

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

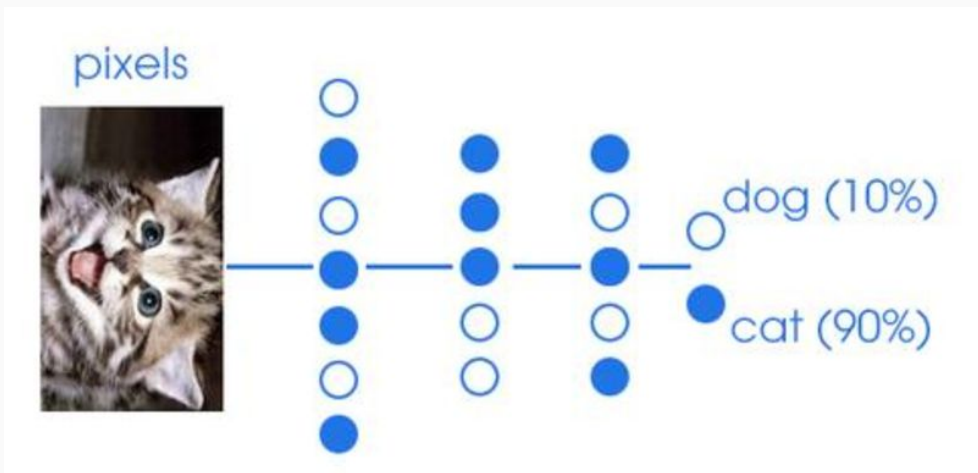
Guide architecture selection

- Increase performance

explain

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

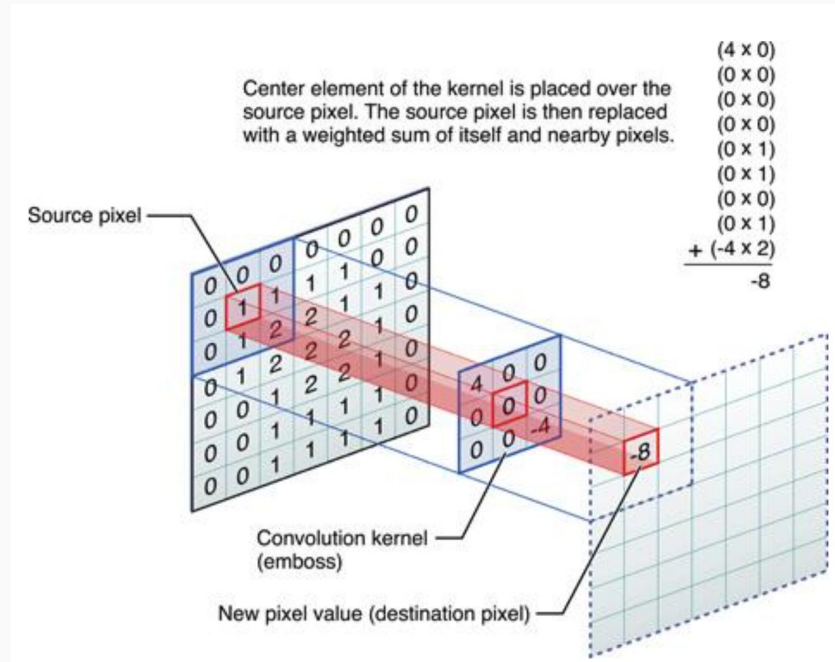
- Richard Feynman



[E]

Convolutional Neural Networks

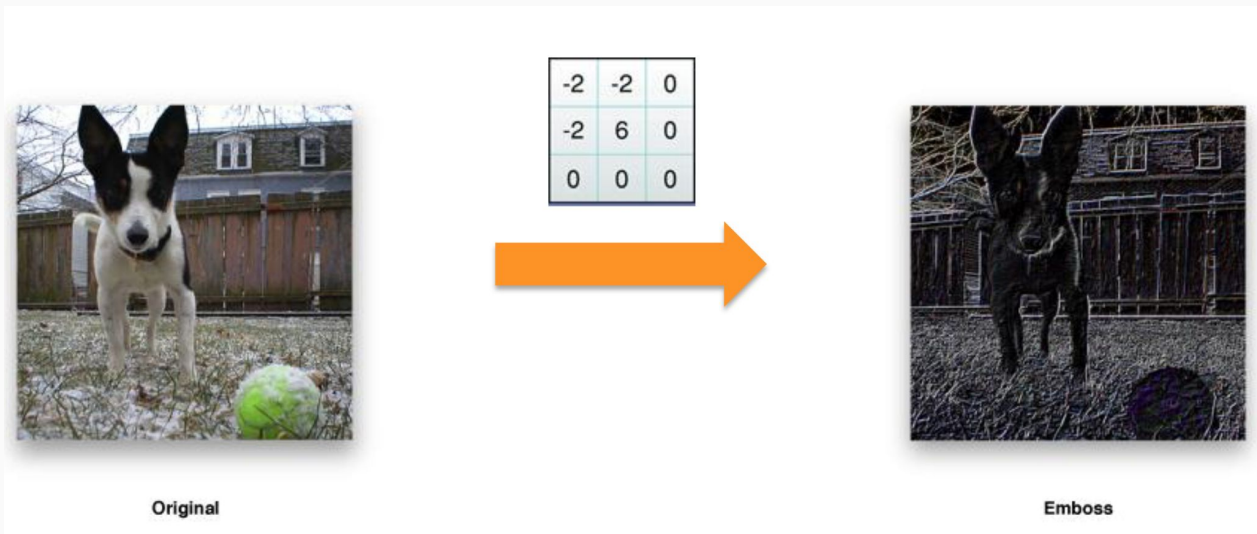
Introduction to CNN



[A]

Convolutional Neural Networks

Feature map



[A]

Convolutional Neural Networks

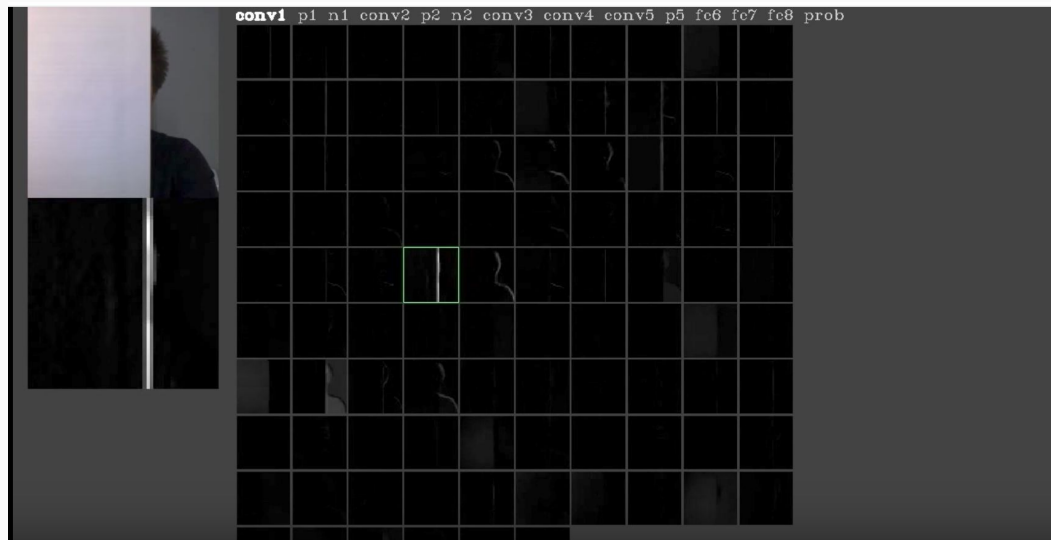
Deep Visualization Toolbox

- <https://github.com/yosinski/deep-visualization-toolbox>
- Synthesized images: Yosinski 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

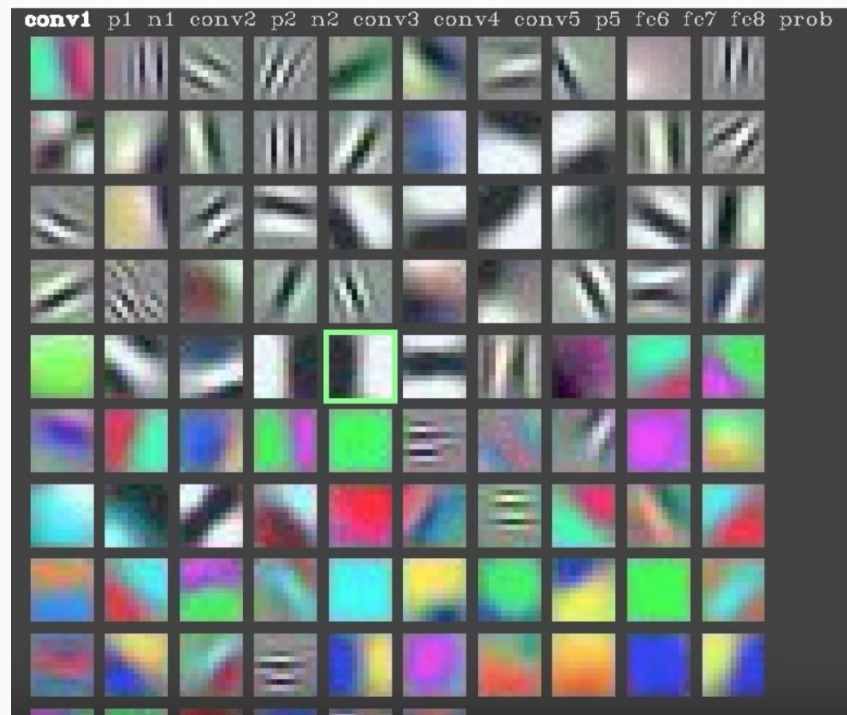
Unit responds to light-dark
edges



[Y]

Convolutional Neural Networks

Feature activation and
images synthesized to
produce high unit activation



[Y]

Convolutional Neural Networks

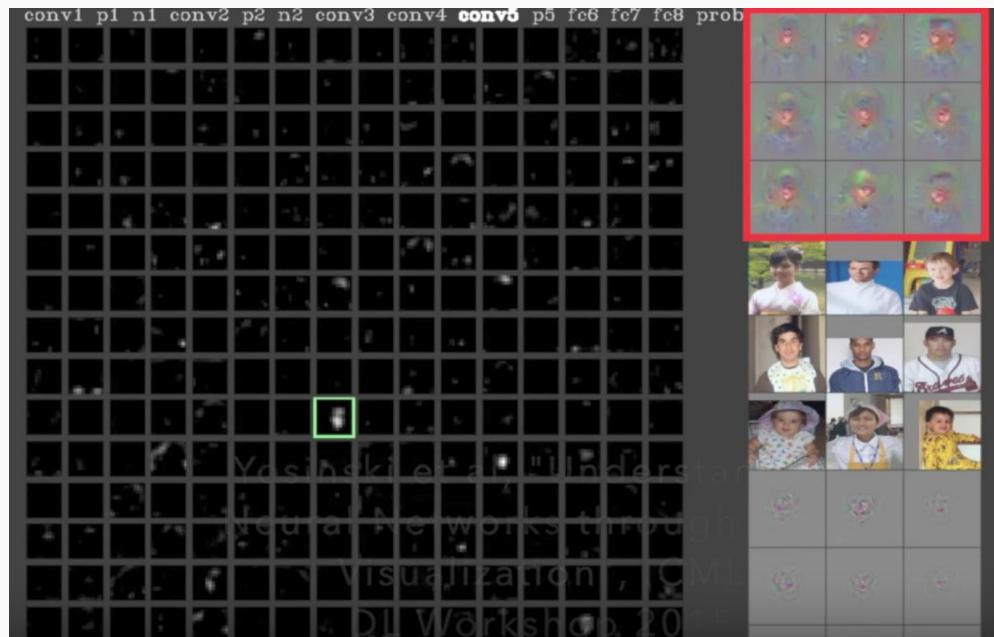
Switching between unit activations
and images synthesized to produce
high activation via backpropagation



Convolutional Neural Networks

5th layer - abstract concepts

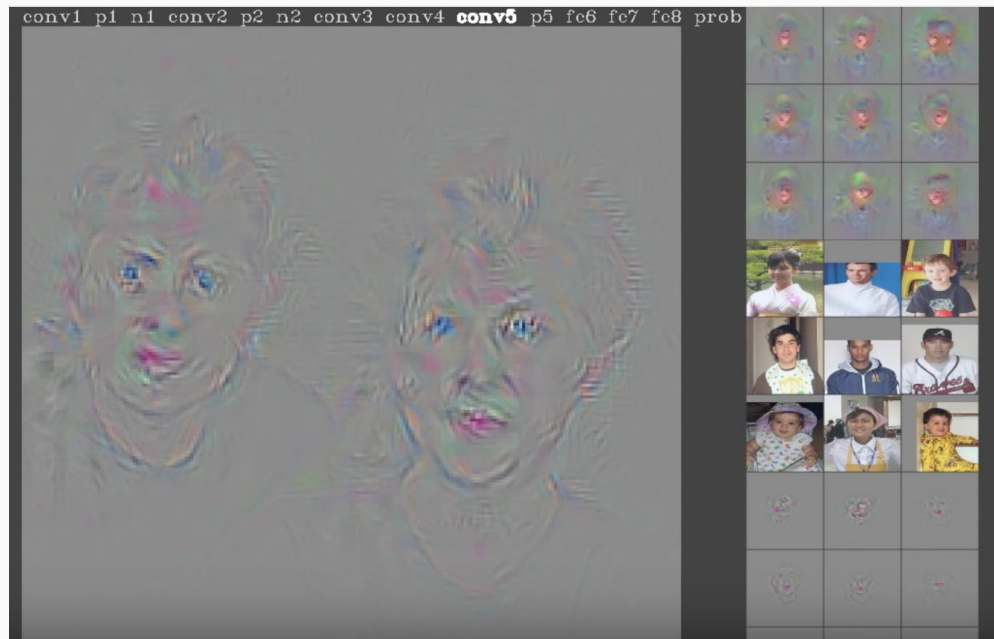
- Face & shoulders unit



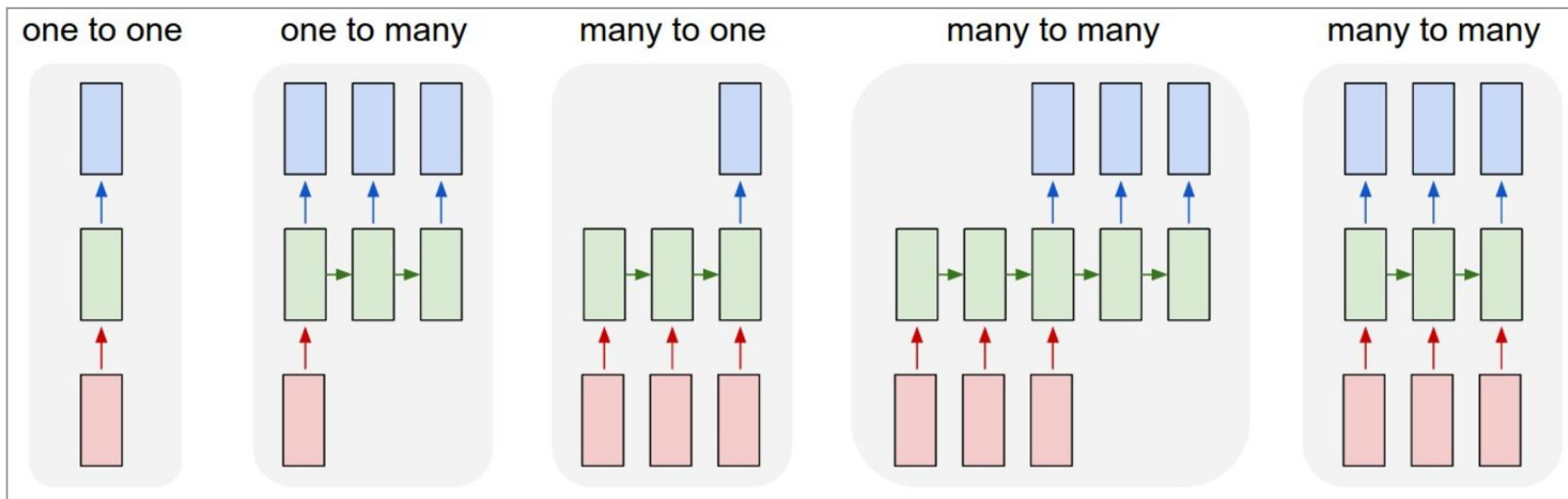
Convolutional Neural Networks

5th layer - abstract concepts

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



Recurrent Neural Networks



[K]

Recurrent Neural Networks

*	'	'	[[J	e	r	u	s	a	l	e	m	R	e	p	o	r	t]]	'	'	
*	'	'	[h	T	o	a	u	s	a	l	e	m	a	o	g	u	r	t]]	'	'	(
	['	[C	a	s	s	m	e	n	e]	B	e	a	o	n	d	s		s	a
	'	s		m	F	u	r	n	i	s		i	a	e	t	a	l	i	s	a		'	:	:
:	*	:	A	q	D	e	n	e	b	i	u	t	n		C	i	p	r	e		e		,	
#	T	&	T	f	S	i	w	r	p		e]]	a	l	u	v	e	l	r	u	,	s	:

[K]

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

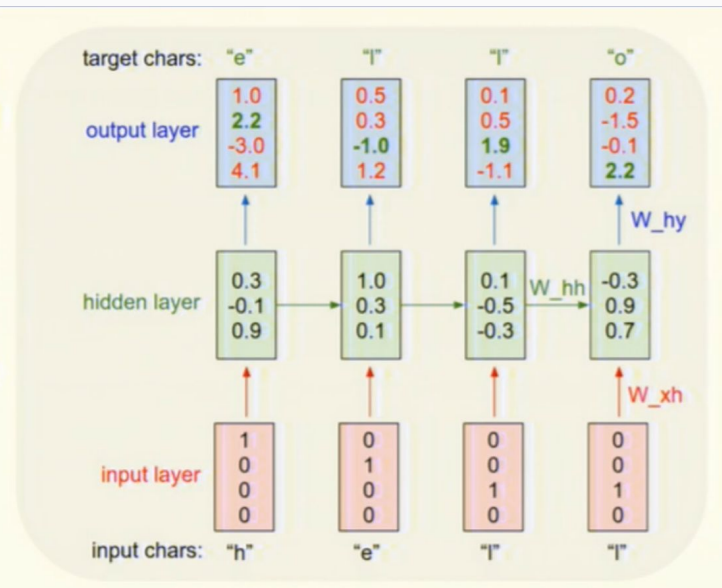
Recurrent Neural Networks

Make predictions based on current state and previous state

Character-level language model example

Vocabulary:
[h,e,l,o]

Example training sequence:
"hello"



Recurrent Neural Networks

Carry memory forward:

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

W is weight matrix

\mathbf{x}_t is input

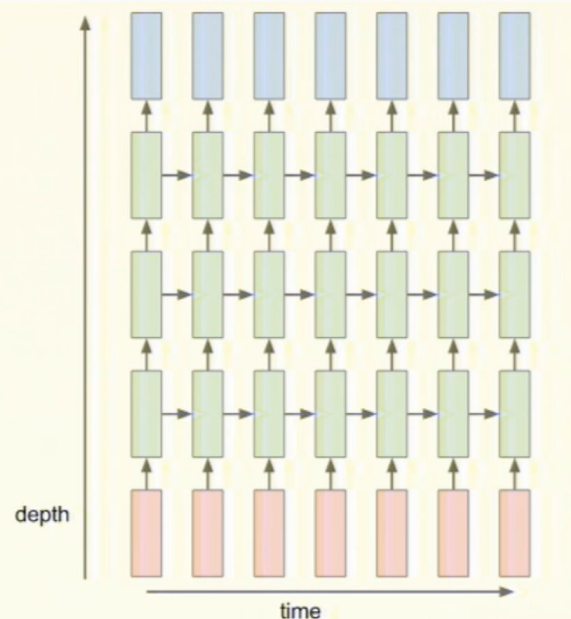
\mathbf{h}_{t-1} is hidden layer at time = $t-1$

U is transition matrix

RNN:

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

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



[K]

Recurrent Neural Networks

Cell activation visualization

Cell sensitive to position in line:

The sole importance of the crossing of the Berezina lies in the fact that it plainly and indubitably proved the fallacy of all the plans for cutting off the enemy's retreat and the soundness of the only possible line of action--the one Kutuzov and the general mass of the army demanded--namely, simply to follow the enemy up. The French crowd fled at a continually increasing speed and all its energy was directed to reaching its goal. It fled like a wounded animal and it was impossible to block its path. This was shown not so much by the arrangements it made for crossing as by what took place at the bridges. When the bridges broke down, unarmed soldiers, people from Moscow and women with children who were with the French transport, all--carried on by vis inertiae--pressed forward into boats and into the ice-covered water and did not, surrender.

Cell that turns on inside quotes:

"You mean to imply that I have nothing to eat out of.... On the contrary, I can supply you with everything even if you want to give dinner parties," warmly replied Chichagov, who tried by every word he spoke to prove his own rectitude and therefore imagined Kutuzov to be animated by the same desire.

Kutuzov, shrugging his shoulders, replied with his subtle penetrating smile: "I meant merely to say what I said."

Cell that robustly activates inside if statements:

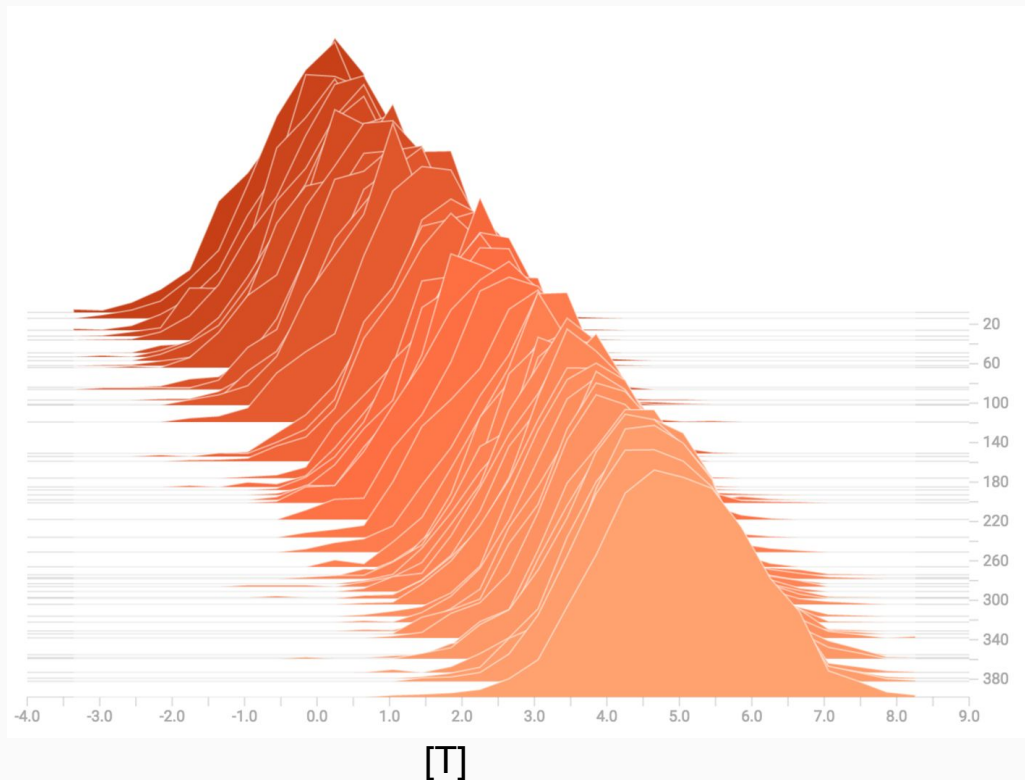
```
static int __dequeue_signal(struct sigpending *pending, sigset_t *mask,
                           sinfo_t *info)
{
    int sig = next_signal(pending, mask);
    if (sig) {
        if (current->notifier) {
            if (sigismember(current->notifier_mask, sig)) {
                if (!current->notifier)(current->notifier_data)) {
                    clear_thread_flag(TIF_SIGPENDING);
                    return 0;
                }
            }
        }
        collect_signal(sig, pending, info);
    }
    return sig;
}
```

TensorBoard

TensorBoard:

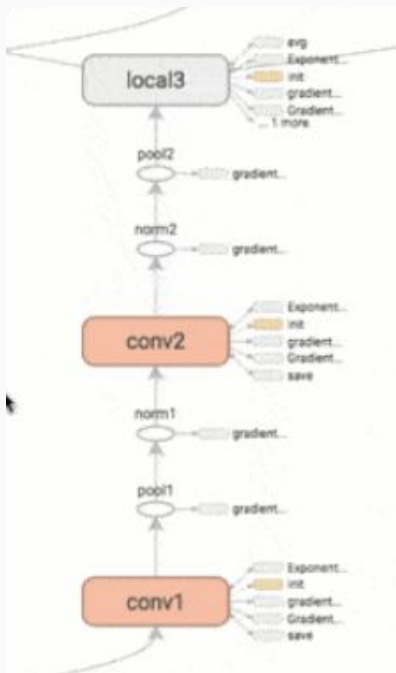
- Weight histogram and accuracy variance during training
- Embedding visualization
- Graph visualization
- Hyperparameter search

Insight from from weight variance



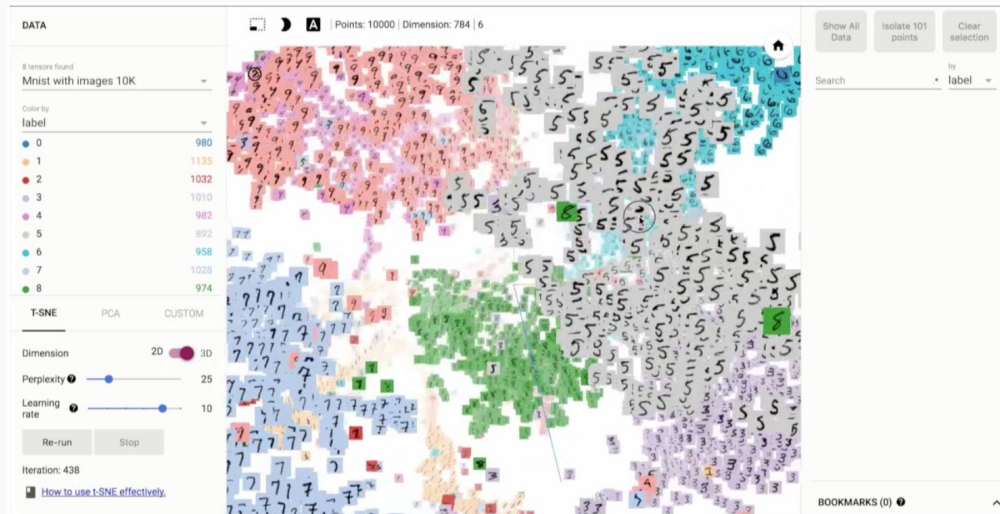
TensorBoard

Graph visualization



TensorBoard

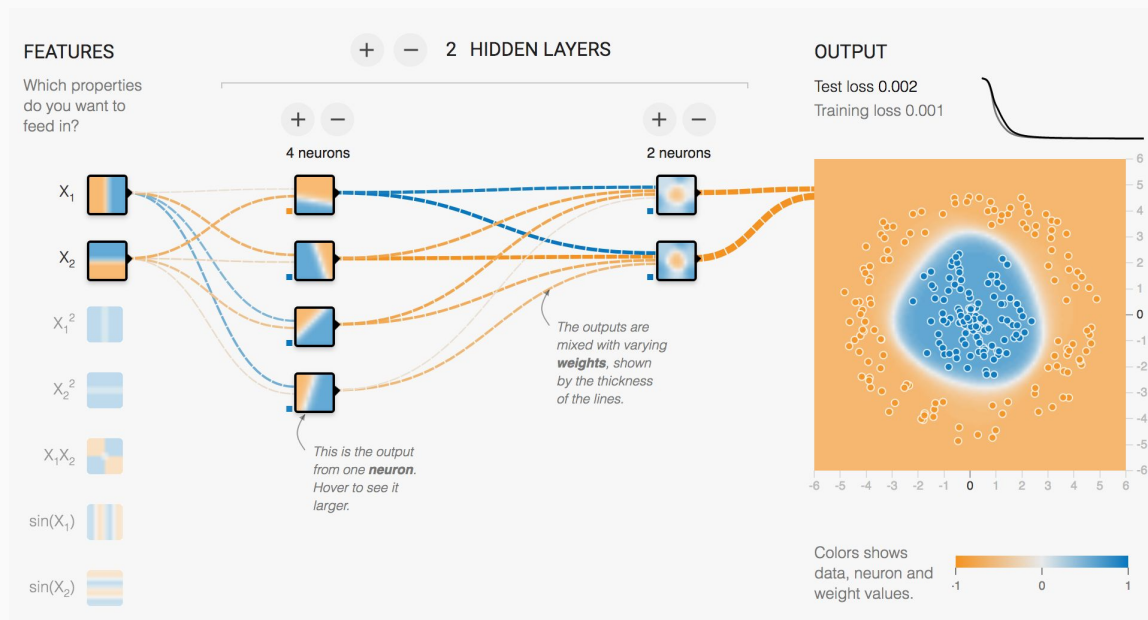
Embedding visualization using
PCA and t-SNE



[1]

TensorBoard

Playground



TensorBoard

TensorBoard plugins to visualize specific datasets

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



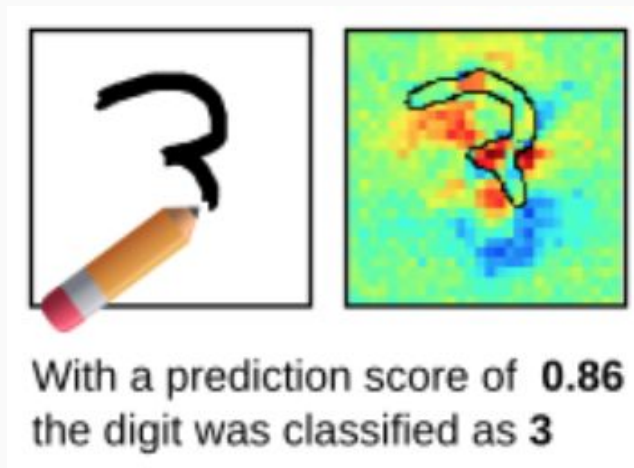
Notable Projects

Heatmapping.org demo - MNIST

Layer-wise relevance

propagation toolbox:

https://github.com/sebastian-lapuschkin/lrp_toolbox



[H]

Notable Projects

Heatmapping.org demo - text

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 **encryption** 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

Class	Prediction Score
sci.crypt	10.0658
sci.electronics	2.9499
comp.sys.ibm.pc.hardware	2.7192

Most Relevant Words

Word	Relevance Value
perry	0.6121
secure	0.3312
government	1.3841
encryption	3.1442

Least Relevant Words

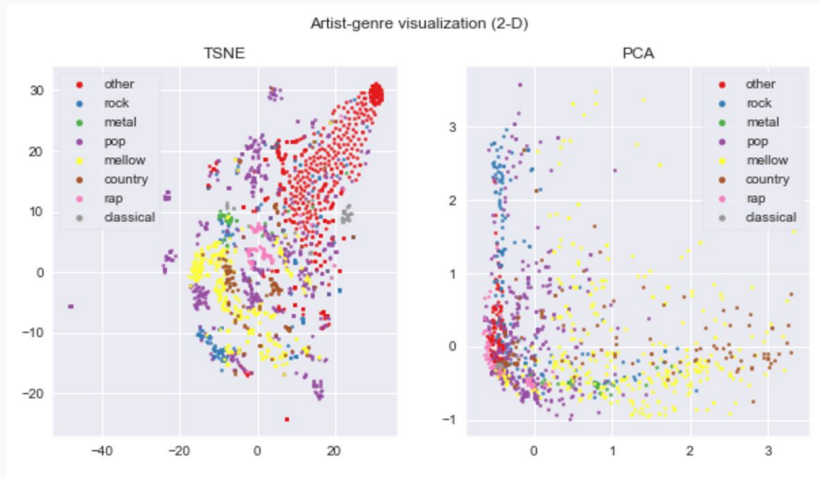
Word	Relevance Value
sell	-0.1962
bulgarian	-0.2189
gear	-0.5796
peace	-0.1281

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>