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

Ankur Mahesh[1,2,3], **Travis O'Brien**[1,4], Mayur Mudigonda[1,2], Karthik Kashinath[1], Sookyung Kim[5], Samira Kahou[7], Vincent Michalski[6], Dean Williams[5], Yunjie Liu[1], Prabhat[1,2], Burlen Loring[1], William D. Collins[1,2]

[1]Lawrence Berkeley National Lab

[2]University of California, Berkeley

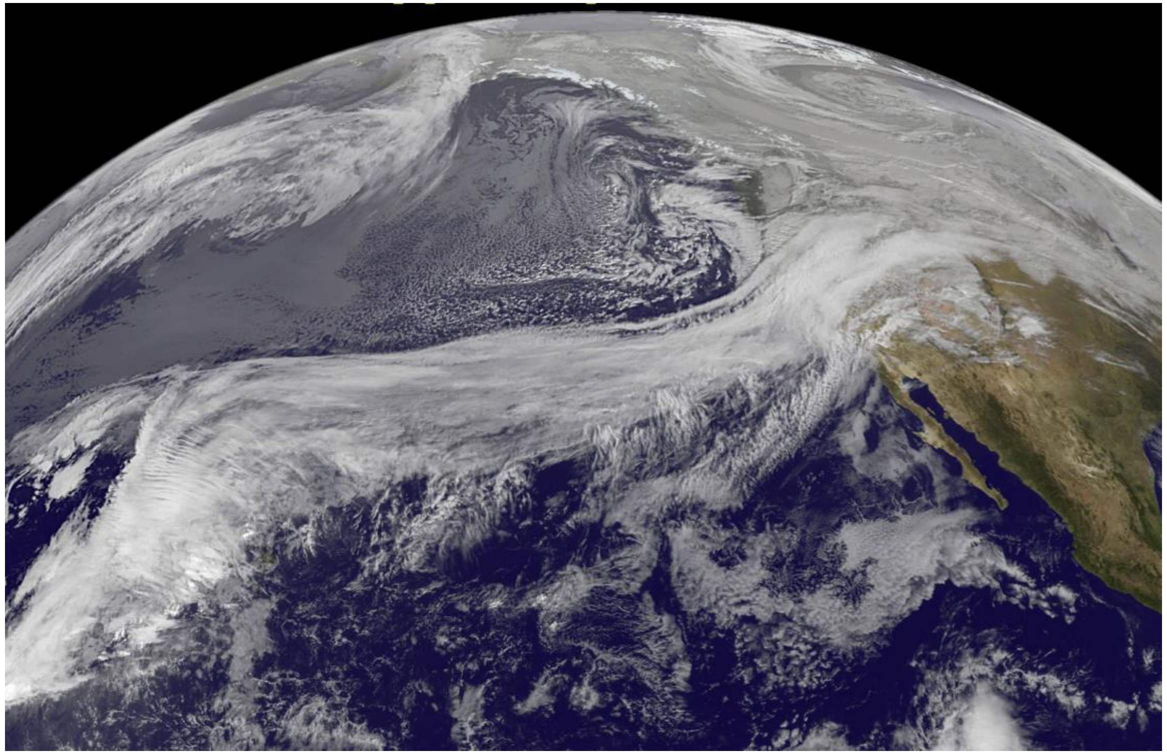
[3]Undergraduate, Department of Electrical Engineering and Computer Science

[4]University of California, Davis

[5]Lawrence Livermore National Laboratory [6]Montreal Institute for Learning Algorithms [7]Microsoft 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



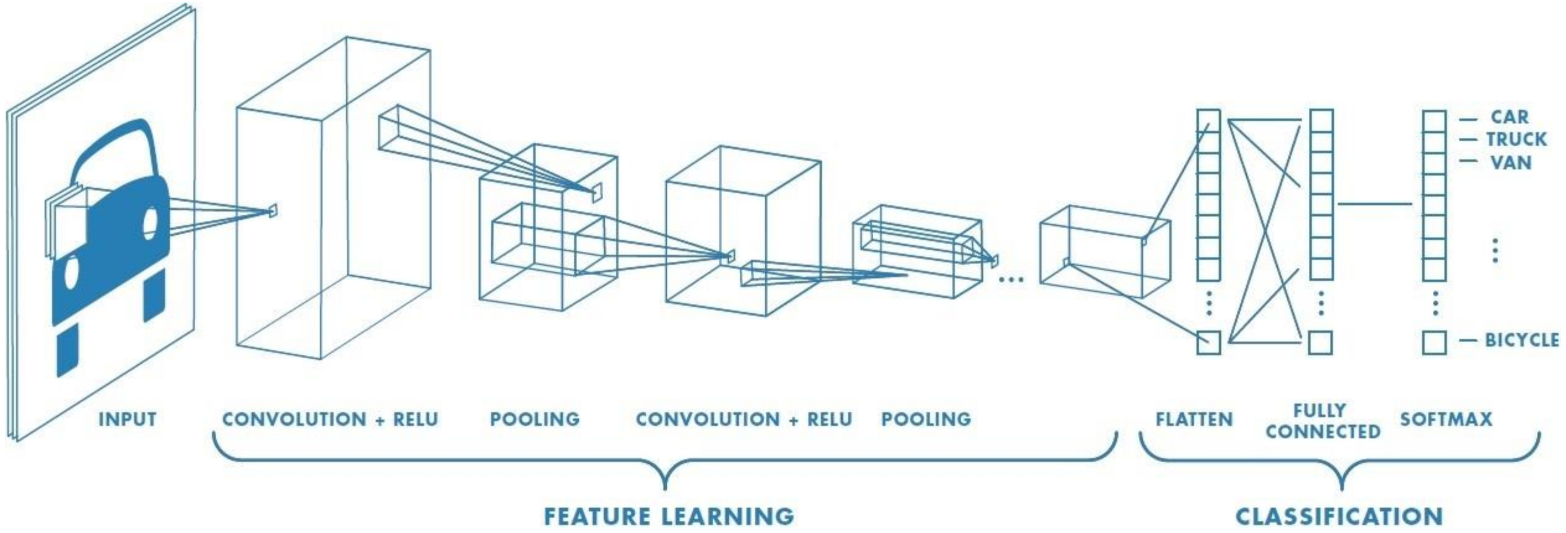
An AR off the coast of California. Source: The Guardian.

Why Deep Learning for Climate Science?

	Logistic Regression		K-Nearest Neighbor		Support Vector Machine		Random Forest		ConvNet (a type of deep learning model)	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Tropical Cyclone	96.8	95.85	98.1	97.85	97.0	95.85	99.2	**99.4**	99.3	**99.1**
Atmospheric Rivers	81.97	82.65	79.7	81.7	81.6	83.0	87.9	88.4	90.5	**90.0**
Weather Fronts	84.9	89.8	72.46	76.45	84.35	90.2	80.97	87.5	88.7	**89.4**

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



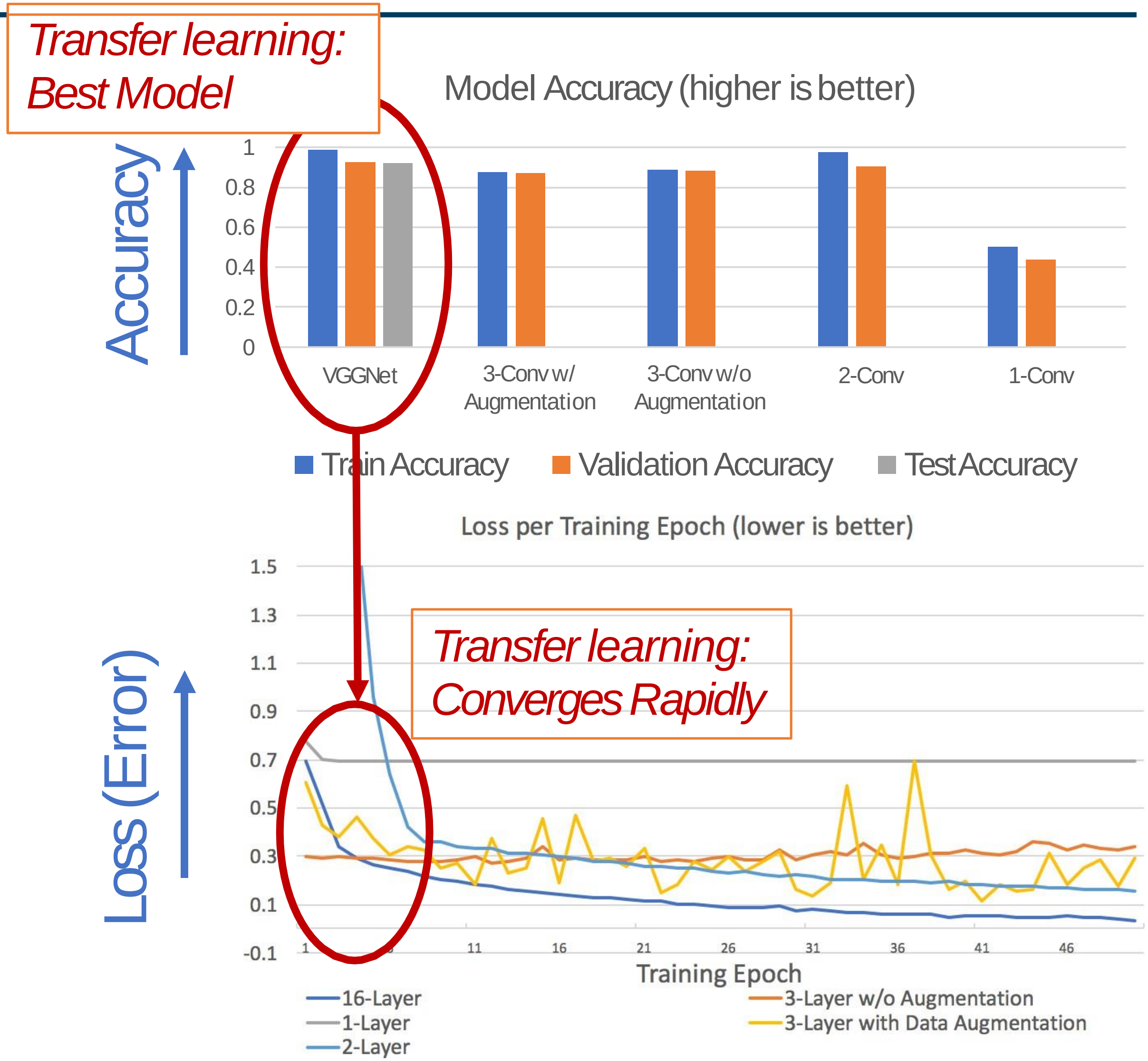
Source: MathWorks

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*

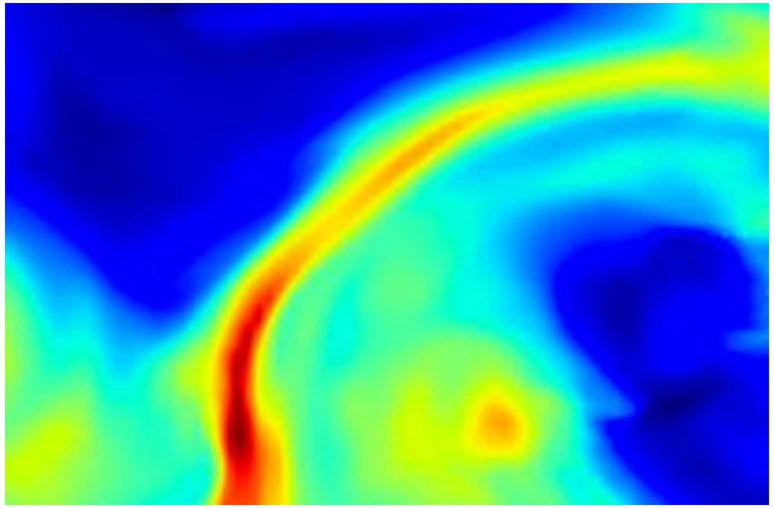
Architecture Uncertainty

- 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

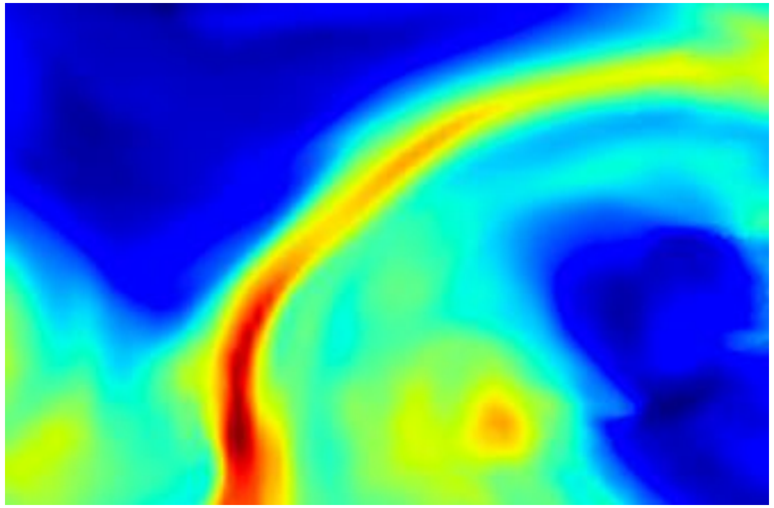
- 16-layer model pre-trained on ImageNet: **92%** accuracy
  - ImageNet: dataset with millions of ordinary images (i.e. dogs, cats, benches, etc.)



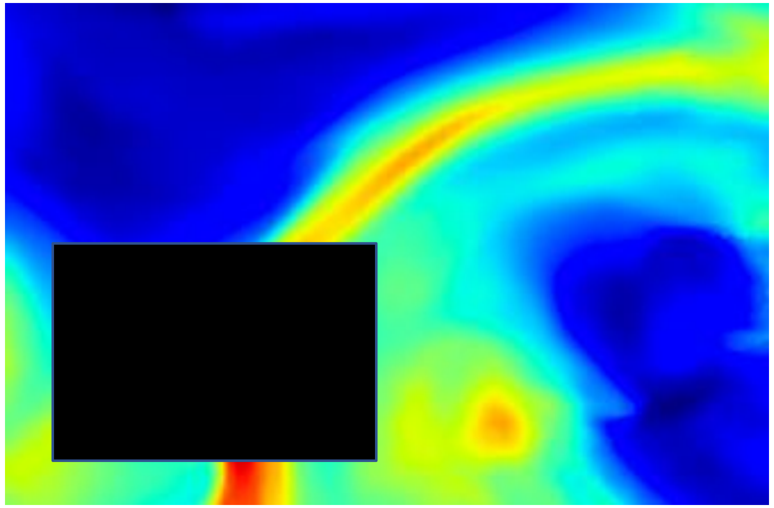
An image of an atmospheric river, correctly classified by the model.

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.)



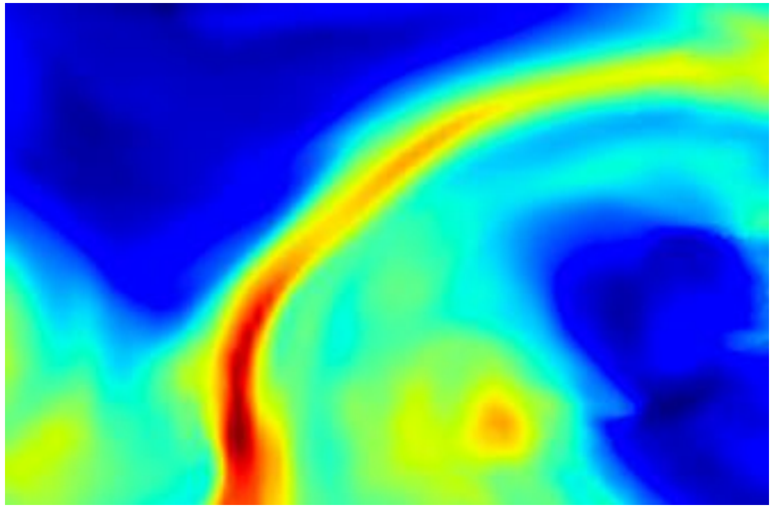
An image of an atmospheric river, correctly classified by the model.



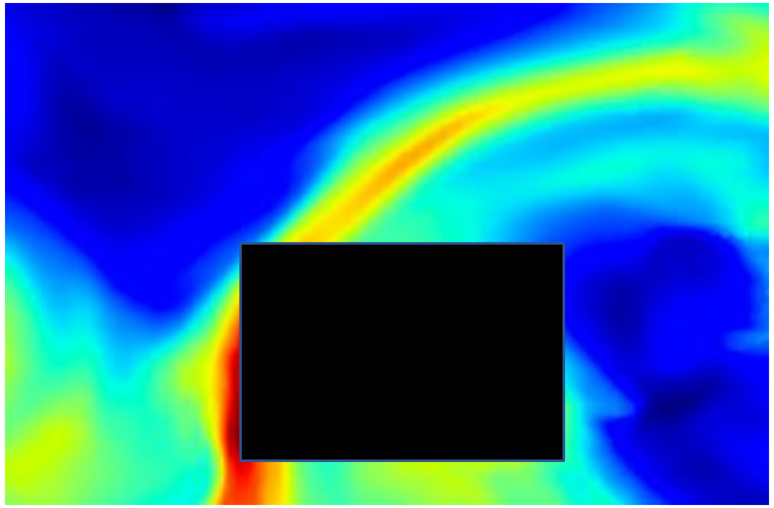
Set some pixels to 0 and record if the model classifies the image as an AR

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.)



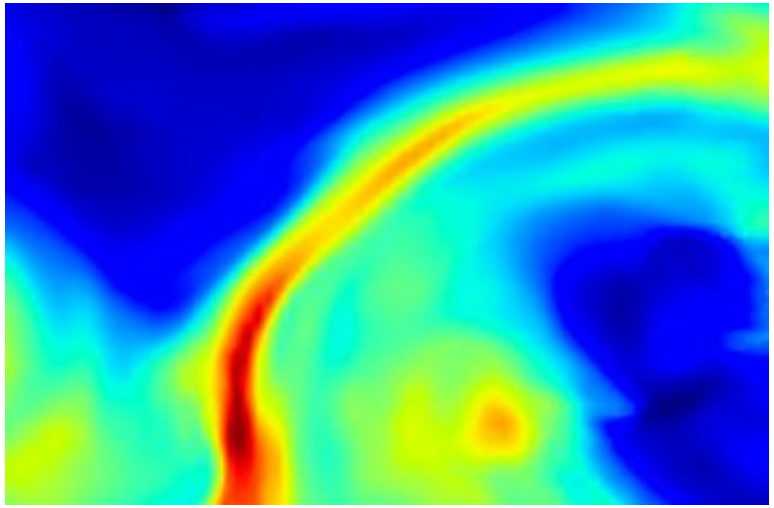
An image of an atmospheric river, correctly classified by the model.



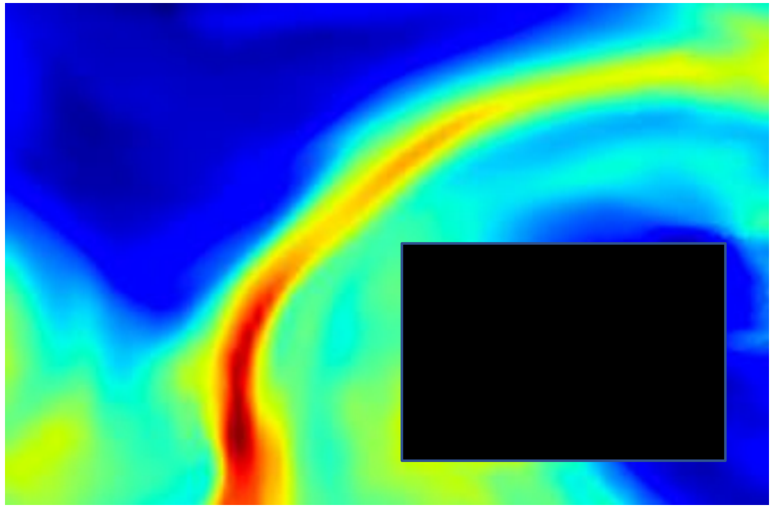
Set some pixels to 0 and record if the model classifies the image as an AR

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.)



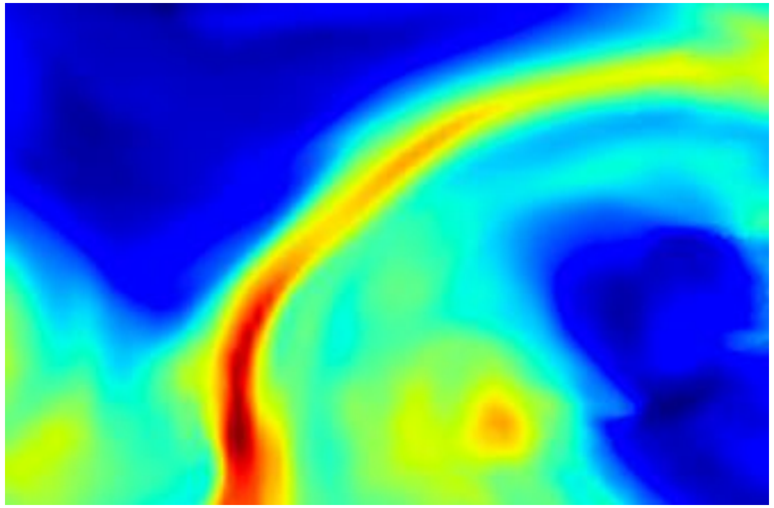
An image of an atmospheric river, correctly classified by the model.



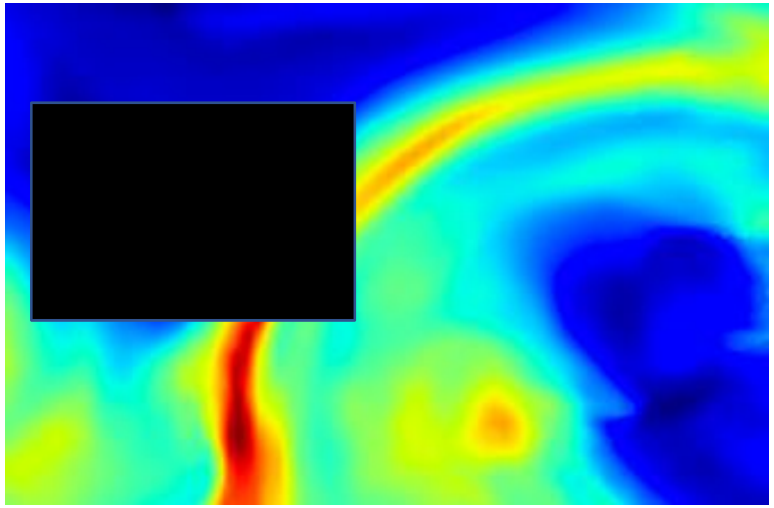
Set some pixels to 0 and record if the model classifies the image as an AR

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.)



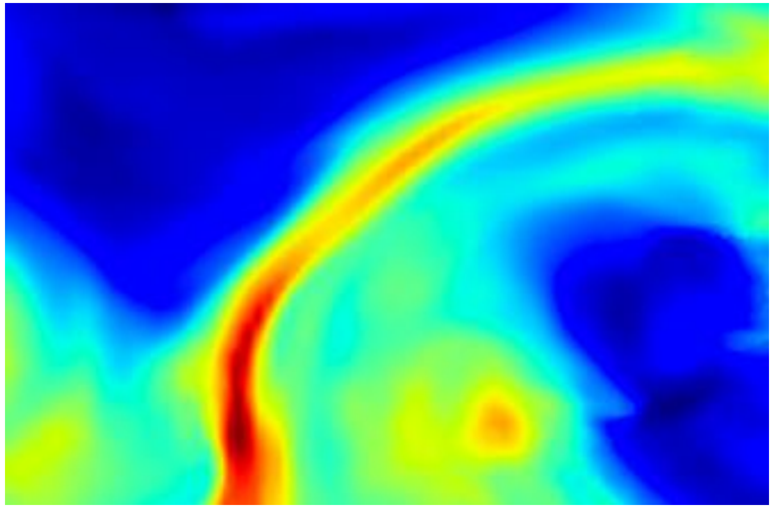
An image of an atmospheric river, correctly classified by the model.



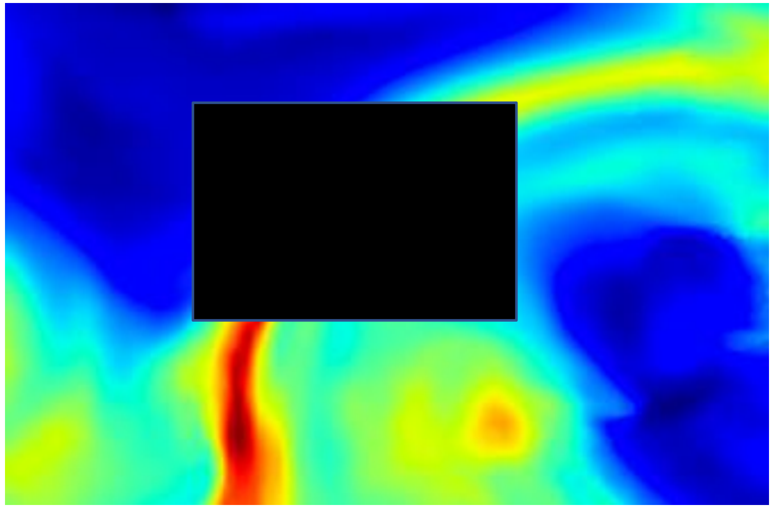
Set some pixels to 0 and record if the model classifies the image as an AR

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.)



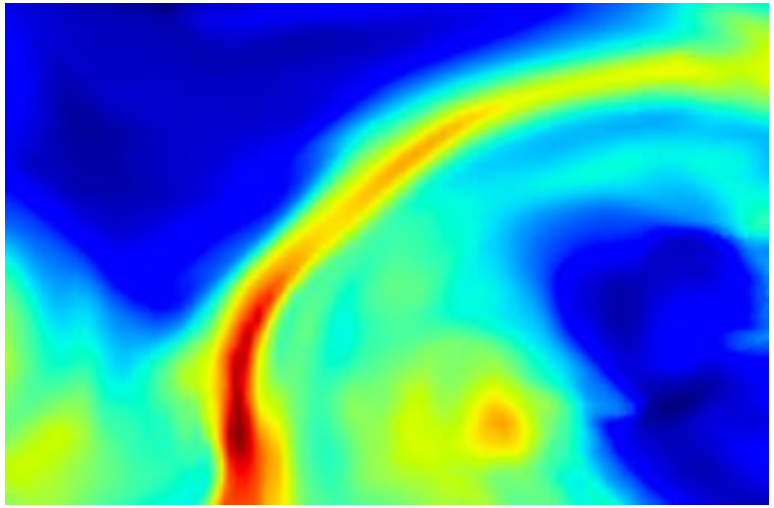
An image of an atmospheric river, correctly classified by the model.



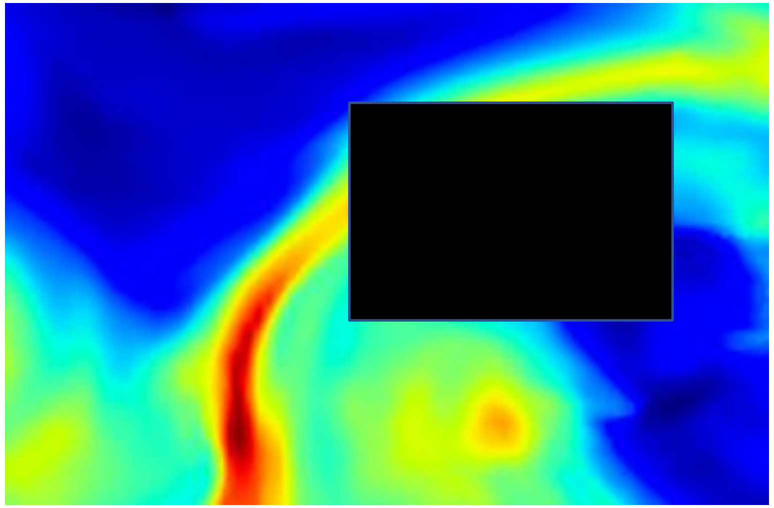
Set some pixels to 0 and record if the model classifies the image as an AR

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.)



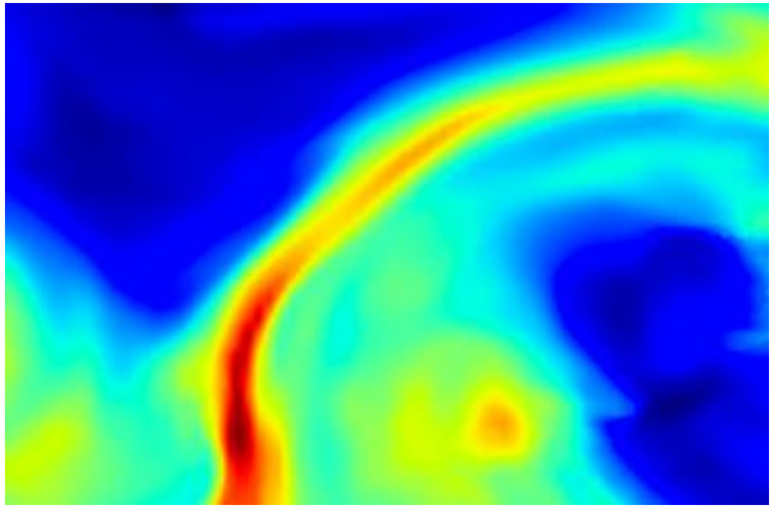
An image of an atmospheric river, correctly classified by the model.



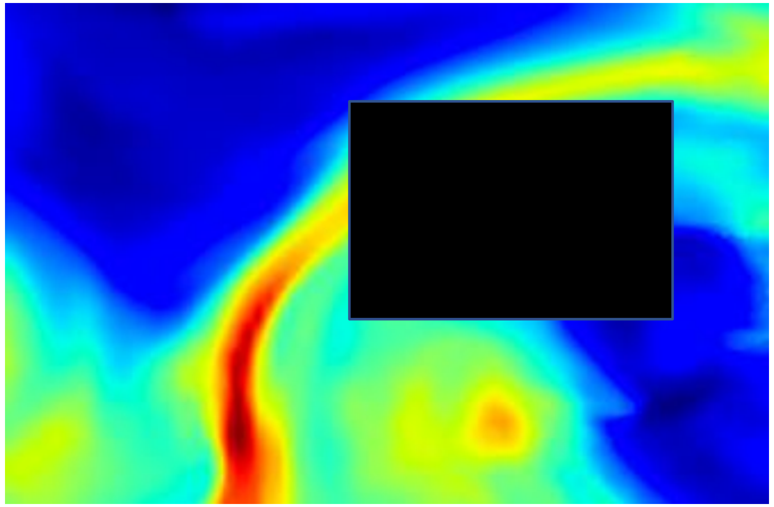
Set some pixels to 0 and record if the model classifies the image as an AR

A Pre-Trained Model for Classifying ARs

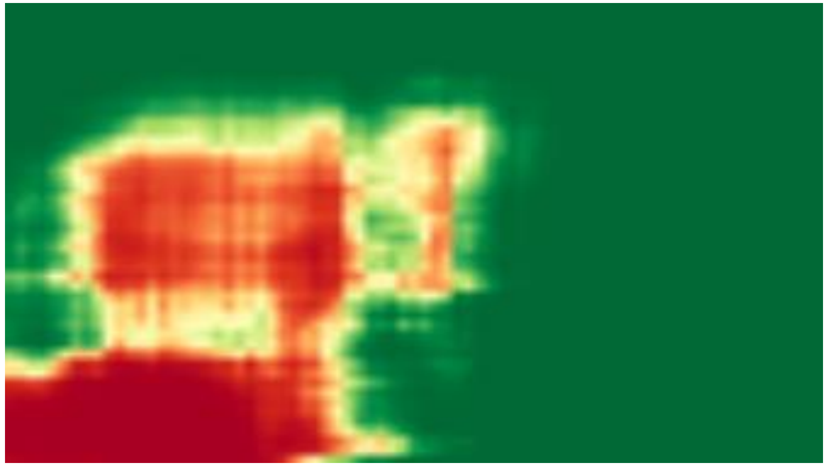
- 16 Layer Model Pre-Trained on ImageNet: **92%** accuracy
  - ImageNet: dataset with millions of ordinary images (i.e. dogs, cats, benches, etc.)



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



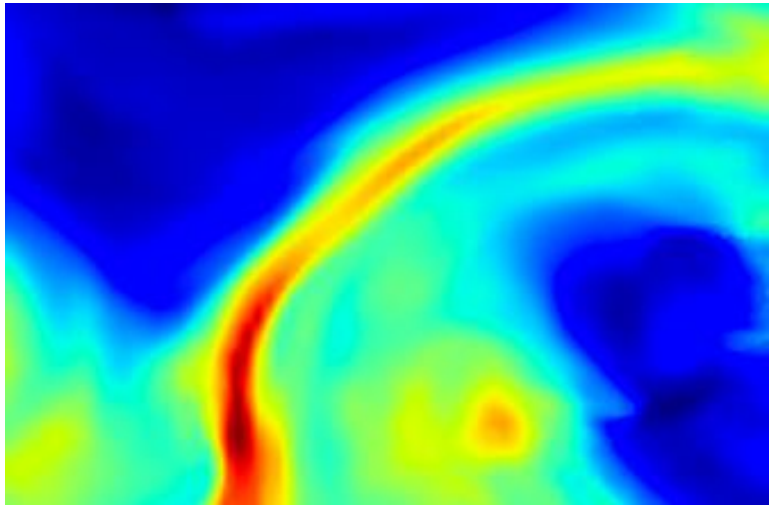
Set some pixels to 0 and record if the model classifies the image as an AR



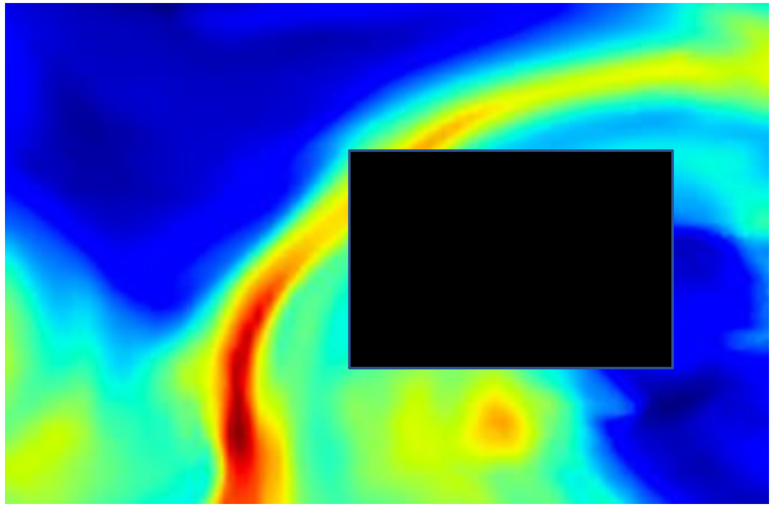
Heat Map: when the bottom left portion of this image is set to 0, the model does not think the image is an AR (**RED**)
If the top left or bottom right portion is set to 0, then the model still thinks the image is an AR (**GREEN**)

A Pre-Trained Model for Classifying ARs

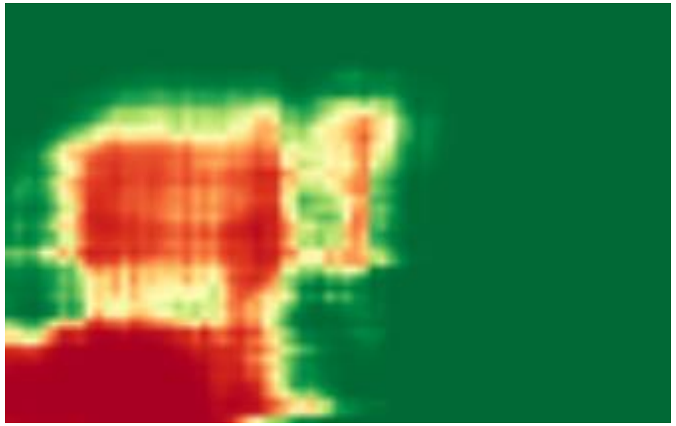
- 16 Layer Model Pre-Trained on ImageNet: **92%** accuracy
  - ImageNet: dataset with millions of ordinary images (i.e. dogs, cats, benches, etc.)
- Conclusion: the model identified the features that make this image an AR!



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**)

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

Classification

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

Segmentation

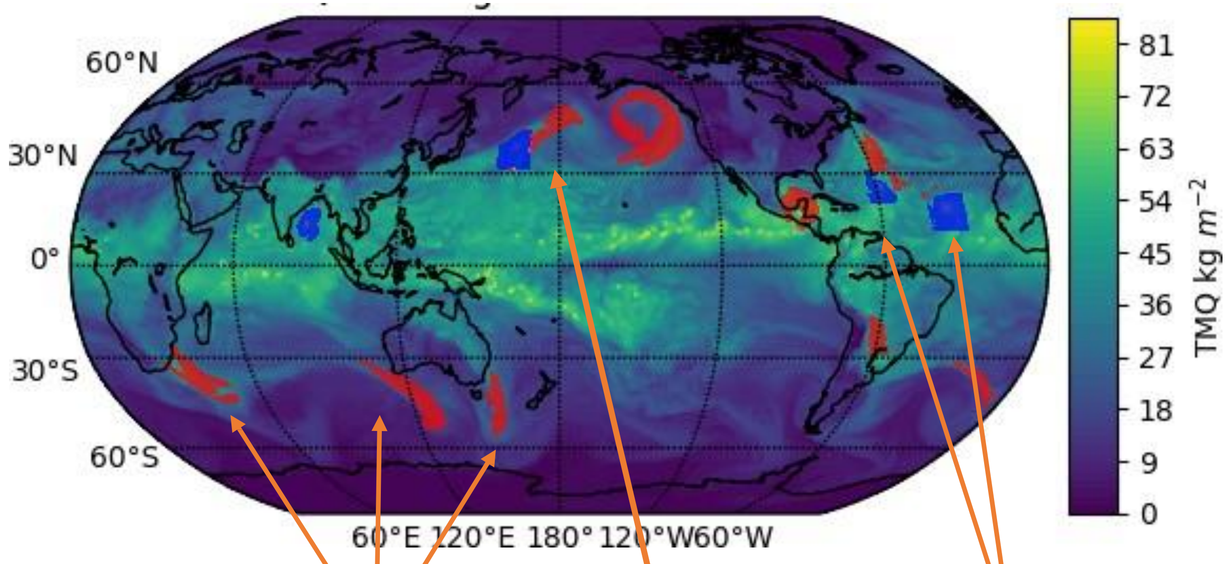
People, bicycles, sidewalk, signposts, roads, and cars are all recognized

Source: Kundu, et al. Feature Space Optimization for Semantic Video Segmentation, 2016.

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

Integrated Water Vapor with Labelled Storms



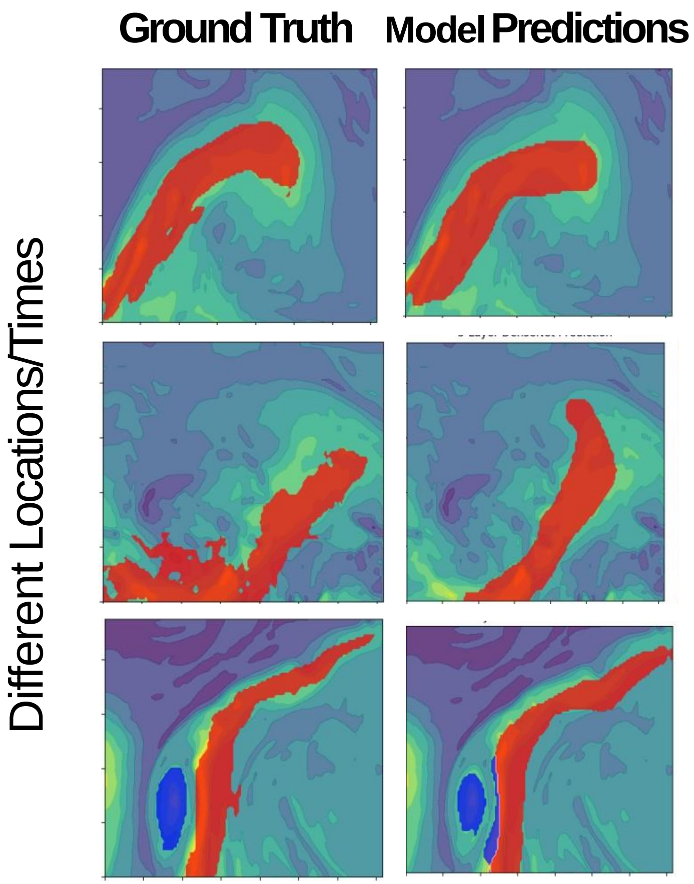
Atmospheric Rivers (RED)

Tropical Cyclones (BLUE)

ARs and TCs often overlap

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

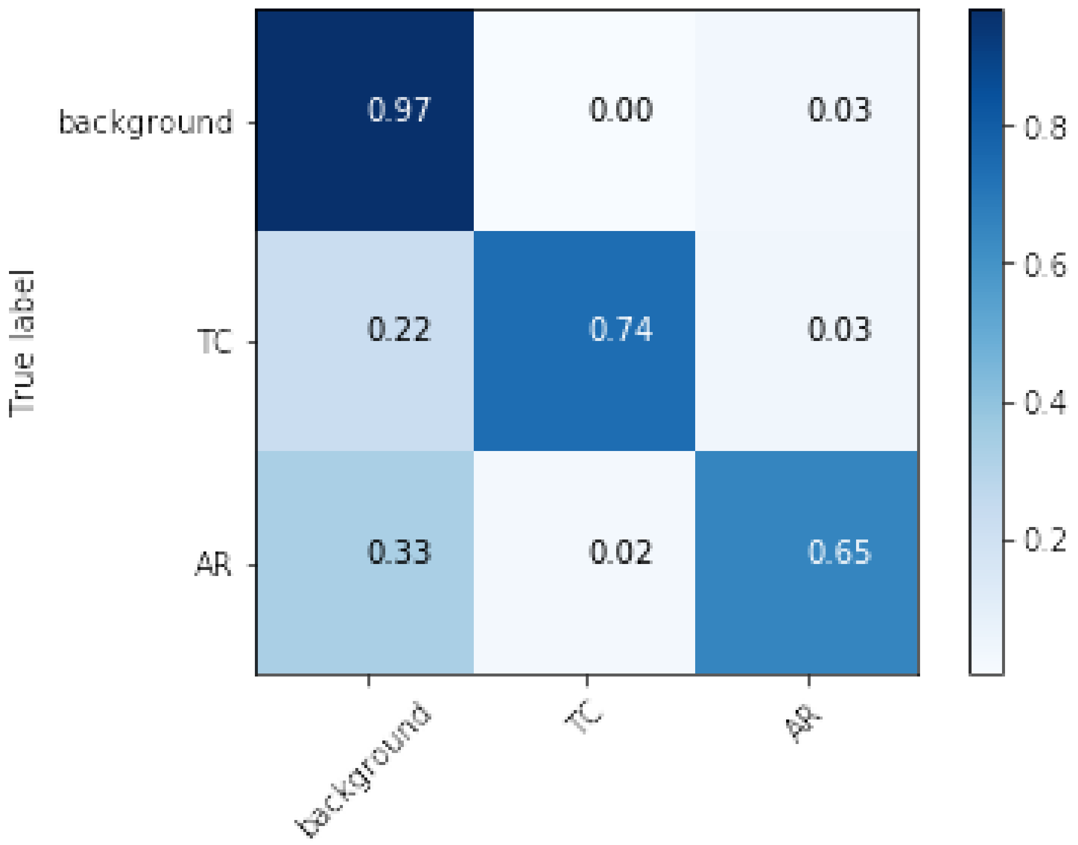


Segmentation of ARs **(RED)** and TCs **(BLUE)** in an IWV image

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?

Segmentation Confusion Matrix



Future Work

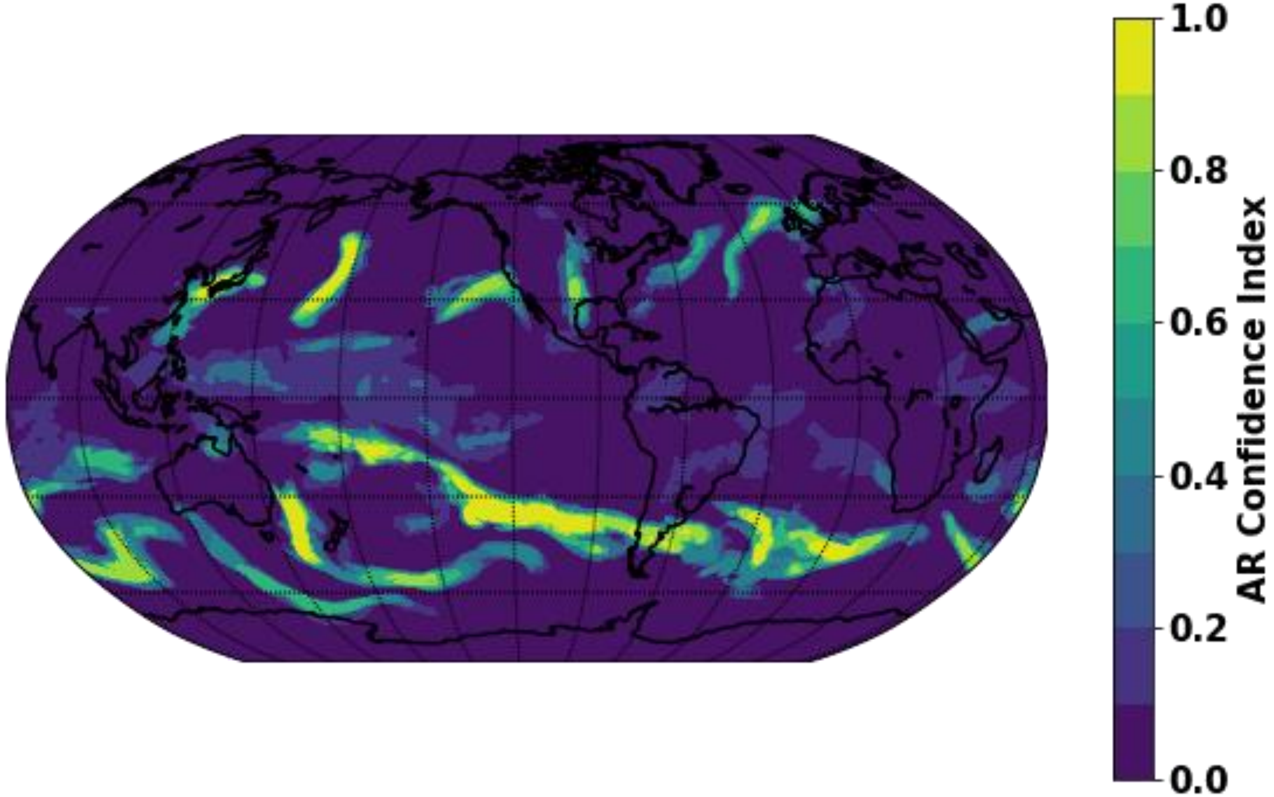
- Investigate how to represent architecture, label, and metric uncertainty
- Ensemble-based extreme event detection
  - Use different labelling strategies to generate multiple ground truth datasets
  - Train a neural network on each ground truth dataset
  - Have each network vote on whether or not an 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