Applications of deep learning in wide-field cosmological surveys

2nd ASTERICS-OBELICS Workshop
16-19 October 2017, Barcelona, Spain

François Lanusse

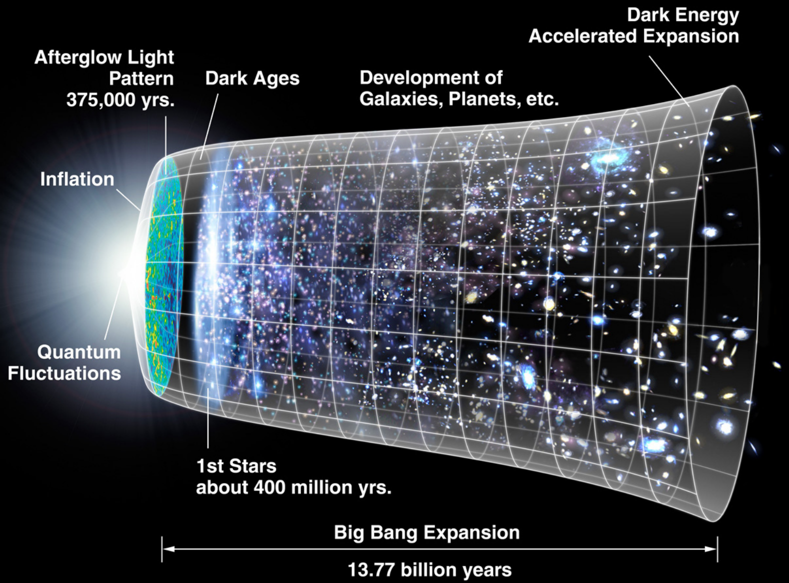
18 October 2017

McWilliams Center for Cosmology
Carnegie Mellon University

Asterics
Astronomy ESFRI & Research Infrastructure Cluster

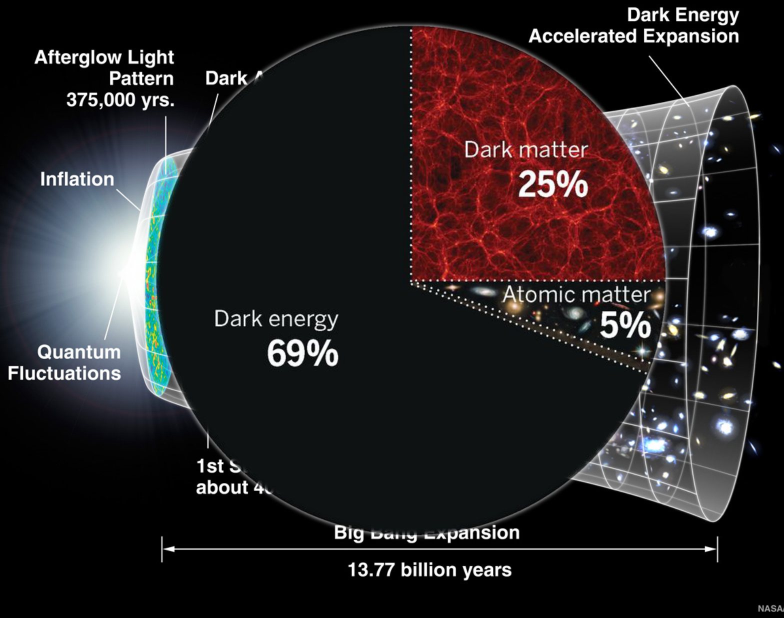
H2020-Astronomy ESFRI and Research Infrastructure Cluster
(Grant Agreement number: 653477).

the ΛCDM view of the Universe

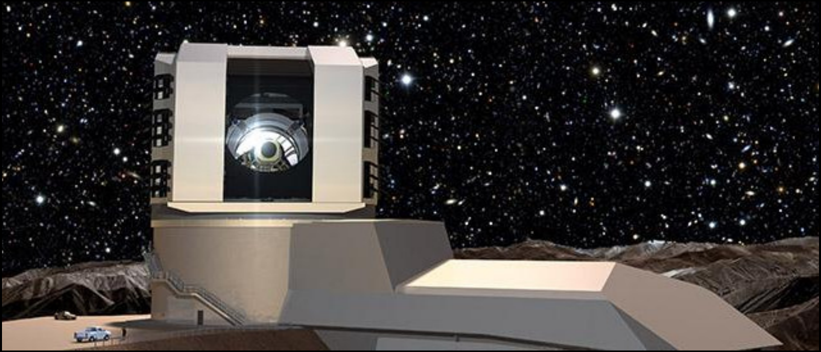


NASA/WMAP Science Team

the ΛCDM view of the Universe



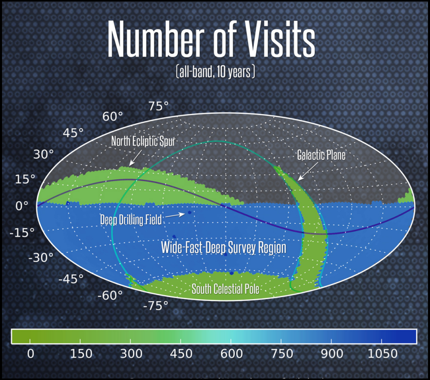
the Large Synoptic Survey Telescope



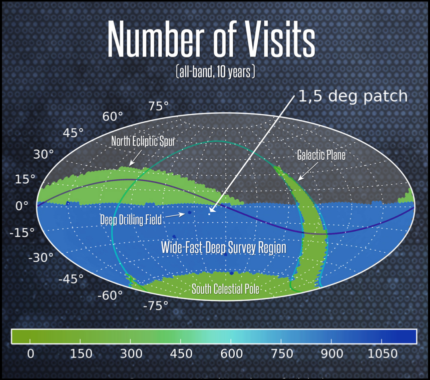
LSST in a few numbers

- 1000 images each night, each one is 3.2 GB and 40 full moons
  $\Longrightarrow$ 15 TB/night for 10 years
- Covers 18,000 square degrees (40% of the sky)
- Tens of billions of objects, each one observed $\sim$ 1000 times

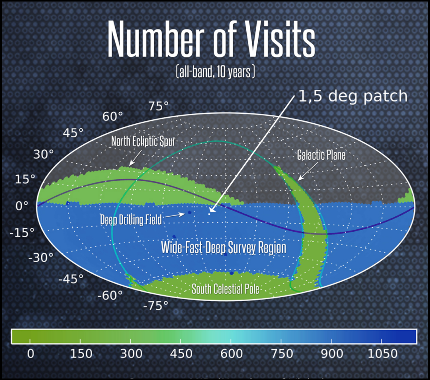
what does it look like ?



what does it look like ?

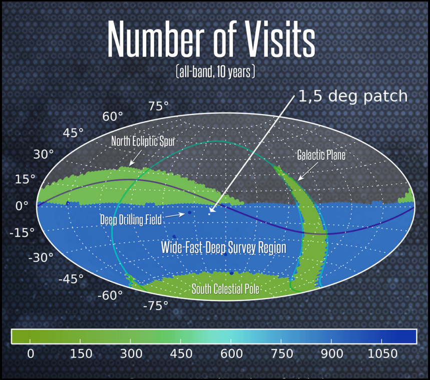


what does it look like ?



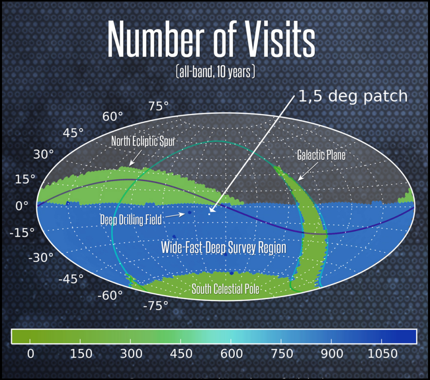
HSC-SPP Data Release 1

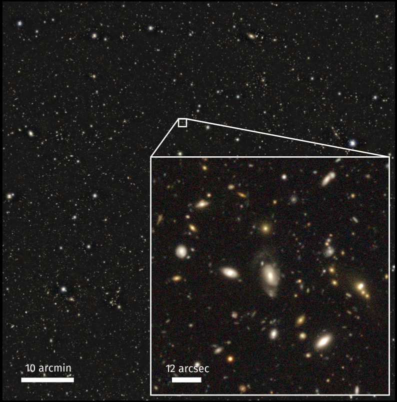
what does it look like ?



HSC-SPP Data Release 1

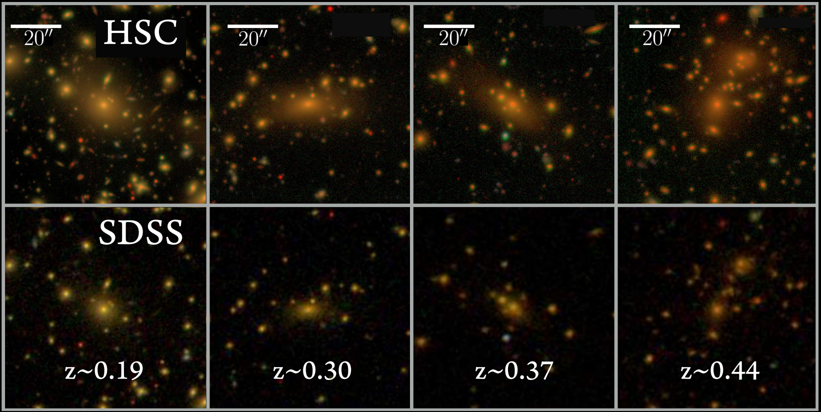
what does it look like ?





HSC-SPP Data Release 1

what does it look like ?



Huang et al. (2017), arXiv:1707.01904

the challenge for modern surveys

$\Longrightarrow$ Modern surveys will provide large volumes of high quality data

the challenge for modern surveys

$\Longrightarrow$ Modern surveys will provide large volumes of high quality data

A Blessing

- Unprecedented statistical power
- Great potential for new discoveries

the challenge for modern surveys

$\Longrightarrow$ Modern surveys will provide large volumes of high quality data

A Blessing

- Unprecedented statistical power
- Great potential for new discoveries

A Curse

- Existing methods are reaching their limits (computational cost, accuracy) at every step of the science analysis
- Control of systematic uncertainties becomes paramount

the challenge for modern surveys

$\Longrightarrow$ Modern surveys will provide large volumes of high quality data

A Blessing

- Unprecedented statistical power
- Great potential for new discoveries

A Curse

- Existing methods are reaching their limits (computational cost, accuracy) at every step of the science analysis
- Control of systematic uncertainties becomes paramount

$\Longrightarrow$ Dire need for novel data analysis techniques to fully realize the potential of modern surveys

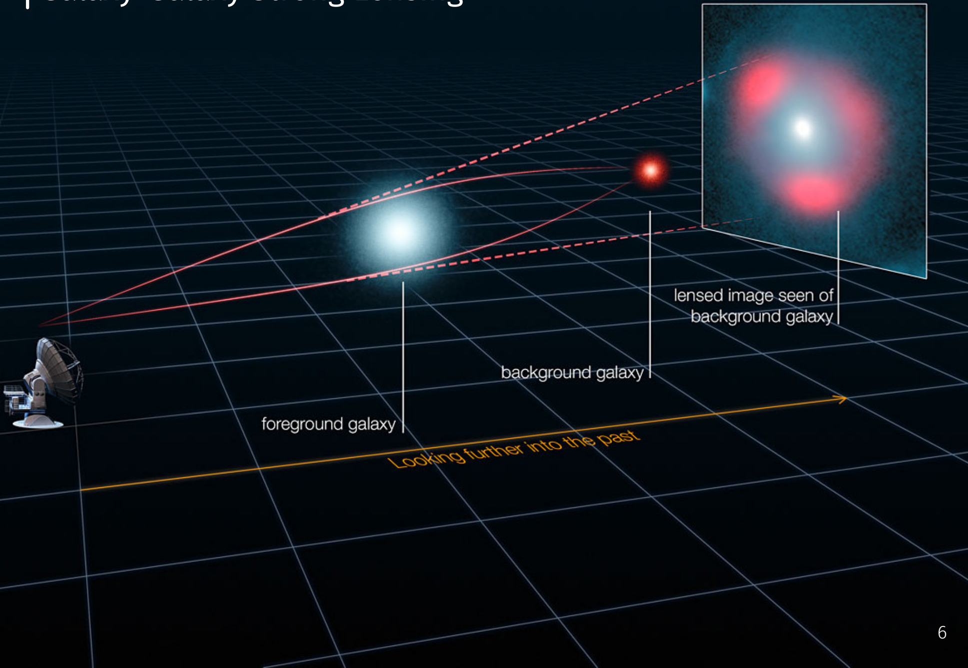
Outline of this talk

1. Finding strong gravitational lenses with Deep Learning
   - Galaxy-Scale strong lensing
   - Finding strong lenses
   - Deep Learning for image classification

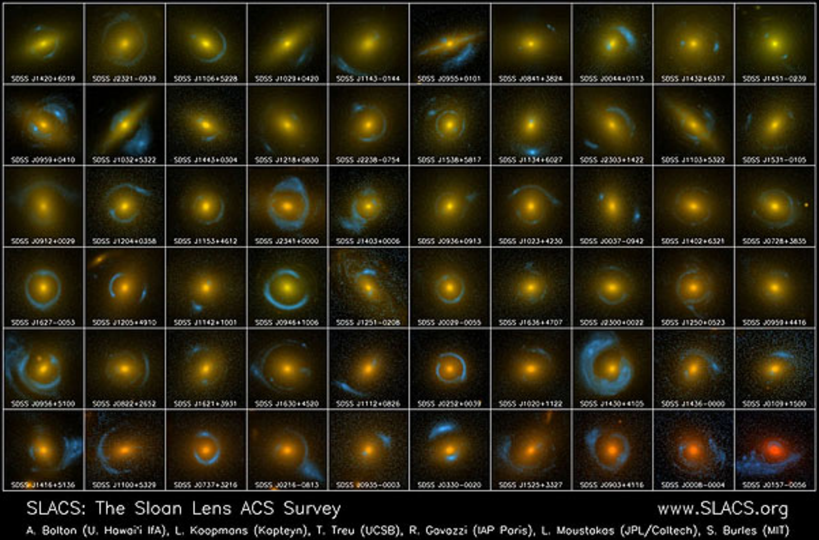
2. Deep Generative Models for weak lensing systematics
   - Weak Gravitational lensing
   - Deep Generative models of galaxy images

Finding strong gravitational lenses with Deep Learning

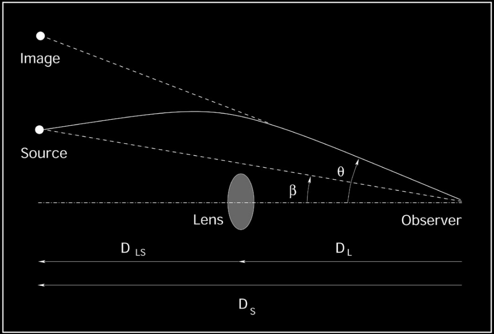
Galaxy-Galaxy Strong Lensing



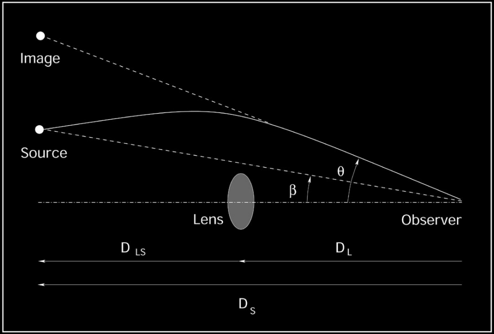
examples of strong lenses



example of application: gravitational time delays

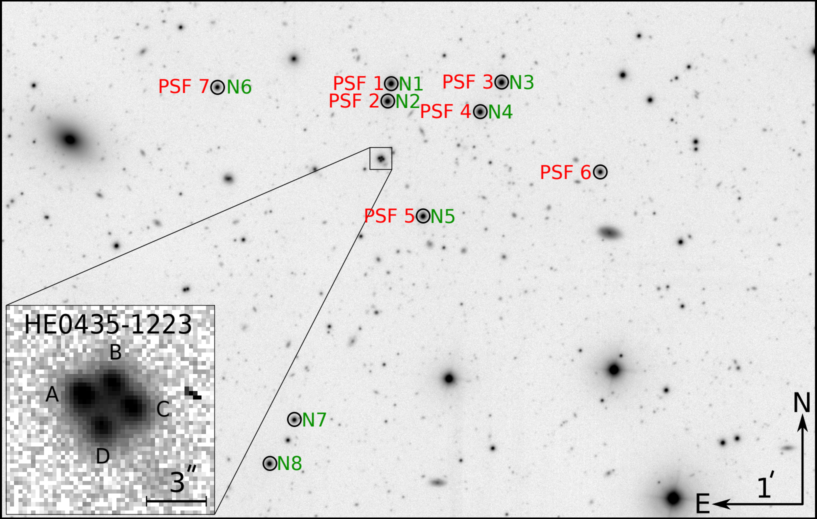


example of application: gravitational time delays

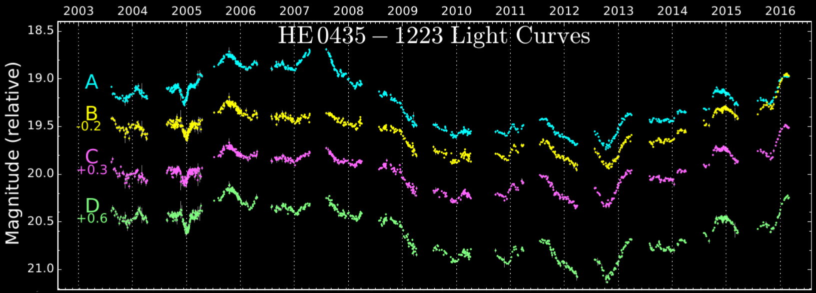


$$\Delta t_{ij} = \frac{1+z_L}{c} \underbrace{\frac{D_L\, D_S}{D_{LS}}}_{\propto H_0^{-1}} \left[\frac{(\boldsymbol{\theta}_i - \boldsymbol{\beta})^2}{2} - \psi(\boldsymbol{\theta}_i) + \frac{(\boldsymbol{\theta}_j - \boldsymbol{\beta})^2}{2} + \psi(\boldsymbol{\theta}_j) \right]$$

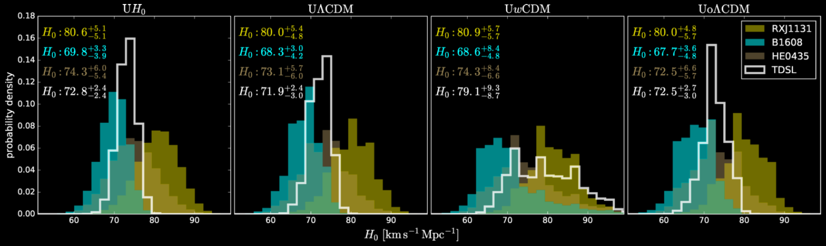
time delays of HE0435-1223 (Bonvin et al. 2017)



time delays of HE0435-1223 (Bonvin et al. 2017)

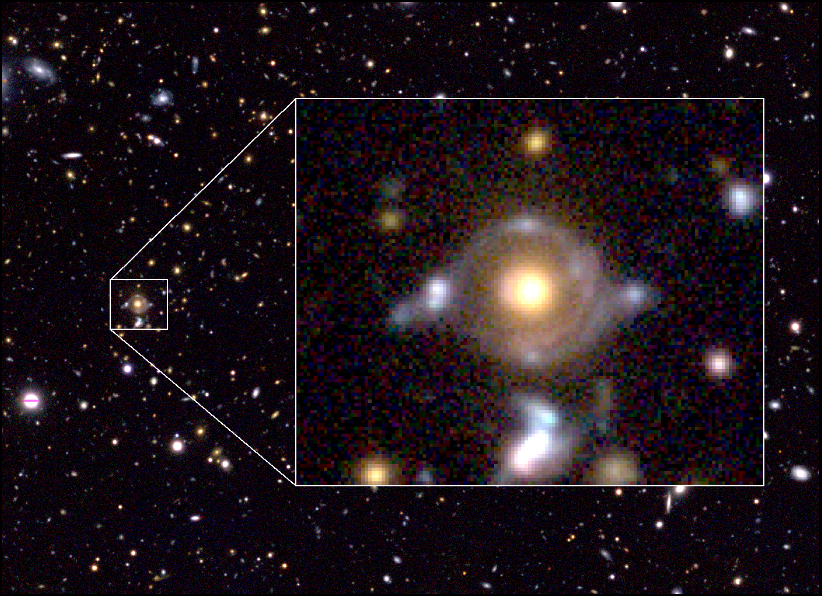


time delays of HE0435-1223 (Bonvin et al. 2017)

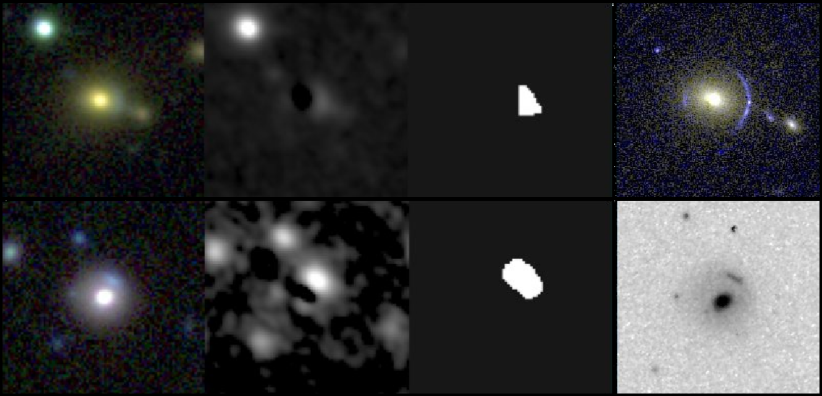


the problem: finding strong lenses

the problem: finding strong lenses

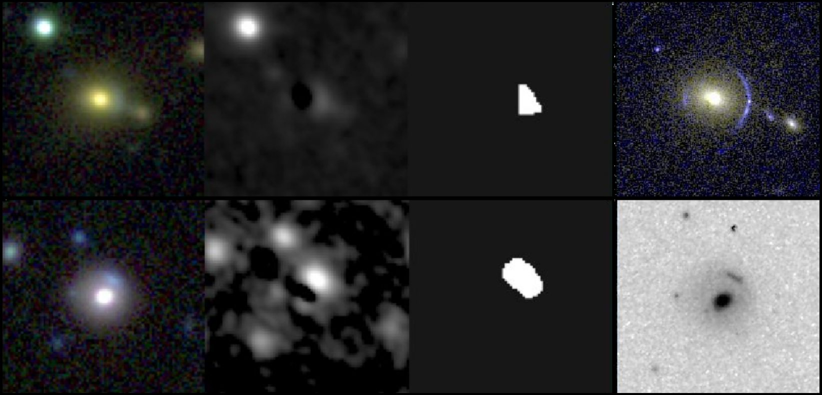


automated lens searches: RingFinder (Gavazzi et al. 2014)



gri composite $g - \alpha i$ detected areas HST images

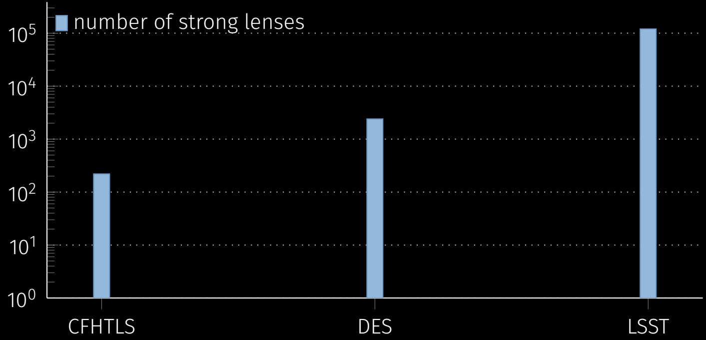
automated lens searches: RingFinder (Gavazzi et al. 2014)



gri composite | $g - \alpha i$ | detected areas | HST images

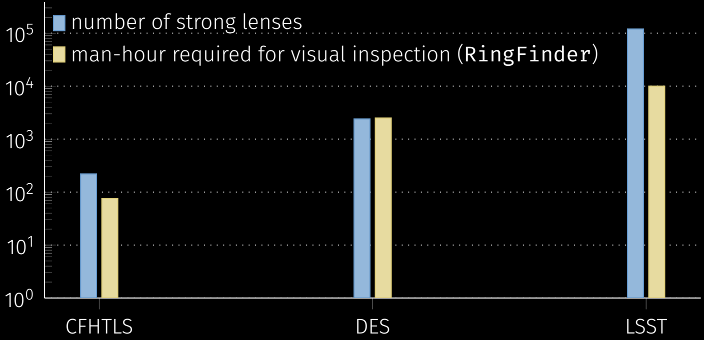
Visual inspection time required
$\sim$ 30 person-minutes / deg^2

extrapolation to future surveys



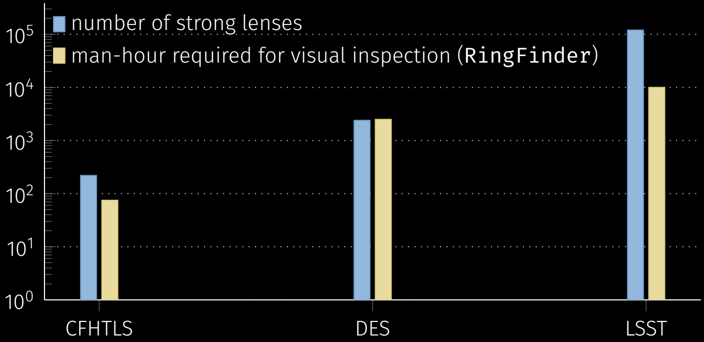
Gavazzi et al. (2014), Collett (2015)

extrapolation to future surveys



Gavazzi et al. (2014), Collett (2015)

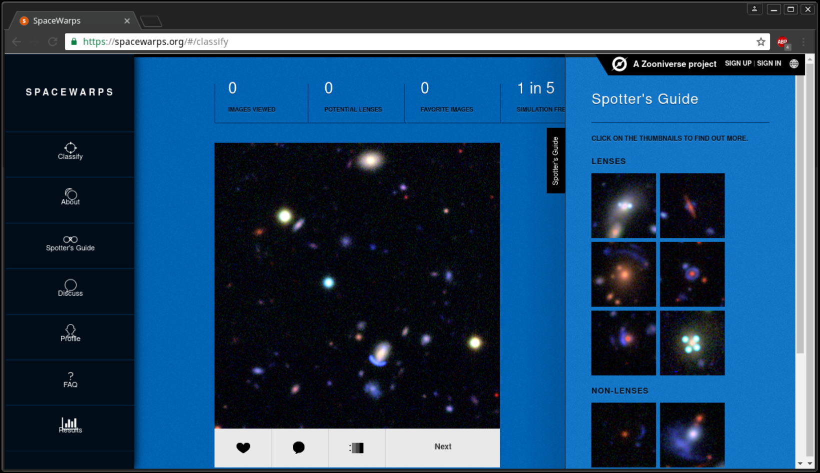
extrapolation to future surveys



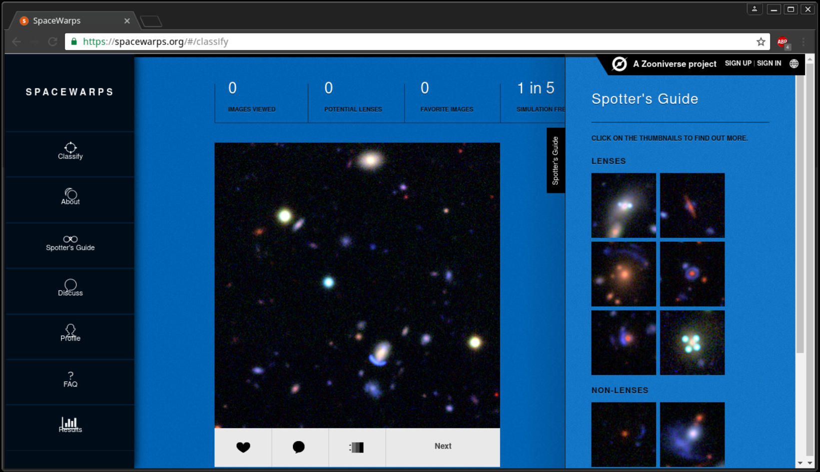
Gavazzi et al. (2014), Collett (2015)

$\Longrightarrow$ LSST would require an estimated 10^4 man-hours.

citizen science, the crowd-sourcing approach



citizen science, the crowd-sourcing approach



$\Longrightarrow$ Classifying all of LSST would take a few weeks with a crowd of 10^6 volunteers

Deep Learning to the rescue

in the news lately...

- Self-driving Uber takes the road in Pittsburgh (Sept. 2016)

in the news lately...

- Self-driving Uber takes the road in Pittsburgh (Sept. 2016)
- CMU's `Libratus` beats top poker players (Jan. 2017)

in the news lately...

- Self-driving Uber takes the road in Pittsburgh (Sept. 2016)
- CMU's `Libratus` beats top poker players (Jan. 2017)
- Google's `AlphaGo` beats world's top Go player (May 2017)

in the news lately...

- Self-driving Uber takes the road in Pittsburgh (Sept. 2016)
- CMU's `Libratus` beats top poker players (Jan. 2017)
- Google's `AlphaGo` beats world's top Go player (May 2017)

in the news lately...

- Self-driving Uber takes the road in Pittsburgh (Sept. 2016)
- CMU's `Libratus` beats top poker players (Jan. 2017)
- Google's `AlphaGo` beats world's top Go player (May 2017)

$\Longrightarrow$ technological revolution brought about by the advancement of deep learning

how did this happen ?

Training deep networks is difficult

Until $\sim$2010, networks were limited to a few layers because of vanishing gradients.

how did this happen ?

Training deep networks is difficult

Until ∼2010, networks were limited to a few layers because of vanishing gradients.

how did this happen ?

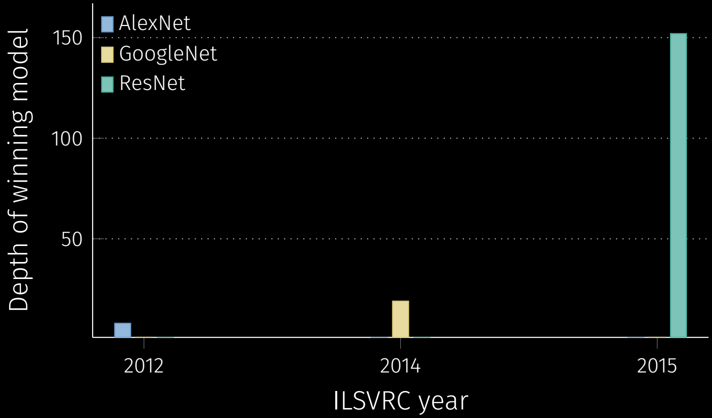
Training deep networks is difficult

Until ∼2010, networks were limited to a few layers because of vanishing gradients.

how did this happen ?

Training deep networks is difficult

Until ∼2010, networks were limited to a few layers because of vanishing gradients.



how did this happen ?

Training deep networks is difficult

Until ∼2010, networks were limited to a few layers because of vanishing gradients.

- Optimization tricks: Rectified Linear Units (ReLU), Dropout

how did this happen ?

Training deep networks is difficult

Until ∼2010, networks were limited to a few layers because of vanishing gradients.

- Optimization tricks: Rectified Linear Units (ReLU), Dropout
- Better architectures: Convolutional Neural Networks (CNN)

how did this happen ?

Training deep networks is difficult

Until $\sim$2010, networks were limited to a few layers because of vanishing gradients.

- Optimization tricks: Rectified Linear Units (ReLU), Dropout
- Better architectures: Convolutional Neural Networks (CNN)
- Powerful hardware: GPUs

how did this happen ?

Training deep networks is difficult

Until ∼2010, networks were limited to a few layers because of vanishing gradients.

- Optimization tricks: Rectified Linear Units (ReLU), Dropout
- Better architectures: Convolutional Neural Networks (CNN)
- Powerful hardware: GPUs
- Huge amount of training data: Internet

how did this happen ?

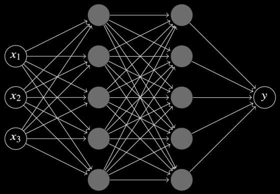
Training deep networks is difficult

Until ∼2010, networks were limited to a few layers because of vanishing gradients.

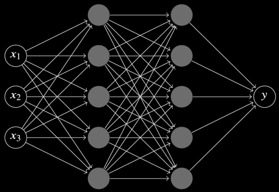
- Optimization tricks: Rectified Linear Units (ReLU), Dropout
- Better architectures: Convolutional Neural Networks (CNN)
- Powerful hardware: GPUs
- Huge amount of training data: Internet

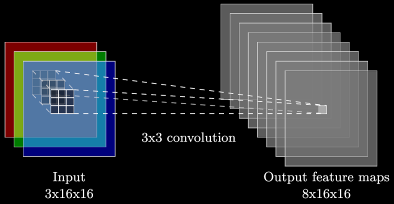
$\Longrightarrow$ State-of-the-art models outperform humans for image detection/classification

Convolutional Neural Network

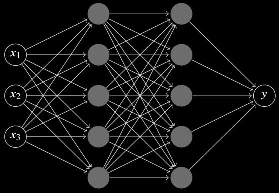


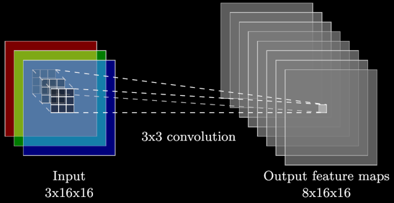
Convolutional Neural Network

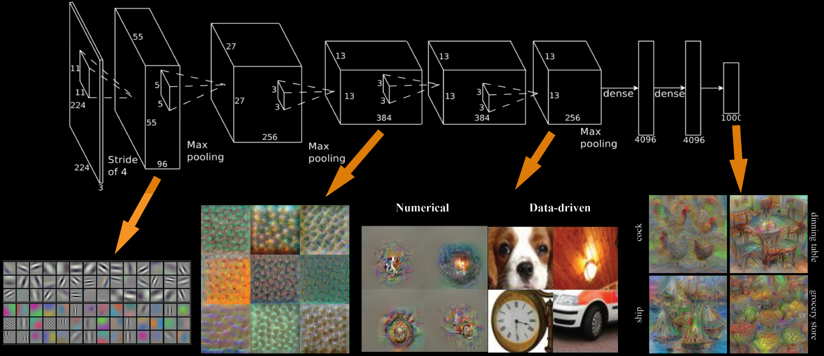




Convolutional Neural Network







Conv 1: Edge+Blob Conv 3: Texture Conv 5: Object Parts Fc8: Object Classes

residual learning

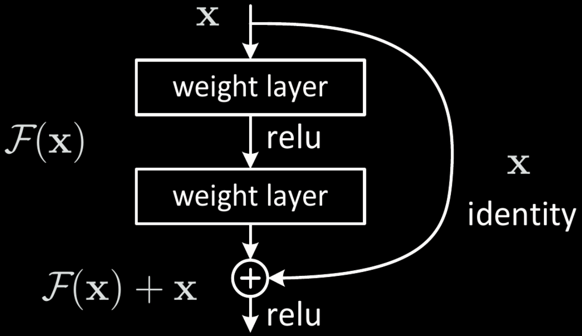


Image credit: He et al. (2015)

- Learning the difference to the identity (He et al. 2015)

residual learning

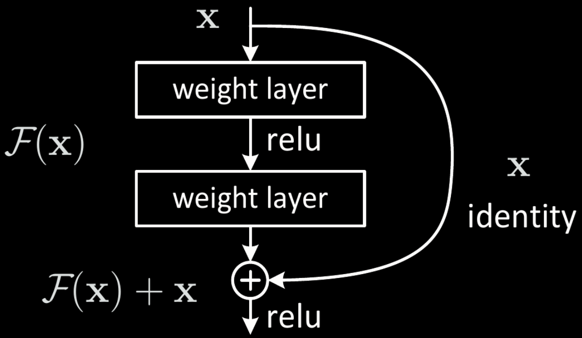
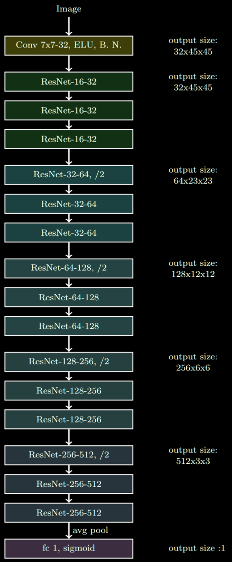


Image credit: He et al. (2015)

- Learning the difference to the identity (He et al. 2015)
- Easier to initialize and to train in deep architectures ($>$ 1000 layers)

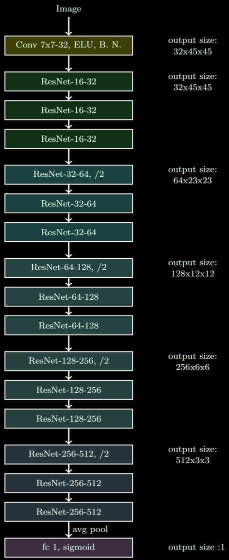
CMU DeepLens: deep residual learning for strong lens finding



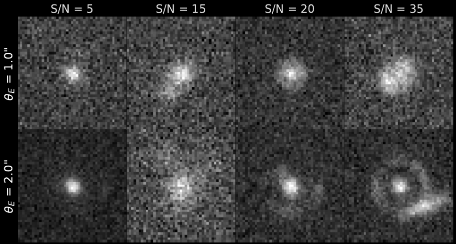
- Deep residual network (46 layers) with pre-activated bottleneck residual units

model architecture

CMU DeepLens: deep residual learning for strong lens finding

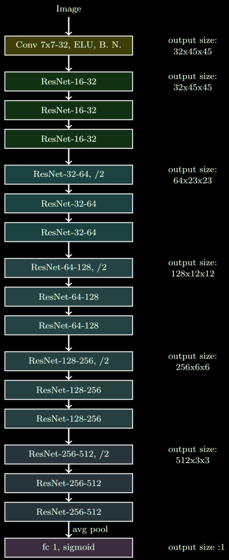


- Deep residual network (46 layers) with pre-activated bottleneck residual units
- Training on simulated LSST lenses:



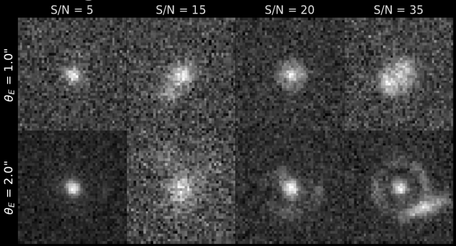
model architecture

CMU DeepLens: deep residual learning for strong lens finding



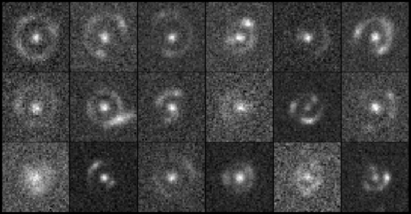
model architecture

- Deep residual network (46 layers) with pre-activated bottleneck residual units
- Training on simulated LSST lenses:

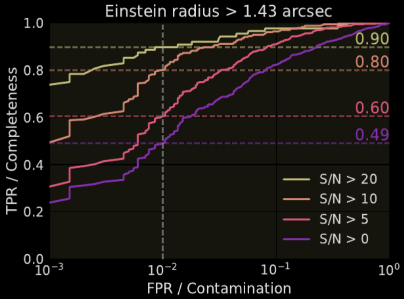


- Classification of 45x45 images in 350 μs $\implies$ 9 hours to classify a sample of 10^8 lens candidates on a single GPU (Nvidia Titan X)

performance on simulations



Highest probability lenses



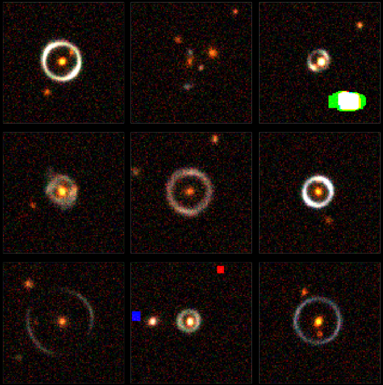
$$\text{True Positive Rate} = \frac{TP}{TP + FN}$$

$$\text{False Positive Rate} = \frac{FP}{FP + TN}$$

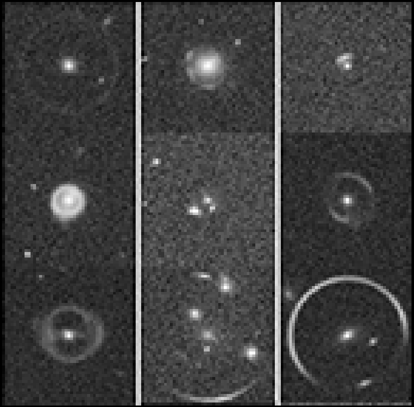
- TP: True Positives
- FN: False Negatives

- FP: False Positives
- TN: True Negatives

Euclid strong lens finding challenge

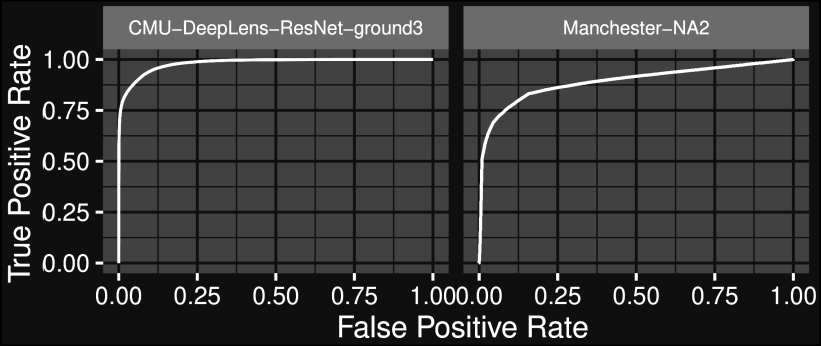


Ground based simulations



Space based simulations

Euclid strong lens finding challenge



- `CMU DeepLens` wins over 24 other methods (including other CNN methods) in space and ground challenge.
- Significantly outperforms human classification accuracy.

outlook for LSST

- Automated lens finder faster and more reliable than human volunteers

outlook for LSST

- Automated lens finder faster and more reliable than human volunteers
- Rethink the pipeline infrastructure if there is no need for visual inspection

outlook for LSST

- Automated lens finder faster and more reliable than human volunteers
- Rethink the pipeline infrastructure if there is no need for visual inspection
- Deep Learning model is only as good as its training data
  $\Longrightarrow$ Need realistic simulations

outlook for LSST

- Automated lens finder faster and more reliable than human volunteers

- Rethink the pipeline infrastructure if there is no need for visual inspection

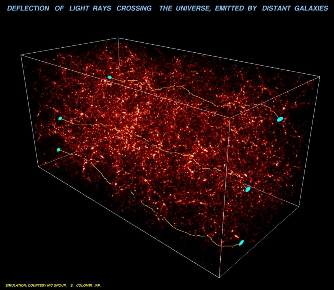
- Deep Learning model is only as good as its training data
  $\Longrightarrow$ Need realistic simulations

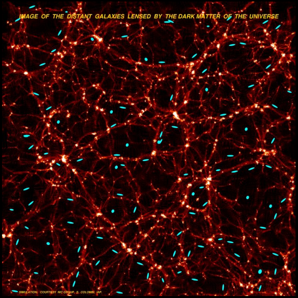
Honorable mentions:

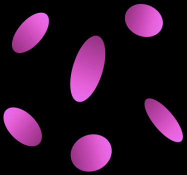
- Classification of time series using Deep Recurrent Neural networks
- Estimation of galaxy redshifts from multi-band images with residual networks

Deep Generative Models for weak lensing systematics

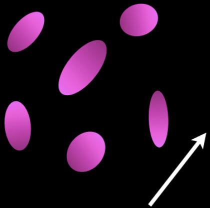
weak gravitational lensing







Galaxies randomly distributed

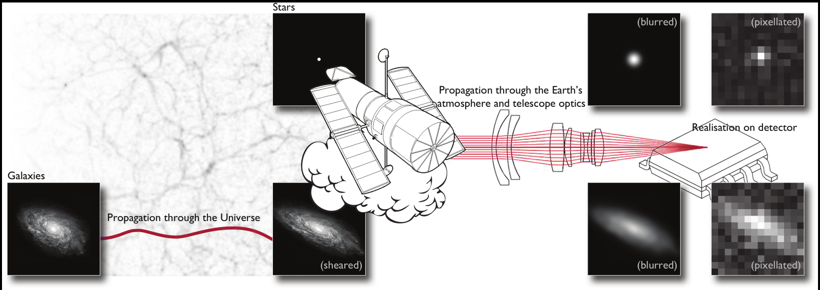


Slight alignment

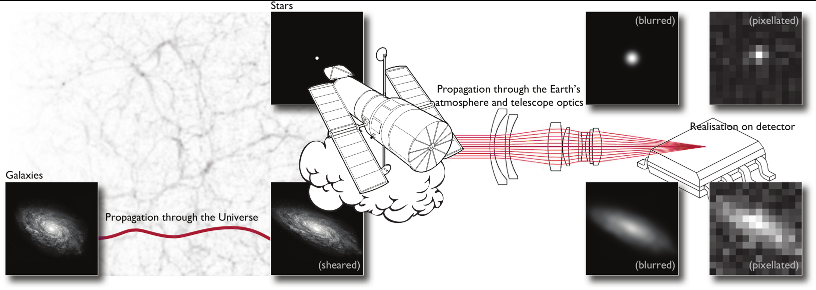
$\epsilon = \epsilon_i + \gamma$ with $< \epsilon_i >= 0$

$\Longrightarrow \quad < \epsilon >= \gamma$

a very delicate measurement



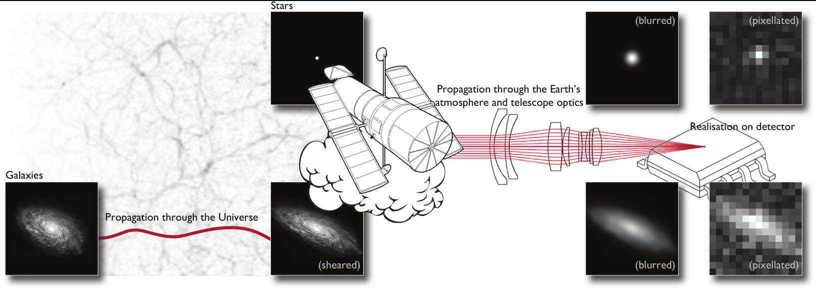
a very delicate measurement



Shape measurement biases

$$< e > \quad = \quad (1+m)\,\gamma + c$$

a very delicate measurement

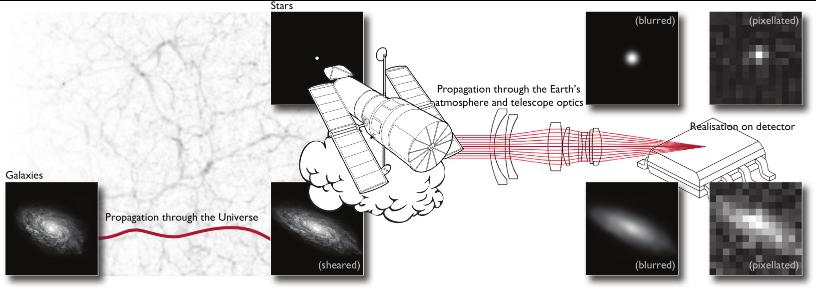


Shape measurement biases

$$< e > \quad = \quad (1 + m)\, \gamma + c$$

- Can be calibrated on image simulations

a very delicate measurement

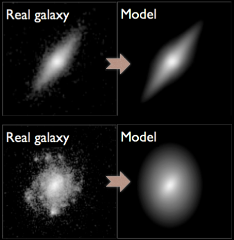


Shape measurement biases

$$< e > \quad = \quad (1 + m)\, \gamma + c$$

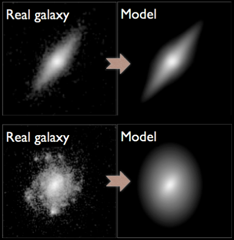
- Can be calibrated on image simulations
- How complex do the simulations need to be ?

impact of galaxy morphology

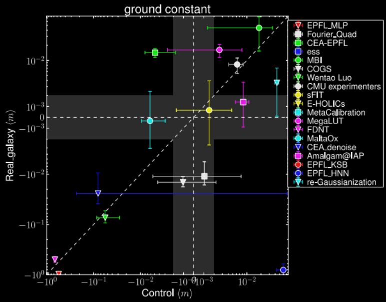


Mandelbaum et al. (2013)

impact of galaxy morphology

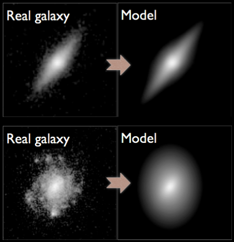


Mandelbaum et al. (2013)

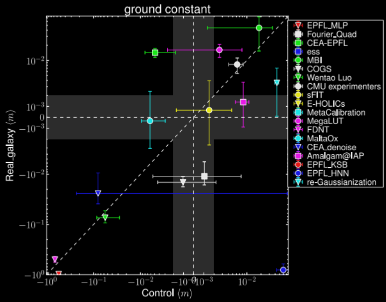


Mandelbaum et al. (2014)

impact of galaxy morphology



Mandelbaum et al. (2013)



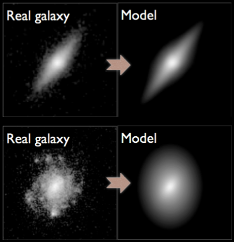
Mandelbaum et al. (2014)

The need for data-driven generative models

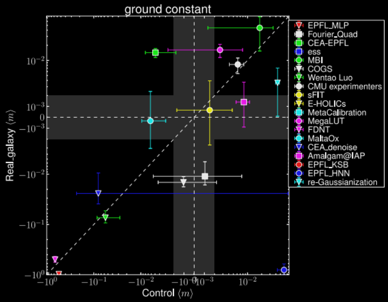
There can be two situations:

- Lack or inadequacy of physical model

impact of galaxy morphology



Mandelbaum et al. (2013)



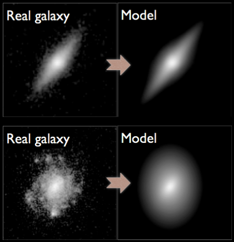
Mandelbaum et al. (2014)

The need for data-driven generative models

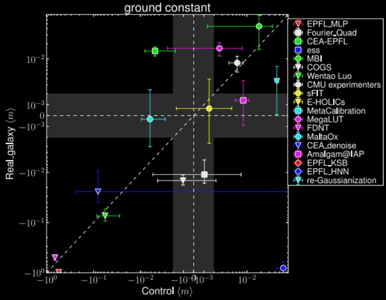
There can be two situations:

- Lack or inadequacy of physical model
- Extremely computationally expensive simulations

impact of galaxy morphology



Mandelbaum et al. (2013)



Mandelbaum et al. (2014)

The need for data-driven generative models

There can be two situations:

- Lack or inadequacy of physical model
- Extremely computationally expensive simulations

$\Longrightarrow$ Learn a model for the signal from the data itself

the evolution of deep generative models

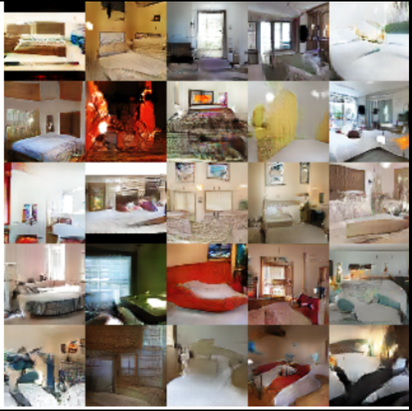
- Deep Belief Network (Hinton et al. 2006)

the evolution of deep generative models

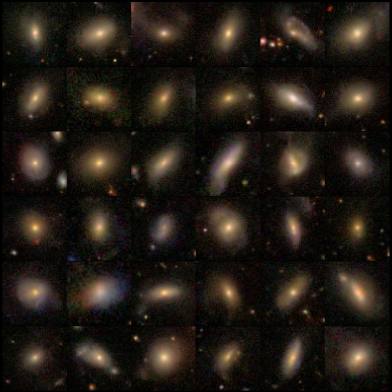
- Deep Belief Network (Hinton et al. 2006)
- Variational AutoEncoder (Kingma & Welling 2013)

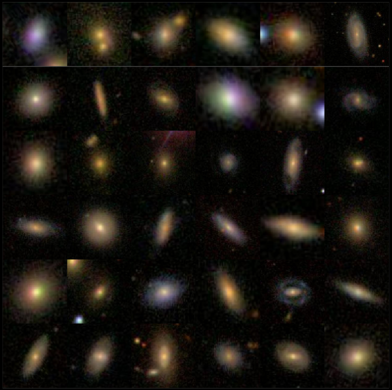
the evolution of deep generative models

- Deep Belief Network (Hinton et al. 2006)
- Variational AutoEncoder (Kingma & Welling 2013)
- Generative Adversarial Network (Radford et al. 2016)

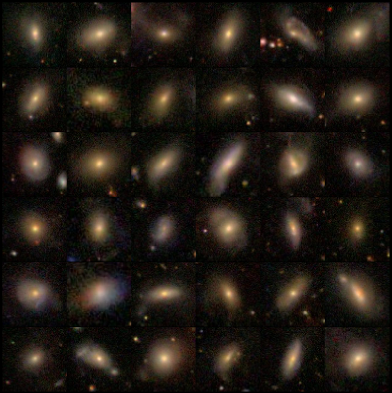


visual Turing test

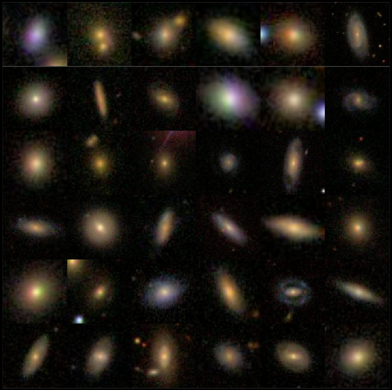




visual Turing test

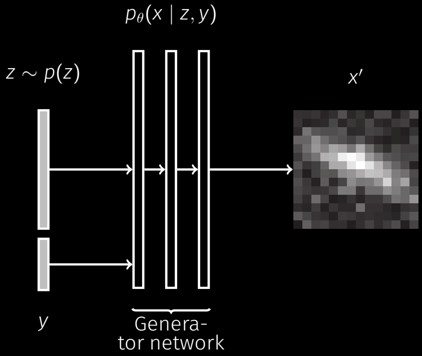


Mock - PixelCNN



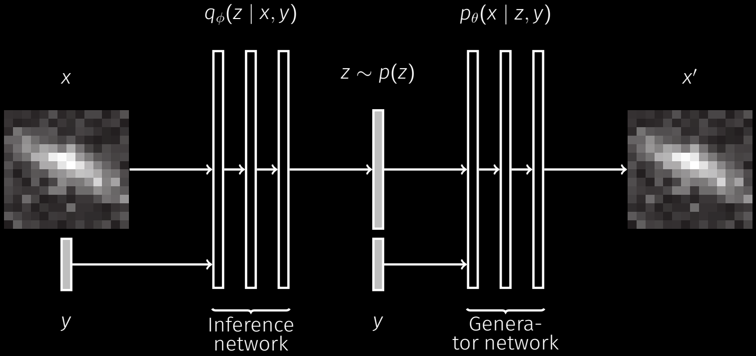
Real - SDSS

Conditional Variational AutoEncoder (CVAE)



Ravanbakhsh, Lanusse et al. (2017)

Conditional Variational AutoEncoder (CVAE)



Ravanbakhsh, **Lanusse** et al. (2017)

$$\log(p_\theta(x \mid y)) \geq -\underbrace{\mathbb{D}_{\mathrm{KL}}(q_\phi(z \mid x, y) \| p_\theta(z \mid y))}_{\text{Code regularisation}} + \underbrace{\mathbb{E}_{z \sim q_\phi(\cdot \mid x, y)}[\log p_\theta(x \mid z, y)]}_{\text{Reconstruction error}}$$

modeling galaxy images from the Hubble Space Telescope

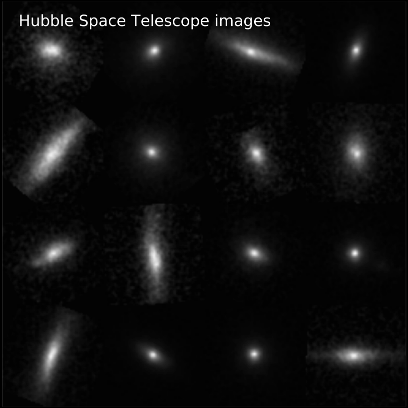
Training parameters

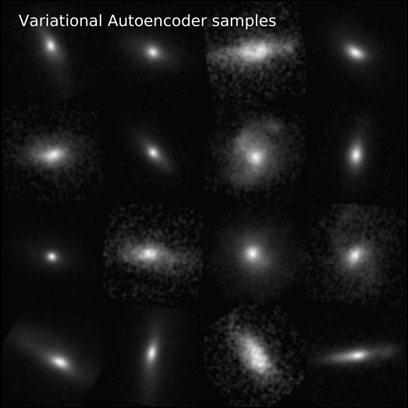
- Training set: postage stamps from COSMOS HST/ACS survey
- Conditional model: Half-light radius, magnitude, redshift

modeling galaxy images from the Hubble Space Telescope

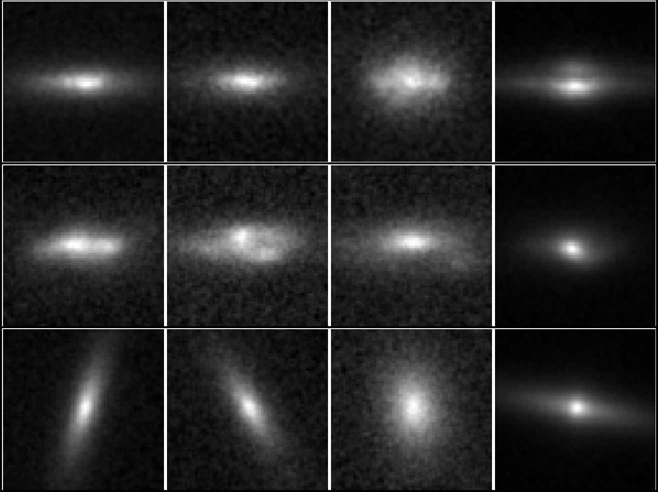
Training parameters

- Training set: postage stamps from COSMOS HST/ACS survey
- Conditional model: Half-light radius, magnitude, redshift



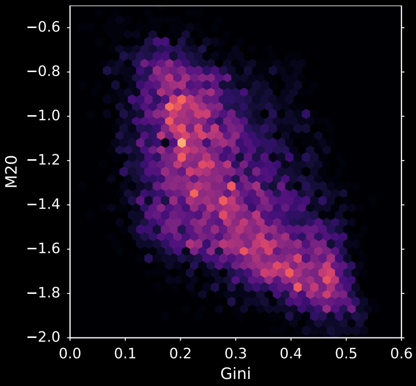


morphological statistics

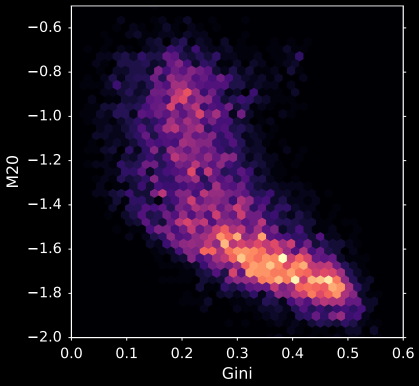


From top to bottom: Real COSMOS galaxies, CVAE samples, Parametric fits

morphological statistics

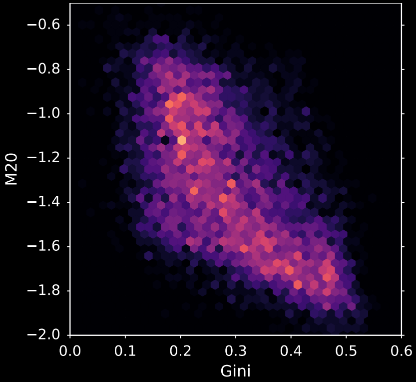


COSMOS images

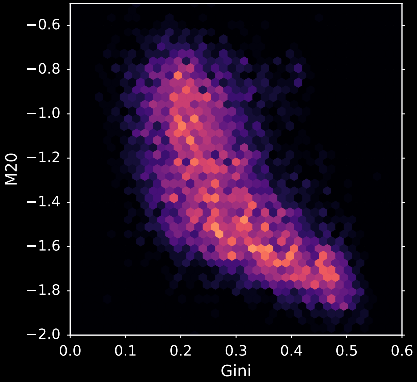


Parametric model

morphological statistics



COSMOS images



CVAE samples

outlook for LSST

- Generative models are a data driven way of completing our physical modeling

outlook for LSST

- Generative models are a data driven way of completing our physical modeling
- Implementing these models inside the image simulation software used to simulate the LSST survey

outlook for LSST

- Generative models are a data driven way of completing our physical modeling
- Implementing these models inside the image simulation software used to simulate the LSST survey
- Will allow for an extra degree of realism in LSST simulations
  $\Longrightarrow$ Essential to the calibration of science pipeline

outlook for LSST

- Generative models are a data driven way of completing our physical modeling
- Implementing these models inside the image simulation software used to simulate the LSST survey
- Will allow for an extra degree of realism in LSST simulations
  $\Longrightarrow$ Essential to the calibration of science pipeline

Honorable mention:

- Modeling galaxy properties in Nbody simulations using deep generative networks on graphs

conclusion

What can machine learning do for cosmology ?

- Model and analyze large volume of complex datasets

conclusion

What can machine learning do for cosmology ?

- Model and analyze large volume of complex datasets
- Open new and powerful ways to look at the data

conclusion

What can machine learning do for cosmology ?

- Model and analyze large volume of complex datasets
- Open new and powerful ways to look at the data
- Help control systematics in conventional analyses

conclusion

What can machine learning do for cosmology ?

- Model and analyze large volume of complex datasets
- Open new and powerful ways to look at the data
- Help control systematics in conventional analyses

Technical solutions

- Hardware: Nvidia Titan X GPUs
- Software: Theano/Lasagne, Tensorflow, CUDNN

conclusion

What can machine learning do for cosmology ?

- Model and analyze large volume of complex datasets
- Open new and powerful ways to look at the data
- Help control systematics in conventional analyses

Technical solutions

- Hardware: Nvidia Titan X GPUs
- Software: Theano/Lasagne, Tensorflow, CUDNN

Acknowledgment

- H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).