How to visualize neural network parameters and activity

Justin Shenk
shenk.justin@gmail.com
justinshenk on GitHub

PyData Warsaw 2017

About Me

- Former neuroscience researcher (San Antonio, Texas)
- Master student in Cognitive Science, University of Osnabrück
- Thesis on “Breaking the black box of deep learning” with Peltarion (Stockholm)
- Intel AI Software Innovator

PELTARION

Overview

Motivation

Convolutional Neural Networks - Jason Yosinski’s Deep Visualization Toolbox

Recurrent Neural Networks - Andrej Karpathy’s text-based examples

TensorFlow’s TensorBoard

Notable Projects

Summary

Motivation

Insight into nature

- Predict and **Understand**

Insight into deep learning

- End-users
- Model developers

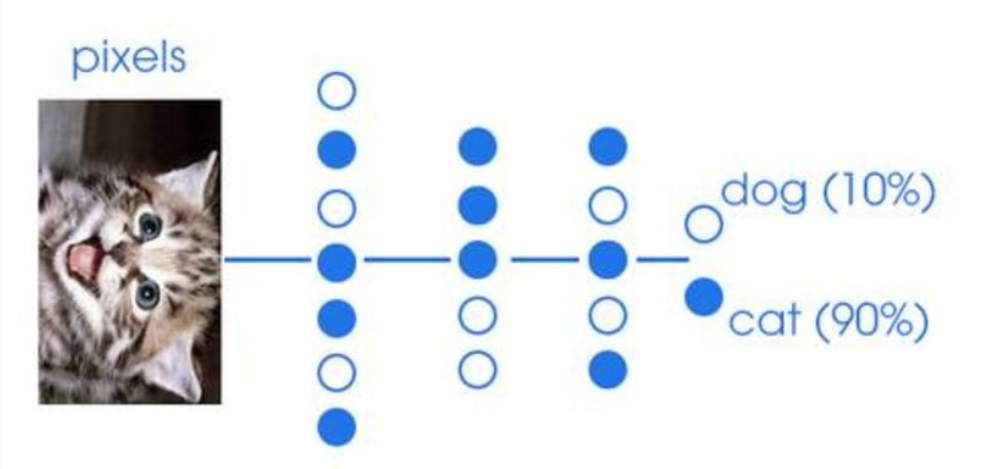
Guide architecture selection

- Increase performance

explain

“What I cannot ~~create,~~ I do not understand.”

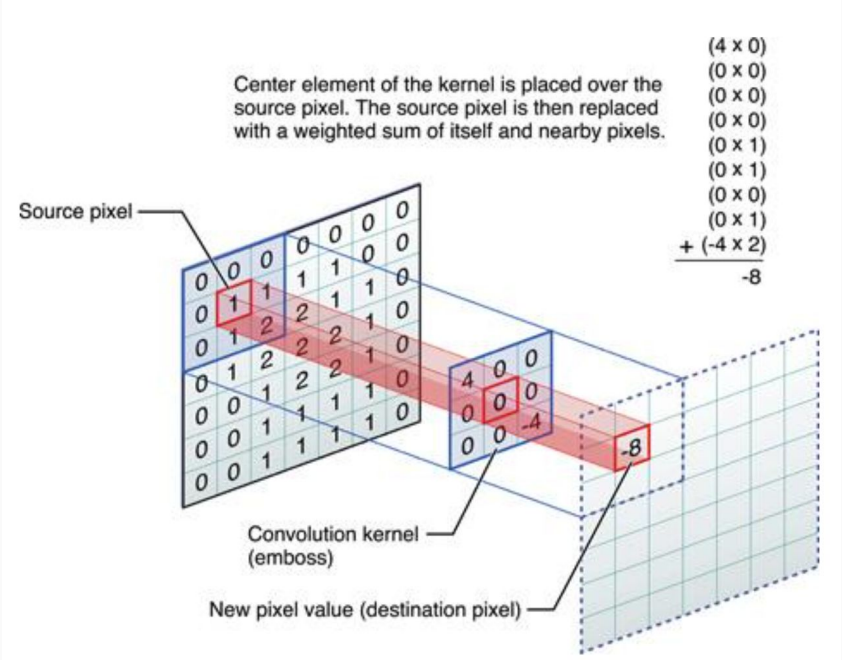
- Richard Feynman



[E]

Convolutional Neural Networks

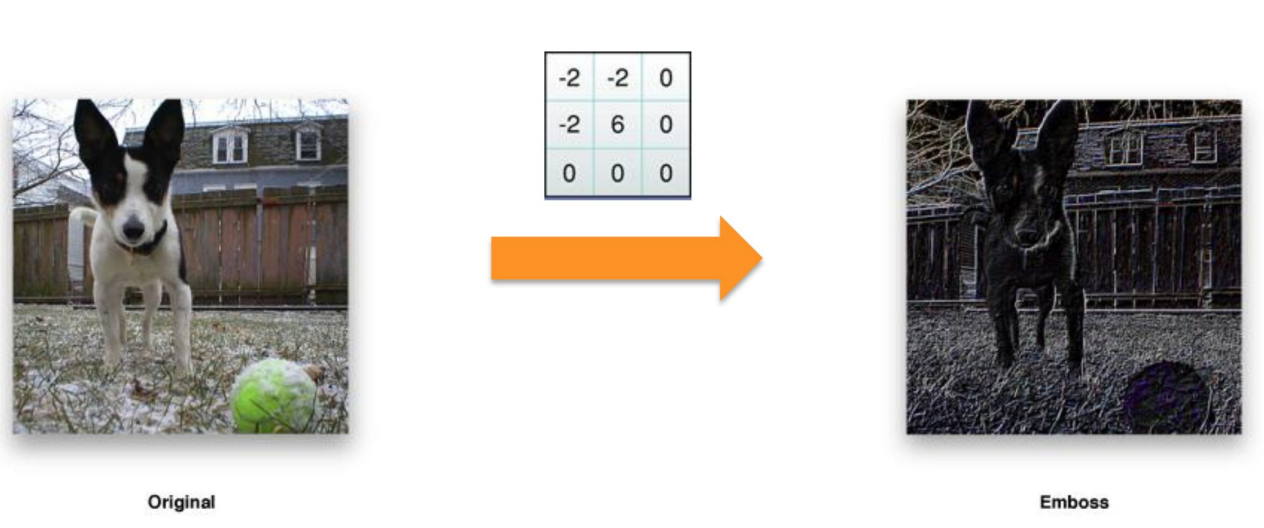
Introduction to CNN



[A]

Convolutional Neural Networks

Feature map



[A]

Convolutional Neural Networks

Deep Visualization Toolbox

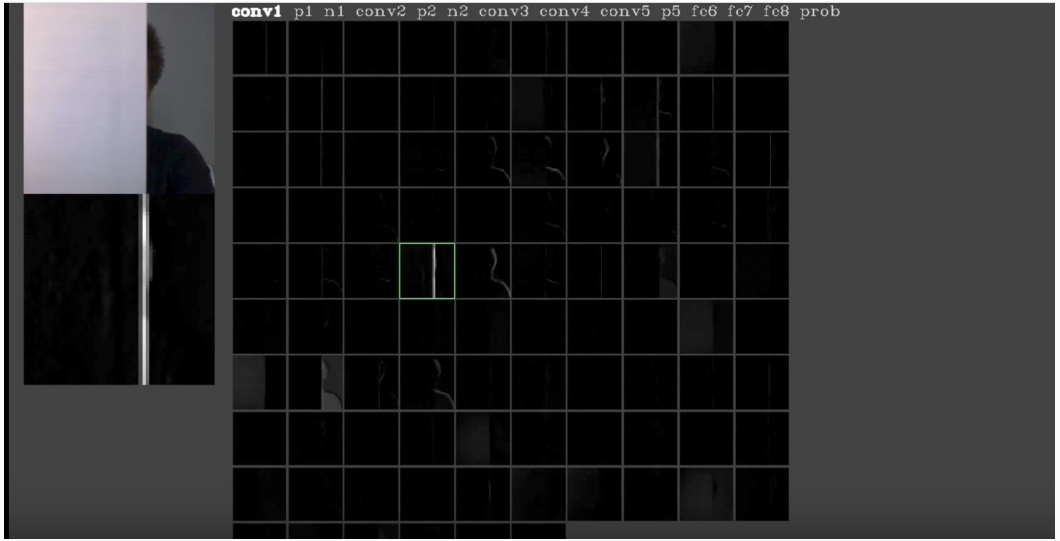
- https://github.com/yosinski/deep-visualization-toolbox

- Synthesized images: Yosinksi et al, Understanding Neural Networks Through Deep Visualization - http://yosinski.com/deepvis
- Important pixels via deconv: Zeiler and Fergus
- AlexNet in Caffe

Convolutional Neural Networks

Low level feature activation

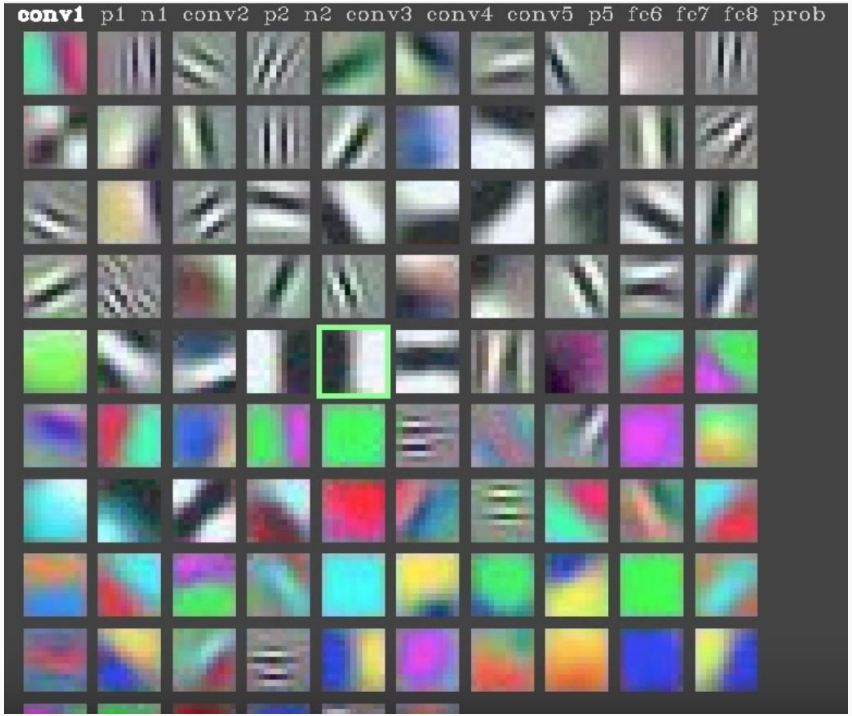
Unit responds to light-dark edges



[Y]

Convolutional Neural Networks

Feature activation and images synthesized to produce high unit activation



[Y]

Convolutional Neural Networks

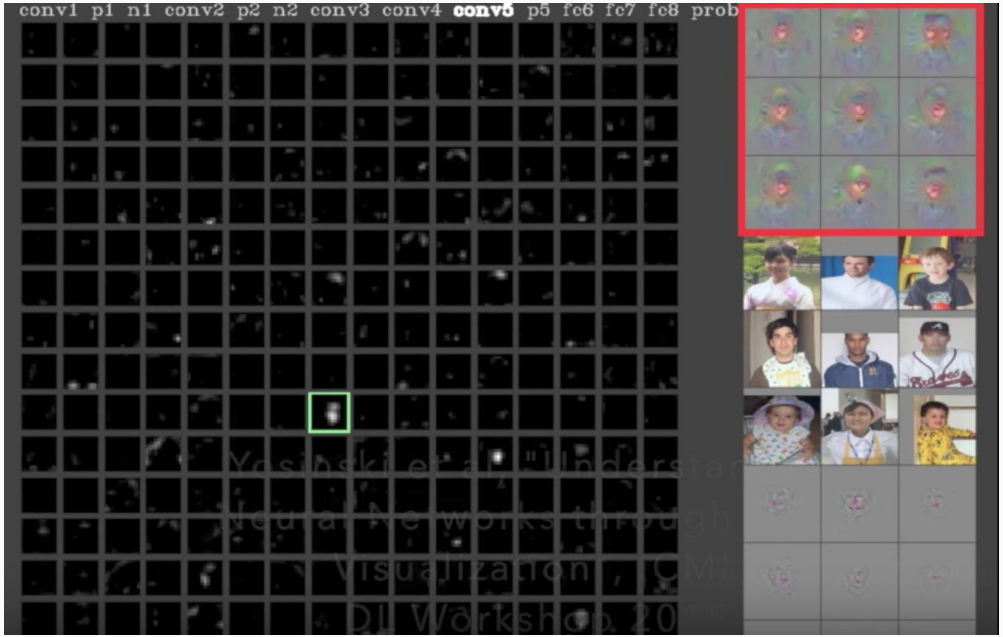
Switching between unit activations and images synthesized to produce high activation via backpropagation

[Y]

Convolutional Neural Networks

5th layer - abstract concepts

- Face & shoulders unit

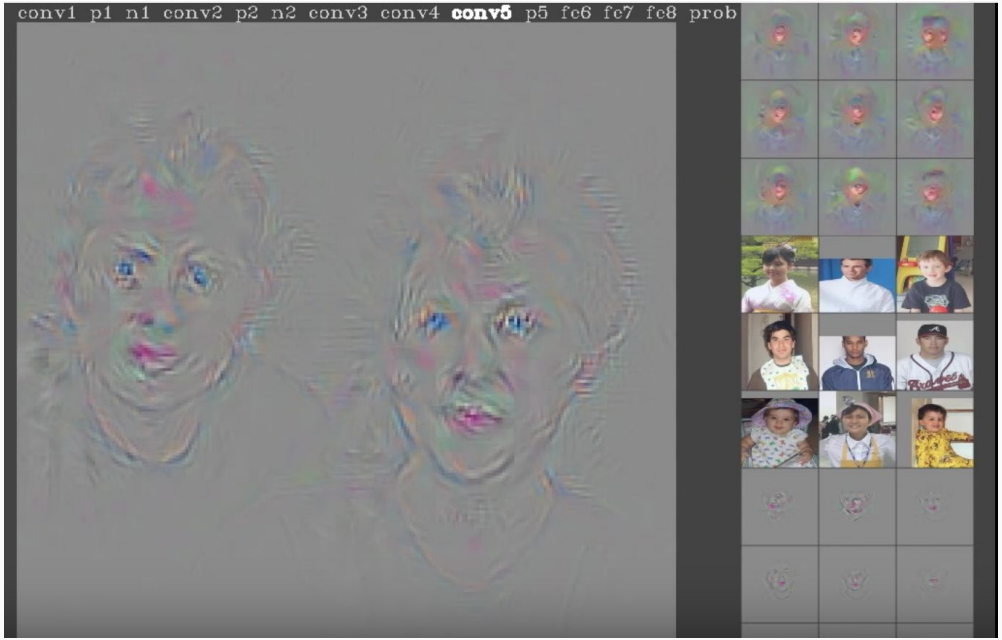


[Y]

Convolutional Neural Networks

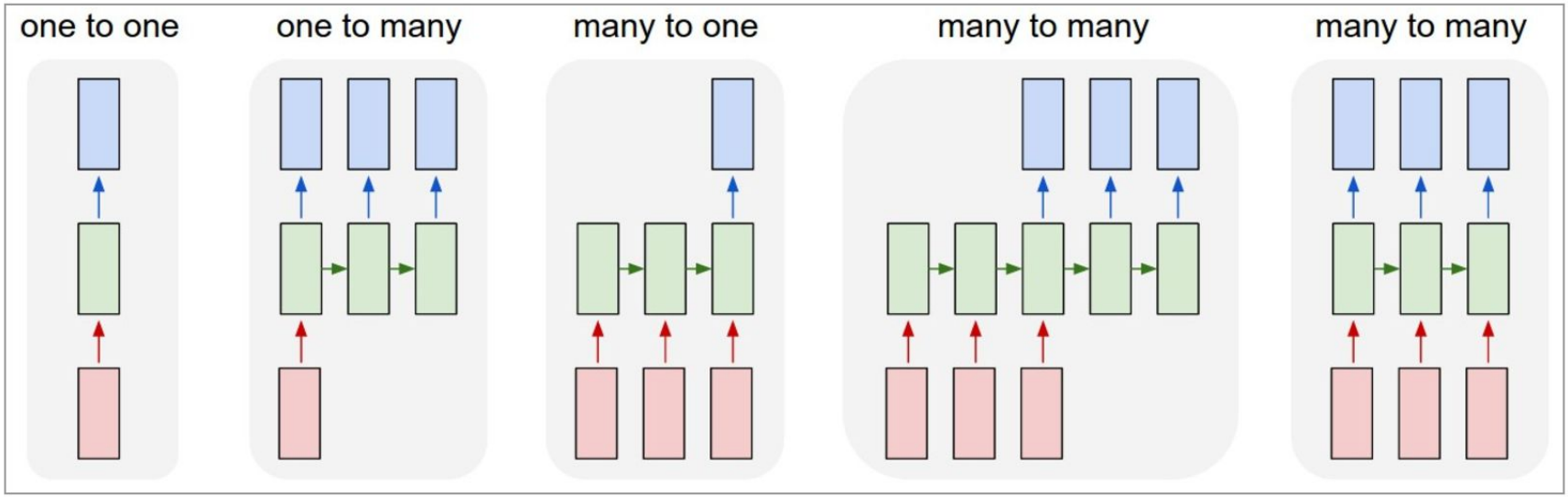
5th layer - abstract concepts

- Face & shoulders unit (like “grandmother cell”)



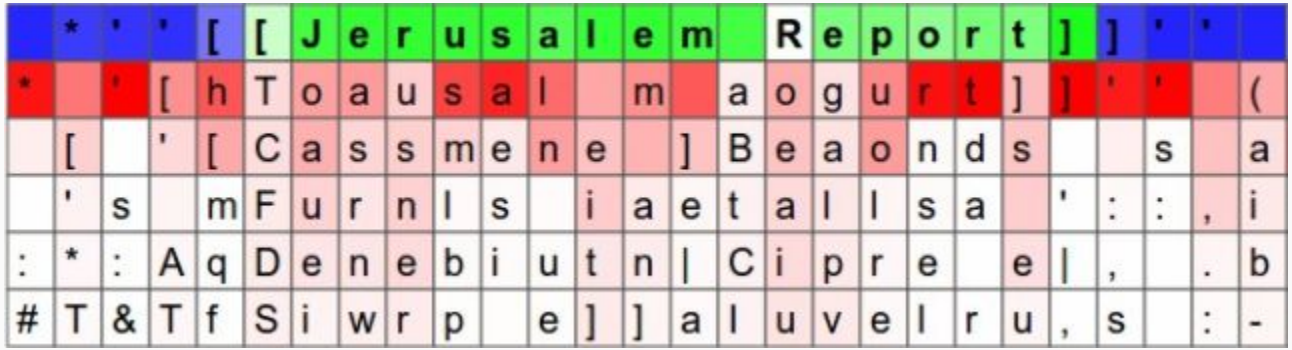
[Y]

Recurrent Neural Networks



[K]

Recurrent Neural Networks

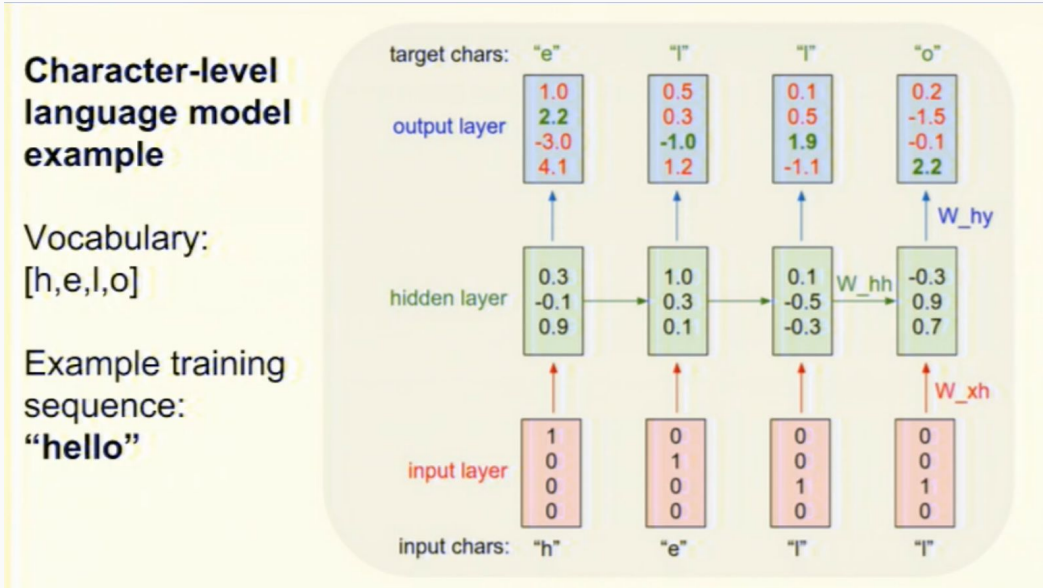


[K]

Andrej Karpathy's seminal [blog post](#) (2015)

Recurrent Neural Networks

Make predictions based on current state and previous state



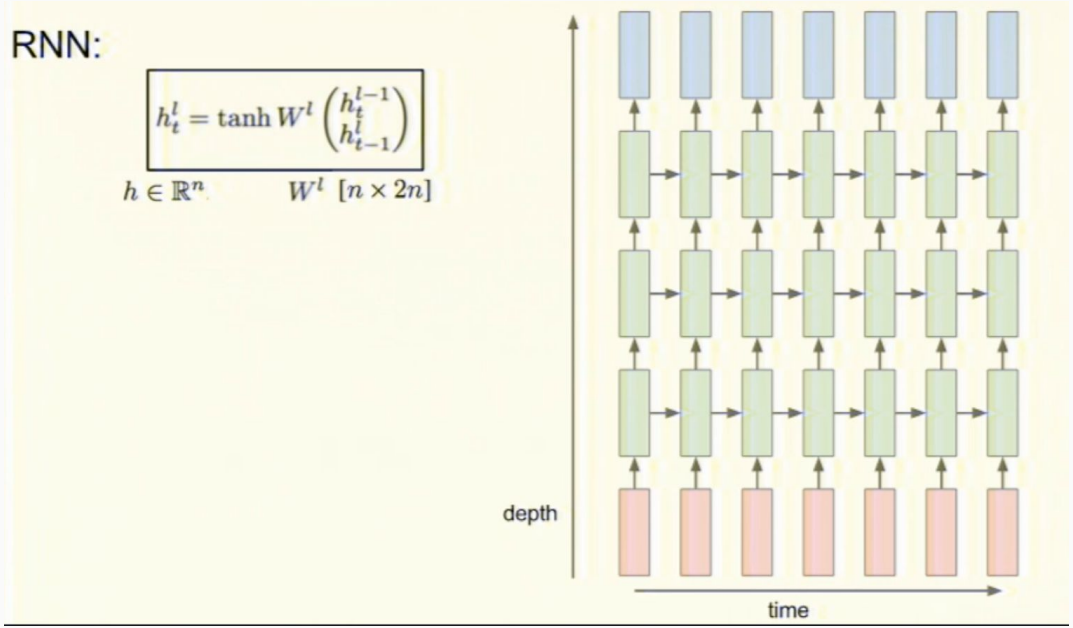
[K]

Recurrent Neural Networks

Carry memory forward:

$$\mathbf{h}_t = \phi\left(W\mathbf{x}_t + U\mathbf{h}_{t-1}\right)$$

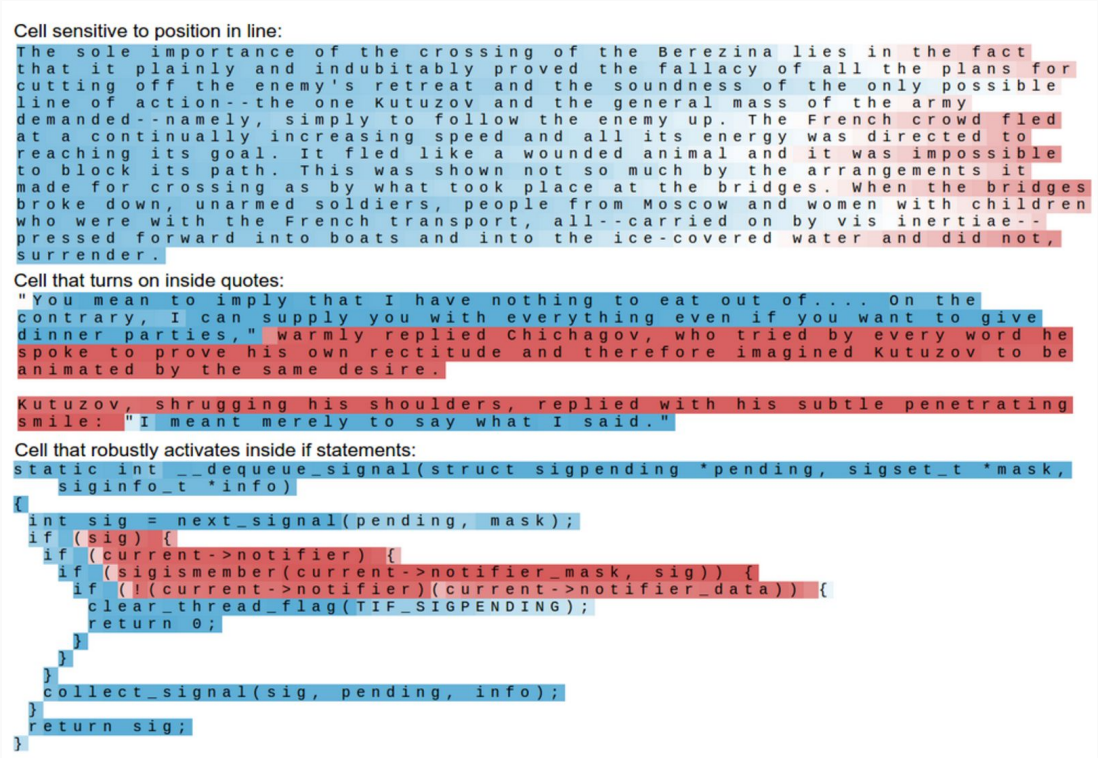
W is weight matrix
x_t is input
h_{t-1} is hidden layer at time = t-1
U is transition matrix



[K]

Recurrent Neural Networks

Cell activation visualization

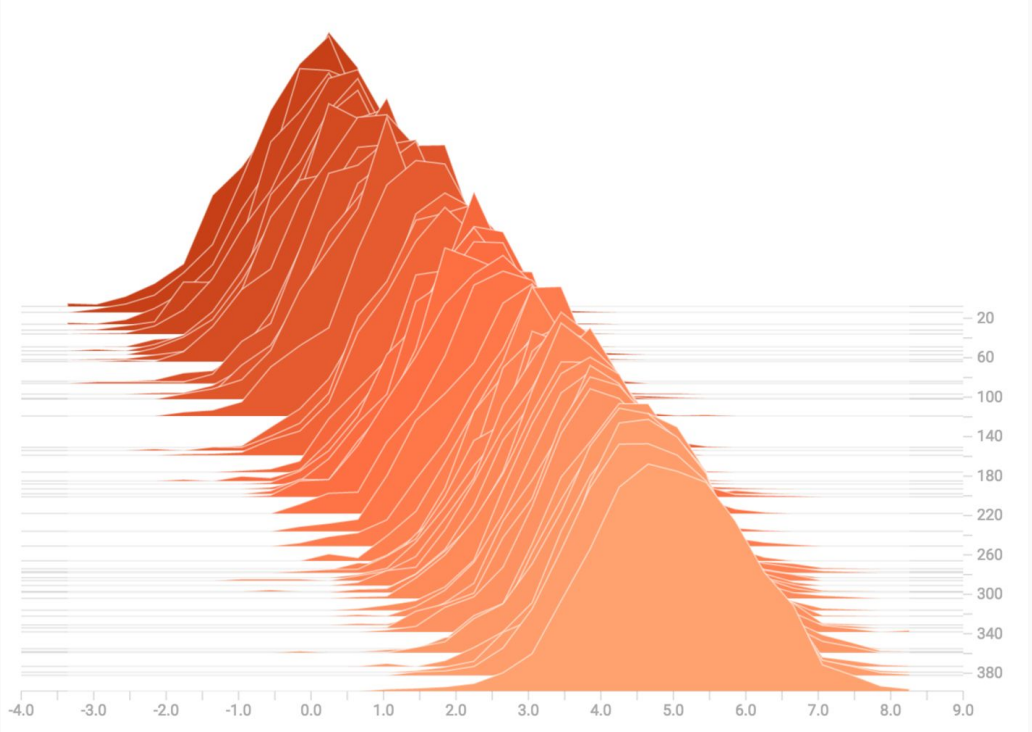


TensorBoard

TensorBoard:

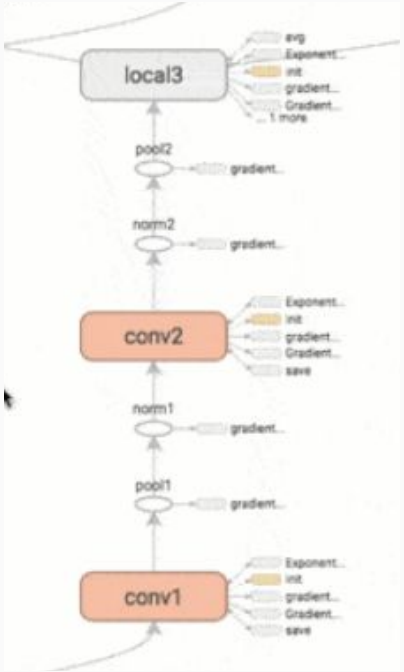
- Weight histogram and accuracy variance during training
- Embedding visualization
- Graph visualization
- Hyperparameter search

Insight from from weight variance



TensorBoard

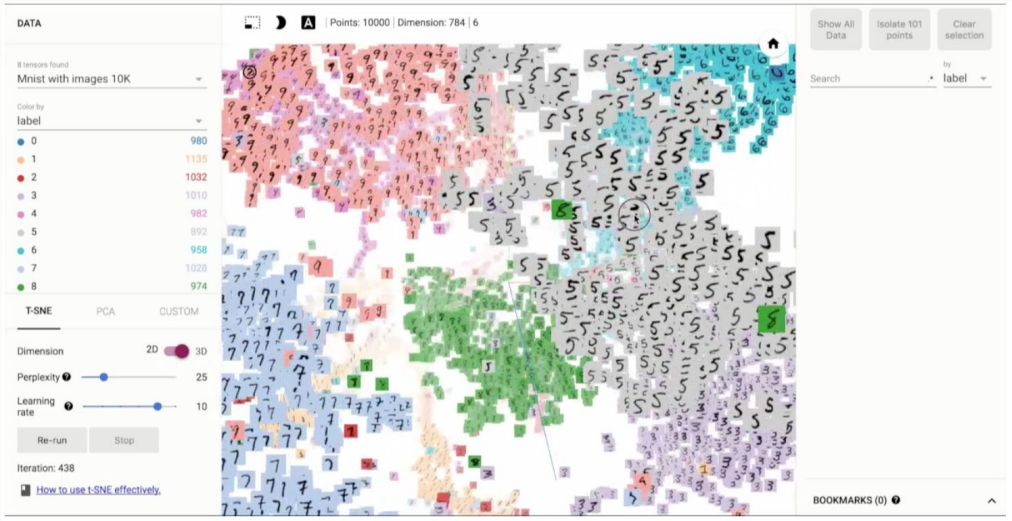
Graph visualization



[T]

TensorBoard

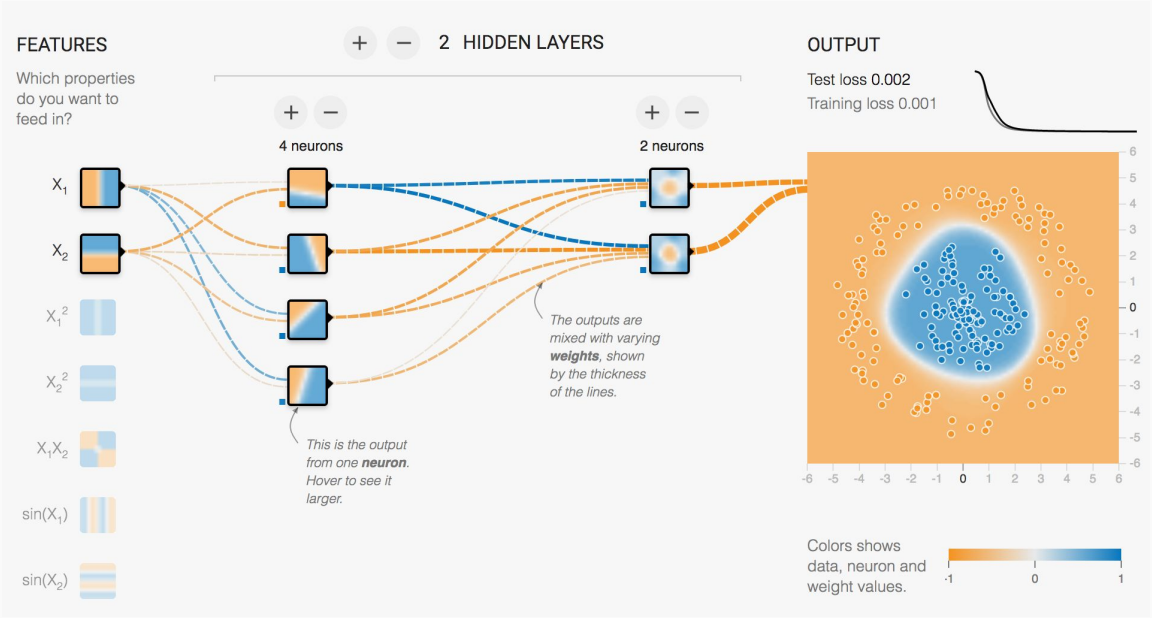
Embedding visualization using PCA and t-SNE



[T]

TensorBoard

Playground

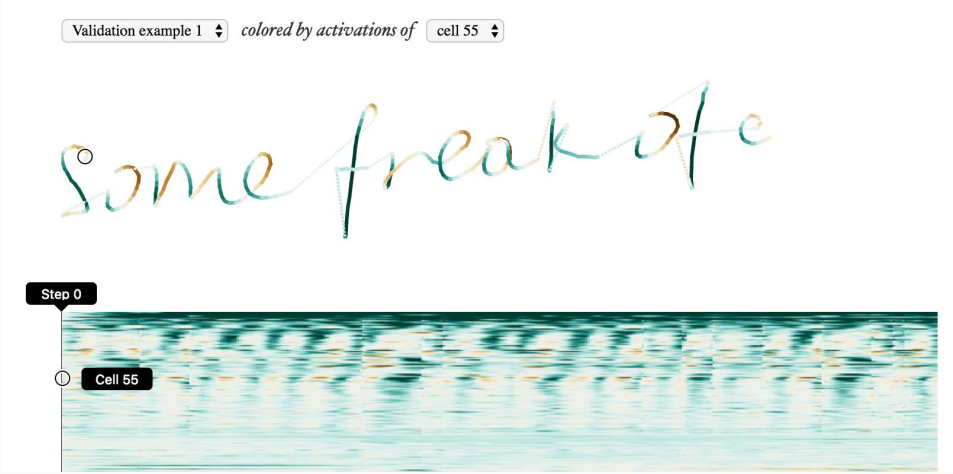


[T]

TensorBoard

TensorBoard plugins to visualize specific datasets

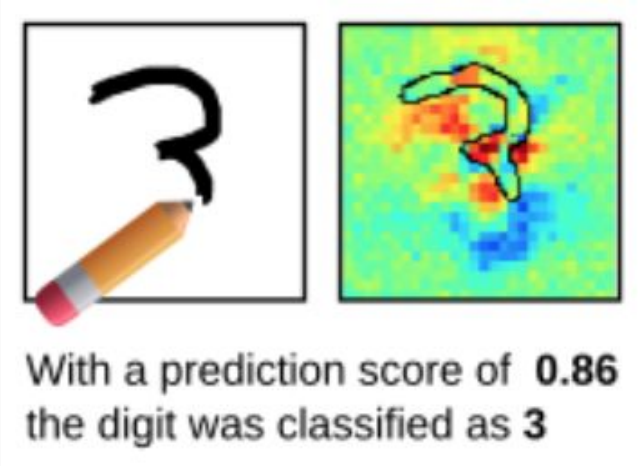
- Eg, handwriting recognition - https://distill.pub/2016/handwriting/



Notable Projects

Heatmapping.org demo - MNIST

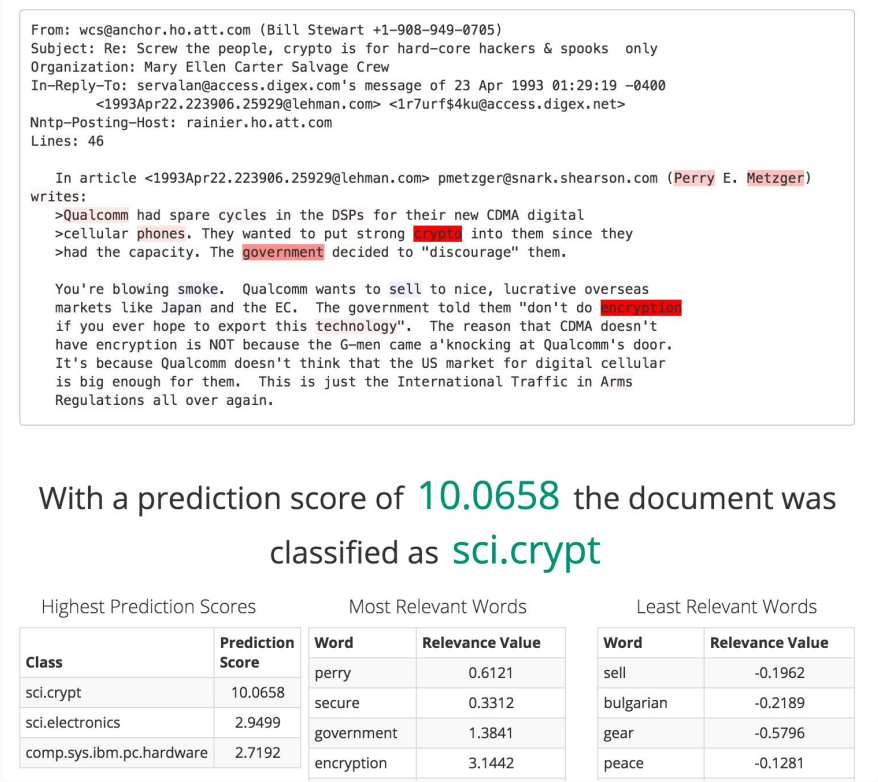
Layer-wise relevance propagation toolbox: https://github.com/sebastian-lapuschkin/lrp_toolbox



[H]

Notable Projects

Heatmapping.org demo - text

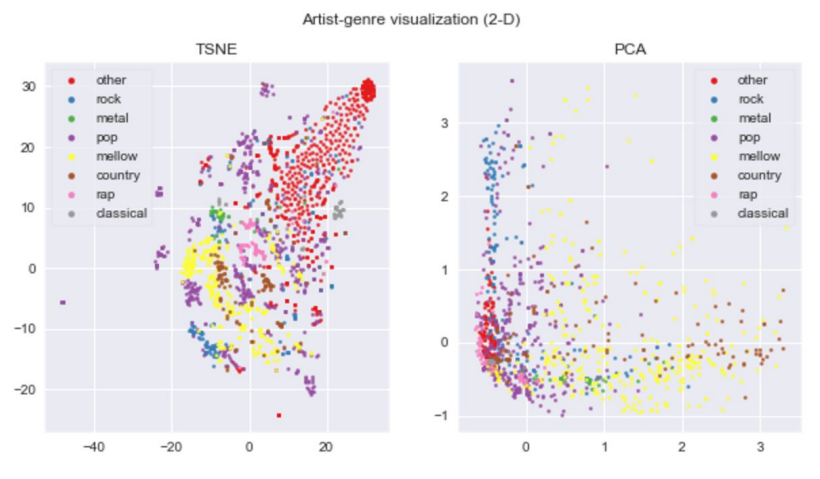


[H]

Notable Projects

- University of Osnabrück Deep Neural Network Visualization (Development Stage) - https://github.com/Petr-By/qtpyvis
- “How to use t-SNE effectively” - https://distill.pub/2016/misread-tsne/
- Eli5 (explanations)
- Lime (explanations)

t-SNE with Genre-based Melody Generation
https://justinshenk.github.io/posts/2017/07/deep-genre/



Summary

Explore deep neural network features and activity using open source tools (in particular, Python)

Discover latent encodings and bases in models

Create tools to guide data scientists to the right model

Scientific research with social impact

Image Sources

[A] - https://developer.apple.com/library/ios/documentation/Performance/Conceptual/vImage/ ConvolutionOperations/ConvolutionOperations.html

[E] - Ersatz Labs, Inc Website

[H] - heatmapping.org

[K] - Andrej Karpathy, https://karpathy.github.io/2015/05/21/rnn-effectiveness/

[T] - TensorFlow/TensorBoard website

[Y] - Jason Yosinski, Deep Visualization Toolbox, YouTube: https://www.youtube.com/watch?v=AgkfIQ4IGaM

Questions

Slides: https://github.com/JustinShenk/pydata

To receive invitations for Intel Machine Learning and Deep Learning webinars, news and tools register in the link below:

bit.ly/Warsawmeetup

$$ - Free Amazon vouchers for posting a project on DevMesh Europe:

https://devmesh.intel.com/groups/447