Addressing key science challenges with Adversarial Neural Networks

HPC User Forum, Tuscon, AZ
April 17 2018

Wahid Bhimji
(feat. Mustafa Mustafa, Michela Paganini, Adrian Albert and others)
Berkeley Lab (LBNL)
Outline

• **Introduction**
  – Deep learning
  – (Generative) Adversarial Networks (GAN)

• **Applications to Science**
  – CosmoGAN
  – CaloGAN
  – Urban Systems

• **Challenges, directions and other applications of adversarial approaches**
Introductions
NERSC

• Mission high-performance computing center for US Dept. of Energy
  – >7000 diverse users across science (e.g. cosmology, climate, biosciences)

• Cori – NERSC’s Newest Supercomputer – Cray XC40 (31.4 PF Peak)
  – 9668 Intel Knights Landing (KNL) nodes

• Data and analytics (DAS) group:
  – Tools for machine learning;
  – Optimize those for large scales;
  – Encourage cutting-edge methods;
  – Collaborative Projects/ Training
  – Example studies and popular deeplearning frameworks available
Deep learning

• **Neutral Networks (NN) with multiple layers**
  – Highly non-linear; many weights – huge capacity
  – Enabled by computing and tractable NN architectures (e.g. convolutional (CNN) weight sharing)

• **Industry/academic interest - improvements in:**
  – Methods – e.g. for image processing
  – Software frameworks (Tensorflow, keras etc.)
  – Hardware – GPUs / TPU/ FPGA/ ASICS..

• **Leverage these for science, e.g.:**
  – Improved sensitivity to new phenomena
  – High-dimensional ‘raw’ full-instrument data
  – Model-agnostic discovery
  – Fast / accurate simulation at high-resolution
Adversarial Neural Networks

What this talk might have been:

• **Adversarial examples:**
  – Big issue in industry – possibly relevant for robust ML in science
  – Not covered here

What this talk is about:

• **Generative Adversarial Networks (GAN)**
  – NNs trained against each other in order to generate new examples

• **Adversarial training generally**
  – One non-generative example in science given here (if time)
  – Many more ideas possible

Generative Adversarial Networks GANs

- **Jointly optimize two NNs:**
  - Generator (G) and Discriminator (D)
  - Loss for G/D in opposition
- **Use CNNs for image like data**

---

Goodfellow *et al.*
https://arxiv.org/abs/1406.2661
https://arxiv.org/abs/1701.00160

GANs and Science

• Can build on industry research
  – e.g. convolutional DCGAN; Wasserstein GAN etc.

• GANs can be unstable
  – Searching in a high-d space;
  – difficult optimization;
  – mode collapse

• Science problems have advantages:
  – Underlying physics structure
  – Existing simulation samples
  – Metrics to evaluate

Science Use Case 1: Celebrities!

Karras et. al. ICLR 2018

CelebA-HQ
1024 × 1024

Latent space interpolations
Science Use-cases
Simulations essential to science
Often computationally expensive (run on HPC)
Sometimes no good simulation exists
Cosmology
CosmoGAN

• Cosmology simulations create ‘virtual universes’.
  – ‘Q-continuum’ HACC sim took ~2 weeks on a Cori-sized HPC resource
  – One product is weak lensing convergence maps, to compare to observations
• Seek to augment simulations with a learned generative neural net
  – Reproduce convergence maps– images
  – Train using existing simulation maps
  – Once trained produce new examples
• Use DCGAN type architecture on 256x256 images

Mustafa Mustafa, Deborah Bard, Wahid Bhimji, Rami Al-Rfou, Zarija Lukić
https://arxiv.org/abs/1706.02390
Summary statistics

- Calculate power spectrum for generated images and validation sample
  - Fourier transform of 2pt correlation
  - Excellent agreement (K-S p_value > 0.995 for 246/248 moments)
  - GAN not explicitly trained to reproduce these distributions
  - Also higher-order Minkowski functionals are reproduced
Interpolating

• Between two random input vectors - varies smoothly
Particle Physics
CaloGAN

- Collisions in detectors e.g. at the Large Hadron Collider (LHC)
- Need detailed micro-physics detector simulations to model
  - >~50% LHC computing budget ($10^9$ CPU hours)
  - Much of this compute time in a ‘shower’ in part of the detector
  - Faster simulation approaches not precise enough for many applications
CaloGAN

- CaloGAN models a 3-layer calorimeter like that of the ATLAS LHC experiment
- Custom NN design
  - Sparsity
  - High dynamic range
  - Highly location-dependent features
- *Conditioned* with particle Energy as input

Michela Paganini, Luke de Oliveira, Benjamin Nachmann
https://arxiv.org/abs/1705.02355
CaloGAN - results

- Realistic average and individual images
- Generation based on particle energy
  - Allowing parameter extrapolation
- Fast
  - 5M particle showers per min on GPU while full simulation on CPU is 30 showers/min

Average energy deposition per calorimeter layer in the GEANT4 training dataset (top) and in the GAN generated dataset (bottom)

Michela Paganini, Luke de Oliveira, Benjamin Nachmann
https://arxiv.org/abs/1705.02355
Urban Systems
Modeling Urban Patterns

CityNet dataset
City = collection of spatial maps (layers)
Urban form = spatial distribution of inter-related socioeconomic data

population density (LandScan/ORNL, 1km)
building density (GUF/DLR, 0.1km)
nighttime luminosity (VIIRS/NASA, 0.75km)
mask of water areas (DLR)
city administrative bounds (GADM L2)

Spatial raster (remote-sensing) data products
Vector data (polygons)

Use DCGAN-like architecture trained on real cities

Adrian Albert, Emanuele Strano, Jasleen Kaur, Marta Gonzalez
https://arxiv.org/abs/1801.02710
Generated images

Real cities

Synthetic cities

Adrian Albert, Emanuele Strano, Jasleen Kaur, Marta Gonzalez

https://arxiv.org/abs/1801.02710
Interpolating: Bamako

- 10% rate of growth (urbanized land) in the last 10 years
- ~1.8M people (14th largest in Africa)
- Not even minimal survey/curated information on demographics, other socio-economic variables

What are possible development paths for Bamako (e.g., into Kumasi, a more developed city in Ghana)?

Forecast possible development paths

Start: Bamako
End: Kumasi

Adrian Albert, Emanuele Strano, Jasleen Kaur, Marta Gonzalez
https://arxiv.org/abs/1801.02710
Future opportunities and challenges
Opportunities

- **GANs can reproduce science features**
  - Including non-Gaussian structures and derived variables
  - With high-accuracy in many cases
  - And fast

- **Many opportunities – e.g.:**
  - Reliably interpolate in parameters for fast emulator
  - Compress simulations to model and generate on the fly when analyzing
  - Create simulations from data without a physics model

\[ \Omega_c, \Omega_m, \sigma_8, \ldots \]
BUT Challenges

Computational
• Significant compute required for training
• Challenges in distributed training on HPC resources

Methodological
• Issues in stability/convergence
• Insufficiently precise in many cases
• Does it learn target distribution? Can it generalize?
• How to make interpolate in *science* parameter?

Active/important/impactful area of research – to augment not throw out your HPC simulation
Other applications for adversarial networks

• Applications not limited to generation

• For example in using deep learning for classification one related set of objections from scientists are
  – The network learns some mis-modelled variables (Z) in training data
  – It is vulnerable to systematic errors in variables (Z) that are difficult to evaluate
  – OR selection is correlated to other variables (Z’) that is needed elsewhere in analysis

• One solution can be to use adversarial approach to ‘pivot’


Conclusions

• Adversarial networks offer new ways to apply deep learning to tackle science problems beyond improving classification accuracy
  – Potentially transformative applications for simulation
  – Other applications too e.g. de-correlation / systematic errors

• Huge promise in simulating science with high-fidelity many orders of magnitude quicker than currently
  – Just beginning to explore applications and to tackle methodological and computational challenges
We’re Hiring! Machine Learning Engineer/Scientist: 
https://lbl.taleo.net/careersection/2/jobdetail.ftl?job=84752
Generative Adversarial Networks – Loss function

Minimax game formulation (saturating):

\[ J^{(D)} = -\frac{1}{2} \mathbb{E}_{x \sim \mathbb{P}_{data}} \log D(x) - \frac{1}{2} \mathbb{E}_{z \sim p_z} \log (1 - D(G(z))) \]

\[ J^{(G)} = -J^{(D)} \]

Heuristic loss function (non-saturating):

\[ J^{(G)} = -\frac{1}{2} \mathbb{E}_{z \sim p_z} \log D(G(z)) \]
GAN not memorizing training images
Manipulating GAN latent space

\[
z^* = \arg \min_{z \sim \mathcal{N}(0,1)} \|x_{\text{real}} - G(z)\|_2^2
\]

[Lipton, 2017] (ICLR)

"inverting" GANs: recover latent vector \( z \)

compare cities & find similar cities \( z_1 \leftarrow z_2 \)

real

start: san francisco de janeiro

recovered

start: san francisco de janeiro

target: san francisco de janeiro

bldg  pop  lum
Measuring cities: average radial profile

Is the building density profile in sync with the population profile?

$\mathbf{x}$

$\hat{x}(d)$

$\mathcal{R}(d)$

$\sum_{(u,v) \in \mathcal{R}(d)} \mathbf{x}(u,v)$

Building density maps of Paris

Layers of spatial maps

mumbai city (pop 3.4M) paris (pop 2.5M) shanghai (pop 22.5M) rio de janeiro (pop 6.4M) barcelona (pop 5.6M) san francisco (pop 930.4k)
Simulated vs real cities: diversity & radial patterns

**Real cities**

**Synthetic cities**