Convolutional Neural Networks and Natural Language Processing

Thomas Delteil – github.com/thomasdelteil – linkedin.com/in/thomasdelteil

Applied Scientist @ AWS Deep Engine

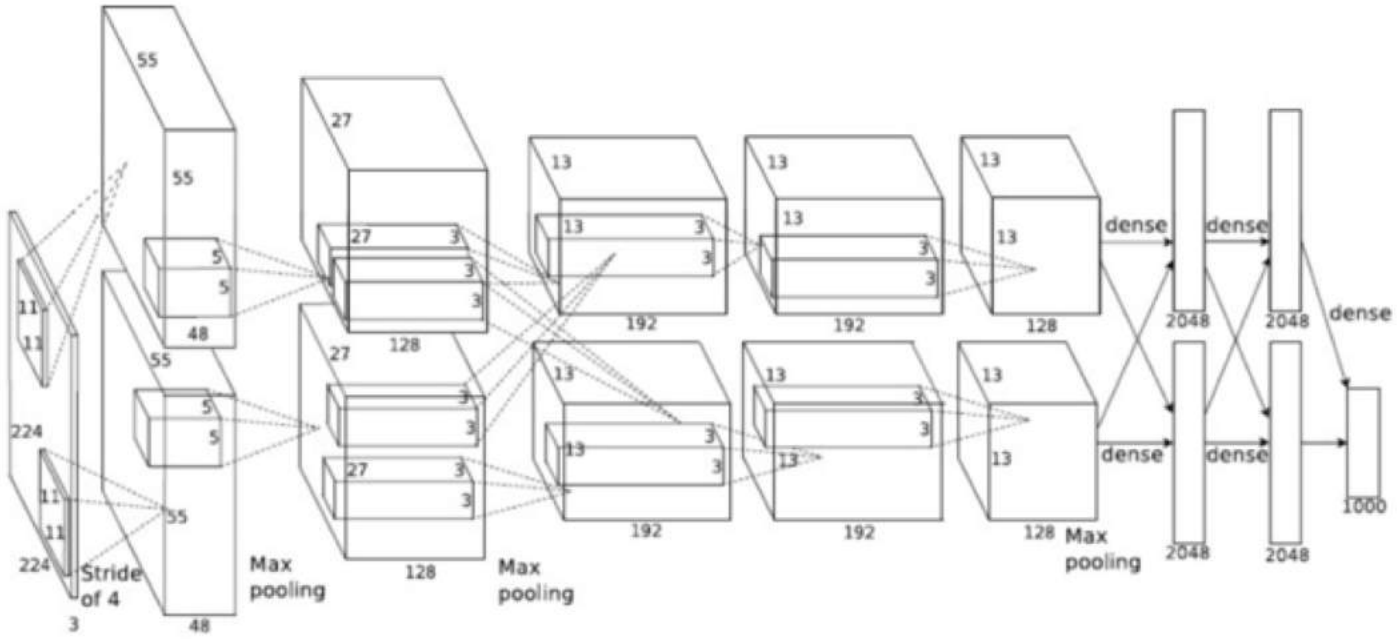
Goals

- Explain what convolutions are
- Show how to handle textual data
- Analyze a reference neural network architecture for text classification
- Demonstrate how to train and deploy a CNN for Natural Language Processing

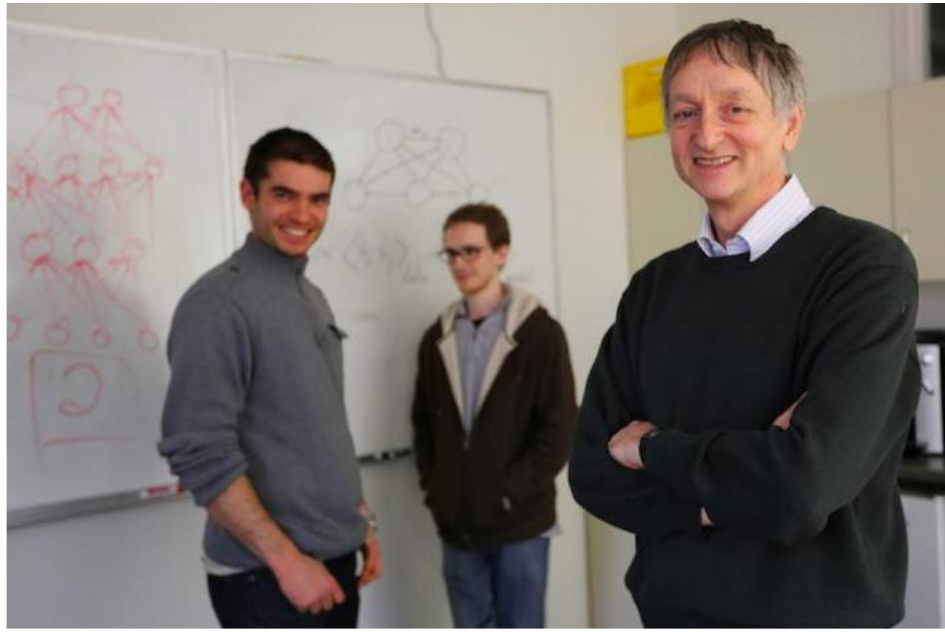
Convolutions

And where to find them

2012 - ImageNet Classification with Deep Convolutional Neural Networks



AlexNet architecture



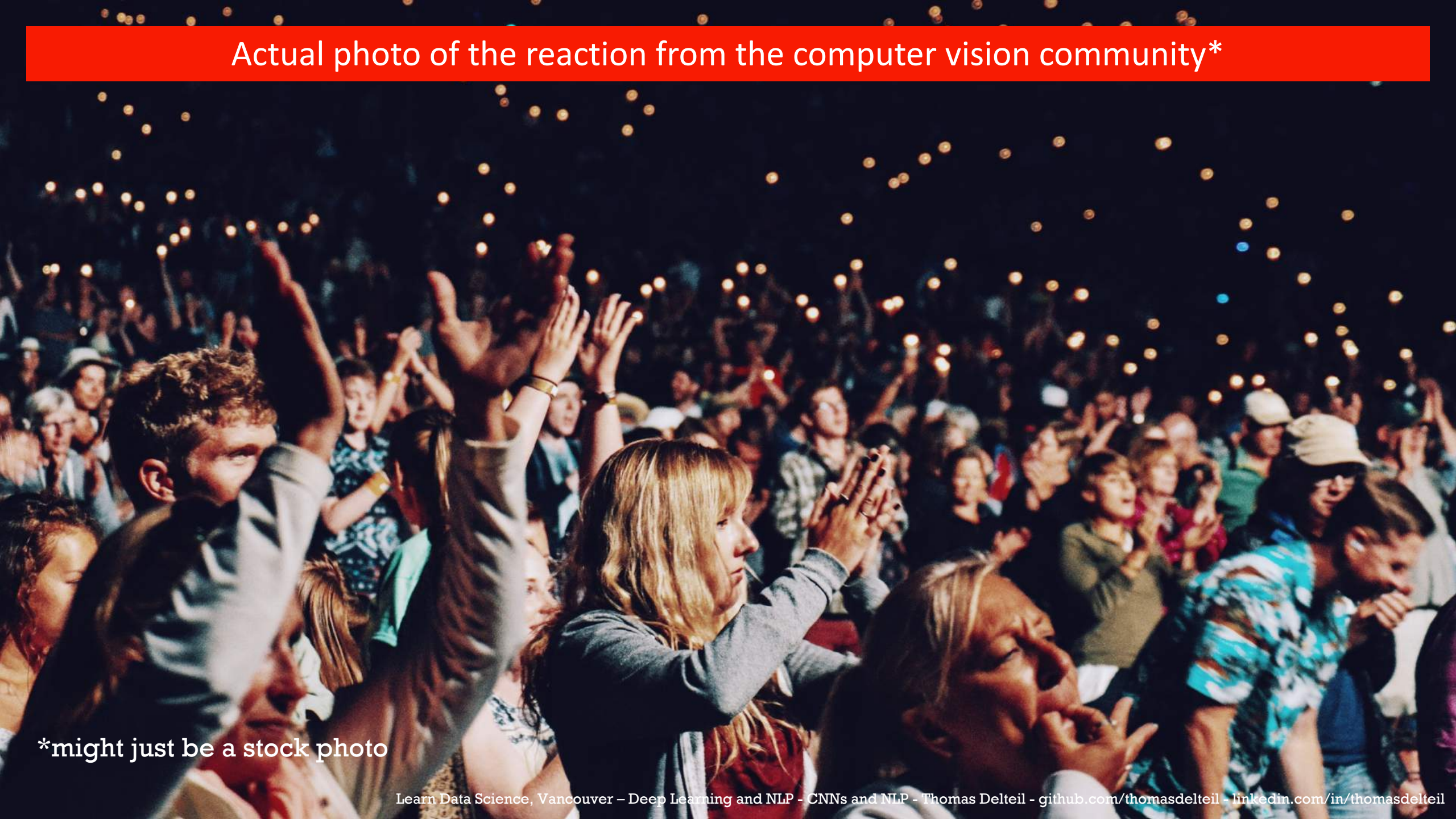
ImageNet classification with Deep Convolutional Neural Networks, Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, Advances in Neural Information Processing Systems, 2012

ImageNet competition

Classify images among 1000 classes:

AlexNet Top-5 error-rate, 25% => **16%**!

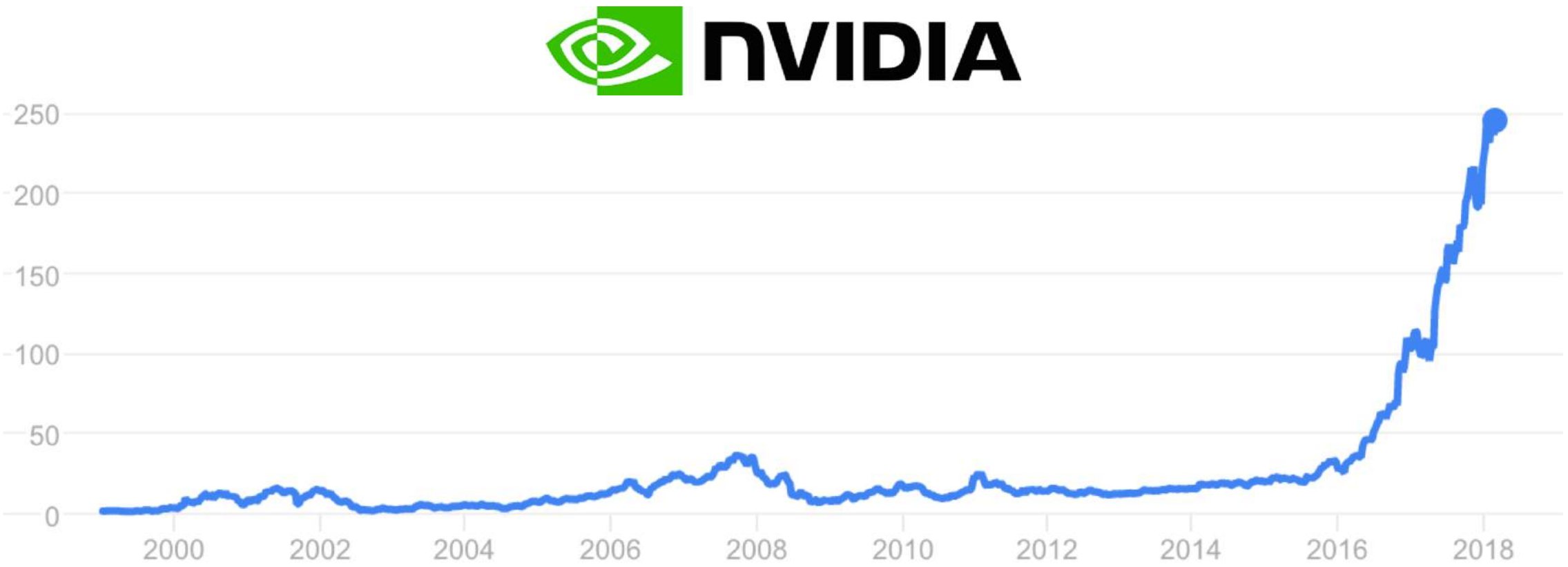
Actual photo of the reaction from the computer vision community*



*might just be a stock photo


I told you so!

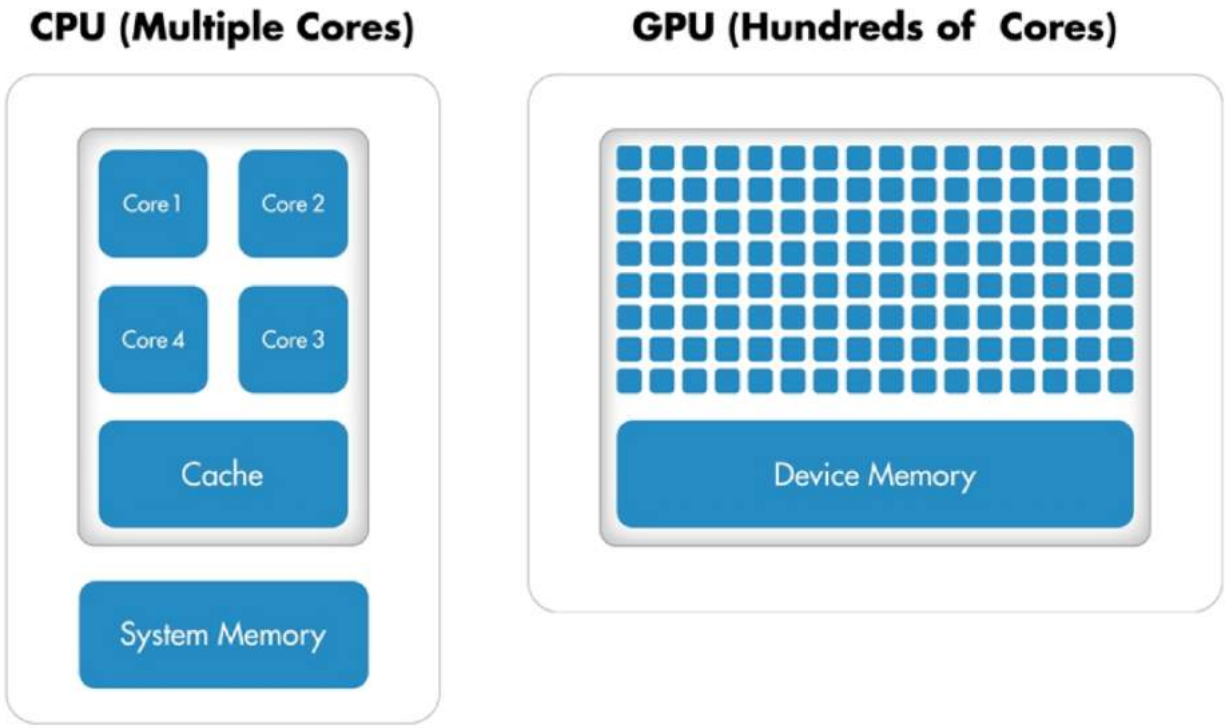
What made Convolutional Neural Networks viable?



GPUs!

- Nvidia V100, float16 Ops:
~ 120 TFLOPS, 5000+ cuda cores

- #1 Super computer 2005 **~135 TFLOPS**

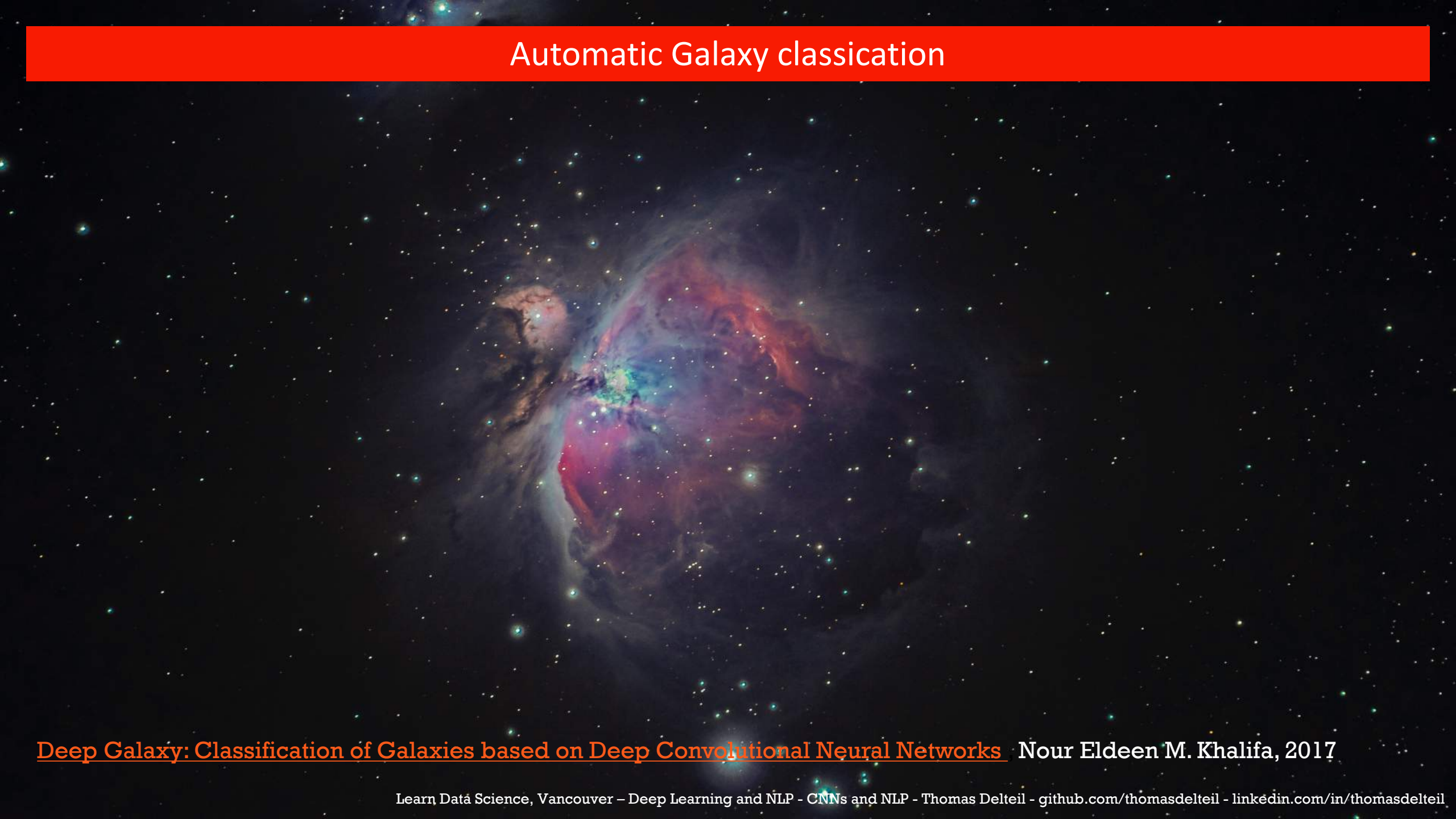


Source: Mathworks

Sea/Land segmentation via satellite images

DeepUNet: A Deep Fully Convolutional Network for Pixel-level Sea-Land Segmentation, Ruirui Li et al, 2017

Automatic Galaxy classication



Deep Galaxy: Classification of Galaxies based on Deep Convolutional Neural Networks Nour Eldeen M. Khalifa, 2017

Medical Imaging, MRI, X-ray, surgical cameras

Review of MRI-based Brain Tumor Image Segmentation Using Deep Learning Methods, Ali Isn et al. 2016

What is a convolution ?

It is the **cross-channel sum** of the **element-wise multiplication** of a convolutional filter (kernel/mask) computed over a sliding window on an input tensor given a certain stride and padding, **plus a bias term**. The result is called a feature map.

1*2 –1*2 –1*3 + 0*1 + 2 = – 1

1*2 –1*2 –1*1 + 0*-1 + 2. = 2

1*3 –1*1 –1*4 + 0*3 + 2 = 0

1*1 – (-1)*1 –1*3 + 0*2 + 2 = 1

2	2	1
3	1	-1
4	3	2

Input matrix (3x3)
no padding
1 channel

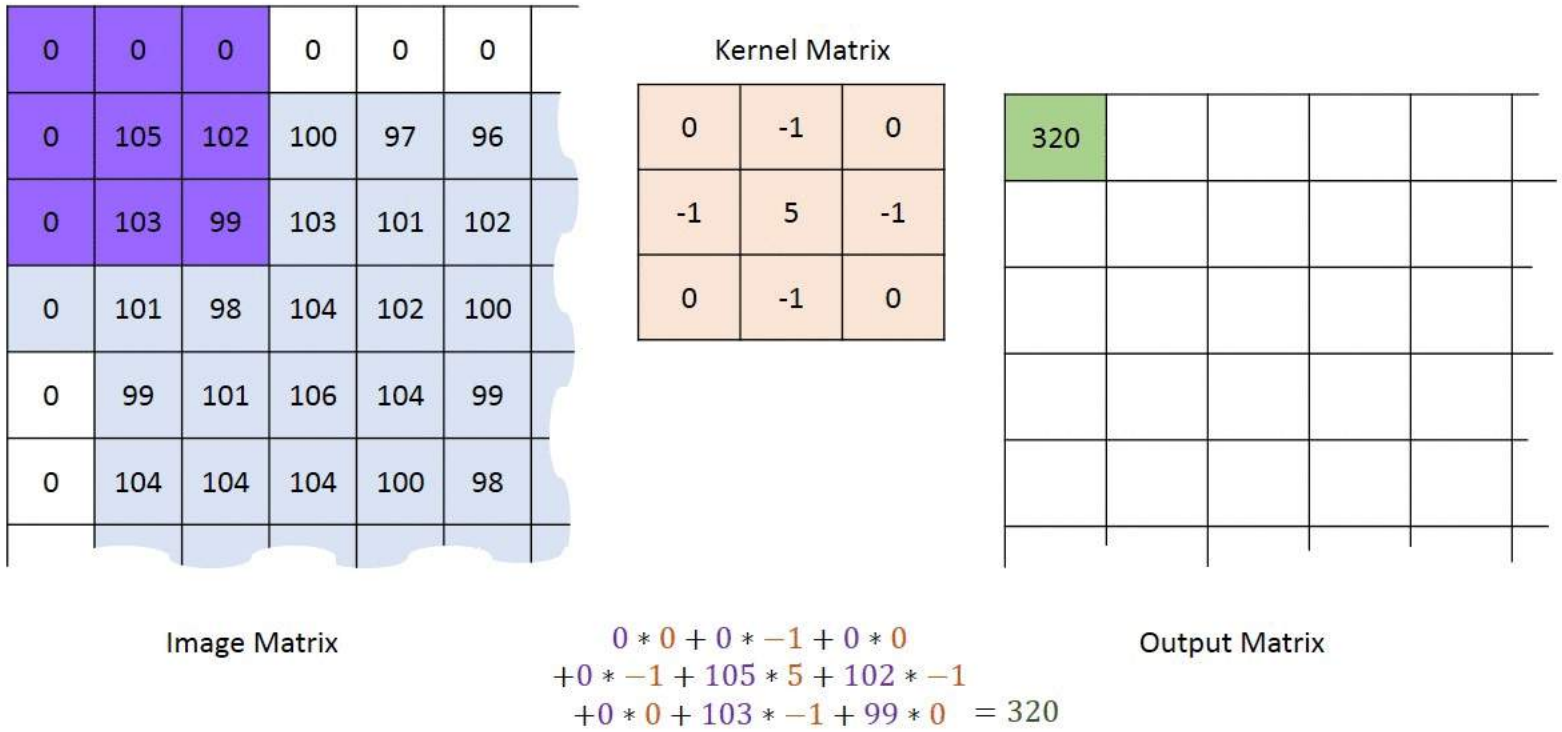
1	-1
-1	0

Kernel (2x2)
Stride 1
Bias = 2

-1	2
0	1

Feature map (2x2)

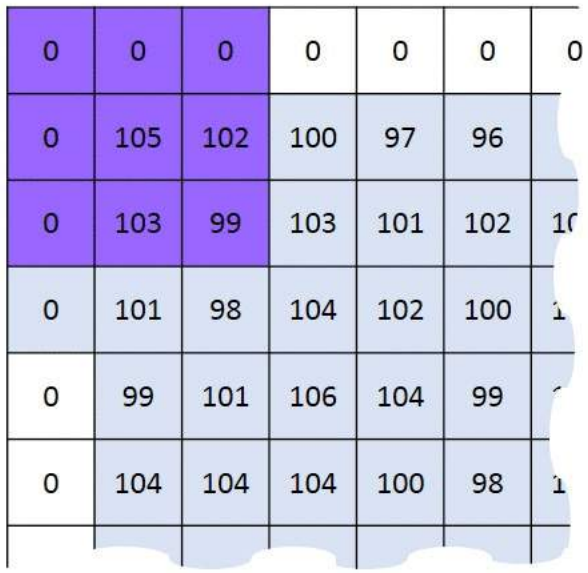
What is a convolution ? Padding



Convolution with horizontal and vertical strides = 1

Source: [Machine Learning guru - Neural Networks CNN]

What is a convolution ? Stride = 2



0	0	0	0	0	0	0
0	105	102	100	97	96	
0	103	99	103	101	102	1(
0	101	98	104	102	100	1
0	99	101	106	104	99	
0	104	104	104	100	98	1

Image Matrix

Kernel Matrix

0	-1	0
-1	5	-1
0	-1	0

320		

Output Matrix

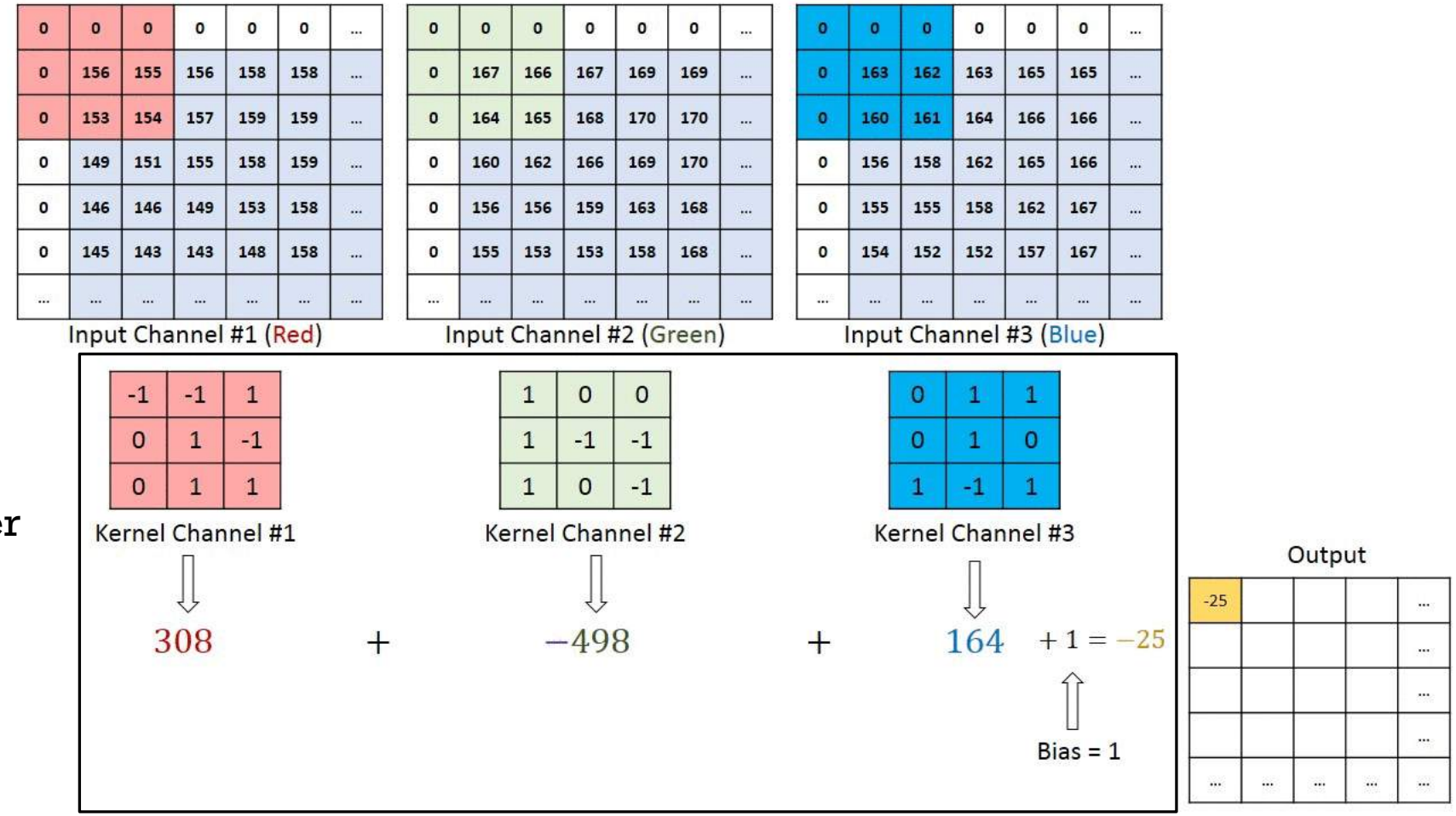
$$0*0+0*-1+0*0$$
$$+0*-1+105*5+102*-1$$
$$+0*0+103*-1+99*0 = 320$$

Convolution with horizontal and vertical strides = 2

Source: Machine Learning guru - Neural Networks CNN

What is a convolution ? Multi Channel

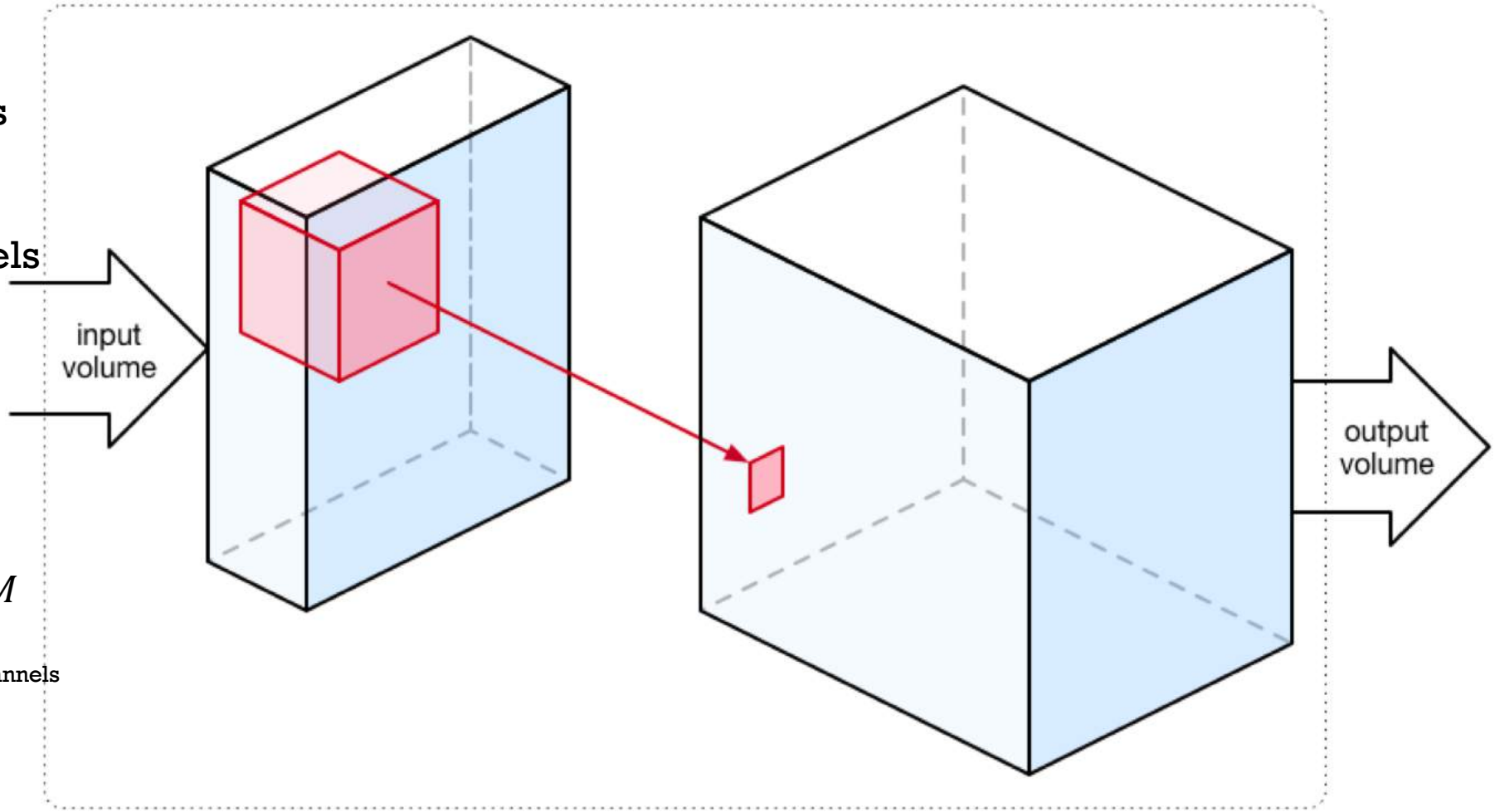
1 convolutional filter
(3)x(3x3)



Source: Machine Learning guru - Neural Networks CNN

What is a convolution ? Multi Channel

N: Number of input channels
W: Width of the kernel
H: Height of the kernel
M: Number of output channels

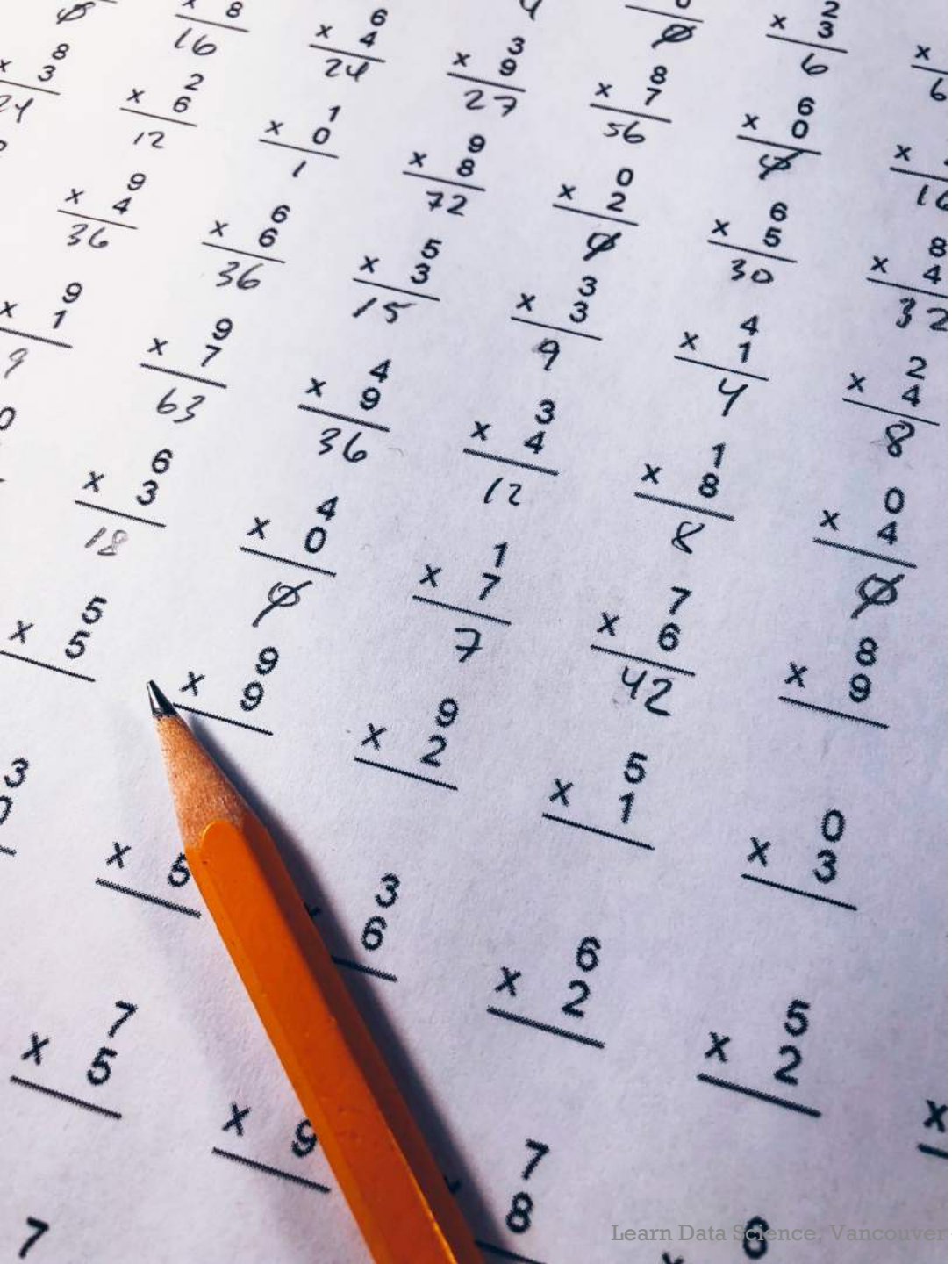


Kernel size = $N * W * H$

#Params = $M * N * W * H + M$

256 convolutions of kernel (3,3) on 256 input channels
256*256*3*3 = **~0.5M**

source: Convolutional Neural Networks on the iphone with vggnet



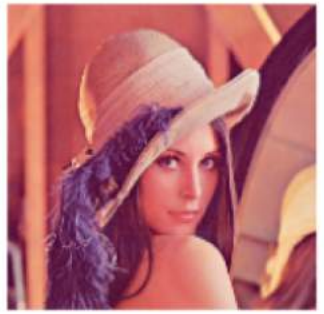
Easily parallelizable

Convolution computations are:

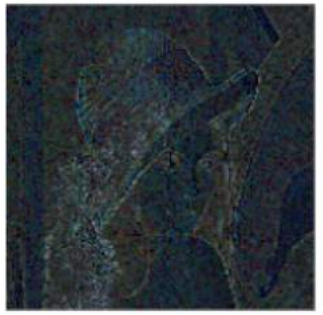
- Independent (across filters and within filter)
- Simple (multiplication and sums)

Why does it work?

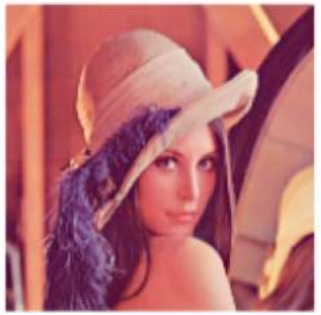
Sharpening filter $\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$

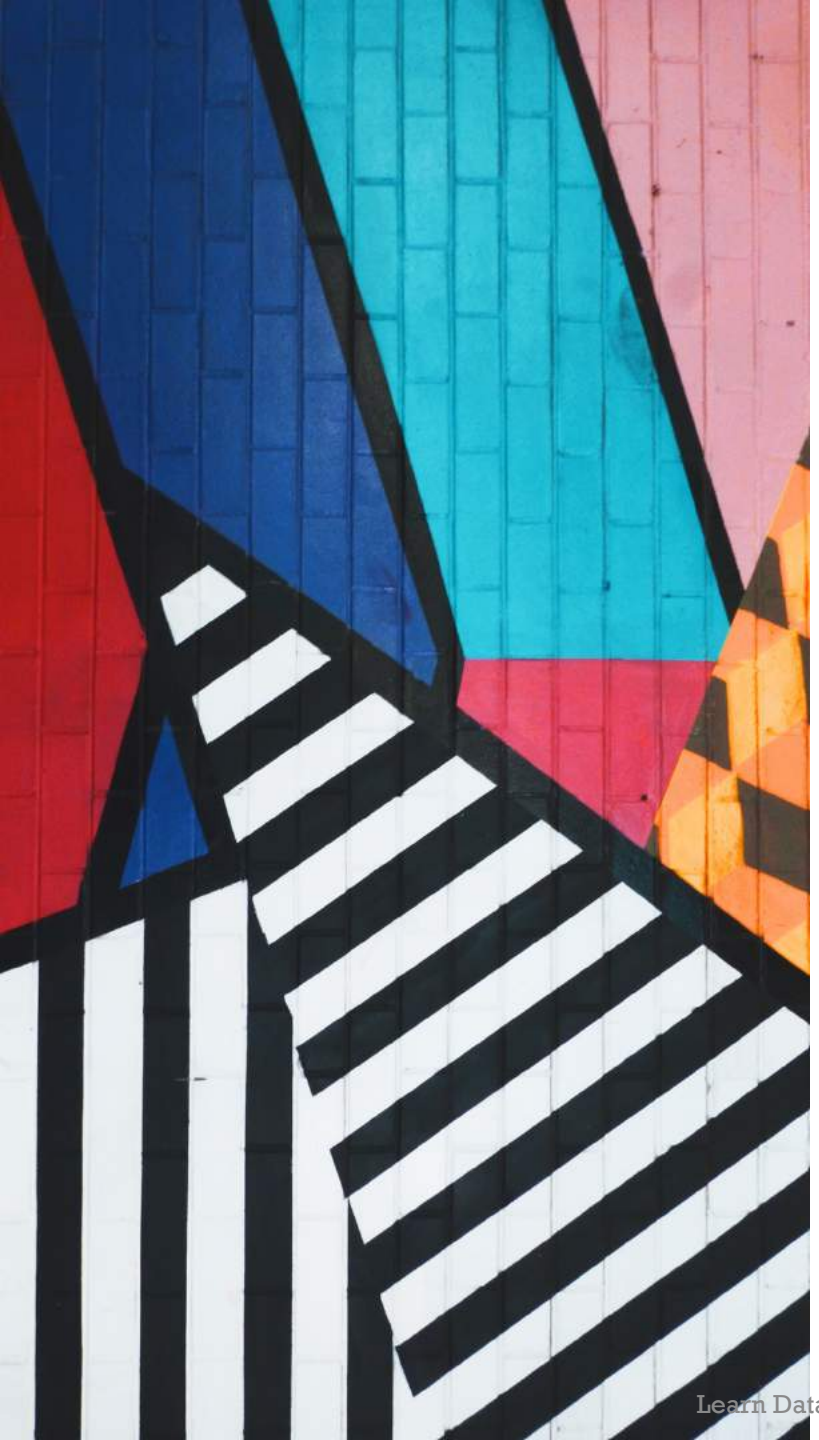


Laplacian filter $\begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$



Sobel x-axis filter $\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$

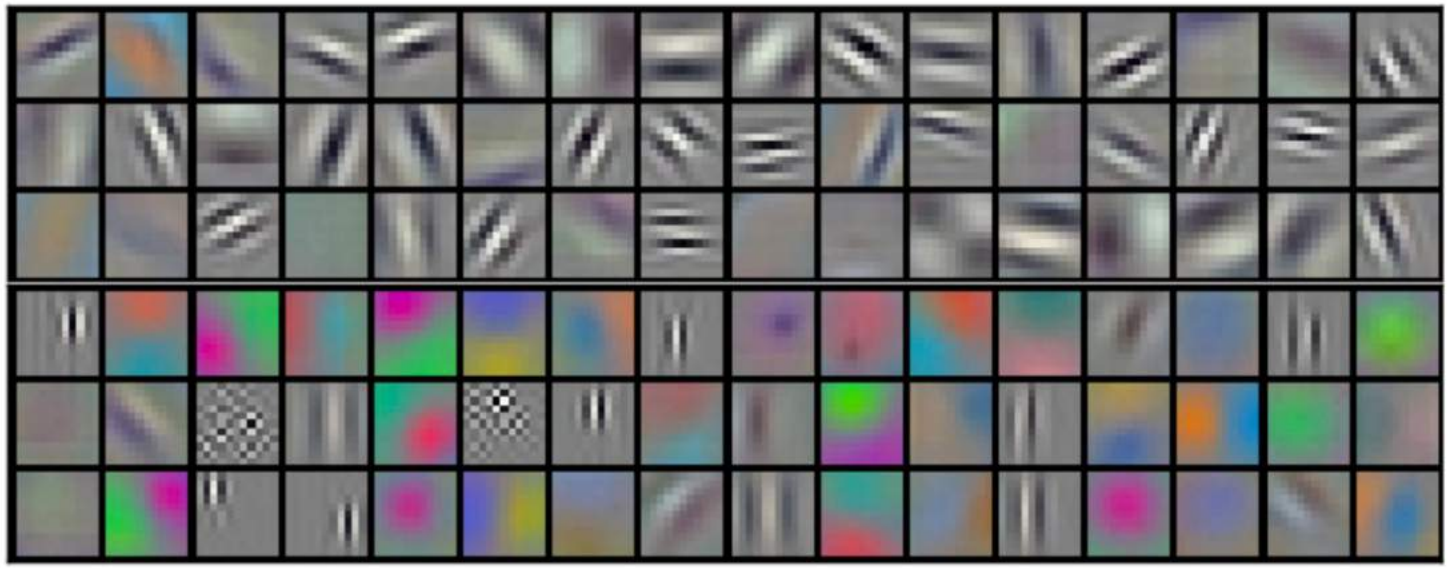




Why does it work?

- Detect patterns at larger and larger scale by stacking convolution layers on top of each others to grow the receptive field

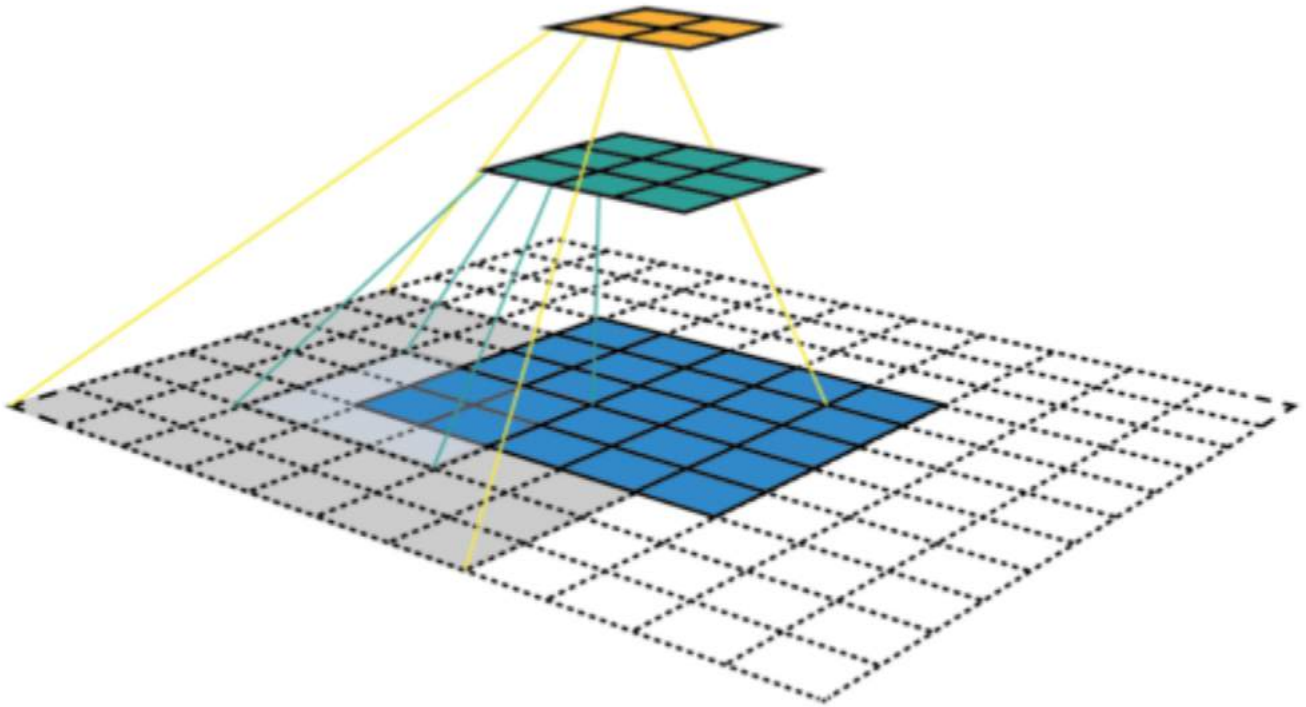
- Applicable to spatially correlated data



Source: AlexNet first 96 (55x55) filters learned represented in RGB space (3 input channels)

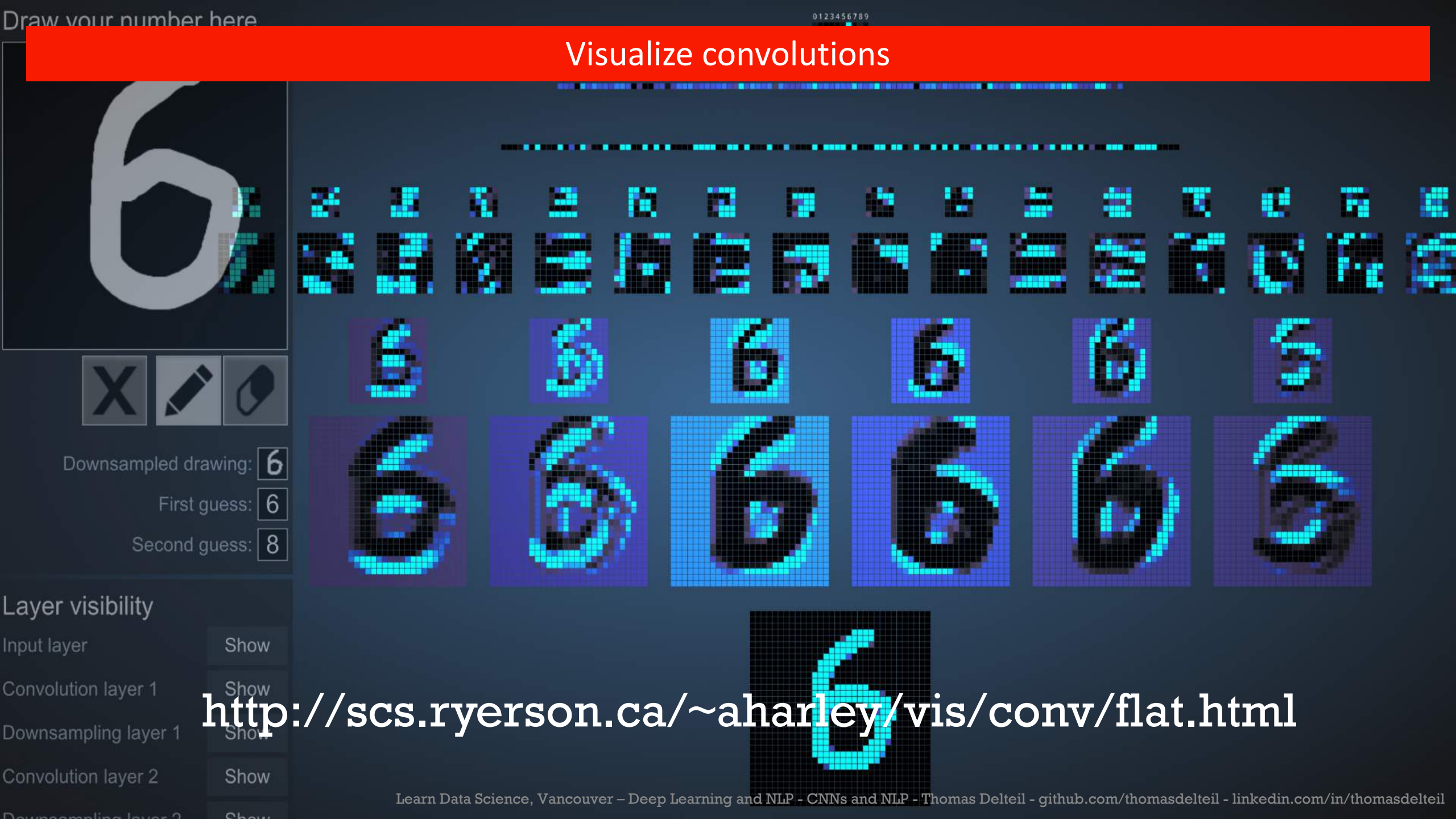
Growing receptive field

Deeper in the network



Source: ML Review, A guide to receptive field arithmetic

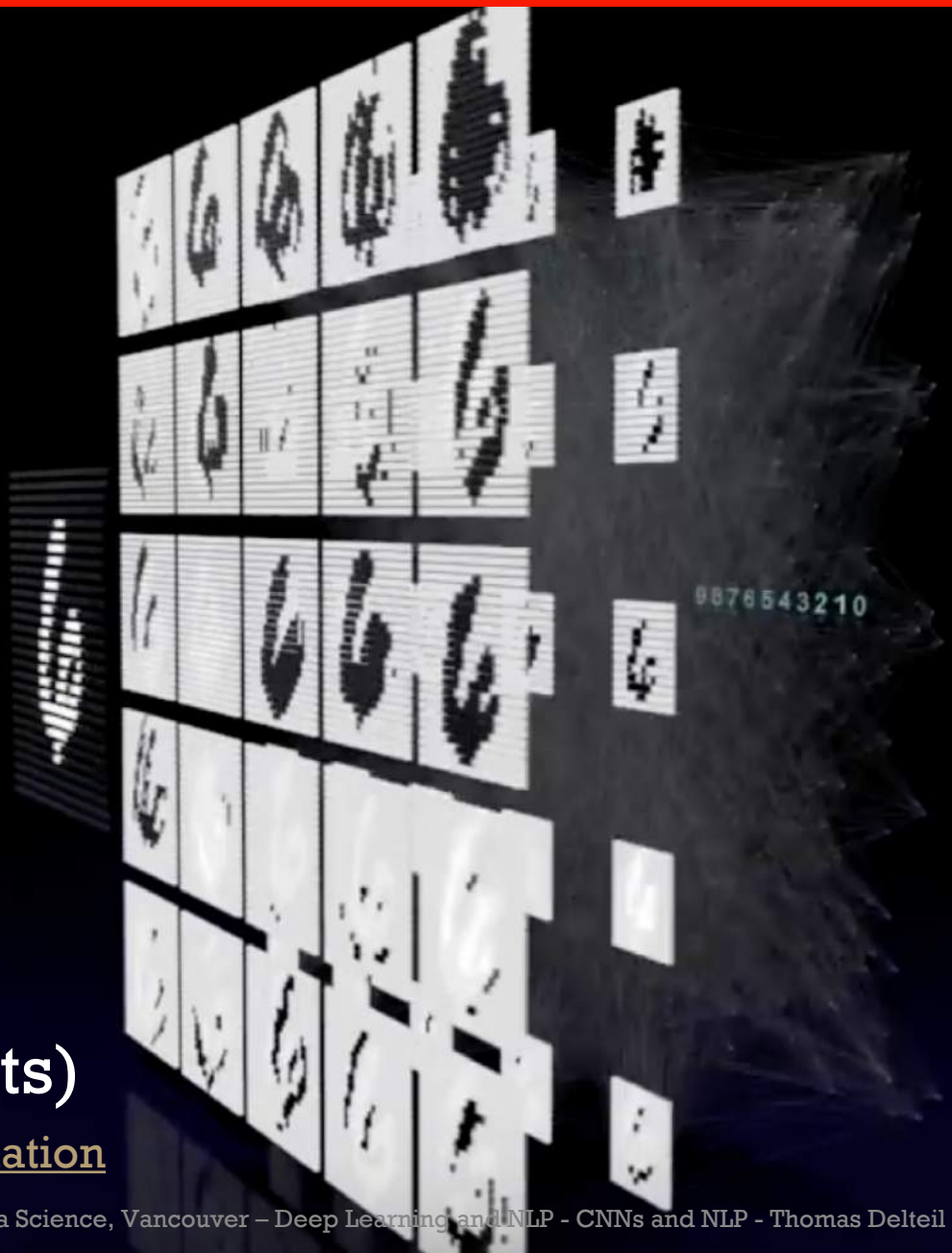
Visualize convolutions



http://scs.ryerson.ca/~aharley/vis/conv/flat.html

Visualize convolutions

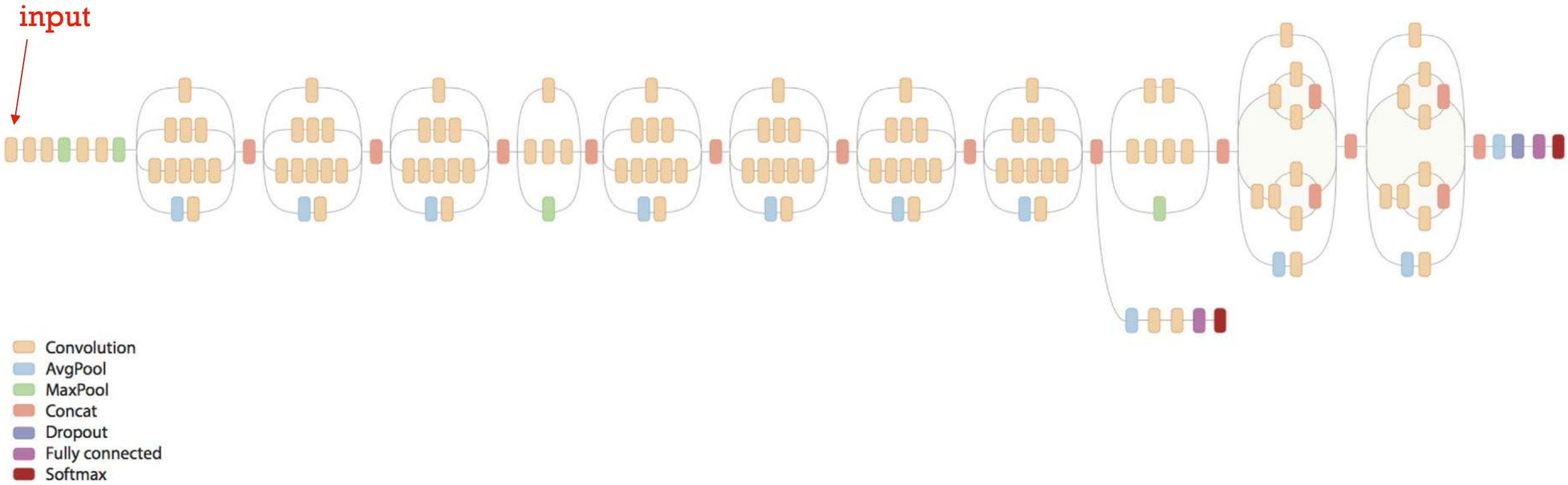
Hidden Neurons: 19794
Synapses: 3610000
Synapses shown: 2%
Learning: BP



(warning flashing lights)

Source: Neural Network 3D Simulation

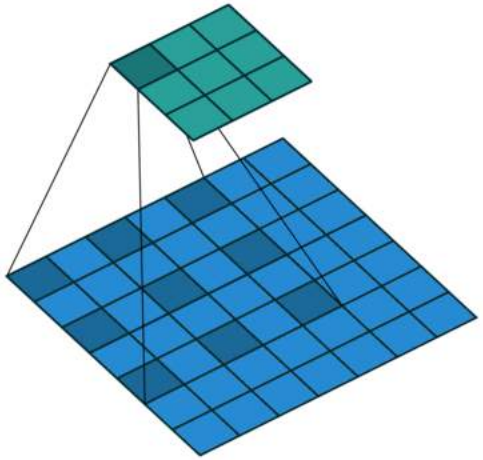
State of the art networks are getting deeper and more complex



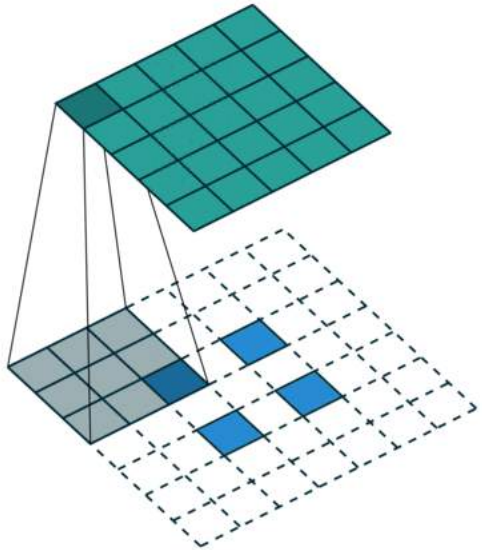
Source: Inception v3

High number of parameters => Requires a lot of data to train

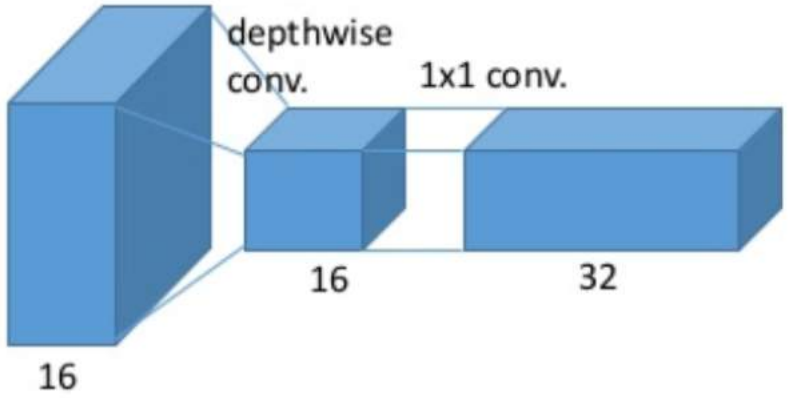
Advanced type of convolutions



Dilated Convolutions
WaveNet



Transposed Convolutions (deconvolution)
EnhanceNet

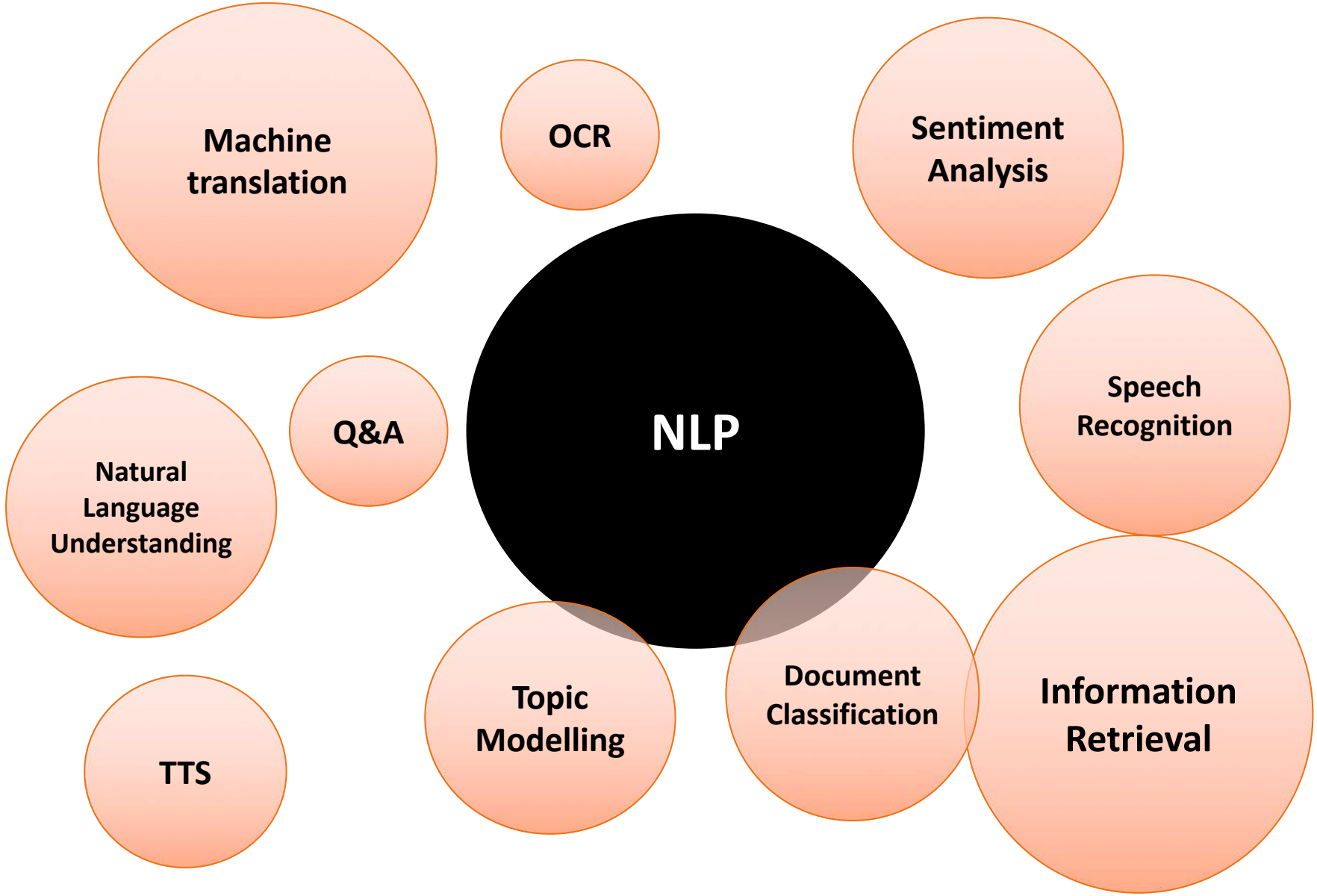


Depth-wise separable Convolutions
MobileNet

Source: An introduction to different types of convolutions

On to Natural Language Processing

NLP Domains


Machine translation
OCR
Sentiment Analysis
NLP
Speech Recognition
Q&A
Natural Language Understanding
Topic Modelling
Document Classification
Information Retrieval
TTS

NLP Industry Facts

£1.3T

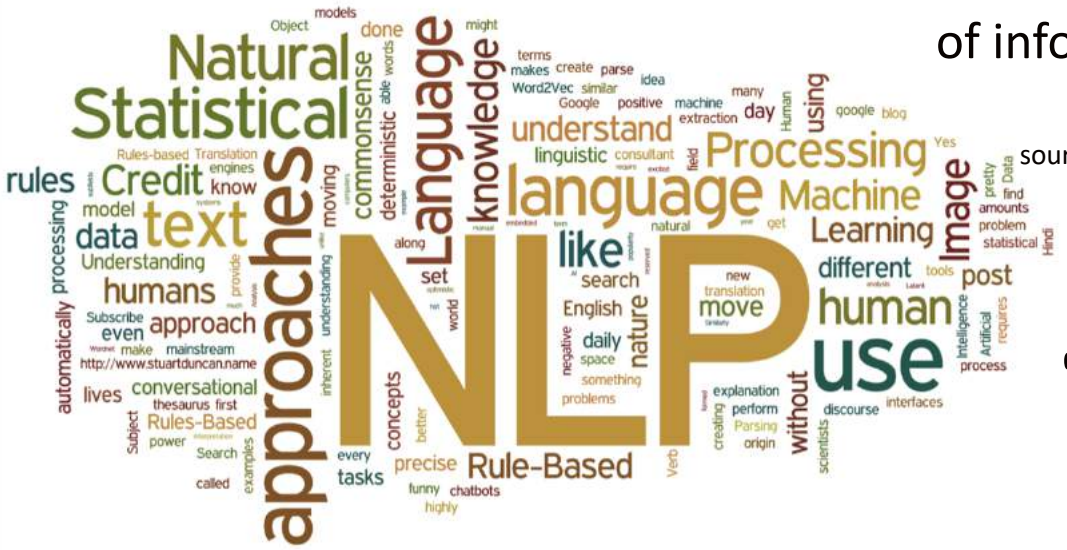
value of company data

source: IDC, 2014

8.4PB

of information per second as of 2020

source: business2comunity, 2016



10%

of organizations expect to commercialise their data by 2020

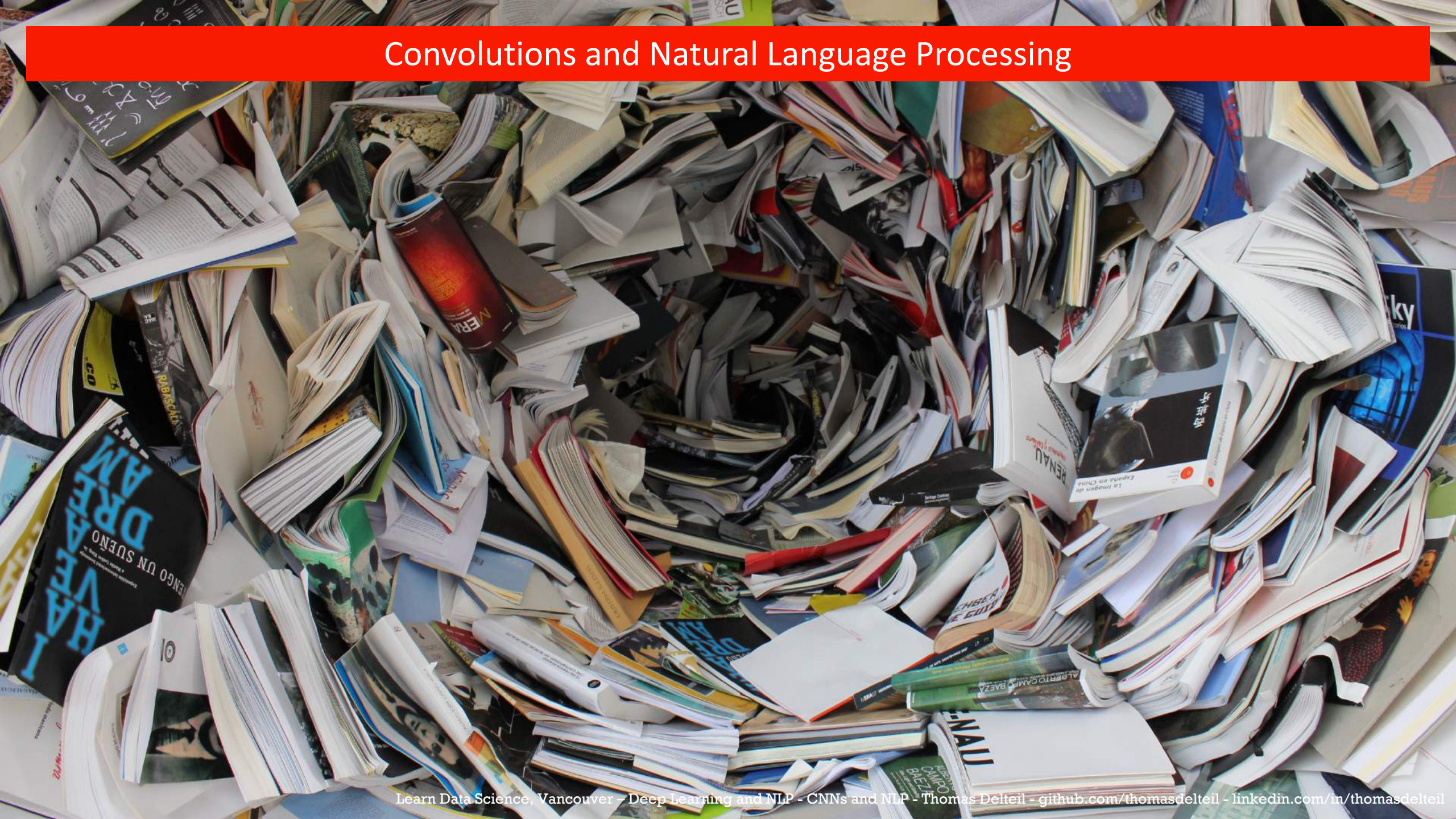
source: Gartner, 2016

70%

of companies use customer feedback

Source: business2comunity, 2016

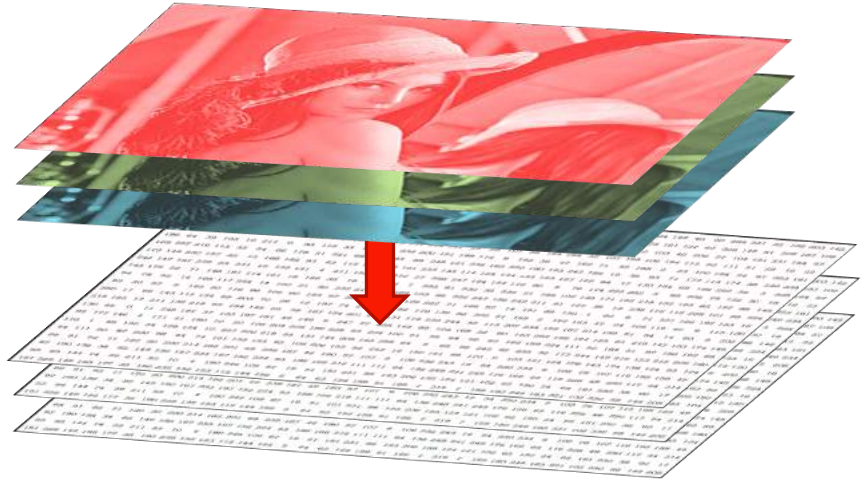
Convolutions and Natural Language Processing

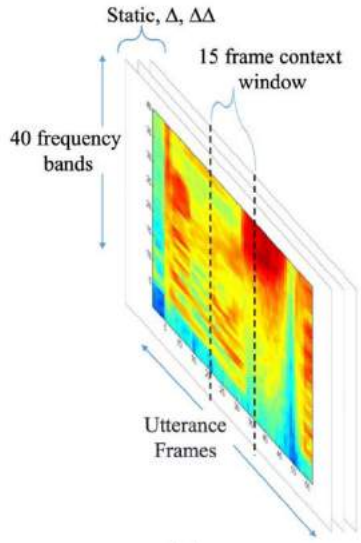


Data Representation

When I read some of the rules for speaking the English language correctly, I think any fool can make a rule, and every fool will mind it

Henry David Thoreau





?

source: Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn,and Dong Yu,. Classification Convolutional Neural Networks for Speech Recognition, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2014

Encoding Data word-level

- Word-level embedding (word2vec). Word -> N-dimensional vector

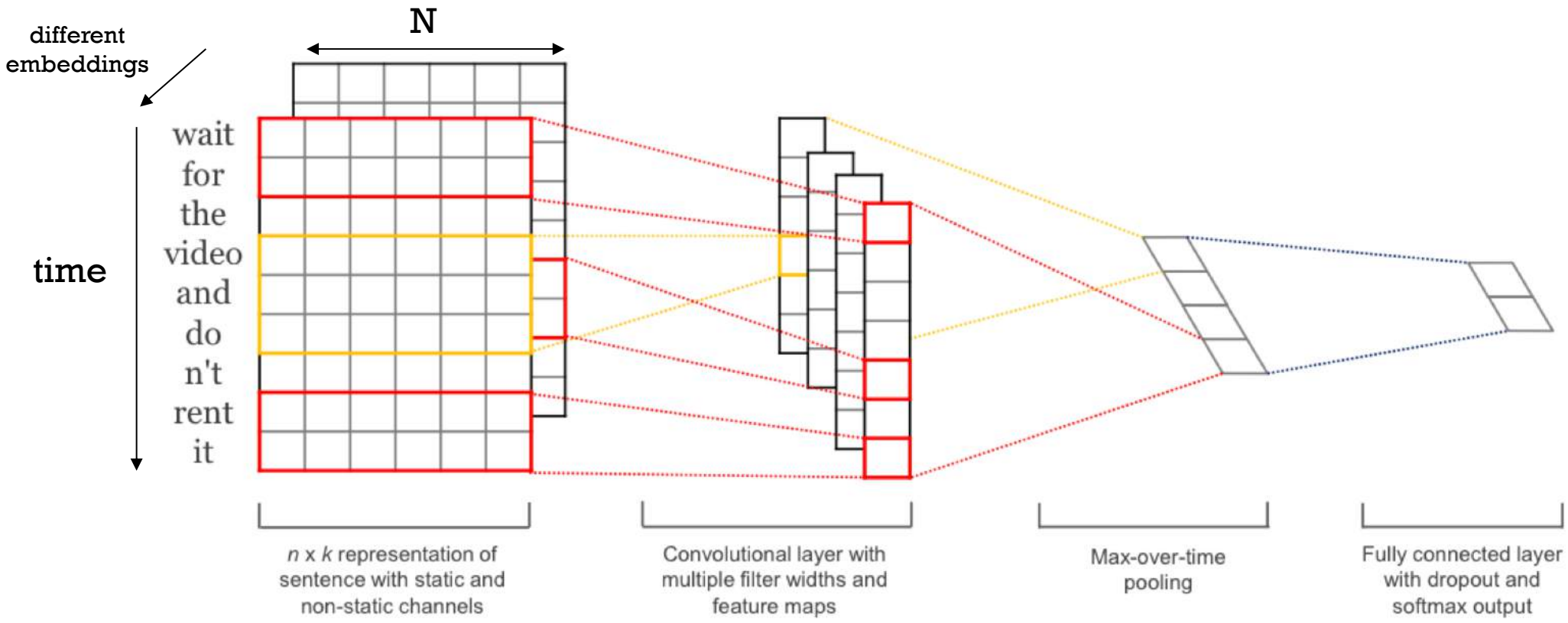
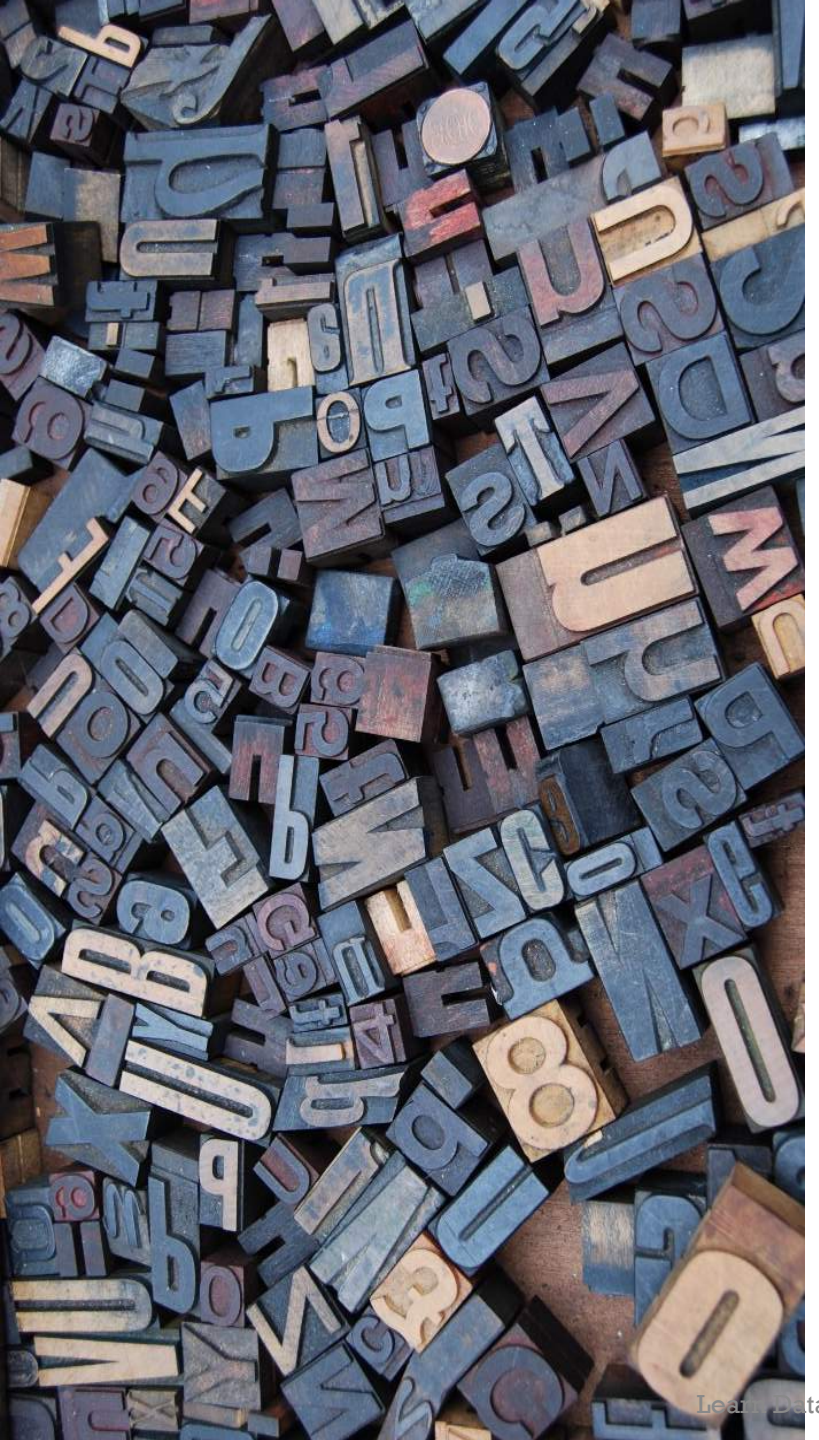


Figure 1: Model architecture with two channels for an example sentence.

Source: Convolutional Neural Networks for Sentence Classification, Yoon Kim, 2014



Encoding Data – Character-level

- One-hot encoding
- Alphabet
- Sparse representation

- Character embedding

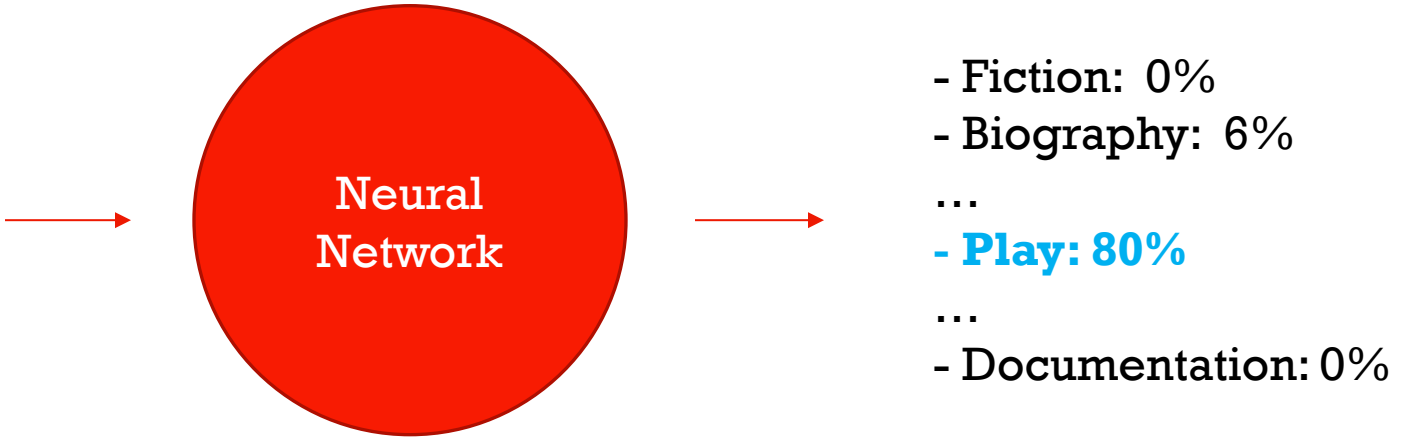
	V	A	N	C	O	U	V	E	R		N	L	P	...
_	0	0	0	0	0	0	0	0	0	1	0	0	0	
-	0	0	0	0	0	0	0	0	0	0	0	0	0	
.	0	0	0	0	0	0	0	0	0	0	0	0	0	
A	0	1	0	0	0	0	0	0	0	0	0	0	0	
B	0	0	0	0	0	0	0	0	0	0	0	0	0	
C	0	0	0	1	0	0	0	0	0	0	0	0	0	
D	0	0	0	0	0	0	0	0	0	0	0	0	0	
E	0	0	0	0	0	0	0	1	0	0	0	0	0	
F	0	0	0	0	0	0	0	0	0	0	0	0	0	
G	0	0	0	0	0	0	0	0	0	0	0	0	0	
H	0	0	0	0	0	0	0	0	0	0	0	0	0	
I	0	0	0	0	0	0	0	0	0	0	0	0	0	
J	0	0	0	0	0	0	0	0	0	0	0	0	0	
K	0	0	0	0	0	0	0	0	0	0	0	0	0	
L	0	0	0	0	0	0	0	0	0	0	0	1	0	
M	0	0	0	0	0	0	0	0	0	0	0	0	0	
N	0	0	1	0	0	0	0	0	0	0	1	0	0	
O	0	0	0	0	1	0	0	0	0	0	0	0	0	
P	0	0	0	0	0	0	0	0	0	0	0	0	1	
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	
R	0	0	0	0	0	0	0	0	1	0	0	0	0	
S	0	0	0	0	0	0	0	0	0	0	0	0	0	
T	0	0	0	0	0	0	0	0	0	0	0	0	0	
U	0	0	0	0	0	1	0	0	0	0	0	0	0	
V	1	0	0	0	0	0	1	0	0	0	0	0	0	
W	0	0	0	0	0	0	0	0	0	0	0	0	0	
X	0	0	0	0	0	0	0	0	0	0	0	0	0	
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	
Z	0	0	0	0	0	0	0	0	0	0	0	0	0	

Text classification, N categories

Text classification, N categories

SCENE FROM "DAN'L DRUCE."

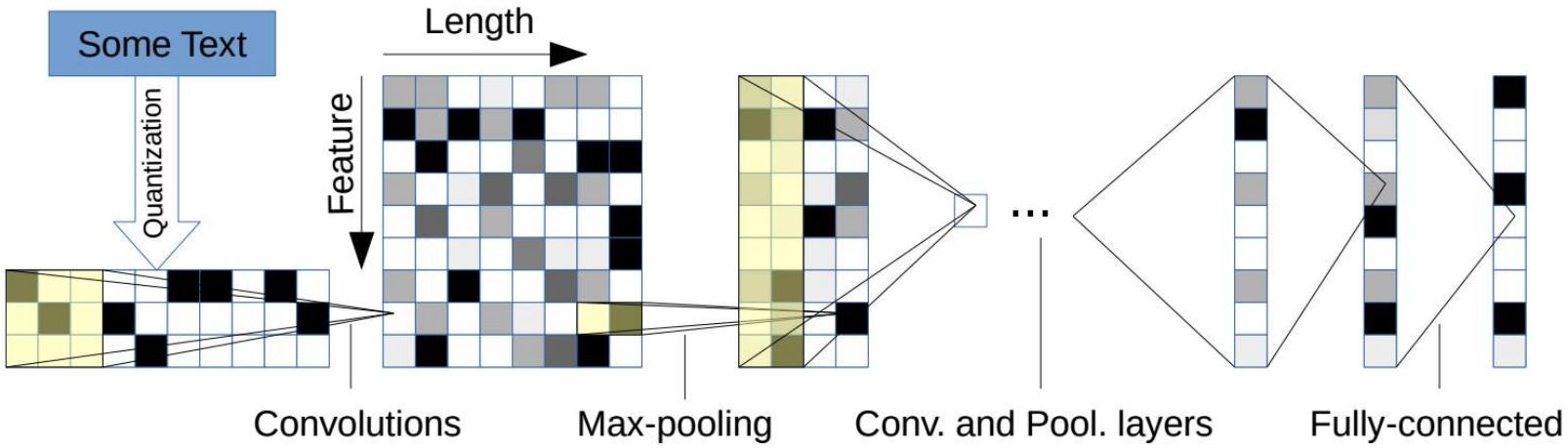
This interesting domestic drama, by Mr. W. S. Gilbert, has continued to engage the sympathies of a nightly sufficient audience at the Haymarket Theatre, where it has now been represented more than sixty times. Its subject and character were described by us, in the ordinary report of theatrical novelties, about two months ago. Our readers will probably not need to be reminded that the hero of the story, Dan'l Druce, the blacksmith, is a solitary recluse dwelling on the coast of Norfolk, where his lone cottage is visited by fugitives from party vengeance during the civil wars of the Commonwealth. His hoard of money is stolen; but a different sort of treasure, a helpless female infant; is left by some mysterious agency, and may be accepted, as in George Eliot's tale of "Silas Marner," for a Divine gift to the sad-hearted misanthrope, far better than riches. In this spirit, at least, he is content to receive the precious human charge; and so to those who would remove it from his home, Dan'l Druce here makes answer with the solemn exclamation, "Touch not the Lord's gift!" This character is well acted by Mr. Hermann Vezin.



Deep Neural Network: Crepe Model

Visualization with Netron

Intuition: convolutions act similarly as n-grams



Layer	Large Feature	Small Feature	Kernel	Pool
1	1024	256	7	3
2	1024	256	7	3
3	1024	256	3	N/A
4	1024	256	3	N/A
5	1024	256	3	N/A
6	1024	256	3	3

Layer	Output Units Large	Output Units Small
7	2048	1024
8	2048	1024
9	Depends on the problem	

source: Xiang Zhang, Junbo Zhao, Yann LeCun. Character-level Convolutional Networks for Text Classification. NIPS 2015

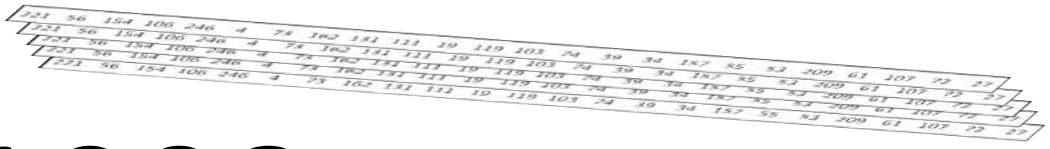
Temporal Convolution (256 69*7/1)

	V	A	N	C	O	U	V	E	R		…	1013
_	0	0	0	0	0	0	0	0	0	1		…
-	0	0	0	0	0	0	0	0	0	0	…	…
.	0	0	0	0	0	0	0	0	0	0	…	…
A	0	1	0	0	0	0	0	0	0	0	…	…
B	0	0	0	0	0	0	0	0	0	0	…	…
C	0	0	0	1	0	0	0	0	0	0	…	…
D	0	0	0	0	0	0	0	0	0	0	…	…
E	0	0	0	0	0	0	0	1	0	0	…	…
F	0	0	0	0	0	0	0	0	0	0	…	…
G	0	0	0	0	0	0	0	0	0	0	…	…
H	0	0	0	0	0	0	0	0	0	0	…	…
I	0	0	0	0	0	0	0	0	0	0	…	…
J	0	0	0	0	0	0	0	0	0	0	…	…
K	0	0	0	0	0	0	0	0	0	0	…	…
L	0	0	0	0	0	0	0	0	0	0	…	…
M	0	0	0	0	0	0	0	0	0	0	…	…
N	0	0	1	0	0	0	0	0	0	0	…	…
O	0	0	0	0	1	0	0	0	0	0	…	…
P	0	0	0	0	0	0	0	0	0	0	…	…
Q	0	0	0	0	0	0	0	0	0	0	…	…
R	0	0	0	0	0	0	0	0	1	0	…	…
S	0	0	0	0	0	0	0	0	0	0	…	…
T	0	0	0	0	0	0	0	0	0	0	…	…
U	0	0	0	0	0	1	0	0	0	0	…	…
V	1	0	0	0	0	0	1	0	0	0	…	…
W	0	0	0	0	0	0	0	0	0	0	…	…
X	0	0	0	0	0	0	0	0	0	0	…	…
Y	0	0	0	0	0	0	0	0	0	0	…	…
Z	0	0	0	0	0	0	0	0	0	0	…	…

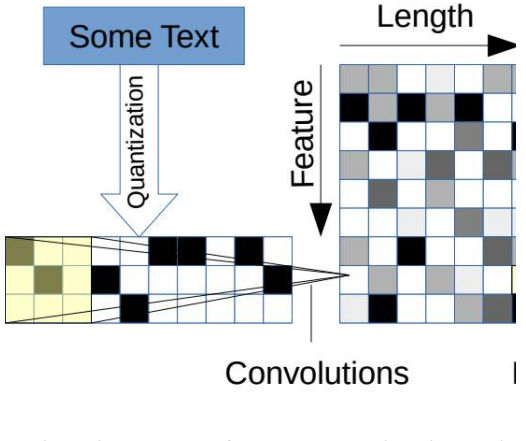
	0	1	2	3	4	…	…	…	…	…	…	…	…	1007
0	6.4	1.1	3.2	0.3	-0.4	…	…	…	…	…	…	…	…	…
1	-2.1	0.2	-3.4	…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
254	…	…	…	…	…	…	…	…	…	…	…	…	…	…
255	1.2	3.4	-1	1.2	3.2	…	…	…	…	…	…	…	…	…

x 256

1x1008x256 = ~256k

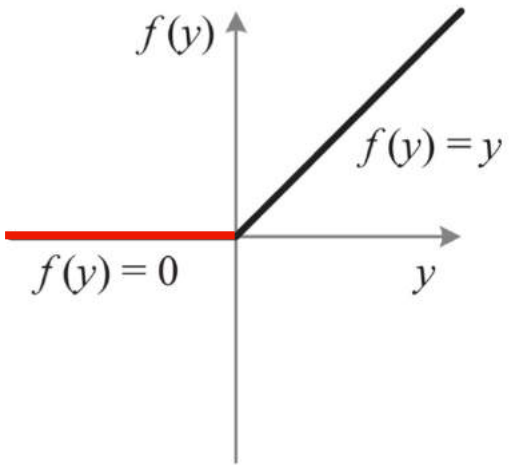


x 1008



69x1014x1 = ~70k

Activation Function: Rectified Linear Unit (ReLU)



$$f(y) = \begin{cases} y, & y \geq 0 \\ 0, & y < 0 \end{cases}$$

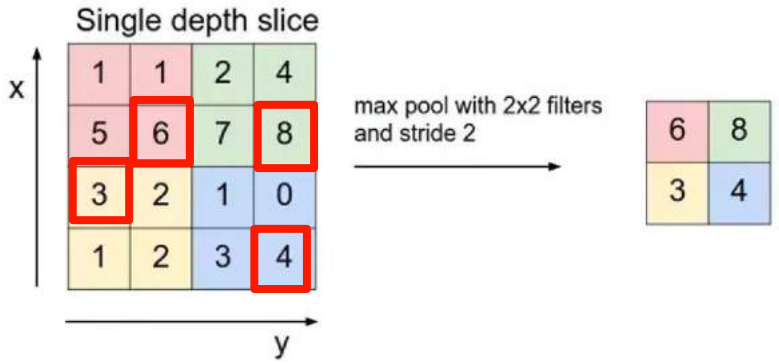
	0	1	2	3	4	5	…	1007
0	6.4	1.1	3.2	0.3	-0.4	0.2	…	…
…	…	…	…	…	…	…	…	…
255	1.2	3.4	-1	1.2	3.2	2.8	…	…

1x1008x256 = ~256k

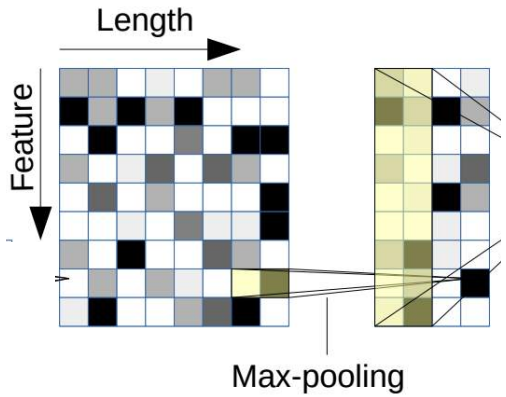
	0	1	2	3	4	5	…	1007
0	6.4	1.1	3.2	0.3	0	0.2	…	…
…	…	…	…	…	…	…	…	…
255	1.2	3.4	0	1.2	3.2	2.8	…	…

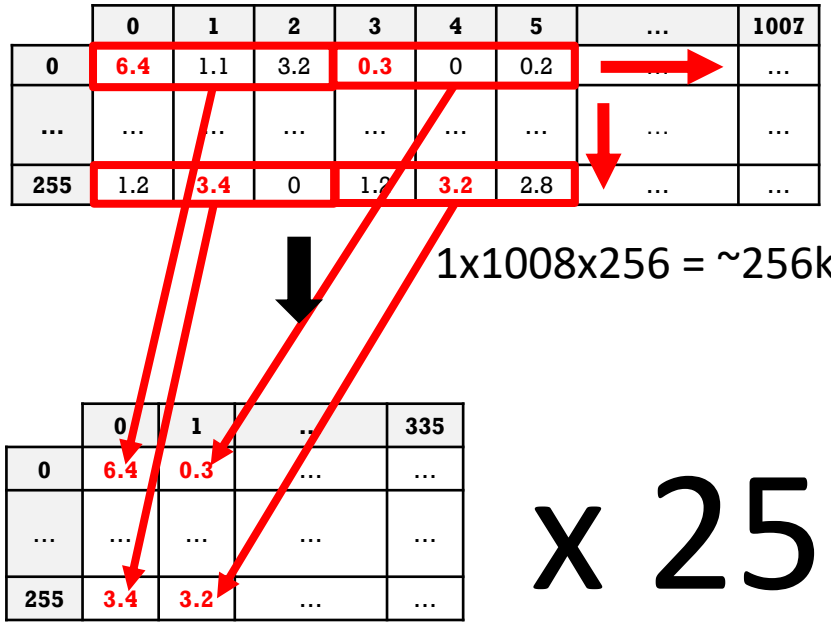
1x1008x256 = ~ 256k

Down-sampling: Max-Pooling (256 1*3/3)



source : Stanford's CS231n





x 336

x 256

Fast forward…

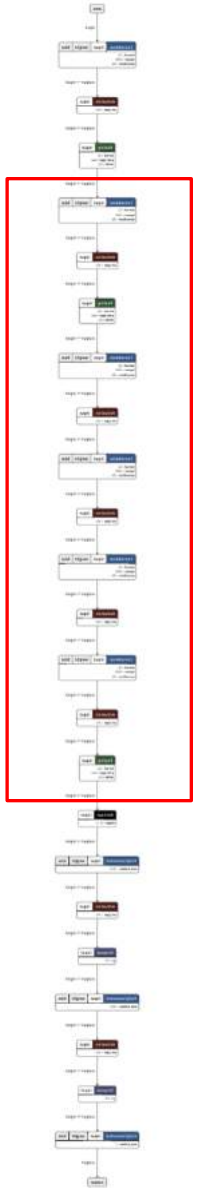
1x336x256 = **~86k** <- after 1 convolution layer (69*7/1) and 1 max pooling (3x1/3)

1x330x256 = ~85k <- after 1 convolution layer (1*7/1)

1x110x256 = ~28k <- 1 max-pooling (1*3/3)

3x102x256 = ~26k <- 4 convolutions layers (1*3/1)

1x34x256 = **~9k** <- 1 max-pooling (1*3/3)



Flattening Layer

	0	1	2	3	4	5	6	7	8	...	33
0	6.4	0.1	...	...	...	...	...	...	...	...	...
1	2.1	24.9	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...
255	...	...	...	...	...	...	...	...	...	...	9.9

1x34x256 = ~9k

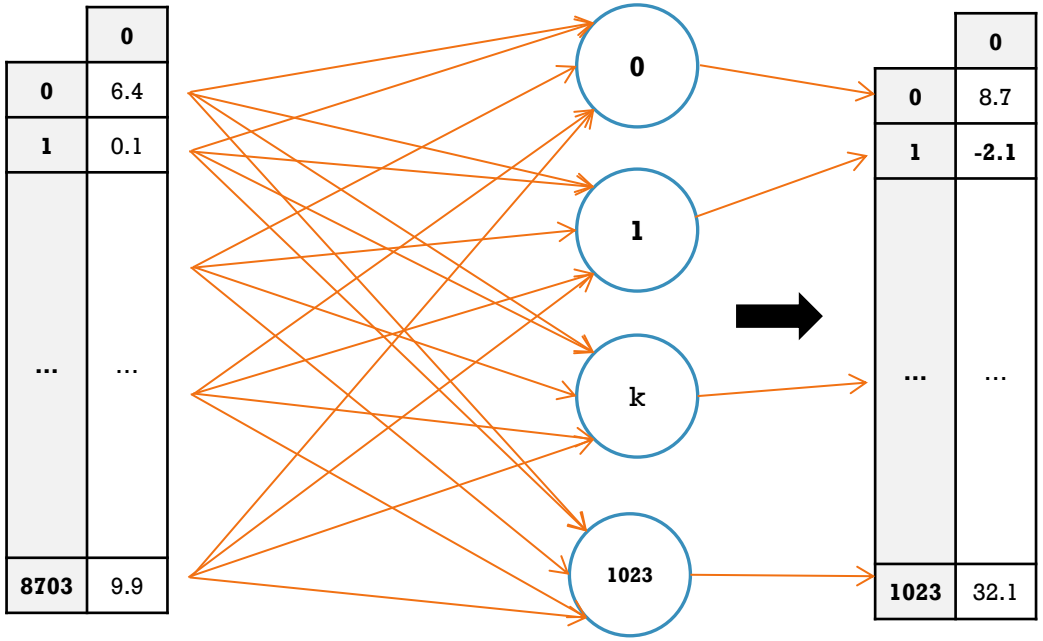
	0
0	6.4
1	0.1
...	...
34	2.1
35	24.9
...	...
...	...
...	...
8703	9.9

x 256

8704x1x1 = ~9k

Fully Connected / Dense layer (1024)

$$f_k(X) = \sum_{i=0}^{8703} w_{ki} * x_i + b_k$$



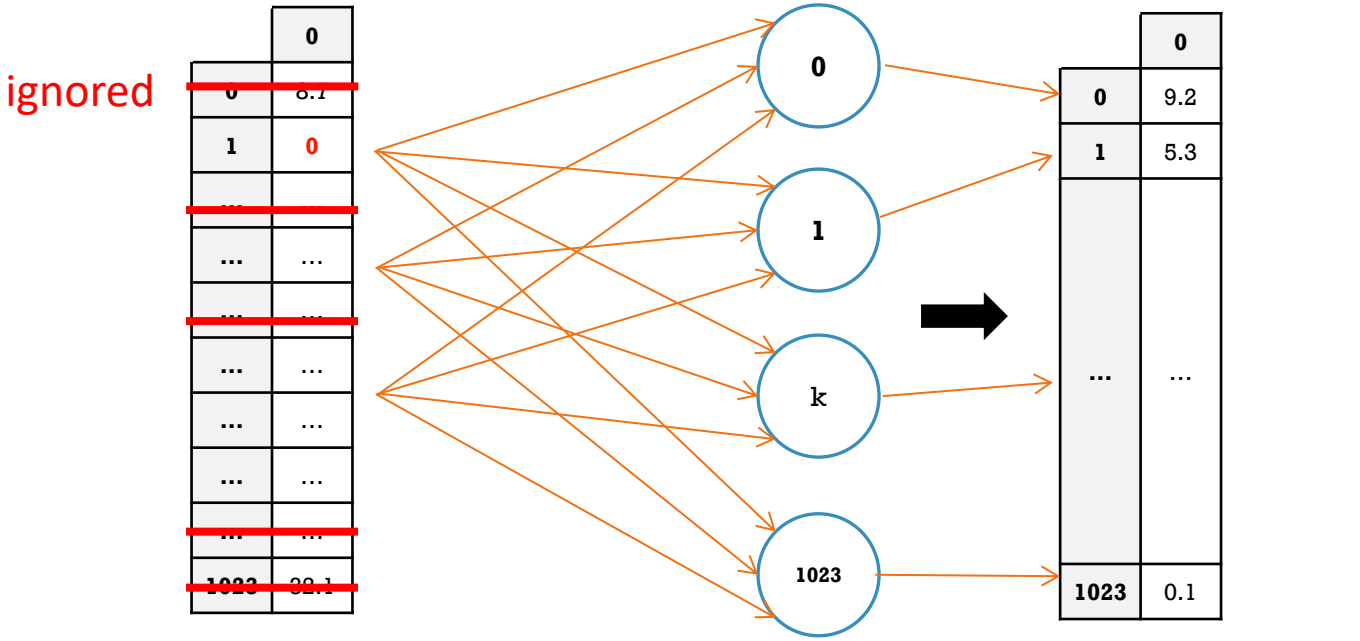
8704x1x1 = ~9k

1024x1x1 = ~1k

x 1024

Dropout (p=0.5) + Fully Connected Layer (1024)

$$f_k(X) = \sum_{i=0}^{8703} w_{ki} * x_i + b_k$$



1024x1x1 = ~1k

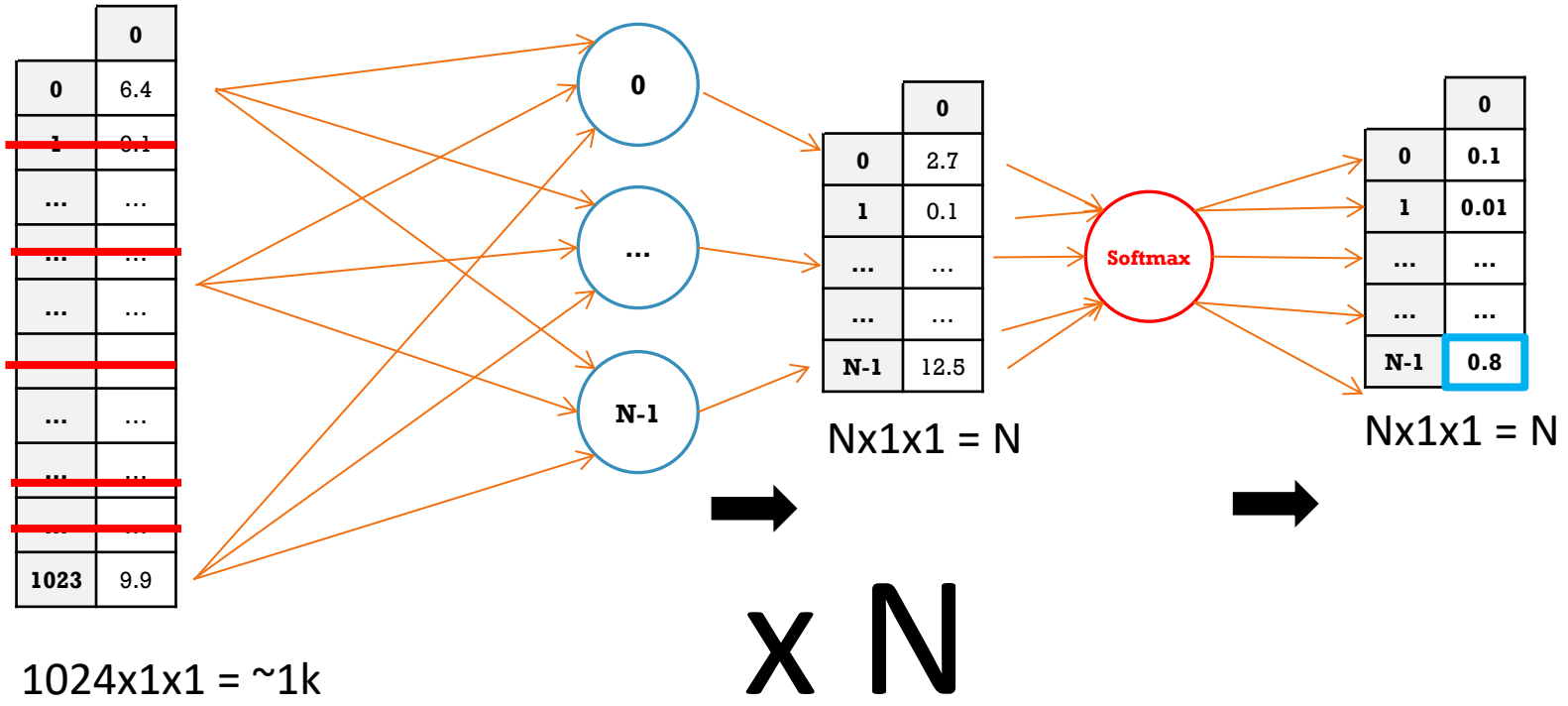
1024x1x1 = ~1k

x 1024

Output: Dropout + Dense + Softmax for **N categories**

$$softmax(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=0}^{N-1} e^{z_j}}$$

ignored



1024x1x1 = ~1k

Nx1x1 = N

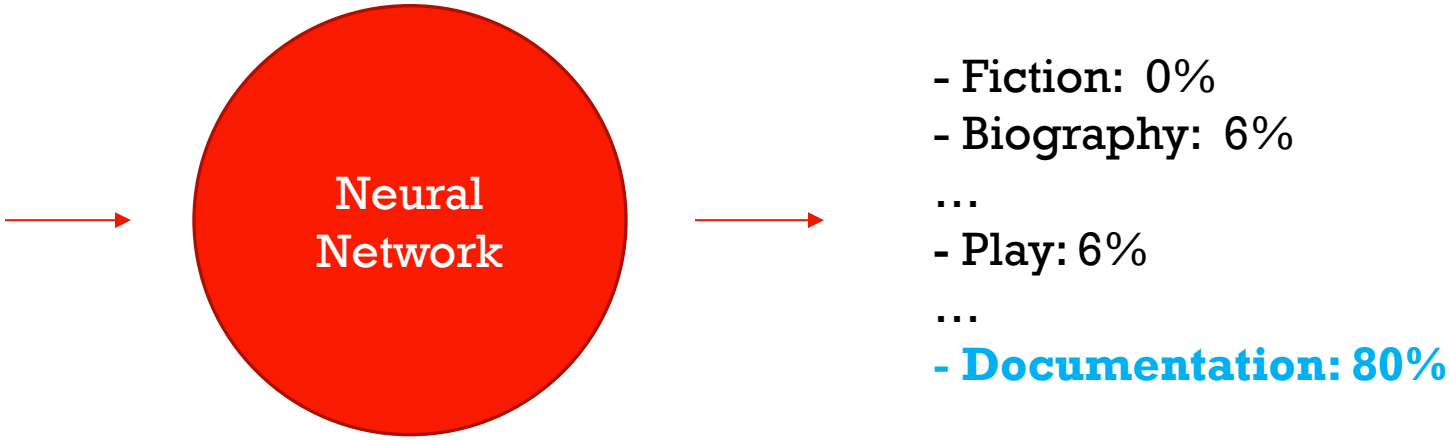
Nx1x1 = N

x N

Text classification, N categories

SCENE FROM "DAN'L DRUCE."

This interesting domestic drama, by Mr. W. S. Gilbert, has continued to engage the sympathies of a nightly sufficient audience at the Haymarket Theatre, where it has now been represented more than sixty times. Its subject and character were described by us, in the ordinary report of theatrical novelties, about two months ago. Our readers will probably not need to be reminded that the hero of the story, Dan'l Druce, the blacksmith, is a solitary recluse dwelling on the coast of Norfolk, where his lone cottage is visited by fugitives from party vengeance during the civil wars of the Commonwealth. His hoard of money is stolen; but a different sort of treasure, a helpless female infant; is left by some mysterious agency, and may be accepted, as in George Eliot's tale of "Silas Marner," for a Divine gift to the sad-hearted misanthrope, far better than riches. In this spirit, at least, he is content to receive the precious human charge; and so to those who would remove it from his home, Dan'l Druce here makes answer with the solemn exclamation, "Touch not the Lord's gift!" This character is well acted by Mr. Hermann Vezin.



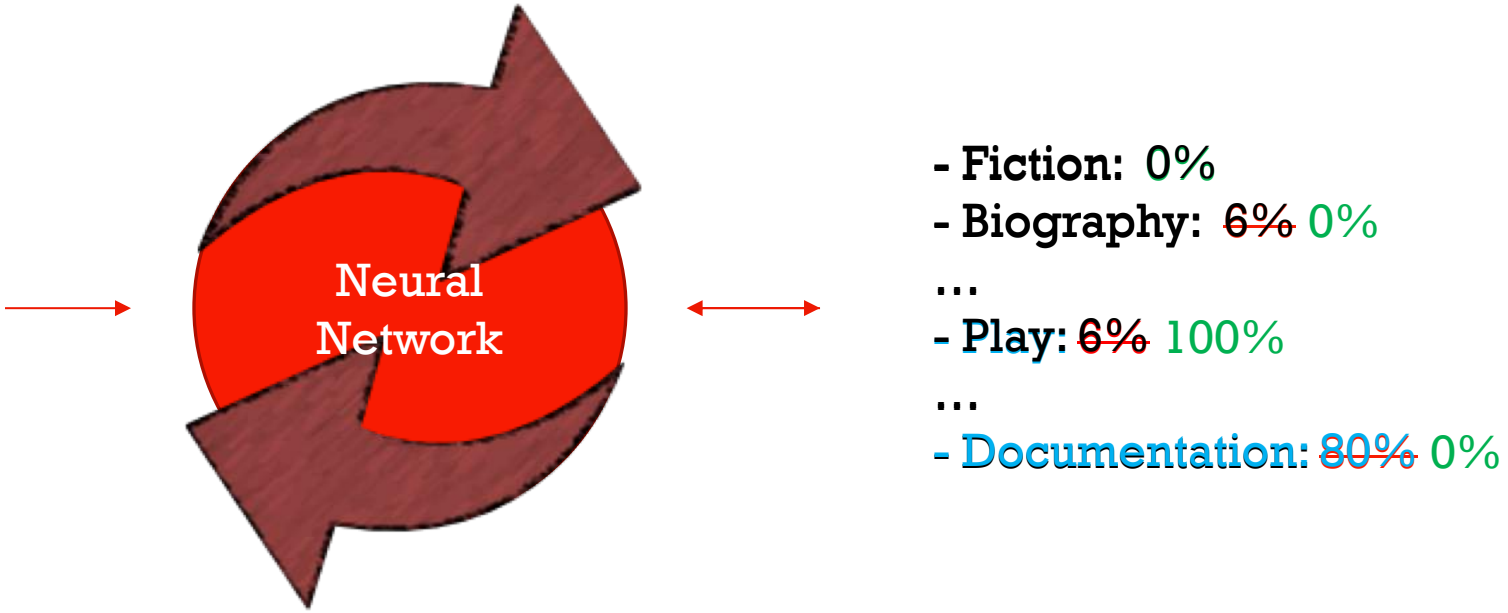
How to train the network? Backward propagation!

Backward propagation – Efficient Gradient Descent

$$w_{ij} = w_{ij} - \eta . \frac{\partial E}{\partial w_{ij}}$$

SCENE FROM "DAN'L DRUCE."

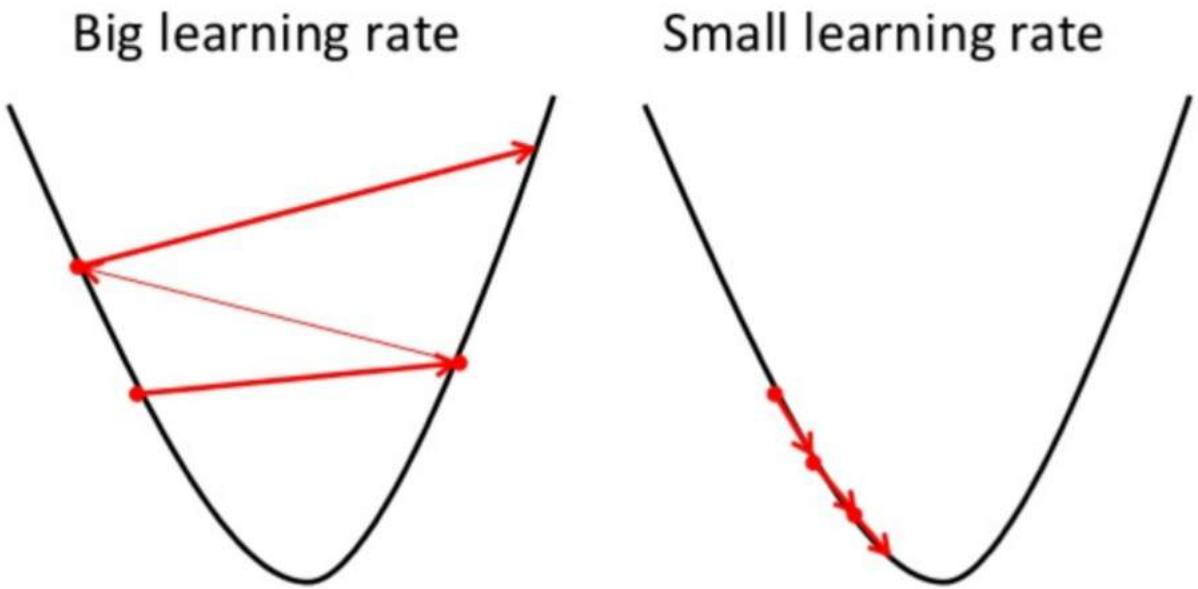
This interesting domestic drama, by Mr. W. S. Gilbert, has continued to engage the sympathies of a nightly sufficient audience at the Haymarket Theatre, where it has now been represented more than sixty times. Its subject and character were described by us, in the ordinary report of theatrical novelties, about two months ago. Our readers will probably not need to be reminded that the hero of the story, Dan'l Druce, the blacksmith, is a solitary recluse dwelling on the coast of Norfolk, where his lone cottage is visited by fugitives from party vengeance during the civil wars of the Commonwealth. His hoard of money is stolen; but a different sort of treasure, a helpless female infant; is left by some mysterious agency, and may be accepted, as in George Eliot's tale of "Silas Marner," for a Divine gift to the sad-hearted misanthrope, far better than riches. In this spirit, at least, he is content to receive the precious human charge; and so to those who would remove it from his home, Dan'l Druce here makes answer with the solemn exclamation, "Touch not the Lord's gift!" This character is well acted by Mr. Hermann Vezin.



Update the weights of the convolutional masks and fully connected units so that the error will be minimized next time

Training Parameters: Learning Rate

Learning Rate η : How much to update the weights for every batch of documents?



Source: Towards data Science: Gradient descent in a nutshell

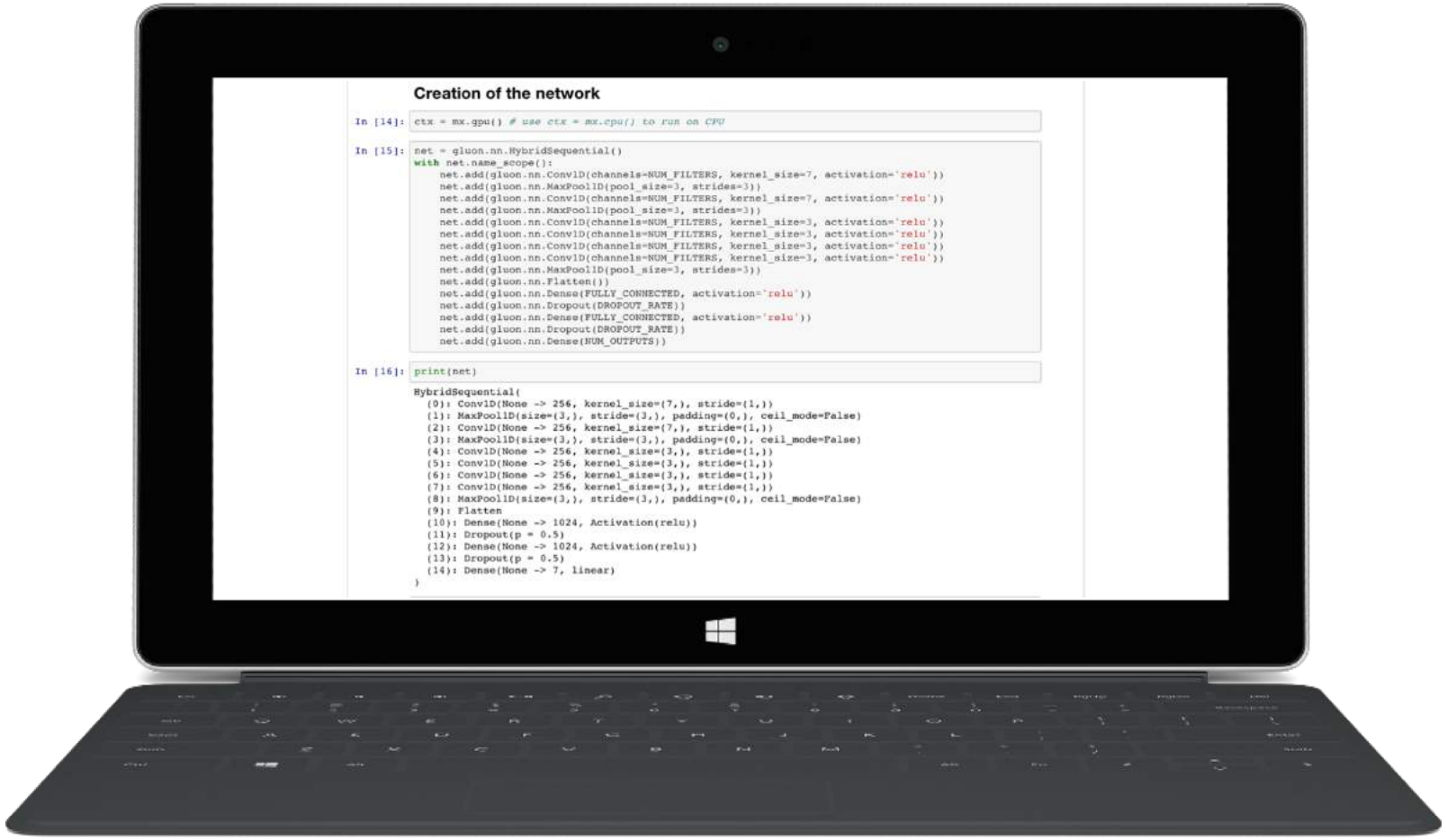
Training parameters: Batch Size

Batch size: How many examples to learn from in one step?

Training parameters: Number of epochs

Number of epochs: How many times should we feed the network the entire training set?

Jupyter notebook demo – Crepe in Apache MXNet/Gluon



https://github.com/ThomasDelteil/CNN_NLP_MXNet

Results

Dataset	Classes	Train Samples
AG's News	4	120,000
Sogou News	5	450,000
DBPedia	14	560,000
Yelp Review Polarity	2	560,000
Yelp Review Full	5	650,000
Yahoo! Answers	10	1,400,000
Amazon Review Full	5	3,000,000
Amazon Review Polarity	2	3,600,000

Traditional approaches

Word-level CNN

Character-level CNN

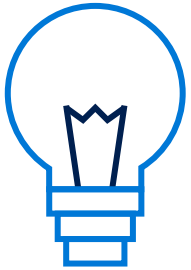
Model	AG	Sogou	DBP.	Yelp P.	Yelp F.	Yah. A.	Amz. F.	Amz. P.
BoW	11.19	7.15	3.39	7.76	42.01	31.11	45.36	9.60
BoW TFIDF	10.36	6.55	2.63	6.34	40.14	28.96	44.74	9.00
ngrams	7.96	2.92	1.37	**4.36**	43.74	31.53	45.73	7.98
ngrams TFIDF	**7.64**	**2.81**	**1.31**	4.56	45.20	31.49	47.56	8.46
Bag-of-means	**16.91**	**10.79**	**9.55**	**12.67**	**47.46**	**39.45**	**55.87**	**18.39**
LSTM	13.94	4.82	1.45	5.26	41.83	29.16	40.57	6.10
Lg. w2v Conv.	9.92	4.39	1.42	4.60	40.16	31.97	44.40	5.88
Sm. w2v Conv.	11.35	4.54	1.71	5.56	42.13	31.50	42.59	6.00
Lg. w2v Conv. Th.	9.91	-	1.37	4.63	39.58	31.23	43.75	5.80
Sm. w2v Conv. Th.	10.88	-	1.53	5.36	41.09	29.86	42.50	5.63
Lg. Lk. Conv.	8.55	4.95	1.72	4.89	40.52	29.06	45.95	5.84
Sm. Lk. Conv.	10.87	4.93	1.85	5.54	41.41	30.02	43.66	5.85
Lg. Lk. Conv. Th.	8.93	-	1.58	5.03	40.52	28.84	42.39	5.52
Sm. Lk. Conv. Th.	9.12	-	1.77	5.37	41.17	28.92	43.19	5.51
Lg. Full Conv.	9.85	8.80	1.66	5.25	38.40	29.90	40.89	5.78
Sm. Full Conv.	11.59	8.95	1.89	5.67	38.82	30.01	40.88	5.78
Lg. Full Conv. Th.	9.51	-	1.55	4.88	38.04	29.58	40.54	5.51
Sm. Full Conv. Th.	10.89	-	1.69	5.42	**37.95**	29.90	40.53	5.66
Lg. Conv.	12.82	4.88	1.73	5.89	39.62	29.55	41.31	5.51
Sm. Conv.	15.65	8.65	1.98	6.53	40.84	29.84	40.53	5.50
Lg. Conv. Th.	13.39	-	1.60	5.82	39.30	**28.80**	40.45	**4.93**
Sm. Conv. Th.	14.80	-	1.85	6.49	40.16	29.84	**40.43**	5.67

Data Augmentation

For images




For text



Humans to rephrase the examples

Synonyms
Similar semantic meaning

Data Augmentation

The quick brown fox jumps over the lazy dog

Data Augmentation

The ~~quick~~ ~~brown~~ fox ~~jumps~~ over the ~~lazy~~ ~~dog~~

fast	hazel	leaps	idle
swift	brunette	springs	indolent
speedy	chestnut	bounds	slothful
		hops	hound
			pup
			mutt

Data Augmentation

The ~~quick~~ ~~brown~~ fox ~~jumps~~ over the ~~lazy~~ ~~dog~~

fast	hazel	leaps	idle
swift	brunette	springs	indolent
speedy	chestnut	bounds	slothful
		hops	hound
			pup
			mutt

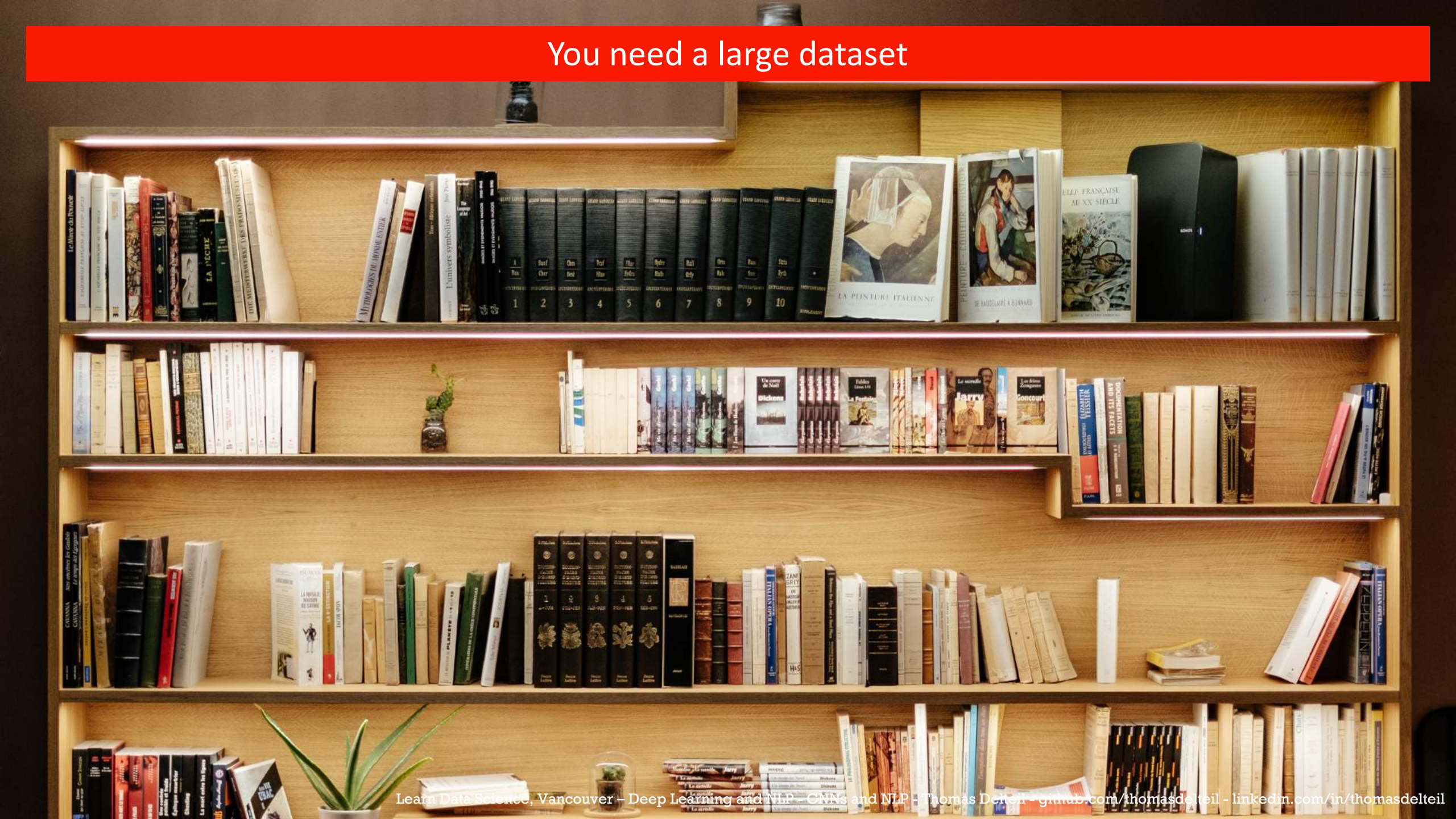
Data Augmentation

The ~~quick~~ ~~brown~~ fox ~~jumps~~ over the ~~lazy~~ ~~dog~~

fast	hazel	leaps	idle
swift	brunette	springs	indolent
speedy	chestnut	bounds	slothful
		hops	hound
			pup
			mutt

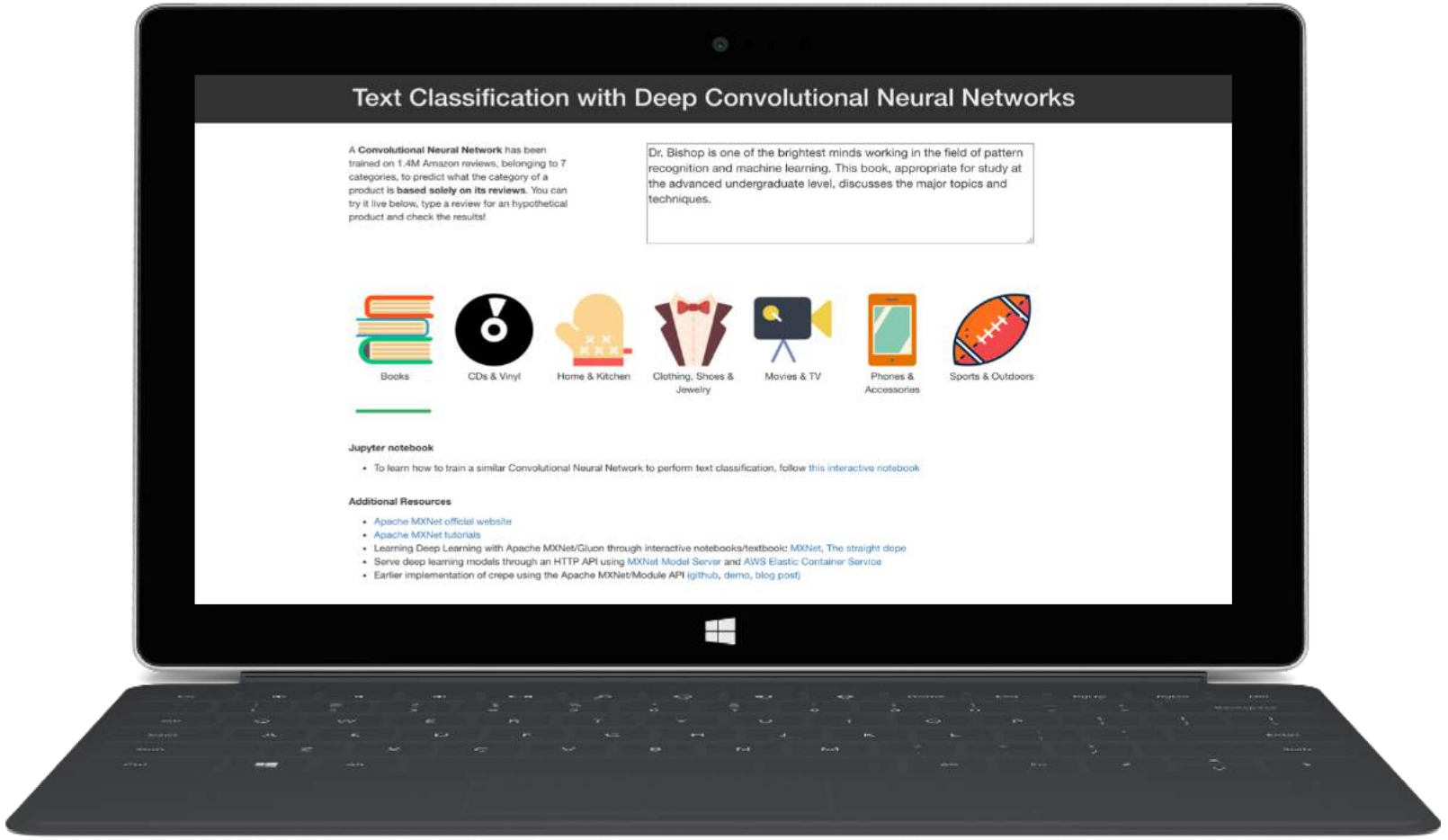
The swift brunette fox leaps over the slothful pup

You need a large dataset



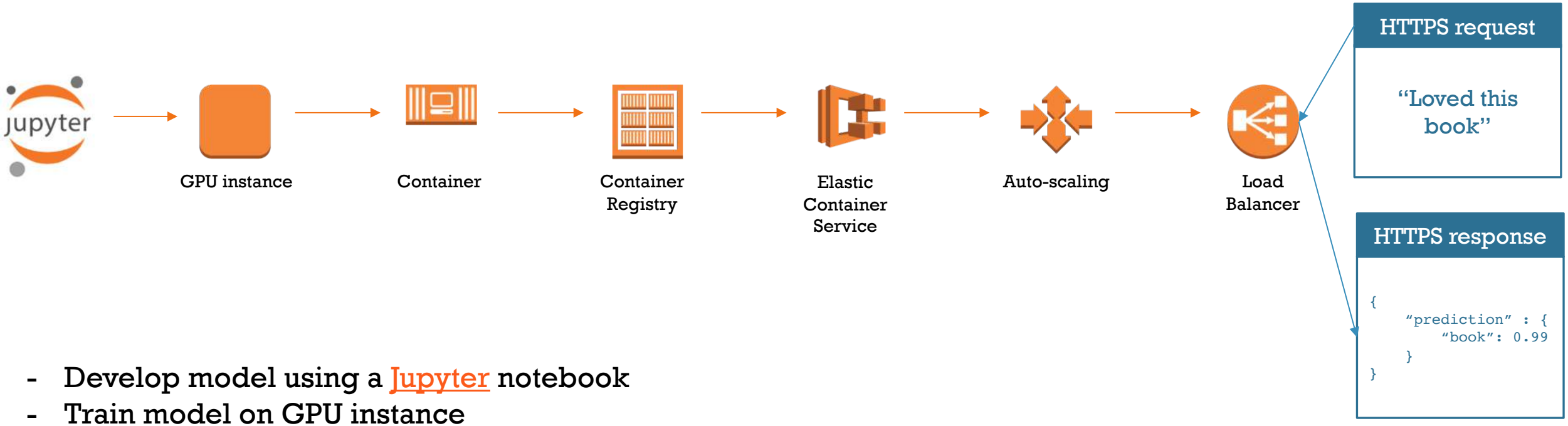
…A very large dataset!

Live Demo – Classification of product category for Amazon Reviews



https://thomasdelteil.github.io/CNN_NLP_MXNet/

Workflow and Operationalization

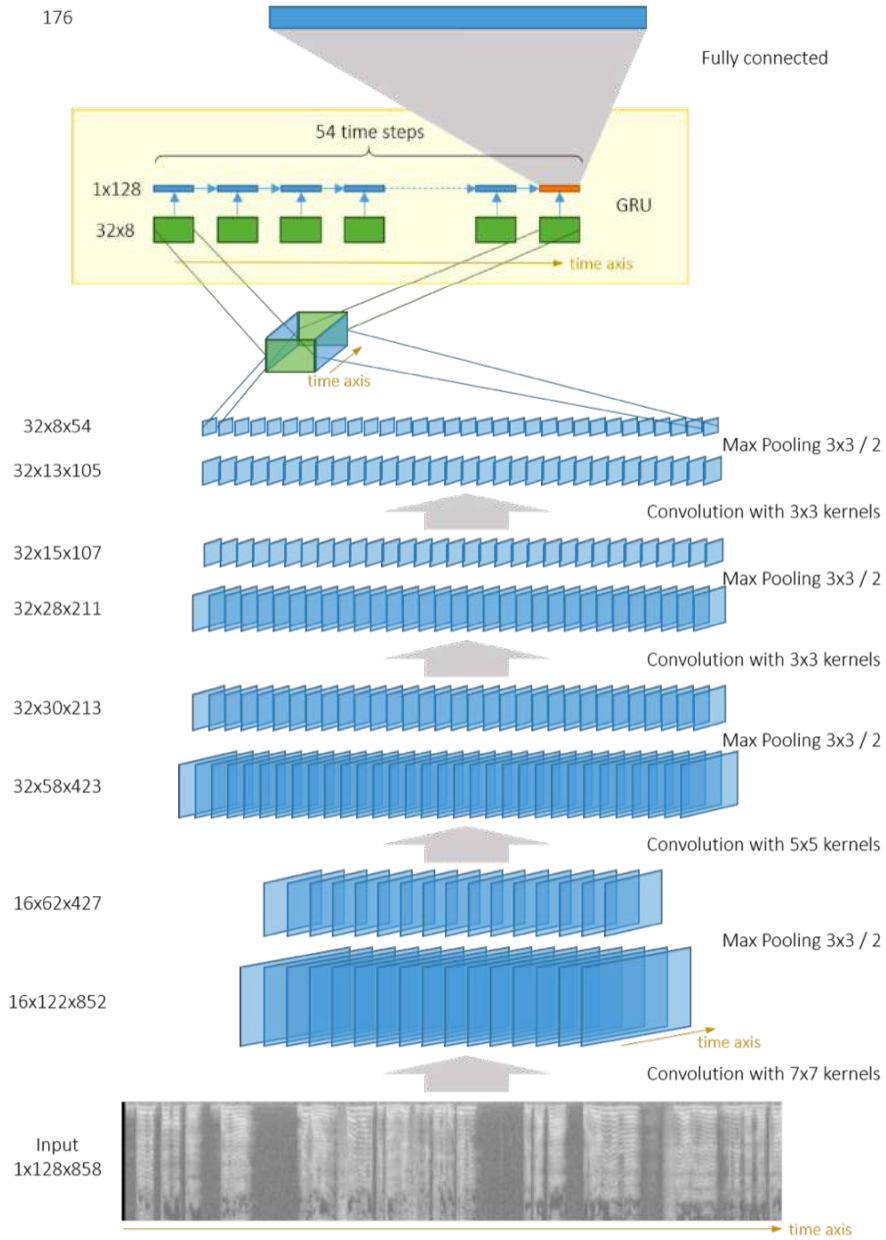


- Develop model using a Jupyter notebook
- Train model on GPU instance
- Package model behind web API in a Docker container, e.g using MXNet Model Server
- Upload container to container registry
- Deploy container to an elastic container service
- Enjoy quick and linear scaling
- Put the API behind a load balancer with SSL termination
- Enjoy ☺

Advanced use-cases for Convolutions and NLP

CNN + LSTM: Spatially and Temporally Deep Neural Networks

- CNN for feature extraction
- LSTM for temporal representation

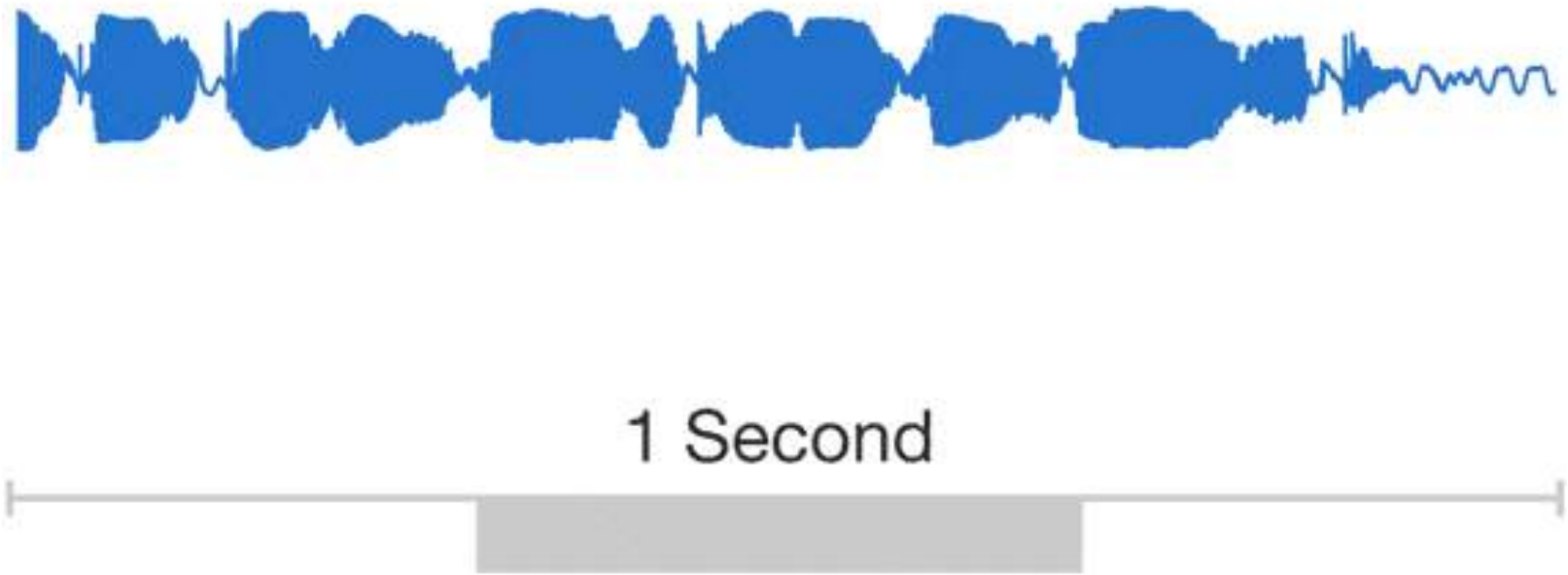


Applications:

- Video (CNN for frames, LSTM to combine them temporally)
- Text tasks
- Audio (Language detection)

Source: Combining CNN and RNN for spoken language detection

Advanced use-case: Speech Generation WaveNet



Source: DeepMind Wavenet generative model raw audio

WaveNet: Dilated Causal Convolution

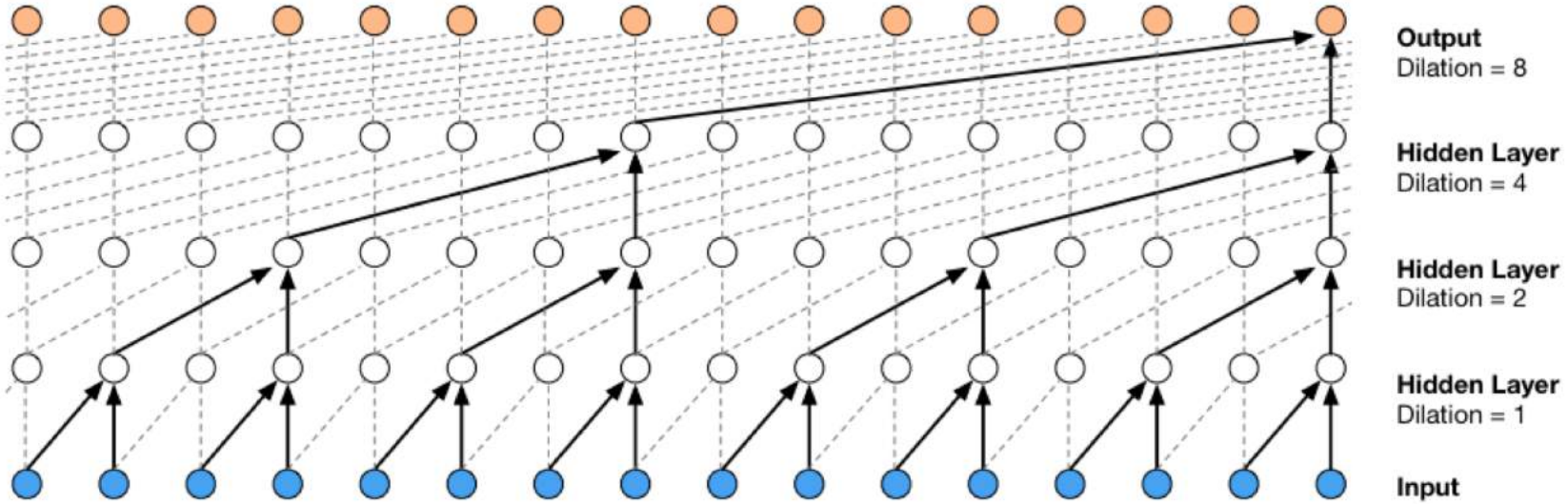
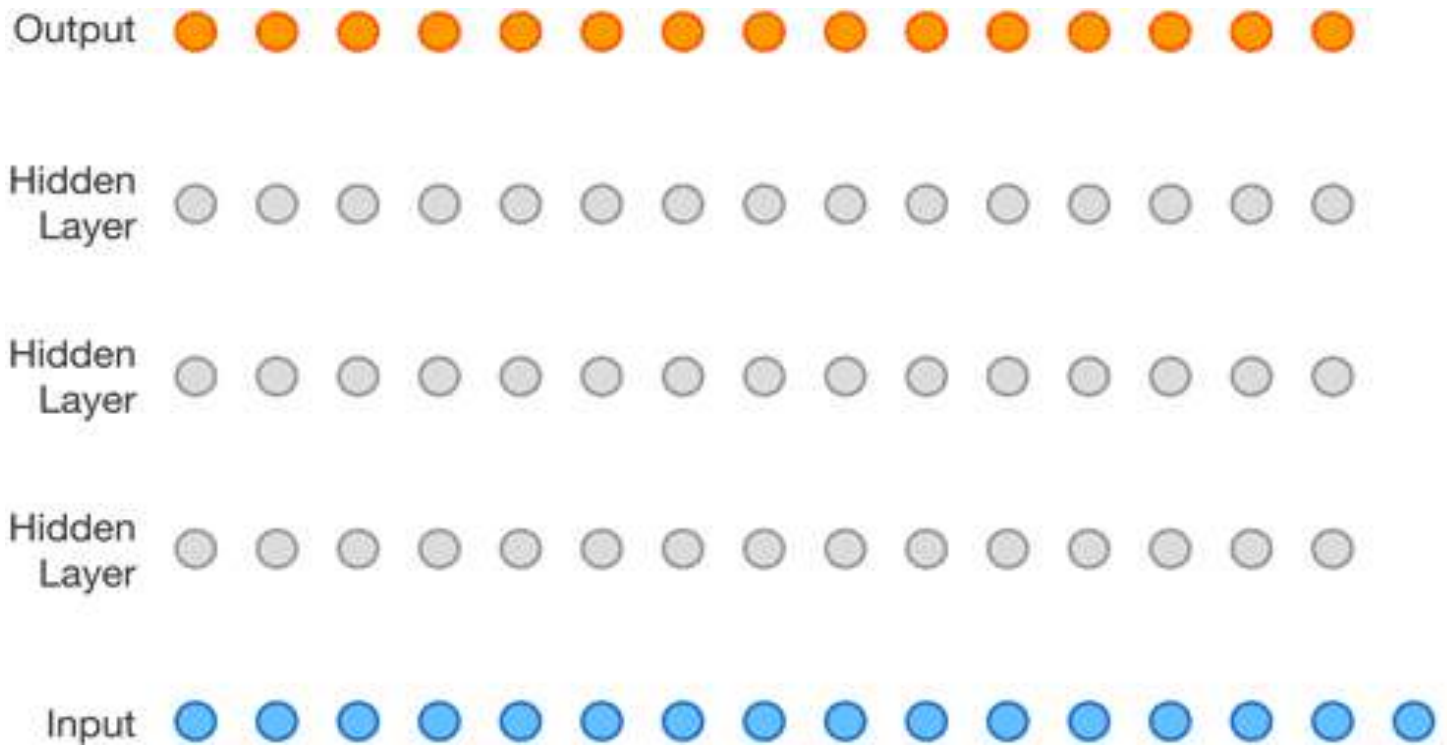


Figure 3: Visualization of a stack of *dilated* causal convolutional layers.

[Source: DeepMind Wavenet generative model raw audio]

WaveNet: Dilated Causal Convolution



Source: DeepMind Wavenet generative model raw audio

Summary

- Learned about convolutions
- Applied them to textual data
- Studied the crepe architecture from Zhang et al. in details
- Learned about advanced use cases and operationalization

Thank you!

Connect here
github.com/thomasdelteil
linkedin.com/in/thomasdelteil
tdelteil@amazon.com

Photos credits: https://pexels.com and https://unsplash.com/