Interpretable Neural Networks for Learning New Science

Elizabeth A. Barnes, Associate Professor, Dept. of Atmospheric Science, CSU

Collaborators for slides in this talk
Benjamin Toms, PhD student, CSU
Imme Ebert-Uphoff, Research Faculty, CSU
Patrick Keys, Research Scientist, CSU
Machine learning for science

data  →  prediction
Machine learning for science

Not a black box!

Visualization tools are a game changer for using machine learning methods for science.
Artificial Neural Networks [ANN]

- linear regression with non-linear mapping by an “activation function”
- training of the network is merely determining the weights “w” and bias/offset “b”

\[ h = f_{activation}(w_1x_1 + w_2x_2 + b) \]

Inputs:
- \( x_1 \)
- \( x_2 \)

Hidden unit:
- \( w_1 \)
- \( w_2 \)

Example:
e.g. gridded sea surface temperatures
Artificial Neural Networks [ANN]

- Linear regression with non-linear mapping by an "activation function"
- Training of the network is merely determining the weights "w" and bias/offset "b"

\[ h = f_{activation}(w_1x_1 + w_2x_2 + b) \]

e.g. gridded sea surface temperatures
Artificial Neural Networks [ANN]

inputs

\( x_1 \)

\( x_2 \)

\( x_3 \)

hidden layers

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)

\( \vdots \)
Artificial Neural Networks [ANN]

e.g. gridded sea surface temperatures

inputs

hidden layers

output/prediction layer

class A (e.g. warmer than average)

class B (e.g. average temperature)

class C (e.g. cooler than average)
Artificial Neural Networks [ANN]

- Complexity and nonlinearities of the ANN allow it to learn many different pathways of predictable behaviour.
- Once trained, you have an array of weights and biases which can be used for prediction on new data.
Complexity and nonlinearities of the ANN allow it to learn many different pathways of predictable behaviour.

Once trained, you have an array of weights and biases which can be used for prediction on new data.

But, how did the network make its prediction? What did it learn?
What to expect from ANN visualization

Not a perfect view, but better than the “black box”.
Two types of visualization tools

Type A: Feature Visualization

Philosophy: Seek to understand all internal components of ANN.

Seek to understand the meaning of all intermediate (blue) nodes.
Two types of visualization tools

Type B: Attribution / Explaining Decisions

Philosophy: Understand the ANN’s overall decision making for specific input.

Seek to understand the meaning of the entire algorithm - for a specific input.

Do NOT worry about meaning of intermediate (blue) nodes.
A visualization tool: Layerwise Relevance Propagation

Prediction of 1 sample

Montavon et al. (2017), Pattern Recognition; Montavon et al. (2018), Digital Signal Processing
A visualization tool: Layerwise Relevance Propagation

Montavon et al. (2017), Pattern Recognition; Montavon et al. (2018), Digital Signal Processing
A visualization tool: Layerwise Relevance Propagation

Prediction of 1 sample

Montavon et al. (2017), Pattern Recognition; Montavon et al. (2018), Digital Signal Processing

where the network looked to determine it was a “cat”
Example use of LRP

**Task:** Decide whether there is a horse in a given image.

**Decision making strategy:** use visualization tools to determine the strategy the network used to make a decision
Example use of LRP

**Task:** Decide whether there is a horse in a given image.

**Decision making strategy:** Use visualization tools to determine the strategy the network used to make a decision.
What does this mean for earth system prediction research?

1. Identifying problematic strategies (i.e. right answer for the wrong reasons)
2. Designing the machine learning methodology
3. Building trust
What does this mean for earth system prediction research?

1. Identifying problematic strategies (i.e. right answer for the wrong reasons)
2. Designing the machine learning methodology
3. Building trust

*LRP showing the relevant regions for the neural network’s prediction of increased human activity*

<table>
<thead>
<tr>
<th>Year</th>
<th>Human Activity Index</th>
<th>Landsat Imagery</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>0.38</td>
<td><img src="image" alt="LRP showing the relevant regions for the neural network’s prediction of increased human activity" /></td>
</tr>
<tr>
<td>2018</td>
<td>0.66</td>
<td><img src="image" alt="LRP showing the relevant regions for the neural network’s prediction of increased human activity" /></td>
</tr>
</tbody>
</table>

*LRP* indicating changes in human activity from Year 2000 to Year 2018.
What does this mean for earth system prediction research?

1. Identifying problematic strategies (i.e. right answer for the wrong reasons)
2. Designing the machine learning methodology
3. Building trust
4. Discovering new science!
   - When our machine learning method is capable of making an accurate prediction we can explore why
Science Applications

1. Multi-year prediction
2. Subseasonal-to-seasonal prediction
3. Indicator patterns of forced change
4. Eddy-mean flow interactions
5. Human impacts on the land surface from Landsat imagery
Science Applications

1. Multi-year prediction
2. Subseasonal-to-seasonal prediction
3. Indicator patterns of forced change
4. Eddy-mean flow interactions
5. Human impacts on the land surface from Landsat imagery
Multi-year prediction network set-up

Time series of sea surface temperature maps

Convolutional neural network

Predicted temperature

Bins of output temperature anomaly (probabilistic)

Predicting 5-year average surface temperature at one grid point
Applied to 1200 years of CESM2 control simulation
Toms et al. (2020; in prep)
Examples of neural network-driven predictions

- Neural network + LRP can be used to identify patterns of earth-system variability that lend predictability

- Here, we predict 5-year average surface temperature using past maps of sea-surface temperature

- Each prediction uses spatially unique information, although dominant patterns emerge

Predicting 5-year average surface temperature at one grid point
Applied to 1200 years of CESM2 control simulation
Toms et al. (2020; in prep)
Examples of neural network-driven predictions

- Neural network + LRP can be used to identify patterns of earth-system variability that lend predictability.

- Here, we predict 5-year average surface temperature using past maps of sea-surface temperature.

- Each prediction uses spatially unique information, although dominant patterns do exist.

Predicting 5-year average surface temperature at one grid point
Applied to 1200 years of CESM2 control simulation
Toms et al. (2020; in prep)
Examples of neural network-driven predictions

- Neural network + LRP can be used to identify patterns of earth-system variability that lend predictability.
- Here, we predict 5-year average surface temperature using past maps of sea-surface temperature.
- Each prediction uses spatially unique information, although dominant patterns do exist.

For us, the science is not the making of a multi-year prediction - it is identifying predictable patterns/regimes of the earth system.

Predicting 5-year average surface temperature at one grid point
Applied to 1200 years of CESM2 control simulation
Toms et al. (2020; in prep)
Wrap-up

- The most basic of neural networks can be viewed as nonlinear regression - climate scientists are well-equipped to think about this architecture.

- Artificial neural networks are no longer black boxes - tools exist to help visualize their decisions. This is a game changer for their use in geoscience research.

- ANNs can be used for more than just prediction. The science can be what the network learns, rather than the prediction. Get creative combining your science with these tools!
References

- **Introduction of LRP to the geosciences:**

- **Use of LRP for identifying patterns of climate change:**

- **Use of LRP for identifying MJO variability:**


Visual Analytics and Interactive Machine Learning for Geospatial Sciences and Cryospheric Research

MORTEZA KARIMZADEH, PH.D.
ASSISTANT PROFESSOR, GEOGRAPHY

ARCUS SIPN2 WEBINAR SERIES
JULY 29, 2020
Labeled Data and Pre-trained Models
Visual Analytics for Machine Learning

1. Real time social media analytics for situational awareness

2. Spambot labeling and behavioral analysis

3. Upcoming NSF EarthCube project on Sea Ice mapping and classification
Situational awareness for first responders:

- Interactive interface
- Visualizations
- Topic modeling
- Advanced filtering
- Trends/anomalies

User-specified filtering based on time, location and topic

Spatial topic lens

Trending topic visualization

Spatial cluster lens

Relevant posts
Harnessing Salient Information in Noisy Text

• How to reduce noise (irrelevant text).
  • Support dynamic phenomena.
    • Spatial dimension.
    • Temporal dimension.
    • Semantic dimension.
  • Support multilingual posts.

• Solution:
  • Interactively incorporate:
    • User knowledge
    • Linguistic context
      • The entire apartment is burning down. → ✓ Relevant
      • Will Bernie feel the burn again? → ✗ Not relevant
Human-in-the-loop Neural Networks

Transform words into a semantic space:

- Word2Vec: A model pre-trained on roughly 100 billion words, provides word embeddings (context of the target word), with each word represented as a 300-dimensional vector.

Word2Vec embeddings

Neural network

Label

Probabilities

Relevant

Not Relevant

Can't Decide

Tweet → Vector matrix → Neural network

Verified labeled tweet

User

Unverified labeled tweet
Evaluation

CrisisLexT26 datasets
- Trained iteratively with 10 tweets

Model reaches its average $F_1$ score after approximately 200 tweets

- 2012 Colorado wildfires
- 2013 Boston bombings
- 2013 NY train crash
The most relevant about weather events:

<table>
<thead>
<tr>
<th>User Name</th>
<th>Creation Date</th>
<th>Tweets Content</th>
<th>Relevance Probability</th>
<th>Relevant Probability</th>
<th>Not Relevant Probability</th>
<th>Can't Decide Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>wHnG5uV5sDF</td>
<td>19-02-19</td>
<td>#DopplerGreg Storm Forecast: Snow, sleet, and rain across NYC &amp; JerseyCity on Wednesday. #NYC #Weather...</td>
<td>Relevant</td>
<td>91.6%</td>
<td>8.4%</td>
<td>0%</td>
</tr>
<tr>
<td>R9DE0E5xRi</td>
<td>19-02-19</td>
<td>#LeeGoldbergABC7 Another snow flap! Another rain... max: sleet!</td>
<td>Relevant</td>
<td>68.9%</td>
<td>25.6%</td>
<td>5.5%</td>
</tr>
<tr>
<td>cMVnHskJw</td>
<td>19-02-19</td>
<td>O'Hare is currently experiencing delays averaging 31 mins due to snow.</td>
<td>Relevant</td>
<td>61.3%</td>
<td>38.7%</td>
<td>0%</td>
</tr>
<tr>
<td>Qy5FvG1G4T</td>
<td>19-02-19</td>
<td>#Snow Snow Snow</td>
<td>Relevant</td>
<td>61.1%</td>
<td>26.9%</td>
<td>12.0%</td>
</tr>
<tr>
<td>ZiSjZU2ZgE</td>
<td>19-02-19</td>
<td>#Snow Snow Snow</td>
<td>Relevant</td>
<td>61.1%</td>
<td>26.9%</td>
<td>12.0%</td>
</tr>
<tr>
<td>s127Z7g2Jo</td>
<td>19-02-19</td>
<td>#winter weather advisory for the Cranford area.</td>
<td>Relevant</td>
<td>60.2%</td>
<td>37.6%</td>
<td>2.2%</td>
</tr>
<tr>
<td>apPH6rmeVv</td>
<td>19-02-19</td>
<td>#winter weather advisory for the Cranford area.</td>
<td>Relevant</td>
<td>58.4%</td>
<td>38.4%</td>
<td>3.2%</td>
</tr>
<tr>
<td>QPdm8vK9MMe</td>
<td>19-02-19</td>
<td>It’s still snowing and snowing still wearing.</td>
<td>Relevant</td>
<td>58.0%</td>
<td>40.7%</td>
<td>1.3%</td>
</tr>
<tr>
<td>0BmXwEv9vz</td>
<td>19-02-19</td>
<td>#winter weather advisory for the Cranford area.</td>
<td>Relevant</td>
<td>55.6%</td>
<td>42.8%</td>
<td>1.6%</td>
</tr>
<tr>
<td>kXOeDn9jYiR</td>
<td>19-02-19</td>
<td>Super Snow Moon tonight...</td>
<td>Relevant</td>
<td>55.5%</td>
<td>43.6%</td>
<td>0.9%</td>
</tr>
</tbody>
</table>

The least relevant about weather events:

<table>
<thead>
<tr>
<th>User Name</th>
<th>Creation Date</th>
<th>Tweets Content</th>
<th>Relevance Probability</th>
<th>Relevant Probability</th>
<th>Not Relevant Probability</th>
<th>Can't Decide Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>TVCJ5m2QmCH</td>
<td>19-02-19</td>
<td>Can you recommend anyone for this IT job in New York, NY?</td>
<td>Not Relevant</td>
<td>69%</td>
<td>31%</td>
<td>0%</td>
</tr>
<tr>
<td>ZFugwCwVw</td>
<td>19-02-19</td>
<td>We're hiring in New York, NY! Click the link to our bio to see it and more. Senior Risk</td>
<td>Not Relevant</td>
<td>83.2%</td>
<td>16.8%</td>
<td>0%</td>
</tr>
<tr>
<td>D787YyX8d</td>
<td>19-02-19</td>
<td>Can you recommend anyone for this job? Manager, FCC Risk</td>
<td>Not Relevant</td>
<td>84.3%</td>
<td>15.7%</td>
<td>0%</td>
</tr>
<tr>
<td>p74h9O9b8</td>
<td>19-02-19</td>
<td>#HarlemBrooklyn# New York, NY</td>
<td>Not Relevant</td>
<td>78.9%</td>
<td>21.1%</td>
<td>0%</td>
</tr>
<tr>
<td>DCD6p9P9eQ</td>
<td>19-02-19</td>
<td>Tune me to help you... I'm playing lonely on New York, NY</td>
<td>Not Relevant</td>
<td>78.4%</td>
<td>21.6%</td>
<td>0%</td>
</tr>
<tr>
<td>D11HYpEq</td>
<td>19-02-19</td>
<td>#<a href="mailto:KatieCameron2@mark.mow">KatieCameron2@mark.mow</a> #AmericanComcast Of course there are.</td>
<td>Not Relevant</td>
<td>78.1%</td>
<td>21.9%</td>
<td>0%</td>
</tr>
<tr>
<td>08b92hK8p</td>
<td>19-02-19</td>
<td>#Way out of #Philadelphia These people will lie right in front of your face. If you told them snow is white or no it's not. #winter #weather</td>
<td>Not Relevant</td>
<td>77.2%</td>
<td>22.8%</td>
<td>0%</td>
</tr>
<tr>
<td>Zg15f4o12nR</td>
<td>19-02-19</td>
<td>@MikeShelshad @syracuse The weather was cold today #0°C almost</td>
<td>Not Relevant</td>
<td>75%</td>
<td>25%</td>
<td>0%</td>
</tr>
<tr>
<td>1G112f32f</td>
<td>19-02-19</td>
<td>#weather is going to be cold tonight #winter almont.</td>
<td>Not Relevant</td>
<td>75%</td>
<td>25%</td>
<td>0%</td>
</tr>
</tbody>
</table>
Social Spambot

A computer algorithm that automatically produces content and interacts with humans on social media, trying to emulate and possibly alter their behavior.

- Spread disinformation
- Manipulate public opinions
- Distribute unsolicited spam
- Propagate malicious links
- Steal personal information

Social Media Accounts

- 40% Spammers

[Ferrara 2016]

[Zhang 2016]
Existing Automated and VA solutions

Issues

- Spambots with natural behavior at individual level ➔ Harder to detect spam groups/campaigns
- Continually Changing Environment ➔ Effort to maintain representative training set
Visual Analytics for Social Spambot Labeling (VASSL)

- Output labels: Spambot or genuine

- Input:
  - Tweet Text
  - Metadata:

Upcoming NSF-funded project: Data Fusion for Sea Ice Classification

- SAR imagery
- Sentinel-1
- NISAR
- IceBridge
- ICESat
- ICESat-2
EarthCube Data Capabilities: Enabling Analysis of Heterogeneous, Multi-source Cryospheric Data

• Morteza Karimzadeh, Geography, Information Science (CU Boulder)
• Farnoush Kashani-Banaei, Computer Science (CU Denver)
• Andrew Barrett (NSIDC)
• Walt Meier (NSIDC)
• Siri Jodha Khalsa (NSIDC)
Thank you!

Q/A

Karimzadeh@colorado.edu

@morteza_kz
IceNet: A seasonal, deep learning-based pan-Arctic sea ice forecasting system

Tom Andersson

Scott Hosking, María Pérez-Ortiz, Brooks Paige, Chris Russell, Andrew Elliott, Stephen Law, Tony Phillips, Jeremy Wilkinson, Yevgeny Askenov, Bablu Sinha, Will Tebbutt, Fruzsina Agocs, and Emily Shuckburgh

British Antarctic Survey, Alan Turing Institute, Cambridge University, UCL Centre for AI, National Oceanography Centre
Two climate forecasting paradigms: Physics-driven vs. data-driven

**Dynamical models (physics-driven)**

- Model the laws of physics directly
- Based on causality
- Computationally expensive

Credit: Schneider et al., Nature Climate Change
Two climate forecasting paradigms: Physics-driven vs. data-driven

**Dynamical models (physics-driven)**
- Model the laws of physics directly
- Based on causality
- Computationally expensive

**Statistical models (data-driven)**
- Automatically learn complex, non-linear relationships between variables from raw data
- Based on correlations
- Computationally cheap (once trained)

Credit: Vinyals et al., CVPR
Credit: Schneider et al., Nature Climate Change
Credit: Shutterstock
Credit: DeepMind
**IceNet data: Observations**

- **NSIDC**
  - sic

- **ERA5**
  - t2m
  - sst
  - msl
  - windspeed
  - u10
  - v10

**Time period:** 1979-present (500 months)
IceNet data: Climate model (MRI-ESM2.0)

Time period: 1850-2100 (3012 months)
*IceNet* design: Inputs and outputs
IceNet design: Inputs and outputs

- Inputs
- Outputs

$t$ (months)
IceNet design: Inputs and outputs

$t$ (months)

Inputs

Outputs

1 month ahead
6 months ahead

IceNet
IceNet design: U-Net Architecture

2D Convolution:
IceNet design: U-Net Architecture

- **Inputs**
- **Outputs**
- **Abstraction increases**
- **Resolution decreases**
- **Convolution + nonlinear function**
- **Downsampling**
- **Upsampling**
- **Concatenate**

Resolution decreases and Abstraction increases through the layers of the network.

1 month ahead
6 months ahead

Convolution:
IceNet design: U-Net Architecture

Inputs

Abstraction increases

Resolution decreases

Convolution + nonlinear function

Downsampling

Upsampling

Concatenate

Activation maps

Convolution:
IceNet design: U-Net Architecture

- Inputs
- Abstraction increases
- Resolution decreases

- Convolution + nonlinear function
- Downsampling
- Upsampling
- Concatenate

Activation maps
*IceNet* design: U-Net Architecture

- **Inputs**
- **Abstraction increases**
- **Resolution decreases**
- **Convolution + nonlinear function**
- **Downsampling**
- **Upsampling**
- **Concatenate**
- **Activation maps**
IceNet design: U-Net Architecture

- Inputs

- Abstraction increases

- Resolution decreases

- Convolution + nonlinear function

- Downsampling

- Upsampling

- Concatenate

Activation maps
**IceNet** design: U-Net Architecture

- **Inputs**
- **Abstraction increases**
- **Resolution decreases**
- **Convolution + nonlinear function**
- **Downsampling**
- **Upsampling**
- **Concatenate**

Activation maps
**IceNet** design: U-Net Architecture

- **Inputs**
- Abstraction increases
- Resolution decreases
- Convolution + nonlinear function
- Downsampling
- Upsampling
- Concatenate

 Activation maps
**IceNet** design: U-Net Architecture

- **Inputs**
- **Resolution decreases**
- **Abstraction increases**
- **Convolution + nonlinear function**
- **Downsampling**
- **Upsampling**
- **Concatenate**

**Activation maps**

**Convolution:**

- **Resolution decreases**
- **Abstraction increases**
- **Convolution + nonlinear function**
- **Downsampling**
- **Upsampling**
- **Concatenate**
**IceNet** design: U-Net Architecture

- **Inputs**
- **Abstraction increases**
- **Resolution decreases**
- **Convolution + nonlinear function**
- **Downsampling**
- **Upsampling**
- **Concatenate**
- **Outputs**

1 month ahead
6 months ahead
*IceNet* design: U-Net Architecture

- Three output classes:
  a. No ice (SIC < 15%)
  b. Marginal ice (15% < SIC < 80%)
  c. Full ice (SIC > 80%)

- # of params: 10,983,434
- Pre-train on >10,000 months of climate model data (MRI-ESM2.0)
- Fine-tune on 1979-2015 observational data
- Validate (hindcast) on 2016-2018
- Ensemble of 3 networks
IceNet predictions: Predict entire second half of 2017 starting in June
**IceNet** predictions: Predict second half of 2017 one month ahead
IceNet predictions: September 2018
**IceNet** predictions: Prediction uncertainty (Aug 2017)

$p(\text{ice}) = p(\text{marginal ice}) + p(\text{full ice})$

- **3 class entropy**
- **2 class entropy**

Observed

1 month ahead

1 month ahead
Hindcast results: 1 month ahead
Hindcast results: 6 months ahead
Validation mean performance vs. lead time

- **Model**: Persistence, IceNet

Left graph: Accuracy (%)
- X-axis: Lead time (months)
- Y-axis: Accuracy (%)
- Persistence: Blue dots, IceNet: Orange crosses

Right graph: Absolute SIE error (km$^2$)
- X-axis: Lead time (months)
- Y-axis: Absolute SIE error (km$^2$)
- Persistence: Blue dots, IceNet: Orange crosses
Thanks for listening!

Contact: tomand@bas.ac.uk
Entropy