Where is Deep Learning Impacting Modern Cosmology ?

19th International ACAT Workshop, Saas-Fee, Switzerland

François Lanusse

March 15, 2019

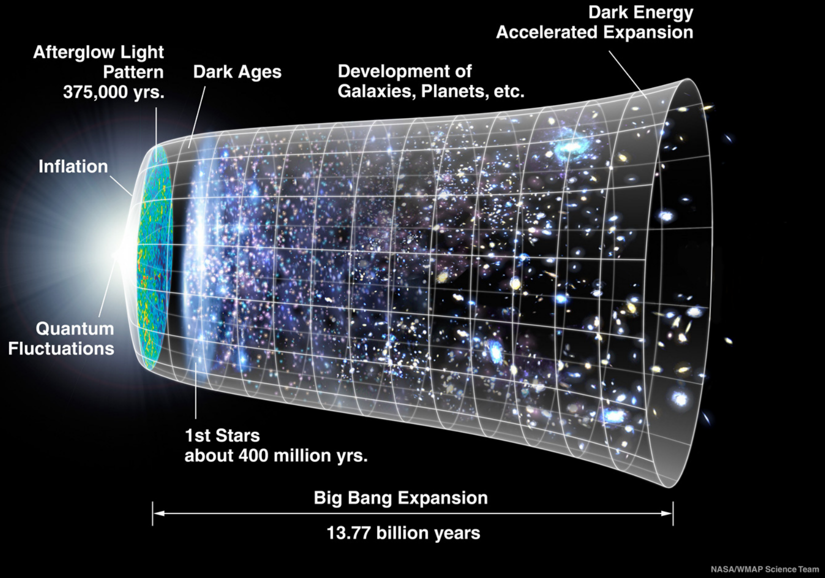
University of California Berkeley
Lawrence Berkeley National Laboratory

BERKELEY CENTER *for* COSMOLOGICAL PHYSICS

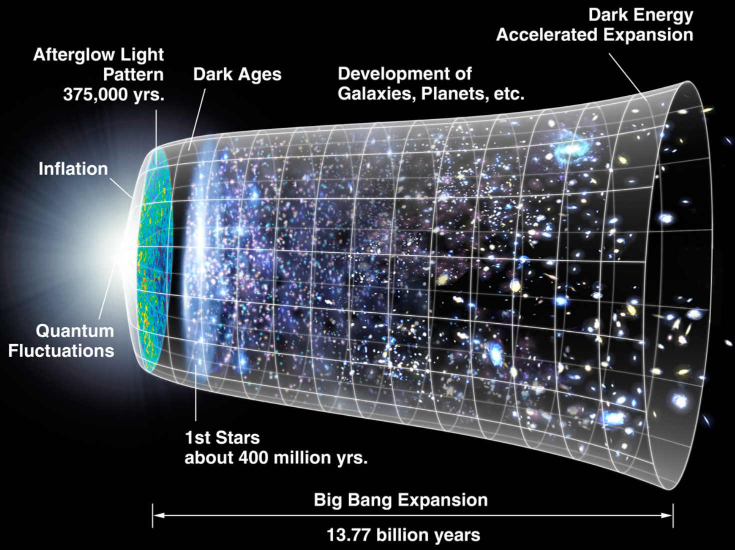
BERKELEY Institute for Data Science

LSST DESC Dark Energy Science Collaboration

the ΛCDM view of the Universe

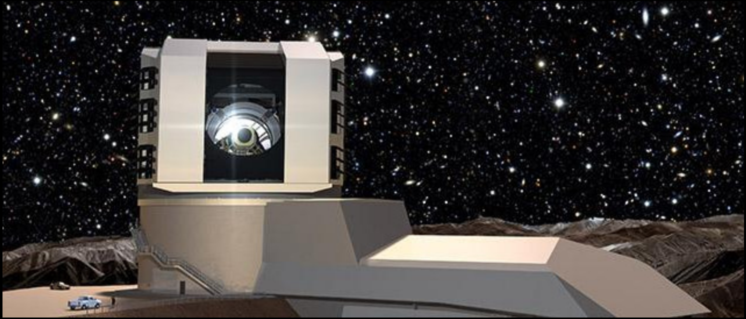


the ΛCDM view of the Universe



NASA/WMAP Science Team

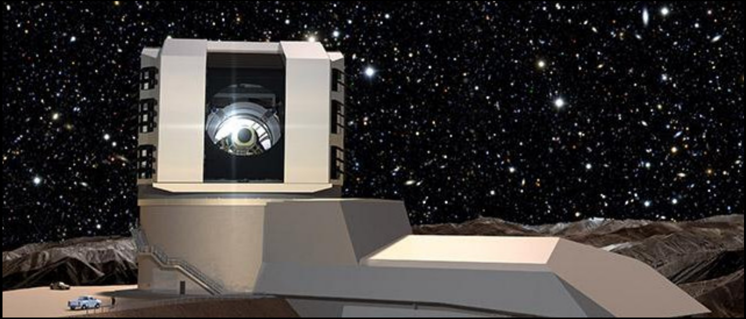
the Large Synoptic Survey Telescope



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

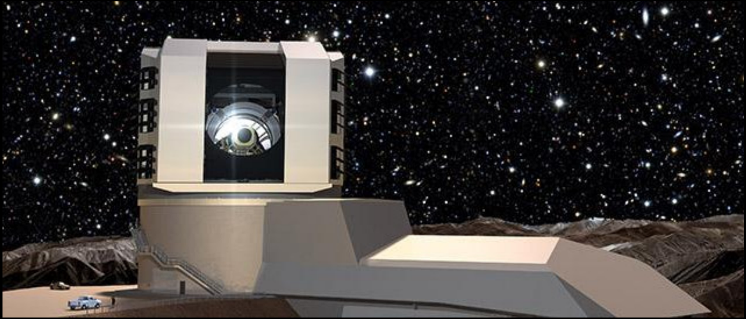
the Large Synoptic Survey Telescope



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)

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

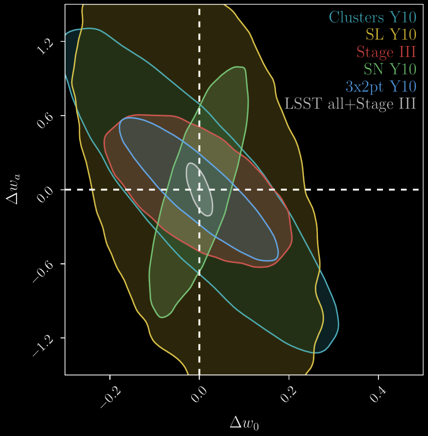
the challenge for modern surveys

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

A Blessing

- Unprecedented statistical power
- Great potential for new discoveries

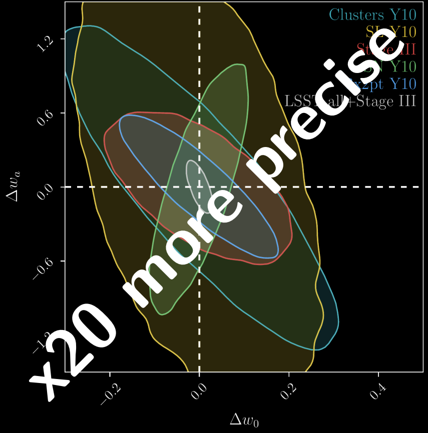
LSST forecast on dark energy parameters



the challenge for modern surveys

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

LSST forecast on dark energy parameters



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

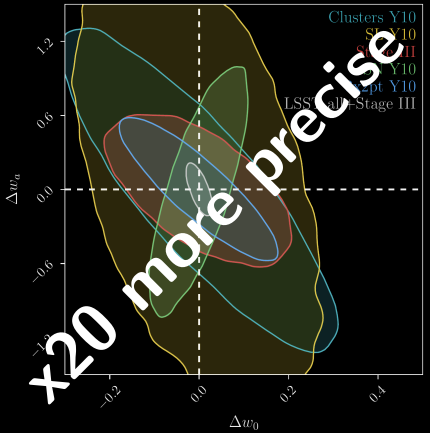
A Blessing

- Unprecedented statistical power
- Great potential for new discoveries

A Curse

- Existing methods are reaching their limits at every step of the science analysis
- Control of systematic uncertainties becomes paramount

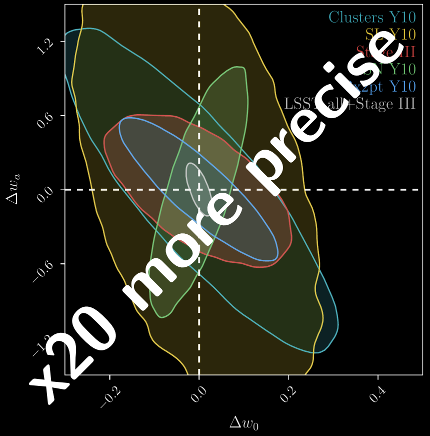
LSST forecast on dark energy parameters



the challenge for modern surveys

⟹ Modern surveys will provide **large volumes** of **high quality** data

LSST forecast on dark energy parameters



A Blessing

- Unprecedented statistical power
- Great potential for new discoveries

A Curse

- Existing methods are reaching their limits at every step of the science analysis
- Control of systematic uncertainties becomes paramount

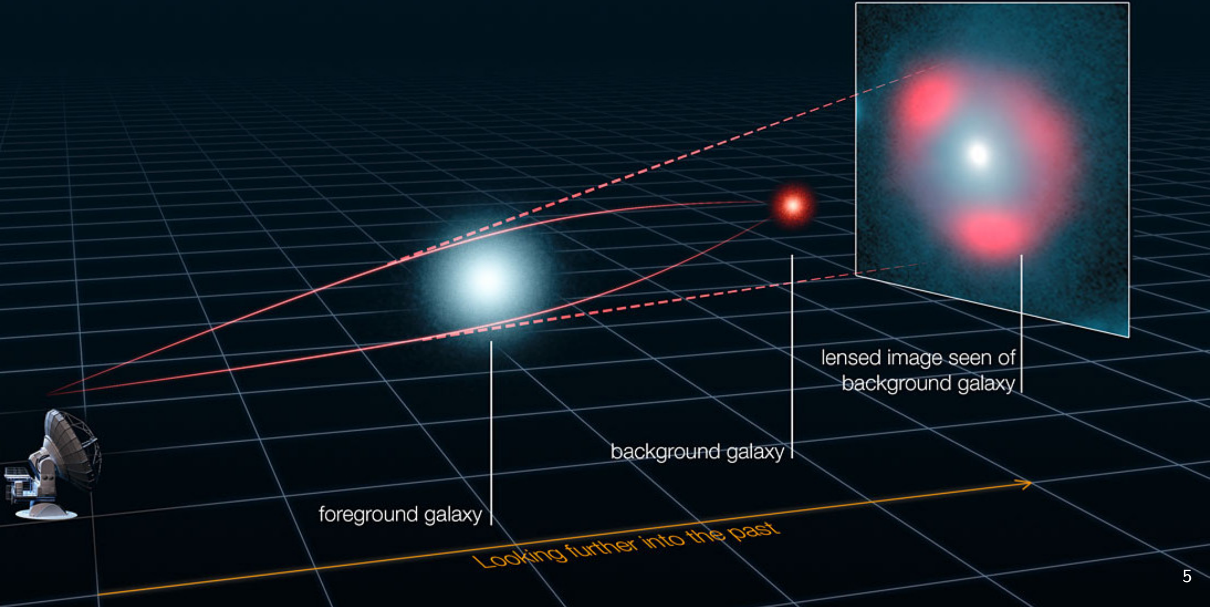
⟹ Dire need for **novel analysis techniques** to fully realize the potential of modern surveys.

Outline of this talk

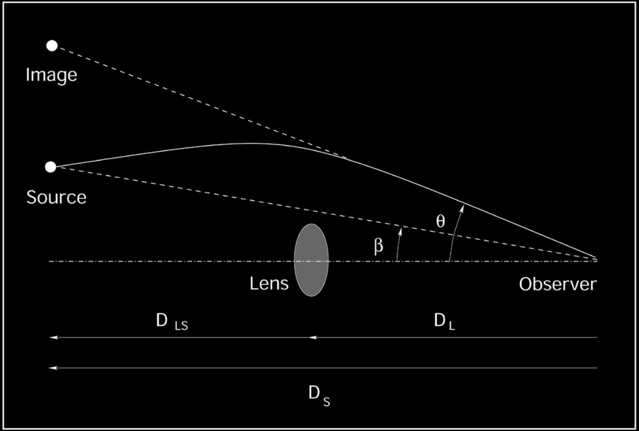
1. Deep residual networks for the detection of gravitational lenses
2. Graph Convolutional Networks for modelling galaxy properties
3. Towards a New Inference Paradigm with Deep Learning
4. Conclusion

Deep residual networks for the detection of gravitational lenses

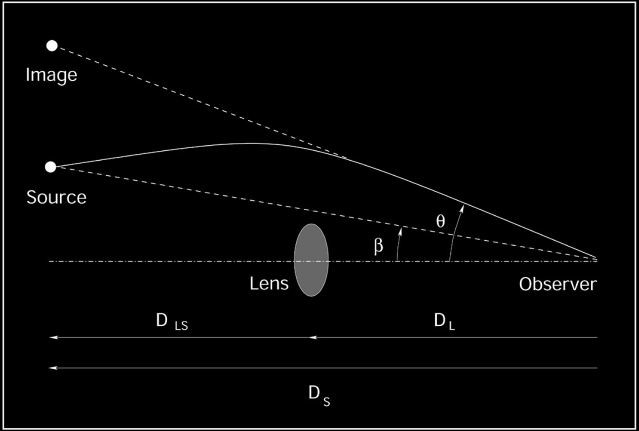
Galaxy-Galaxy Strong Lensing



example of application: gravitational time delays

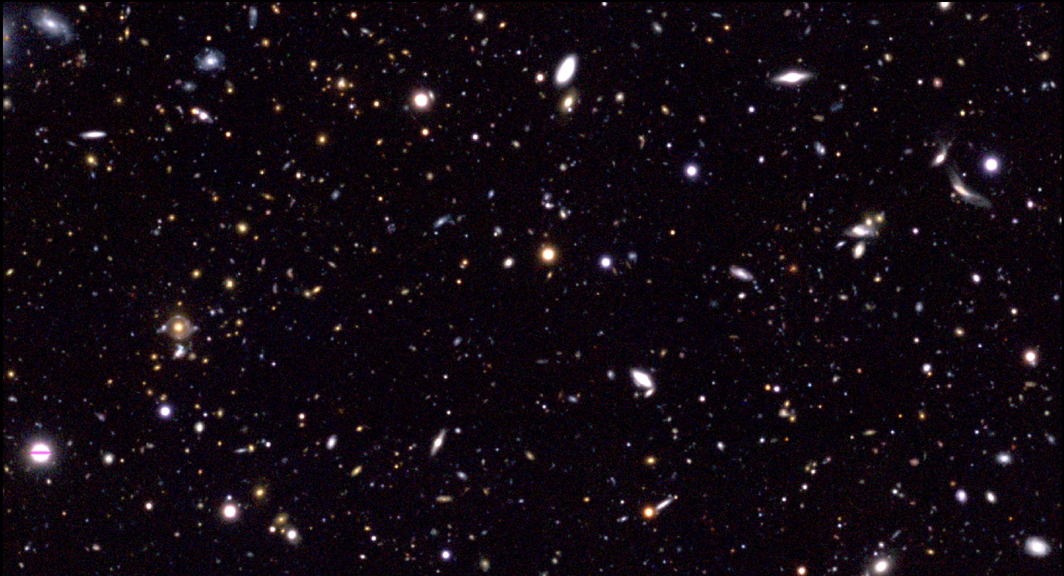


example of application: gravitational time delays



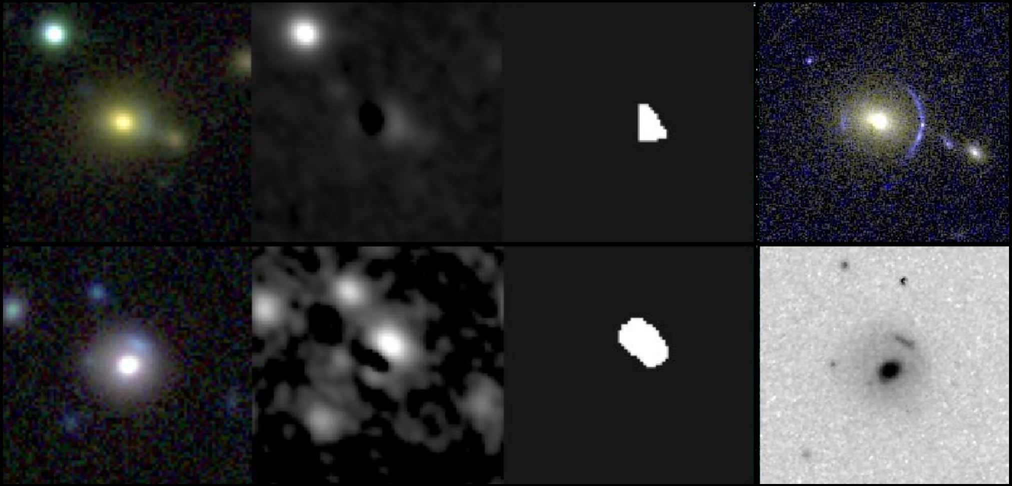
$$\Delta t_{ij} = \frac{1+z_L}{c} \underbrace{\frac{D_L\ D_S}{D_{LS}}}_{\propto H_0^{-1}} \left[\frac{(\boldsymbol{\theta}_i - \boldsymbol{\beta})^2}{2} - \psi(\boldsymbol{\theta}_i) + \frac{(\boldsymbol{\theta}_j - \boldsymbol{\beta})^2}{2} + \psi(\boldsymbol{\theta}_j) \right]$$

the problem: finding strong lenses



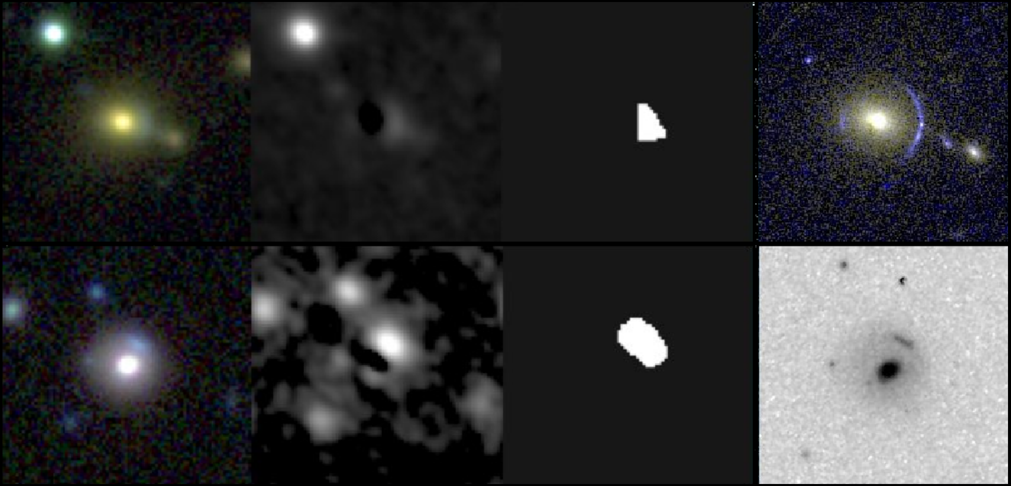
the problem: finding strong lenses

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



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

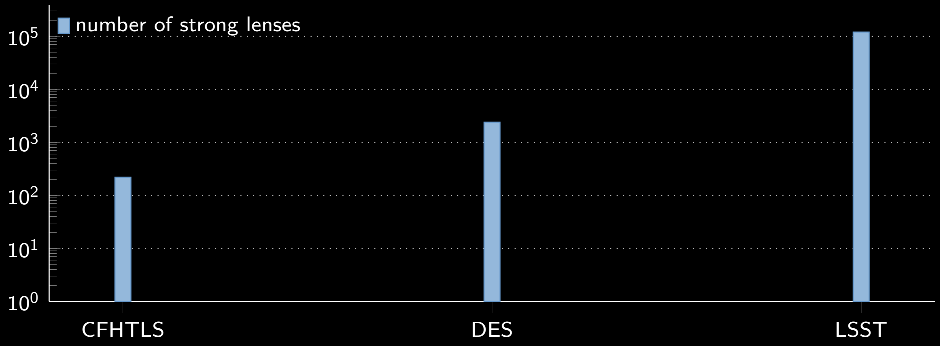
automated lens searches: RingFinder (Gavazzi et al. 2014)



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

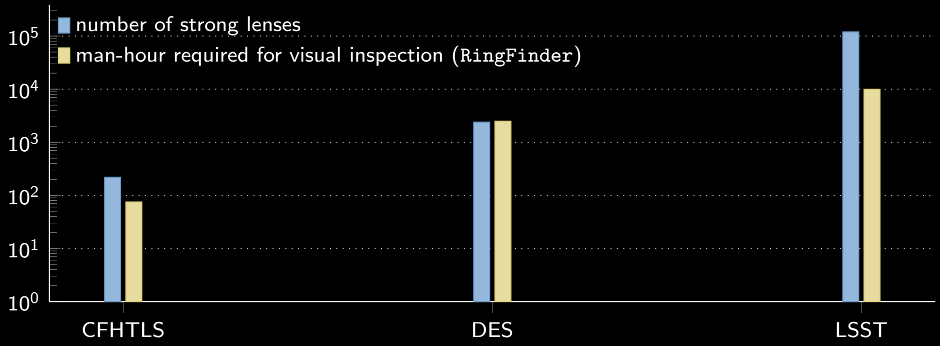
Visual inspection time required: $\sim$ 30 person-minutes / deg^2

extrapolation to future surveys



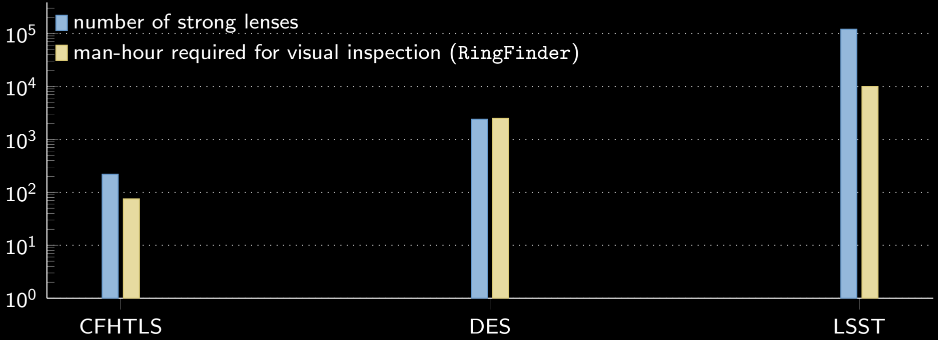
Gavazzi et al. (2014), Collett (2015)

extrapolation to future surveys



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

extrapolation to future surveys

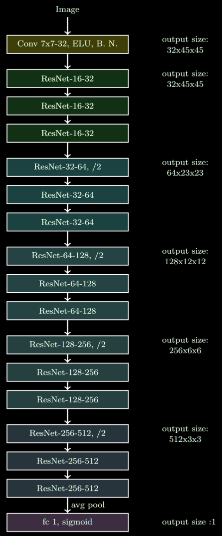


Gavazzi et al. (2014), Collett (2015)

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

How can we robustly detect these rare objects without needing an army of grad students ?

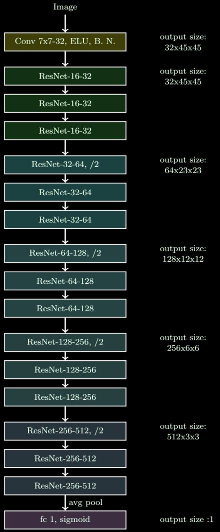
CMU DeepLens: deep residual learning for strong lens finding



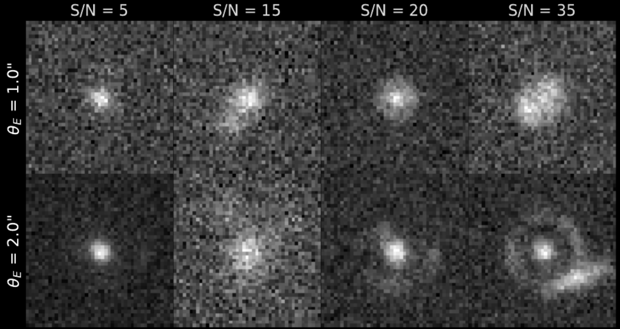
- Deep ResNet (46 layers) with pre-activated bottleneck residual units

Lanusse et al. (2017)

CMU DeepLens: deep residual learning for strong lens finding

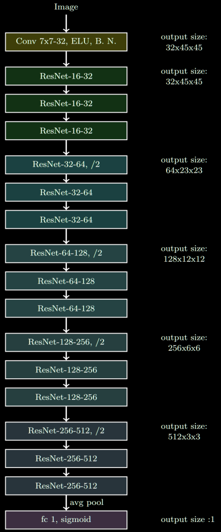


- Deep ResNet (46 layers) with pre-activated bottleneck residual units
- Training on simulated LSST lenses:

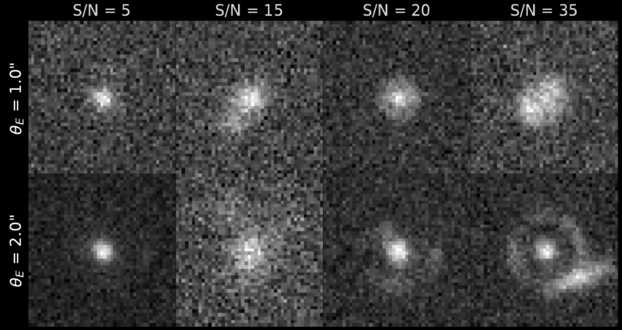


Lanusse et al. (2017)

CMU DeepLens: deep residual learning for strong lens finding



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

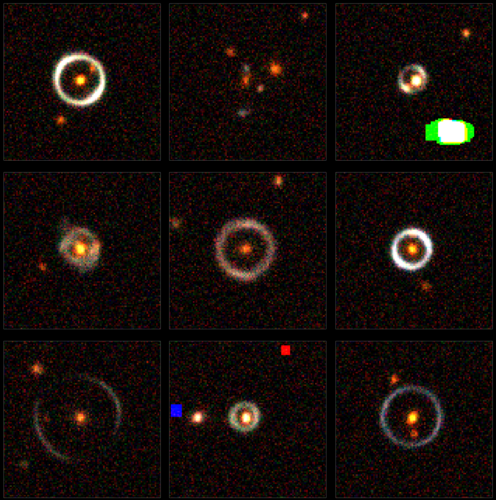


- Classification of 45x45 images in 350 μs
  $\Longrightarrow$ 9 hours to classify a sample of 10^8 lens candidates on single GPU

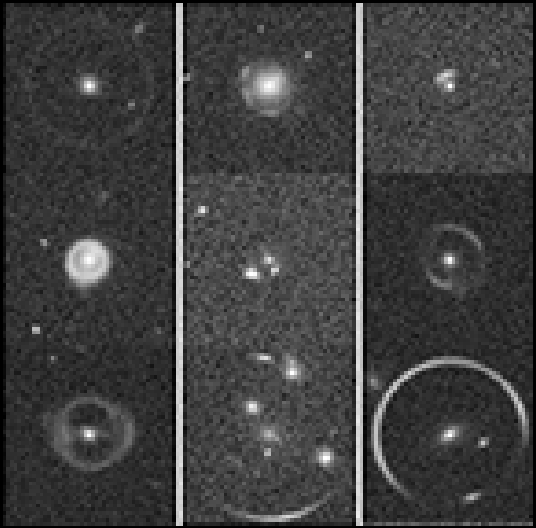
Lanusse et al. (2017)

Euclid strong lens finding challenge

Metcalf, ..., **Lanusse**, et al. (2018)



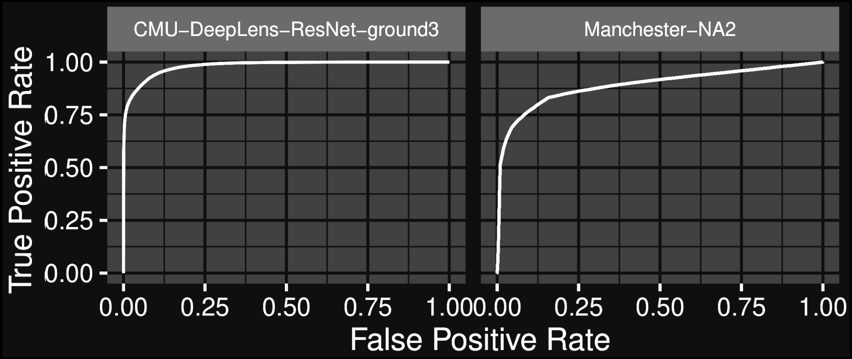
Ground based simulations



Space based simulations

Euclid strong lens finding challenge

Metcalf, . . ., **Lanusse**, et al. (2018)



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

takeaway message

Deep Learning for Low Level Processing

- An example of Deep Learning allowing us to handle the volume and data rate at the **image level**

takeaway message

Deep Learning for Low Level Processing

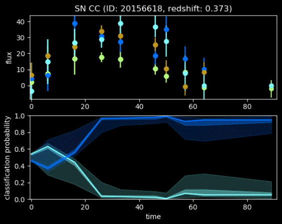
- An example of Deep Learning allowing us to handle the volume and data rate at the **image level**
- Our automated lens finder is faster and more reliable than human volunteers.
  - Larger and more robust samples for the science analysis.

takeaway message

Deep Learning for Low Level Processing

- An example of Deep Learning allowing us to handle the volume and data rate at the **image level**
- Our automated lens finder is faster and more reliable than human volunteers.
    - Larger and more robust samples for the science analysis.

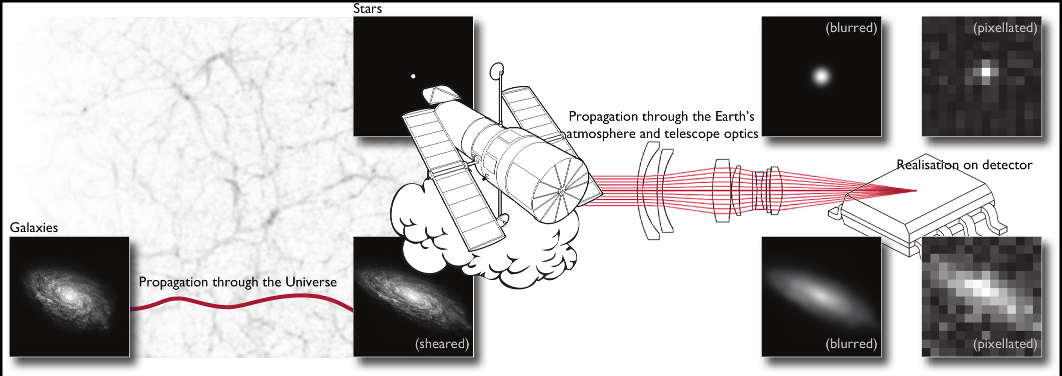
Many other applications of classifications, for instance for time series classification:



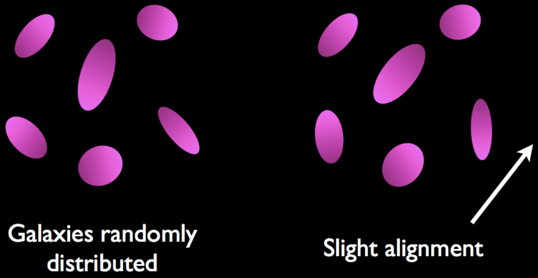
- Bayesian Recurrent Neural Networks for supernovae detection (Moller & De Boissiere, 2019), arXiv:1901.06384

Graph Convolutional Networks for modelling galaxy properties

weak gravitational lensing and the intrinsic alignment of galaxies

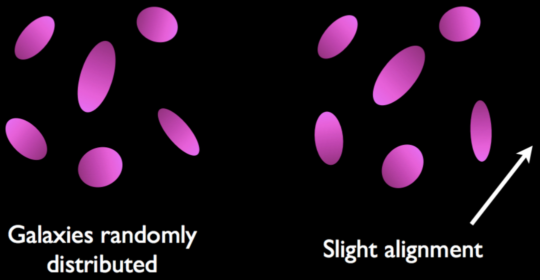


weak gravitational lensing and the intrinsic alignment of galaxies



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

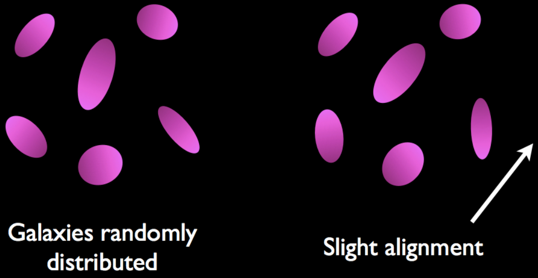
weak gravitational lensing and the intrinsic alignment of galaxies



$\epsilon = \epsilon_i + \gamma$ with $\underbrace{< \epsilon_i \epsilon_i' >= 0}_{\text{not completely true}}$

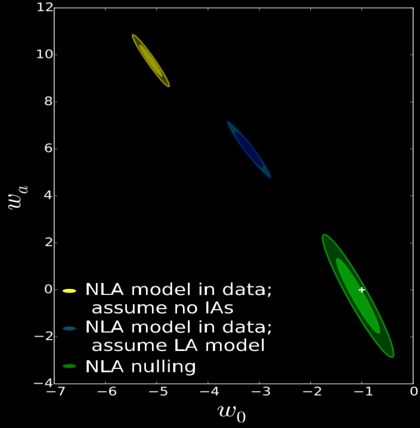
$$\underbrace{< \epsilon\epsilon' >}_{\text{measured}} = \underbrace{< \gamma\gamma' >}_{\text{cosmological signal}} + \underbrace{< \epsilon_i \epsilon_i' >}_{\text{II}} + \underbrace{< \gamma\epsilon_i' > + < \epsilon_i \gamma' >}_{\text{GI}}$$

weak gravitational lensing and the intrinsic alignment of galaxies



$\epsilon = \epsilon_i + \gamma$ with $\underbrace{< \epsilon_i \epsilon_i' >= 0}_{\text{not completely true}}$

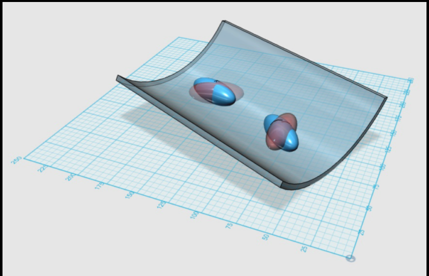
Impact on dark energy constraints



Kirk et al. (2015)

$$\underbrace{< \epsilon\epsilon' >}_{\text{measured}} = \underbrace{< \gamma\gamma' >}_{\text{cosmological signal}} + \underbrace{< \epsilon_i \epsilon_i' >}_{\text{II}} + \underbrace{< \gamma\epsilon_i' > + < \epsilon_i \gamma' >}_{\text{GI}}$$

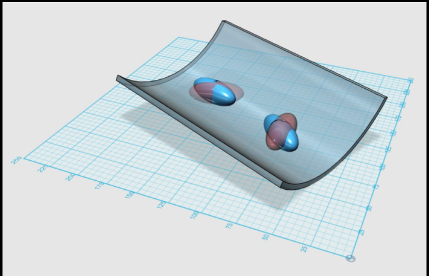
why does this happen ?



Kiessling et al. (2015)

- Tidal interactions with local gravitational potential
  $\Longrightarrow$ Can be analytically modeled on large scales

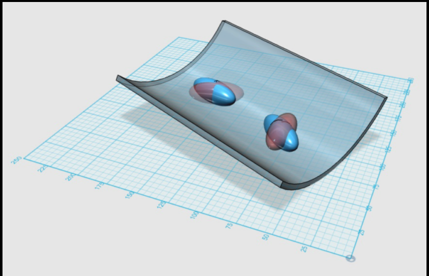
why does this happen ?



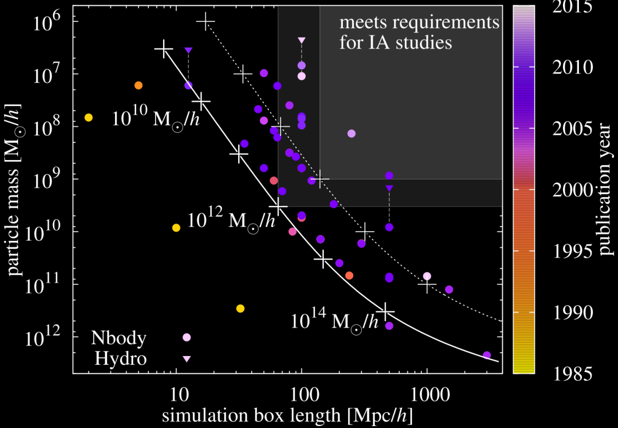
Kiessling et al. (2015)

- Tidal interactions with local gravitational potential
  $\Longrightarrow$ Can be analytically modeled on large scales
- Much more complicated in details, impacted by baryonic physics

why does this happen ?

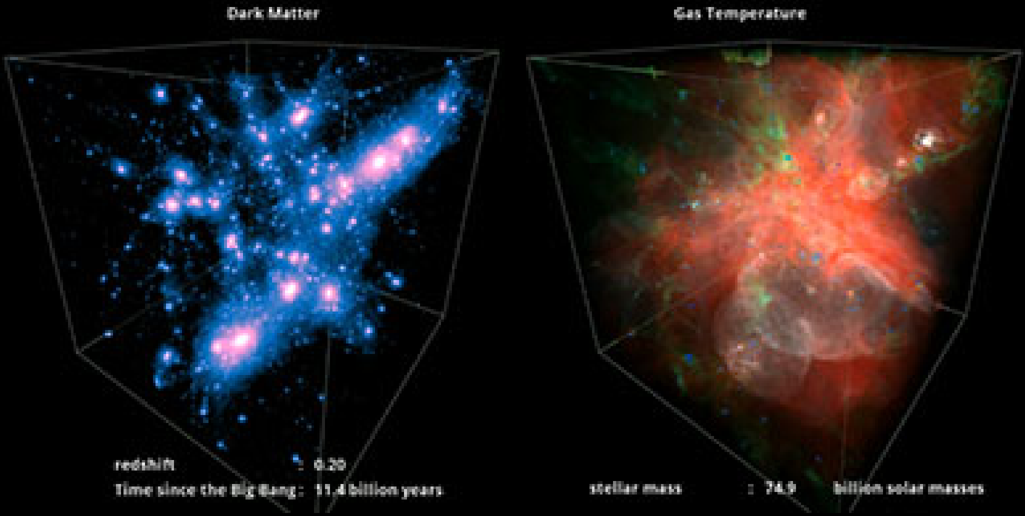


Kiessling et al. (2015)



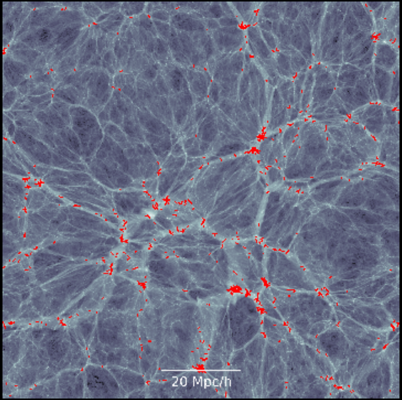
- Tidal interactions with local gravitational potential
  $\Longrightarrow$ Can be analytically modeled on large scales
- Much more complicated in details, impacted by baryonic physics
  $\Longrightarrow$ Study requires expensive hydrodynamical simulations

why galaxy alignments are complicated



How to produce mock galaxy catalogs on large cosmological volumes with realistic alignments ?

inpainting intrinsic aligments on N-body simulations

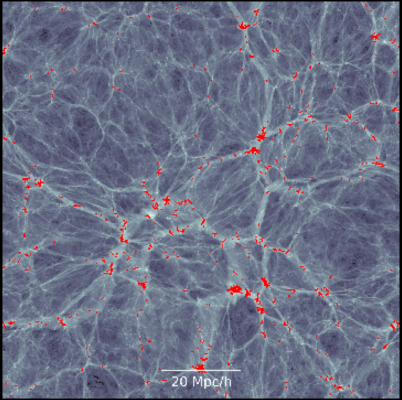


Massive Black II (Khandai et al, 2015)

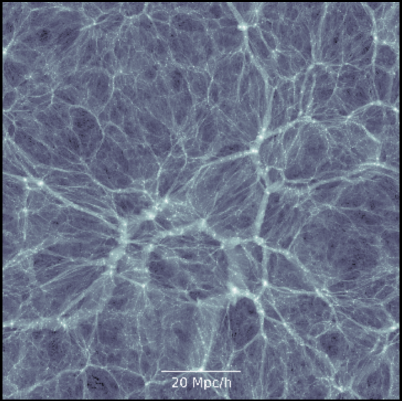
Image credit: Tenneti et al. (2015)

$$p\left(\vec{a}_{3D} \mid x_{DM}, M_{DM}, \ldots\right)$$

inpainting intrinsic aligments on N-body simulations



Massive Black II (Khandai et al, 2015)

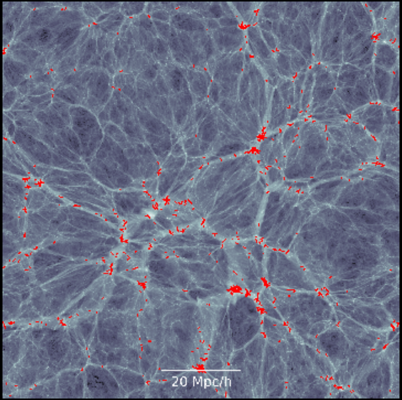


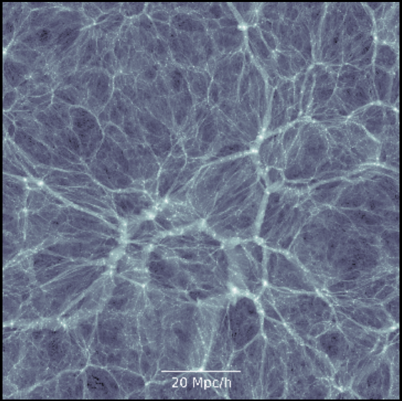
Dark Matter Only

Image credit: Tenneti et al. (2015)

$$\text{gal} \sim p\left(\vec{a}_{3D} \mid x_{DM}, M_{DM}, \ldots\right)$$

inpainting intrinsic aligments on N-body simulations



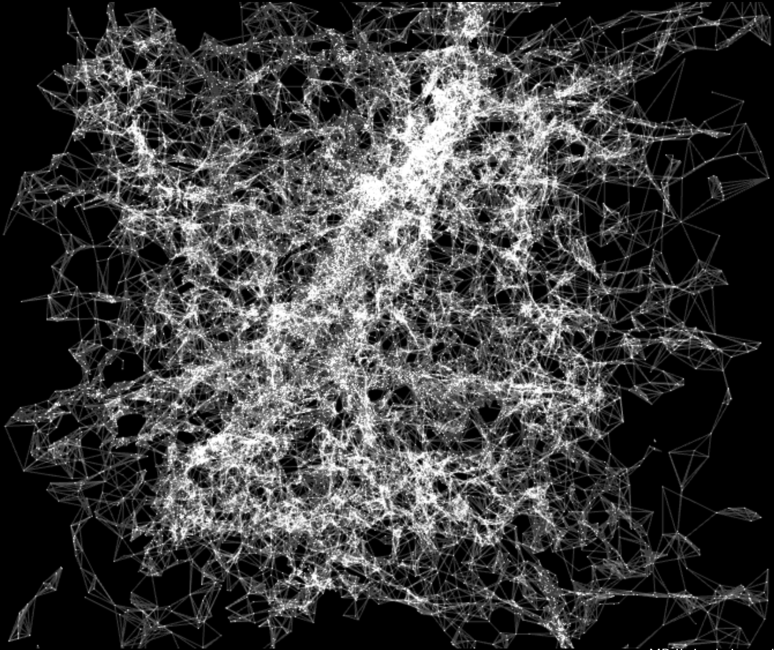


Massive Black II (Khandai et al, 2015)

Image credit: Tenneti et al. (2015)

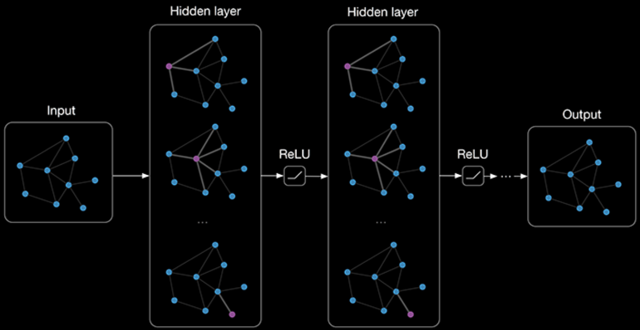
$$\text{gal} \sim p\left(\vec{a}_{3D} \mid x_{DM}, M_{DM}, \ldots\right)$$

$\Longrightarrow$ How to **model** and **sample** from this conditional distribution ?



MB II simulation, animation credit: Kim Albrecht

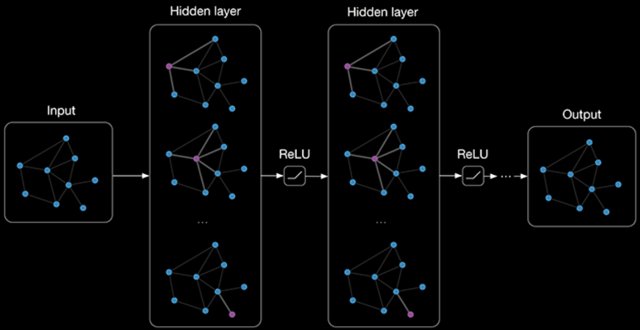
Graph Convolutional Networks (Kipf & Welling, 2017)



Computation of the activation y_i for a node i in the graph:

$$y_i = b + \underbrace{\mathbf{W}_0 h_i}_{\text{self-connection}} + \underbrace{\sum_{j \in \mathcal{N}_i} w_{i,j} \mathbf{W}_1 h_j}_{\text{average over neighbors}}$$

Graph Convolutional Networks (Kipf & Welling, 2017)

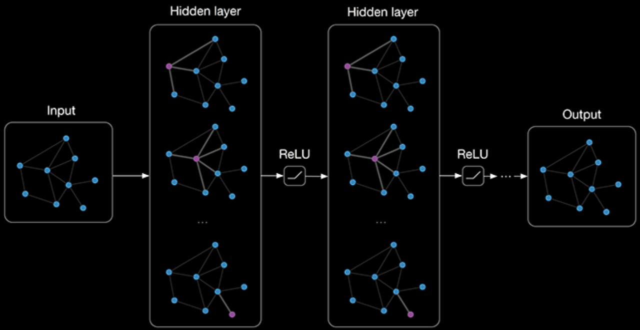


Computation of the activation y_i for a node i in the graph:

$$y_i = b + \underbrace{\mathbf{W}_0 h_i}_{\text{self-connection}} + \underbrace{\sum_{j \in \mathcal{N}_i} w_{i,j} \mathbf{W}_1 h_j}_{\text{average over neighbors}}$$

- Approximation of a spectral convolution on the graph, restricted to first neighbors
  $\implies$ Only requires one multiplication by the sparse graph adjacency matrix

Graph Convolutional Networks (Kipf & Welling, 2017)



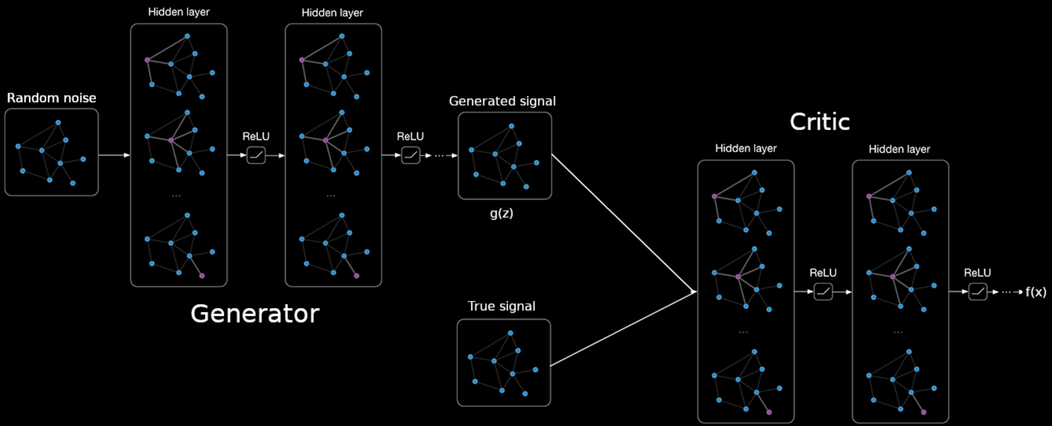
Computation of the activation y_i for a node i in the graph:

$$y_i = b + \underbrace{\mathbf{W}_0 h_i}_{\text{self-connection}} + \underbrace{\sum_{j \in \mathcal{N}_i} w_{i,j} \mathbf{W}_1 h_j}_{\text{average over neighbors}}$$

- Approximation of a spectral convolution on the graph, restricted to first neighbors
  $\Longrightarrow$ Only requires one multiplication by the sparse graph adjacency matrix
- To preserve information about the respective 3D positions of the nodes, we introduce a directional graph convolution:

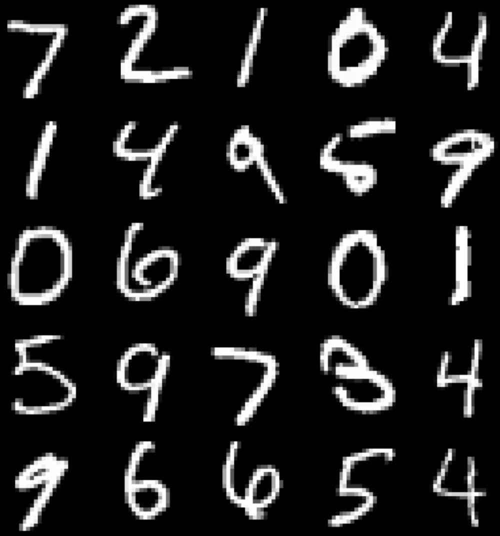
$$y_i = b + \mathbf{W}_0 h_i + \sum_{m=1}^{M} \sum_{j \in \mathcal{N}_i} q_m(\mathbf{x}_i, \mathbf{x}_j)\ w_{i,j}\ \mathbf{W}_m h_j$$

Wasserstein Generative Adversarial Networks on graphs

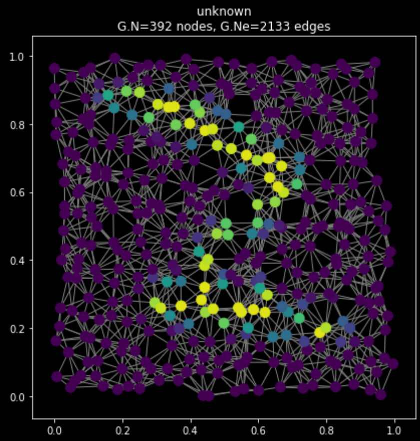


- Simple extension to the graph of a standard Wasserstein GAN, using our graph convolutions

proof of concept on MNIST

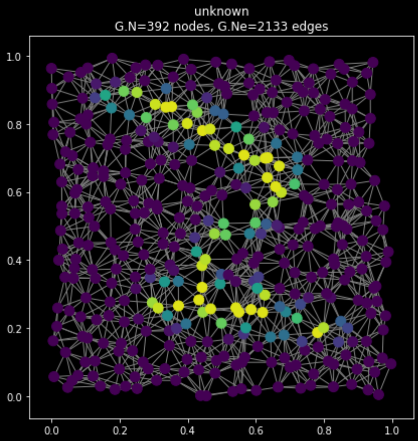


proof of concept on MNIST

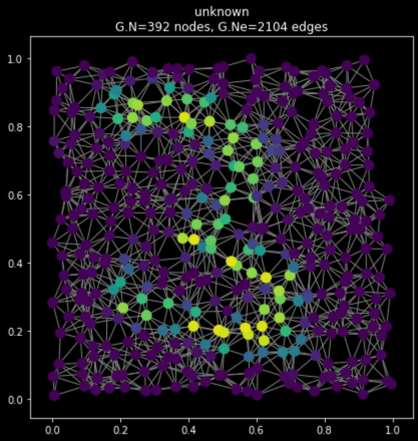


Example of training data

proof of concept on MNIST

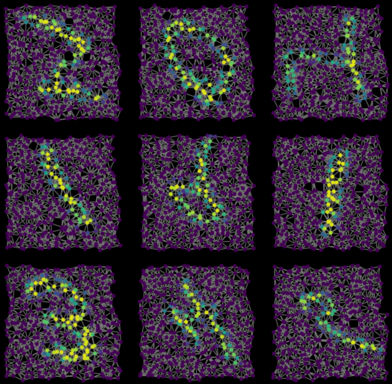


Example of training data

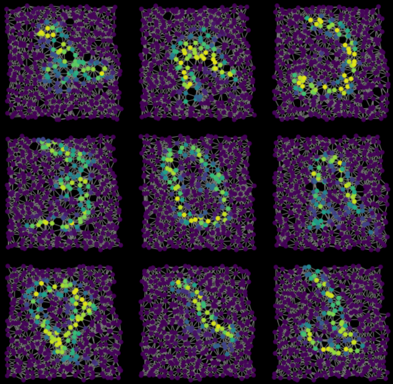


Example of WGAN sample

proof of concept on MNIST

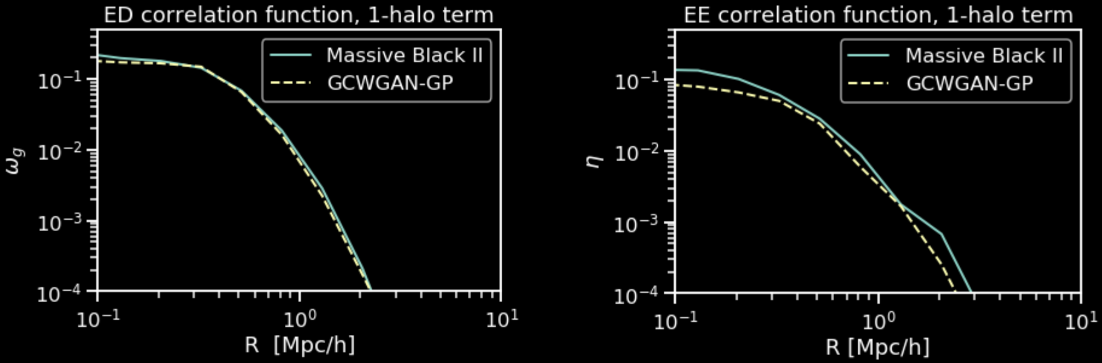


Example of training data



Example of WGAN sample

application to intrinsic alignments



- Successfully samples 3D galaxy orientations with the correct alignment, just from dark matter information

takeaway message

Deep Learning for Improving Cosmological Simulation

- Exciting new framework to empirically populate large volume simulations with realistic galaxy populations

takeaway message

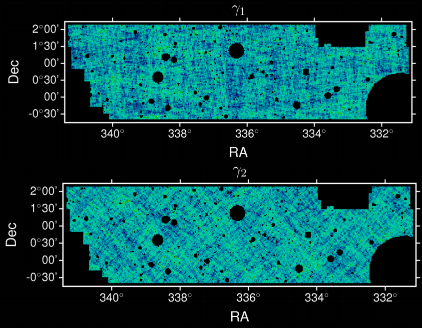
Deep Learning for Improving Cosmological Simulation

- Exciting new framework to empirically populate large volume simulations with realistic galaxy populations

- Will add to the realism of cosmological simulations and allow us to test IA mitigation
  - Being implemented as part of the simulation pipeline for the LSST DESC Second Data Challenge

Towards a New Inference Paradigm with Deep Learning

traditional cosmological inference

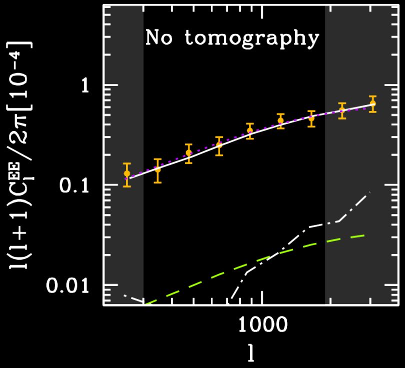


(Alonso et al. 2018)

- Measure the ellipticity $\epsilon = \epsilon_i + \gamma$ of all galaxies
  $\Longrightarrow$ Noisy tracer of the weak lensing shear γ

traditional cosmological inference

HSC cosmic shear power spectrum

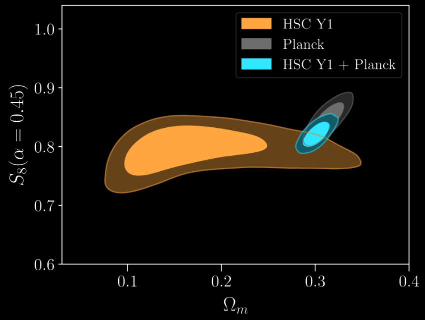


(Hikage, . . . , Lanusse, et al. 2018)

- Measure the ellipticity $\epsilon = \epsilon_i + \gamma$ of all galaxies
  $\Longrightarrow$ Noisy tracer of the weak lensing shear γ
- Compute summary statistics based on 2pt functions, e.g. the **power spectrum**

traditional cosmological inference

HSC Y1 constraints on (S_8, Ω_m)



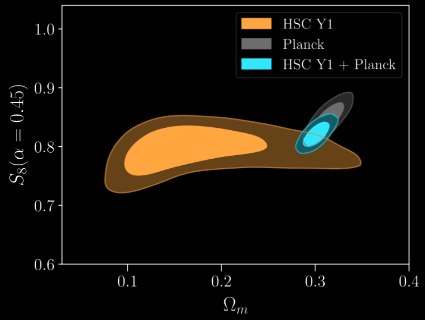
(Hikage, . . . , Lanusse, et al. 2018)

- Measure the ellipticity $\epsilon = \epsilon_i + \gamma$ of all galaxies
  $\Longrightarrow$ Noisy tracer of the weak lensing shear γ
- Compute summary statistics based on 2pt functions, e.g. the **power spectrum**
- Run an MCMC to recover a posterior on model parameters, using an analytic likelihood

$$p(\theta|x) \propto \underbrace{p(x|\theta)}_{\text{likelihood}} \underbrace{p(\theta)}_{\text{prior}}$$

traditional cosmological inference

HSC Y1 constraints on (S_8, Ω_m)



(Hikage, . . . , Lanusse, et al. 2018)

- Measure the ellipticity $\epsilon = \epsilon_i + \gamma$ of all galaxies
  $\Longrightarrow$ Noisy tracer of the weak lensing shear γ
- Compute summary statistics based on 2pt functions, e.g. the **power spectrum**
- Run an MCMC to recover a posterior on model parameters, using an analytic likelihood

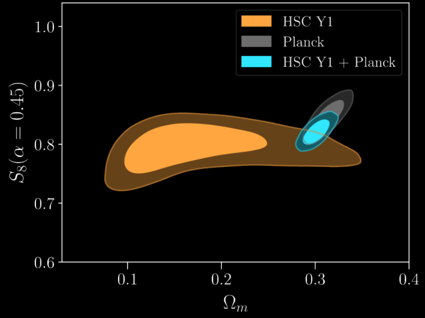
$$p(\theta|x) \propto \underbrace{p(x|\theta)}_{\text{likelihood}} \underbrace{p(\theta)}_{\text{prior}}$$

Main limitation: the need for an explicit likelihood

We can only compute the likelihood for **simple summary statistics** and on **large scales**

traditional cosmological inference

HSC Y1 constraints on (S_8, Ω_m)



(Hikage, . . . , Lanusse, et al. 2018)

- Measure the ellipticity $\epsilon = \epsilon_i + \gamma$ of all galaxies
  $\Longrightarrow$ Noisy tracer of the weak lensing shear γ
- Compute summary statistics based on 2pt functions, e.g. the **power spectrum**
- Run an MCMC to recover a posterior on model parameters, using an analytic likelihood

$$p(\theta|x) \propto \underbrace{p(x|\theta)}_{\text{likelihood}} \underbrace{p(\theta)}_{\text{prior}}$$

Main limitation: the need for an explicit likelihood

We can only compute the likelihood for **simple summary statistics** and on **large scales**

$\Longrightarrow$ We are dismissing most of the information!

Can I use a Deep Learning to perform a proper Bayesian inference without likelihoods?

let us rephrase the question

- I assume a forward model of the observations:

$$p(x) = p(x|\theta)\ p(\theta)$$

  All I ask is the ability to sample from the model, to obtain $\mathcal{D} = \{x_i, \theta_i\}_{i \in \mathbb{N}}$

let us rephrase the question

- I assume a forward model of the observations:

$$p(x) = p(x|\theta)\ p(\theta)$$

  All I ask is the ability to sample from the model, to obtain $\mathcal{D} = \{x_i, \theta_i\}_{i \in \mathbb{N}}$

- I am going to assume $q_\phi(\theta|x)$ a **parametric conditional density**

let us rephrase the question

- I assume a forward model of the observations:

$$p(x) = p(x|\theta)\ p(\theta)$$

  All I ask is the ability to sample from the model, to obtain $\mathcal{D} = \{x_i, \theta_i\}_{i \in \mathbb{N}}$

- I am going to assume $q_\phi(\theta|x)$ a **parametric conditional density**

- Optimize the parameters ϕ of q_ϕ according to

$$\min_\phi \sum_i - \log q_\phi(\theta_i|x_i)$$

  In the limit of **large number of samples** and **sufficient flexibility**

$$\boxed{q_{\phi^*}(\theta|x) \approx p(\theta|x)}$$

let us rephrase the question

- I assume a forward model of the observations:

$$p(x) = p(x|\theta)\ p(\theta)$$

  All I ask is the ability to sample from the model, to obtain $\mathcal{D} = \{x_i, \theta_i\}_{i \in \mathbb{N}}$

- I am going to assume $q_\phi(\theta|x)$ a **parametric conditional density**

- Optimize the parameters ϕ of q_ϕ according to

$$\min_\phi \sum_i - \log q_\phi(\theta_i | x_i)$$

  In the limit of **large number of samples** and **sufficient flexibility**

$$\boxed{q_{\phi^*}(\theta|x) \approx p(\theta|x)}$$

$\Longrightarrow$ One can asymptotically recover the posterior by optimizing a parametric estimator over the Bayesian joint distribution

let us rephrase the question

- I assume a forward model of the observations:

$$p(x) = p(x|\theta)\ p(\theta)$$

  All I ask is the ability to sample from the model, to obtain $\mathcal{D} = \{x_i, \theta_i\}_{i \in \mathbb{N}}$

- I am going to assume $q_\phi(\theta|x)$ a **parametric conditional density**

- Optimize the parameters ϕ of q_ϕ according to

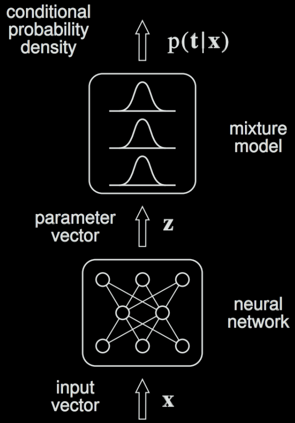
$$\min_{\phi} \sum_i - \log q_\phi(\theta_i | x_i)$$

  In the limit of **large number of samples** and **sufficient flexibility**

$$\boxed{q_{\phi^*}(\theta|x) \approx p(\theta|x)}$$

$\Longrightarrow$ One can asymptotically recover the posterior by optimizing a Deep Neural Network over a simulated training set

Neural Density Estimation



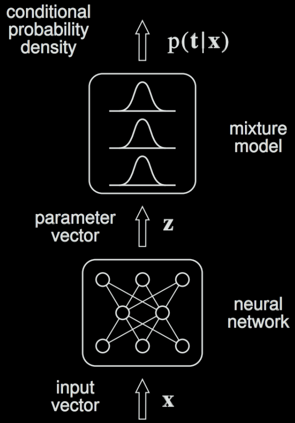
Bishop (1994)

- Mixture Density Networks (MDN)

$$p(\theta|x) = \prod_i \pi_i(x)\, \mathcal{N}\left(\mu_i(x),\ \sigma_i(x)\right)$$

DELFI method (Alsing et al. 2018) based on Papamakarios & Murray (2016)

Neural Density Estimation



Bishop (1994)

- Mixture Density Networks (MDN)

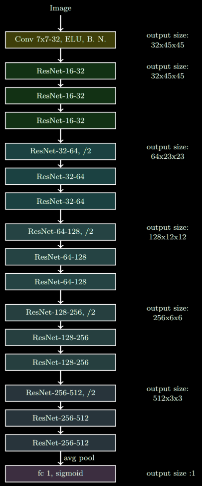
$$p(\theta|x) = \prod_i \pi_i(x)\, \mathcal{N}\left(\mu_i(x),\ \sigma_i(x)\right)$$

DELFI method (Alsing et al. 2018) based on Papamakarios & Murray (2016)

- Flourishing Machine Learning literature on density estimators

GLOW, (Kingma & Dhariwal, 2018)

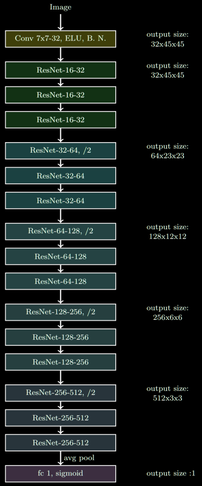
deep residual network for amortized inference



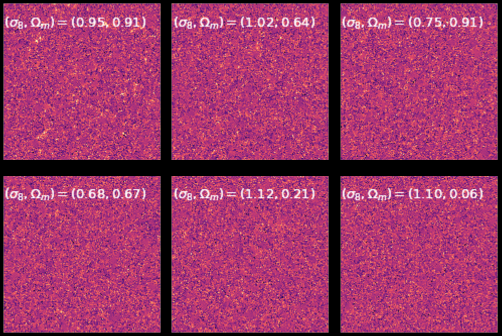
- Deep Residual Network with mixture density output

(Lanusse & Lin, in prep.)

deep residual network for amortized inference



- Deep Residual Network with mixture density output
- Training on raw weak lensing maps simulated for different cosmologies

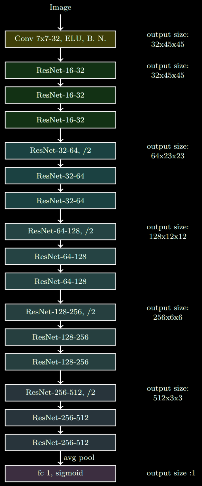


TensorFlow

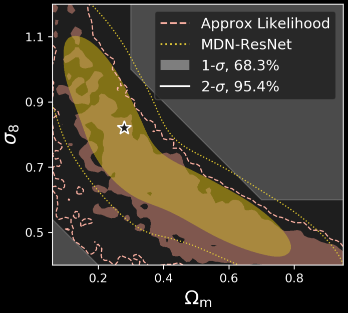
Google Cloud Platform

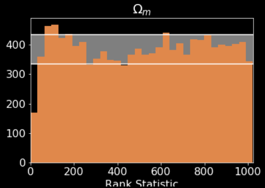
(Lanusse & Lin, in prep.)

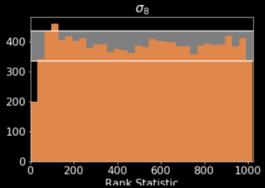
deep residual network for amortized inference



- Deep Residual Network with mixture density output
- Training on raw weak lensing maps simulated for different cosmologies
- Parameter constraints and posterior validation by Simulation-Based Calibration (Talts et al. 2018):

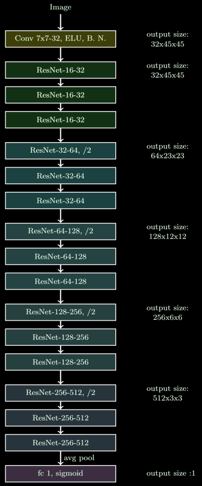




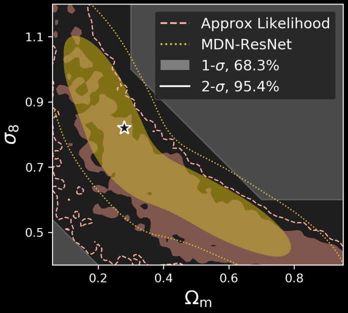


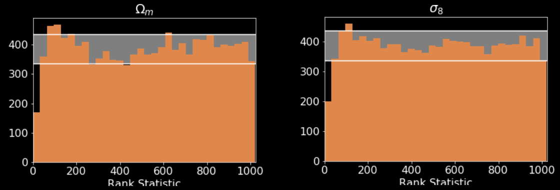
(Lanusse & Lin, in prep.)

deep residual network for amortized inference



- Deep Residual Network with mixture density output
- Training on raw weak lensing maps simulated for different cosmologies
- Parameter constraints and posterior validation by Simulation-Based Calibration (Talts et al. 2018):

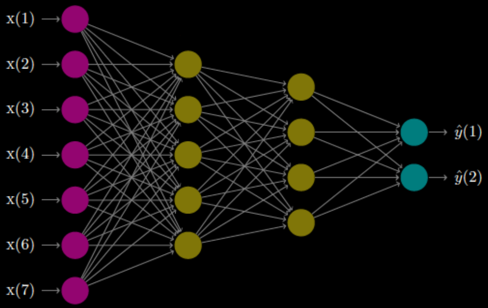




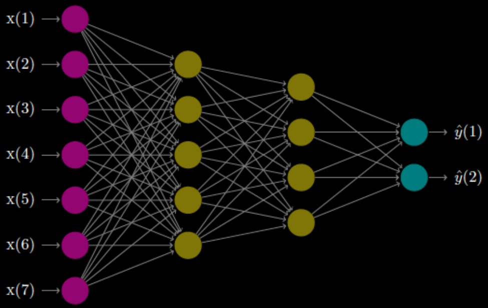
(Lanusse & Lin, in prep.)

$\Longrightarrow$ **Completely automated end-to-end inference methodology**

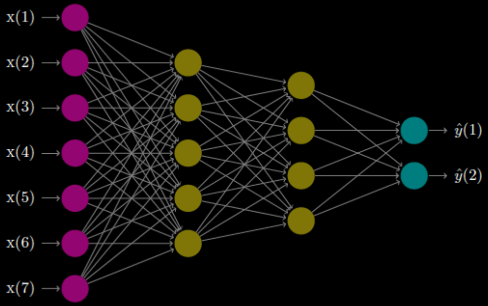
wait.... what about summary statistics?



wait.... what about summary statistics?



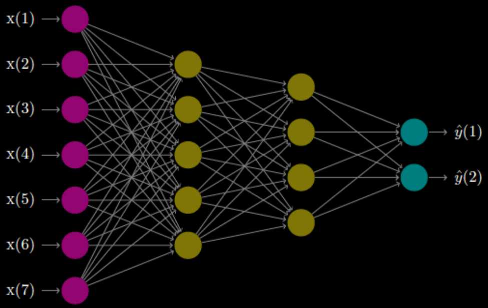
wait.... what about summary statistics?



Variational Mutual Information Maximization

$$I(y;\theta)$$

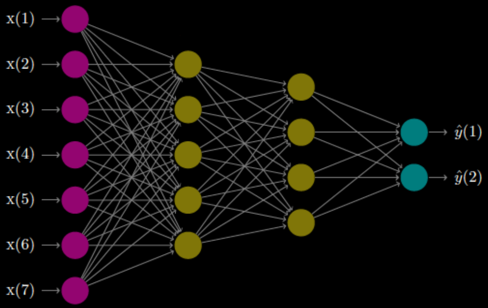
wait.... what about summary statistics?



Variational Mutual Information Maximization

$$I(y;\theta) = \mathbb{E}_{(y,\theta)}[\ \log p(\theta \mid y)\ ] \qquad + H(\Theta)$$

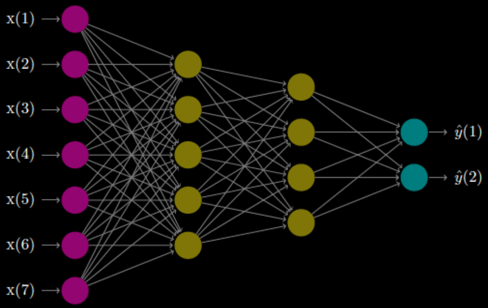
wait.... what about summary statistics?



Variational Mutual Information Maximization

$$I(y;\theta) = \mathbb{E}_{(y,\theta)}[\ \log p(\theta \mid y)\ ] \qquad + H(\Theta)$$
$$\geq \mathbb{E}_{(y,\theta)}[\ \log q_\phi(\theta \mid y)\ ] \qquad + H(\Theta)$$

wait.... what about summary statistics?

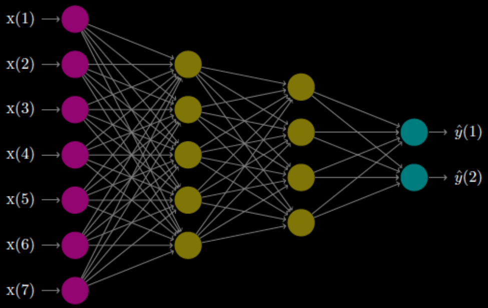


Variational Mutual Information Maximization

$$I(y;\theta) = \mathbb{E}_{(y,\theta)}[\ \log p(\theta \mid y)\ ] \qquad + H(\Theta)$$

- Not derived from Fisher information around a fiducial value, asymptotically optimal over the entire parameter space

wait.... what about summary statistics?



Variational Mutual Information Maximization

$$I(y;\theta) = \mathbb{E}_{(y,\theta)}[\ \log p(\theta \mid y)\ ] \qquad + H(\Theta)$$

- Not derived from Fisher information around a fiducial value, asymptotically optimal over the entire parameter space
- Comes for free by training a deep MDN with a bottleneck
  $\Longrightarrow$ The learned statistics can then be reused with different Likelihood-Free techniques

takeaway message

Deep Learning For Cosmological Inference

- This is part of the broader class of **Likelihood-Free Inference** methods
  - Shifts the physics from signal modeling and statistics extraction to simulation

$\Longrightarrow$ Will be essential to maximize the scientific return of Stage IV surveys.

Conclusion

conclusion

What can deep learning do for cosmology ?

- Open new and powerful ways to look at the data
  - Image detection for finding rare astrophysical objects

conclusion

What can deep learning do for cosmology ?

- Open new and powerful ways to look at the data
    - Image detection for finding rare astrophysical objects
- Data driven way of complementing our physical models
    - Modeling galaxy properties in numerical simulations

conclusion

What can deep learning do for cosmology ?

- Open new and powerful ways to look at the data
  - Image detection for finding rare astrophysical objects
- Data driven way of complementing our physical models
  - Modeling galaxy properties in numerical simulations
- New strategies for inference for increasingly complex surveys

Thank you !