Science use case 1: Deep Learning in astrophysics

Elise Jennings ALCF Data Science group



ejennings@anl.gov

Deep Learning in Cosmology/Astronomy

Big data challenges:

- Volume of data (Gigabytes, Terabytes, Petabytes, Exabytes)
- Complexity of data
- Rate of data (massive data streams, real time knowledge extraction)
- Big scientific data visualization
- Hardware and software co-design

LSST Operations: Sites & Data Flows

Data and compute sizes: Final volume of raw image data = 60 PB Final catalog size (DR11) = 15 PB Peak compute power in LSST data centers = about 2 PFLOPS

Network bandwidths: Summit (Cerro Pachón) - Base (La Serena = 600 Gbps Base (La Serena) to Archive (NCSA) = 2 x 100 Gbps

Alert Production: Real-time alert latency = 60 seconds Estimated number of alerts per night = up to about 10 million

Data Releases: Number of Data Releases = 11 Images collected = 5.5 million 3.2 Gigapixel images

Estimated counts for DR1 (produced from first 6 months of observing) . Objects = 18 billion; Sources = 350 billion (single epoch); Forced Sources = 0.75 trillion

Estimated counts for DR11 Objects = 37 billion; Sources = 7 trillion (single epoch); Forced Sources = 30 trillion **HQ Site** Tucson, AZ Science Operations Observatory Management Education & Public Outreach

French Site CC-IN2P3, Lyon, France

Satellite Processing Center Data Release Production Long-term Storage (copy 3)

LSST Data Facility National Center for Supercomputing plications (NCSA), Urbana-Champagne,

Processing Center Alert Production Data Release Production Calibration Products Production EPQ Infrastructure Long-term Storage (copy 2) Data Access Center Data Access and User Services

> Summit Site Cerro Pachón, Chile Telescope & Camera Data Acquisition Crosstalk Correction

Base Site La Serena, Chile Base Center Long-term storage (copy 1) Data Access Center Data Access & User Services

Argonne Leadership Computing Facility



Deep Learning in Cosmology/Astronomy

Applications:

- Mock catalogue creation
- Augment N-body simulations
- Detect gravitational lenses
- Detect transients in real time
- Detect gravitational waves in real time
- Classify galaxy images
- Parameter estimation from time series data



Deep Transfer Learning to classify galaxy images





Deep Transfer Learning to classify galaxy images

Network:

Xception + a few custom defined fully connected layers Weights:

pretrained weights with the ImageNet dataset

Data:

resized all the galaxy images 299 × 299 pixels

Method:

Progressively unfreeze earlier layers of the whole network Fine tune their weights for a few epochs of training Retain earlier layers of a trained network: versatile filters for features like lines and edges





Using the network as a feature extractor

Network output:

output activation values from second-to-last layer

3D representation t-Distributed Stochastic Neighbor Embedding (t-SNE)

Addresses common problem of large unlabeled datasets







ADSP mmaADSP Khan et al 2019

FIG. 4: t-SNE visualization of the clustering of HP SDSS and DES test sets, and unlabelled DES test.





Deep Learning driven science

Multimessenger Astrophysics through the NCSA-Argonne Collaboration Pls: Huerta, Zhao, Haas, Saxton (NCSA)

Detecting Gravitational Waves in Real-Time with Deep Learning NCSA WOLFRAM **NVIDIA**. Data from a LIGO Interferometer around the first event (GW150914) see half the provide section of the statement of t معاور والاطريق والمرافقة والمرافقة والمرافعة والمراجعة والمراجعة والمراجعة والمراجعة والاطريقي والاطريقي والمراجعة والم - ▶ K + Convolution Laver 2 Convolution Laver 3 Gravitational Waves Not Detected ection and Parameter Estimation: Results with Advanced LIGO Data – Daniel George and E. A. Huerta (2017)

Novel data-parallel deep learning fusing HPC and AI for MultiMessenger Astrophysics (MMA).

Huge potential for scientific discovery

- Convergence of all-sky GW observations (LIGO) with deep, highcadence electromagnetic observations (LSST)
- Novel visualization of Neural Networks

Deep Learning for Multi-Messenger Astrophysics. A Gateway for Discovery in the Big Data Era, Huerta et al., Nature Review Physics



Deep Learning at Scale for Gravitational Wave Parameter Estimation of Binary Black Hole Mergers

- DL at scale for parameter estimation of Binary Black Hole (BH) mergers (spins are aligned or anti-aligned, evolve on quasi-circular orbits)
- Densely sample 4-D signal manifold ~300,000 simulated waveforms
- Enhance dataset to 10⁷ samples time invariance in the data stream detectors scale invariance in range of SNR Add non-Gaussian and non-stationary noise.
- Distinct NN models to estimate Individual BH masses BH remnants - final spin, GW quasi-normal frequencies Curriculum learning with decreasing SNR
- Inference carried out for each binary BH merger observed so far from LIGO and Virgo detectors Parameters reconstructed within 2 milliseconds Consistent with Bayesian analyses (days to weeks)



Data

Number of features:

8192Hz (nsteps) downsampled compared to real data frequency

Number of samples:

300,000 simulated unique waveforms (1D time series data), augmented to 10^7 waveforms ~28GB. Collaboration with Princeton 50M waveforms

NN for spin of BHs – 300M waveforms

Encoding the data:

HDF5, TFRecords

Feature selection

- in network Root and Leaf model

 $m_1 \in [9M_{\odot}, 65M_{\odot}]$ $m_2 \in [5.2M_{\odot}, 42M_{\odot}]$ $a_{\{1,2\}} \in [-0.8, 0.8]$



Models

Model 1:

- BH remnants final spin, GW quasi-normal frequencies
- HSD-CNN Hierarchically self decomposing CNN (SaiRam et al 2018), subnetworks for specific set of classes
- EraseReLU (Dong et al 2017): erasing ReLUs of certain layers to enhance propagation of useful information



(kernel size, # of output channels, stride, dilation rate, max pooling kernel, max pooling stride)

Root Layer: Convolutional	$\begin{array}{c}(16, 64, 1, 1, 4, 4)\\(16, 128, 1, 2, 4, 4)\\(16, 256, 1, 2, 4, 4)\\(32, 256, 1, 2, 4, 4)\\(4, 128, 1, 2, 0, 0)\\(4, 128, 1, 2, 0, 0)\\(2, 64, 1, 1, 0, 0)\end{array}$	ReLU	
------------------------------	---	------	--

Leaf Layer : FC (1024, 0.0) ReLU FC (1024, 0.0) ReLU FC (1024, 0.0) Identity FC (1, 0.0) Tanh



Models

Model 2:

BH masses •

- Squeeze-Excitation structure(Hu et al 2018) models interdependencies in channels
- 'Highway' block (Srivastava et al 2015) learns the residual components, short cut for learning features



(kernel size, # of output channels, stride, dilation rate, max pooling kernel, max pooling stride)

Layer	Layer	Activation
Component	Configurations	Functions
Root Layer: Convolutional	$\begin{array}{c}(16,64,1,2,4,4)\\(16,128,1,2,4,4)\\(16,128,1,2,4,4)\end{array}$	ReLU
Leaf Layer: SE	$(128,3) \\ (128,3)$	ReLU
Leaf Layer: Highway	(4, 128, 2, 30)	ReLU

Leaf Layer :

Х

Inception ĩ

FC(1024, 0.1) ReLU

FC(1024, 0.1) ReLU

FC(1024, 0.0) Identity

FC(1, 0.0) Tanh



Bayesian Neural Network implementation

L2loss re-defined to be ELBO loss

 Posterior distribution = Normal with network outputs as mean, std could be implemented also

labels_distribution = tfd.Normal(loc=outputs,scale=fixed_var*tf.ones(1))

```
sample_distribution = labels_distribution.sample()
```

Compute the -ELBO as the loss, averaged over the batch size. neg_log_likelihood = tf.reduce_mean(input_tensor=labels_distribution.log_prob(output_vector))

```
KL = sum(model.losses)/ tf.cast(self.N,dtype=tf.float32)
```

```
elbo_Loss = neg_log_likelihood + alpha_KL * KL
```







Thank you !

