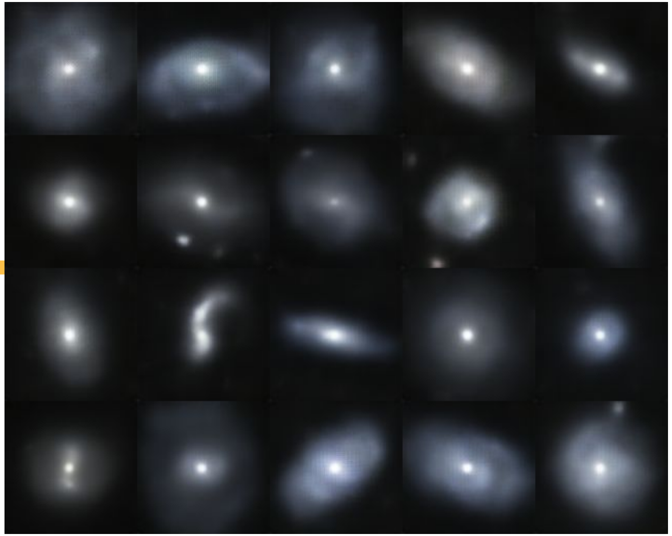
This galaxy does not exist

Matias Carrasco Kind, NCSA & Astronomy



Astrofest, April 23rd, 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

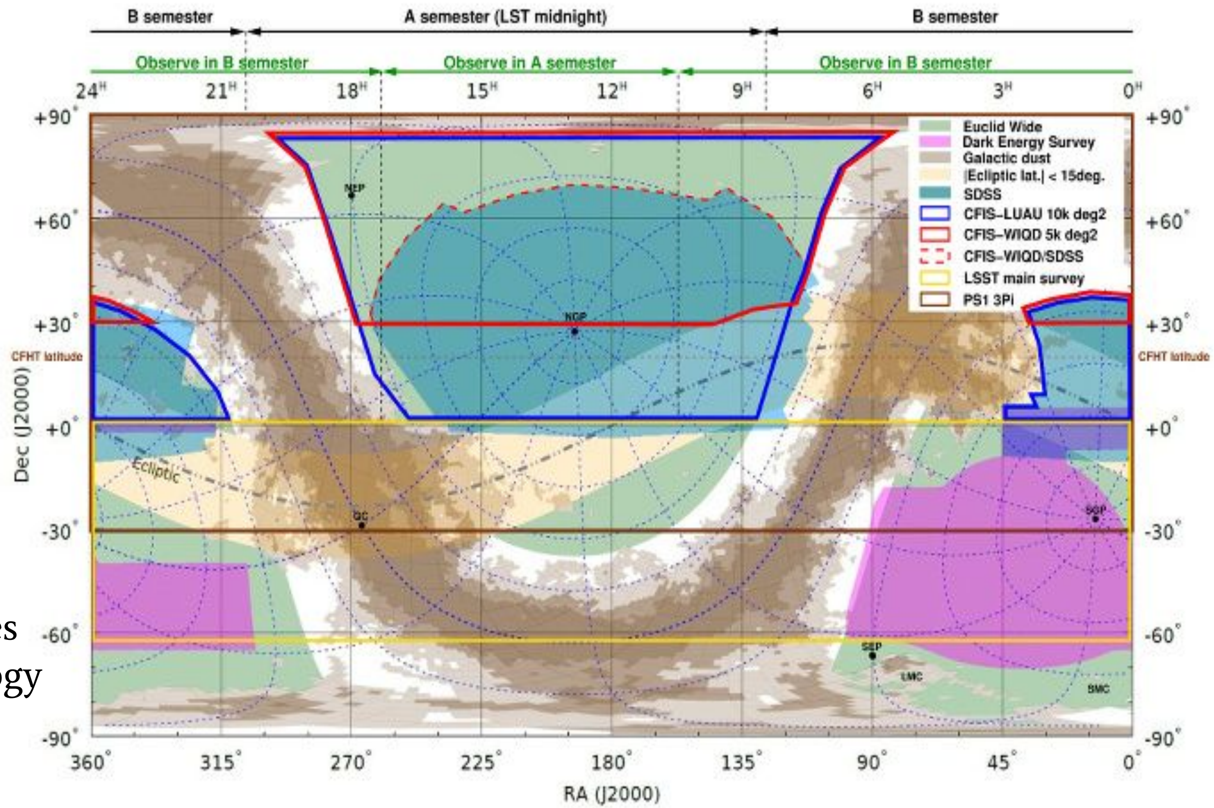
Introduction

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

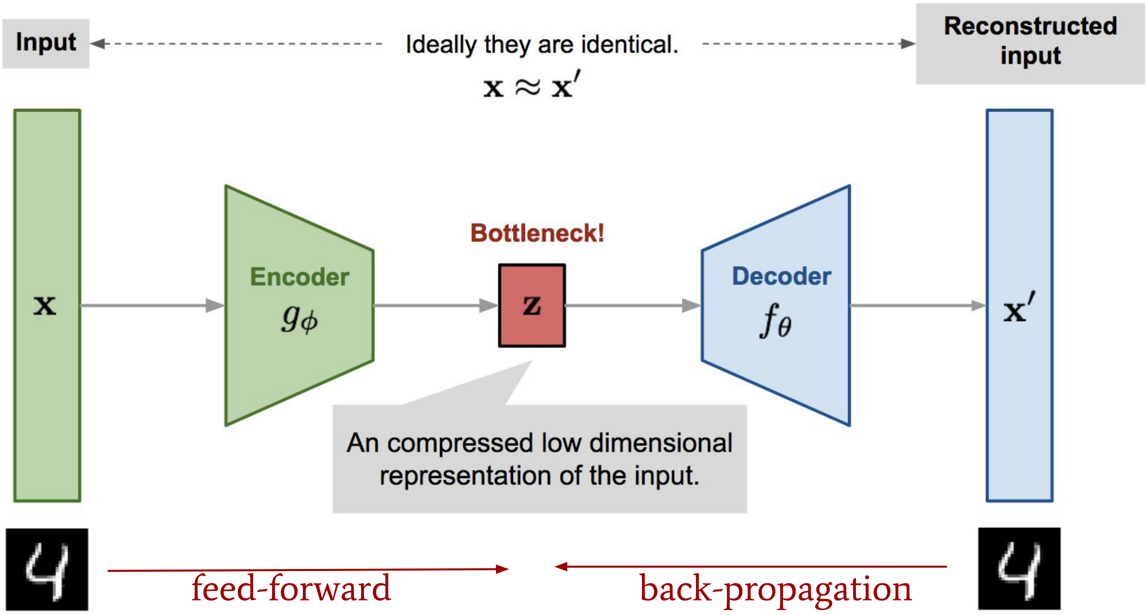
- For galaxies with i < 17,
  DR1 ~ 60K galaxies and
  LSST DR1 ~ 250K

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

- LSST overlaps DES completely
- Error estimation, denoising, deconv.
- Sampling technique, training samples
- Photo-z still only option for cosmology
- Big data problem → compresion techniques
- Galaxy Census and anomaly detection



Autoencoders



$$z = g_\phi(x)$$
$$x' = f_\theta(g_\phi(x))$$
$$\mathcal{L}(x, x') + reg.$$

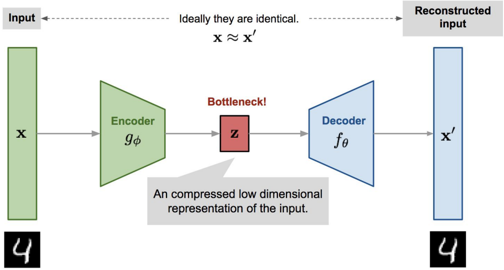
$$\ell_1 = \lambda \sum_i |a_i^{(h)}|$$

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

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

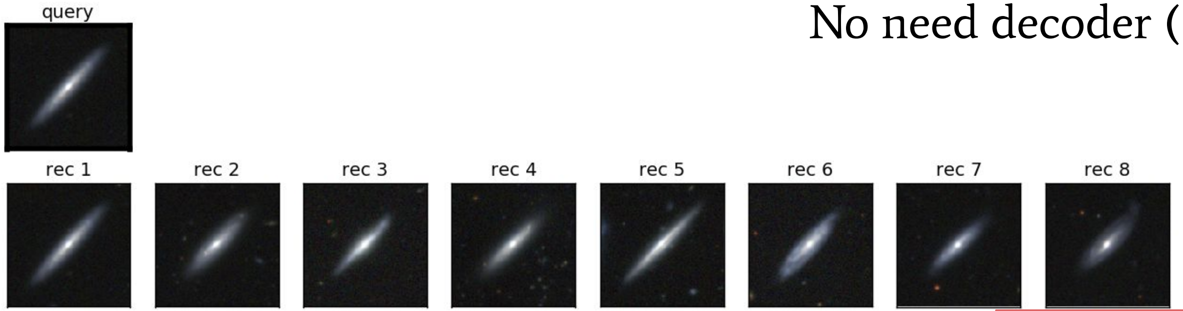
Many variants: Sparse AE, Contractive AE, Stacked AE, etc...

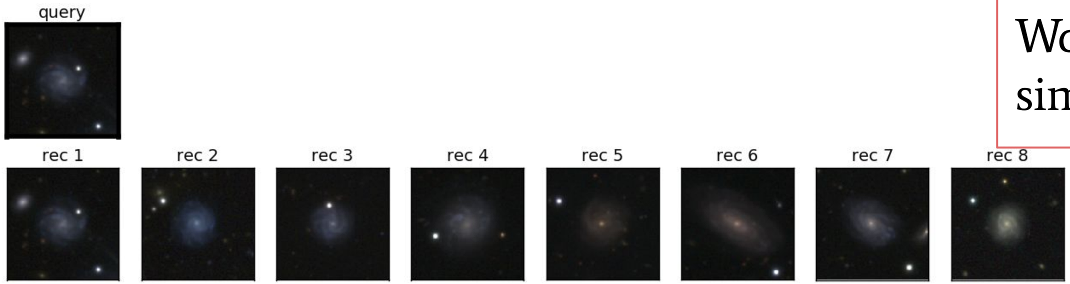
Autoencoders Applications: Image compression



Compress images from 220x220x3 pixels to 50-vector (2000x), for fast similarity search, anomaly detection, etc…

No need decoder (only for Loss)





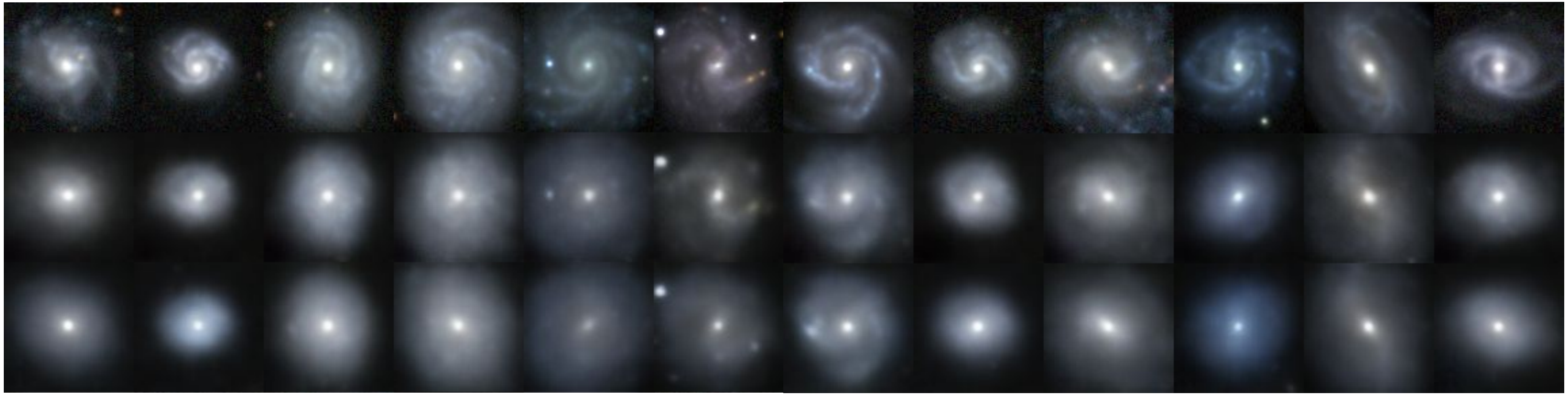
Working on a service for image similarity ranking for DES images

Autoencoders Applications: Reconstruction

Original

Model 1

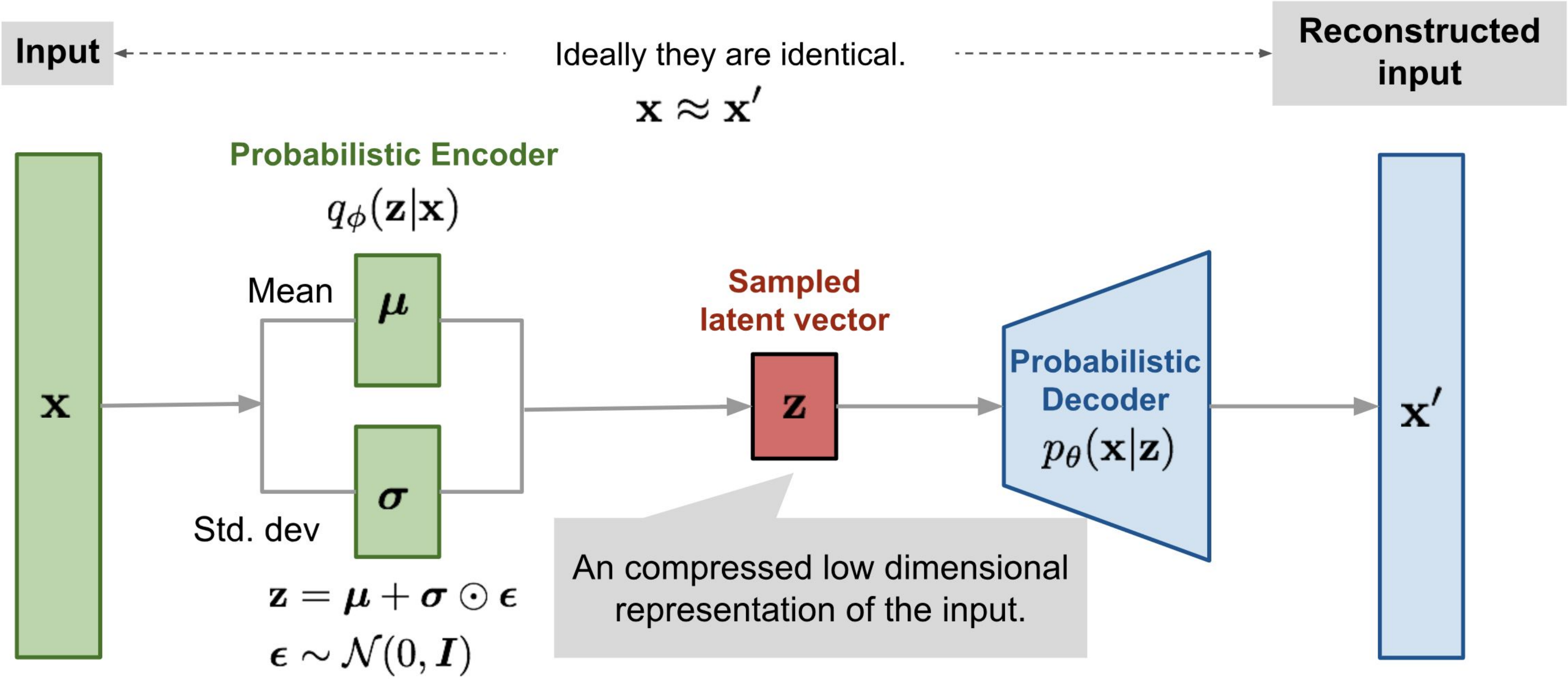
Model 2



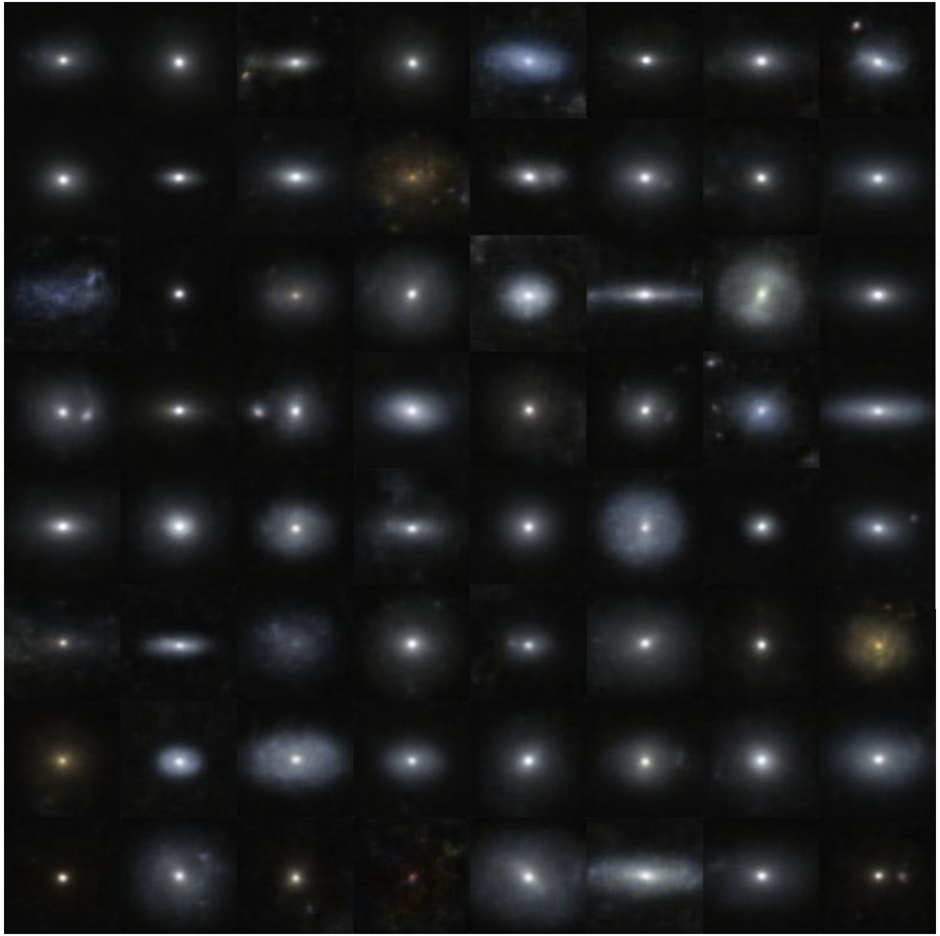
Blurry images and structure is lost, but angular sizes, radial profiles and brightness are a match.

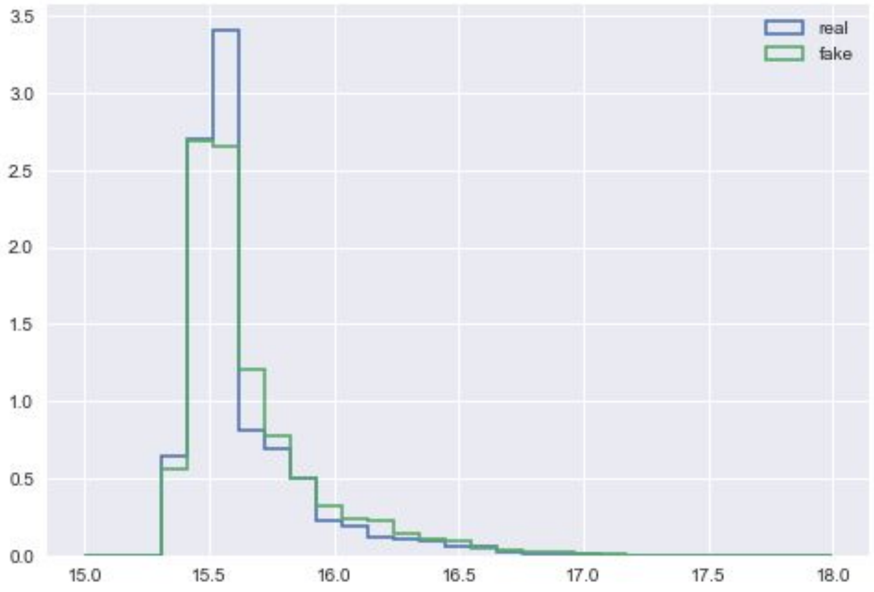
What if we can make the model learn properties at the same time as images.? What if can sample from the latent space?

Summarizing VAE basics, more complex models built on top



VAE Sampling example (no reconstruction)





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

Multimodal VAE: Training modalities

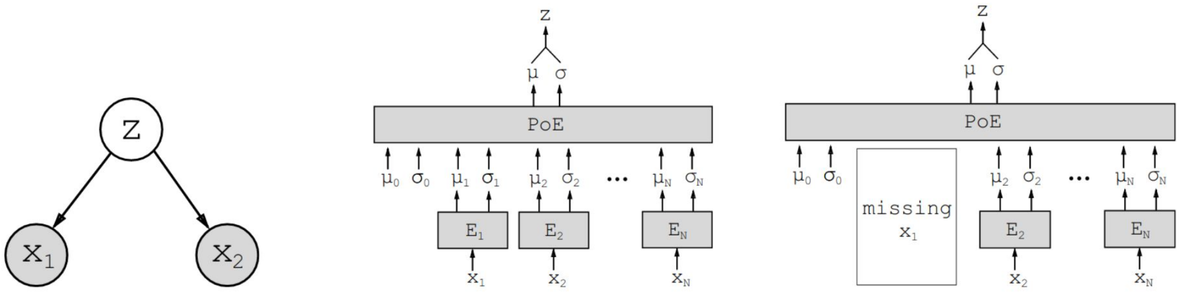
Multimodal Generative Models for Scalable Weakly-Supervised Learning

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.

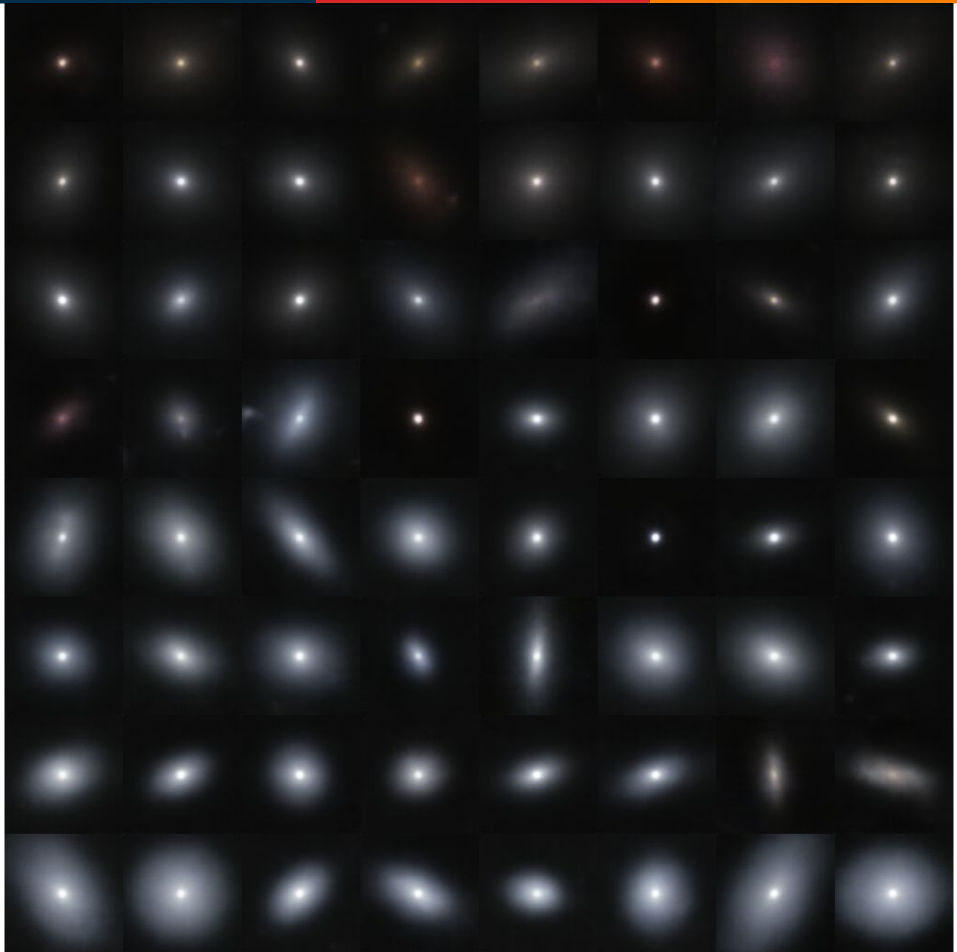


(a) (b) (c)

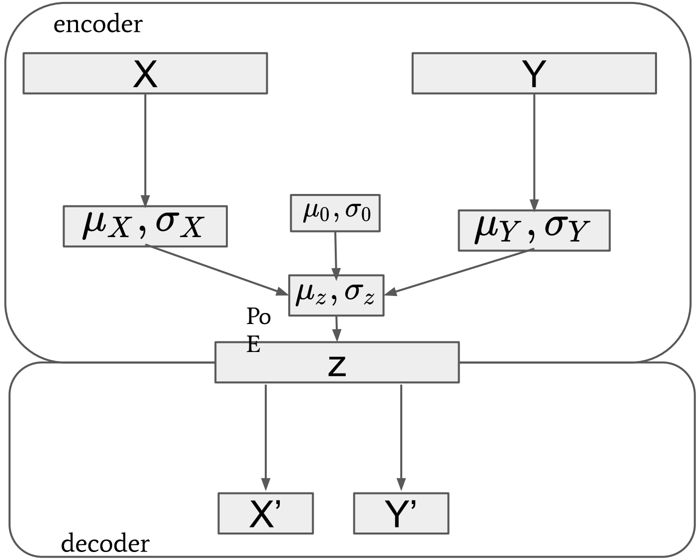
We can:

- 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

MVAE: Examples

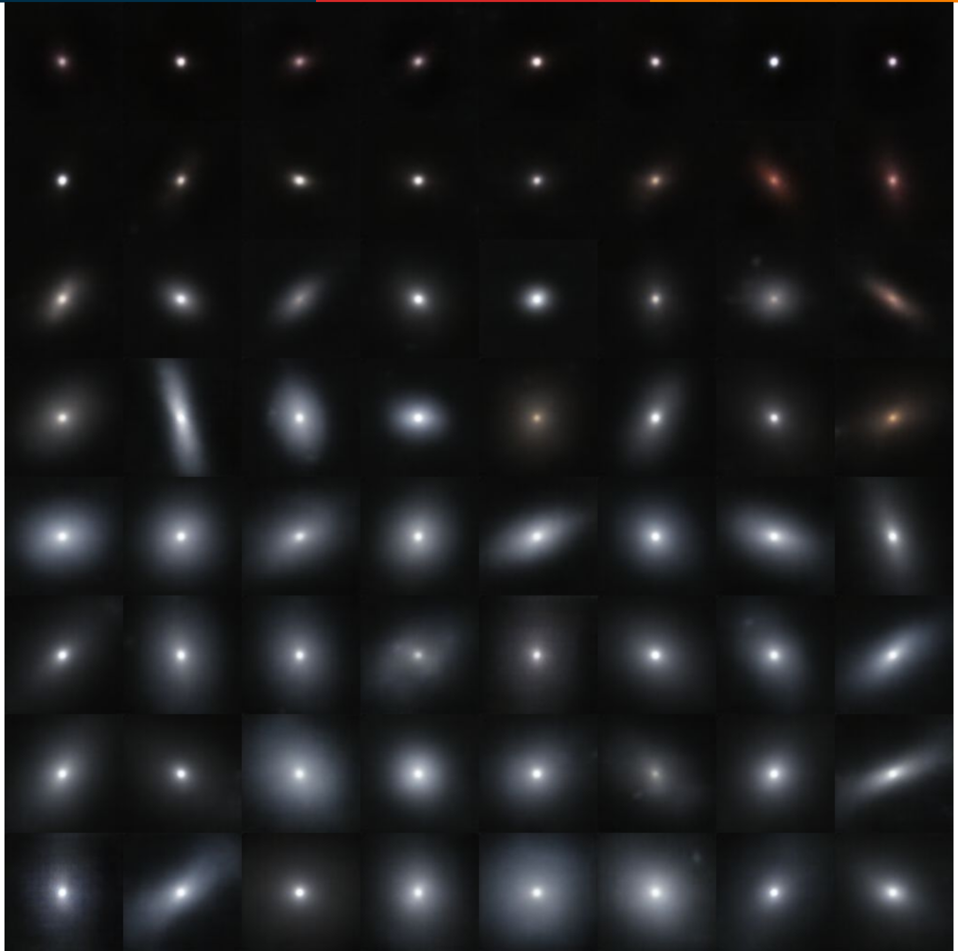


Samples with changing brightness (increasing downwards)

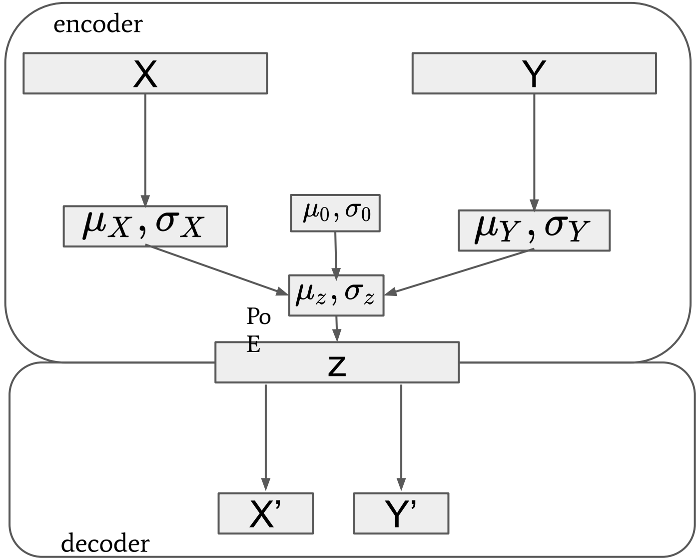


Bowen, Carrasco-Kind, et al. in prep.

MVAE: Examples

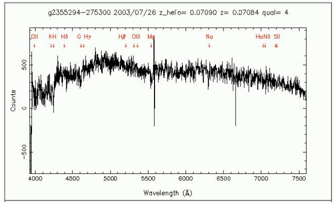


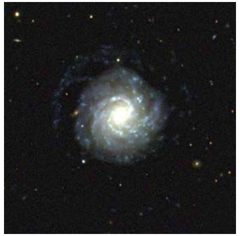
Samples with changing area
(increasing downwards)

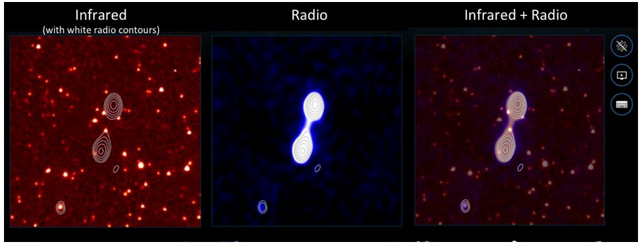


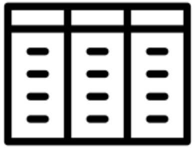
Bowen, Carrasco-Kind, et al. in prep.

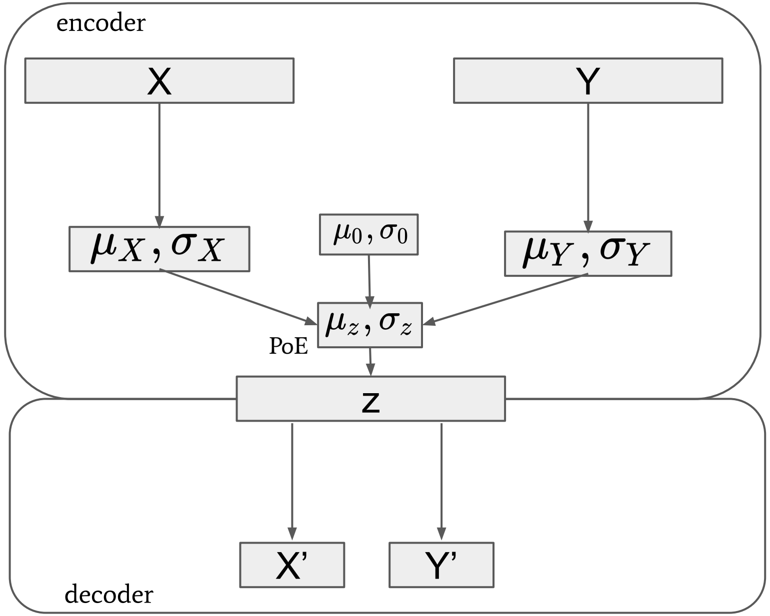
MVAE: Opens very interesting options







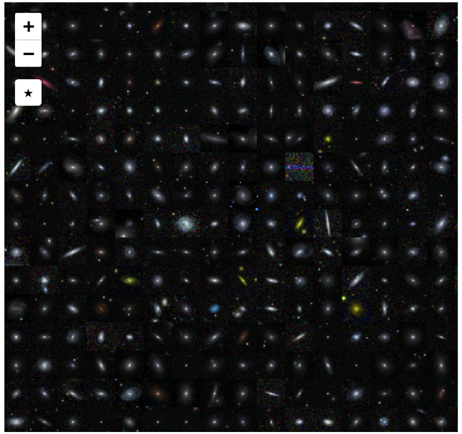


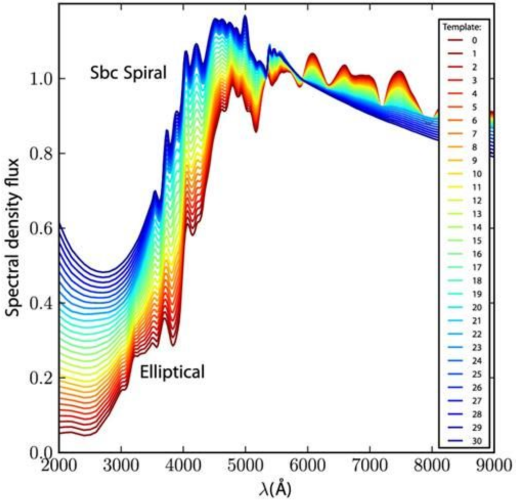


Bowen, Carrasco-Kind, et al. in prep.

Conclusions and Future plan

- 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
- Simulate SN events and galaxy images for training
- There is plans to classify DES images (prototype)

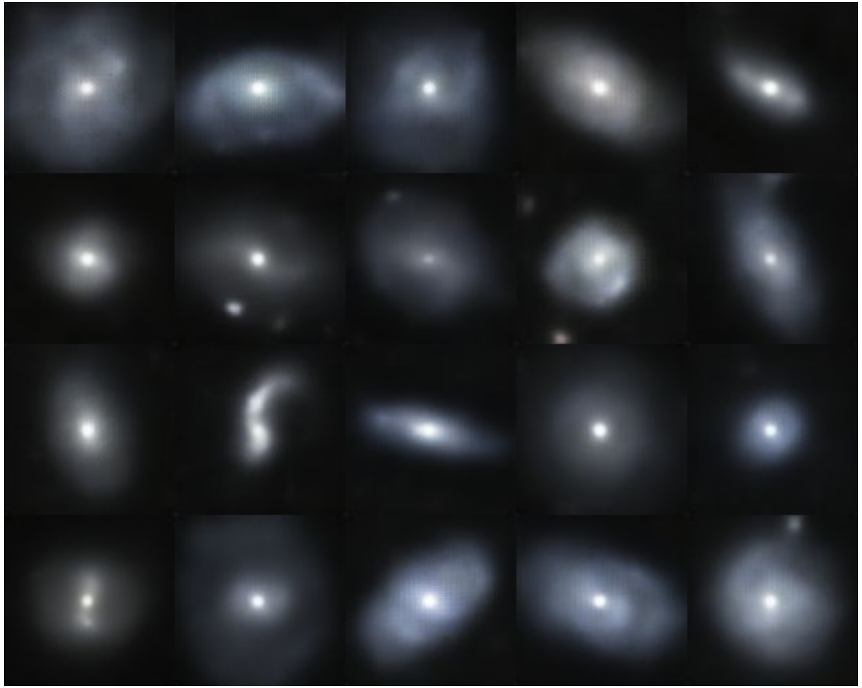




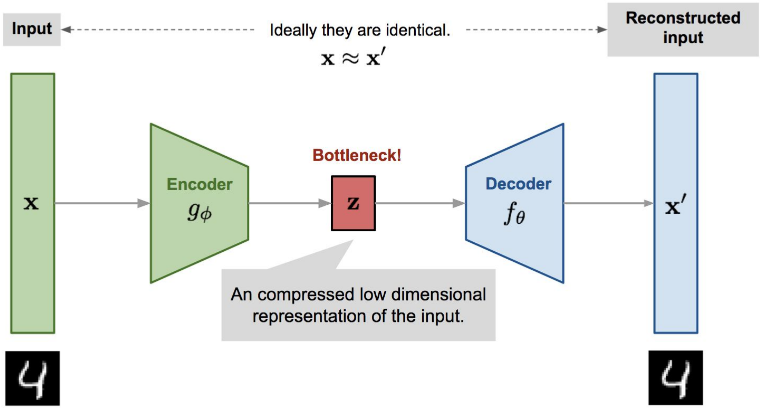
Thank you!

Questions?

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



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)

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

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

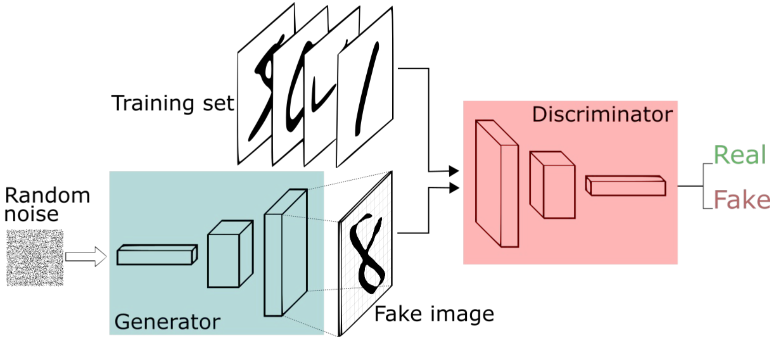
- Map x to a distribution $p(z|x)$
- Sample from distribution $z_i \sim p(z)$
- Generate fake data $x_i' \sim p_\theta(x'|z)$
- Probabilistic approach

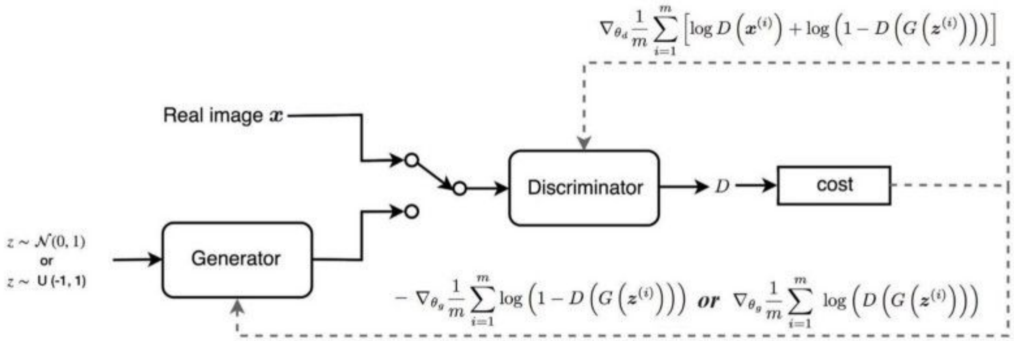
$$p(x,z) = p(x|z)p(z)$$

Variational "Autoencoder"

Generative Adversarial Networks (GAN)

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



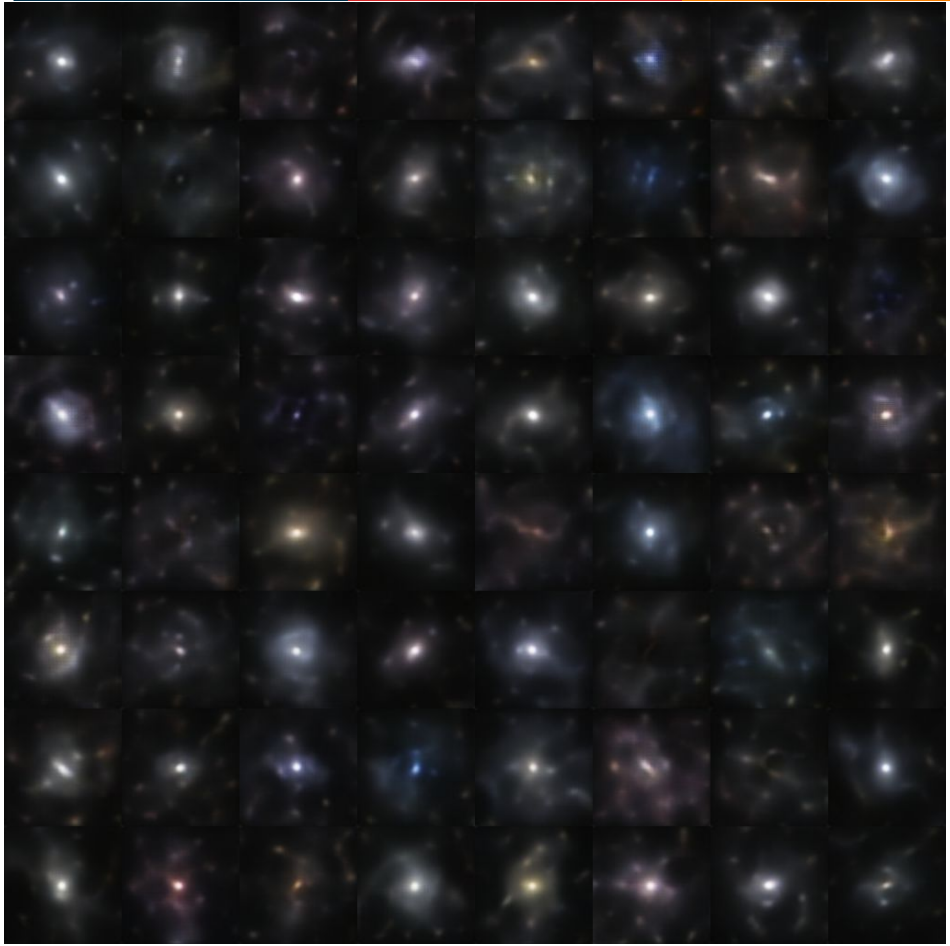


But:

- 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

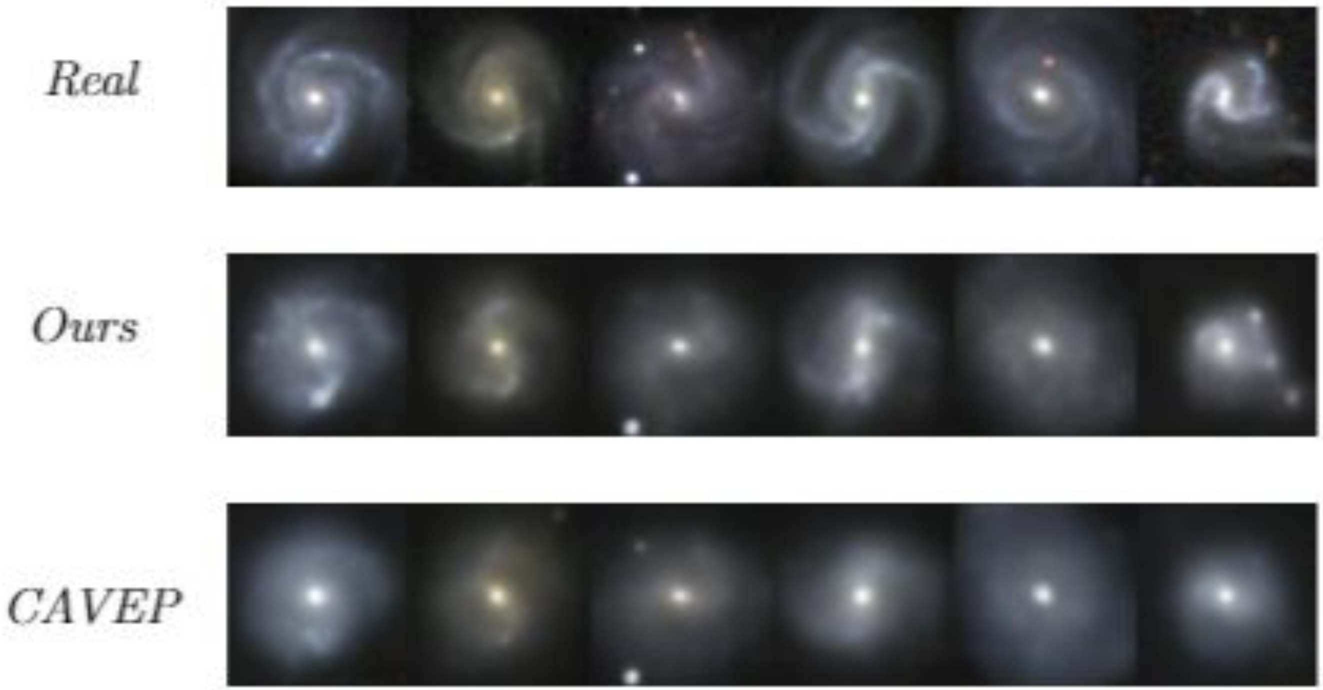
MVAE: Examples



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

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

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