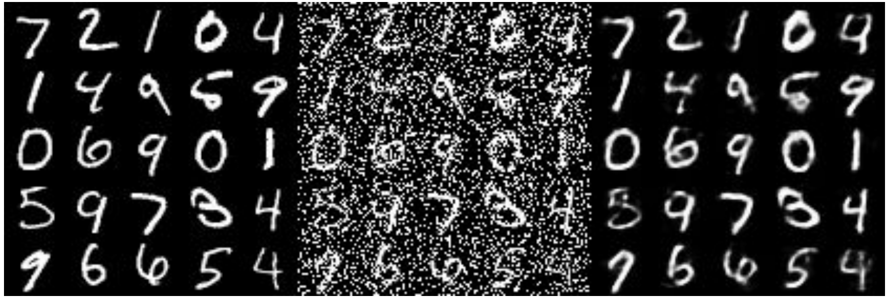
Autoencoders Applications: Denoising

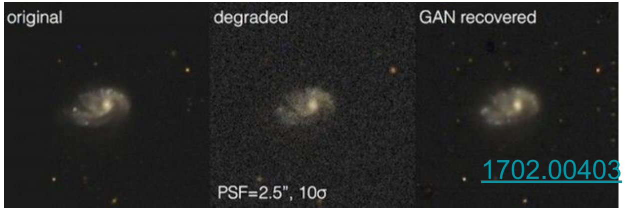




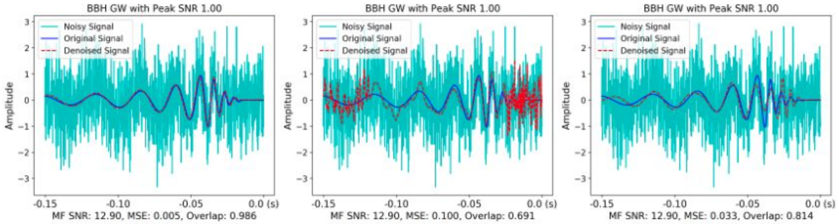
Autoencoders Applications: Denoising







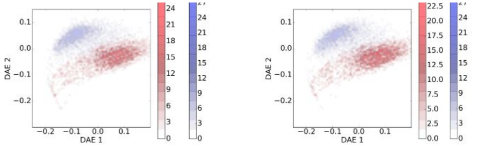
1702.00403



(a)SMTDAE (b)Dictionary Learning (c)PCA

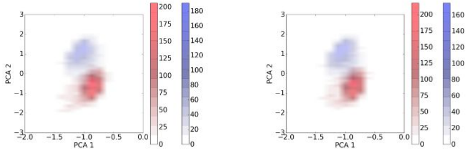
1711.09919

It doesn't need to be noise, but any (even non linear) transformation



(a) Training set (b) Test set

Fig. 1. Histogram of the training set and the test set on the DAE diagram. In the new plane, galaxies classified as star-forming galaxies in the COSMOS2015 catalog are depicted in blue, while the quiescent galaxies are shown in red.



(a) Training set (b) Test set

Fig. 2. Histogram of the training set and the test set projected on the two PCA components.

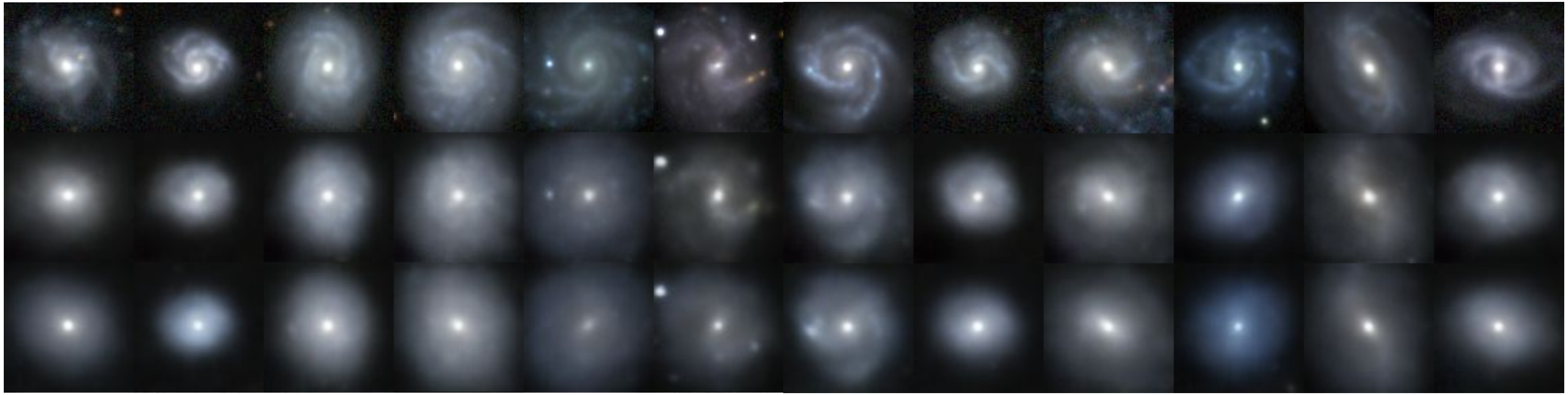
1705.05620

Autoencoders Applications: Reconstruction

Original

Model 1

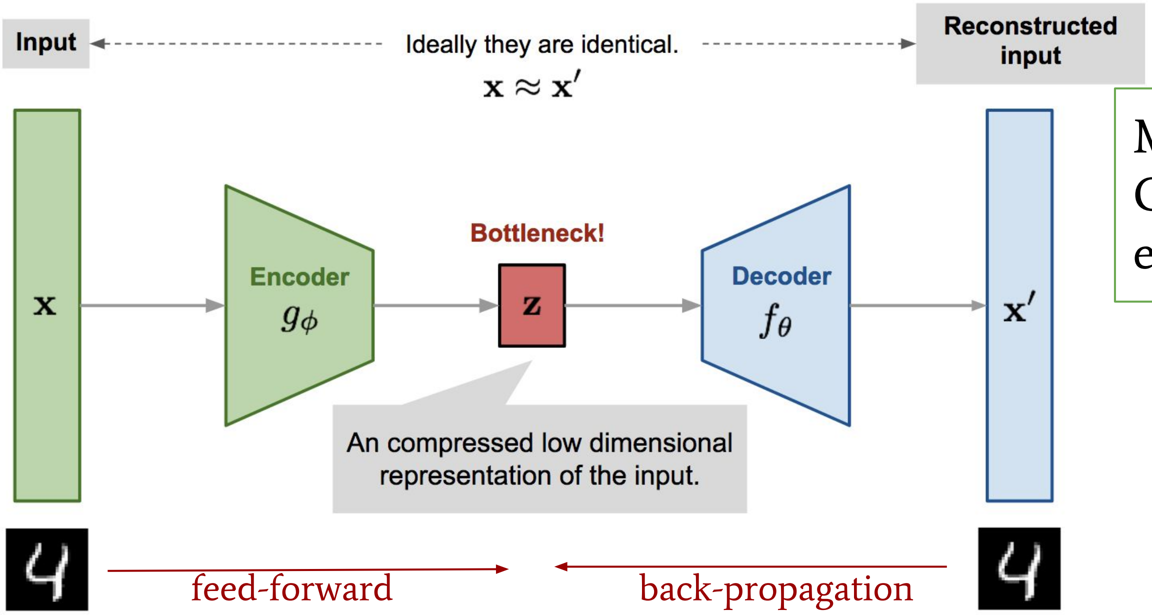
Model 2



Blurry images and structure is lost, but angular sizes, radial profiles and brightness are a match.

What if we can make the model learn properties at the same time as images.

Autoencoders Summary



Many variants: Sparse AE, Contractive AE, Stacked AE, etc...

$$z = g_\phi(x)$$

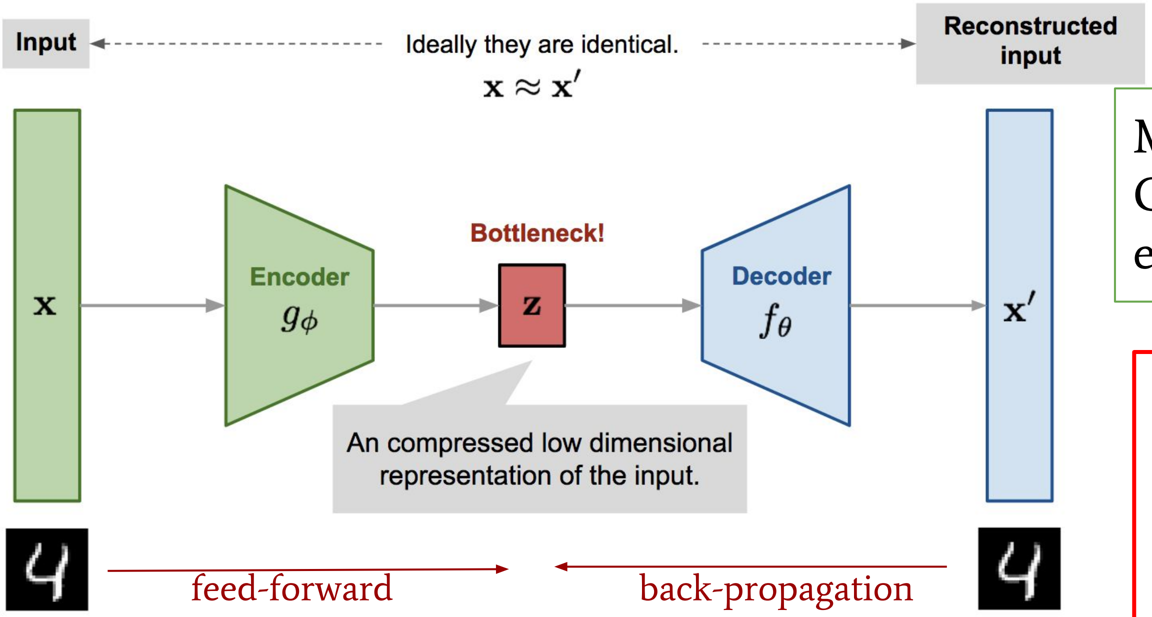
$$x' = f_\theta(g_\phi(x))$$

$$\mathcal{L}(x, x') + reg.$$

$$\ell_1 = \lambda \sum_i |a_i^{(h)}|$$

$$\mathcal{L}(x, x') = ||x - x'||^2$$

Autoencoders Summary



$$z = g_\phi(x)$$

$$x' = f_\theta(g_\phi(x))$$

$$\mathcal{L}(x, x') + reg.$$

$$\ell_1 = \lambda \sum_i |a_i^{(h)}|$$

$$\mathcal{L}(x, x') = ||x - x'||^2$$

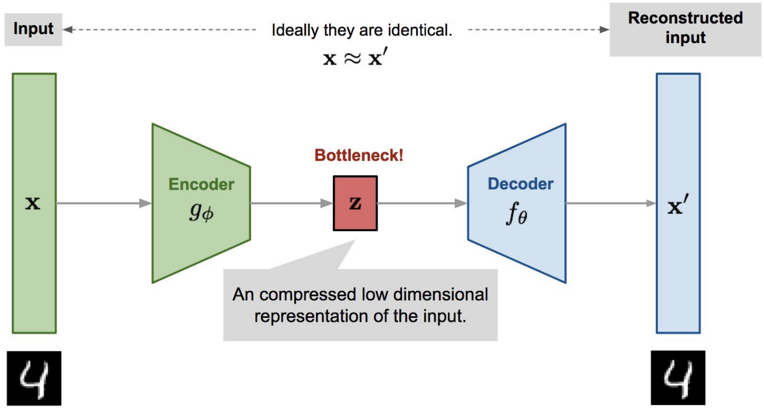
Many variants: Sparse AE, Contractive AE, Stacked AE, etc...

We have no control over z which is randomly distributed.

If we can manipulate z we can generate data that in the space of X.

$$p(z) = \mathcal{N}(0, I)$$

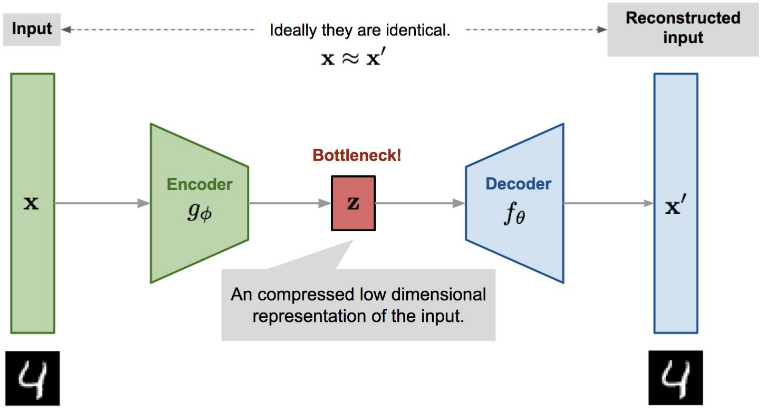
Can we sample z to generate fake data?



- Map x to a distribution $p(z|x)$
- Sample from distribution $z_i \sim p(z)$
- Generate fake data $x_i' \sim p_\theta(x'|z)$
- Probabilistic approach

$$p(x, z) = p(x|z)p(z)$$

Can we sample z to generate fake data?



Exist θ for max the likelihood

$$p(x) = \int p(x|z,\theta)p(z)dz$$

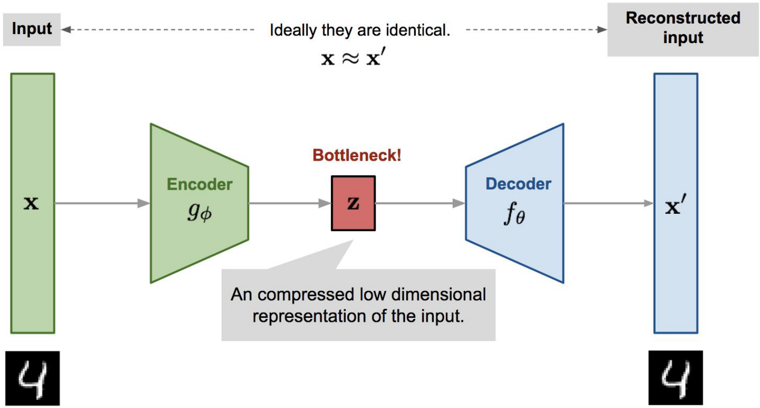
$$p(z|x) = \frac{p(x|z)p(z)}{p(x)}$$

Too expensive

- Map x to a distribution $p(z|x)$
- Sample from distribution $z_i \sim p(z)$
- Generate fake data $x_i' \sim p_\theta(x'|z)$
- Probabilistic approach

$$p(x,z) = p(x|z)p(z)$$

Can we sample z to generate fake data?



Exist θ for max the likelihood

$$p(x) = \int p(x|z,\theta)p(z)dz$$

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)}$$

Too expensive

We need an approximate posterior (prob. encoder)

$$q_\lambda(z|x) \approx p(z|x)$$

And we can use q to be Gaussian (there are other alternatives)

$$q_\lambda(z|x) = \mathcal{N}(z; \mu_\lambda(x), \sigma_\lambda(x))$$

$$p(z) = \mathcal{N}(0, I)$$

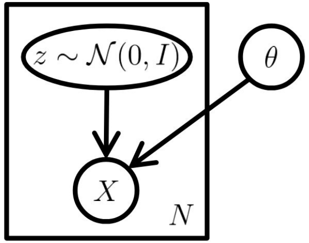
- Map x to a distribution $p(z|x)$
- Sample from distribution $z_i \sim p(z)$
- Generate fake data $x'_i \sim p_\theta(x'|z)$
- Probabilistic approach

$$p(x,z) = p(x|z)p(z)$$

Variational "Autoencoder"

Variational AutoEncoders

How to compare q(z|x) and p(z|x)?
Using the Kullback-Leibler (KL) divergence which compares distributions. $\mathbb{KL}(p||q) = -\sum p(x) \log \frac{q(x)}{p(x)}$



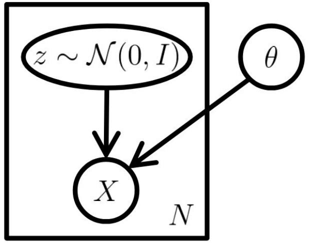
$$\mathbb{KL}(q_\lambda(z|x)||p(z|x)) = \mathbb{E}_q[\log q_\lambda(z|x)] - \mathbb{E}_q[\log p(z|x)]$$

$$\mathbb{KL}(q_\lambda(z|x)||p(z|x)) = \mathbb{E}_q[\log q_\lambda(z|x)] - \mathbb{E}_q[\log p(x|z) + \log p(z) - \log p(x)]$$

Variational AutoEncoders

How to compare q(z|x) and p(z|x)?
Using the Kullback-Leibler (KL) divergence which compares distributions.

$$\mathbb{KL}(p||q) = -\sum p(x)\log\frac{q(x)}{p(x)}$$



$$\mathbb{KL}(q_\lambda(z|x)||p(z|x)) = \mathbb{E}_q[\log q_\lambda(z|x)] - \mathbb{E}_q[\log p(z|x)]$$

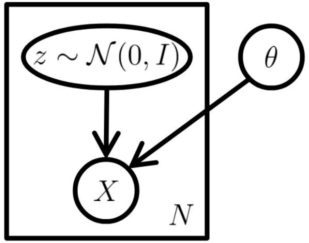
$$\mathbb{KL}(q_\lambda(z|x)||p(z|x)) = \mathbb{E}_q[\log q_\lambda(z|x)] - \mathbb{E}_q[\log p(x|z) + \log p(z) - \log p(x)]$$

$$\underbrace{\log p(x) - \mathbb{KL}(q_\lambda(z|x)||p(z|x))} = \underbrace{\mathbb{E}_q[\log p(x|z)] - \mathbb{KL}(q_\lambda(z|x)||p(z))}$$

Variational AutoEncoders

How to compare q(z|x) and p(z|x)?
Using the Kullback-Leibler (KL) divergence which compares distributions.

$$\mathbb{KL}(p||q) = -\sum p(x) \log \frac{q(x)}{p(x)}$$



$$\mathbb{KL}(q_\lambda(z|x)||p(z|x)) = \mathbb{E}_q[\log q_\lambda(z|x)] - \mathbb{E}_q[\log p(z|x)]$$

$$\mathbb{KL}(q_\lambda(z|x)||p(z|x)) = \mathbb{E}_q[\log q_\lambda(z|x)] - \mathbb{E}_q[\log p(x|z) + \log p(z) - \log p(x)]$$

$$\log p(x) - \mathbb{KL}(q_\lambda(z|x)||p(z|x)) = \mathbb{E}_q[\log p(x|z)] - \mathbb{KL}(q_\lambda(z|x)||p(z))$$

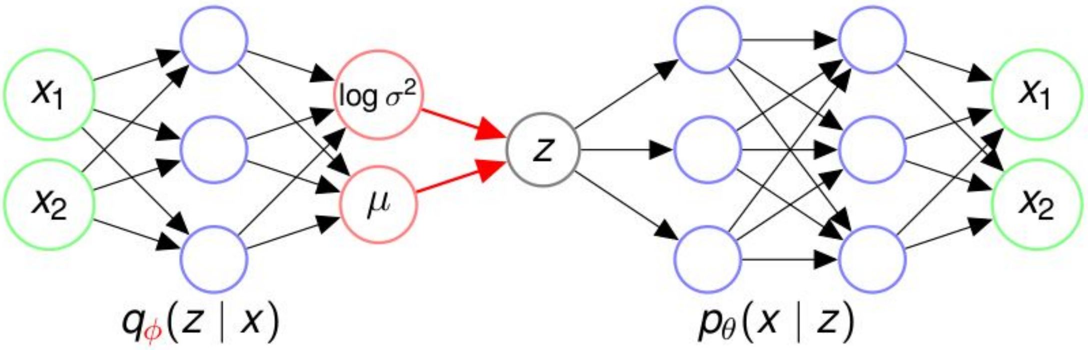
We want to maximize!

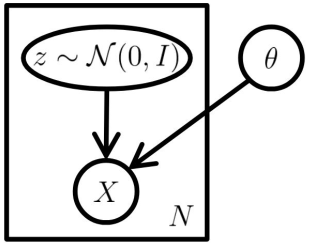
By construction the KL divergence is always >= 0

Evidence Lower Bound (ELBO)
Loss function = - ELBO

Loss function = “Reconstruction” + “Regularization”
But is not really an autoencoder

Variational AutoEncoders





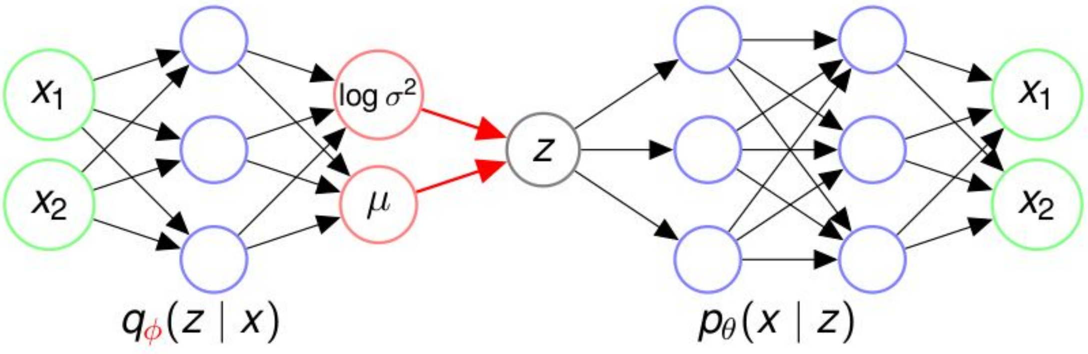
We need:

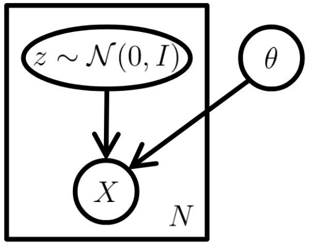
$$p(z) = \mathcal{N}(0, I)$$

$$q_\lambda(z|x) = \mathcal{N}(z; \mu_\lambda(x), \sigma_\lambda(x))$$

$$Loss_{VAE} = -\mathbb{E}_q[\log p(x|z)] + \mathbb{KL}(q(z|x)||p(z))$$

Variational AutoEncoders





We need:

$$p(z) = \mathcal{N}(0, I)$$

$$q_\lambda(z|x) = \mathcal{N}(z; \mu_\lambda(x), \sigma_\lambda(x))$$

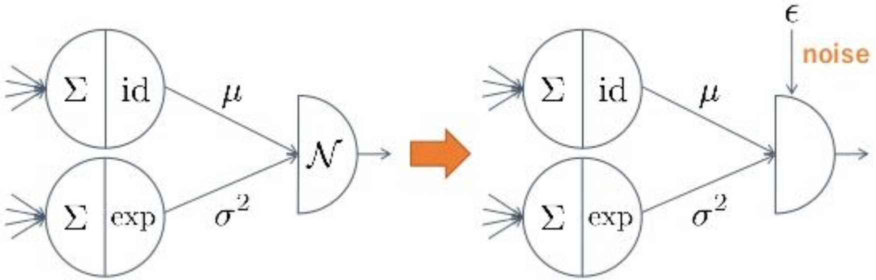
One last problem: Sampling z directly from μ and σ introduces an stochastic node which is untracktable for backpropagation, i.e., can't use SGD to tune parameters since info is lost.

$$Loss_{VAE} = -\mathbb{E}_q[\log p(x|z)] + \mathbb{KL}(q(z|x)||p(z))$$

Reparameterization trick

$$Loss_{VAE} = -\mathbb{E}_q[\log p(x|z)] + \mathbb{KL}(q(z|x)||p(z))$$

Write the sampling procedure as a differentiable computation



$z \sim \mathcal{N}(\mu, \sigma^2)$ $\quad$ $z = \mu + \sigma \cdot \epsilon, \ \epsilon \sim \mathcal{N}(0,1)$

We need:

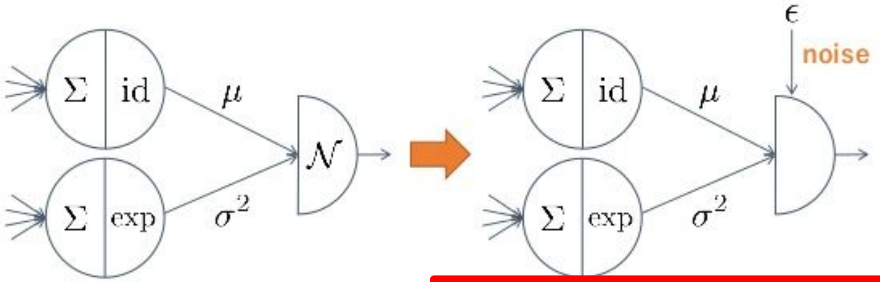
$$p(z) = \mathcal{N}(0, I)$$

$$q_\lambda(z|x) = \mathcal{N}(z; \mu_\lambda(x), \sigma_\lambda(x))$$

Reparameterization trick

$$Loss_{VAE} = -\mathbb{E}_q[\log p(x|z)] + \mathbb{KL}(q(z|x)||p(z))$$

Write the sampling procedure as a differentiable computation



$z \sim \mathcal{N}(\mu, \sigma^2)$

$\boxed{z = \mu + \sigma \cdot \epsilon, \ \epsilon \sim \mathcal{N}(0,1)}$

We need:

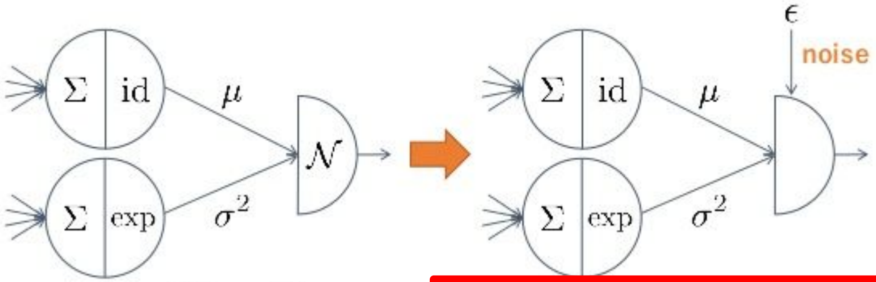
$$p(z) = \mathcal{N}(0, I)$$

$$q_\lambda(z|x) = \mathcal{N}(z; \mu_\lambda(x), \sigma_\lambda(x))$$

Reparameterization trick

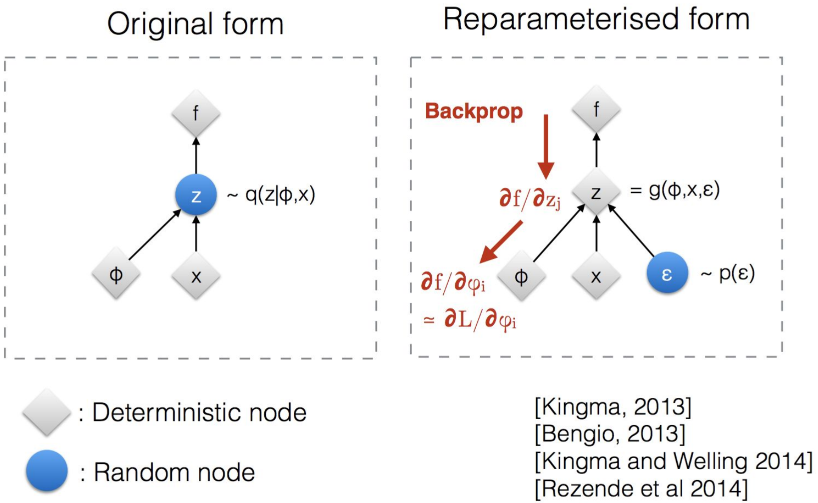
$$Loss_{VAE} = -\mathbb{E}_q[\log p(x|z)] + \mathbb{KL}(q(z|x)||p(z))$$

Write the sampling procedure as a differentiable computation



$$z \sim \mathcal{N}(\mu, \sigma^2)$$

$$\boxed{z = \mu + \sigma \cdot \epsilon, \;\; \epsilon \sim \mathcal{N}(0,1)}$$



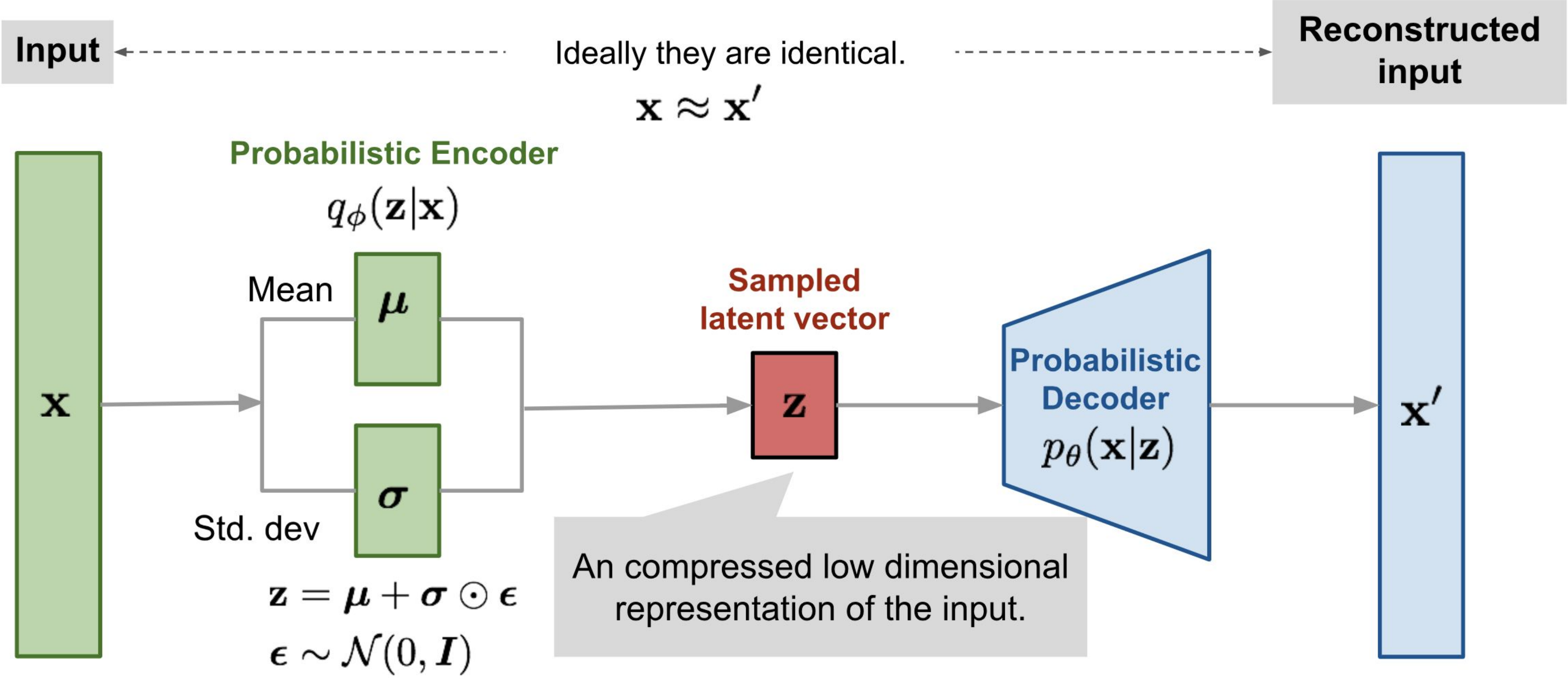
We need:

$$p(z) = \mathcal{N}(0, I)$$

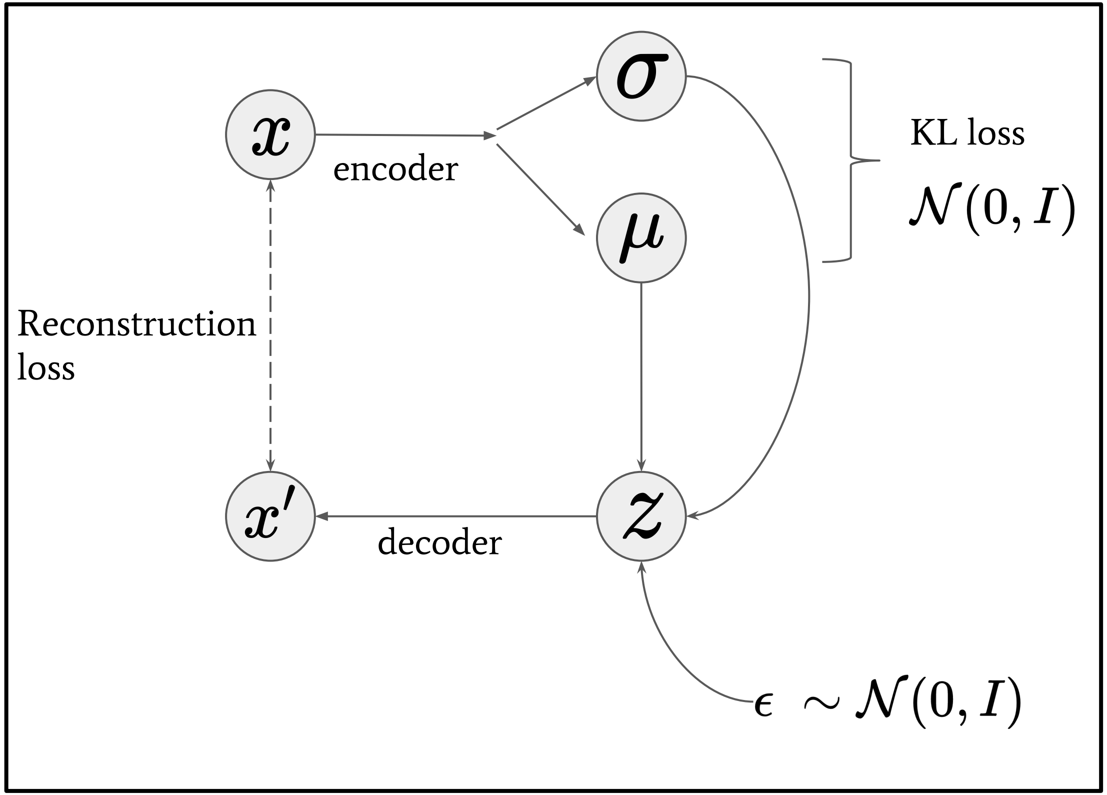
$$q_\lambda(z|x) = \mathcal{N}(z; \mu_\lambda(x), \sigma_\lambda(x))$$

[Kingma, 2013]
[Bengio, 2013]
[Kingma and Welling 2014]
[Rezende et al 2014]

Summarizing VAE

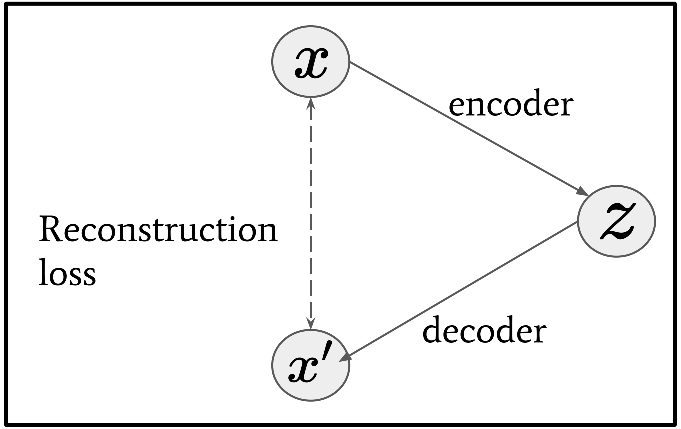


Summarizing VAE



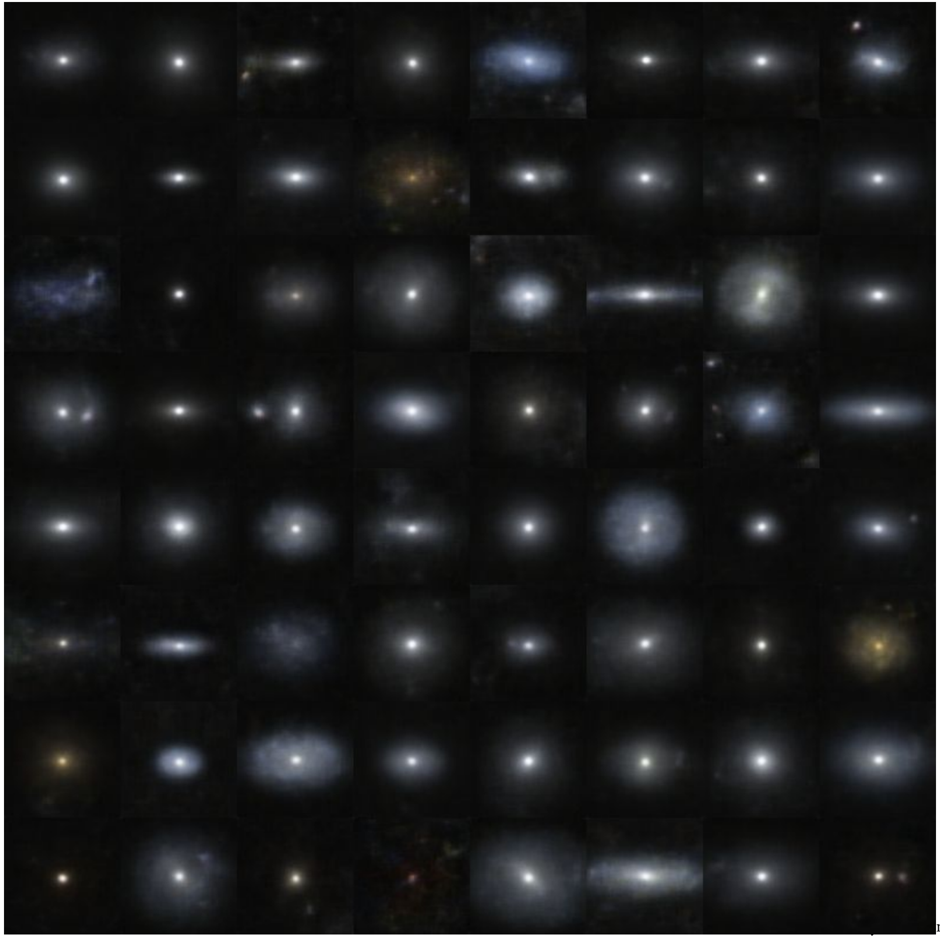
Variational Autoencoder

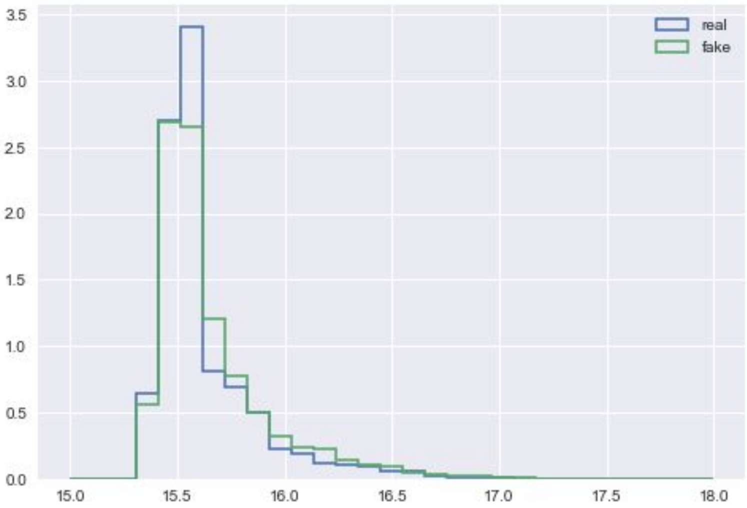
Autoencoder



Many extensions:
PixelVAE, MGVAE, etc....

VAE Sampling example (no reconstruction)

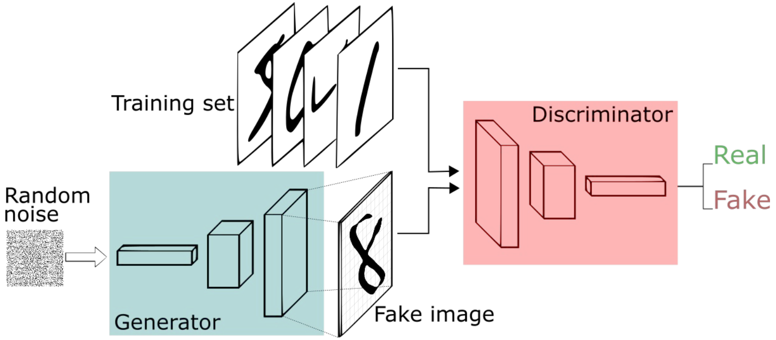


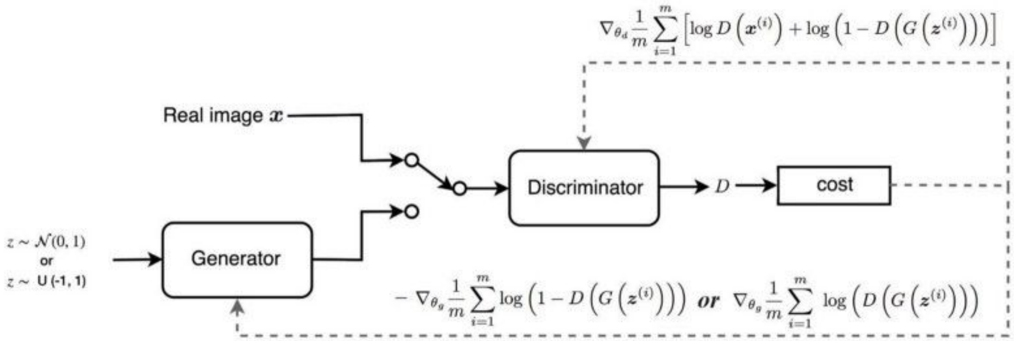


We can generate samples from z, next step is can we constrain what's being sampled?

Generative Adversarial Networks (GAN)

Gan can be VERY good to specific image generation and create realistic images. Very powerful discriminator



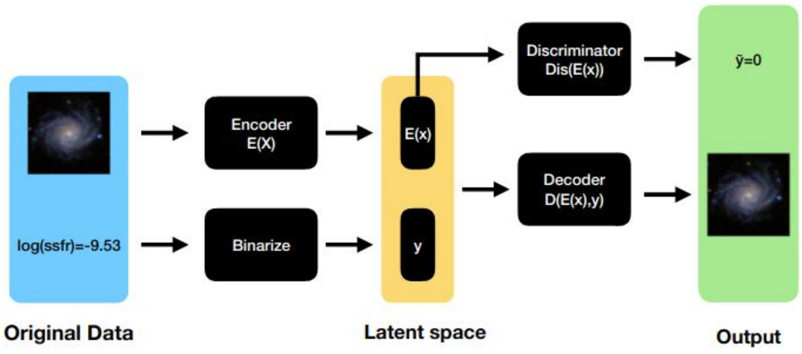


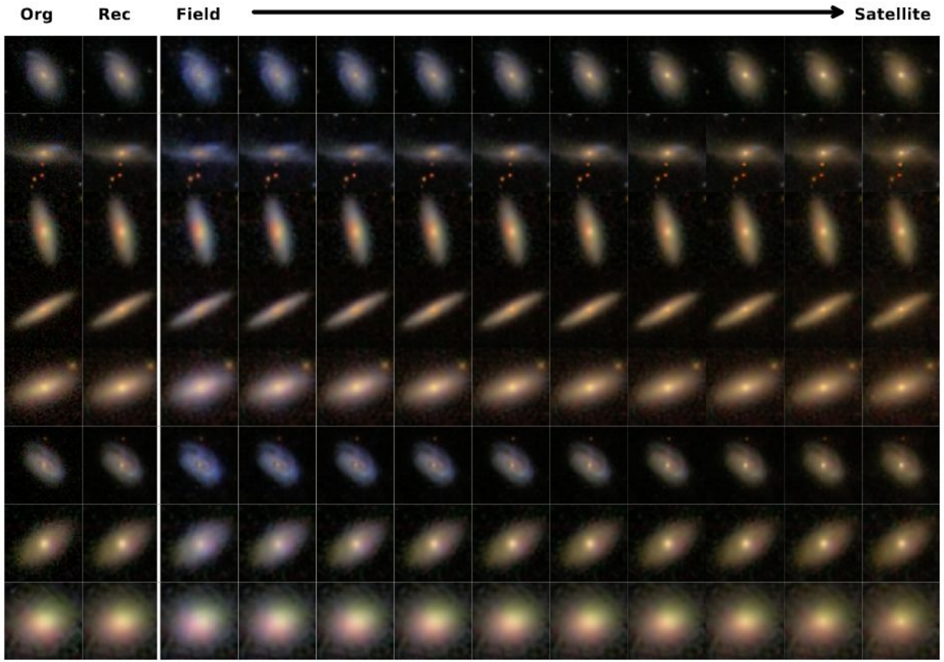
But:

- Very hard to train (unstable)
- Not really sampling methods
- Hard to evaluate likelihood of data p(x)
- Tend to underfit data distribution
- Main goal is to fool the discriminator

Very powerful if combined with VAE

AE + Discriminator

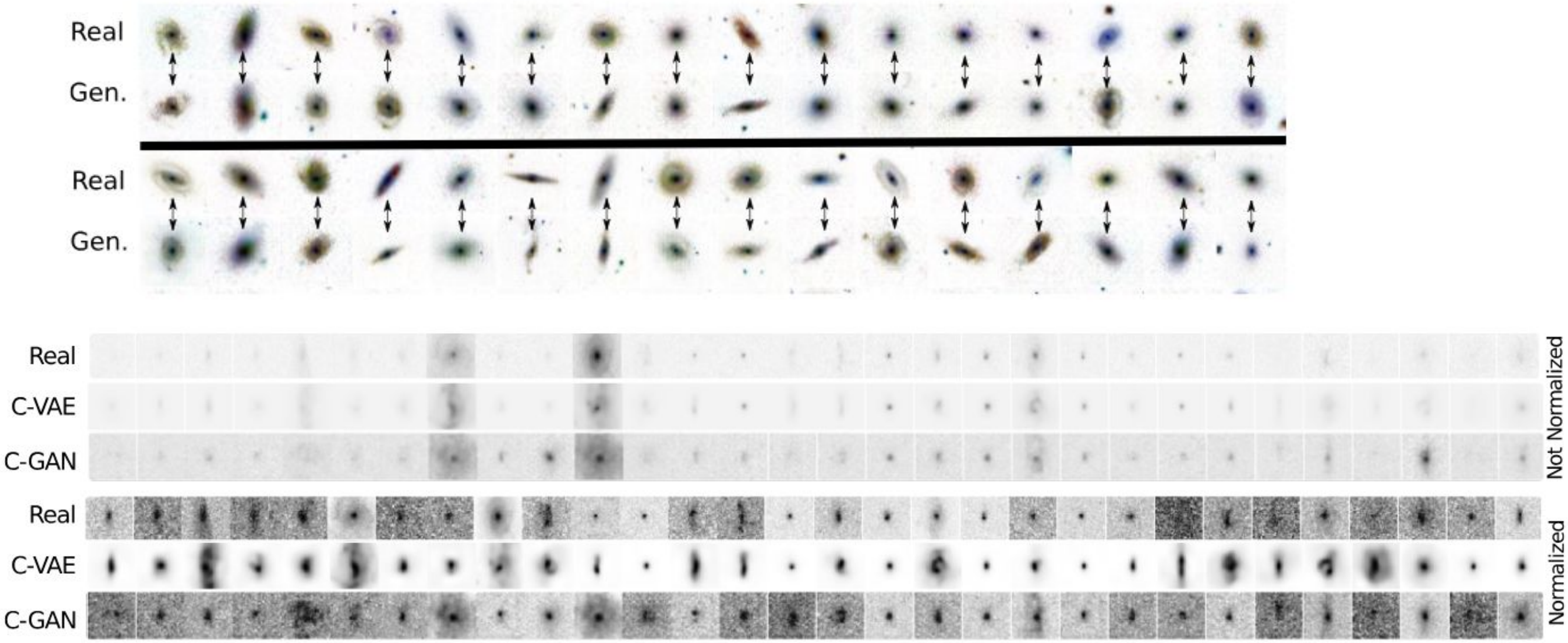




[Schawinski K., Turp M.~D., Zhang C., 2018, A&A, 616, L16]()

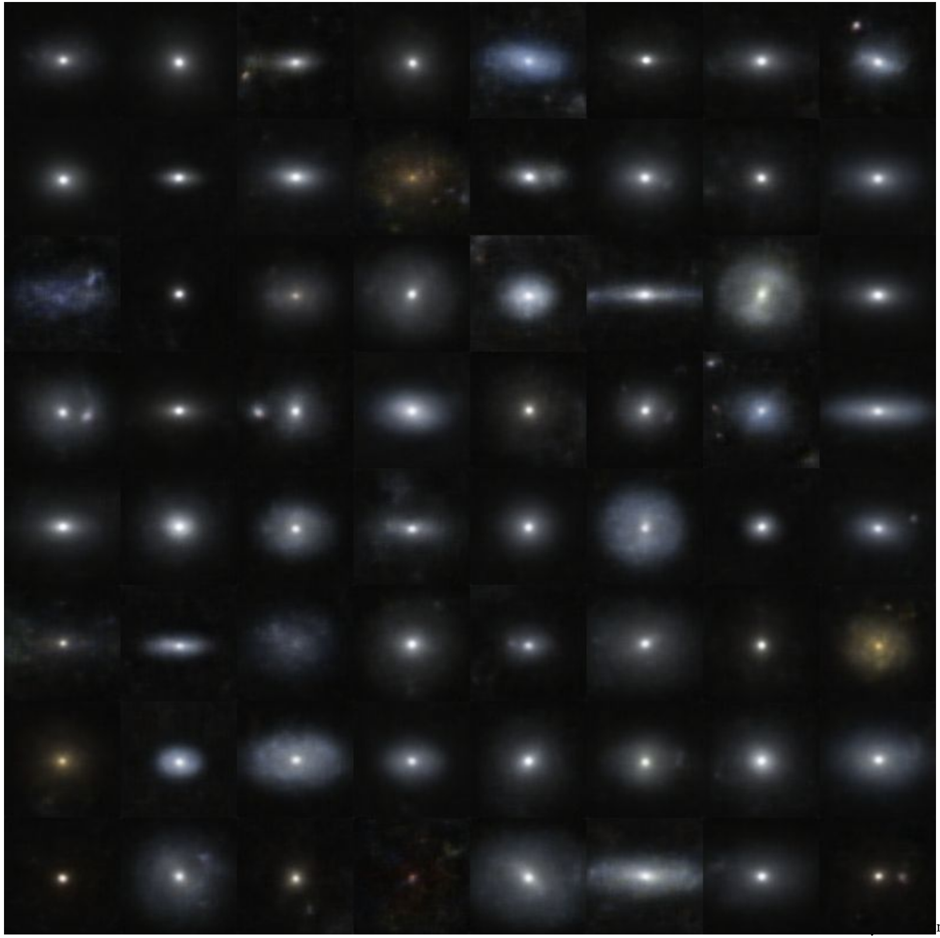
Fader Network: arXiv:1706.00409

GAN + VAE



[Ravanbakhsh S., Lanusse F., Mandelbaum R., Schneider J., Poczos B., 2016, arXiv e-prints, arXiv:1609.05796]

VAE Sampling example (no reconstruction)



We can generate samples from z, next step is can we constrain what’s being sampled?

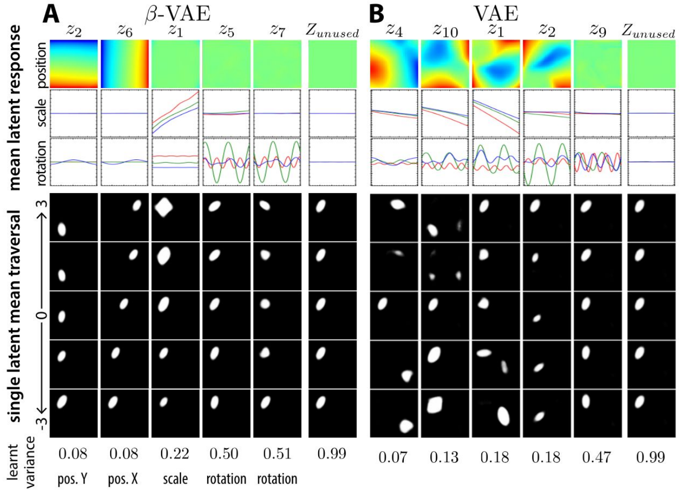
β-VAE: Disentangle VAE

In VAE we have some control over the latent space but every dimension in z can represent multiple-features (entanglement), how can we reduce the correlation between those dimensions?

$$Loss_{VAE} = -\mathbb{E}_q[\log p(x|z)] + \mathbb{KL}(q(z|x)||p(z))$$

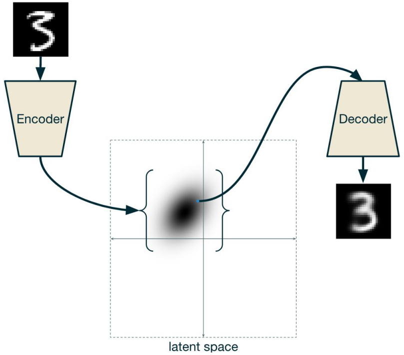
$$Loss_{\beta-VAE} = -\mathbb{E}_q[\log p(x|z)] + \beta\mathbb{KL}(q(z|x)||p(z))$$

Trade-off between reconstruction loss and penalty loss in q(z|x). → The higher β the higher is the independence of z-space component but less realistic. No labels just feature disentanglement.

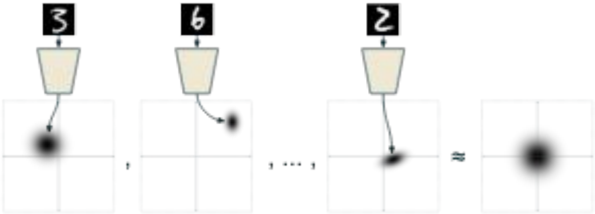


[beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework]

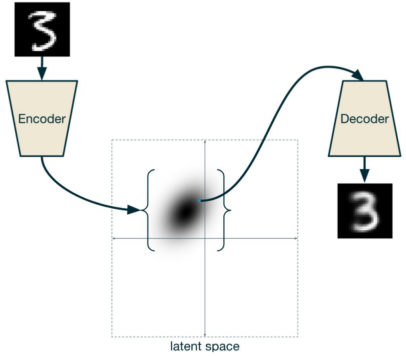
Conditional VAE: Adding labels



p(z) tend to look like this

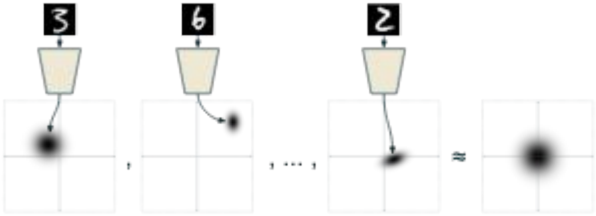


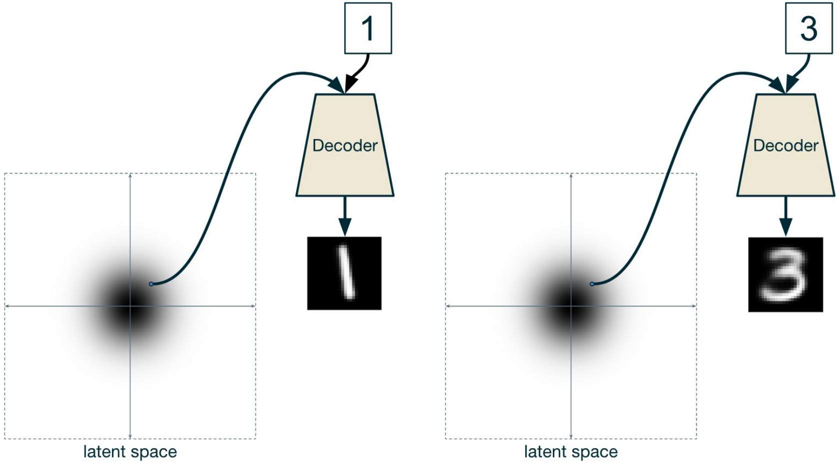
Conditional VAE: Adding labels



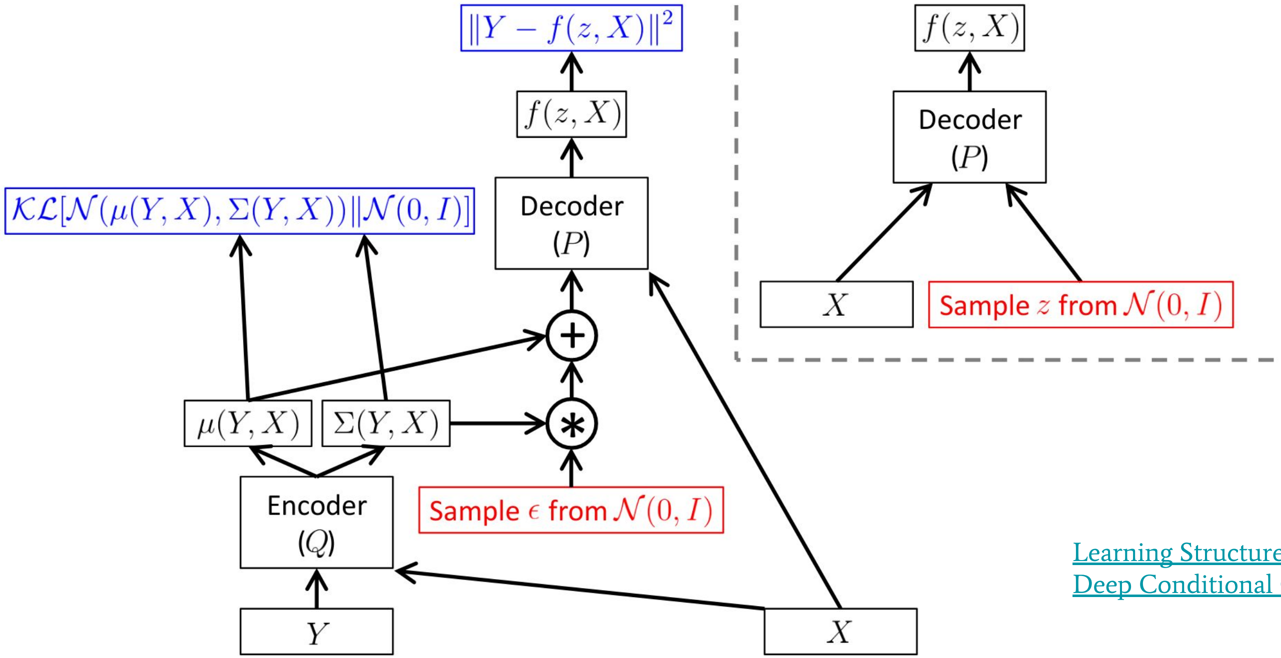
Conditional VAE tries to generate labeled samples. Labels are not used directly in training.

p(z) tend to look like this





CVAE: Adding labels



Y: Image
X: Label

[Learning Structured Output Representation using Deep Conditional Generative Models]

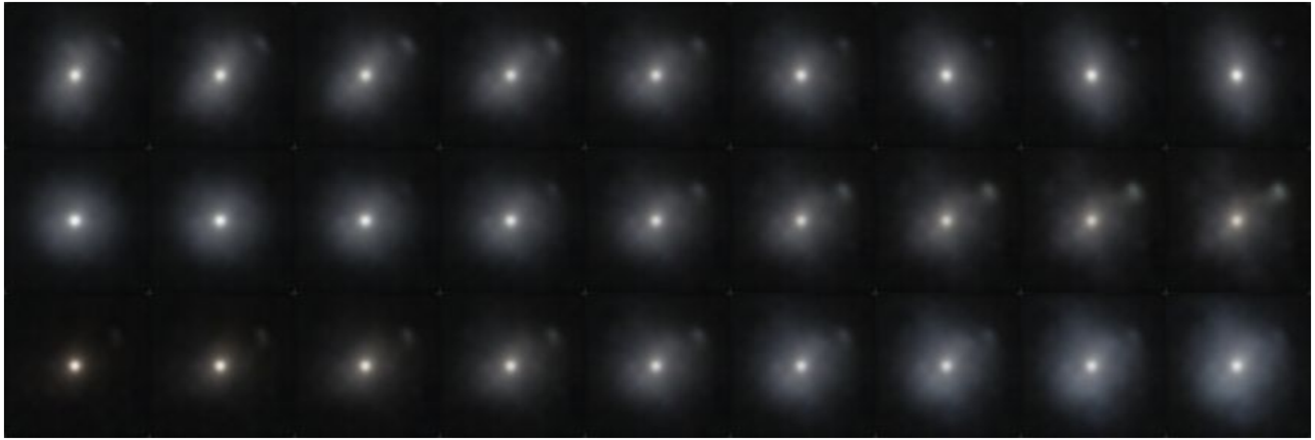
$$Loss_{CVAE} = -\mathbb{E}_q[\log p(x|z,c)] + \mathbb{KL}(q(z|x,c)||p(z,c))$$

CVAE: Examples

angle

eccentricity

brightness



angle

eccentricity

brightness

Bowen, Carrasco-Kind, et al. in prep.

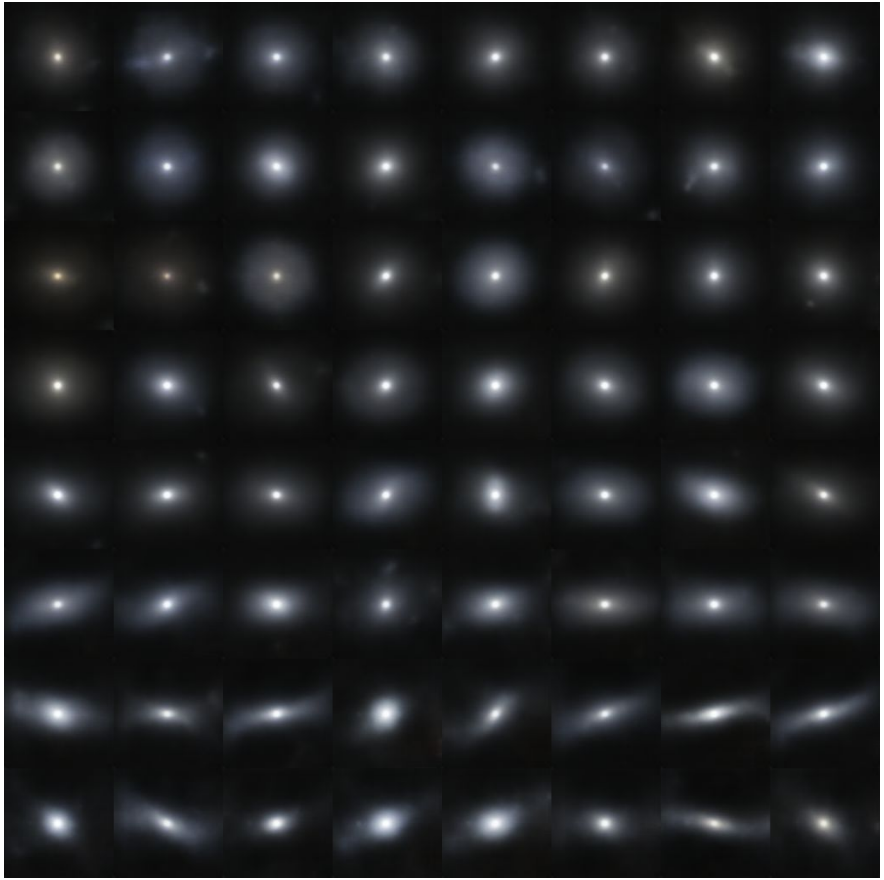
CVAE: Examples

25,000 galaxies for training
5,000 galaxies for validation

Sampling by concentration

Bowen, Carrasco-Kind, et al. in prep.

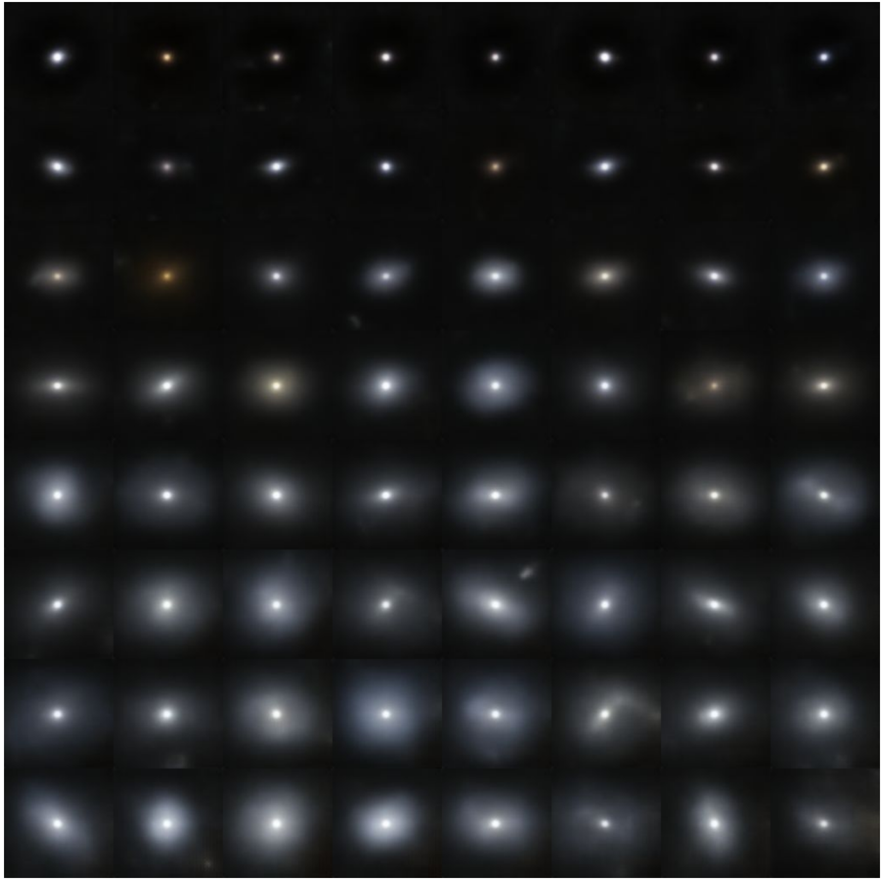
CVAE: Examples



Sampling by eccentricity

Bowen, Carrasco-Kind, et al. in prep.

CVAE: Examples



Sampling by area, not same as concentration

Bowen, Carrasco-Kind, et al. in prep.

Multimodal VAE: Training modalities

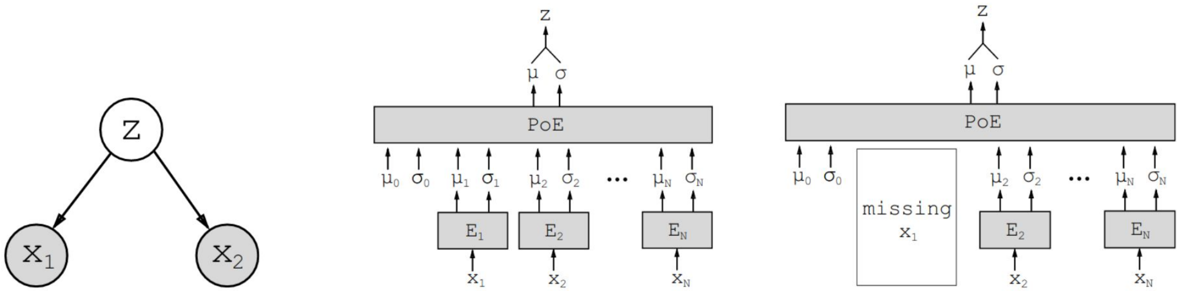
Multimodal Generative Models for Scalable Weakly-Supervised Learning

Mike Wu
Department of Computer Science
Stanford University
Stanford, CA 94025
wumike@stanford.edu

Noah Goodman
Departments of Computer Science and Psychology
Stanford University
Stanford, CA 94025
ngoodman@stanford.edu

[multimodal-generative-models-for-scalable-weakly-supervised-learning]()

Learning joint representation of conditionally independent modalities using product of experts.



(a) (b) (c)

Multimodal VAE: Training modalities

Multimodal Generative Models for Scalable Weakly-Supervised Learning

Mike Wu
Department of Computer Science
Stanford University
Stanford, CA 94025
wumike@stanford.edu

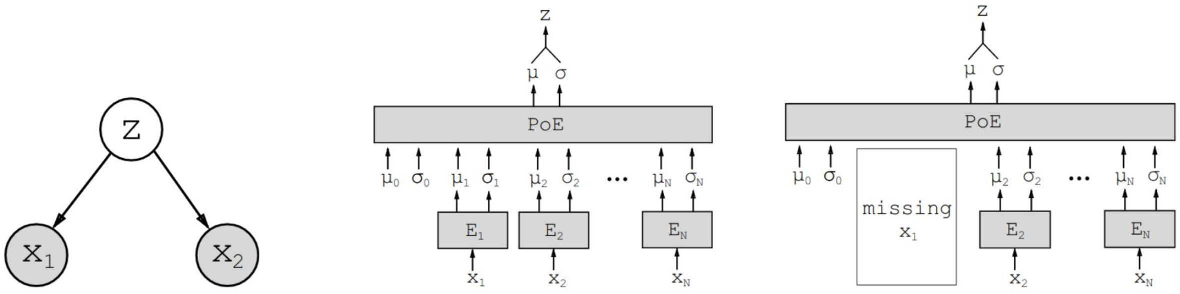
Noah Goodman
Departments of Computer Science and Psychology
Stanford University
Stanford, CA 94025
ngoodman@stanford.edu

multimodal-generative-models-for-scalable-weakly-supervised-learning

Learning joint representation of conditionally independent modalities using product of experts.

We can:

- Conditional sample with certain attributes
- Sample without any limitations
- Change the attribute of an existing input data
- Similarity search and anomaly detection
- Predict one modality from the others
- Sample and train with missing modalities



(a) (b) (c)

MVAE

Learning joint representation of conditionally independent modalities.

$$p(x_1, x_2, \ldots, x_n, z) = p(z)p(x_1|z)p(x_2|z)\cdots p(x_n|z)$$

$$\text{ELBO}(X) \triangleq \mathbb{E}_{q_\phi(z|X)}\Big[\sum_{x_i \in X} \lambda_i \log p_\theta(x_i|z)\Big] - \beta\, \text{KL}[q_\phi(z|X), p(z)]$$

Join probability

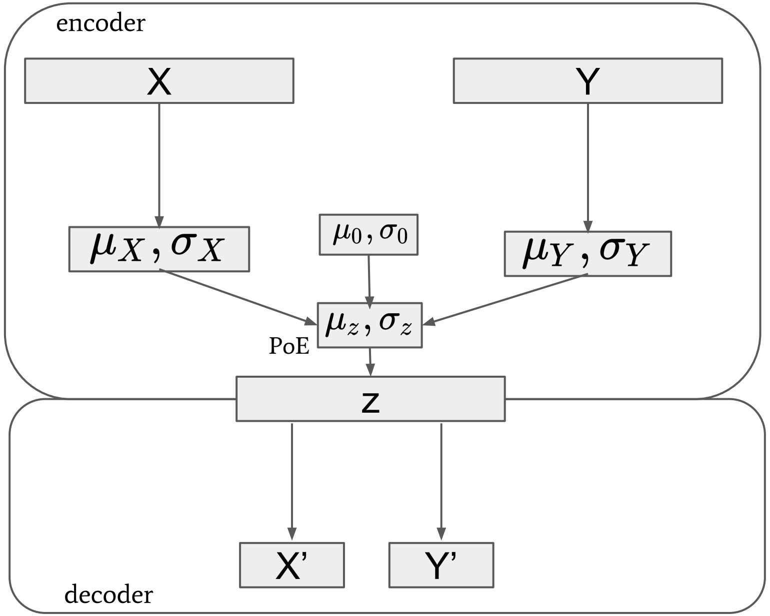
$$p(z|x_1, ..., x_N) \propto \frac{\prod_{i=1}^N p(z|x_i)}{\prod_{i=1}^{N-1} p(z)} \approx \frac{\prod_{i=1}^N [\tilde{q}(z|x_i)p(z)]}{\prod_{i=1}^{N-1} p(z)} = p(z) \prod_{i=1}^N \tilde{q}(z|x_i).$$

$$q(z|X) \propto p(z) \prod_{x_i \in X} \tilde{q}(z|x_i)$$

Evaluate ELBO for the whole sample, and for individual modalities

$$\text{ELBO}(x_1, ..., x_N) + \sum_{i=1}^N \text{ELBO}(x_i) + \sum_{j=1}^k \text{ELBO}(X_j)$$

MVAE 2 Modality Example



2 modalities example,
X are input images
Y are features (magnitude, area, concentration, etc...)
z is latent space

$$p(X, Y, z)$$

We can sample independently

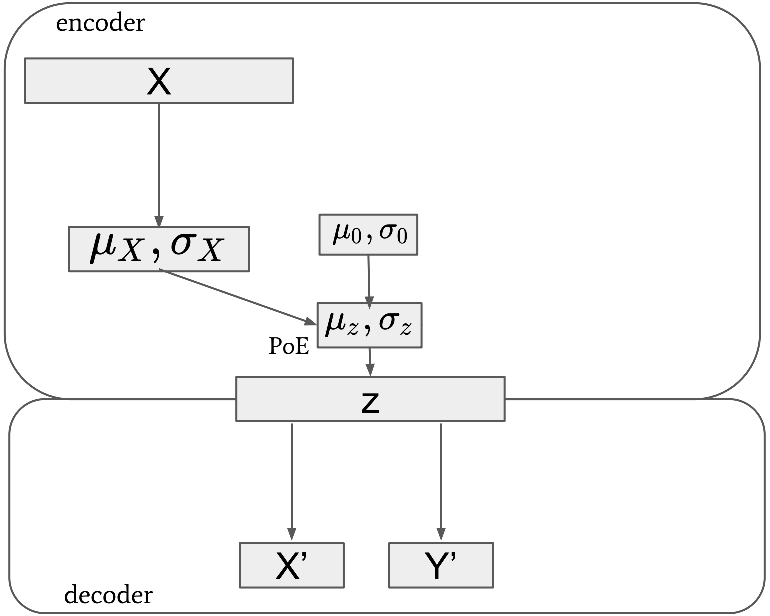
$$p(X|z) \quad p(Y|z) \quad p(X, Y|z)$$

$$p(z, X, Y) \propto p(z)\hat{q}(z|X)\hat{q}(z|Y)$$

$$ELBO(X, Y) + ELBO(X) + ELBO(Y)$$

$$\mathrm{ELBO}(X) \triangleq \mathbb{E}_{q_\phi(z|X)}[\sum_{x_i \in X} \lambda_i \log p_\theta(x_i|z)] - \beta\, \mathrm{KL}[q_\phi(z|X), p(z)]$$

MVAE 2 Modality Example



2 modalities example,
X are input images
Y are features (magnitude, area, concentration, etc...)
z is latent space

$$p(X, Y, z)$$

We can sample independently

$$p(X|z) \quad p(Y|z) \quad p(X, Y|z)$$

$$p(z, X, Y) \propto p(z)\hat{q}(z|X)\hat{q}(z|Y)$$

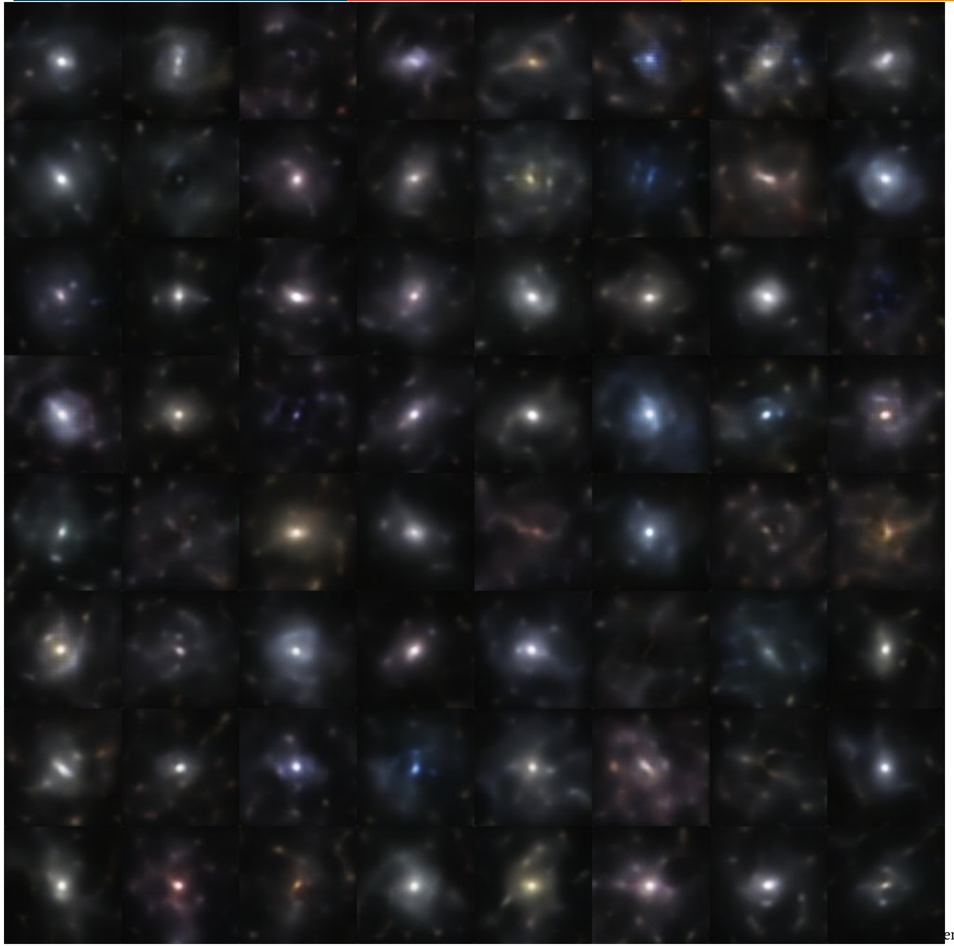
$$ELBO(X,Y) + ELBO(X) + ELBO(Y)$$

$$\text{ELBO}(X) \triangleq \mathbb{E}_{q_\phi(z|X)}\Big[\sum_{x_i \in X} \lambda_i \log p_\theta(x_i|z)\Big] - \beta \, \text{KL}[q_\phi(z|X), p(z)]$$

MVAE: Examples

Sampling for a fixed image

MVAE: Examples



Unconstrained sampling, very small β. Lots of structure, less control over modalities

$$\mathrm{ELBO}(X) \triangleq \mathbb{E}_{q_\phi(z|X)}\Big[\sum_{x_i \in X} \lambda_i \log p_\theta(x_i|z)\Big] - \beta\, \mathrm{KL}[q_\phi(z|X), p(z)]$$

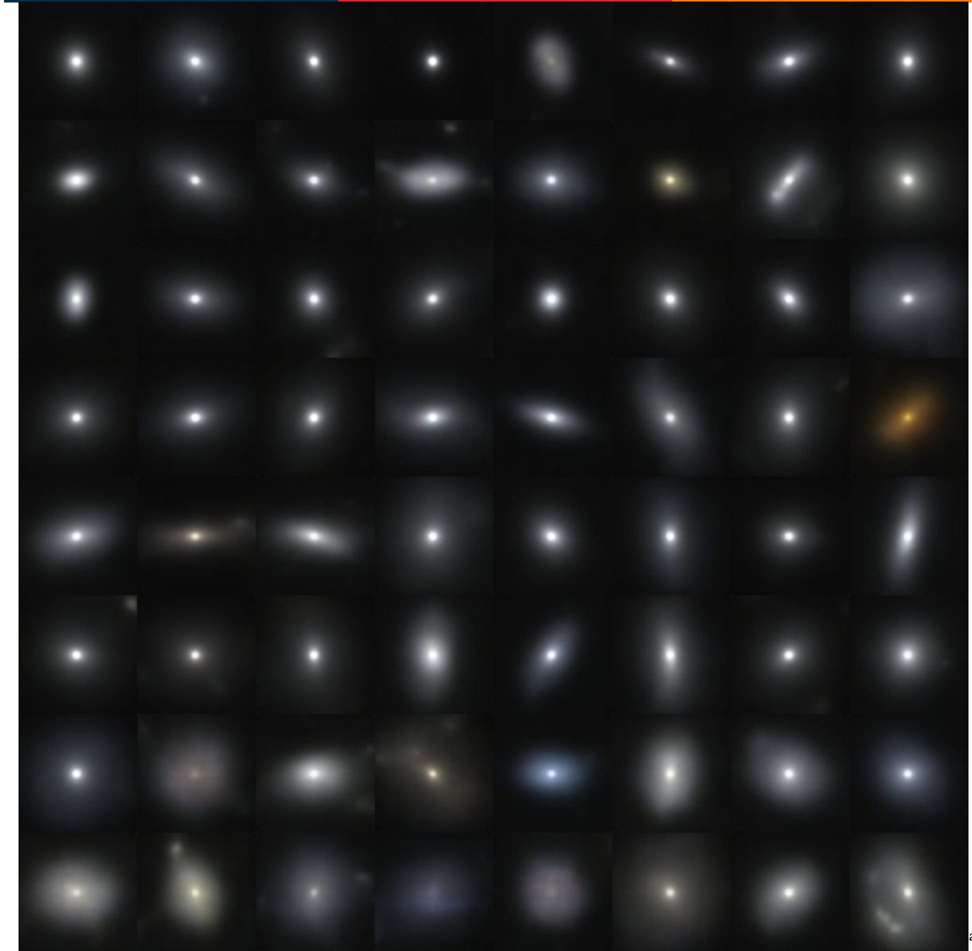
MVAE: Examples

Fix galaxy, sampling over features.

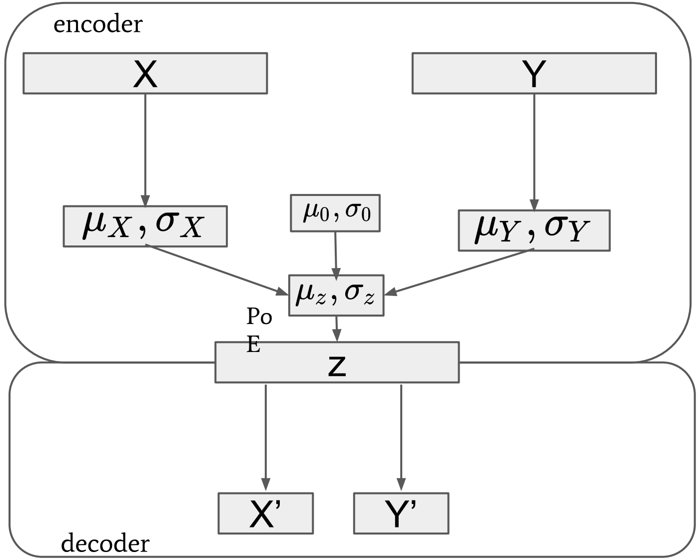
MAG_R, MAG_I, ECC, Concentration, Angle, Area

Bowen, Carrasco-Kind, et al. in prep.

MVAE: Examples



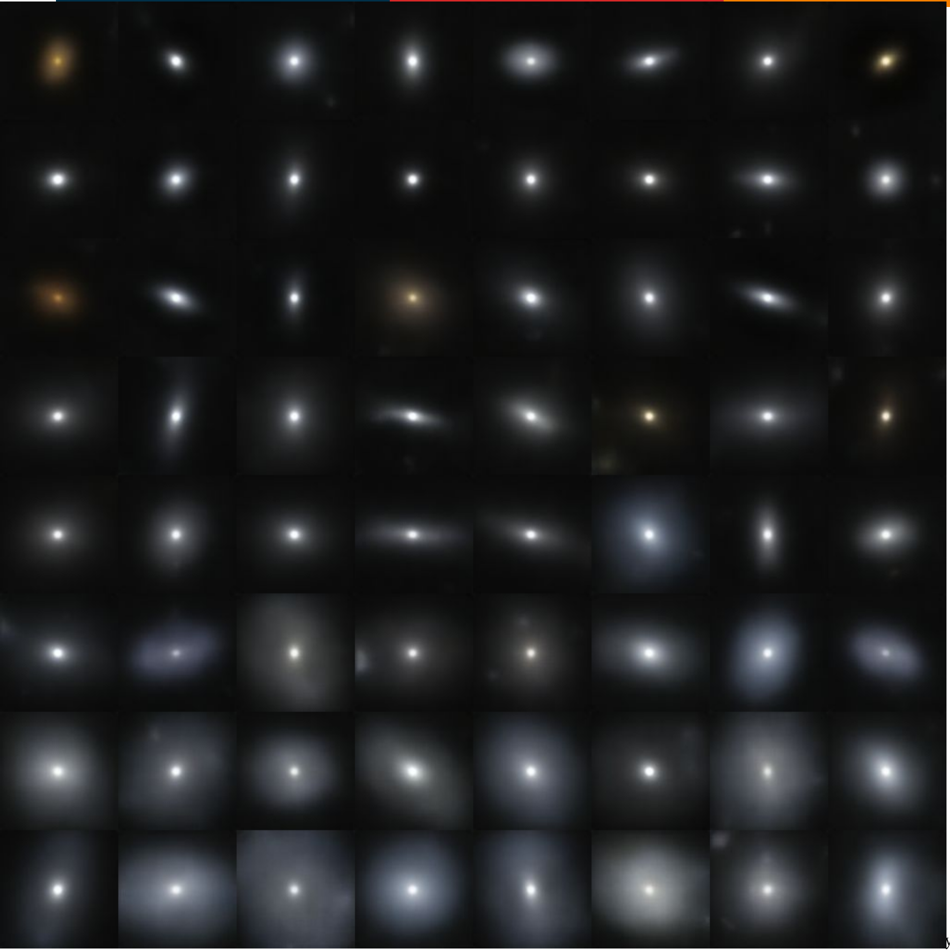
Samples with changing concentration (increasing downwards)



Bowen, Carrasco-Kind, et al. in prep.

ar

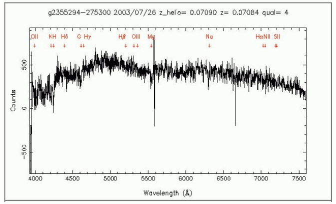
MVAE: Examples

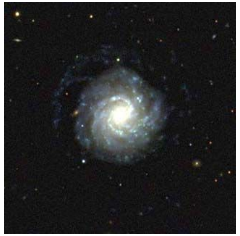


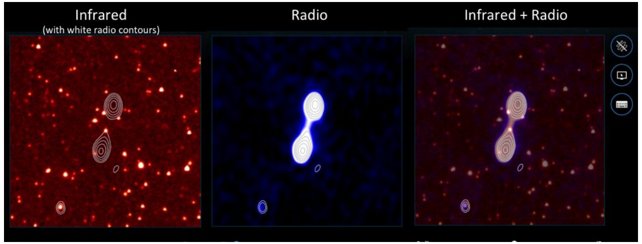
Samples with changing area
(increasing downwards)

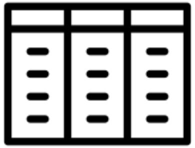
Bowen, Carrasco-Kind, et al. in prep.

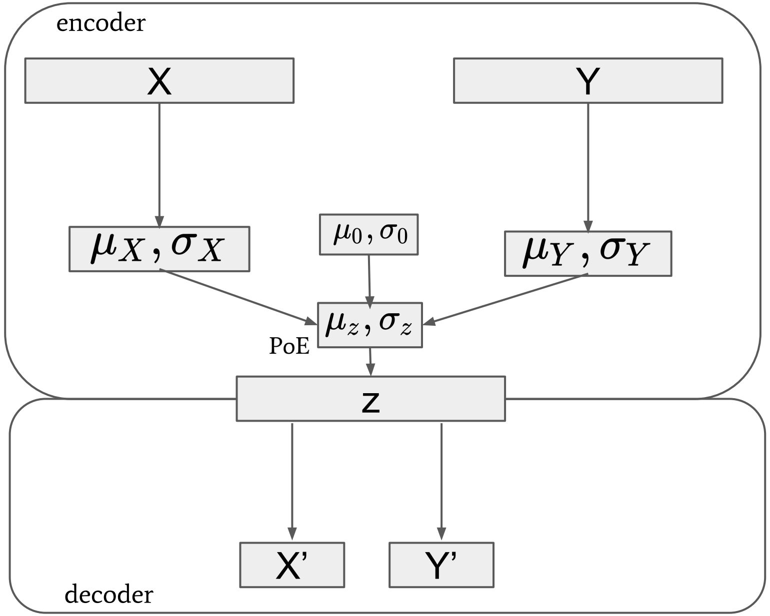
MVAE: It opens interesting options







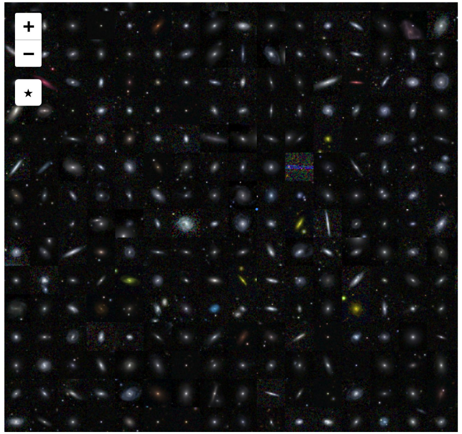


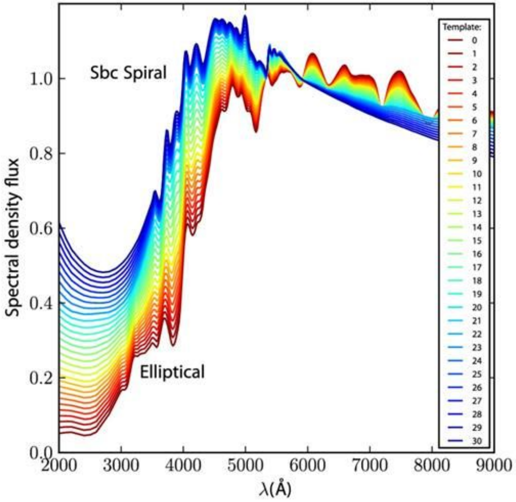


Bowen, Carrasco-Kind, et al. in prep.

Conclusions / Potential open projects

- Use VAE for data/image compression and similarity search of galaxy images for comparative studies
  - Elliptical stacking for diffuse extended halo S/N
  - Modern Hubble sequence and census
  - Search of similar galaxies and anomalies
  - Denoising/deconvolving
- Use CVAE/MVAE and TL to generate sampling of realistic galaxy images conditioned to diverse priors
  - Create a uniform training sample for photo-z's
  - Break color/redshift degeneracy
  - Conditioning on redshift, template, brightness,
  - Galaxy evolutions studies
  - Sampling procedures for galaxy images
  - Combine multiple modalities
- Develop photo-z compression techniques considering information from galaxy images
- Anomaly detection with extended Isolation Forest
- There is plans to classify DES images (prototype)



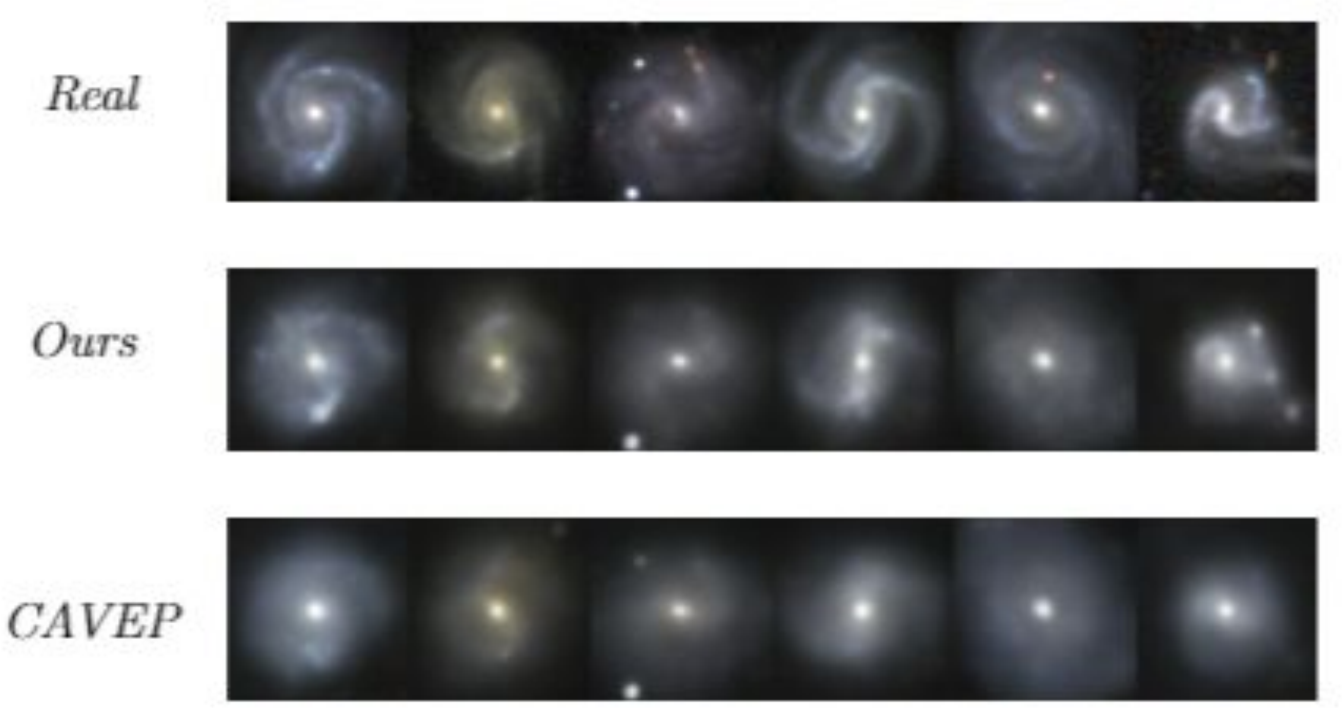


Thank you!

Questions?

Matias Carrasco Kind -- NCSA
mcarras2@illinois.edu
github.com/mgkind
matias-ck.com

Create realistic samplings of galaxy images with no prior (DES)



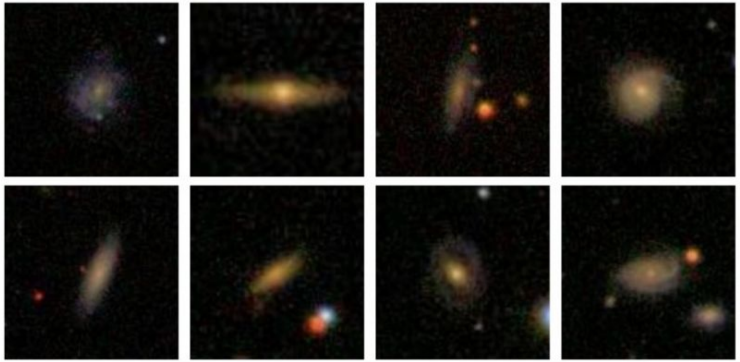
Bowen, Carrasco-Kind, et al. in prep.

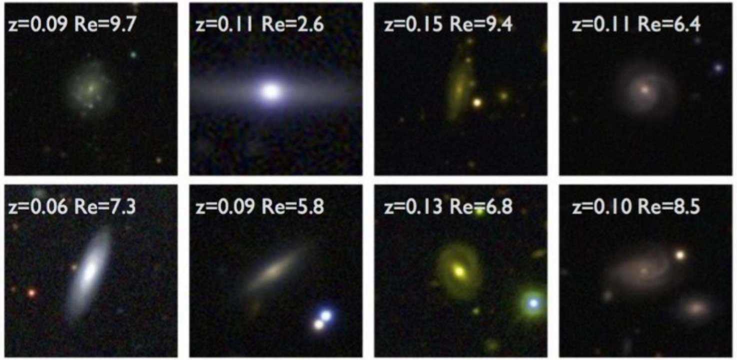
Learn from one survey to another

Knowledge transfer of Deep Learning for galaxy morphology from one survey to another

H. Domínguez Sánchez[1★], M. Huertas-Company [1,2,3], M. Bernardi [1], S. Kaviraj[4], J.L. Fischer[1], T. M. C. Abbott[5], F. B. Abdalla[6,7], J. Annis[8], S. Avila[9], D. Brooks[6], E. Buckley-Geer[8], A. Carnero Rosell[10,11], M. Carrasco Kind[12,13], J. Carretero[14], C. E. Cunha[15], C. B. D'Andrea[1], L. N. da Costa[10,11], C. Davis[15], J. De Vicente[16], P. Doel[6], A. E. Evrard[17,18], P. Fosalba[19,20], J. Frieman[8,21], J. García-Bellido[22], E. Gaztanaga[19,20], D. W. Gerdes[17,18], D. Gruen[15,23], R. A. Gruendl[12,13], J. Gschwend[10,11], G. Gutierrez[8], W. G. Hartley[6,24], D. L. Hollowood[25], K. Honscheid[26,27], B. Hoyle[28,29], D. J. James[30], K. Kuehn[31], N. Kuropatkin[8], O. Lahav[6], M. A. G. Maia[10,11], M. March[1], P. Melchior[32], F. Menanteau[12,13], R. Miquel[14,33], B. Nord[8], A. A. Plazas[34], E. Sanchez[16], V. Scarpine[8], R. Schindler[23], M. Schubnell[18], M. Smith[35], R. C. Smith[5], M. Soares-Santos[36], F. Sobreira[37,10], E. Suchyta[39], M. E. C. Swanson[13], G. Tarle[18], D. Thomas[9], A. R. Walker[5], and J. Zuntz[40]

SDSS (GalaxyZoo) transfer learning to DES images for morphological classification





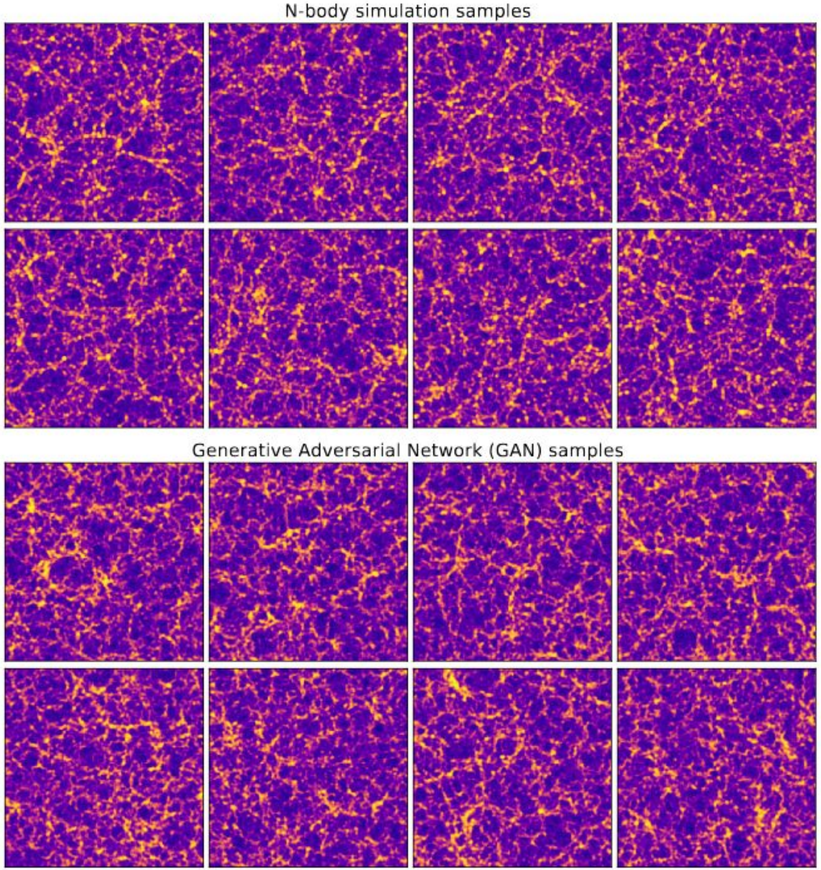
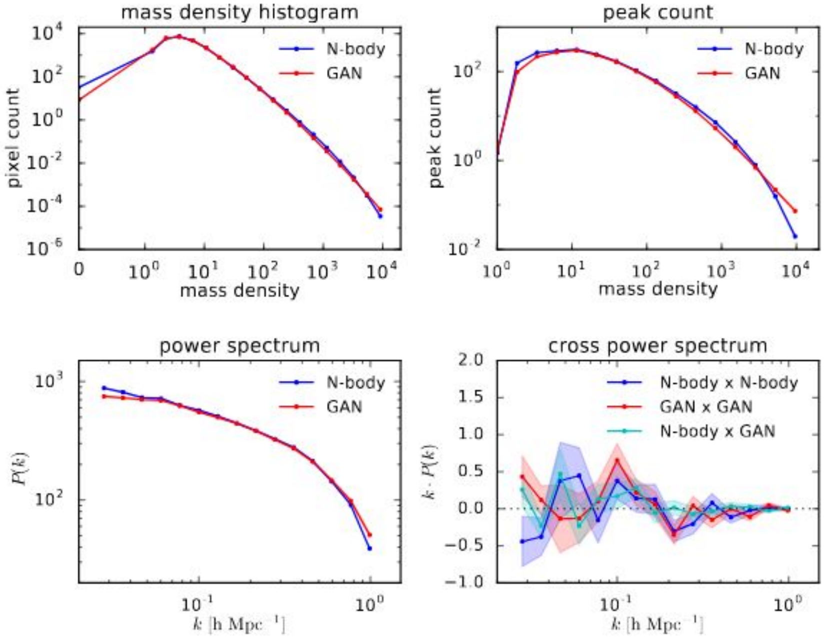
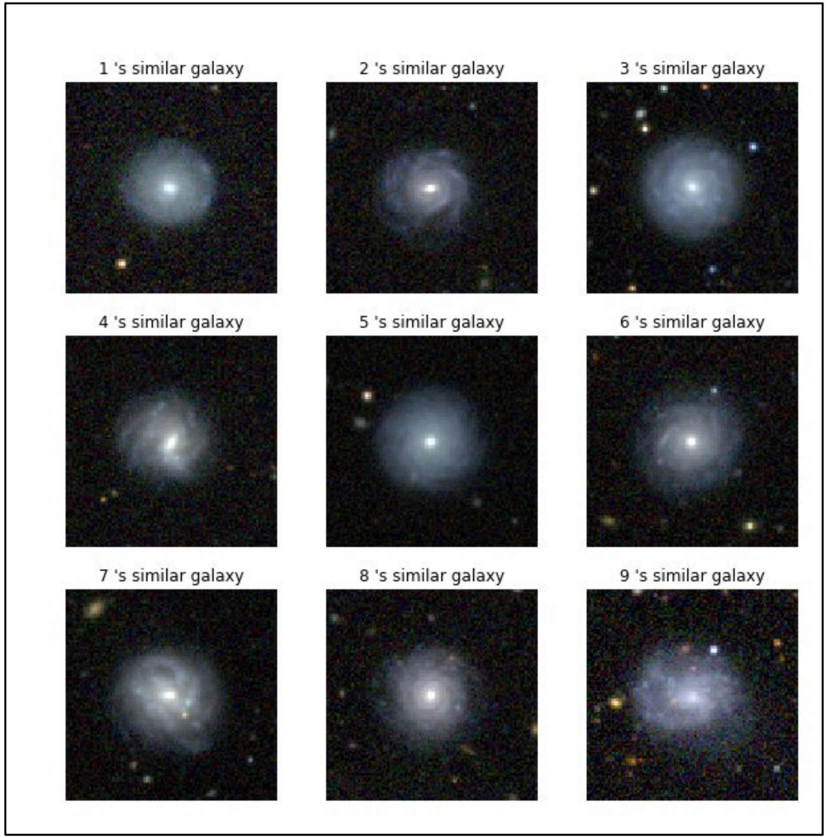


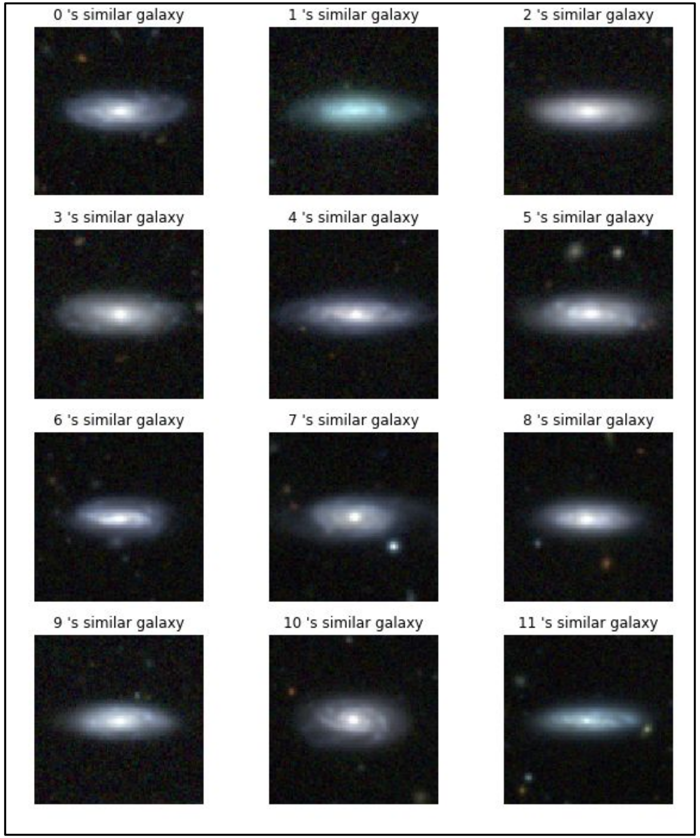
Figure 1: Samples from N-body simulation and from GAN for the box size of 500 Mpc. Note that the transformation in Equation 3.1 with $a = 20$ was applied to the images shown above for better clarity.



Rodriguez A.C., et al., 2018, Computational Astrophysics and Cosmology, 5, 4

Galaxy selection and similarity search using VAE → DES



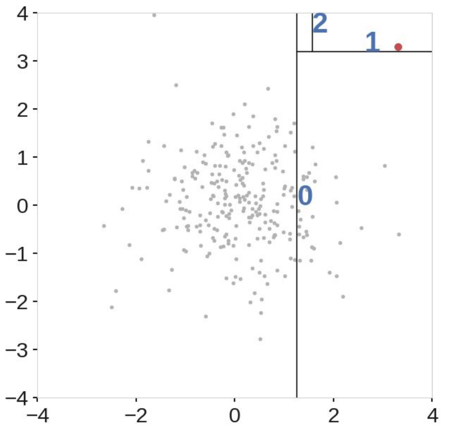


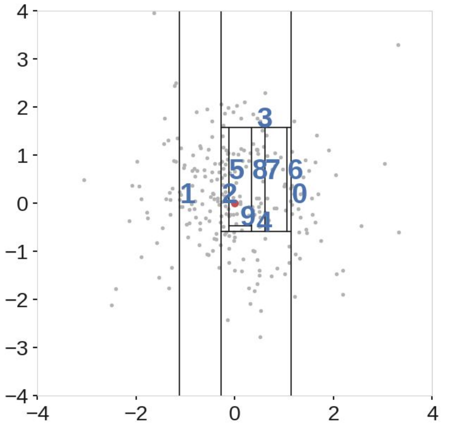
Bowen, Carrasco-Kind, et al. in prep.

Anomaly Detection with Isolation Forest

- Few and different to be isolated quicker
- For each tree:
  - Get a sample of the data
  - Randomly select a dimension
  - Randomly pick a value in that dimension
  - Draw a straight line through the data at that value and split data
  - Repeat until tree is complete
- Generate multiple trees → forest
- Anomalies will be isolated in only a few steps
- Nominal points in more

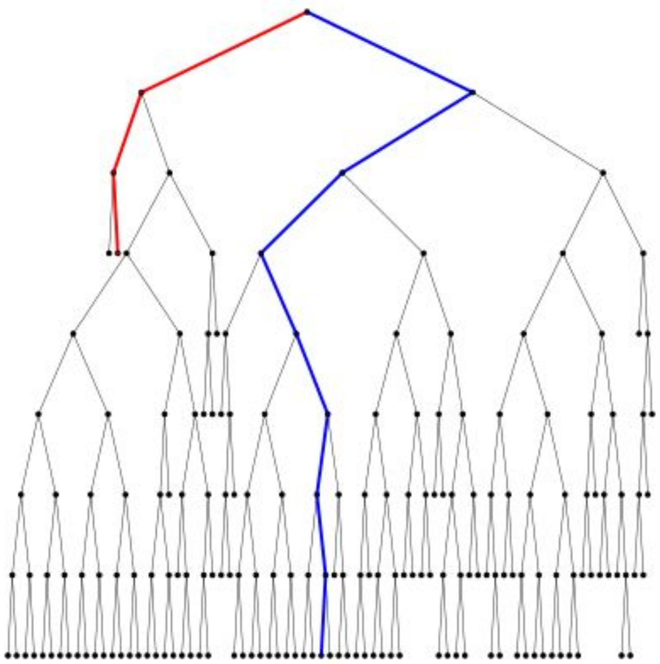
- To score points:
  - Run point down tree, record path
  - Repeat for each tree, aggregate scores $s(x,n) = 2^{-\frac{E(h(x))}{c(n)}}$
  - Score distribution



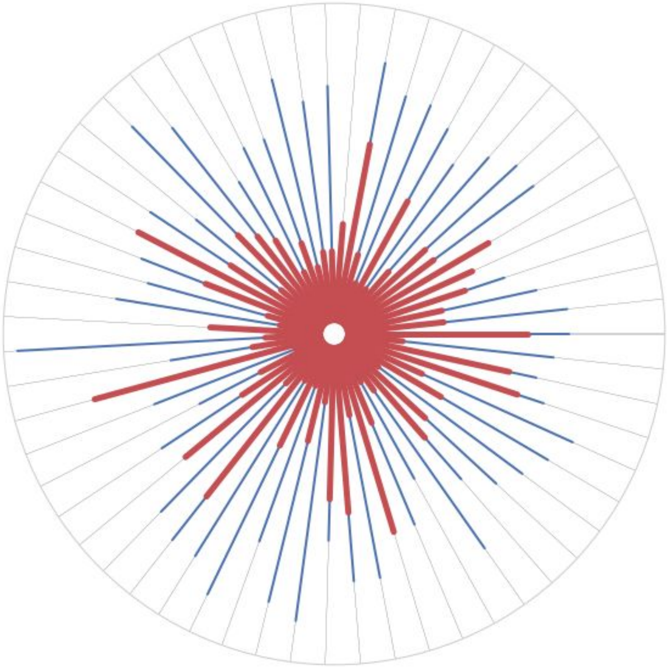


Anomaly Detection with Isolation Forest

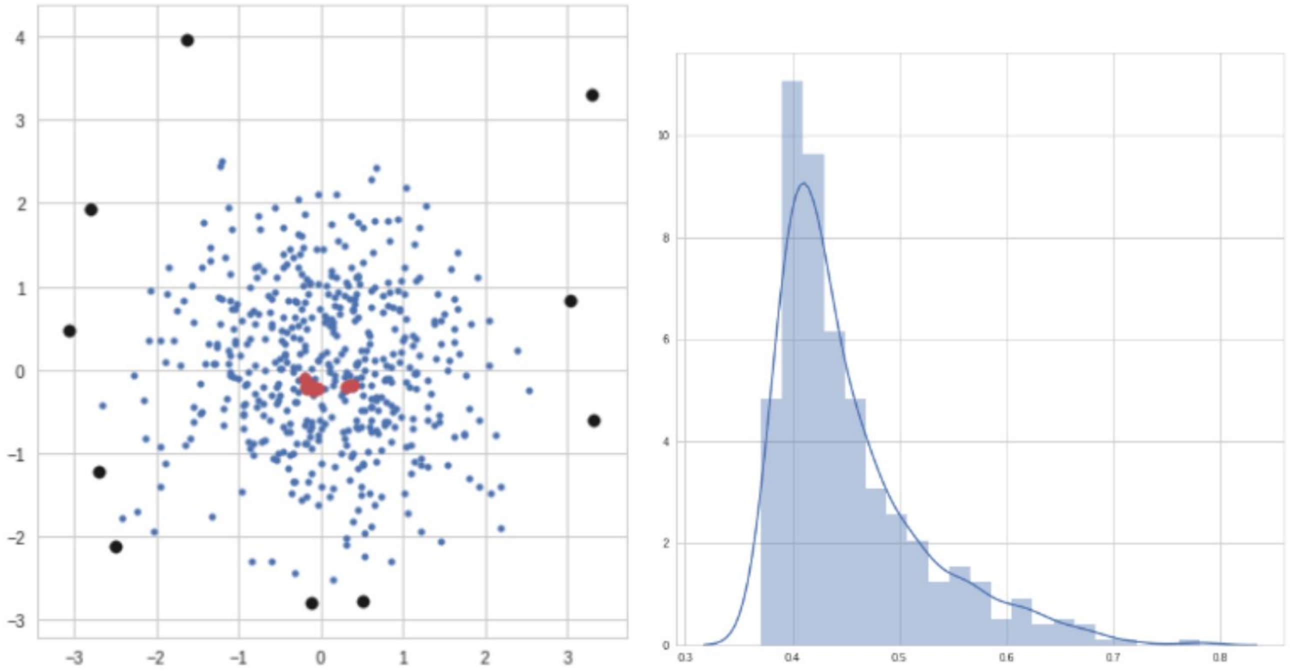
Single Tree scores for anomaly and nominal points



Forest plotted radially. Scores for anomaly and nominal shown as lines



Anomaly Detection with Isolation Forest



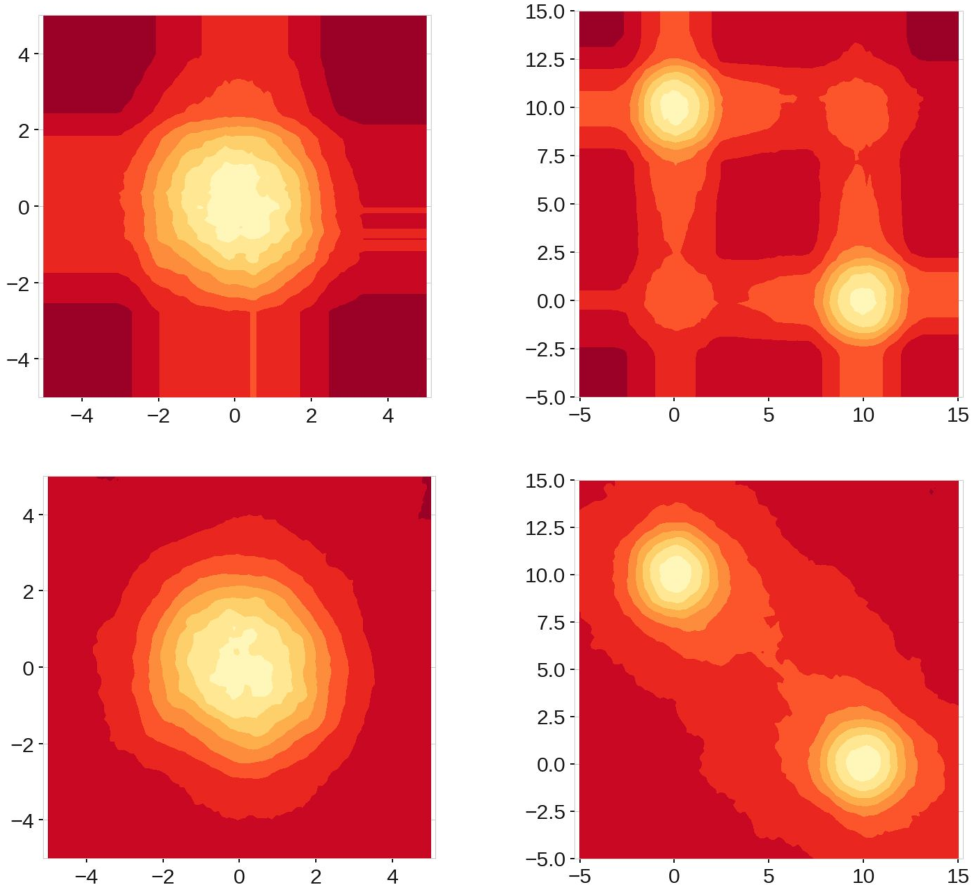
Anomaly Detection with Extended Isolation Forest

Isolation Forest:

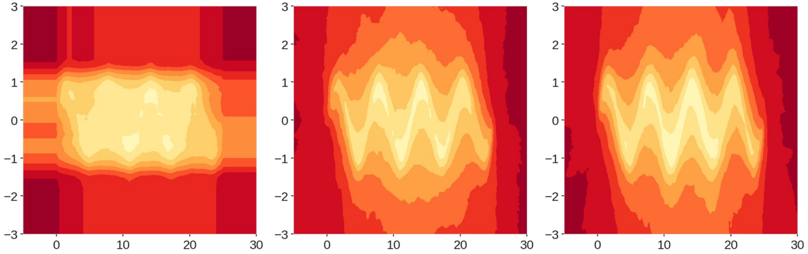
- ✓ Model free
- ✓ Computationally efficient
- ✓ Readily applicable to parallelization
- ✓ Readily application to high dimensional data
- ✗ Inconsistent scoring seen in score maps

Extended Isolation Forest:

- ✓ Model free
- ✓ Computationally efficient
- ✓ Readily applicable to parallelization
- ✓ Readily application to high dimensional data
- ✓ Consistent scoring



Anomaly Detection with Extended Isolation Forest



(a) Standard IF (b) Rotated IF (c) Extended IF

Algorithm 2 $iTree(X, e, l)$

Require: X - input data, e - current tree height, l - height limit

Ensure: an iTree

1: **if** $e \geq l$ or $|X| \leq 1$ **then**
2: **return** $exNode\{Size \leftarrow |X|\}$
3: **else**
4: randomly select a normal vector $n \in \mathbb{R}^{|X|}$ by drawing each coordinate of $\vec{n}$ from a uniform distribution.
5: randomly select an intercept point $p \in \mathbb{R}^{|X|}$ in the range of X
6: set coordinates of n to zero according to extension level
7: $X_l \leftarrow filter(X, (X - p) \cdot n \leq 0)$
8: $X_r \leftarrow filter(X, (X - p) \cdot n > 0)$
9: **return** inNode{ $Left \leftarrow iTree(X_l, e + 1, l)$,
$Right \leftarrow iTree(X_r, e + 1, l)$,
$Normal \leftarrow n$,
$Intercept \leftarrow p\}$
10: **end if**

Hariri, Carrasco Kind & Brunner, 2018