NASA FDL OVERVIEW - DR LIKA GUHATHAKURTA (on behalf of the FDL team.)
What is NASA doing with Big Data today?

October 04, 2012 by Nick Skytland

In the time it took you to read this sentence, NASA gathered approximately 1.73 gigabytes of data from our nearly 100 currently active missions! We do this every hour, every day, every year – and the collection rate is growing exponentially. Handling, storing, and managing this data is a massive challenge. Our data is one of our most valuable assets, and its strategic importance in our research and science is huge. We are committed to making our data as accessible as possible, both for the benefit of our work and for the betterment of humankind through the innovation and creativity of the over seven billion other people on this planet who don’t work at NASA.
“An AI R&D accelerator that tackles knowledge gaps useful to the space program. The program is an intense 8-week concentrated study on topics not only important to NASA, but also to humanity’s future.”
SETI enables the public / private partnership
FDL IS INTERDISCIPLINARY:
50% DATA SCIENCE / 50% SPACE SCIENCES
3 projects in 2016
6 projects in 2017
PLANETARY DEFENSE

ADJACENT BUT RELATED PROBLEM DOMAINS

ALLOWS USEFUL OVERLAP OF EXPERTISE AND TALENT
FDL also benefits from vast GPU compute from the private sector.
And expertise.
IBM’s Executive Project Manager briefs the FDL team on the compute resource available for each team.
Google’s Francois Chollet - inventor of the Keras.io framework briefs the FDL team. (Python for machine learning.)
Current operational flare forecasting relies on human morphological analysis of active regions and the persistence of solar flare activity.

The FDL team performed analyses of solar magnetic complexity and deployed convolutional neural networks to connect solar UV images taken by SDO/AIA into forecasts of maximum x-ray emissions.

The technique has the potential to improve both the reliability and accuracy of solar flare predictions.
Interdisciplinary Collaboration

**Heliophysicist’s view of ML**

**Data scientist's view of HP**

*This is your machine learning system?*

Yup! You pour the data into this big pile of linear algebra, then collect the answers on the other side.

What if the answers are wrong?

Just stir the pile until they start looking right.

*The Sun's atmosphere is a superhot plasma governed by magnetohydrodynamic forces...*

Ah, yes, of course.

Whenever I hear the word "magnetohydrodynamic" my brain just replaces it with "magic."
SPACE WEATHER: SOLAR STORM PREDICTION

Types of Space Weather

**FLARES**
- Electromagnetic Radiation

**ENERGETIC PARTICLES**
- Particle Radiation

**MASS EJECTIONS**
- Massive Magnetic Ropes
SPACE WEATHER: SOLAR STORM PREDICTION

Types of Space Weather

FLARES
Disruption of Communications

ENERGETIC PARTICLES
Satellite Damage

MASS EJECTIONS
Power grid Disruption
Why Solar Flare prediction is important?

- **FLARES**
  - Speed of Light
  - No warning

- **ENERGETIC PARTICLES**
  - Relativistic speeds
  - 20 minute warning

- **MASS EJECTIONS**
  - 20 hour warning
How is a flare defined?

Using X-ray flux as measured by the GOES satellite
Deep learning has revolutionized the way we do image classification.
1. **Dataset Preparation:** *Take advantage of big data*
2. **Software:** *Build scientific process*
3. **Prediction:** *Enable Flare Forecasting*
4. **Science:** *Visualize Results*
   - Discover Flare Precursors
   - Providing new physical insight
   - New Physics?
Can we use deep learning to connect AIA images with flare strength?
Deep Learning: Convolutional Networks

Several convolutional layers allow the neural network to recognize features of increased complexity.
FlareNet

Total parameters: 518,667
Our first goal was to see if the neural network could connect AIA images with flare X-ray amplitude.

The concern is whether the neural network is simply memorizing the images.
Our current neural network seems to be able to generalize for weak flares (C-class), but not yet for stronger flares. Only flares observed prior to 2015 used for training.
Our current biggest challenge is class imbalance!
What does a convolutional neural network pay attention to?

Analysis Scripts: Saliency

FlareNet is paying attention to the relative location of structures in different channels.
FlareNet’s filter activations

Several convolutional layers allow the neural network to recognize features of increased complexity.
SPACE WEATHER: SOLAR STORM PREDICTION

FlareNet’s filter activations

Block 1
Filter 7

Block 2
Filter 8

Block 3
Filter 7
FlareNet’s filter activations

FlareNet learned the importance of active regions
Achievements

- Developed a framework to apply CNNs to heliophysics problems.
- Developed a CNN visualization framework to mine trained networks for physical insight.
- Demonstrated the capability of CNNs to identify structures of flaring relevance.
Future Work

1. Expand our data enhancement capabilities.

1. Explore the possibility of adding other instruments to increase our flare pool (Stereo, SOHO, GOES.)

1. Try alternative problem definitions besides regression (distribution, classification.).
The vast amounts of data collected by satellites and observatories operated by government agencies such as NASA, NOAA and the US Geological Survey remains a largely untapped resource for discovering how the Sun interacts with Earth.

The FDL team built a knowledge discovery module named STING (Solar Terrestrial Interactions Neural Network Generator) on top of industry-standard, open source machine learning frameworks to allow researchers to further explore these complex datasets.

STING showed the ability to accurately predict the variability of Earth’s geomagnetic fields in response to solar driving - specifically the KP index.

In the process the tool discovered the imprint of the magnetospheric ring current in precursors of geomagnetic storms - an example of an AI derived discovery.
SPACE WEATHER: SOLAR TERRESTRIAL INTERACTIONS

DATA SOURCES
Planetary Kp Index

(Bartels, 1938)

Kp INDEX

Use this table on the right to convert the difference in the maximum and minimum x-values for today to a K index. The larger the K index, the stormier it is in Earth’s magnetic field.

<table>
<thead>
<tr>
<th>K index</th>
<th>nT diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0-5</td>
</tr>
<tr>
<td>1</td>
<td>5-10</td>
</tr>
<tr>
<td>2</td>
<td>10-20</td>
</tr>
<tr>
<td>3</td>
<td>20-40</td>
</tr>
<tr>
<td>4</td>
<td>40-70</td>
</tr>
<tr>
<td>5</td>
<td>70-120</td>
</tr>
<tr>
<td>6</td>
<td>120-200</td>
</tr>
<tr>
<td>7</td>
<td>200-330</td>
</tr>
<tr>
<td>8</td>
<td>330-500</td>
</tr>
<tr>
<td>9</td>
<td>&gt;500</td>
</tr>
</tbody>
</table>

Petersburg, AK magnetometer data with a 75 nT change in the X-direction (Magnetic North)
SPACE WEATHER: SOLAR TERRESTRIAL INTERACTIONS

GRADIENT BOOSTING RESULTS

Prediction of $K_p$ 3 hours ahead
Most important: Current Kp index

Other important predictors:
- Solar wind magnetic field strength and Bz,
- Solar wind speed and proton density,
- **Unexpected Result:** N-S component of the geomagnetic field at low latitude stations (Guam, Hawaii, Puerto Rico). This points to the importance of the magnetospheric ring current.

Machine learning extracted important physical parameters without a priori knowledge of the system.
• The FDL team tackled the task of automating the task of creating 3D shape models of NEOs from sparse radar data.

• The process currently takes up to four weeks of manual interventions by experts using established software.

• The team demonstrated a pipeline for automation that allows NEOs to be modelled in several hours.

• This result will hopefully support researchers render 3D models of the current backlog of radar imaged asteroids.
We are observing NEAs faster than we can analyze them!

https://echo.jpl.nasa.gov/~lance/Radar_detected_neas.html
PLANETARY DEFENSE: RADAR 3D SHAPE MODELING

SHAPE MODELING PIPELINE

- **2/3 days** in hours
- **50 - 100 calls** in days
- **1 week** and **20 parameters to fit** in weeks
- **1 to 2 months** and **> 700 parameters to fit** in weeks

- **Raw radar data** → **Preprocessing** → **Simplified shape** → **Spin state search** → **Initial triangular mesh** → **Vertex fit** → **Refined triangular mesh** → **Final result**

- **0.16 (16.03)**
- **0.43 (19.03)**
- **0.75 (18.01)**
PLANETARY DEFENSE: RADAR 3D SHAPE MODELING

VARIATIONAL AUTOENCODER

Generative model

Sample

Condition on delay-Doppler images

Kingma and Welling, 2013
OUR SOLUTIONS

a) Pre-processing is faster
b) Spin axis determination is faster
c) Training data generation is improved
d) Neural network is improved
Meteor showers caused by the previous-return ejecta of long period comets can guide deep searches, and improve warning time, for potentially hazardous long period comets that passed near Earth’s orbit in the past ten millennia.

The FDL team showed how the data reduction of the ‘CAMS’ meteor shower survey program could be successfully automated by using deep learning approaches.

By using dimensionality reduction (t-SNEs) the team were able to identify yet uncatalogued meteor shower clusters - a promising direction for further investigation.
PLANETARY DEFENSE: LONG-PERIOD COMETS

Meteor shower surveys
Convolutional Neural Network (CNN)

Results: Precision: 88.6% Recall: 90.3%
PLANTARY DEFENSE: LONG-PERIOD COMETS

Mapping meteors in the sky
PLANETARY DEFENSE: LONG-PERIOD COMETS

Established meteor showers
Maps that detail the regions of interest in the dark polar regions are plagued by artefacts and shadow variability that severely hamper the planning of future prospecting missions.

A large dataset was compiled for the south polar region and high-level feature extraction was performed. Results showed an impressive speed-up of 100x compared to human experts, with more than 98.4% agreement when approaching a crater labelling.

This work represents a potential keystone to facilitate accessing water on the Lunar surface and future traverse planning.
Building lunar maps at the poles is problematic.

Lunar Orbiter Laser Altimeter
Digital Elevation Model (LOLA DEM), 20 m resolution

Narrow Angle Camera (NAC)
Optical images, 0.5 m resolution
PLANETARY DEFENSE: RADAR 3D SHAPE MODELING

Improving Maps Conventionally.
Most algorithms remain unused.
### Timing Comparison of FDL Technique

<table>
<thead>
<tr>
<th>Group</th>
<th>Human</th>
<th>Single-Layer</th>
<th>CNN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
<td>-</td>
<td>Poor</td>
<td>98.4%</td>
</tr>
<tr>
<td><strong>Time (1000 Images)</strong></td>
<td>1-3 hours</td>
<td>10 hours</td>
<td>1 minute</td>
</tr>
<tr>
<td><strong>Person-hours</strong></td>
<td>1-3 hours</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
AI & SPACE SCIENCES
Closing Thoughts

• Focus on applied AI solutions using mainstream deep learning tools, thereby complementing and informing the research into novel AI technology being undertaken by other NASA teams.

• Strong incentive for the private sector to participate due to commercial opportunities that are implicit in the outcome;

• Clear risk/cost reduction benefit to manned activities beyond LEO, and for cis-lunar operations in particular;

• Problem definitions for which relevant data has already been collected and is available for use under an open license.
By way of example, consider the application of AI to Space Weather

- Solar flares and associated proton storms pose a significant risk to astronauts beyond LEO, and offer little or no warning. The Apollo “near miss” of the August 1972 solar flare provides a dramatic example of this concern.

- Multiple industry sectors have a vested commercial interest in seeing improvements to solar flare predictions and better heliophysics modeling in general. Examples include the power utilities, insurance companies, communications and satellite operators, and the military.

- There are hundreds terabytes of well structured heliophysics data highly suited to deep learning applications, including the archives from SDO/AIA, ACE, and SOHO.

- The image-centric nature of solar data (e.g. SDO – HMI and AIA) makes it easy to leverage the rapid advances in image analysis that the AI community has contributed into open source.

- There are tantalizing indications that machine learning techniques can offer better predicative capabilities over physics-based models, which leads many to believe that the use of neural net deep learning will prove to be even more effective.