CHALLENGES IN MACHINE LEARNING FOR COMPLEX PHYSICAL SYSTEMS
Dr. Christoph Angerer, 09.01.2017
Monitoring Effects of Carbon and Greenhouse Gas Emissions

Minute-by-minute AI Weather Forecasting

92% believe AI will impact their work

93% using deep learning seeing positive results

insideHPC.com Survey November 2016
WHY THE EXCITEMENT?
GPUs as Enablers of Breakthrough Results

(a) Stage-I images
This bird has a yellow belly and tarsus, grey back, wings, and brown throat, nape with a black face

(b) Stage-II images
This bird is white with some black on its head and wings, and has a long orange beak
This flower has overlapping pink pointed petals surrounding a ring of short yellow filaments

ImageNet — Accuracy %

Huma n
Deep Learning
Hand-coded CV

2010 2011 2012 2013 2014 2015
72% 74% 74% 79% 93% 98%

65x in 3 Years

We can generate photorealistic images from textual descriptions now!

Achieve super-human accuracy in classification

And we are getting faster fast

AGENDA

A Quick Introduction to Neural Networks
Four Questions and Partial Answers
Concluding Remarks
1-SLIDE INTRO TO CONVOLUTIONAL NEURAL NETS

Forward/Backward Propagation

- Input
- Convolution
- Activation
- Fully-connected
- Loss-Function (Cross Entropy)
- Classification (Softmax)
- Backward-propagation (gradient computation)
- All layers are differentiable

SGD

Optimization Algorithm

weight updates
1-SLIDE INTRO TO RECURRENT NEURAL NETS

Network + Internal State ⇒ Dependencies Over Time

Diagrams from: http://colah.github.io/posts/2015-08-Understanding-LSTMs/
CATEGORIZATION BY SIGNAL

Supervised learning
- Training Data
- Labels (expected results)
- Model
- Testing Data

Unsupervised learning
- Unlabeled Training Data
- Model

Reinforcement learning
- Model
- Environment

One-shot learning
- Very small set of training data
- Model
- Use
CATEGORIZATION BY INPUT/OUTPUT

- **one to one**: Image Classification
- **one to many**: Image Captions
- **many to one**: Sentiment Analysis
- **many to many**: Text Recognition
- **many to many**: Generative (diabolo and others)

Diagram from: http://karpathy.github.io/2015/05/21/rnn-effectiveness/
WHAT IF I DON’T HAVE ENOUGH TRAINING DATA?
HOW MUCH TRAINING DATA IS NEEDED?

A recursive answer

• No general answer, need to experiment
  • Test error >> training error: probably more data (overfitting?)
  • Test error \approx training error: more data probably doesn’t help
  • Look at learned filters: noisy filters generally want more training
• For N functions, need > \log(N)+c training cases (see: A Theory of the Learnable, L.G. Valiant, 1984)
  • Example: N parameters of type float32 = max 2^{32N} distinct networks > 32N samples.
• Rough Guideline: some constant (e.g. 10) multiple of # parameters to avoid overfitting
  • Batch normalization, Regularization, etc can give improvement
HOW LARGE SHOULD MY NETWORK BE?

A recursive answer

• Depends on the amount of training data available
  • Too small: bad generalization; Too large: overfitting
• And the complexity of the function to be learned\(^1\)
  • 1-hidden layer (grows exponentially) vs. deep networks (may grow linearly)
• Rough Design Guideline:
  • First and last layer are given by model
  • Number of nodes of a hidden layer somewhere between the size of its input and output layer
  • Number of nodes in layer should be \(< 2 \times \text{#input nodes}\) to avoid overfitting
• The rest is Art(?)

\(^1\)Y. Bengio, Y. LeCun. Scaling learning algorithms towards AI. Large-scale Kernel Machines, 2007
HOW TO GET MORE TRAINING DATA?
And their Labels

• Data Augmentation and Data Synthesis
  • e.g., adding artificial background noise to speech samples (10x increase for Baidu)
  • e.g., adding shifts, rotations, distortions to images

• Training and Testing on Simulators
  • Google DeepMind Lab
  • Self-Driving Vehicles Playing for Data: Ground Truth from Computer Games, S.Richter et al., ECCV, 2016)
  • Robotics

• One-shot Learning, GANs, Autoencoders?
EXAMPLE: PARTICLE PHYSICS (CERN)

1) We begin with Quantum Field Theory

2) Theory gives detailed prediction for high-energy collisions

3) The interaction of outgoing particles with the detector is simulated.

4) Finally, we run particle identification and feature extraction algorithms on the simulated data as if they were from real collisions.

MY DATA IS SYMMETRIC OR INVARIANT IN XYZ?
INVARIANTS AND SYMMETRIES IN DATA

Pattern Recognition

- CNNs don’t understand Invariants and Symmetries out of the box
  - Pooling and downsampling helps with some transformations
- (Training and Test-time) Data augmentation may explode the training set
  - Scale/Rotate/Transform/Perturbate each training image many times?
- Approaches:
  - Teach networks about certain symmetries (e.g. rotation)
  - Normalize/preprocess data to ensure well-known layout
  - Find encoding of the data that is invariant to certain operations
Figure 3. Schematic representation of the effect of the cyclic slice, roll and pool operations on the feature maps in a CNN. Arrows represent network layers. Each square represents a minibatch of feature maps. The letter ‘R’ is used to clearly distinguish orientations. Different colours are used to indicate that feature maps are qualitatively different, i.e. they are not rotations of each other. Feature maps in a column are stacked along the batch dimension in practice; feature maps in a row are stacked along the feature dimension.
NORMALIZING AND PRE-PROCESSING

DL trained on jet images vs. physically-motivated feature driven approaches

Figure 2: The average jet image for signal $W$ jets (top) and background QCD jets (bottom) before (left) and after (right) applying the rotation, re-pixelation, and inversion steps of the pre-processing. The average is taken over images of jets with $240 \text{ GeV} < p_T < 260 \text{ GeV}$ and $65 \text{ GeV} < \text{mass} < 95 \text{ GeV}$. 

Molecules are encoded as Vectors of Nuclear Charges and Inter-atomic Distance Matrices

=> Translation and rotation Invariant Representation
HOW DO I REPRESENT MY DATA IN NEURAL NETWORKS?
SIMPLE EXAMPLE: CLASSIFICATION

One-Hot Encoding

Training Data

Scalar Encoding

One-Hot Encoding

<table>
<thead>
<tr>
<th>Data</th>
<th>Scalar Encoding</th>
<th>One-Hot Encoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>[0.0]</td>
<td>[1,0,0,0,0,0,0,0,0,0]</td>
</tr>
<tr>
<td>1.0</td>
<td>[1.0]</td>
<td>[0,1,0,0,0,0,0,0,0,0]</td>
</tr>
<tr>
<td>2.0</td>
<td>[2.0]</td>
<td>[0,0,1,0,0,0,0,0,0,0]</td>
</tr>
<tr>
<td>3.0</td>
<td>[3.0]</td>
<td>[0,0,0,1,0,0,0,0,0,0]</td>
</tr>
<tr>
<td>4.0</td>
<td>[4.0]</td>
<td>[0,0,0,0,1,0,0,0,0,0]</td>
</tr>
<tr>
<td>5.0</td>
<td>[5.0]</td>
<td>[0,0,0,0,0,1,0,0,0,0]</td>
</tr>
<tr>
<td>6.0</td>
<td>[6.0]</td>
<td>[0,0,0,0,0,0,1,0,0,0]</td>
</tr>
<tr>
<td>7.0</td>
<td>[7.0]</td>
<td>[0,0,0,0,0,0,0,1,0,0]</td>
</tr>
<tr>
<td>8.0</td>
<td>[8.0]</td>
<td>[0,0,0,0,0,0,0,0,1,0]</td>
</tr>
<tr>
<td>9.0</td>
<td>[9.0]</td>
<td>[0,0,0,0,0,0,0,0,0,1]</td>
</tr>
</tbody>
</table>
IMAGE SEGMENTATION & BOUNDING BOXES
Creative use of feature channels

ORGANIZING SPEECH INTO FEATURE MAPS
Reducing Problems to Image Recognition

ENCODING TIME SERIES AS IMAGES
Gramian Angular Fields (GAF) and Markov Transition Fields (MTF)

Looking for Gravitational Waves

**Classifier:** Detect Presence of GWs

**Regression:** Parameter Estimation (i.e., masses of the two black holes)

**Figure 2.** Left panel: The blue curve is a sample of an input to our DNN algorithm. It contains a BBH GW signal (red) which was whitened with aLIGO’s PSD design sensitivity (see Figure 3) and superimposed in noisy data with SNR = 0.5.

Right panel: The corresponding spectrogram showing that the BBH GW signal on the left is not visible and thus cannot be detected by any algorithm trained for image recognition. Nevertheless, our DNN detects the presence of this signal from the time-series data, and reconstructs the source’s parameters with excellent accuracy.

HOW CAN I TRUST THE NETWORK?
“DEEP NEURAL NETS ARE BLACK BOXES”
... even if you can look at the internals...

• If a network performs well on the test data and appears to work reasonably well on real data...
  • Can we trust it?
  • Are there formal error bounds on the recognition accuracy?
  • E.g., would you trust a trained NN to operate your nuclear power plant?
• Field of active research (DARPA, MIT, Capital One, many others)
  • Debugging and Understanding NN behavior
  • Rationales for network decisions
ATTACKING NEURAL NETWORKS
Spoofing and Malicious Misclassification

1. Run input $x$ through the classifier model
2. Derive a perturbation tensor that maximizes chances of misclassification:
   1. Find blind spots in input space; or
   2. Linear perturbation in direction of neural network’s cost function gradient; or
   3. Select only input dimensions with high saliency*
3. Apply scaled effective perturbation ($\delta x$) to $x$
   1. Larger perturbation == higher probability for misclassification
   2. Smaller perturbation == less likely for human detection

* Saliency refers to the importance of individual input dimensions in the classification process.
LOOKING INSIDE NEURAL NETS
Debugging, Understanding, Verifying

- Inspecting the NN
  - Visualize activations, filters, generate input that maximizes activation of a neuron
  - Occlude parts of the input and check expectations
  - (e.g., http://cs231n.github.io/understanding-cnn/)
- Capture Model Confidence, Estimate Uncertainty
  - Place Gaussian Distribution over Weights => Bayesian Neural Networks
- How to gain scientific insight from a trained network?
CONCLUDING REMARKS
DEEP LEARNING — A NEW COMPUTING MODEL

“Software that writes software”
PIONEERS ADOPTING HPC FOR DEEP LEARNING

“Investments in computer systems — and I think the bleeding-edge of AI, and deep learning specifically, is shifting to HPC — can cut down the time to run an experiment from a week to a day and sometimes even faster.”

— Andrew Ng, Baidu

Dr. Andrew Ng, Chief Scientist, Baidu
ROOM FOR FUTURE WORK

The Four Questions Revisited

• Data Acquisition
  • How to get enough high-quality labeled data (or unlabeled learning or less need for input data); how much simulated data is okay without generating artefacts?

• Exploiting Properties of the Data (Symmetries, Invariants, ...)
  • To speed up learning, improve precision, guarantee properties

• Data Representations, especially for non-image data

• Trusting the Network
  • Formal verification, attack models, error bounds, cost of misclassification, gaining scientific insight

• Other topics: new use cases (signal processing), how to design networks, new layer types, generative models, debugging, optimizing
NVIDIA EXPERTISE AT EVERY STEP

Solution Architects
- 1:1 support
- Network training setup
- Network optimization

Deep Learning Institute
- Certified expert instructors
- Worldwide workshops
- Online courses

GTC Conferences
- Epicenter of industry leaders
- Onsite training
- Global reach

Global Network of Partners
- NVIDIA Partner Network
- OEMs
- Startups
# NVIDIA DEEP LEARNING PARTNERS

<table>
<thead>
<tr>
<th>Graph Analytics</th>
<th>DL Frameworks</th>
<th>Enterprise DL</th>
<th>Enterprises</th>
<th>Data Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>blazegraph</td>
<td>Caffe</td>
<td>bonsai</td>
<td>IBM, SAP</td>
<td>Hortonworks</td>
</tr>
<tr>
<td>graphistry</td>
<td>Chainer</td>
<td>deep instinct</td>
<td></td>
<td>splunk</td>
</tr>
<tr>
<td>kinesis</td>
<td>CNTK</td>
<td>SKYMIND</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAPD</td>
<td>TensorFlow</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>theano</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>torch</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>syner scope</td>
<td>accenture, deepsense.io</td>
<td></td>
</tr>
</tbody>
</table>
REFERENCES
