

# ShapeShop

Open-sourced on Github  
[github.com/fredhohman/shapeshop](https://github.com/fredhohman/shapeshop)

## Towards Understanding Deep Learning Representations via Interactive Experimentation

Fred Hohman  
fredhohman@gatech.edu

Nathan Hodas  
nathan.hodas@pnnl.gov

Polo Chau  
polo@gatech.edu

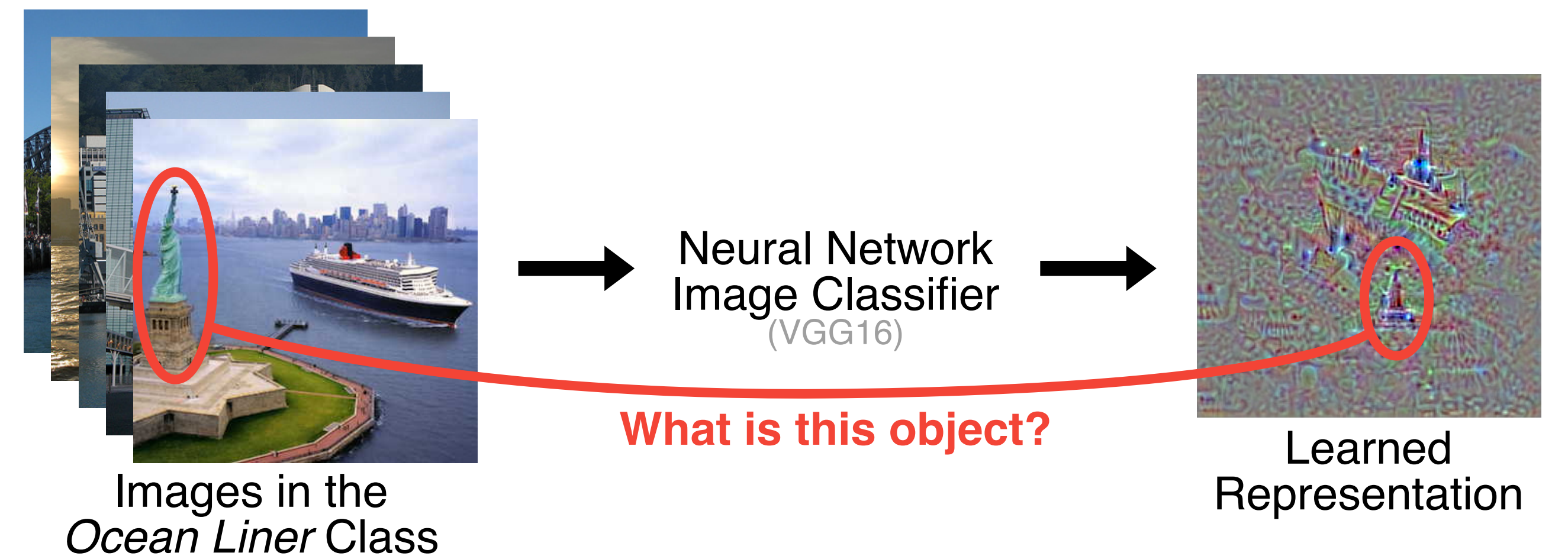
Pacific Northwest  
NATIONAL LABORATORY

Georgia  
Tech

## Summary

ShapeShop is an interactive system for visualizing and understanding what representations a neural network model has learned.

Built using standard web technologies, ShapeShop allows users to experiment with and compare deep learning models to help explore and understand the relationship between data and its learned representations.



### ShapeShop

Visualizing Deep Learning Representations in Simple Shapes via Interactive Experimentation

Classifier Model Builder

- Select Training Data
- Select Model
- Select Hyperparameters
- Train and Visualize

MLP: multilayer perceptron. CNN: convolutional network.

Initial Image: [Black square] Step-size: [0.001] [0.005] [0.01] Epoch Count: [50] [100] [150] [200] [250]

Train classifier and visualize its learned representations by performing class activation maximization on each class.

Train and Visualize

#### Experiment Results

[1]: Model: MLP, Step-size: 0.005, Epoch Count: 100

0.35	0.59	0.34	0.45	0.03	0.06
------	------	------	------	------	------

[2]: Model: CNN, Step-size: 0.005, Epoch Count: 100

0.57	0.85	0.68	0.75	0.01	0.19
------	------	------	------	------	------

#### Class activation maximization of the vertical line class

Original Image	Correlation Coefficient
[Image]	0.59

## Example

Addition of *diverse data* produces more human recognizable representations.

- Original binary classifier
- Correlation coefficient improves to 0.52 after adding both circular classes
- Correlation coefficient improves to 0.59 after adding both noise classes

0.39	0.39
0.52	0.53
0.38	0.39
0.59	0.57
0.36	0.46
0.01	0.02

## Ongoing Work

Incorporating human feedback to improve interpretability via semantic highlighting.

Yellow = ✓  
Red = ✗



This work supported in part by grants NSF IIS-1563816 and NIH BD2K U54EB020404.

ShapeShop: Towards Understanding Deep Learning Representations via Interactive Experimentation.

Fred Hohman, Nathan Hodas, Duen Horng (Polo) Chau.

Extended Abstracts, ACM Conference on Human Factors in Computing Systems (CHI), May 6-11, 2017, Denver, CO, USA.