Assessing Uncertainty in Deep Learning Techniques that Identify Atmospheric Rivers in Climate Simulations

Ankur Mahesh¹,²,³, Travis O’Brien¹,⁴, Mayur Mudigonda¹,², Karthik Kashinath¹, Sookyung Kim⁵, Samira Kahou⁷, Vincent Michalski⁶, Dean Williams⁵, Yunjie Liu¹, Prabhat¹,², Burlen Loring¹, William D. Collins¹,²

¹Lawrence Berkeley National Lab
²University of California, Berkeley
³Undergraduate, Department of Electrical Engineering and Computer Science
⁴University of California, Davis
⁵Lawrence Livermore National Laboratory
⁶Montreal Institute for Learning Algorithms
⁷Microsoft Research
Why Deep Learning for Climate Science?

- **Atmospheric Rivers (ARs):**
  "long, narrow, and transient corridors of strong horizontal water vapor transport…" *(AMS Glossary)*

- **No community-accepted standard** for identifying atmospheric rivers
  "You know one when you see one" – *ARTMIP 2018 Participant*

- **Poses a unique machine learning problem:** uncertainty with ground truth "label" data

An AR off the coast of California. Source: The Guardian.
Why Deep Learning for Climate Science?

<table>
<thead>
<tr>
<th></th>
<th>Logistic Regression</th>
<th>K-Nearest Neighbor</th>
<th>Support Vector Machine</th>
<th>Random Forest</th>
<th>ConvNet (a type of deep learning model)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atmospheric Rivers</td>
<td>Train: 81.97, Test: 82.65</td>
<td>Train: 79.7, Test: 81.7</td>
<td>Train: 81.6, Test: 83.0</td>
<td>Train: 87.9, Test: 88.4</td>
<td>Train: 90.5, Test: 90.0</td>
</tr>
<tr>
<td>Weather Fronts</td>
<td>Train: 84.9, Test: 89.8</td>
<td>Train: 72.46, Test: 76.45</td>
<td>Train: 84.35, Test: 90.2</td>
<td>Train: 80.97, Test: 87.5</td>
<td>Train: 88.7, Test: 89.4</td>
</tr>
</tbody>
</table>
Convolutional Neural Networks

• **Feature Learning**: a filter slides, or convolves, over the image and extracts features

• **Classification**: probabilistically map the features to the likelihood that an image belongs to a class
Transfer Learning

• **Transfer learning**: take a model trained to solve one problem and use it to solve a different problem

• When trained with millions of images, neural networks are generic feature extractors

• In transfer learning, neural networks use one dataset to train the feature learning part of the model

• Using this feature learning strategy, neural networks classify images in another dataset

• *Reduces the need for large labelled training datasets in climate science*
Architecture Uncertainty

- Tested several architectures with 1, 2, 3, and 16 layers for classifying images of ARs (16-layer model = VGGNet)

- How a neural network is trained: minimization of a loss function that quantifies model performance

- 16 layer architecture used transfer learning, which led to higher accuracy and more rapid convergence

- Uncertainty: which type of architecture yields best results?
Transfer Learning for Classifying ARs

• 16-layer model pre-trained on ImageNet: **92%** accuracy
  • ImageNet: dataset with millions of ordinary images (i.e. dogs, cats, benches, etc.)

An image of an atmospheric river, correctly classified by the model.
A Pre-Trained Model for Classifying ARs

• 16 Layer Model Pre-Trained on ImageNet: **92%** accuracy
  • ImageNet: dataset with millions of ordinary images (i.e. dogs, cats, benches, etc.)

An image of an atmospheric river, correctly classified by the model.

Set some pixels to 0 and record if the model classifies the image as an AR.
A Pre-Trained Model for Classifying ARs

• 16 Layer Model Pre-Trained on ImageNet: 92% accuracy
  • ImageNet: dataset with millions of ordinary images (i.e. dogs, cats, benches, etc.)

An image of an atmospheric river, correctly classified by the model.

Set some pixels to 0 and record if the model classifies the image as an AR.
A Pre-Trained Model for Classifying ARs

- 16 Layer Model Pre-Trained on ImageNet: **92% accuracy**
  - ImageNet: dataset with millions of ordinary images (i.e. dogs, cats, benches, etc.)

An image of an atmospheric river, correctly classified by the model.

Set some pixels to 0 and record if the model classifies the image as an AR.
A Pre-Trained Model for Classifying ARs

- 16 Layer Model Pre-Trained on ImageNet: **92%** accuracy
  - ImageNet: dataset with millions of ordinary images (i.e. dogs, cats, benches, etc.)

An image of an atmospheric river, correctly classified by the model.

Set some pixels to 0 and record if the model classifies the image as an AR.
A Pre-Trained Model for Classifying ARs

- 16 Layer Model Pre-Trained on ImageNet: **92%** accuracy
  - ImageNet: dataset with millions of ordinary images (i.e. dogs, cats, benches, etc.)

An image of an atmospheric river, correctly classified by the model.

Set some pixels to 0 and record if the model classifies the image as an AR.
A Pre-Trained Model for Classifying ARs

- 16 Layer Model Pre-Trained on ImageNet: 92% accuracy
  - ImageNet: dataset with millions of ordinary images (i.e. dogs, cats, benches, etc.)

An image of an atmospheric river, correctly classified by the model.

Set some pixels to 0 and record if the model classifies the image as an AR.
A Pre-Trained Model for Classifying ARs

- 16 Layer Model Pre-Trained on ImageNet: **92%** accuracy
  - ImageNet: dataset with millions of ordinary images (i.e. dogs, cats, benches, etc.)

An image of an atmospheric river, correctly classified by the model.

Set some pixels to 0 and record if the model classifies the image as an AR.

Heat Map: when the bottom left portion of this image is set to 0, the model does not think the image is an AR (**RED**).
If the top left or bottom right portion is set to 0, then the model still thinks the image is an AR (**GREEN**).
A Pre-Trained Model for Classifying ARs

- 16 Layer Model Pre-Trained on ImageNet: **92%** accuracy
  - ImageNet: dataset with millions of ordinary images (i.e. dogs, cats, benches, etc.)
- Conclusion: the model identified the features that make this image an AR!

**Heat Map**: when the bottom left portion of this image is set to 0, the model does not think the image is an AR (**RED**)
If the top left or bottom right portion is set to 0, then the model still thinks the image is an AR (**GREEN**)

An image of an atmospheric river, correctly classified by the model.
Set some pixels to 0 and record if the model classifies the image as an AR.
Classification vs. Segmentation

• **Classification:** classify each image as a member of a class

• **Semantic Segmentation:** classify each pixel as a member of class

• Semantic segmentation does not distinguish between multiple instances of the same class

*Top:* this is a picture of a car  
*Bottom:* this is a picture of a crowd

People, bicycles, sidewalk, signposts, roads, and cars are all recognized

Label Uncertainty

- **ARs**: Isolate areas 1500 km long with 95th percentile *Integrated Vapor Transport*

- **Tropical Cyclones**: Use the Toolkit for Extreme Climate Analysis (TECA) to generate labels

- There is uncertainty with these labels, which rely on arbitrary thresholds
Segmentation Model Results

- Model has successfully learned the structure of ARs and TCs
- Segmentations are smoother than current “ground-truth” labelling methodologies
  - Model predictions remove reliance on arbitrary thresholds by finding patterns from thousands of training images
- The model can detect TCs and ARs, despite their close proximity

Segmentation of ARs (RED) and TCs (BLUE) in an IWV image
Segmentation Results: Metric Uncertainty

- Overall accuracy: 92%
- The “ground truth” labels were generated using much more information than the model was provided
  - Ground-truth-labelling input: integrated vapor transport, geopotential height, wind velocity, and sea surface temperature
  - Model input: integrated water vapor
- Metric Uncertainty: how do we evaluate the model when ground truth is imperfect?
Future Work

• Investigate how to represent architecture, label, and metric uncertainty

• Ensemble-based extreme event detection
  • Use different labelling strategies to generate multiple ground truth datasets
  • Train a neural network on each ground truth dataset
  • Have each network vote on whether or not an image is a particular type of extreme

• Test neural networks with an expert-hand-labelled dataset

• Use neural networks to detect other classes of extremes
Explicit Uncertainty in Training CNNs

Example Training Data:
Average AR Mask from ARTMIP algorithms.

Possible approach:

• Modify *loss function* used in training CNNs
• Explicitly account for uncertainty in training data
• Applicable to expert-labeled datasets w/ input from multiple experts
Thank You

This work was supported in part by the U.S. Department of Energy, (DOE) Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internship (SULI) program, and the DOE Regional and Global Climate Modeling Program as part of the Calibrated And Systematic Characterization, Attribution, and Detection of Extremes Scientific Focus Area.

For more information, contact:

Ankur Mahesh  
amahesh@lbl.gov

Travis O’Brien  
TAOBrien@lbl.gov