Vienna Deep Learning Meetup

October 24, 2017 @ Marx Palast Wien

Thomas Lidy

Jan Schlüter

Alex Schindler

Our Sponsor

Organizing Hosts:
Christoph Toeglhofer
Tamara Berger-Feichter
Boris Marte

Our Streaming Partner

STREEMED

Organizer:
Maria Vasilevich

https://www.youtube.com/watch?v=yJiBqVxqfmA

Vienna Deep Learning Meetup

Vienna 13th Deep Learning Meetup

Agenda:

- Welcome by the Organizers (Tom Lidy)
- Introduction by Erste Group, the host of today (Boris Marte)
- **TensorFlow Wide & Deep**: Data Classification the easy way (**Yufeng Guo, Google** Cloud, NY)

30 minutes break

- **One Model to Learn them All**: Lightning talk (Valentyn Boreiko)
- Hot Topics & Latest News (Tom Lidy, Alex Schindler, Jan Schlüter)
- Discussions and Networking

Vienna Deep Learning Meetup

Hot Topics &
Latest News

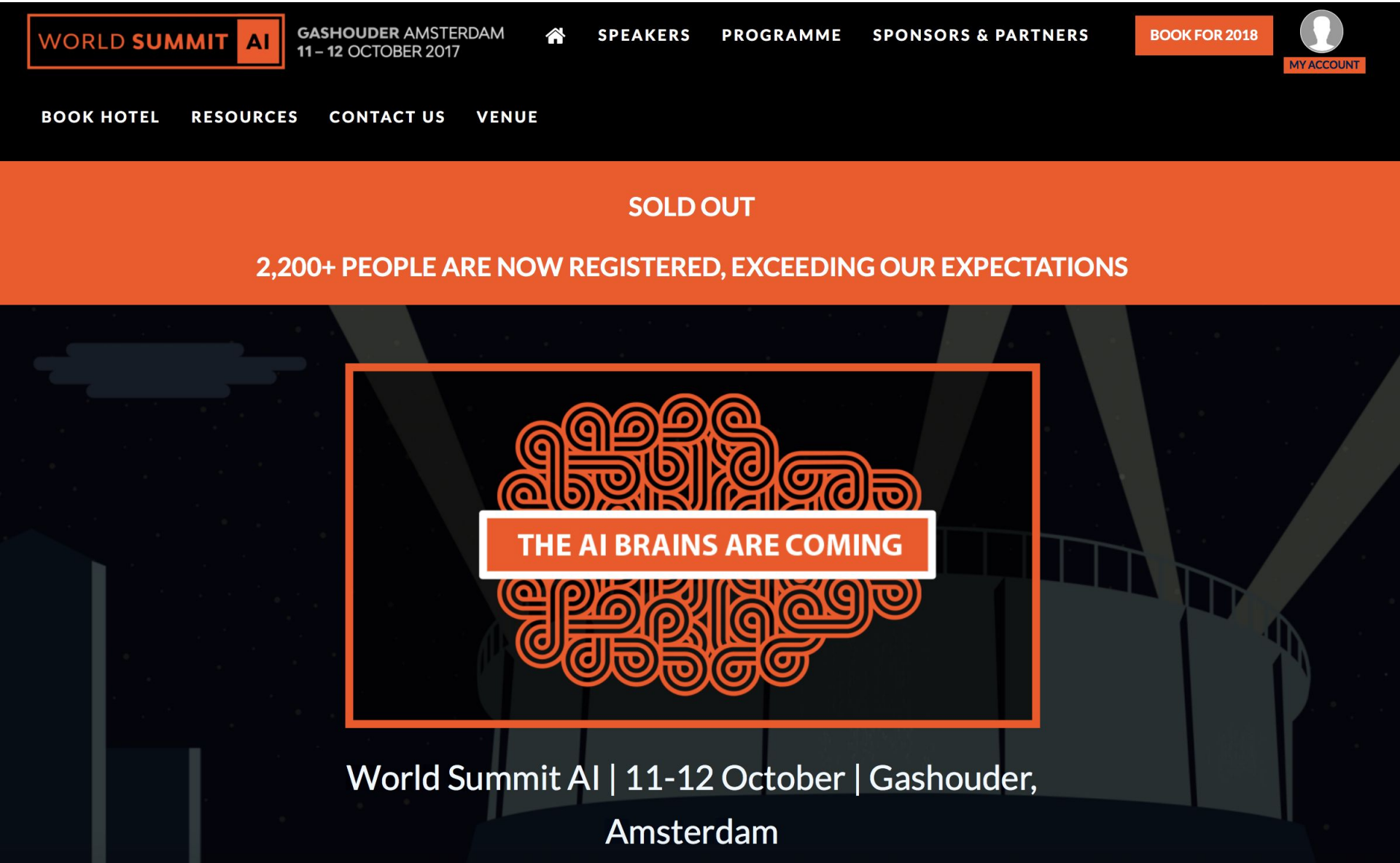
Tom Lidy,
Alexander Schindler,
Jan Schlüter

a 10-15 min block at every meetup
to briefly present recent papers and news

Send us contributions (tom.lidy@gmail.com)
or come with slides to do a short block yourself!

Vienna
Deep Learning
Meetup

Review: World Summit AI (Amsterdam, Oct 2017)



Review: World Summit AI

Main Take-Aways:

- AI is taking off - but we are not quite there yet
- Artificial General Intelligence will take time
- Ethics and Dangers need to be discussed + tackled
- Too much is focused on “narrow” AI
- Deep Learning solves a lot of “narrow” AI problems but does not generalize well for a “thinking” system

Vienna Deep Learning Meetup

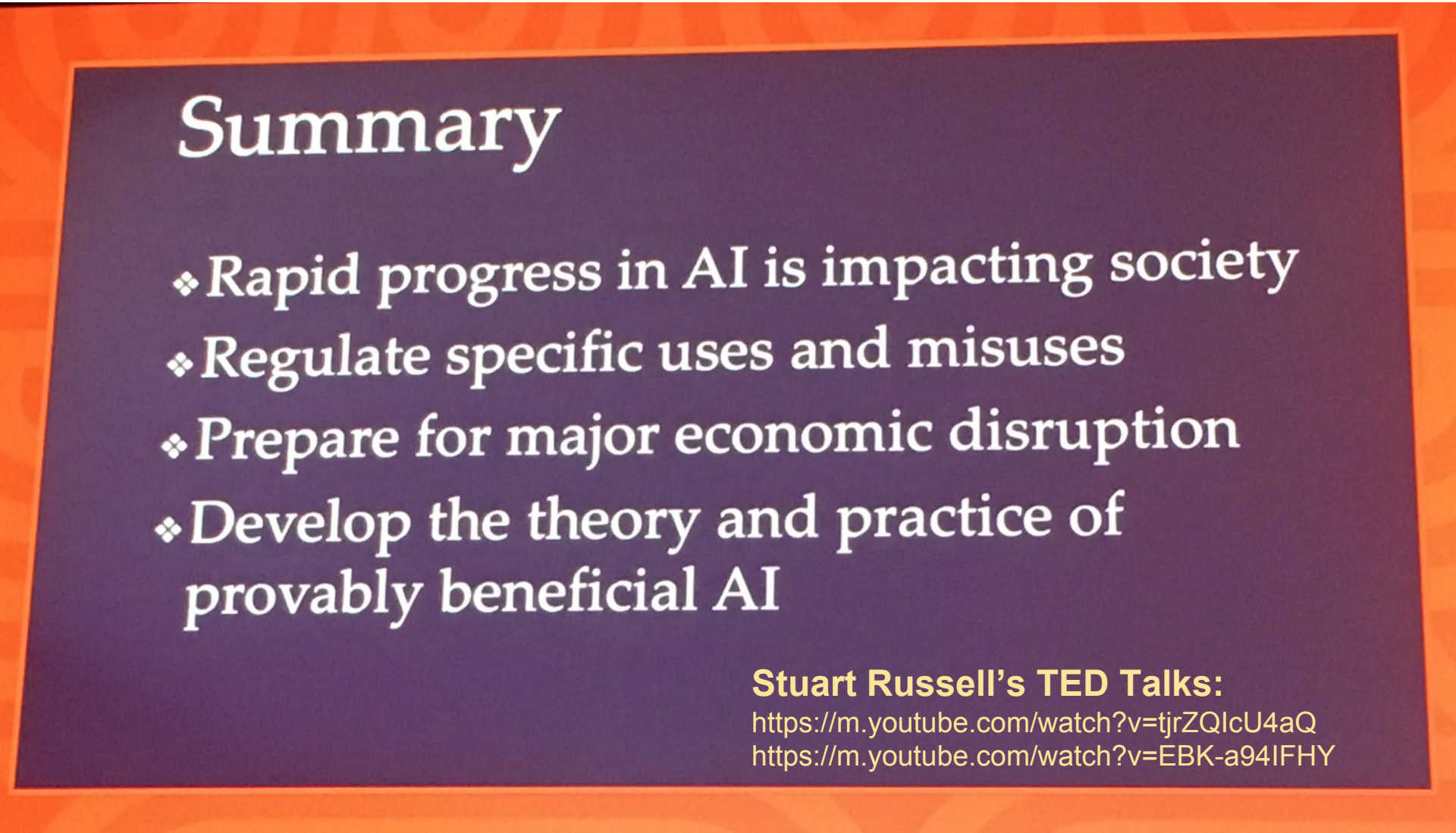
Stuart Russell

Prof. in AI at University of California, Berkeley

"The value created by AI is bigger than the GDP of the entire planet“

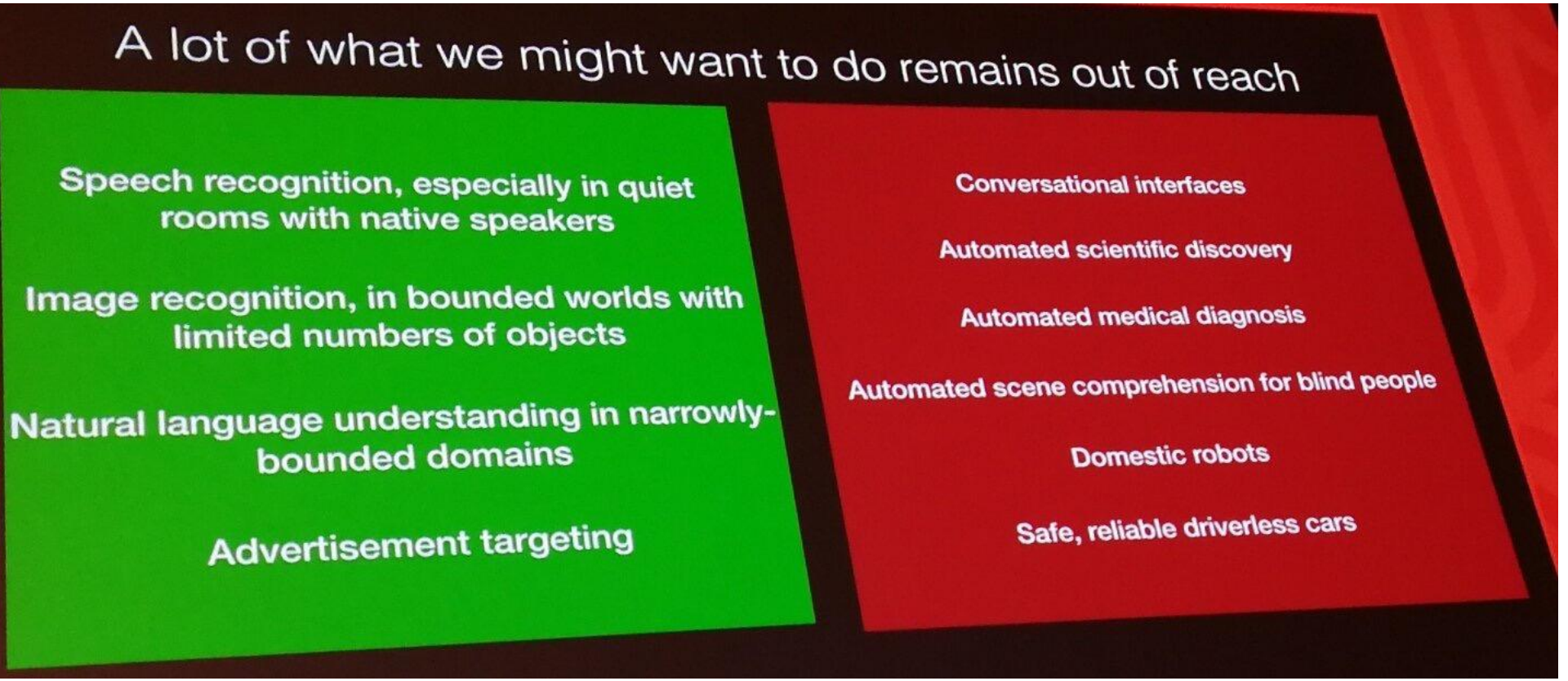
Stuart Russell

Prof. in AI at University of California, Berkeley



Review: World Summit AI

We have ... **We are far from ...**

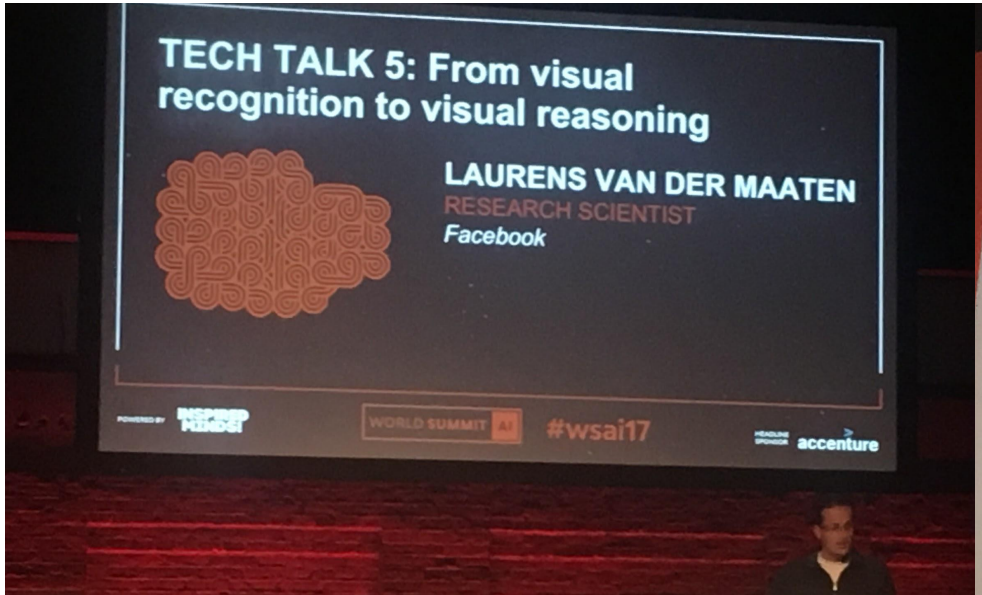


Vienna Deep Learning Meetup

Review: World Summit AI

From Recognition to Reasoning

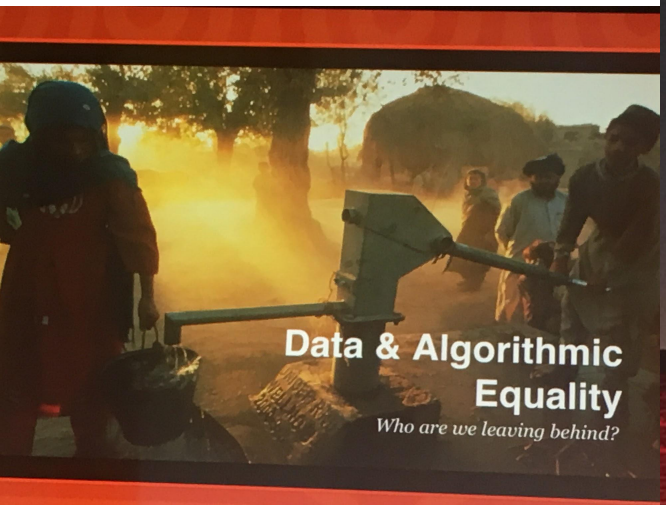
- Reasoning, Attention, Memory are key challenges
- Multimodal Learning is in its early days
- Deep Reinforcement Learning (with memory) is a hot topic

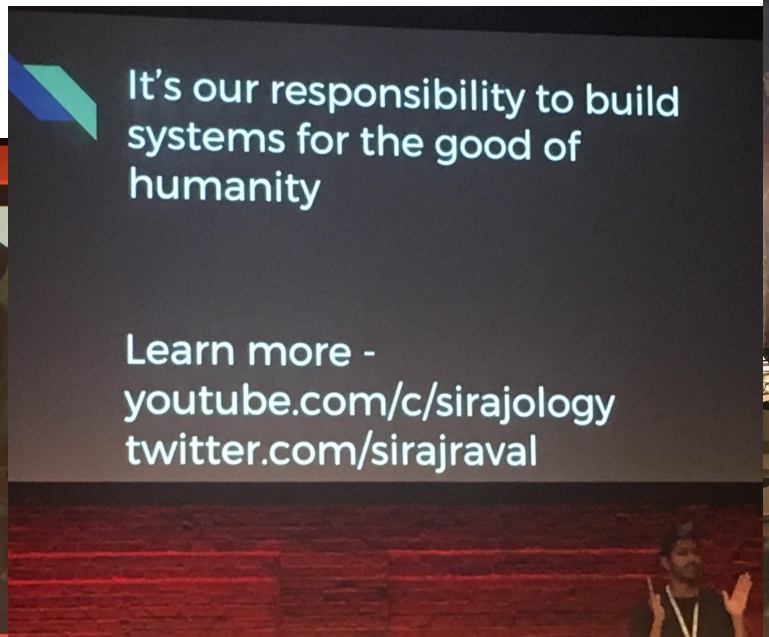


Review: World Summit AI

Questions discussed

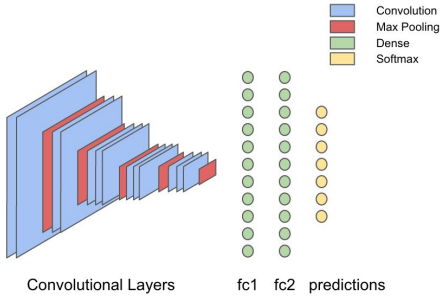
- Biased AI - Data & **Algorithmic Equality**
- Reasoning and **understanding intent**
- **Uncertainty in AI:** AI systems should know when they don't know!
- The power and **limits of Deep Learning** (Yann LeCun)

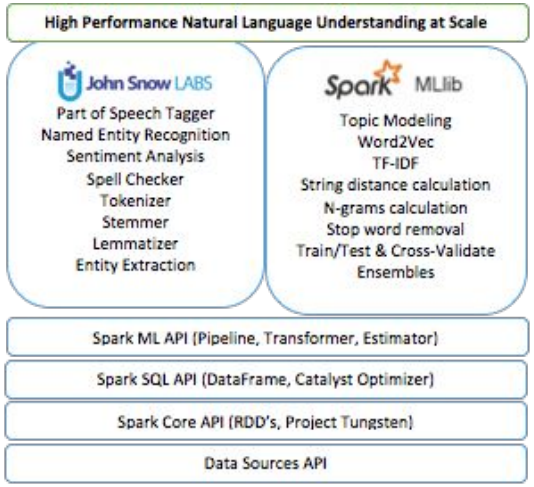




Deep Learning on Apache Spark

1. Object classification on the Caltech-256 image dataset
   - pre-trained VGG16
   - Apache Spark
   - Apache Hadoop
   - Deeplearning4j
2. Introducing the Natural Language Processing Library for Apache Spark



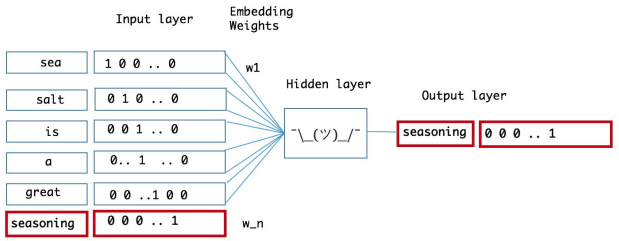


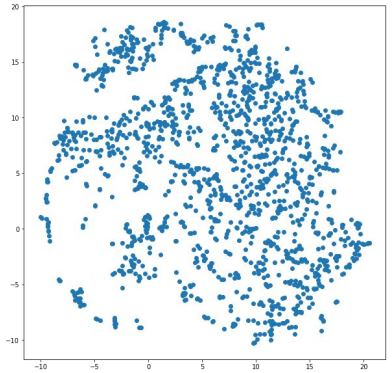
1. https://blog.cloudera.com/blog/2017/06/deep-learning-on-apache-spark-and-hadoop-with-deeplearning4j
2. https://databricks.com/blog/2017/10/19/introducing-natural-language-processing-library-apache-spark.html

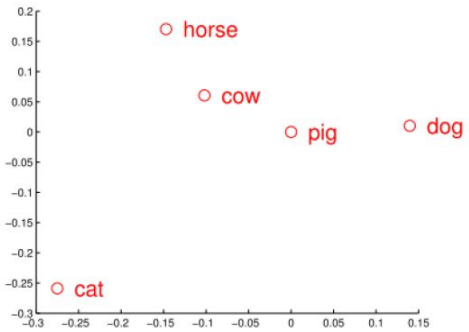
Vienna Deep Learning Meetup

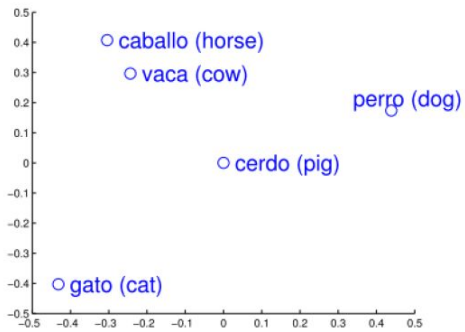
Word Embeddings & NLP Crash Course

1. Introduction to NLP and Word Embeddings
   - preprocessing
   - modelling
   - visualization with T-SNE
   - Well explained code examples
     - NLTK (NLP, preprocessing)
     - Gensim (NLP, word2vec)
     - sklearn (T-SNE)
     - bokeh (interactive plotting)
2. A Survey Of Cross-lingual Word Embedding Models
   - Well curated survey by Sebastian Ruder, Ivan Vulić, Anders Søgaard







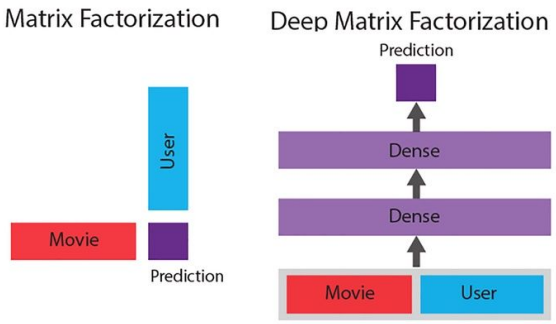


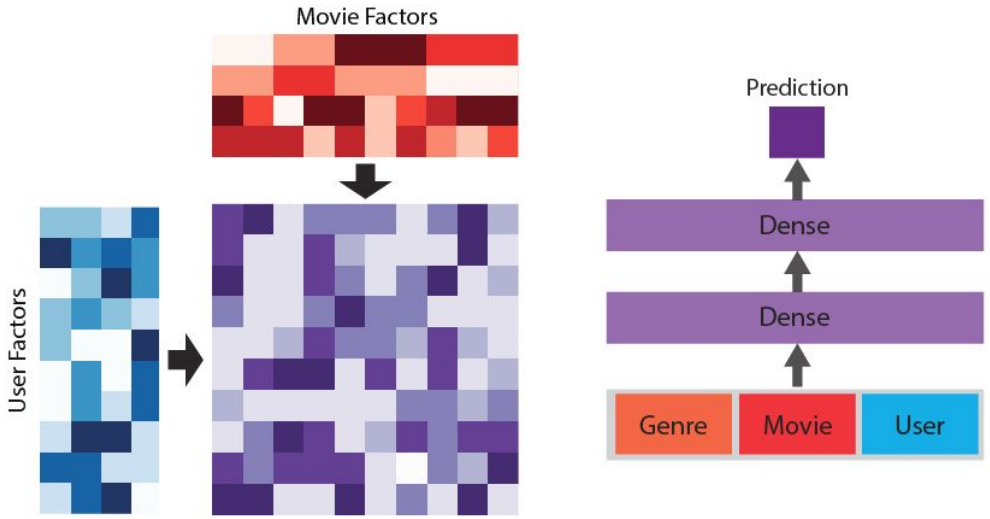
1. https://www.datascience.com/resources/notebooks/word-embeddings-in-python
2. https://arxiv.org/abs/1706.04902

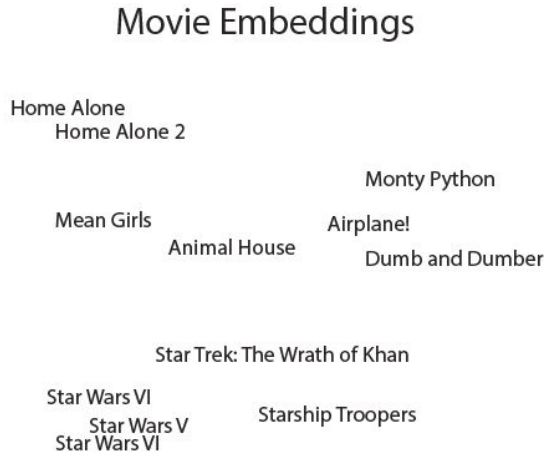
Vienna Deep Learning Meetup

Deep matrix factorization using Apache MXNet

1. Excellent Introduction to Recommender Systems
   - Recommending Movies to Users
   - Deep Matrix Factorization
     - Apache mxnet
     - Word Embeddings





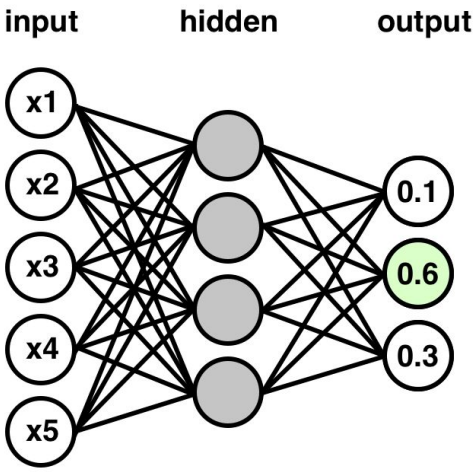


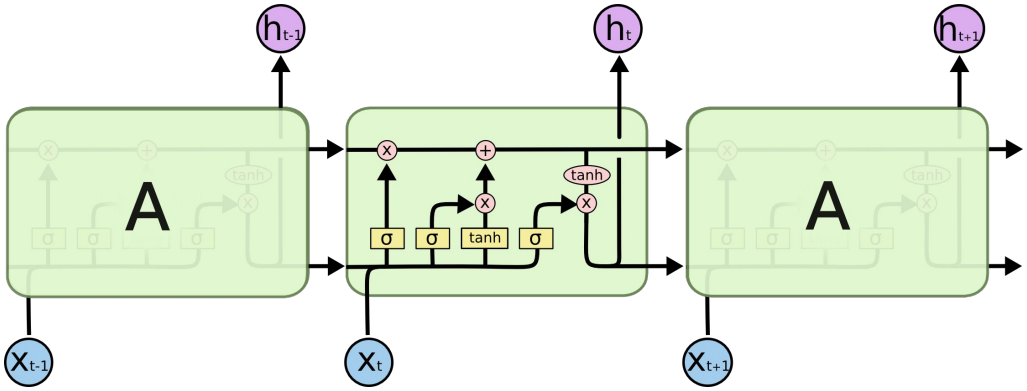
1. https://www.oreilly.com/ideas/deep-matrix-factorization-using-apache-mxnet
2. Jupyter notebook: https://github.com/jmschrei/notebooks/tree/master/MXNet%20Deep%20Matrix%20Factorization

Vienna Deep Learning Meetup

Learning by Coding

1. Coding up a Neural Network classifier from scratch
   - Fully connected Layer in Python
   - Forward- Back-propagation
   - Prediction
2. Vanilla LSTM
   - RNN implementation with Numpy
   - Jupyter Notebook
3. Recurrent Neural Networks
   - Introduction
   - Lists of
     - Blogs
     - Papers
     - Tutorials
     - Theses





1. https://github.com/ankonzoid/NN-from-scratch
2. http://blog.varunajayasiri.com/numpy_lstm.html
3. https://blog.recast.ai/ml-spotlight-rnn

Vienna Deep Learning Meetup

Blockchain meets Deep Learning

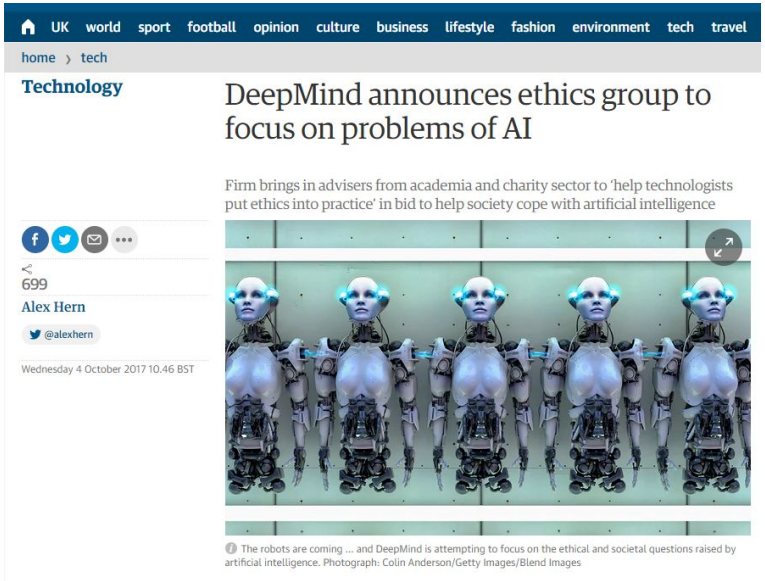
1. Blockchain and Machine Learning Workshop at START Summit 2017 in Switzerland
   - Smart Contracts + Machine Learning
   - Chat Client
     - Transmit / Receive to Blockchain
       - Messages, Images
     - Automatic Tagging of message
       - Keras + Tensorflow
     - History remains in the Blockchain

1. https://github.com/thoschm/START-Summit-2017-Blockchain-Machine-Learning-Workshop

Vienna Deep Learning Meetup

Politics

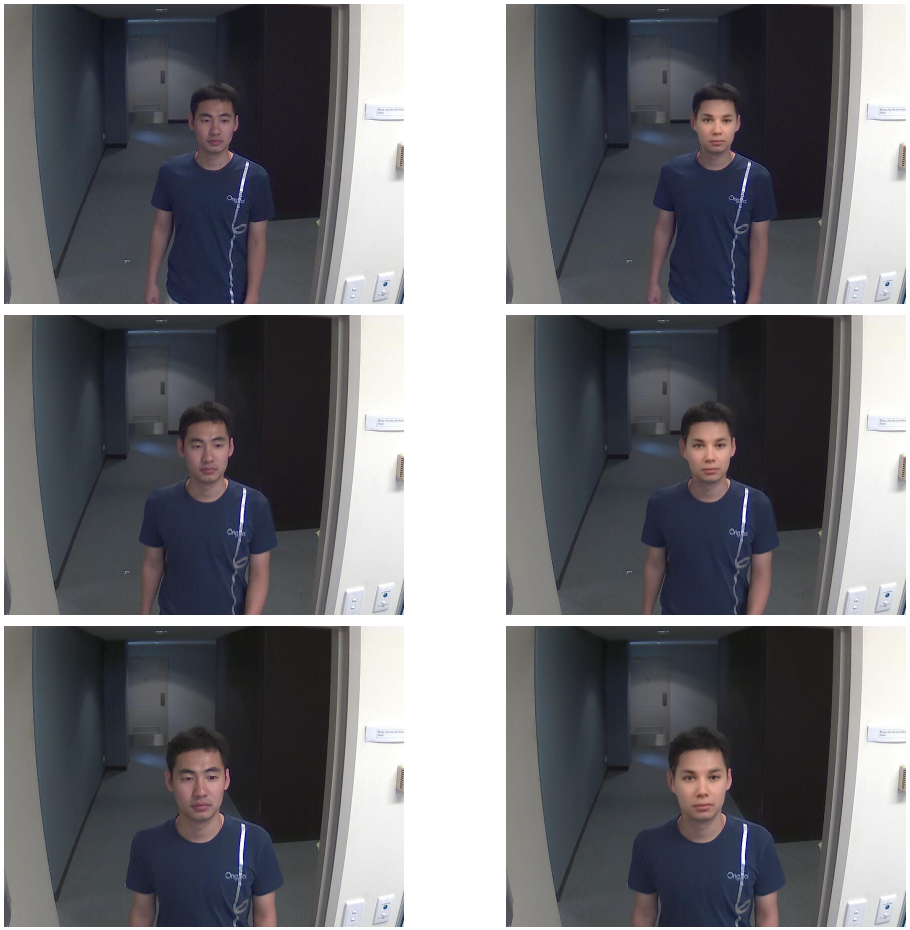
1. DeepMind announces ethics group to focus on problems of AI
   - ethical and societal questions raised by artificial intelligence
   - Fears
     - AI will lead 3rd world war (Elon Musk)
     - Critical decisions by Black-Boxes
       - Credit
       - Job
         - Health
         - Sexual preference
       - Law Enforcement
         - Public observation
         - Predictive policing



1. https://www.theguardian.com/technology/2017/oct/04/google-deepmind-ai-artificial-intelligence-ethics-group-problems

Face Deidentification with Generative Deep Neural Networks

- **Goal:** Protect privacy in surveillance videos
- Pixelation, blurring: destroys too much information
- **New approach:**
  Face detection
  + generative model
  = face deidentification
- Can retain facial expression without revealing identity

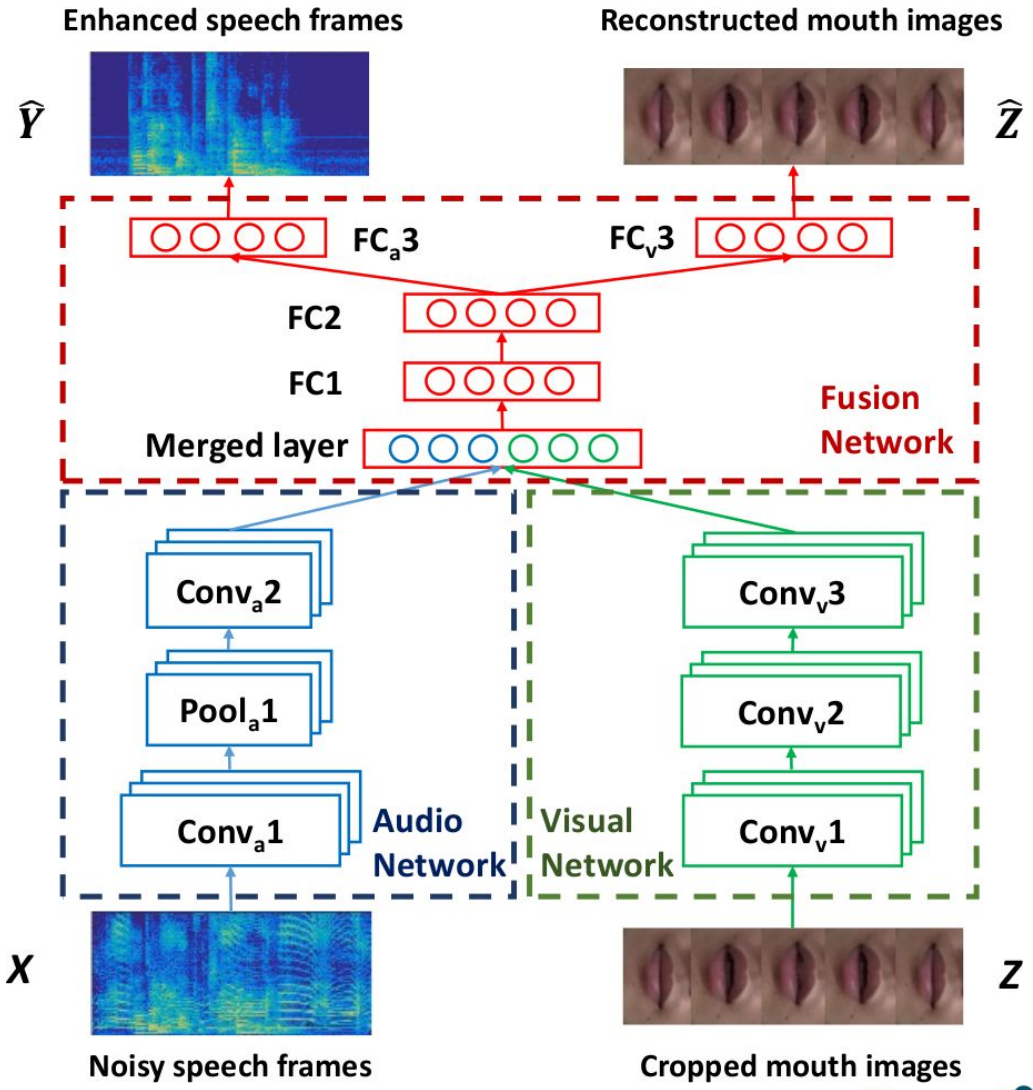


https://arxiv.org/abs/1707.09376

Vienna Deep Learning Meetup

Audio-Visual Speech Enhancement based on Multimodal Deep Convolutional Neural Networks

- Humans can understand speech better when watching the speaker's face
- **Idea:** Train network to improve noisy audio using visual cues
- **Result:** better than audio only. (Main cue: mouth open vs. closed.)

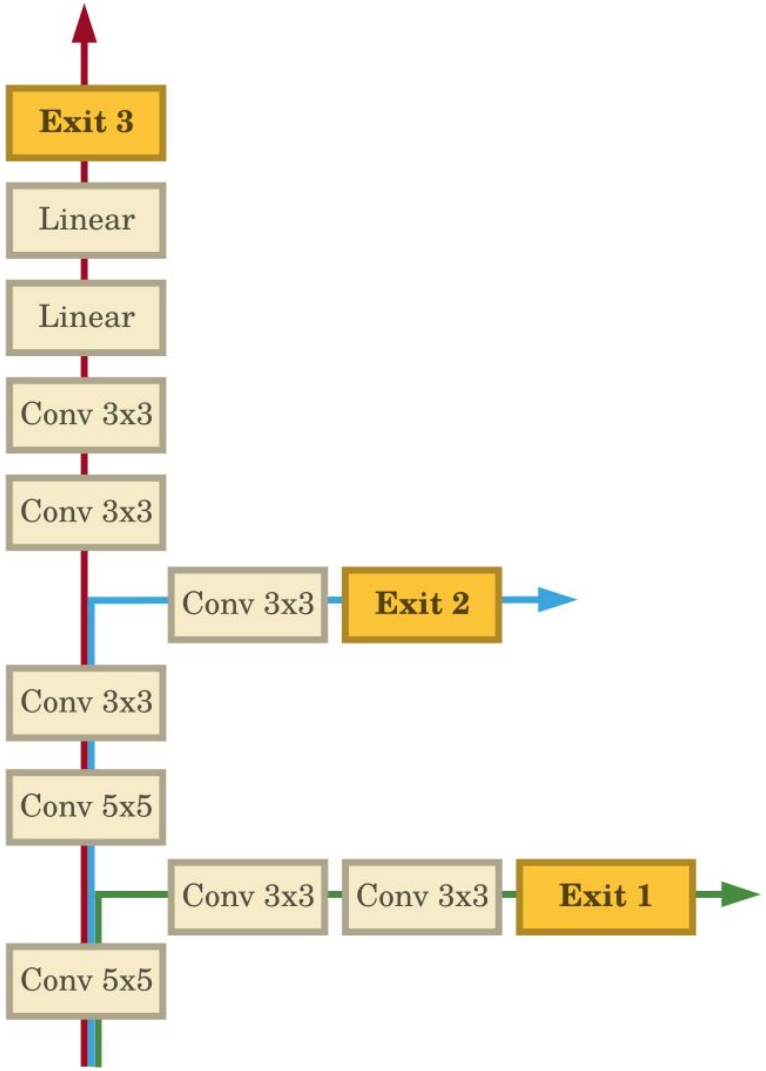


https://arxiv.org/abs/1709.00944

Vienna Deep Learning Meetup

BranchyNet: Fast Inference via Early Exiting from Deep Neural Networks

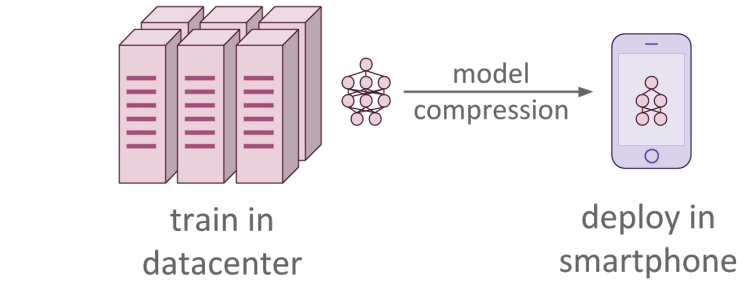
- GoogLeNet (2014): multiple classification heads to improve gradient flow
- **Idea:** use same principle to accelerate inference (deployment)
- Only evaluate higher layers if shallow exit is not confident (as measured by entropy)
- About 2x faster
- Complementary to other approaches, but less useful when evaluating in batches

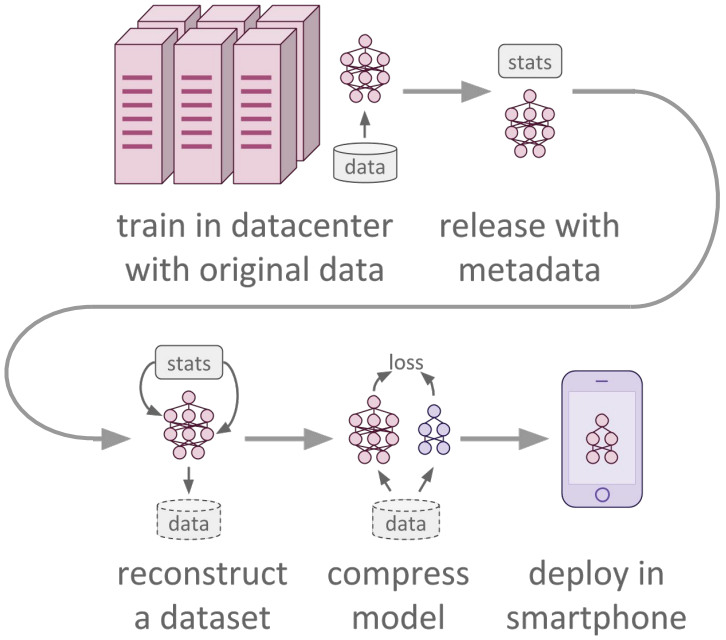


https://arxiv.org/abs/1709.01686

Data-Free Knowledge Distillation for Deep Neural Networks

- Old trick for deployment (Hinton, 2015): Train small network to reproduce large network outputs
- Reproducing probabilities of wrong answers teaches about generalization ("dark knowledge")
- Caveat: Requires access to training data
- **Idea:** Recreate training data from the network + side information about activation statistics
- **Result:** Sort of works.



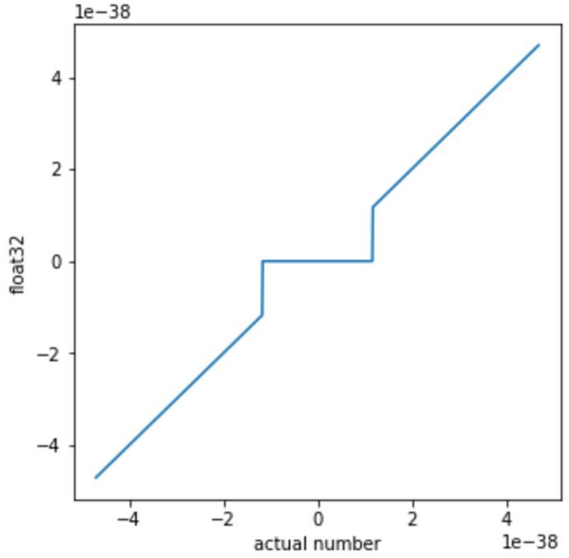


https://arxiv.org/abs/1710.07535

Vienna Deep Learning Meetup

Nonlinear Computation in Linear Networks

- Deep network without nonlinearities reduces to single-layer network
- IEEE-754 float32 numbers have a large gap between zero and the smallest positive and negative numbers (when excluding denormals)
- Computation results falling into that gap get mapped to zero (when excluding denormals) ⇒ nonlinear computation around 10^{-38}

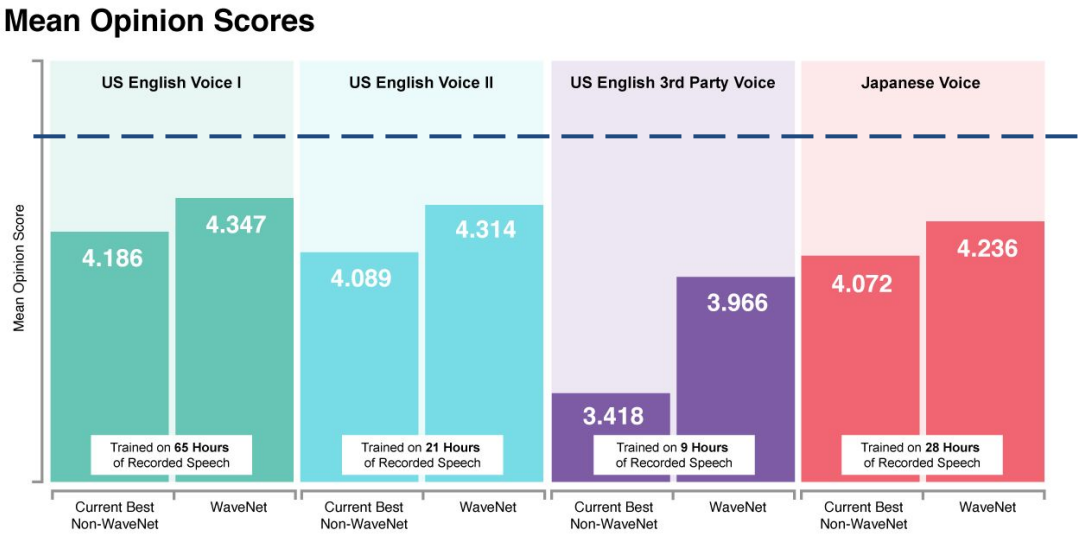


- Backpropagation does not know about this, but evolution strategies can find and exploit this behavior
- Limited use (free nonlinearity in widespread hardware), but a nice hack

https://blog.openai.com/nonlinear-computation-in-linear-networks/

Text-to-Speech WaveNet in production

- Previously discussed:
  - WaveNet (DeepMind, Sep 2016, arXiv 1609.03499)
  - Deep Voice 1 (Baidu, Feb 2017, arXiv 1702.07825)
  - Tacotron (Google, Mar 2017, arXiv 1703.10135)
  - Deep Voice 2 (Baidu, May 2017, arXiv 1705.08947)
- Oct 2017: Improved WaveNet used for Google Assistant
  - Generate 20 s of audio in 1 s (old WaveNet: 20 ms in 1 s)
  - 24 kHz, 16 bit (old WaveNet: 16 kHz, 8 bit)

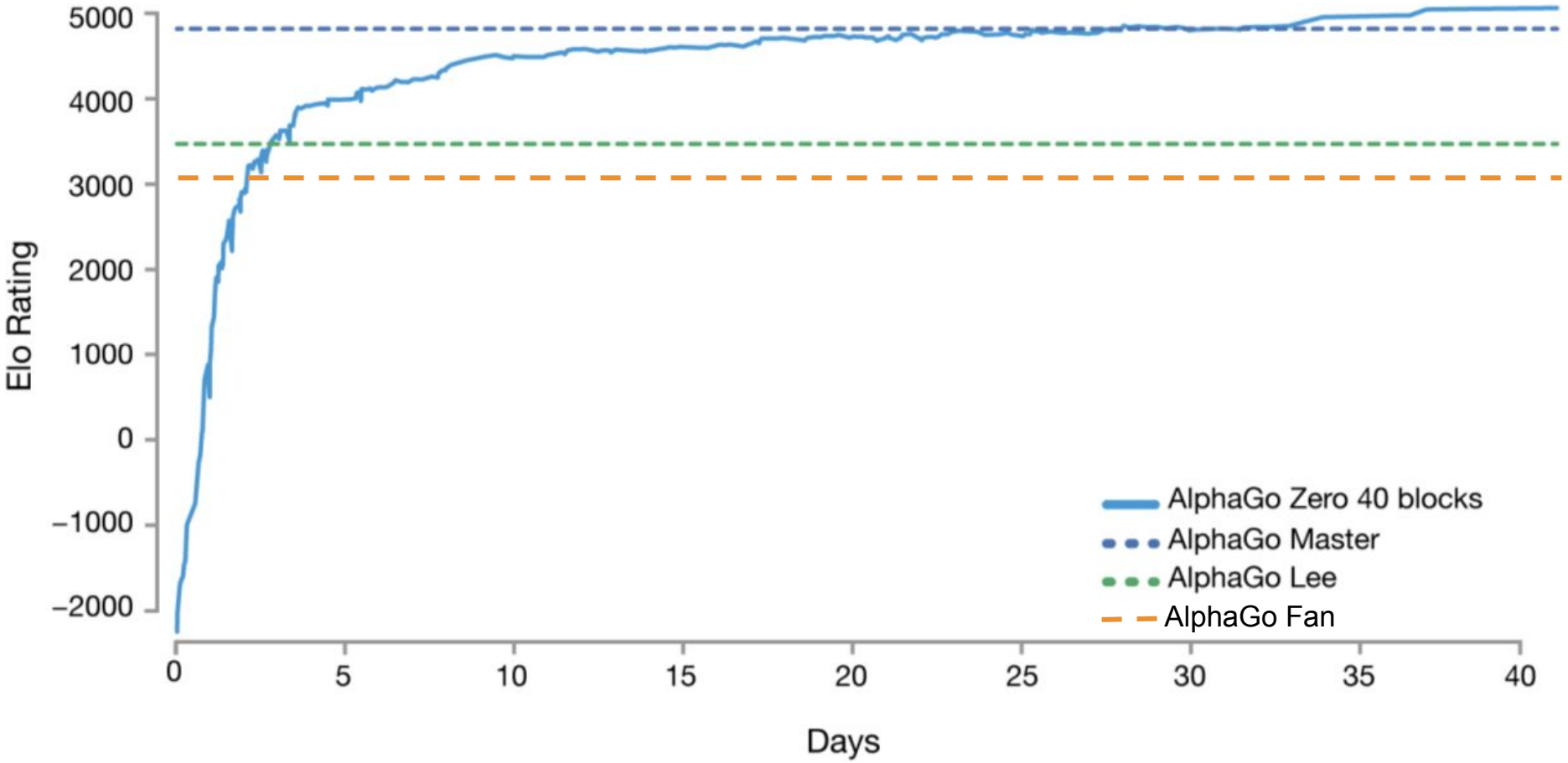


Human voice scores at 4.667.

Side note: Univ. of Hamburg just released aligned Wikipedia audio corpus with hundreds of hours (German, English, Dutch): http://nats.gitlab.io/swc/

https://deepmind.com/blog/wavenet-launches-google-assistant/

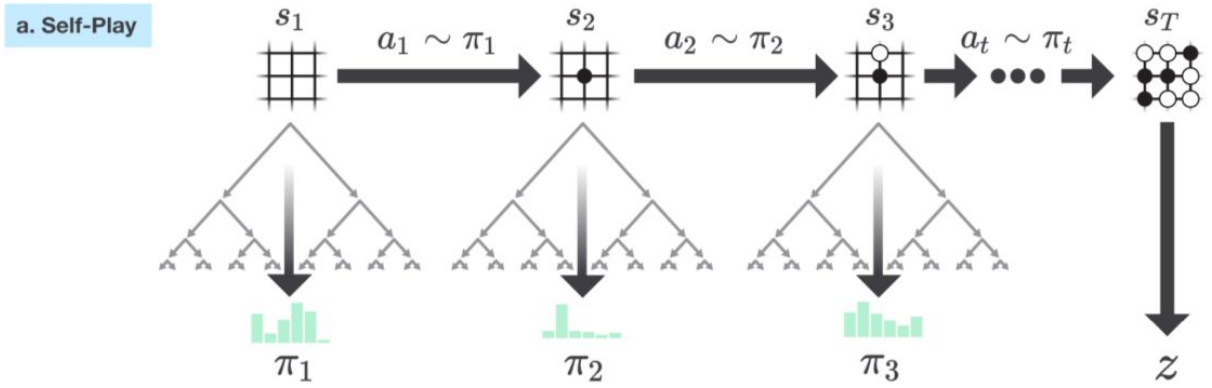
AlphaGo Zero



https://deepmind.com/blog/alphago-zero-learning-scratch/
https://deepmind.com/documents/119/agz_unformatted_nature.pdf

Vienna Deep Learning Meetup

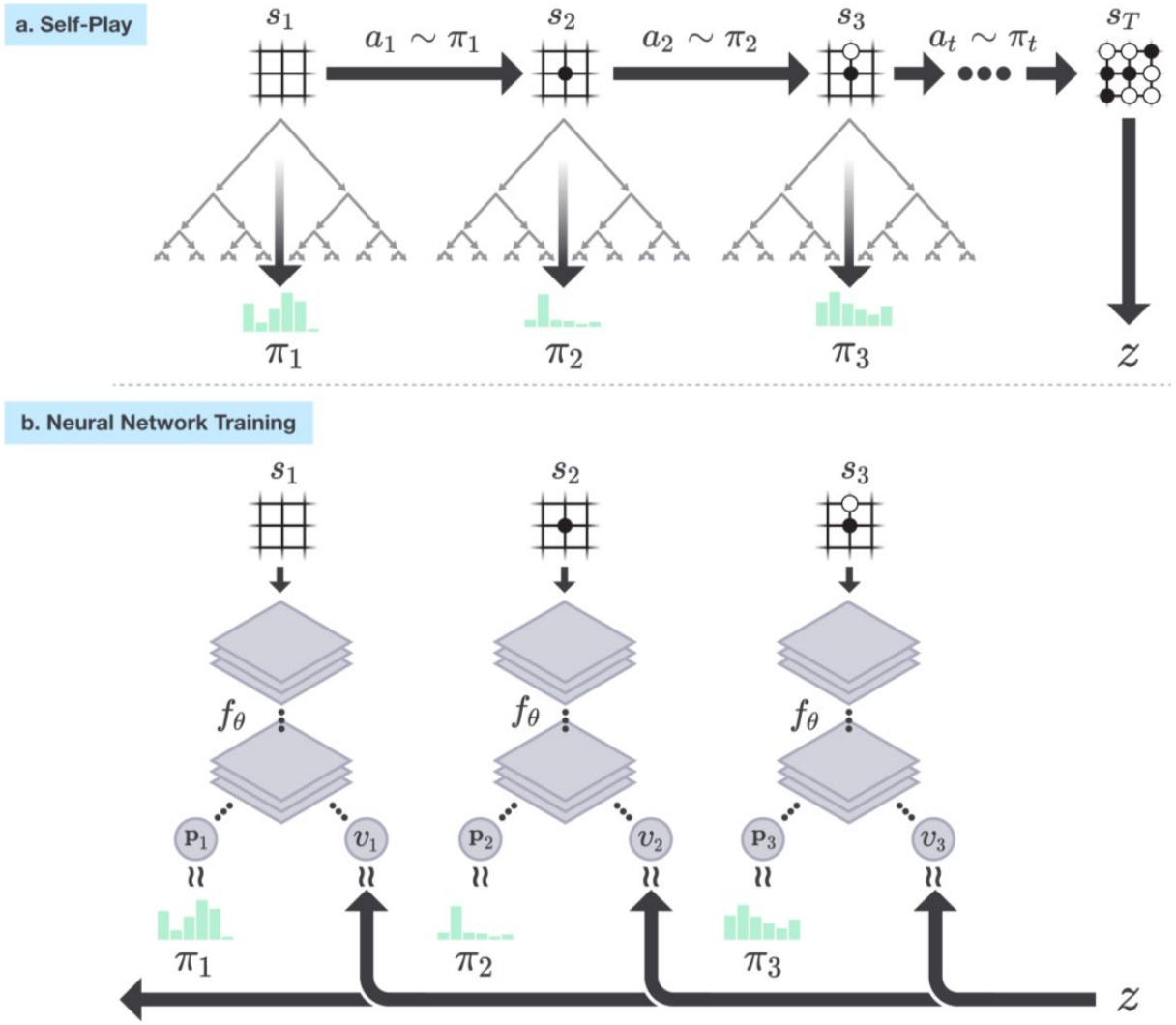
AlphaGo Zero



- AlphaGo:
  - policy network to predict moves
  - value network to predict winner
  - Monte-Carlo tree search
- Main novelties:
  - single network for policy and value
  - no roll-outs in tree search
  - trained from scratch, without human games or engineered features
  - used tree search results for training

https://deepmind.com/blog/alphago-zero-learning-scratch/
https://deepmind.com/documents/119/agz_unformatted_nature.pdf

AlphaGo Zero



https://deepmind.com/blog/alphago-zero-learning-scratch/
https://deepmind.com/documents/119/agz_unformatted_nature.pdf

Vienna Deep Learning Meetup

Orthogonal Recurrent Neural Networks with Scaled Cayley Transform

- Problem in deep or recurrent networks: Exploding / vanishing gradient
- Glorot/Bengio (2011): Scale weights to keep variance of propagated data

- Remaining problem: Ill-conditioned weights (vastly different singular val.)
- Saxe et al. (2014): Initialize as random orthogonal matrices

- Remaining problem: Well-conditioned only initially, can change in training
- Arjovsky et al. (2016): Unitary RNNs, forced to unitary singular values

- Remaining problem: Unitary RNNs are expensive, use complex values
- Idea: Use **Cayley transform** to create orthogonal from skew-sym. matrix:

$$W = (I + A)^{-1} (I - A), \quad \text{where } A^T = -A$$

Creates orthogonal W, and saves 50% of params.
Eigenvalues on unit circle in complex plane, but excluding -1 + 0j.
Multiply some columns with -1 to also allow reaching eigenvalue -1 + 0j.

https://arxiv.org/abs/1707.09520

Mathematical underpinning

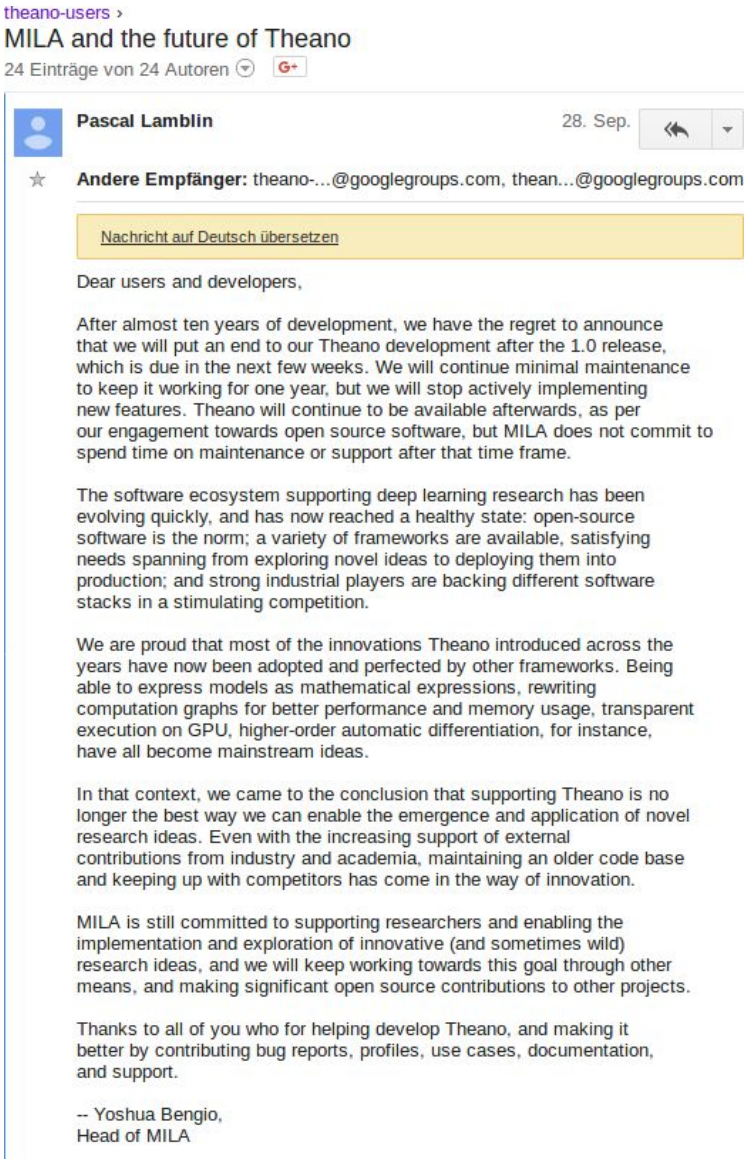
- David Duvenaud (Toronto): Deep learning ≈ Engineering without Physics
  - Paper A: “I made this bridge and it stood up!”
  - Paper B: “I made this bridge and it fell down—but then I added pillars, and then it stayed up.”
  - Pillars become popular, somebody discovers arches, and so on.

- Recent work on the missing “physics” side:
  - Kawaguchi/Kaelbing/Bengio (Oct 2017), Generalization in Deep Learning (https://arxiv.org/abs/1710.05468): Explains recent seemingly paradoxical findings, proposes regularizer
  - Fok/An/Wang (Oct 2017), Spontaneous Symmetry Breaking in Neural Networks (https://arxiv.org/abs/1710.06096): Also explains some recent findings using quantum field theory

photograph: Stephen Edmonds, cc-by-sa 2.0

Theano

- Theano will be discontinued after next release
- Minimal maintenance for one more year
- Reasons:
  - deep learning software ecosystem is healthy and diverse enough now
  - maintenance takes away too much time from innovative research



theano-users ›
MILA and the future of Theano
24 Einträge von 24 Autoren

Pascal Lamblin 28. Sep.

Andere Empfänger: theano-...@googlegroups.com, thean...@googlegroups.com

Nachricht auf Deutsch übersetzen

Dear users and developers,

After almost ten years of development, we have the regret to announce that we will put an end to our Theano development after the 1.0 release, which is due in the next few weeks. We will continue minimal maintenance to keep it working for one year, but we will stop actively implementing new features. Theano will continue to be available afterwards, as per our engagement towards open source software, but MILA does not commit to spend time on maintenance or support after that time frame.

The software ecosystem supporting deep learning research has been evolving quickly, and has now reached a healthy state: open-source software is the norm; a variety of frameworks are available, satisfying needs spanning from exploring novel ideas to deploying them into production; and strong industrial players are backing different software stacks in a stimulating competition.

We are proud that most of the innovations Theano introduced across the years have now been adopted and perfected by other frameworks. Being able to express models as mathematical expressions, rewriting computation graphs for better performance and memory usage, transparent execution on GPU, higher-order automatic differentiation, for instance, have all become mainstream ideas.

In that context, we came to the conclusion that supporting Theano is no longer the best way we can enable the emergence and application of novel research ideas. Even with the increasing support of external contributions from industry and academia, maintaining an older code base and keeping up with competitors has come in the way of innovation.

MILA is still committed to supporting researchers and enabling the implementation and exploration of innovative (and sometimes wild) research ideas, and we will keep working towards this goal through other means, and making significant open source contributions to other projects.

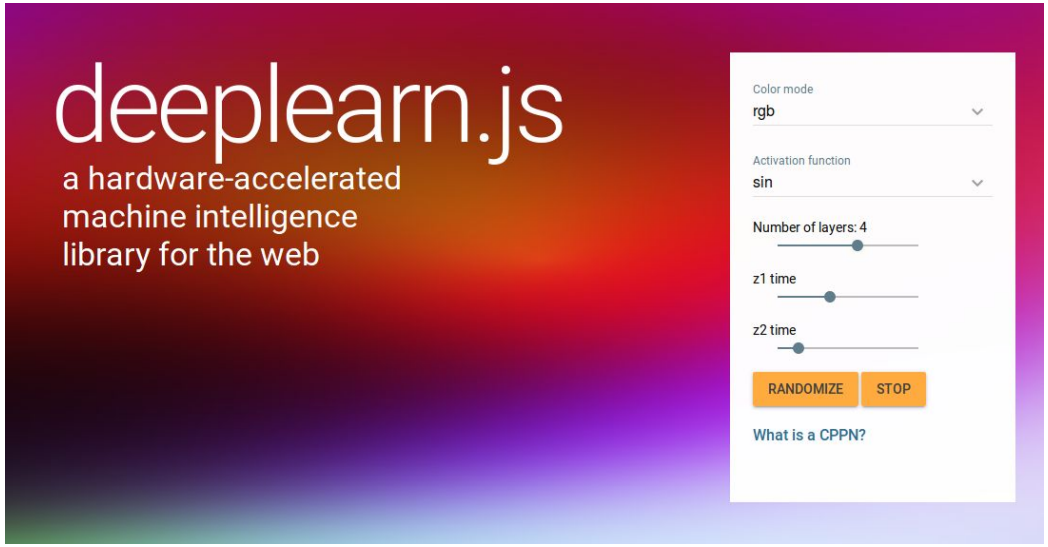
Thanks to all of you who for helping develop Theano, and making it better by contributing bug reports, profiles, use cases, documentation, and support.

-- Yoshua Bengio,
Head of MILA

https://groups.google.com/forum/#!topic/theano-users/7Poq8BZutbY

deeplearn.js

- Train and evaluate neural networks in web browser
- Supports MLPs, RNNs, CNNs
- Runs on GPU using WebGL (with CPU fallback)
  - convert input data to 2D texture
  - run pixel shader(s)
  - copy back results



https://deeplearnjs.org/

Announcements

ARE YOU **READY** to be a DUALISTA?

DUALISTA is a job initiative launched by the TEC GRUPPE in cooperation with INiTS. Goal is to provide a maximum of flexibility, individuality and job security.

The DUALISTA will be hired by one of the companies that are part of the TEC GRUPPE. The program consists of three focus areas:

- **DUALISTA start-up:** work hours will be equally spent between the company and the start-up
- **DUALISTA education:** work processes will be covered within dual studies
- **DUALISTA life-balance:** in collaboration with your employer you create your own individual job format that fits perfektly to what you need.

Further information and application: www.dualista.at


Hey,
I am ONDEWO's Deep Learning Dream Machine
and want to be built ☺
CORE i9
SILENT WINGS 3
SILENT WINGS 3
H115i
PRIME X299-DELUXE
AURA SYNC
thermaltake
3x "free" tickets to build together with us our deep learning server - old school with pizza and beer :-)

Vienna Data Science Group

- **Mission**
  - NPO, aiming to promote knowledge about data science

- **Diverse memberbase**
  - Academics, professionals, students and Data Science enthusiasts
  - Mathematics, physics, econometrics, electrical engineering, medical science, finance, real estate, computer science, social sciences...

- **International scope**
  - Founded in Vienna - aiming to connect Data Scientists all across Europe!

VDSG
VIENNA DATA SCIENCE GROUP

So… What does the VDSG actually do?

We aspire to provide benefit to the DS Community via…

- Various event formats
  - Knowledge Feed
  - Data Science Café
  - Hackathons
- Cooperations with companies
- Networking opportunities
- … a couple of other ways which are yet to come :)

VDSG
VIENNA DATA SCIENCE GROUP

Get in touch with us!

Homepage: www.vdsg.at
Email: contact@vdsg.at
Twitter: @ViennaDSG
Meetup: “Vienna Data Science Group”

VDSG
VIENNA DATA SCIENCE GROUP

THANKS TO

Our Sponsor

Organizing Hosts:
Christoph Toeglhofer
Tamara Berger-Feichter
Boris Marte

Our Streaming Partner

STREEMED

Organizer:
Maria Vasilevich

https://www.youtube.com/watch?v=yJiBqVxqfmA

Vienna Deep Learning Meetup

Next

Vienna Deep Learning Meetup

November 20, 2017 @ Telekom A1
"Evolution of Image Search @ Seznam.cz"


Thomas Lidy


Jan Schlüter


Alex Schindler