Machine Learning for Science

Kathy Yelick
Associate Laboratory Director for Computing Sciences
Machine Learning on Images and Videos

Self-driving cars interact with human drivers
AliBaba’s deep learning software beats humans at reading comprehension test
Machine Learning for Robotics

Robots
Artificial Intelligence, Machine Learning and Deep Learning

Artificial Intelligence

Machine Learning

Deep Learning

Statistics and mathematics, including optimization and linear algebra
Three ingredients for machine learning

Data

Algorithms

Machines

Thus, we change the channel image computation of an MS-D network with dilations and biases $b_{ij}$ and the weights $w_{ijkl}$ of Eq. $(3)$. A graphical representation of the result:

$$ y = g(c) = X_i l=0 D_{ij} z_i 1 + s_{ijkl} z_i 0, \ldots, z_i 1, z_i 1, y = g(c) = X_i l=0 D_{ij} z_i 1 + s_{ijkl} z_i 0, \ldots, z_i 1, $$

where $y$ is the output image, $s_{ijkl}$ is a channel-specific dilation, $D_{ij}$ is a convolution filter, and $g(c)$ is an application-specific activation function. Note that all feature maps are used for the final output computation.

By combining mixed-scale dilated convolutions with dense connections, we can define a DCNN architecture that uses dilated convolutions. A dilated convolution is defined as:

$$ y = g(c) = X_i l=0 D_{ij} z_i 1 + s_{ijkl} z_i 0, \ldots, z_i 1, $$

where $y$ is the output image, $s_{ijkl}$ is a channel-specific dilation, $D_{ij}$ is a convolution filter, and $g(c)$ is an application-specific activation function. Note that all feature maps are used for the final output computation.

By combining mixed-scale dilated convolutions and dense connections, we can define a DCNN architecture that uses dilated convolutions. A dilated convolution is defined as:

$$ y = g(c) = X_i l=0 D_{ij} z_i 1 + s_{ijkl} z_i 0, \ldots, z_i 1, $$

where $y$ is the output image, $s_{ijkl}$ is a channel-specific dilation, $D_{ij}$ is a convolution filter, and $g(c)$ is an application-specific activation function. Note that all feature maps are used for the final output computation.

Furthermore, information at a certain layer with a different dilation can be used to compute the layer output. For example, consider the equation:

$$ y = g(c) = X_i l=0 D_{ij} z_i 1 + s_{ijkl} z_i 0, \ldots, z_i 1, $$

where $y$ is the output image, $s_{ijkl}$ is a channel-specific dilation, $D_{ij}$ is a convolution filter, and $g(c)$ is an application-specific activation function. Note that all feature maps are used for the final output computation.

By combining mixed-scale dilated convolutions, dense connections, and enabling effective training with relatively small training sets, we can define a DCNN architecture that uses dilated convolutions. A dilated convolution is defined as:

$$ y = g(c) = X_i l=0 D_{ij} z_i 1 + s_{ijkl} z_i 0, \ldots, z_i 1, $$

where $y$ is the output image, $s_{ijkl}$ is a channel-specific dilation, $D_{ij}$ is a convolution filter, and $g(c)$ is an application-specific activation function. Note that all feature maps are used for the final output computation.

By combining mixed-scale dilated convolutions, dense connections, and enabling effective training with relatively small training sets, we can define a DCNN architecture that uses dilated convolutions. A dilated convolution is defined as:

$$ y = g(c) = X_i l=0 D_{ij} z_i 1 + s_{ijkl} z_i 0, \ldots, z_i 1, $$

where $y$ is the output image, $s_{ijkl}$ is a channel-specific dilation, $D_{ij}$ is a convolution filter, and $g(c)$ is an application-specific activation function. Note that all feature maps are used for the final output computation.
Google’s AlphGo Zero beats humans after self-training for 70 hours
The Data
Scientific data needs machine learning

Image / Video Processing

Text

Genomics

Signal Processing

Graph Analytics (Relationships)

Simulation Analytics
Berkeley Lab advances detector technology for many fields of science, including (above CryoEM) biology, cosmology, material science, physics, and more.
Salmonella causes 1 million foodborne illnesses in the US, with 19,000 hospitalizations and 380 deaths.

Reconstructed Bacteriophage P22 which can attack Salmonella in foods
Embedded sensors in infrastructure

Fiber-optic cables can be used as sensors for urban seismic hazard analysis, to monitor soil layer changes, detect nuclear explosions, and do global seismic imaging.
Embedded sensors in infrastructure

Data analysis of power grid micro sensors trained to identify intrusion and other events
Preliminary CAM5 hi-resolution simulations (0.25°, prescribed aerosols)

Michael Wehner, Prabhat, Chris Algieri, Fuyu Li, Bill Collins
Lawrence Berkeley National Laboratory

Kevin Reed, University of Michigan

Andrew Gettelman, Julio Bacmeister, Richard Neale
National Center for Atmospheric Research

June 1, 2011
The Algorithms
Finding Storms in Simulations

Ground Truth Prediction

Climate Science Tasks

Classification

Classification + Localization

Object Detection

Instance Segmentation
Learn the relationship between features with Graphical Model Estimator
New Algorithm for HPC discovering regions and co-regions

Automatic parcellation from fMRI data alone

Baseline parcellation from Glasser [Glasser et al. 2016]

listening

controlling the eyes

receiving info from the senses

First of kind analysis at this scale using new algorithm and high performance computing at LBNL

Wearable MRI sensors + HPC Analytics

Goals:
1) reduce time in MRI
2) improve patient experience
3) better quality of images

Wearable MIR sensors [Arias UCB]

Many Types of MIR Scans
- Cartesian
- Spiral
- 2D Radial
- Stochastic
- 3D Radial
- Stack of Spirals

Compressed sensing algorithms [Lustig, UCB]
Real-Time Analytics in Health

Compressed Sensing Approach by Mike Lustig et al
MRI results Wenwen Jiang

3 min goal (1 sec/iteration)
Michael Driscoll HPC optimization
Understanding energy and human behavior

How well do various algorithms work?
What are the particular challenges in science?

**Scale**
- Data rates from detectors
- Machine scale, novelty and performance

**Complexity**
- Adaptive, hierarchical
- Multi-modal, noisy

**Interpretability**
- Explainable, understandable, robust
- Physically realizable
The Machines
State-of-the art computing for the broad DOE science community – over 7000 users, 700 applications
Deep Learning at 15 Petaflops

Berkeley Lab scientists new parallel algorithm for deep learning on climate and particle physics data at 15 Petaflops

= 4 x
Microbiome analysis uses high performance computing
Microbiome analysis uses high performance computing
Microbiome analysis with machine learning

Similarities between genes (proteins) Clusters of related ones

New science, impossible without HPC

Graph showing time (minutes) versus number of nodes (24 cores/node) for Eukarya with #V=3M, #E=360M, comparing HipMCL and MCL methods.
The Berkeley National Lab Advantage in ML for Science

HPC and networking
Systems and Expertise

Applied Math
Driven by Science

Team Science
End-to-end Solutions

Advanced Detectors
Data source
AI is also revolutionizing science revolution