Information Visualization

eXplainable Artificial Intelligence

Dr. David Koop
High-Dimensional Data Exploration

• What are the tasks?
  - Discovering data configurations according to personal preference
  - Understanding the tradeoffs involved in such configurations
  - Partition the data (or views) to help with exploration

• Goals of TripAdvisorND & Subspace Voyager:
  - Facilitate examination of key projection and key clustering
  - Let users explore and tweak

TripAdvisor-ND: Global Sight Map & Local Sight Explorer

Sight map with sight glyphs of sub-space projections

Control panel

Dynamic multivariate scatterplot display

N-D touchpad polygon

Touchpad configuration controls

Glyph currently displayed in the local sight explorer

Vector component bar chart display

Data axes vector display

Scatterplot controls

Vector component bar chart display
Coordination of Views

(a) 
(b) 
(c) 

N-D Touchpad Polygon

- 2 polygons, one for each axis (inner = x, outer = y)
- Controls the orientation of the two PPA vectors
- Shading of vertices indicates weight
- Move the vertices around to change the weights

Problems with TripAdvisor-ND

• Have to keep track of two views at once
• …so single window
• Have to move around two points in ND trackpad
• …so trackball interface
• Hard to map axes
• …so direct labeling
Subspace Voyager Interface

Subspace Explorer

Subspace Trail Map

[Subspace Voyager Interface diagram by B. Wang & K. Mueller, 2018]
• Can use different mouse buttons
• Left: rotation
• Right: transition by changing axis
• Middle: travel along orthogonal vector (a z-axis), can't see until rotation
Fix Labels

(a) Naïve implementation causing label overlap; (b) Using our angular label overlap prevention scheme; (c) Illustration of our label overlap prevention scheme.
The Building Blocks of Interpretability

Interpretability techniques are normally studied in isolation. We explore the powerful interfaces that arise when you combine them—and the rich structure of this combinatorial space.

For instance, by combining feature visualization (what is a neuron looking for?) with attribution (how does it affect the output?), we can explore how the network decides between labels like Labrador retriever and tiger cat.

Several floppy ear detectors seem to be important when distinguishing dogs, whereas pointy ears are used to classify “tiger cat”.

Channels that most support...

<table>
<thead>
<tr>
<th>Feature visualization of channel</th>
<th>Labrador retriever</th>
<th>Tiger cat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net evidence for “Labrador retriever”</td>
<td>1.63 1.51 1.19</td>
<td>1.62 1.54 1.72</td>
</tr>
<tr>
<td>Net evidence for “tiger cat”</td>
<td>-0.40 -0.27 0.13</td>
<td>0.62 0.30 1.29</td>
</tr>
</tbody>
</table>
Introduction to eXplainable AI (XAI)

Q. V. Liao, M. Singh, Y. Zhang, and R. Bella
Survey of visualization in deep learning

Visual Analytics in Deep Learning | Interrogative Survey Overview

§4 WHY
Why would one want to use visualization in deep learning?
- Interpretability & Explainability
- Debugging & Improving Models
- Comparing & Selecting Models
- Teaching Deep Learning Concepts

§6 WHAT
What data, features, and relationships in deep learning can be visualized?
- Computational Graph & Network Architecture
- Learned Model Parameters
- Individual Computational Units
- Neurons in High-dimensional Space
- Aggregated Information

§8 WHEN
When in the deep learning process is visualization used?
- During Training
- After Training

§5 WHO
Who would use and benefit from visualizing deep learning?
- Model Developers & Builders
- Model Users
- Non-experts

§7 HOW
How can we visualize deep learning data, features, and relationships?
- Node-link Diagrams for Network Architecture
- Dimensionality Reduction & Scatter Plots
- Line Charts for Temporal Metrics
- Instance-based Analysis & Exploration
- Interactive Experimentation
- Algorithms for Attribution & Feature Visualization

§9 WHERE
Where has deep learning visualization been used?
- Application Domains & Models
- A Vibrant Research Community

[Graphic representation of visual analytics concepts]

Interrogative Survey Overview

In this survey, we summarize a large number of deep learning visualization works using the Five W’s and How (Why, Who, What, How, When, and Where). Figure 1 presents a visual overview of how these interrogative questions reveal and organize the various facets of deep learning visualization research and their related topics. By framing the survey in this way, many existing works fit a description as the following fictional example:

To interpret representations learned by deep models (why), model developers (who) visualize neuron activations in convolutional neural networks (what) using t-SNE embeddings (how) after the training phase (when) to solve an urban planning problem (where).

This framing captures the needs, audience, and techniques of deep learning visualization, and positions new work’s contributions in the context of existing literature.

We conclude by highlighting prominent research directions and open problems. We hope that this survey acts as a companion text for researchers and practitioners wishing to understand how visualization supports deep learning research and applications.
### Survey Landscape

<table>
<thead>
<tr>
<th>Work</th>
<th>WHY</th>
<th>WHO</th>
<th>WHAT</th>
<th>HOW</th>
<th>WHEN</th>
<th>WHERE</th>
</tr>
</thead>
</table>


[Hohman et al.]
The Building Blocks of Interpretability

Interpretability techniques are normally studied in isolation. We explore the powerful interfaces that arise when you combine them—and the rich structure of this combinatorial space.

For instance, by combining feature visualization (what is a neuron looking for?) with attribution (how does it affect the output?), we can explore how the network decides between labels like **Labrador retriever** and **tiger cat**.

Several floppy ear detectors seem to be important when distinguishing dogs, whereas pointy ears are used to classify “tiger cat”.

![Feature visualization and attribution maps for Labrador retriever and tiger cat](image)

<table>
<thead>
<tr>
<th>CHANNELS THAT MOST SUPPORT...</th>
<th>LABRADOR RETRIEVER</th>
<th>TIGER CAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>feature visualization of channel</td>
<td><img src="image" alt="Images of Labrador and tiger for feature visualization" /></td>
<td><img src="image" alt="Images of Labrador and tiger for feature visualization" /></td>
</tr>
<tr>
<td>hover for attribution maps</td>
<td><img src="image" alt="Hoverable attribution maps for Labrador and tiger" /></td>
<td><img src="image" alt="Hoverable attribution maps for Labrador and tiger" /></td>
</tr>
<tr>
<td>net evidence for “Labrador retriever”</td>
<td><img src="image" alt="Net evidence values for Labrador" /></td>
<td><img src="image" alt="Net evidence values for Labrador" /></td>
</tr>
<tr>
<td>net evidence for “tiger cat”</td>
<td><img src="image" alt="Net evidence values for tiger" /></td>
<td><img src="image" alt="Net evidence values for tiger" /></td>
</tr>
<tr>
<td>1.63</td>
<td>1.51</td>
<td>1.19</td>
</tr>
<tr>
<td>1.22</td>
<td>1.24</td>
<td>1.32</td>
</tr>
<tr>
<td>-0.40</td>
<td>-0.27</td>
<td>0.13</td>
</tr>
</tbody>
</table>

| 1.32 | 1.54 | 1.72 |
| -0.70 | -1.24 | -0.43 |
| 0.62 | 0.30 | 1.29 |

[C. Olah et al.]

D. Koop, CSCI 628, Fall 2021