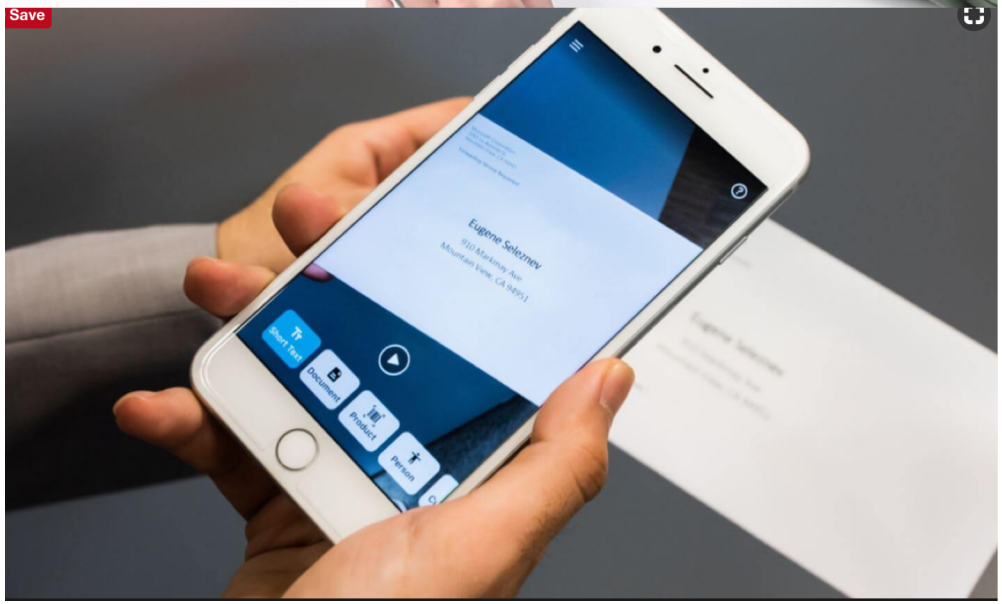
Apache MXNet (Incubating)

Steffen Rochel
steroche@amazon.com; @srochel



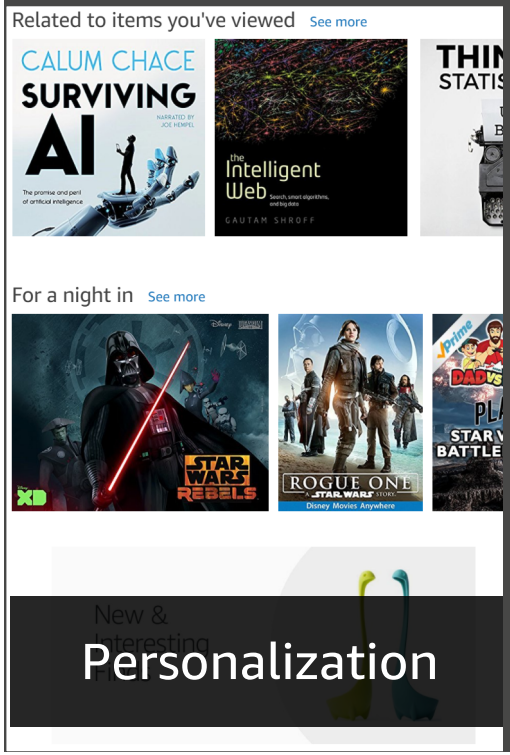
Why Amazon's Alexa Is 'Life Changing' for the Blind

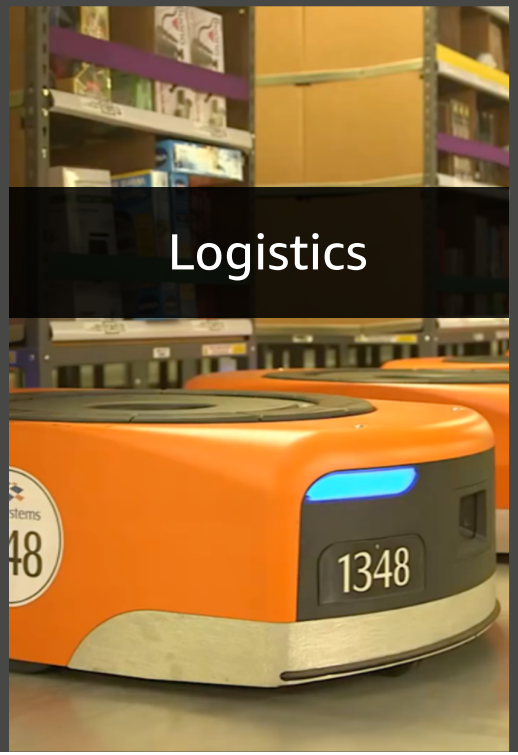
For the blind, Amazon's Echo and Alexa is more than just neat technology; it's a lifeline.

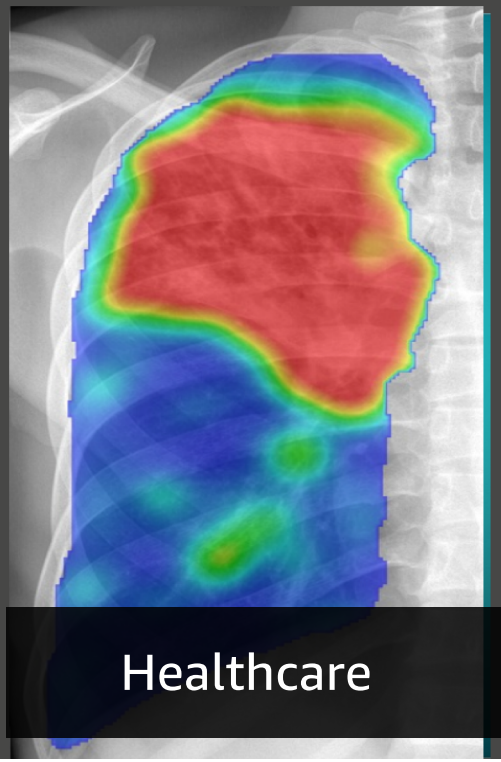
By Jon Kalish January 8, 2018 8:00AM EST

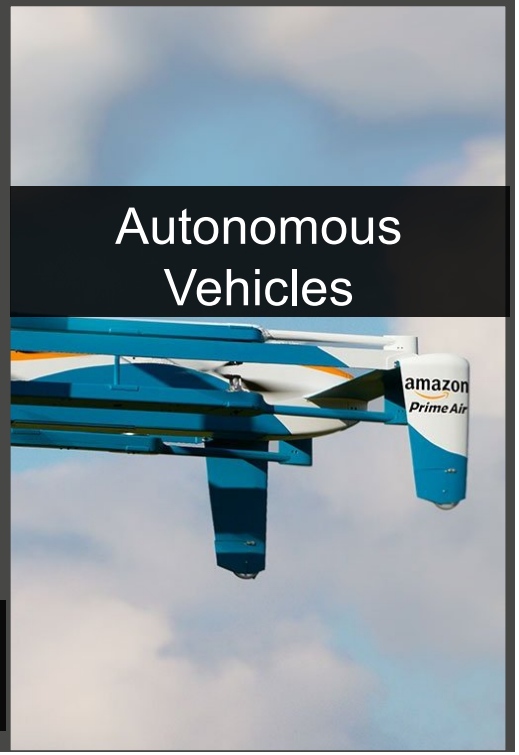
2.7K SHARES

Why Deep Learning?

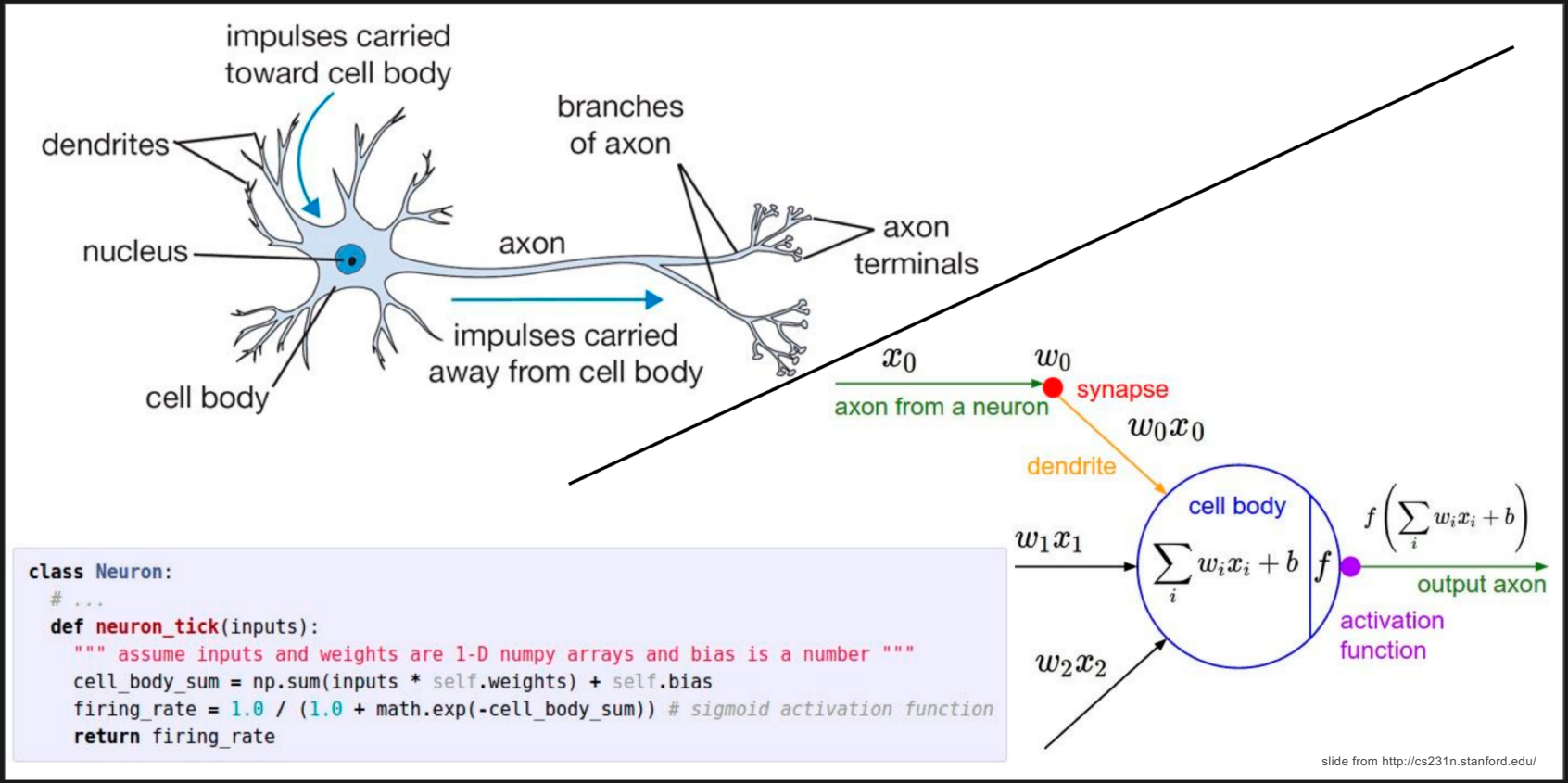








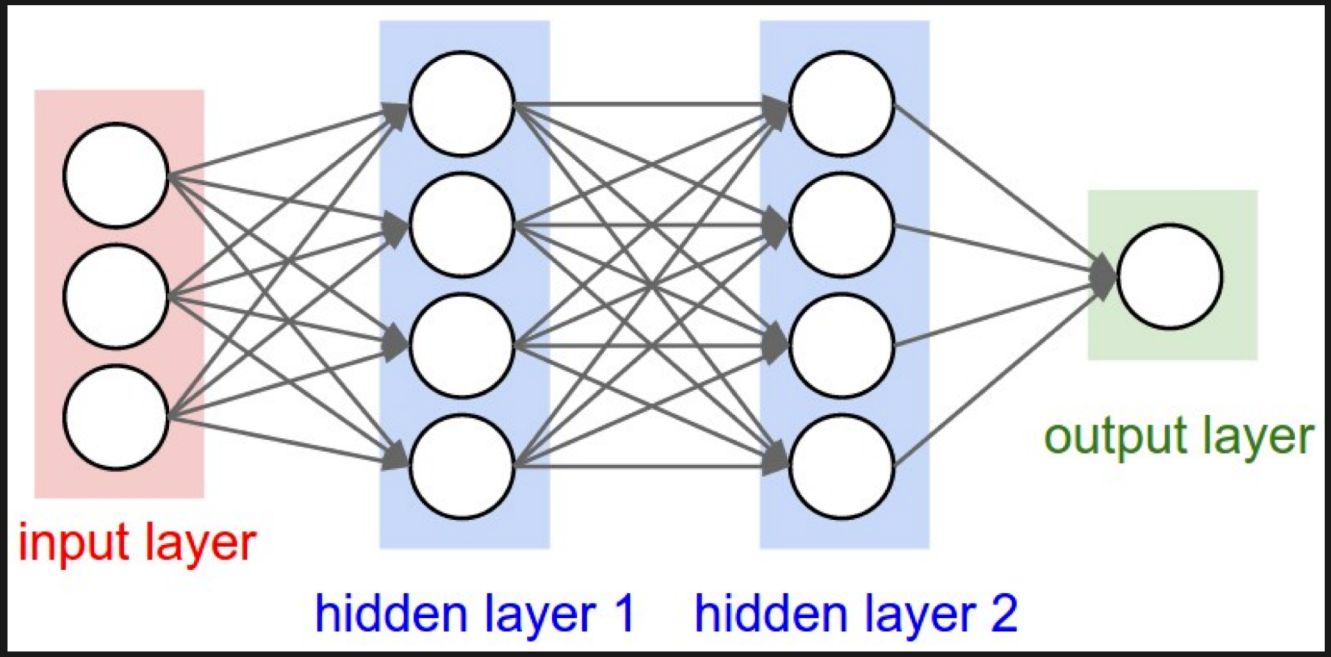
Biological & Artificial Neuron




```
class Neuron:
  # ...
  def neuron_tick(inputs):
    """ assume inputs and weights are 1-D numpy arrays and bias is a number """
    cell_body_sum = np.sum(inputs * self.weights) + self.bias
    firing_rate = 1.0 / (1.0 + math.exp(-cell_body_sum)) # sigmoid activation function
    return firing_rate
```

Source: http://cs231n.github.io/neural-networks-1/

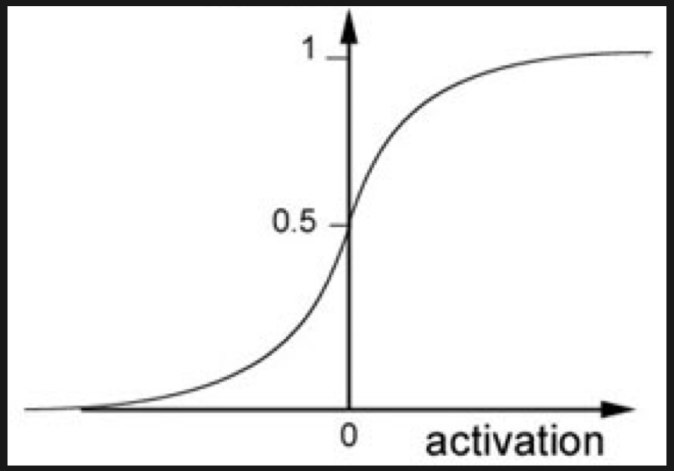
Fully Connected Layer

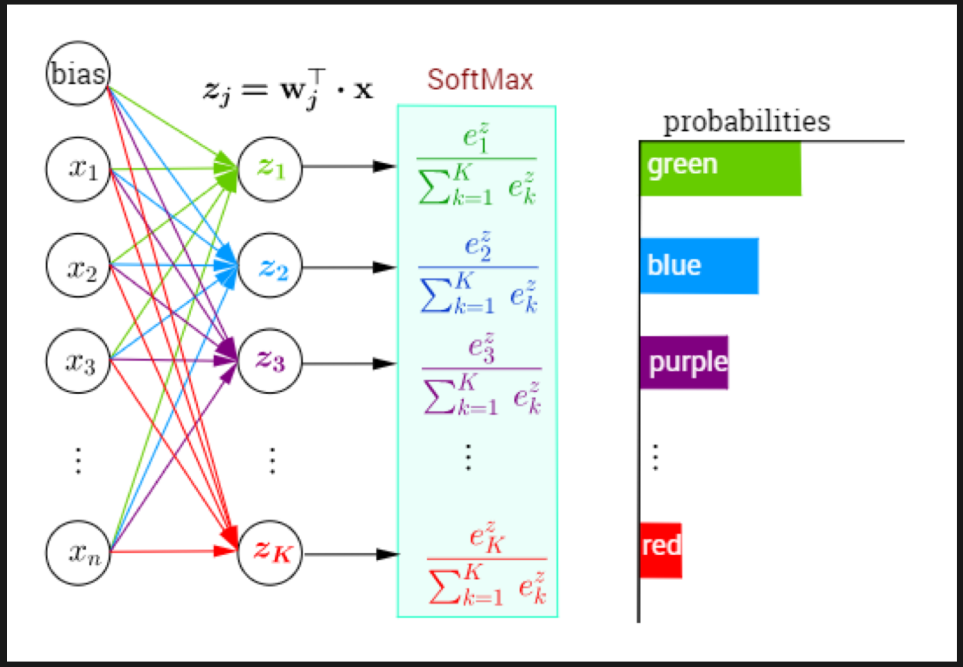


Each node (“neuron”) in a layer is connected to every node in the previous layer

Classification with the Softmax Function

Softmax Function

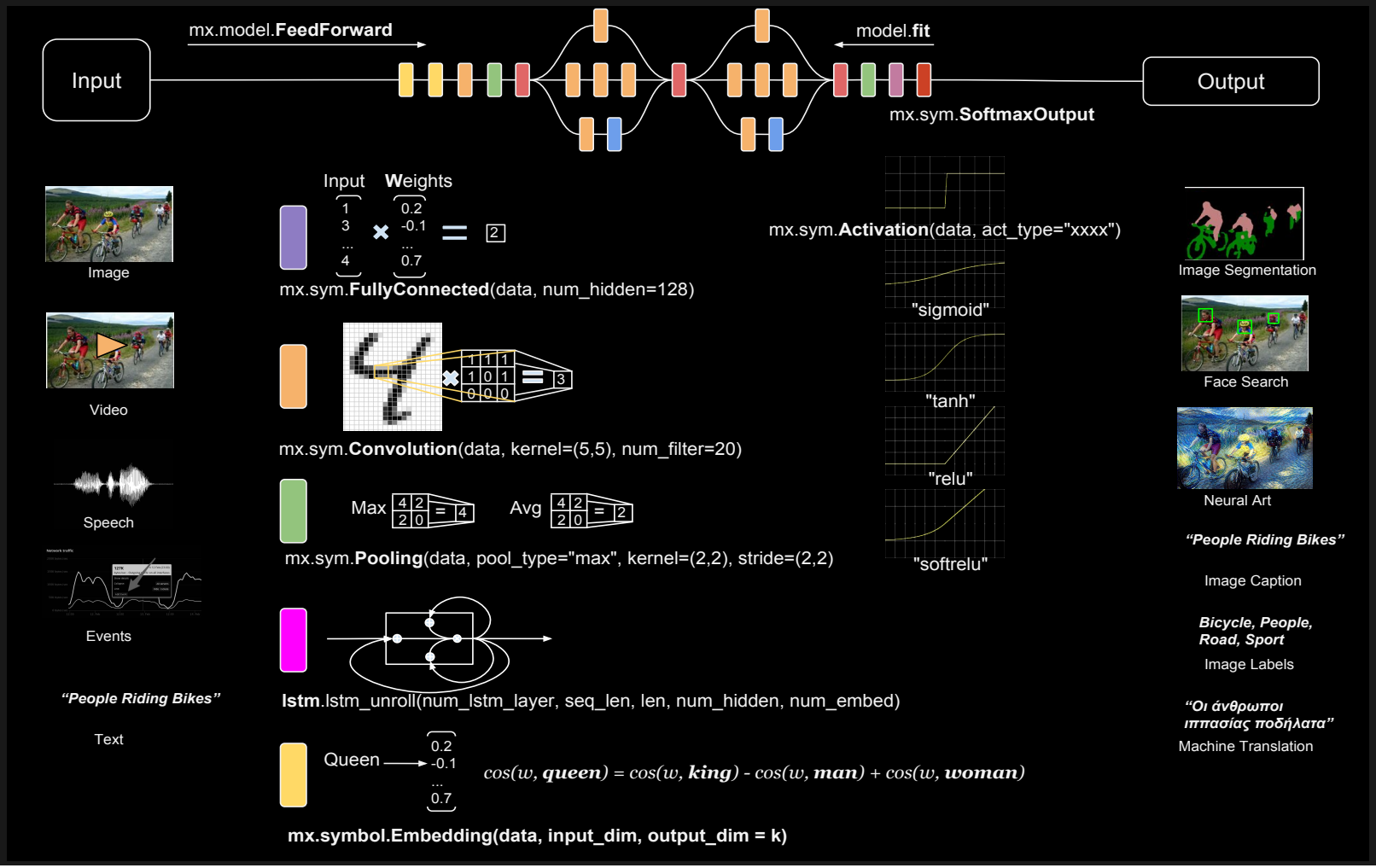




Softmax converts the output layer into probabilities – necessary for classification

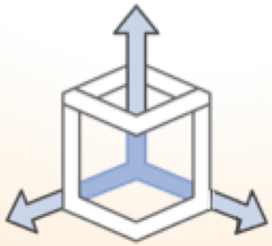
Source: https://stats.stackexchange.com/questions/273465/neural-network-softmax-activation

Deep Learning Models

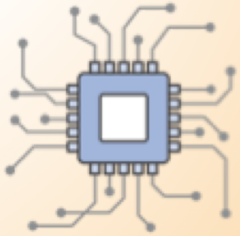


MXNet – a Deep Learning Framework

Programmable
Simple Syntax
Imperative/Declarative
Multiple languages



Portable
Highly efficient models
for Mobile and IOT



High Performance
Near linear scaling across
hundreds of GPUs

Open Source
Incubating at Apache

ONNX Support

Easily and quickly build high
performance models with
Imperative APIs

MXNet EcoSystem

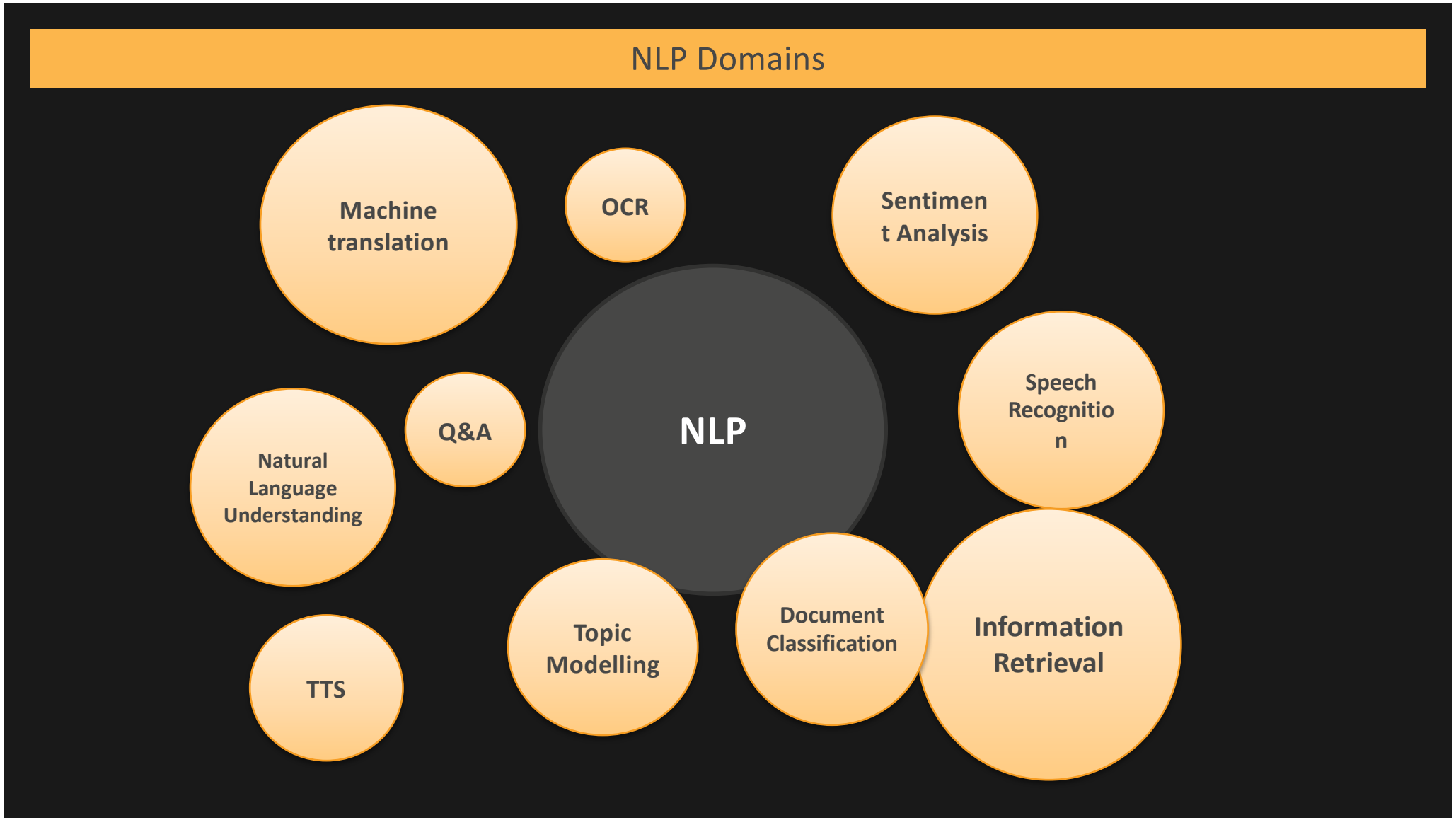
- Gluon Model Zoo
  https://mxnet.incubator.apache.org/api/python/gluon/model_zoo.html
- Sockeye: A Toolkit for Neural Machine Translation
  https://arxiv.org/abs/1712.05690
- GluonCV: A Deep Learning Toolkit for Computer Vision
  https://gluon-cv.mxnet.io/
- GluonNLP: a Deep Learning Toolkit for Natural Language Processing
  http://gluon-nlp.mxnet.io/
- DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks
  https://arxiv.org/abs/1704.04110v2
- MXNet Model Server
  https://github.com/awslabs/mxnet-model-server

Apache MXNet: Deep Learning for Enterprise Developers

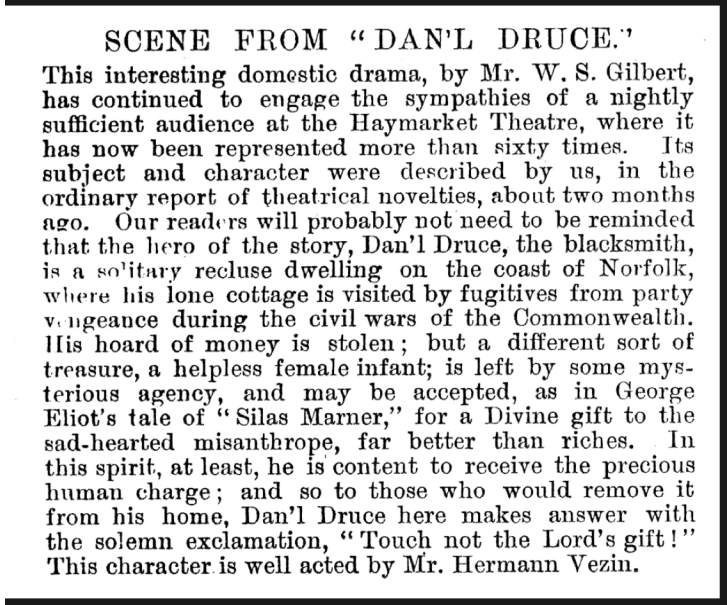
- Start with pre-trained high-quality models
  - Gluon CV and Gluon NLP
- Refine with fast and scalable training
  - Keras-MXNet up to 2x faster than Keras-TensorFlow
  - Near-linear scalability up to 256 GPUs
  - Dynamic Training
- Deploy using familiar tools
  - Java/Scala APIs
  - MXNet Model Server

MXNET.IO

Steffen Rochel
@srochel

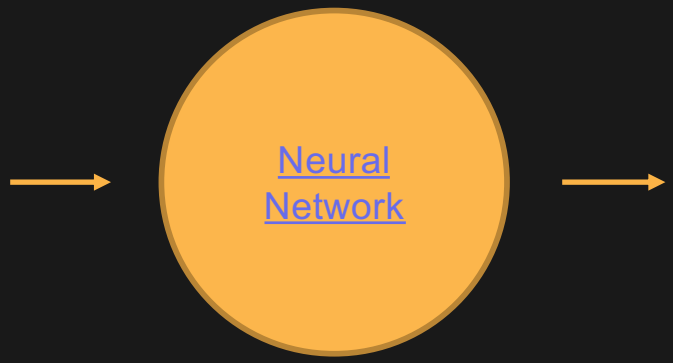

NLP Domains
Machine translation
OCR
Sentimen t Analysis
Speech Recognitio n
Q&A
NLP
Natural Language Understanding
Topic Modelling
Document Classification
Information Retrieval
TTS

Text classification, N categories



SCENE FROM "DAN'L DRUCE."

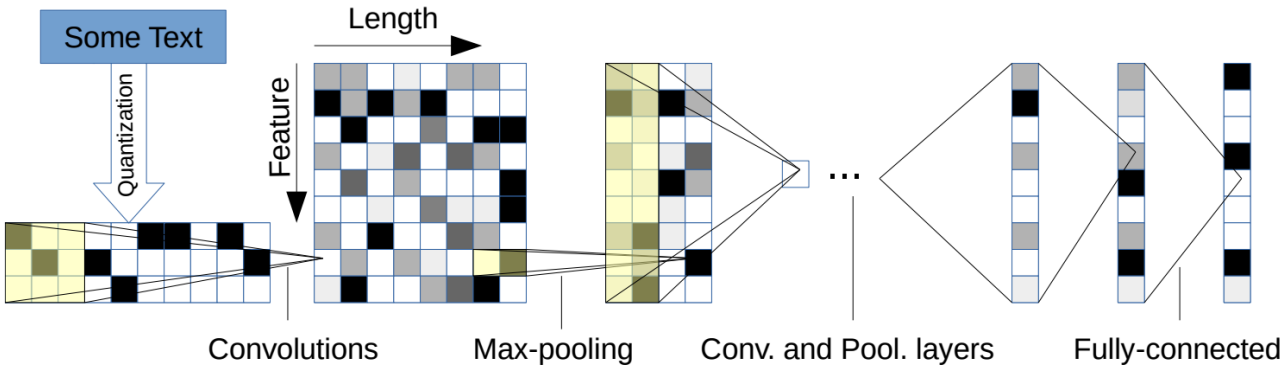
This interesting domestic drama, by Mr. W. S. Gilbert, has continued to engage the sympathies of a nightly sufficient audience at the Haymarket Theatre, where it has now been represented more than sixty times. Its subject and character were described by us, in the ordinary report of theatrical novelties, 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.



- Fiction: 0%
- Biography: 6%
…
- **Play: 80%**
…
- Documentation: 0%

Deep Neural Network: Crepe Model

Visualization with Netron
Intuition: convolutions act similarly as n-grams



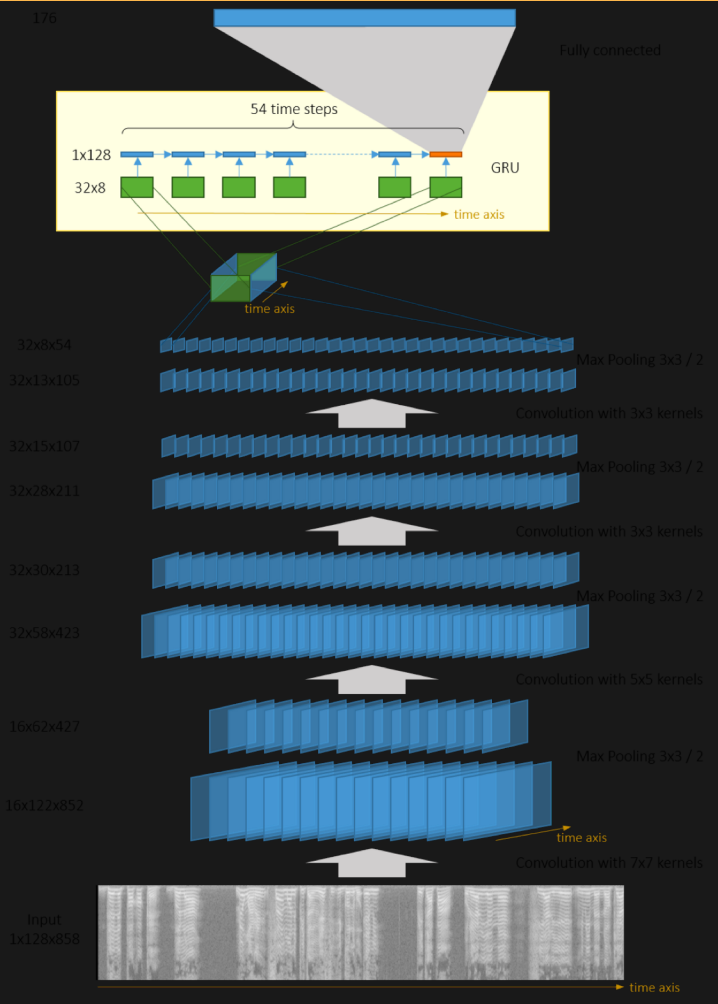
Layer	Large Feature	Small Feature	Kernel	Pool
1	1024	256	7	3
2	1024	256	7	3
3	1024	256	3	N/A
4	1024	256	3	N/A
5	1024	256	3	N/A
6	1024	256	3	3

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

CNN + LSTM: Spatially and Temporally Deep Neural Networks

- CNN for feature extraction
- LSTM for temporal representation

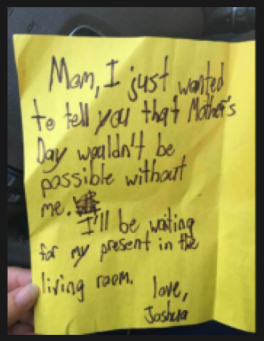


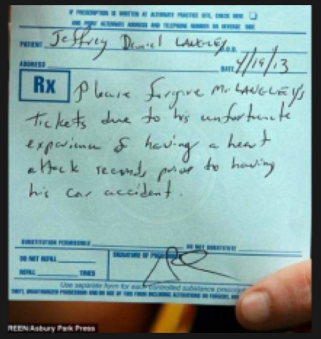
Applications:

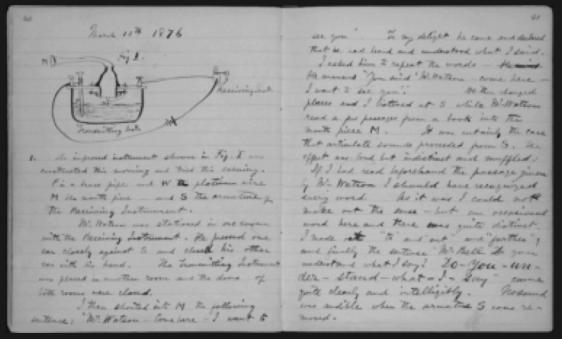
- Video (CNN for frames, LSTM to combine them temporally)
- Text tasks
- Audio (Language detection)

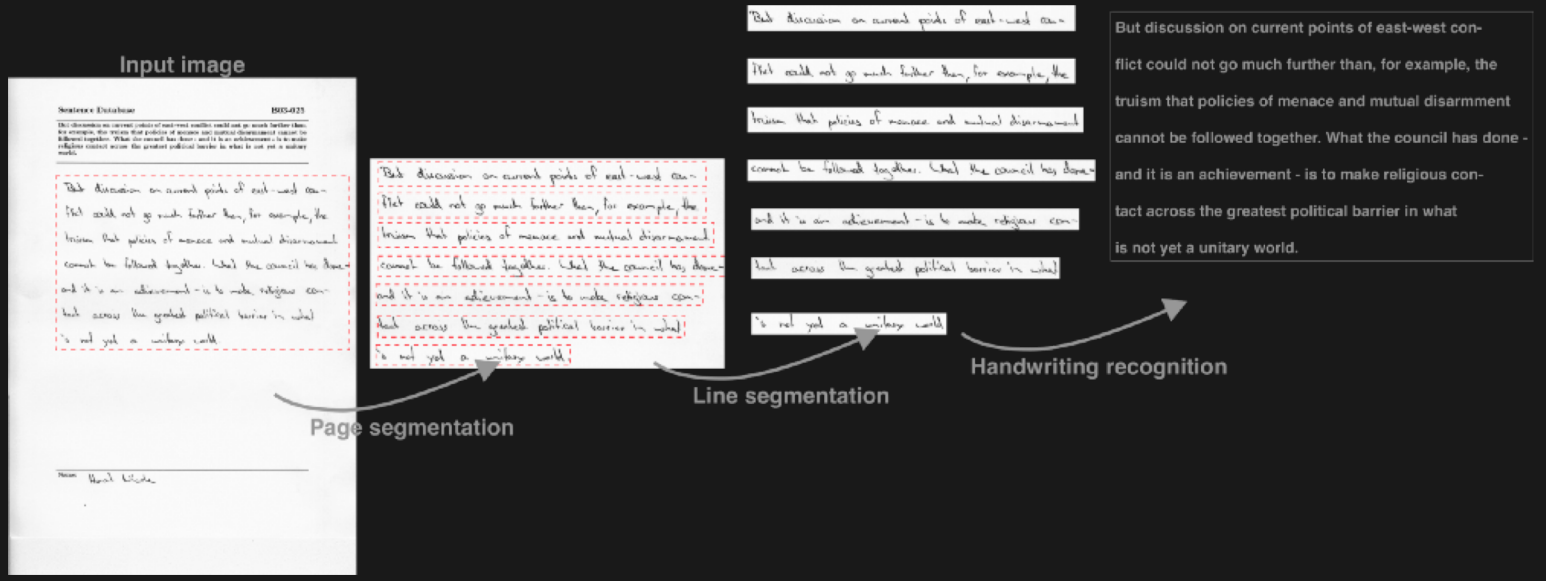
Source: Combining CNN and RNN for spoken language detection

Handwriting OCR










Thank you!
Connect here
github.com/srochel
linkedin.com/in/steffenrochel
steroche@amazon.com
mxnet-info@amazon.com
Photos credits: https://pexels.com and https://unsplash.com/