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
the $\Lambda$CDM view of the Universe

- Inflation
- Quantum Fluctuations
- Afterglow Light Pattern 375,000 yrs.
- Dark Ages
- Development of Galaxies, Planets, etc.
- Dark Energy Accelerated Expansion

Big Bang Expansion
13.77 billion years
the $Λ$CDM view of the Universe

- Afterglow Light Pattern 375,000 yrs.
- Dark Ages
- Development of Galaxies, Planets, etc.
- Dark Energy Accelerated Expansion
- Inflation
- Quantum Fluctuations
- 1st Stars about 400 million yrs.
- Big Bang Expansion 13.77 billion 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
  ⇒ 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
  $\Rightarrow$ 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
  \[\Rightarrow 15 \text{ TB/night for 10 years}\]
- 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

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

**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.
the challenge for modern surveys

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

x20 more precise
the challenge for modern surveys

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

x20 more precise
the challenge for modern surveys

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

<table>
<thead>
<tr>
<th>A Blessing</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Unprecedented statistical power</td>
</tr>
<tr>
<td>• Great potential for new discoveries</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A Curse</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Existing methods are reaching their limits at every step of the science analysis</td>
</tr>
<tr>
<td>• Control of systematic uncertainties becomes paramount</td>
</tr>
</tbody>
</table>

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

LSST forecast on dark energy parameters

x20 more precise
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

foreground galaxy

background galaxy

lensed image seen of background galaxy

Looking further into the past
example of application: gravitational time delays

\[ \Delta t_{ij} = 1 + \frac{z L}{c D_L D_S} \propto H^{-10} \left[ (\theta_i - \beta)^2 - \psi(\theta_i) + (\theta_j - \beta)^2 + \psi(\theta_j) \right] \]
example of application: gravitational time delays

\[ \Delta t_{ij} = \frac{1 + z_L}{c} \frac{D_L D_S}{D_{LS}} \left( \frac{(\theta_i - \beta)^2}{2} - \psi(\theta_i) + \frac{(\theta_j - \beta)^2}{2} + \psi(\theta_j) \right) \]

\[ \propto H_0^{-1} \]
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

Visual inspection time required: $\sim 30$ person-minutes / deg$^2$
automated lens searches: RingFinder (Gavazzi et al. 2014)

Visual inspection time required: \(\sim 30 \text{ person-minutes} / \text{deg}^2\)
extrapolation to future surveys

Gavazzi et al. (2014), Collett (2015)
extrapolation to future surveys

- CFHTLS
- DES
- LSST

10
0
10
1
10
2
10
3
10
4
10
5

Gavazzi et al. (2014, Collett (2015))
extrapolation to future surveys

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

⇒ 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
CMU DeepLens: deep residual learning for strong lens finding

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

![Image of lens candidates]

- S/N = 5
- S/N = 15
- S/N = 20
- S/N = 35

\[ \theta_E = 1.0'' \]

\[ \theta_E = 2.0'' \]

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 $\mu$s
  - $\Rightarrow$ 9 hours to classify a sample of $10^8$ lens candidates on single GPU

Lanusse et al. (2017)
Euclid strong lens finding challenge


Ground based simulations

Space based simulations
• 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

- 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
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

$\epsilon = \epsilon_i + \gamma$

Impact on dark energy constraints

Kirk et al. (2015)

$\epsilon \epsilon' > \epsilon_i \epsilon'_i + \gamma \gamma' + \epsilon_i \gamma'_i + \epsilon'_i \gamma$
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 \[ \langle \epsilon_i \epsilon_i' \rangle = 0 \]
not completely true

\[
\begin{align*}
\langle \epsilon' \epsilon' \rangle &= \langle \gamma' \gamma' \rangle + \langle \epsilon_i \epsilon_i' \rangle + \langle \gamma \epsilon_i' \rangle + \langle \epsilon_i \gamma' \rangle \\
&= \text{measured} + \text{cosmological signal} + \text{II} + \text{GI}
\end{align*}
\]
weak gravitational lensing and the intrinsic alignment of galaxies

\[ \epsilon = \epsilon_i + \gamma \quad \text{with} \quad \langle \epsilon_i \epsilon_i' \rangle = 0 \]

not completely true

\[ \langle \epsilon \epsilon' \rangle = \langle \gamma \gamma' \rangle + \langle \epsilon_i \epsilon_i' \rangle_{\text{II}} + \langle \gamma \epsilon_i' \rangle + \langle \epsilon_i \gamma' \rangle \]

Impact on dark energy constraints

Kirk et al. (2015)
why does this happen?

Kiessling et al. (2015)

- Tidal interactions with local gravitational potential
  \[\Rightarrow\] Can be analytically modeled on large scales
why does this happen?

Kiessling et al. (2015)

- Tidal interactions with local gravitational potential
  \[ \Rightarrow \] 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
  \[ \implies \text{Can be analytically modeled on large scales} \]
- Much more complicated in details, impacted by baryonic physics
  \[ \implies \text{Study requires expensive hydrodynamical simulations} \]
why galaxy alignments are complicated
How to produce mock galaxy catalogs on large cosmological volumes with realistic alignments?
inpainting intrinsic alignments on N-body simulations

Massive Black II (Khandai et al, 2015)

Image credit: Tenneti et al. (2015)

\[ p (\ddot{a}_{3D} \mid x_{DM}, M_{DM}, \ldots) \]
inpainting intrinsic aligments on N-body simulations

Massive Black II (Khandai et al, 2015)

Dark Matter Only

Image credit: Tenneti et al. (2015)

\[ \text{gal} \sim p (\vec{a}_{3D} | x_{DM}, M_{DM}, \ldots) \]
inpainting intrinsic alignments on N-body simulations

Massive Black II (Khandai et al, 2015)

Image credit: Tenneti et al. (2015)

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

⇒ How to **model** and **sample** from this conditional distribution?
Graph Convolutional Networks (Kipf & Welling, 2017)

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

$$y_i = b + W_0 h_i + \sum_{j \in \mathcal{N}_i} w_{i,j} W_1 h_j$$

- Approximation of a spectral convolution on the graph, restricted to first neighbors
- 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:
Graph Convolutional Networks (Kipf & Welling, 2017)

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

$$y_i = b + \underbrace{W_0 h_i}_\text{self-connection} + \sum_{j \in N_i} w_{i,j} \underbrace{W_1 h_j}_\text{average over neighbors}$$

- 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:

$$y_i = b + \mathbf{W}_0 h_i + \sum_{j \in \mathcal{N}_i} w_{i,j} \mathbf{W}_1 h_j$$

- Approximation of a spectral convolution on the graph, restricted to first neighbors
  $\implies$ 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:

$$y_i = b + \mathbf{W}_0 h_i + \sum_{m=1}^{M} \sum_{j \in \mathcal{N}_i} q_m(x_i, x_j) w_{i,j} \mathbf{W}_m h_j$$
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
Deep Learning for Improving Cosmological Simulation

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

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

(Alonso et al. 2018)
traditional cosmological inference

HSC cosmic shear power spectrum

- Measure the ellipticity \( \epsilon = \epsilon_i + \gamma \) of all galaxies
  \( \Rightarrow \) Noisy tracer of the weak lensing shear \( \gamma \)

- Compute summary statistics based on 2pt functions, e.g. the power spectrum

(Hikage, . . . , Lanusse, et al. 2018)
traditional cosmological inference

HSC Y1 constraints on \((S_8, \Omega_m)\)

- Measure the ellipticity \(\epsilon = \epsilon_i + \gamma\) of all galaxies \(\Rightarrow\) Noisy tracer of the weak lensing shear \(\gamma\)
- 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

\[
p(\theta|x) \propto p(x|\theta) p(\theta)
\]

Main limitation: the need for an explicit likelihood
We can only compute the likelihood for simple summary statistics and on large scales \(\Rightarrow\) We are dismissing most of the information!

(Hikage, . . . , Lanusse, et al. 2018)
traditional cosmological inference

HSC Y1 constraints on \((S_8, \Omega_m)\)

- Measure the ellipticity \(\epsilon = \epsilon_i + \gamma\) of all galaxies
  \(\implies\) Noisy tracer of the weak lensing shear \(\gamma\)
- 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
  \[
  p(\theta|x) \propto p(x|\theta) p(\theta)
  \]

\(\text{(Hikage, \ldots, Lanusse, et al. 2018)}\)

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

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

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

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

**Main limitation: the need for an explicit likelihood**

We can only compute the likelihood for simple summary statistics and on large scales $\implies$ 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 \( D = \{x_i, \theta_i\}_{i \in \mathbb{N}} \)
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 \( D = \{x_i, \theta_i\}_{i \in \mathbb{N}} \)

- 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 \( D = \{ x_i, \theta_i \}_{i \in \mathbb{N}} \)

• I am going to assume \( q_{\phi}(\theta|x) \) a **parametric conditional density**

• Optimize the parameters \( \phi \) of \( q_{\phi} \) according to

\[
\min_{\phi} \sum_{i} - \log q_{\phi}(\theta_i|x_i)
\]

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

\[
q_{\phi^*}(\theta|x) \approx p(\theta|x)
\]
let us rephrase the question

• I assume a forward model of the observations:

\[ p(x) = p(x \mid \theta) \ p(\theta) \]

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

• I am going to assume \( q_{\phi}(\theta \mid x) \) a **parametric conditional density**

• Optimize the parameters \( \phi \) of \( q_{\phi} \) according to

\[
\min_{\phi} \sum_i - \log q_{\phi}(\theta_i \mid x_i)
\]

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

\[
q_{\phi}^*(\theta \mid x) \approx p(\theta \mid x)
\]

\[\implies\] 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 \( D = \{ x_i, \theta_i \}_{i \in \mathbb{N}} \)

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

- Optimize the parameters \( \phi \) of \( q_{\phi} \) according to

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

In the limit of large number of samples and sufficient flexibility

\[ q_{\phi^*}(\theta|x) \approx p(\theta|x) \]

\[ \implies \text{One can asymptotically recover the posterior by optimizing a Deep Neural Network over a simulated training set} \]
Neural Density Estimation

- Mixture Density Networks (MDN)

\[ p(\theta|x) = \prod_i \pi_i(x) \mathcal{N}(\mu_i(x), \sigma_i(x)) \]


Bishop (1994)
Neural Density Estimation

- **Mixture Density Networks (MDN)**

\[
p(\theta|x) = \prod_i \pi_i(x) \mathcal{N}(\mu_i(x), \sigma_i(x))
\]


- Flourishing Machine Learning literature on density estimators

Bishop (1994)

GLOW, (Kingma & Dhariwal, 2018)
• Deep Residual Network with mixture density output
Deep residual network for amortized inference

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

(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.)

⇒ Completely automated end-to-end inference methodology
wait.... what about summary statistics?

\[
I(y; \theta) = E(y, \theta) [ \log p(\theta | y) ] + H(\Theta) \geq E(y, \theta) [ \log q_{\phi}(\theta | y) ] + H(\Theta)
\]

• 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

The learned statistics can then be reused with different Likelihood-Free techniques
wait.... what about summary statistics?

\[ I(y; \theta) = \mathbb{E}(y, \theta) \left[ \log p(\theta | y) \right] + \text{H}(\Theta) \geq \mathbb{E}(y, \theta) \left[ \log q_{\phi}(\theta | y) \right] + \text{H}(\Theta) \]

• 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

The learned statistics can then be reused with different Likelihood-Free techniques
wait.... what about summary statistics?

Variational Mutual Information Maximization

\[ I(y; \theta) \]
wait.... what about summary statistics?

Variational Mutual Information Maximization

\[ I(y; \theta) = \mathbb{E}_{(y, \theta)} \left[ \log p(\theta \mid y) \right] + H(\Theta) \]
wait.... what about summary statistics?

Variational Mutual Information Maximization

\[ I(y; \theta) = \mathbb{E}_{(y, \theta)}[ \log p(\theta | y) ] + H(\Theta) \]
\[ \geq \mathbb{E}_{(y, \theta)}[ \log q_\phi(\theta | y) ] + H(\Theta) \]
Variational Mutual Information Maximization

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

- 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

\[ I(y; \theta) = \mathbb{E}_{(y, \theta)} \left[ \log p(\theta | y) \right] + H(\Theta) \]

- 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
  \[ \implies \] The learned statistics can then be reused with different Likelihood-Free techniques
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

⇒ Will be essential to maximize the scientific return of Stage IV surveys.
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
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
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!