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

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
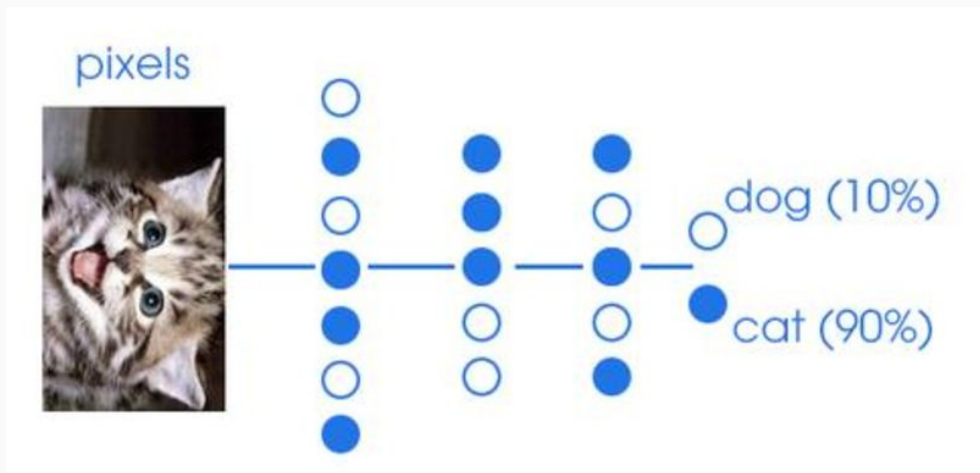
- End-users
- Model developers

Guide architecture selection

- Increase performance

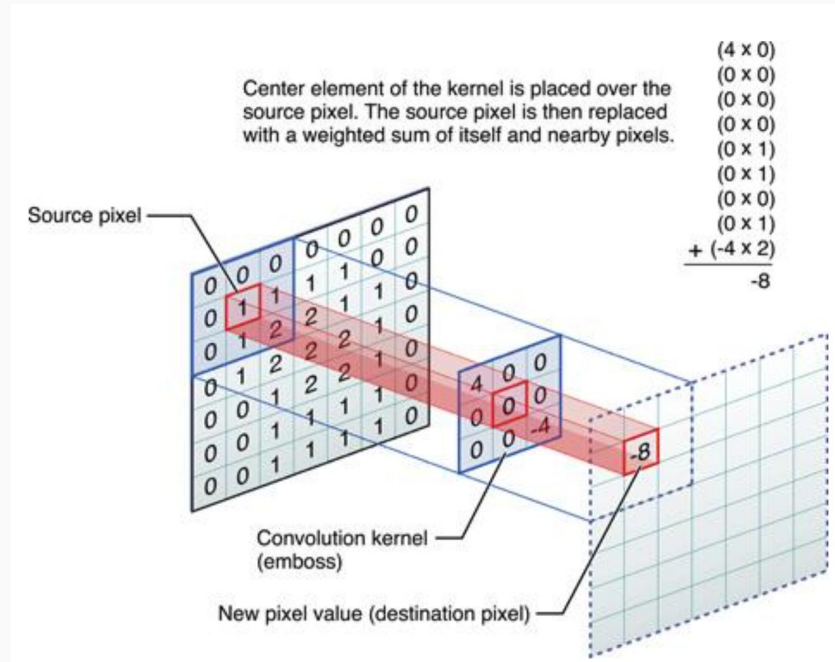explain

"What I cannot ~~create~~, I do not understand."
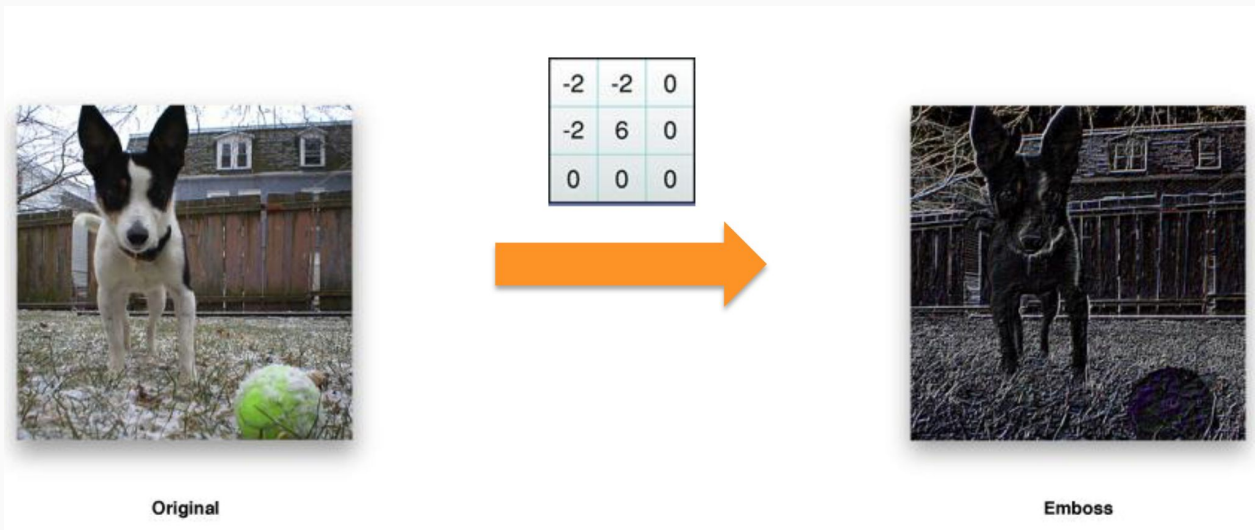- Richard Feynman

[E]

Convolutional Neural Networks

Introduction to CNN

[A]

Convolutional Neural Networks

Feature map

Original   Emboss

[A]

Convolutional Neural Networks

Deep Visualization Toolbox

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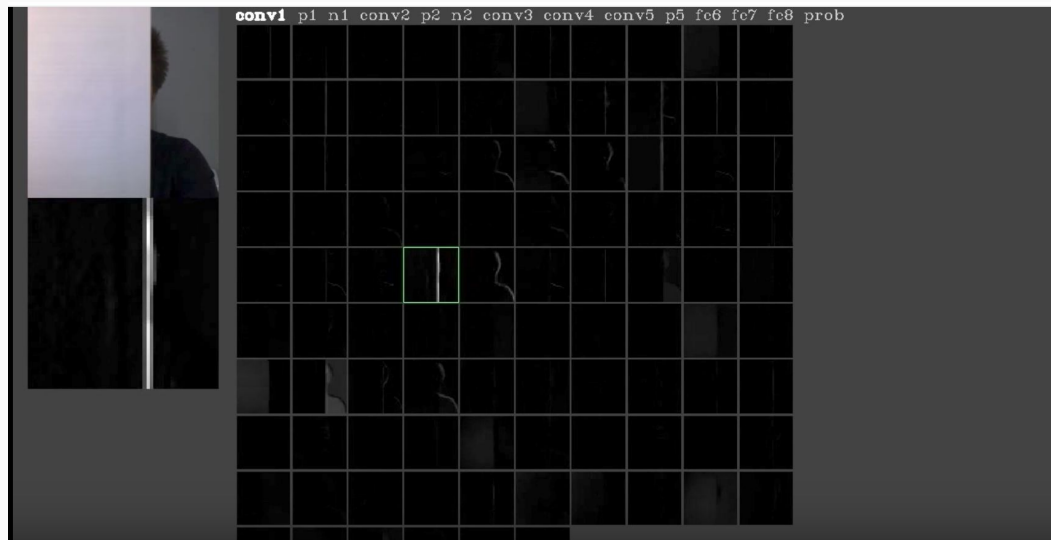
- Synthesized images: Yosinksi et al, Understanding Neural Networks Through Deep Visualization - http://yosinski.com/deepvis
- Important pixels via deconv: Zeiler and Fergus
- AlexNet in Caffe

Convolutional Neural Networks

Low level feature activation
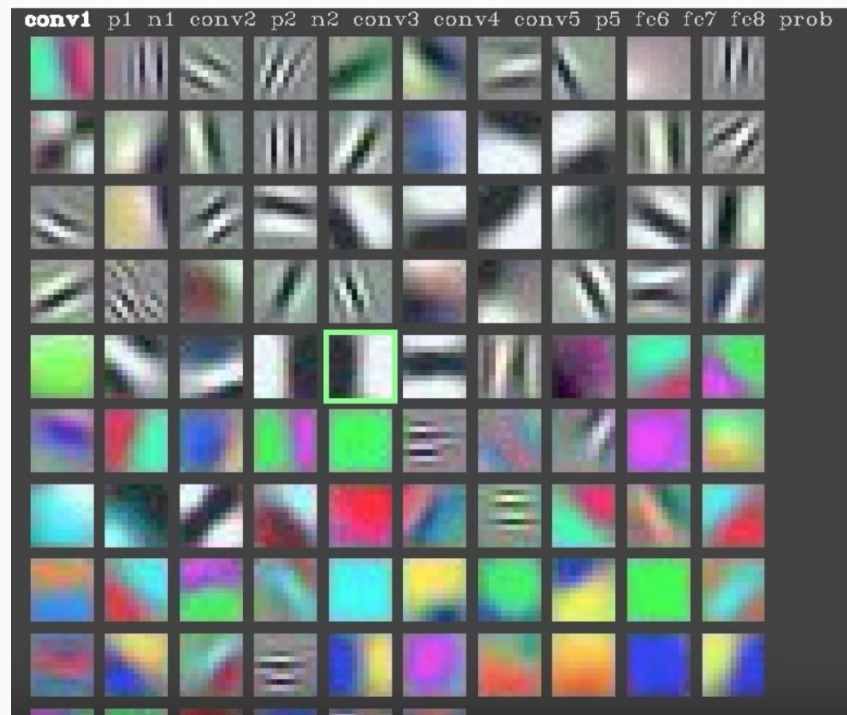
Unit responds to light-dark edges

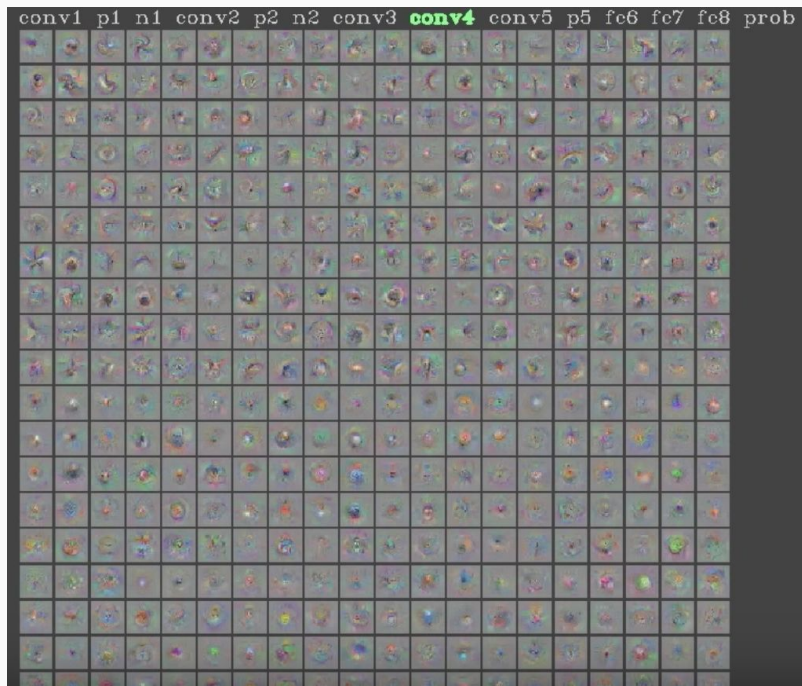

[Y]

Convolutional Neural Networks

Feature activation and images synthesized to produce high unit activation

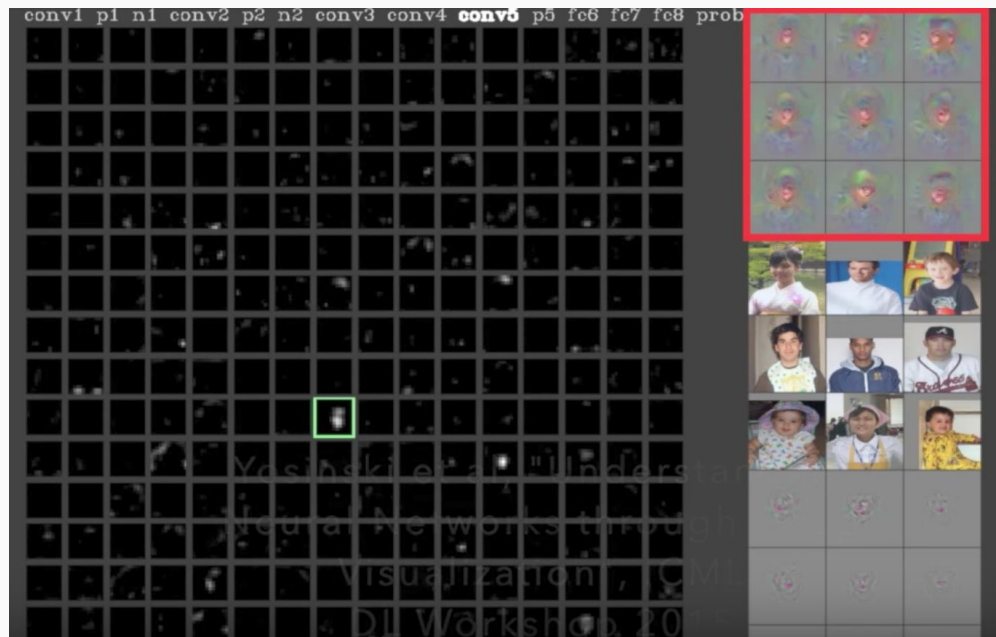[Y]

Convolutional Neural Networks

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

[Y]

Convolutional Neural Networks
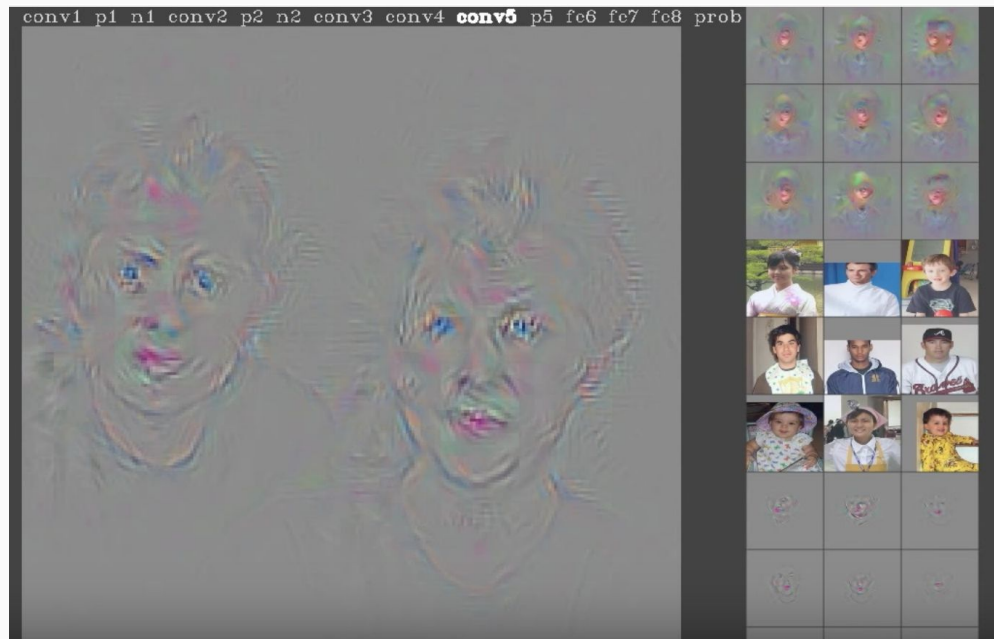
5th layer - abstract concepts
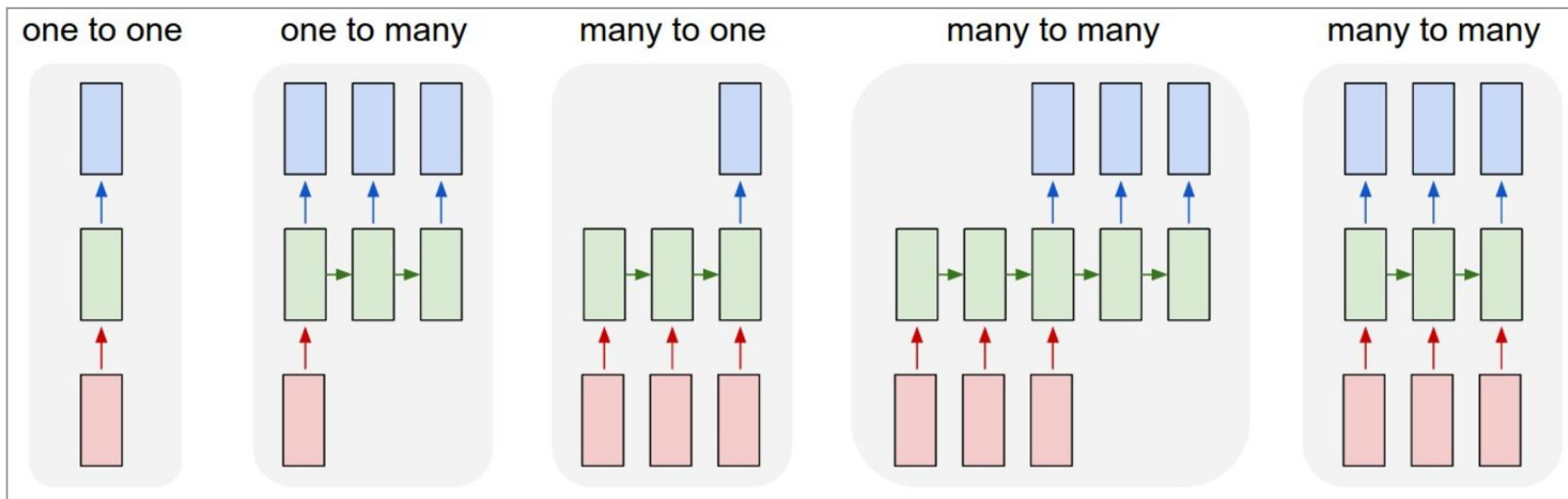
- Face & shoulders unit

[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]
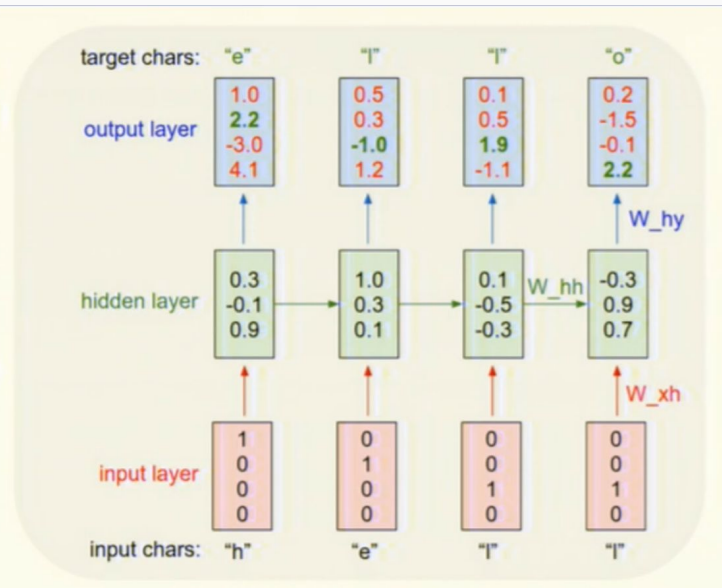
Recurrent Neural Networks

Carry memory forward:

$$\mathbf{h}_t = \phi\left(W\mathbf{x}_t + U\mathbf{h}_{t-1}\right)$$
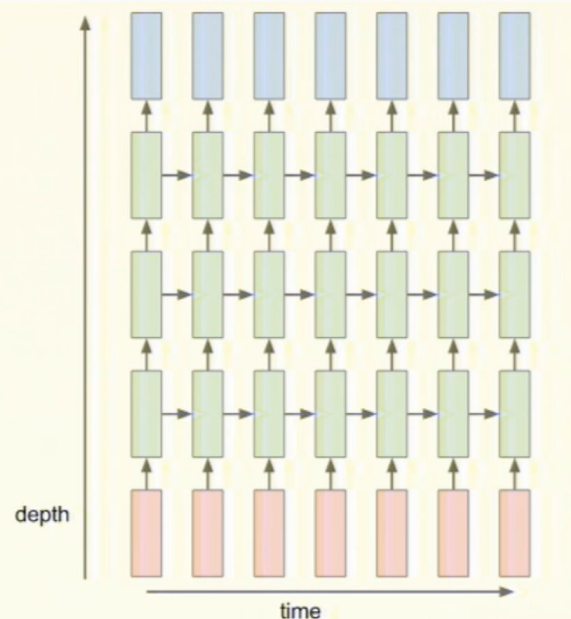
W is weight matrix
x_t is input
h_{t-1} is hidden layer at time = t-1
U is transition matrix

RNN:

$$h_t^l = \tanh W^l \begin{pmatrix} h_t^{l-1} \\ h_{t-1}^l \end{pmatrix}$$

$h \in \mathbb{R}^n \qquad W^l \ [n \times 2n]$

depth

time

[K]

Recurrent Neural Networks

Cell activation visualization

[K]

TensorBoard

[TensorBoard](#):

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

Insight from from weight variance
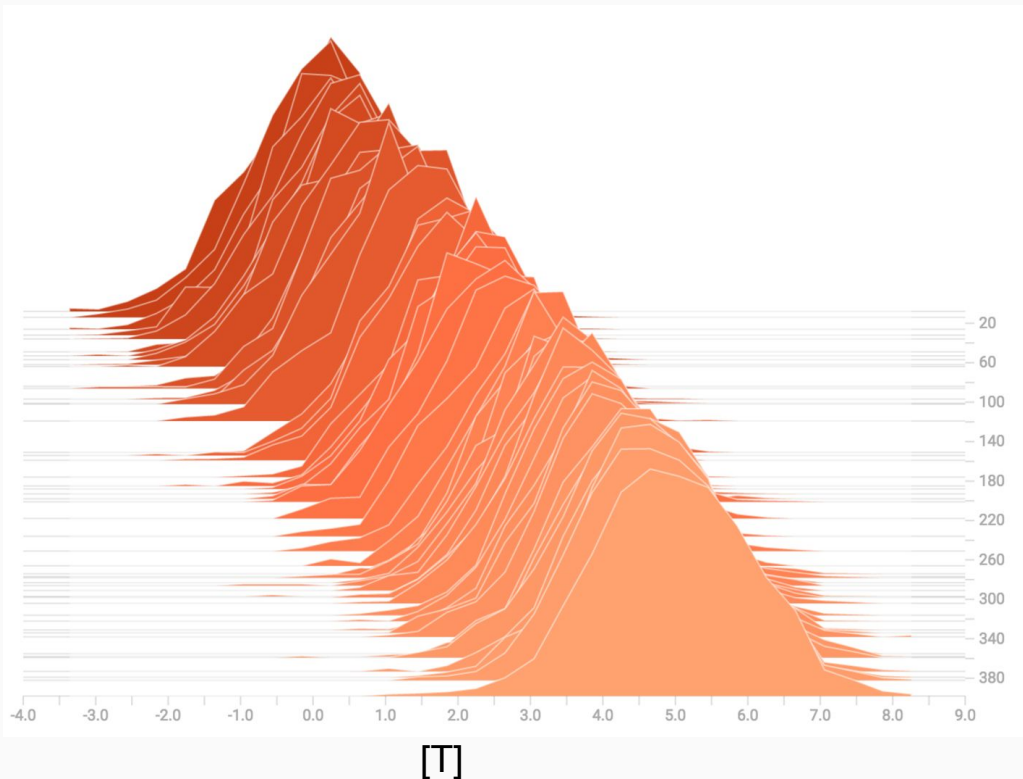
TensorBoard

Graph visualization
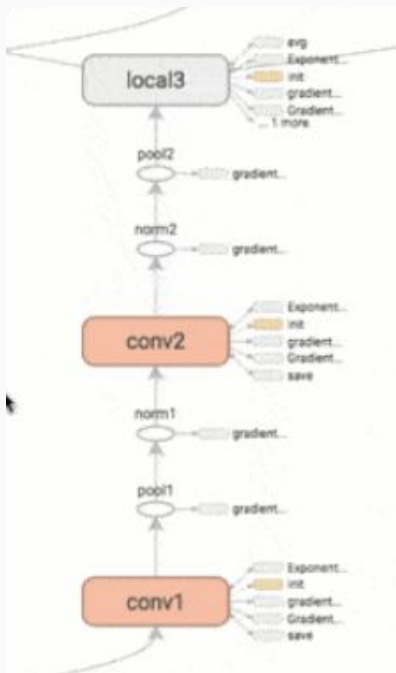
[T]

TensorBoard

[Embedding visualization](#) using PCA and t-SNE

[T]
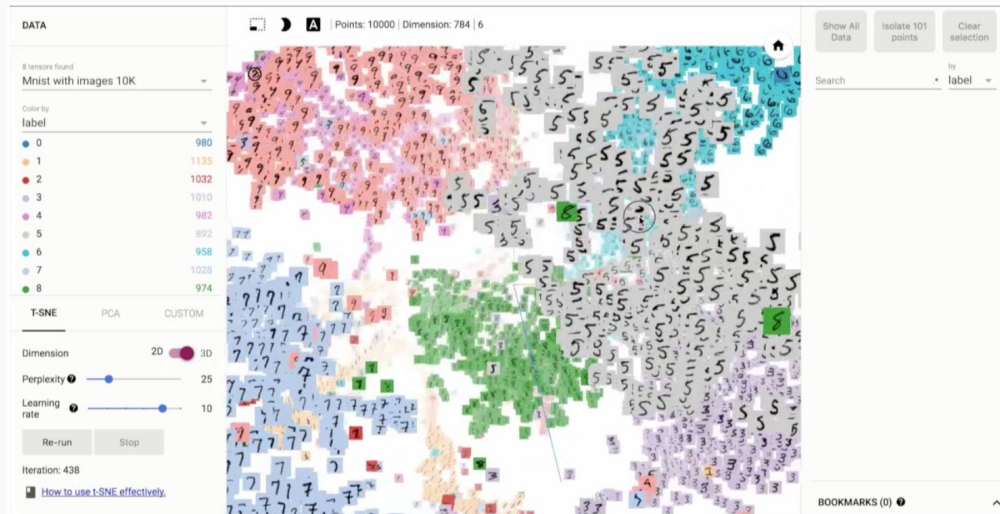
TensorBoard

Playground

[T]
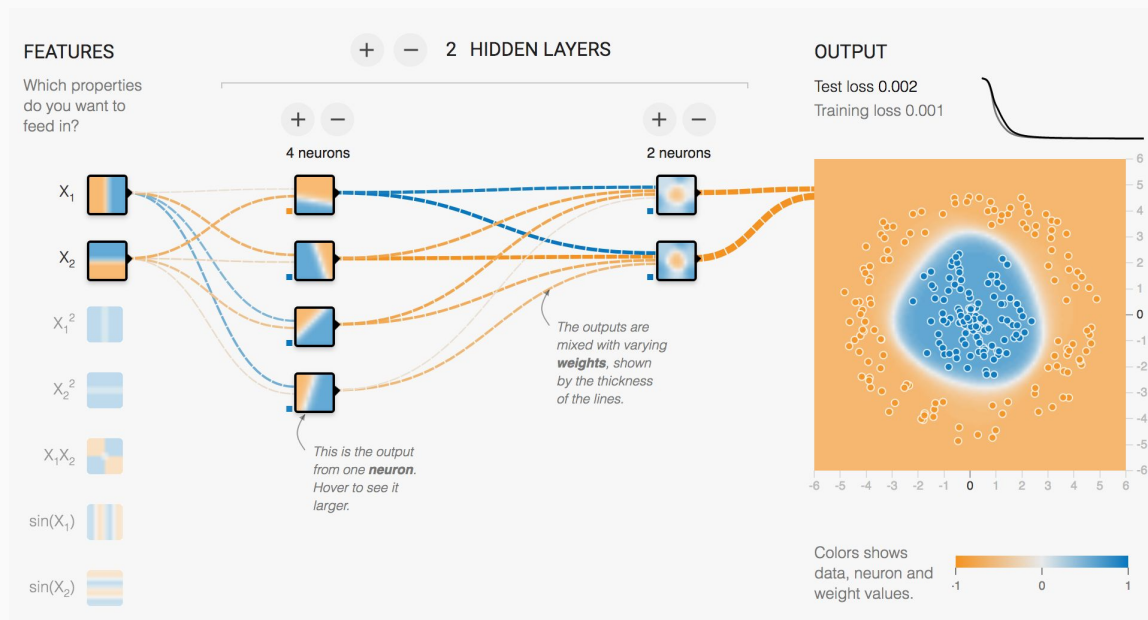
TensorBoard

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

With a prediction score of **0.86** the digit was classified as **3**
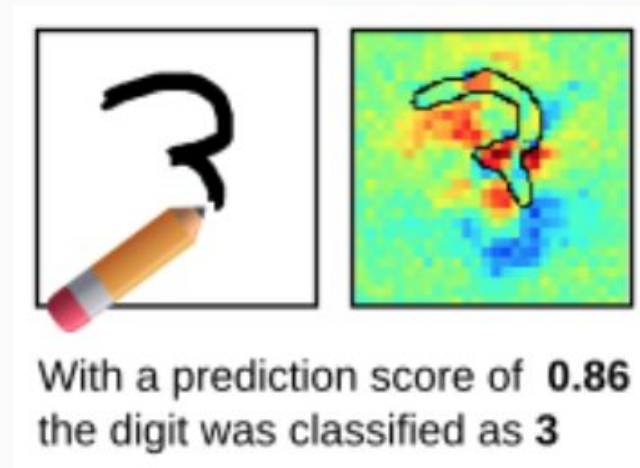
[H]

Notable Projects

Heatmapping.org demo - text


```
From: wcs@anchor.ho.att.com (Bill Stewart +1-908-949-0705)
Subject: Re: Screw the people, crypto is for hard-core hackers & spooks  only
Organization: Mary Ellen Carter Salvage Crew
In-Reply-To: servalan@access.digex.com's message of 23 Apr 1993 01:29:19 -0400
        <1993Apr22.223906.25929@lehman.com> <1r7urf$4ku@access.digex.net>
Nntp-Posting-Host: rainier.ho.att.com
Lines: 46

    In article <1993Apr22.223906.25929@lehman.com> pmetzger@snark.shearson.com (Perry E. Metzger)
writes:
    >Qualcomm had spare cycles in the DSPs for their new CDMA digital
    >cellular phones. They wanted to put strong crypto into them since they
    >had the capacity. The government decided to "discourage" them.

    You're blowing smoke.  Qualcomm wants to sell to nice, lucrative overseas
markets like Japan and the EC.  The government told them "don't do encryption
if you ever hope to export this technology".  The reason that CDMA doesn't
have encryption is NOT because the G-men came a'knocking at Qualcomm's door.
It's because Qualcomm doesn't think that the US market for digital cellular
is big enough for them.  This is just the International Traffic in Arms
Regulations all over again.
```

[H]

With a prediction score of **10.0658** the document was classified as **sci.crypt**

Highest Prediction Scores			Most Relevant Words			Least Relevant Words	
Class	**Prediction Score**		**Word**	**Relevance Value**		**Word**	**Relevance Value**
sci.crypt	10.0658		perry	0.6121		sell	-0.1962
sci.electronics	2.9499		secure	0.3312		bulgarian	-0.2189
comp.sys.ibm.pc.hardware	2.7192		government	1.3841		gear	-0.5796
			encryption	3.1442		peace	-0.1281

Notable Projects

- deeplearn.js - https://github.com/PAIR-code/deeplearnjs
- 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/

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

Artist-genre visualization (2-D)
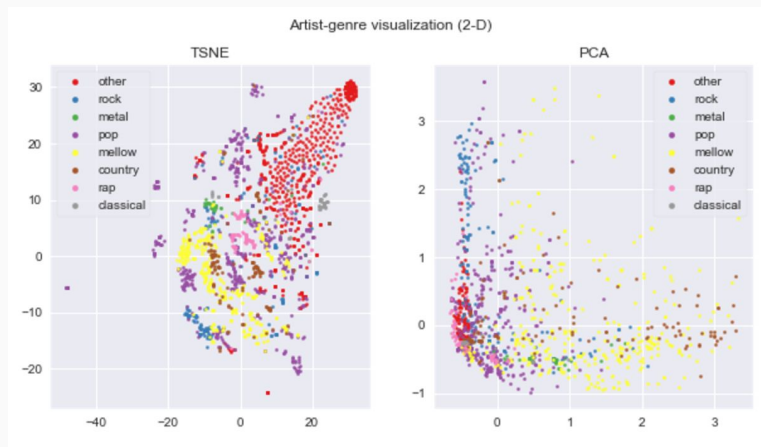
Summary

Explore deep neural network features and activity using open source tools (particularly Python)

Discover latent encodings and bases in models

Promising area of 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

Contact me:

    email: shenk.justin@gmail.com

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