Machine Learning

«A Gentle Introduction»

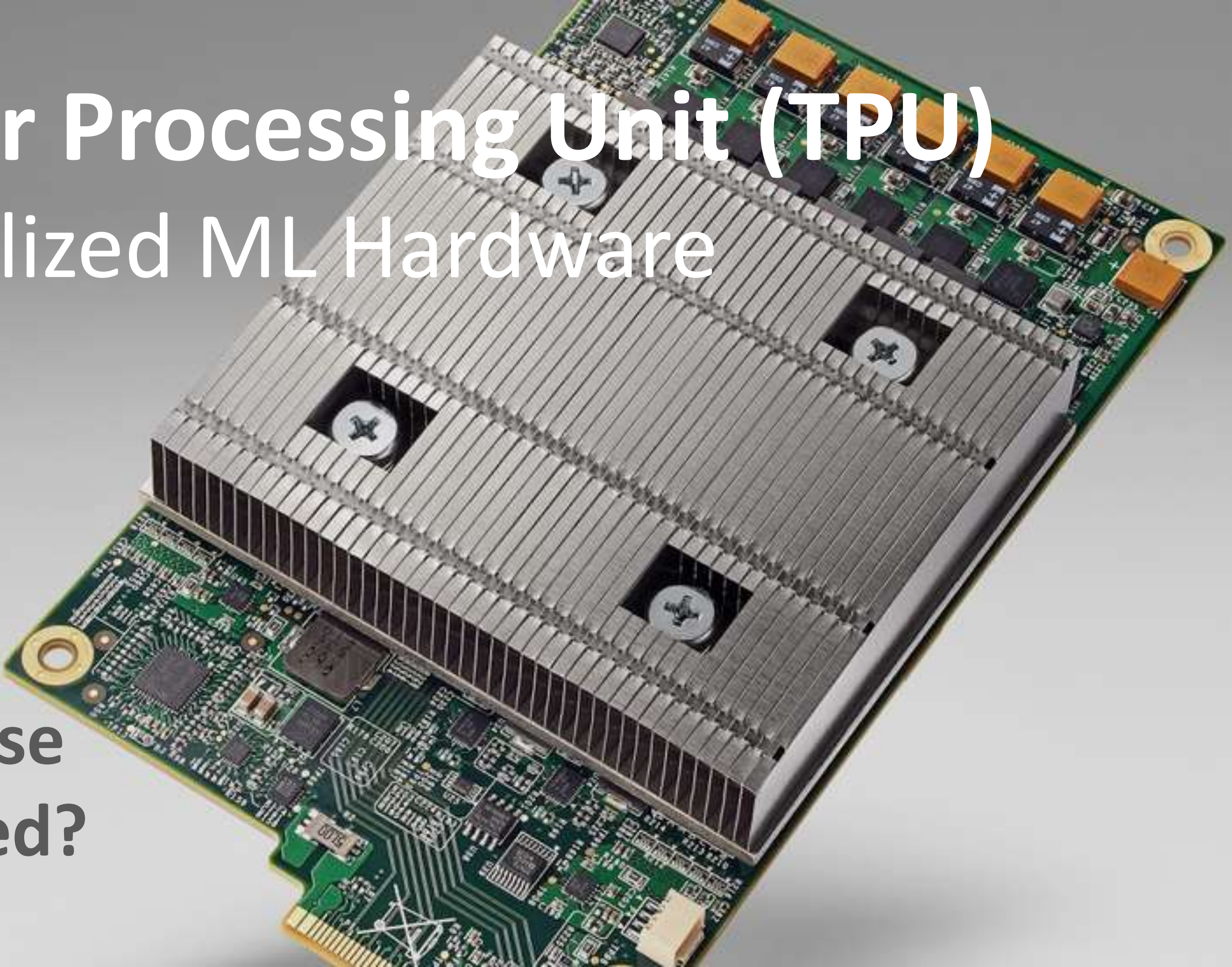
@ZimMatthias Matthias Zimmermann
BSI Business Systems Integration AG

«What is the difference to IBM’s chess system 20 years ago?»

AlphaGo Hardware
Powered by TPUs

Tensor Processing Unit (TPU)

Specialized ML Hardware



What else is needed?

nature International weekly journal of science

NATURE | LETTER

日本語要約

Human-level control through deep r learning

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjelar Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, He Kumaran, Daan Wierstra, Shane Legg & Demis Hassabis

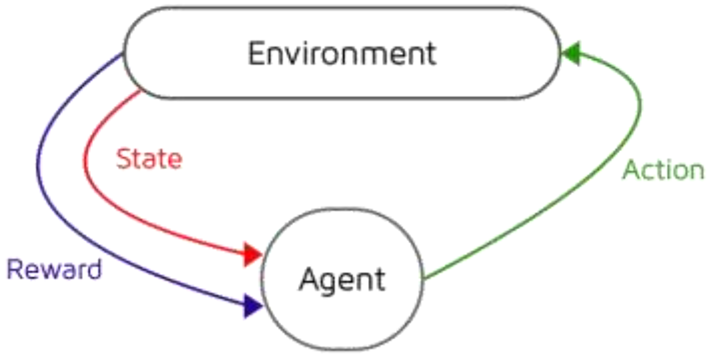
Affiliations | Contributions | Corresponding authors

Nature **518**, 529–533 (26 February 2015) | doi:10.1038/nature14236
Received 10 July 2014 | Accepted 16 January 2015 | Published online 25 February 2015

Deep Reinforcement Learning

Markov Decision Process

- ➔ **Environment** (Atari Breakout)
- ➔ **Agent** performing **Actions** (Left, Right, Release Ball)
- ➔ **State** (Bricks, location / direction of ball, …)
- ➔ **Rewards** (A Brick is hit)



Deep Reinforcement Learning

Q-Learning (simplified)

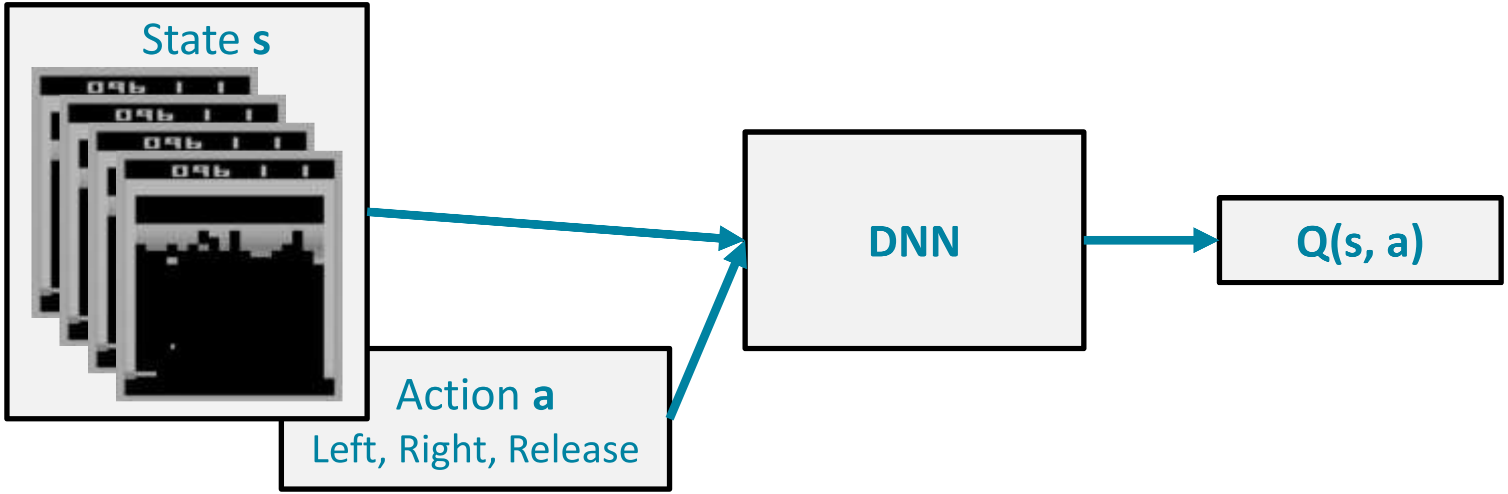
- ➔ **Markov Decision Process**
- ➔ Q(s, a) Highest sum of future **Rewards** for action a in state s

```
initialize Q randomly
set initial state s_0
repeat
   execute a to maximize Q(s_i, a)
   observe r and new state s_i+1
   set Q = update(Q, r, a, s_i+1)
   set s_i = s_i+1
until terminated
```

Deep Reinforcement Learning

Deep Q Learning (DQN)

- ➔ Q Learning
- ➔ Q(s, a) = **Deep Neural Network** (DNN)
- ➔ Retrain DNN regularly (using it’s own experience)



Machine Learning Concepts

Data
Models
Training and Evaluation
ML Topics

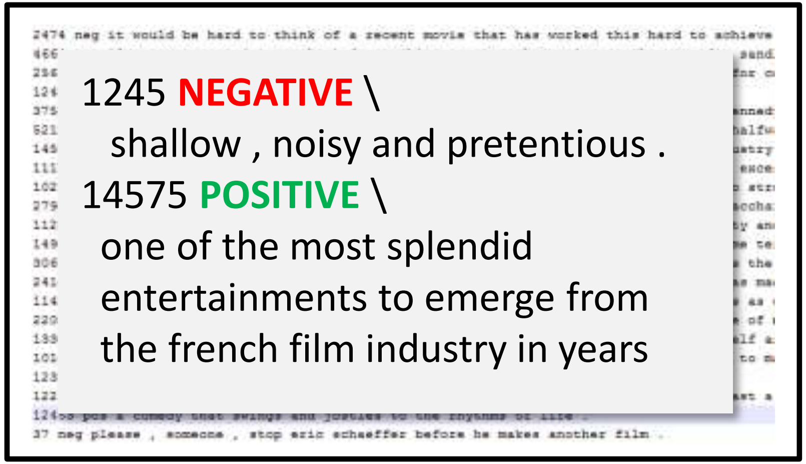
Getting the Data

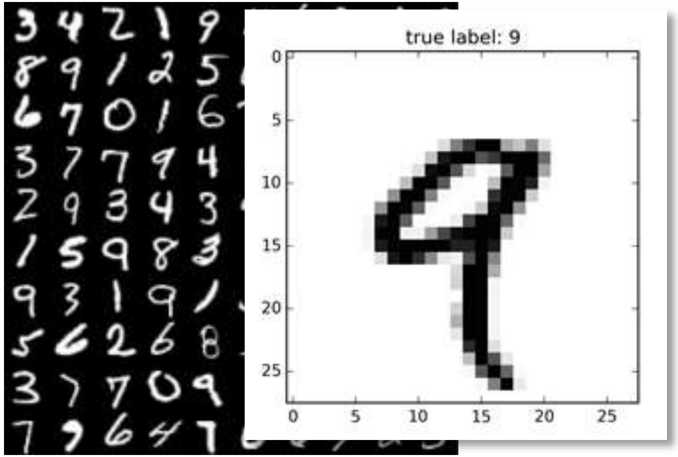
Challenges

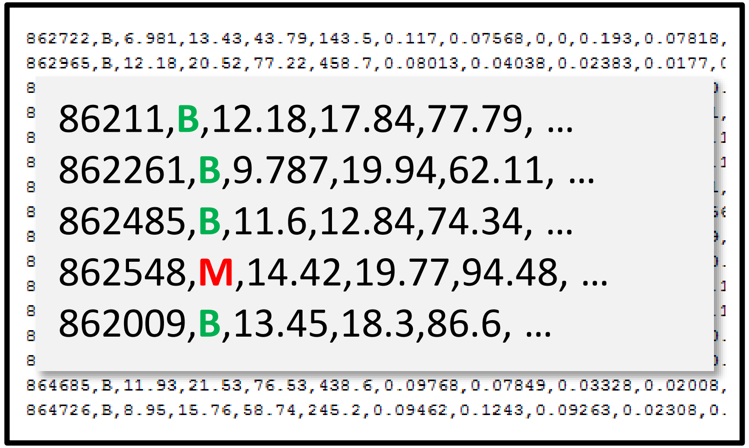
→ Getting the **RIGHT** data for the task
→ And **LOTs** of it
→ There is never enough data …

Real World Lessons

→ Data is crucial for successful ML projects
→ Most boring and timeconsuming task
→ Most underestimated task

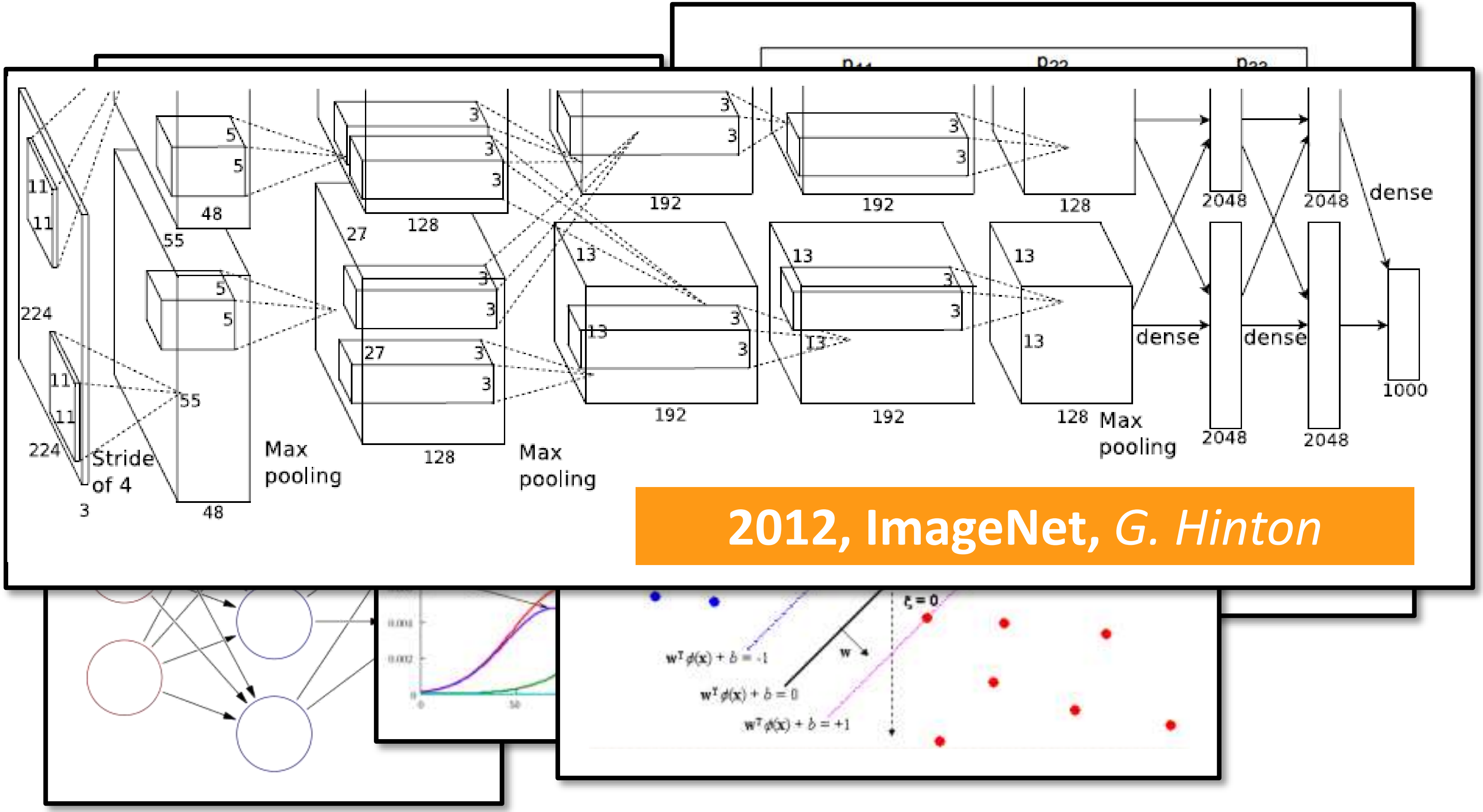

2474 neg it would be hard to think of a recent movie that has worked this hard to achieve
1245 NEGATIVE \
shallow , noisy and pretentious .
14575 POSITIVE \
one of the most splendid
entertainments to emerge from
the french film industry in years
37 neg please , someone , stop eric schaeffer before he makes another film .


true label: 9
0
5
10
15
20
25
0
5
10
15
20
25


862722,B,6.981,13.43,43.79,143.5,0.117,0.07568,0,0,0.193,0.07818,
862965,B,12.18,20.52,77.22,458.7,0.08013,0.04038,0.02383,0.0177,
86211,B,12.18,17.84,77.79, …
862261,B,9.787,19.94,62.11, …
862485,B,11.6,12.84,74.34, …
862548,M,14.42,19.77,94.48, …
862009,B,13.45,18.3,86.6, …
864685,B,11.93,21.53,76.53,438.6,0.09768,0.07849,0.03328,0.02008,
864726,B,8.95,15.76,58.74,245.2,0.09462,0.1243,0.09263,0.02308,0.

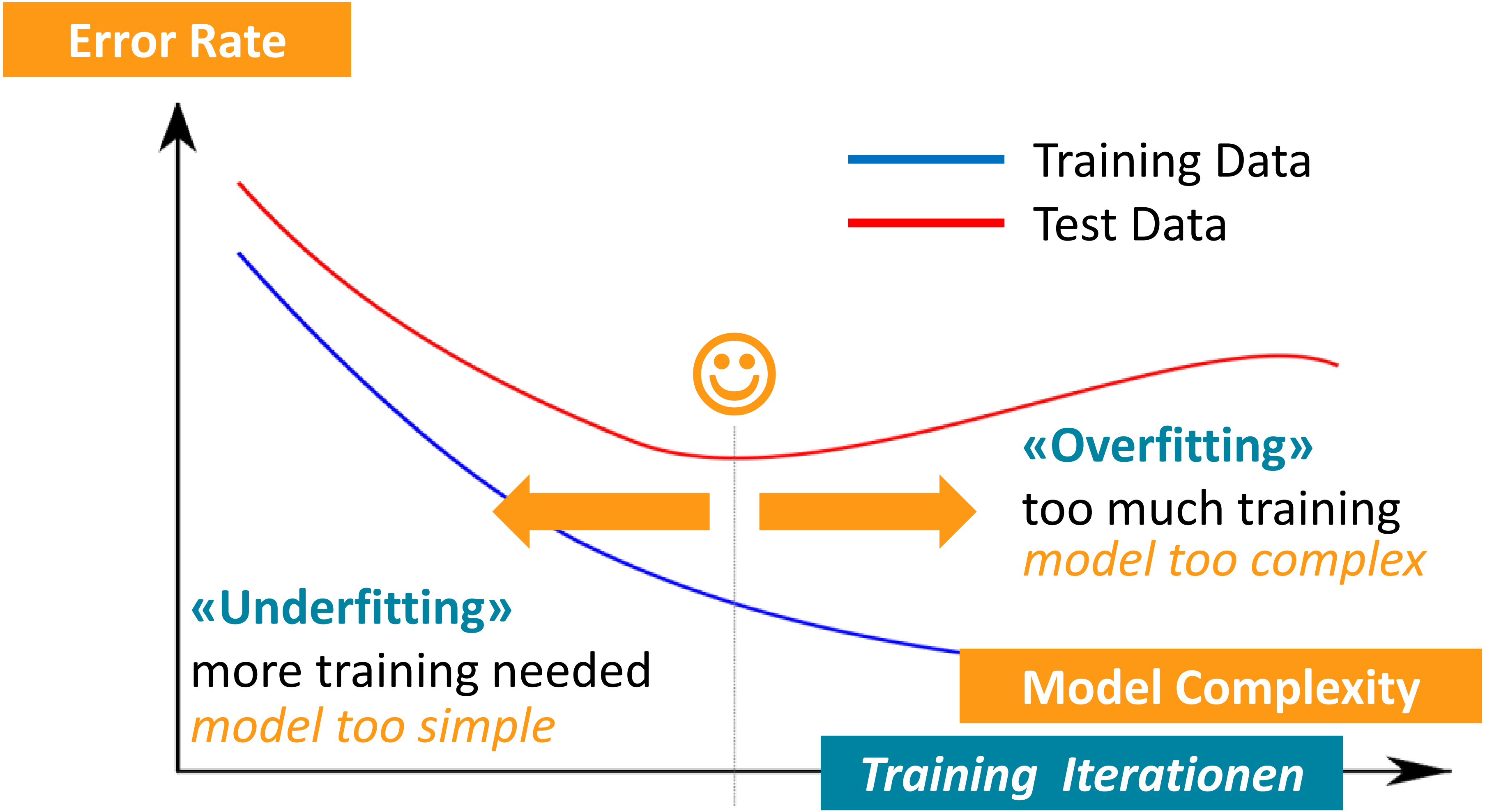

Rosemary, Rosmarinus officinalis

Data
Models
Training and Evaluation
ML Topics



224
224
11
11
Stride of 4
3
55
48
5
Max pooling
27
128
3
Max pooling
13
192
128
Max pooling
dense
2048
1000
$w^T\phi(x) + b = -1$
$w^T\phi(x) + b = 0$
$w^T\phi(x) + b = +1$
w
$\xi = 0$
2012, ImageNet, *G. Hinton*

Data
Models
Training and Evaluation
ML Topics


Error Rate
Training Data
Test Data
«Overfitting»
too much training
model too complex
«Underfitting»
more training needed
model too simple
Model Complexity
Training Iterationen

Data
Models
Training and Evaluation
ML Topics

CS 229
Machine Learning
Handout #2: Course Schedule

Syllabus

- **Introduction** (1 class)
  Basic concepts.

- **Supervised learning.** (7 classes)
  Supervised learning setup. LMS.
  Logistic regression. Perceptron. Exponential family.
  Generative learning algorithms. Gaussian discriminant analysis. Naive Bay
  Support vector machines.
  Model selection and feature selection.
  Ensemble methods: Bagging, boosting.
  Evaluating and debugging learning algorithms.

- **Learning theory.** (3 classes)
  Bias/variance tradeoff. Union and Chernoff/Hoeffding bounds.
  VC dimension. Worst case (online) learning.
  Practical advice on how to use learning algorithms.

- **Unsupervised learning.** (5 classes)
  Clustering. K-means.
  EM. Mixture of Gaussians.
  Factor analysis.
  PCA (Principal components analysis).
  ICA (Independent components analysis).

- **Reinforcement learning and control.** (4 classes)
  MDPs. Bellman equations.
  Value iteration and policy iteration.
  Linear quadratic regulation (LQR). LQG.
  Q-learning. Value function approximation.
  Policy search. Reinforce. POMDPs.

Supervised Learning

- Learning from Examples
- Right Answers are known

Unsupervised Learning

- Discover Structure in Data
- Dimensionality Reduction

Reinforcement Learning

- Interaction with Dynamic Environment

Demo Time

Demos

Demo 1 Supervised Learning

- → **Pattern recognition**
- → Handwritten character recognition
- → Convolutional neural network

Demo 2 Unsupervised Learning

- → **Natural language processing** (NLP)
- → Neural word embeddings
- → Word2vec

Pattern Recognition
Handwritten Digits

Data

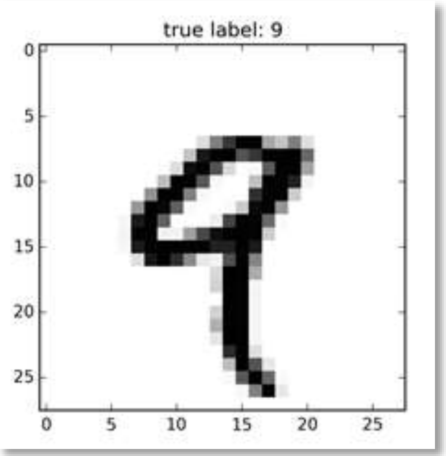
- ➔ Which digit is this?
- ➔ Collect our own data

Model

- ➔ Deep Neural Network (LeNet-5)

Deeplearning4j

- ➔ Deep Learning Library
- ➔ Open Source (Apache)
- ➔ Java



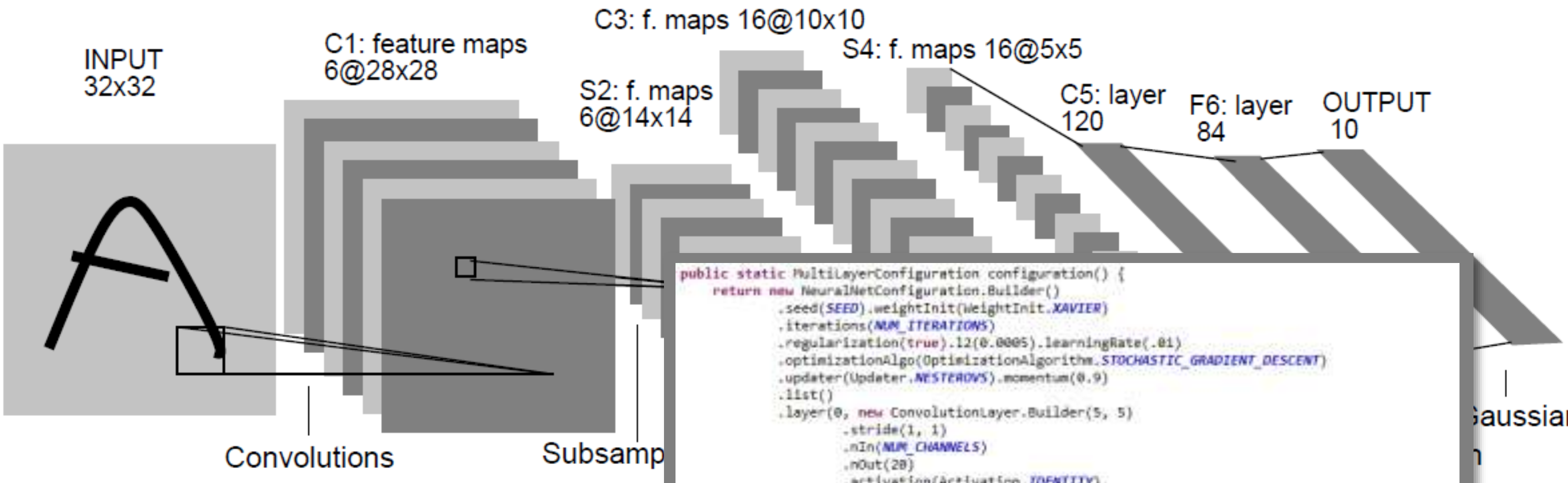
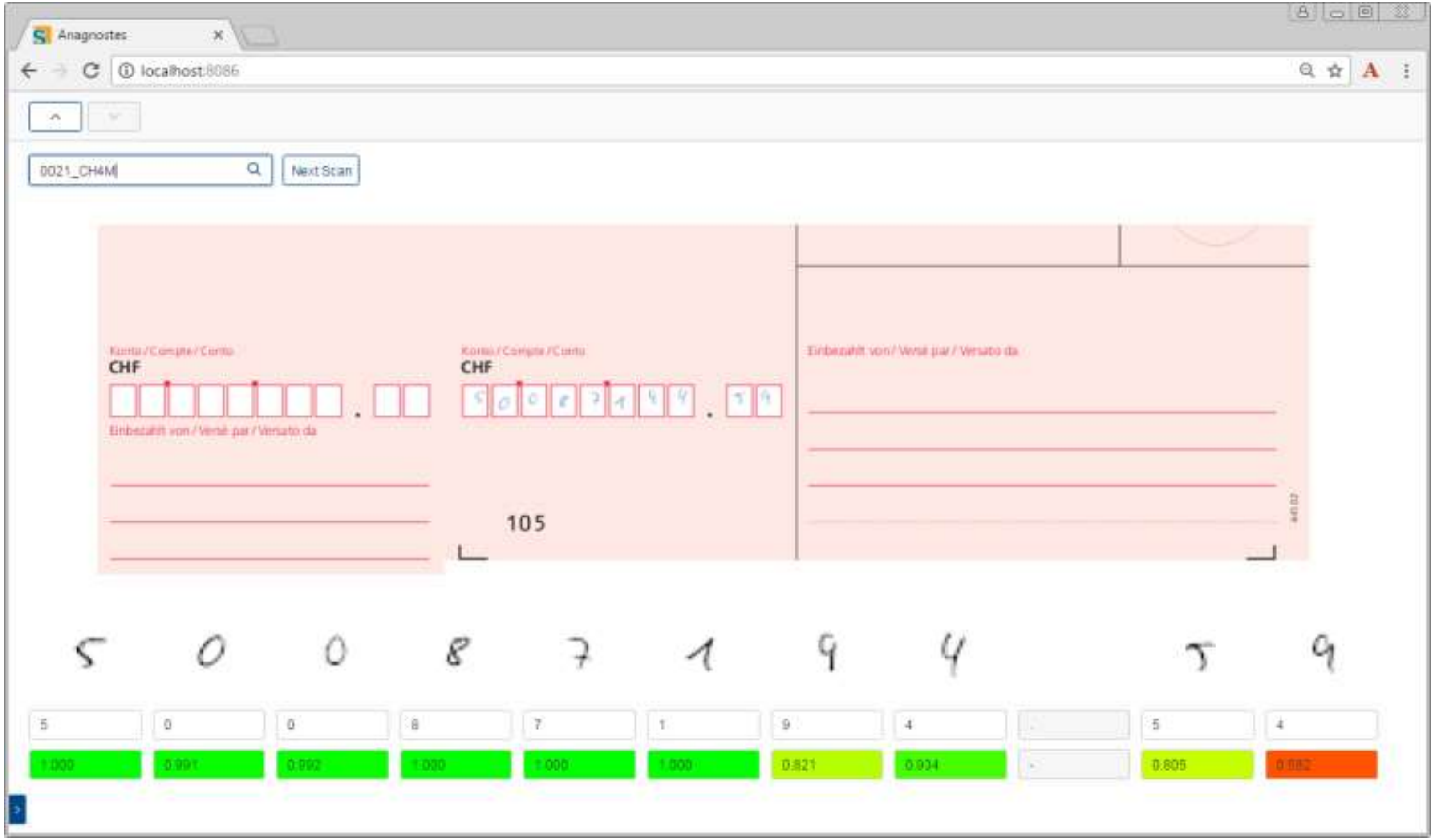
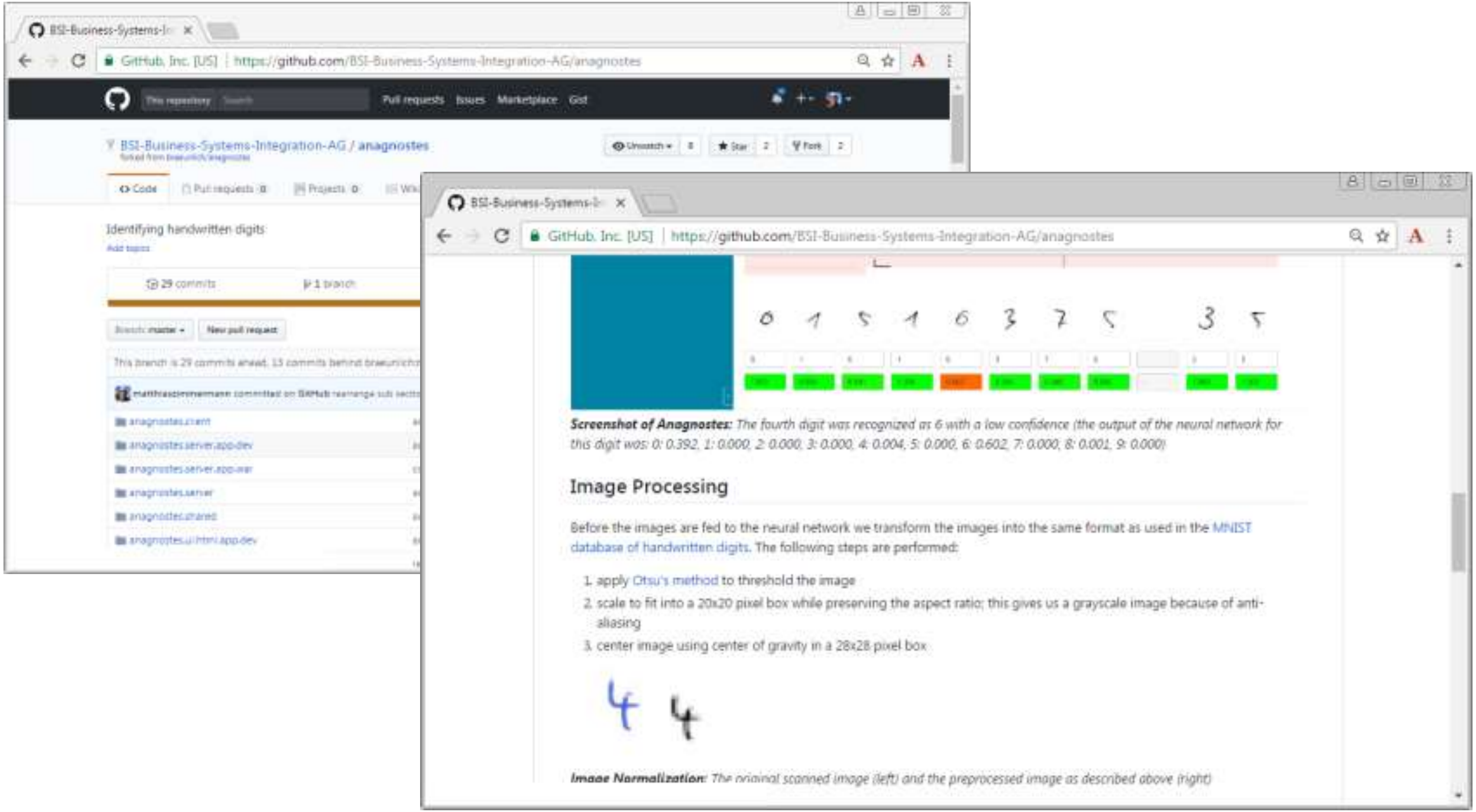


Fig. 2. Architecture of LeNet-5, a Convolutional Neu … map, i.e. a set of units whose weights are constrained to be identical.

```
public static MultiLayerConfiguration configuration() {
    return new NeuralNetConfiguration.Builder()
            .seed(SEED).weightInit(WeightInit.XAVIER)
            .iterations(NUM_ITERATIONS)
            .regularization(true).l2(0.0005).learningRate(.01)
            .optimizationAlgo(OptimizationAlgorithm.STOCHASTIC_GRADIENT_DESCENT)
            .updater(Updater.NESTEROVS).momentum(0.9)
            .list()
            .layer(0, new ConvolutionLayer.Builder(5, 5)
                    .stride(1, 1)
                    .nIn(NUM_CHANNELS)
                    .nOut(20)
                    .activation(Activation.IDENTITY)
                    .build())
            .layer(1, new SubsamplingLayer.Builder(SubsamplingLayer.PoolingType.MAX)
                    .kernelSize(2, 2)
                    .stride(2, 2)
                    .build())
            .layer(2, new ConvolutionLayer.Builder(5, 5).stride(1, 1)
                    .nOut(50)
                    .activation(Activation.IDENTITY)
                    .build())
            .layer(3, new SubsamplingLayer.Builder(SubsamplingLayer.PoolingType.MAX)
                    .kernelSize(2, 2)
                    .stride(2, 2)
                    .build())
            .layer(4, new DenseLayer.Builder()
                    .activation(Activation.RELU)
                    .nOut(500)
                    .build())
            .layer(5, new OutputLayer.Builder(LossFunctions.LossFunction.NEGATIVELOGLIKELIHOOD)
                    .activation(Activation.SOFTMAX)
                    .nOut(NUM_OUTPUTS)
                    .build())
            .setInputType(InputType.convolutionalFlat(28, 28, 1))
            .backprop(true)
            .pretrain(false).build();
}
```

1998 Gradient-based Lear … *Y. LeCun*


Anagnostes
localhost:8086
0021_CH4M
Next Scan
Konto / Compte / Conto
CHF
Einbezahlt von / Versé par / Versato da
Konto / Compte / Conto
CHF
105
Einbezahlt von / Versé par / Versato da
5
0
0
8
7
1
9
4
5
4
1.000
0.991
0.992
1.000
1.000
1.000
0.821
0.934
0.805


BSI-Business-Systems-Integration-AG / anagnostes
Identifying handwritten digits
29 commits
1 branch
New pull request
anagnostes.client
anagnostes.server.app-dev
anagnostes.server.app-war
anagnostes.server
anagnostes.shared
anagnostes.ui.html.app.dev
GitHub, Inc. [US] https://github.com/BSI-Business-Systems-Integration-AG/anagnostes
Screenshot of Anagnostes: The fourth digit was recognized as 6 with a low confidence (the output of the neural network for this digit was: 0: 0.392, 1: 0.000, 2: 0.000, 3: 0.000, 4: 0.004, 5: 0.000, 6: 0.602, 7: 0.000, 8: 0.001, 9: 0.000)
Image Processing
Before the images are fed to the neural network we transform the images into the same format as used in the MNIST database of handwritten digits. The following steps are performed:
1. apply Otsu's method to threshold the image
2. scale to fit into a 20x20 pixel box while preserving the aspect ratio; this gives us a grayscale image because of anti-aliasing
3. center image using center of gravity in a 28x28 pixel box
Image Normalization: The original scanned image (left) and the preprocessed image as described above (right)

Unsupervised Learning
Natural Language Processing

Data

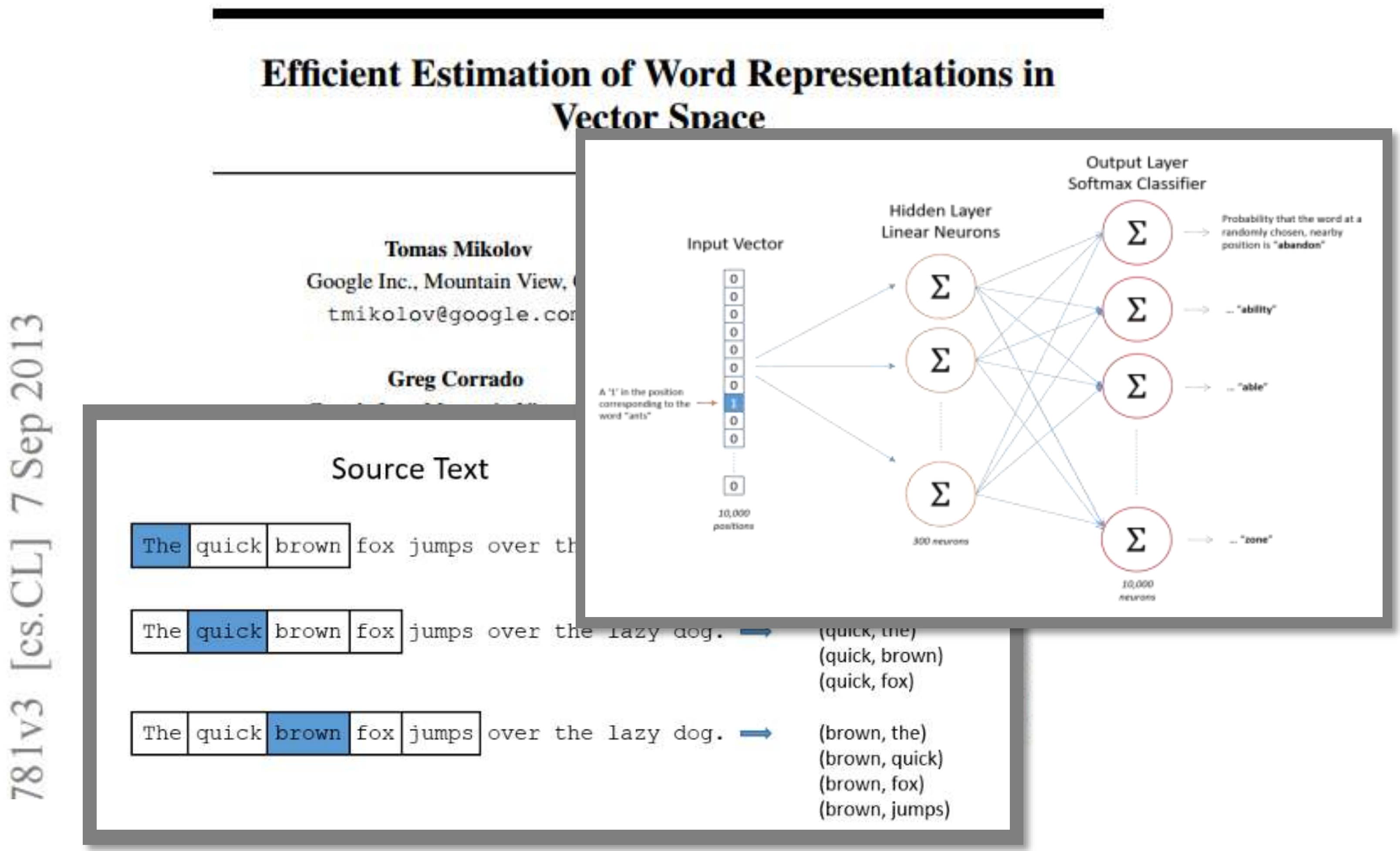
- → Google News text training dataset
- → Texts with total of 3’000’000’000 words
- → Lexicon: 3’000’000 words/phrases

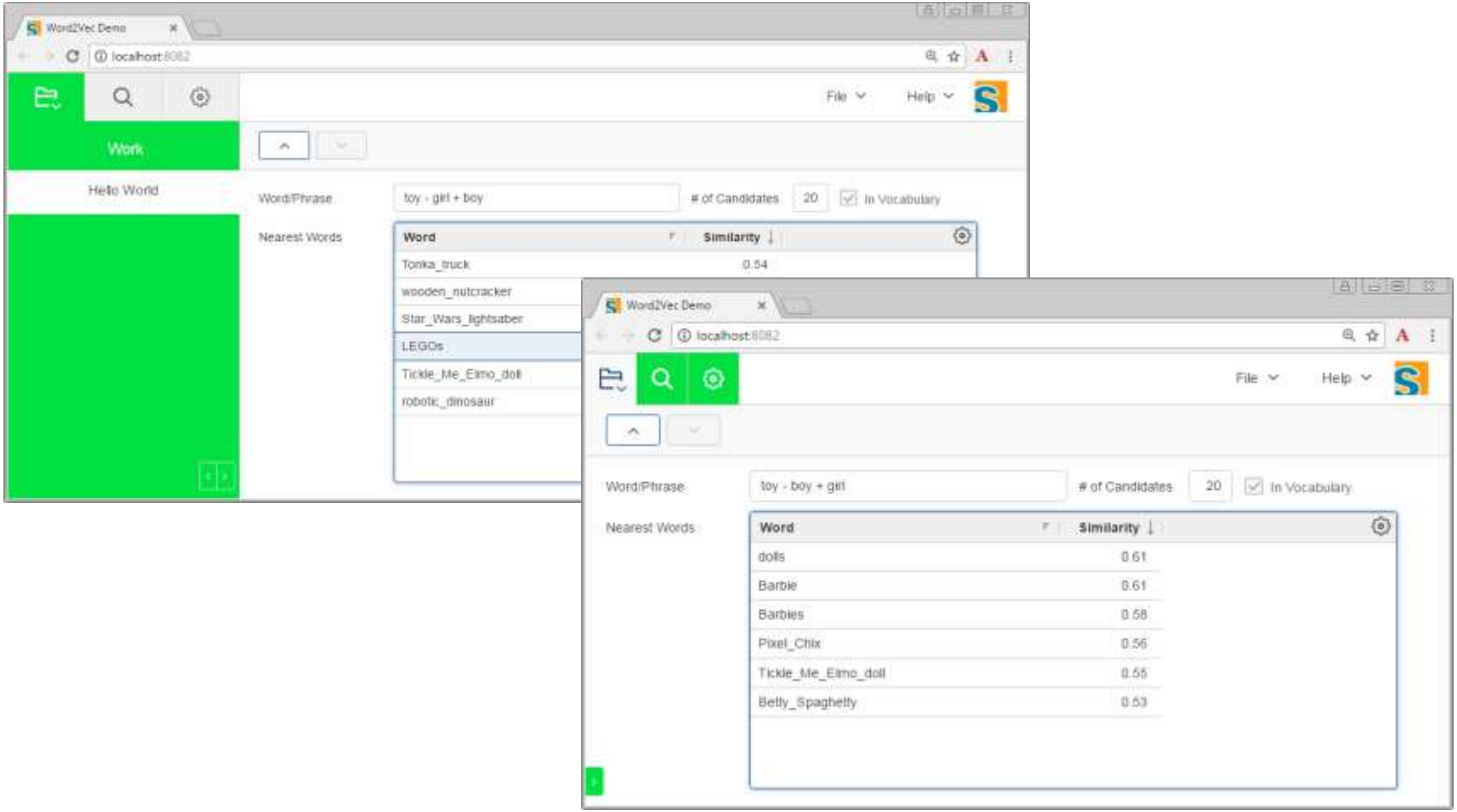
Model

- → Word2Vec Skip-gram
- → Mapping: Word → 300-dimensional number space
- → Many useful properties (word clustering, syntax, semantics)

Deeplearning4j

- → (Train) load and use Google News word2vec model


Efficient Estimation of Word Representations in Vector Space
Tomas Mikolov
Google Inc., Mountain View,
tmikolov@google.co
Greg Corrado
781v3 [cs.CL] 7 Sep 2013
Source Text
The quick brown fox jumps over th
The quick brown fox jumps over the lazy dog.
(quick, the)
(quick, brown)
(quick, fox)
The quick brown fox jumps over the lazy dog.
(brown, the)
(brown, quick)
(brown, fox)
(brown, jumps)
Input Vector
A '1' in the position corresponding to the word "ants"
10,000 positions
Hidden Layer
Linear Neurons
300 neurons
Output Layer
Softmax Classifier
Probability that the word at a randomly chosen, nearby position is "abandon"
... "ability"
... "able"
... "zone"
10,000 neurons


Word2Vec Demo
localhost:8082
File
Help
Work
Hello World
Word/Phrase
toy - girl + boy
# of Candidates
20
In Vocabulary
Nearest Words
Word
Similarity
Tonka_truck
0.54
wooden_nutcracker
Star_Wars_lightsaber
LEGOs
Tickle_Me_Elmo_doll
robotic_dinosaur
Word2Vec Demo
localhost:8082
File
Help
Word/Phrase
toy - boy + girl
# of Candidates
20
In Vocabulary
Nearest Words
Word
Similarity
dolls
0.61
Barbie
0.61
Barbies
0.58
Pixel_Chix
0.56
Tickle_Me_Elmo_doll
0.55
Betty_Spaghetty
0.53

Recent Advances

ML performance >= Human Levels (2017)

Games Backgammon 1979, chess 1997, Jeopardy! 2011, Atari games 2014, Go 2016, Poker (Texas Hold'em) 2017

Visual CAPTCHAs 2005, face recognition 2007, traffic sign reading 2011, ImageNet 2015, lip-reading 2016

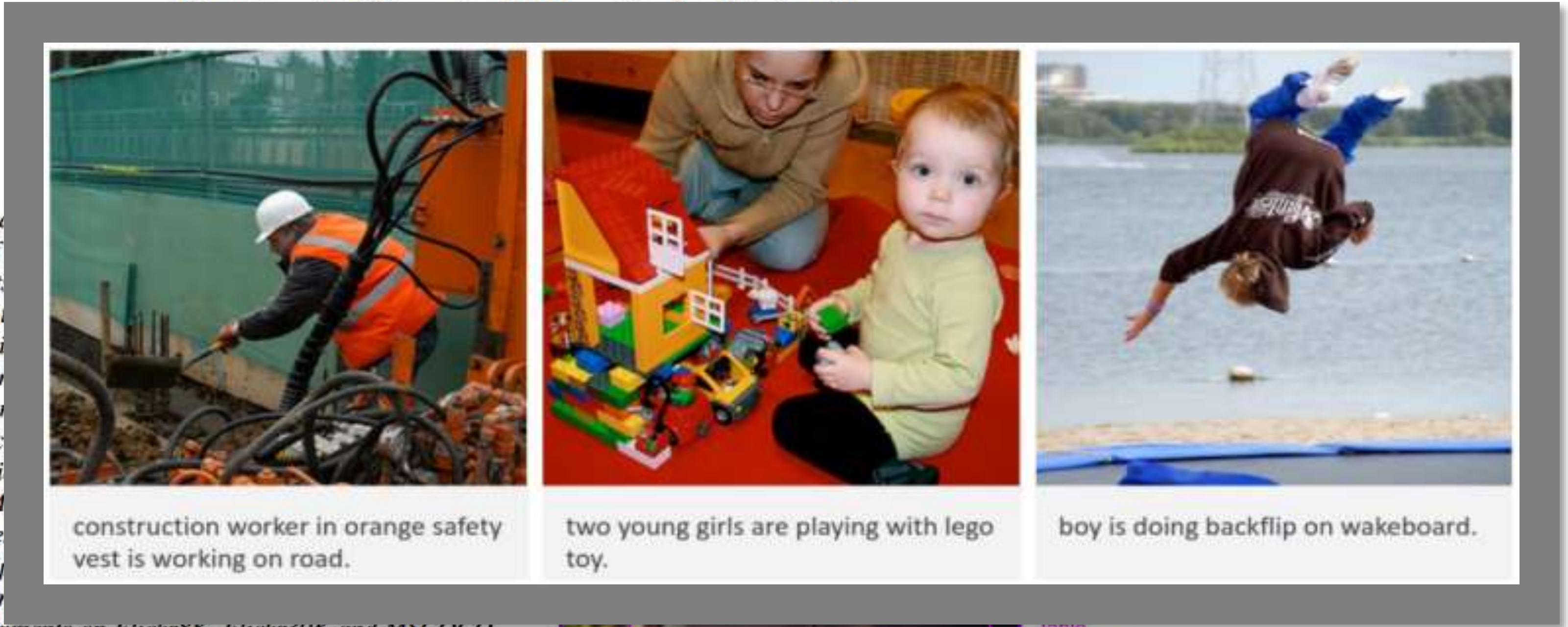
Other Age estimation from pictures 2013, personality judgement from Facebook «likes» 2014, conversational speech recognition 2016

https://finnaarupnielsen.wordpress.com/2015/03/15/status-on-human-vs-machines/

2014, Stanford

Deep Visual-Semantic Alignments for Generating Image Descriptions

Andrej Karpathy Li Fei-Fei



We present
scriptions of
ages dataset
learn about
guage and vi
novel combi
image regio
over sentenc
two modaliti
describe a M
ture that use
novel descri
our alignme
trieval experiments on Flickr8K, Flickr30K and MSCOCO
datasets. We then show that the generated descriptions significantly outperform retrieval
and on a new dataset of regio

table

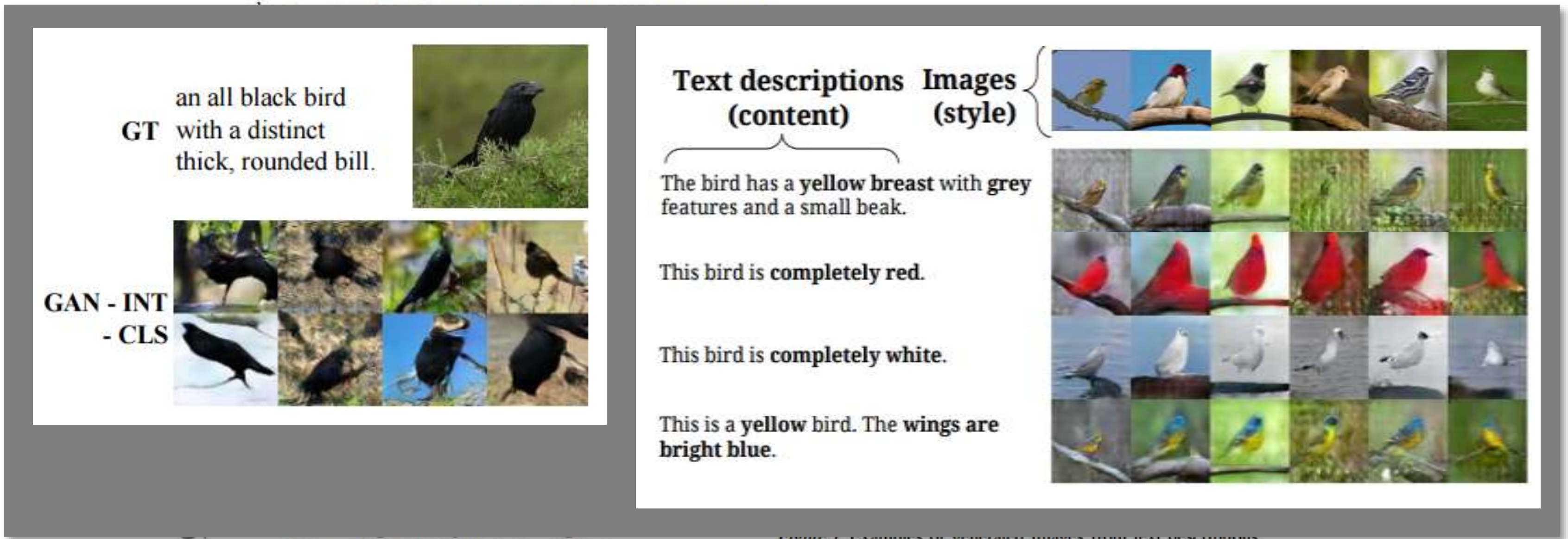
Figure 1. Motivation/Concept Figure: Our model treats language

http://cs.stanford.edu/people/karpathy/deepimagesent/devisagen.pdf
https://gigaom.com/2014/11/18/google-stanford-build-hybrid-neural-networks-that-can-explain-photos/

Generative Adversarial Text to Image Synthesis

2016, UMich+Max Plank

Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran REEDSCOT[1], AKATA[2], XCYAN[1]
Bernt Schiele, Honglak Lee SCHIELE[2],HONGLAK[1]



flowers from detailed text descriptions.

Left: captions are from zero-shot (held out) categories, unseen text. Right: captions are from the training set.

1. Introduction

In this work we are interested in translating text in the form of single-sentence human-written descriptions directly into image pixels. For example, "this small bird has a short, pointy orange beak and white belly" or "the petals of this

properties of attribute representations are attractive, attributes are also cumbersome to obtain as they may require domain-specific knowledge. In comparison, natural language offers a general and flexible interface for describing

the discrimi-

https://arxiv.org/pdf/1605.05396.pdf

2017, Cornell + Adobe

Deep Photo Style Transfer

Fujun Luan
Cornell Univ
fujun@cs.corn

Sylvain Paris

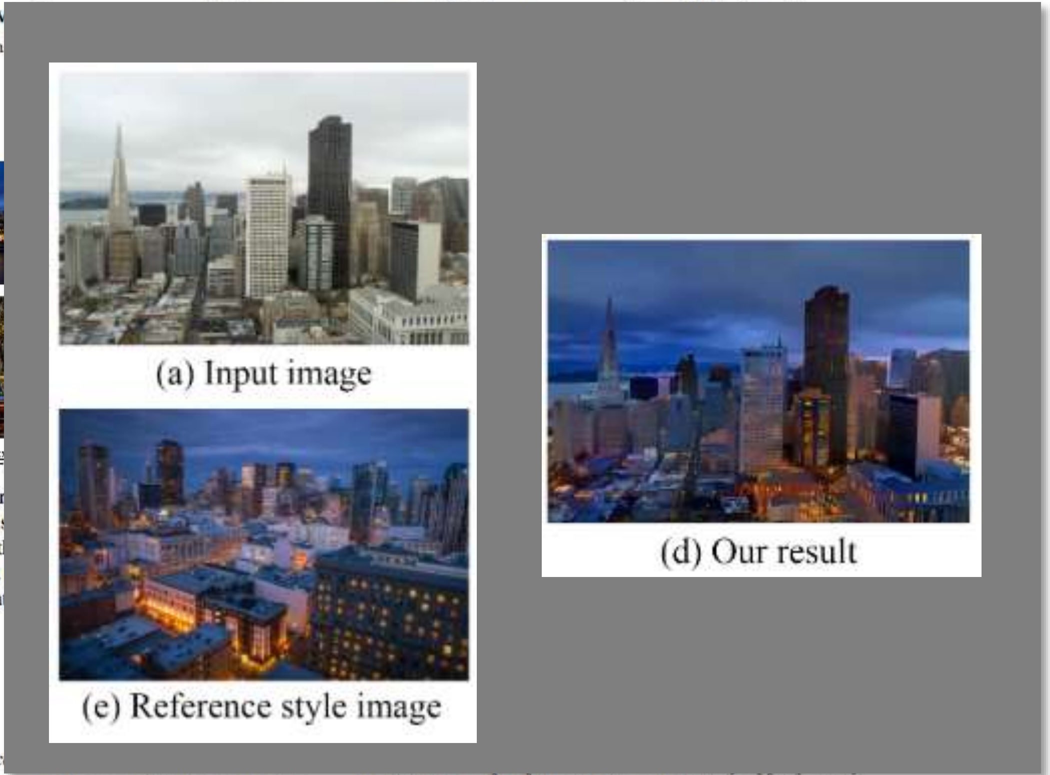
Eli Shechtman

Kavita Bala

arXiv:1703.07511v3 [cs.CV] 11 Apr 2017

(a) Reference style imag

Figure 1: Given a refer
the input, but with the s
introduces distortions t
comparison, our result
the output. On the righ



(a) Input image

(e) Reference style image

(d) Our result

This paper introduc
tographic style transfer that handles a large variety of image content while faithfully transferring the reference style. Our
a
ti

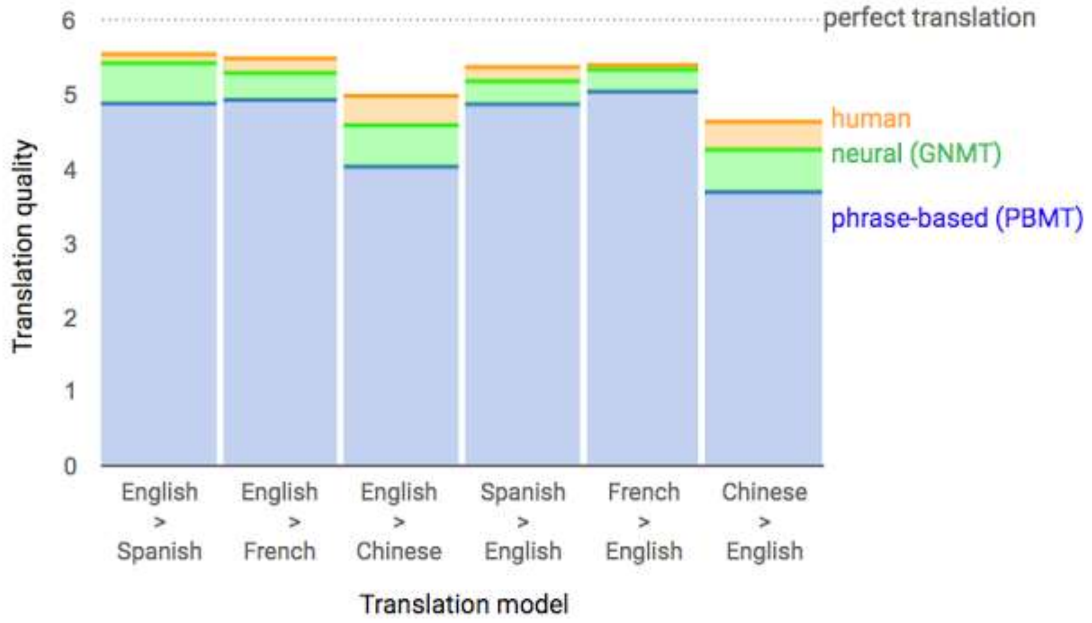
choosing the reference style photo, one can make the input picture look like it has been taken under a different illumina-
rtistically
echniques

https://arxiv.org/abs/1703.07511

Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation

2016, Google

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi



Data from side-by-side evaluations, where human raters compare the quality of translations for a given source sentence. Scores range from 0 to 6, with 0 meaning "completely nonsense translation", and 6 meaning "perfect translation."

English-to-German benchmarks, GNMT achieves competitive results to state-of-the-art. Using a human

https://research.googleblog.com/2016/09/a-neural-network-for-machine.html

Face2Face: Real-time Face Capture and Reenactment of RGB Videos

2016, Erlangen, Max-Plank, Stanford

Justus Thies[1] Michael Zollhöfer[2] Marc Stamminger[1] Christian Theobalt[2] Matthias Nießne
[1]University of Erlangen-Nuremberg [2]Max-Planck-Institute for Informatics [3]Stanford Universit

Proposed online reenactment setup: a monocular target video sequence (e.g., from Youtube) is reenacted based on the expressions of a source actor who is recorded live with a commodity webcam.

Abstract

We present a novel approach for real-time facial reenactment of a monocular target video sequence (e.g., Youtube video). The source sequence is also a monocular video stream, captured live with a commodity webcam. Our goal is to animate the facial expressions of the target video by a source actor and re-render the manipulated output video in a photo-realistic fashion. To this end, we first address the under-constrained problem of facial identity recovery from monocular video by non-rigid model-based bundling. At run time, we track facial expressions of both source and tar-

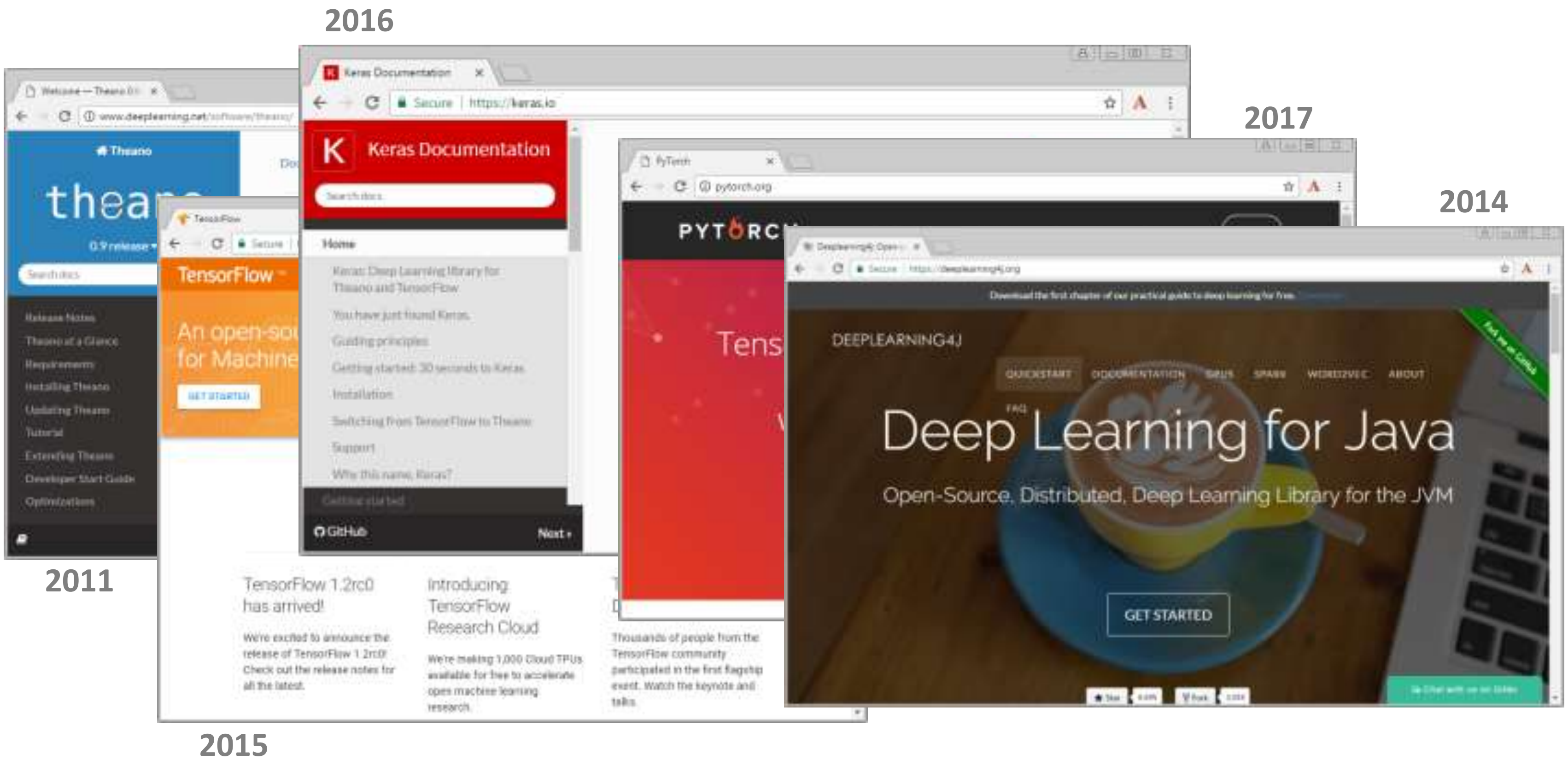
on RGB [8, 6] as well as RGB-D data [31, 10, 21, 4, 16]. These techniques have become increasingly popular for the animation of virtual CG avatars in video games and movies. It is now feasible to run these face capture and tracking algorithms from home, which is the foundation for many VR and AR applications, such as teleconferencing.

In this paper, we employ a new dense markerless facial performance capture method based on monocular RGB data, similar to state-of-the-art methods. However, instead of transferring facial expressions to virtual CG characters, our main contribution is monocular *facial reenactment* in real-time. In contrast to previous reenactment approaches

e *online* transfer
ured by an RGB
ence can be any

http://www.graphics.stanford.edu/~niessner/papers/2016/1facetoface/thies2016face.pdf
https://www.youtube.com/watch?v=ttGUiwfTYvg

ML Libraries



Food for Thought + Next Steps

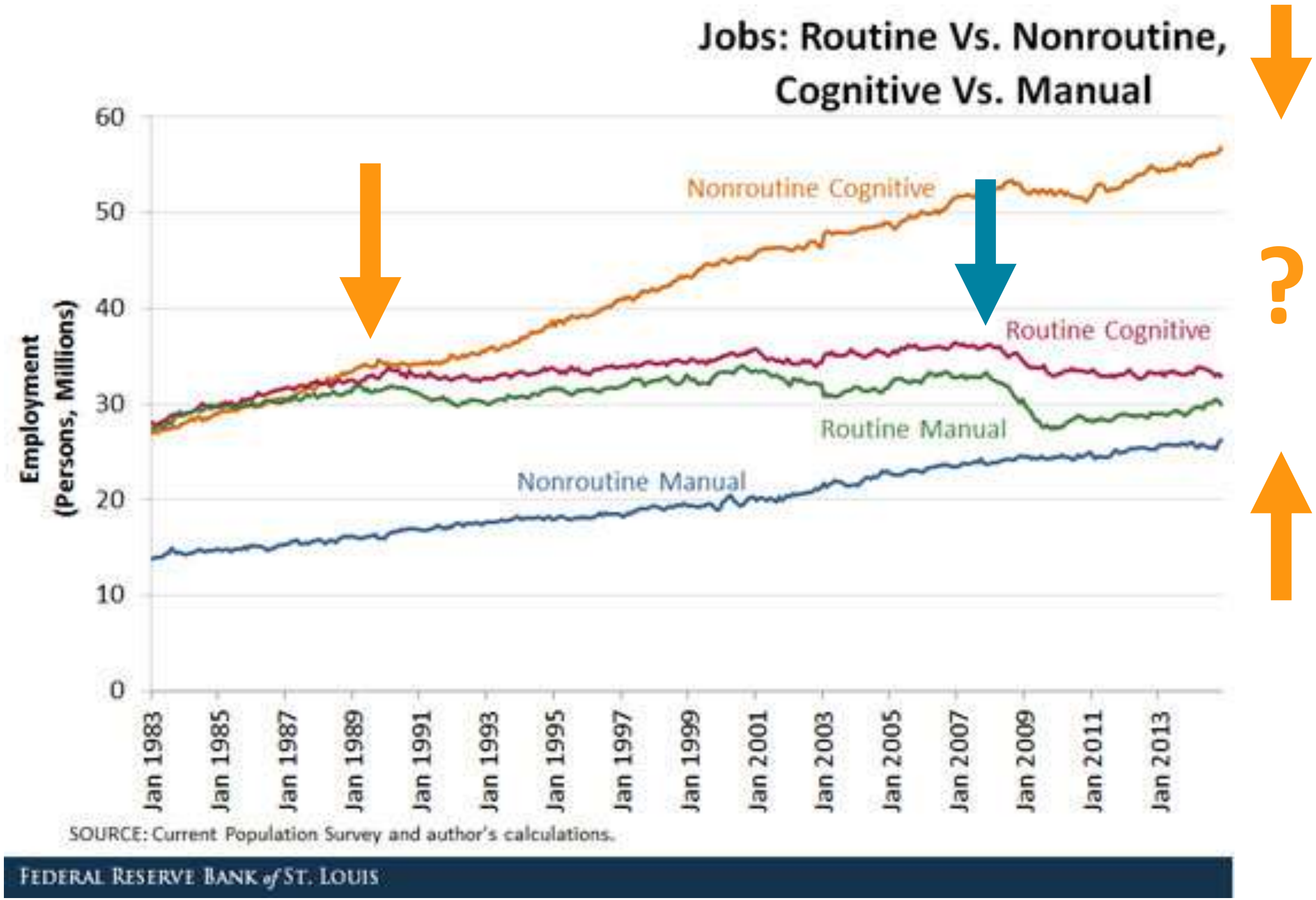
FINANCIAL TIMES

Artificial Intelligence and Robotics + Add to myFT

AI and robots threaten to unleash mass unemployment, scientists warn

Intelligent machines will soon replace human workers in all sectors of economy

Read latest:
Market grows for 'regtech', or AI for regulation
OCTOBER 14, 2016


Jobs: Routine Vs. Nonroutine,
Cognitive Vs. Manual
Employment
(Persons, Millions)
60
50
40
30
20
10
0
Nonroutine Cognitive
Routine Cognitive
Routine Manual
Nonroutine Manual
Jan 1983
Jan 1985
Jan 1987
Jan 1989
Jan 1991
Jan 1993
Jan 1995
Jan 1997
Jan 1999
Jan 2001
Jan 2003
Jan 2005
Jan 2007
Jan 2009
Jan 2011
Jan 2013
SOURCE: Current Population Survey and author's calculations.
FEDERAL RESERVE BANK of ST. LOUIS
?

Positive Outcomes

Statement by Lee Sedol

Lee replied that playing against the machine had rekindled his passion for Go. As with Fan Hui, AlphaGo had opened his eyes to a new side of the game. “I have improved already,” Lee said. “It has given me new ideas.” He has not lost a match since.

Like to learn more?

Socalizing

➔ Go to talks, conferences

➔ Visit meetups (Zurich Machine Learning and Data Science, …)

Increase Context

➔ Blogs, Twitter, arxiv.org, …

Doing

➔ GitHub (deeplearning4j/deeplearning4j,
BSI-Business-Systems-Integration-AG/anagnostes, …)

➔ Learn Python ;-)

Thanks!

@ZimMatthias