1. BERT Applications

Gerhard Binder

Why BERT performe so well

- For BERT architecture see the notes
- Reasons for the performance on NLP-tasks is yet not well understood; some research works focus the analysis on:
  - the way input sequences where choosing,
  - the internal vector representations,
  - the relationships represented by attention weights.

BERT Applications

- Applying BERT models for search queries (google adopted for 70 languages)
- Create contextualized word embeddings
- Sentence classification, tagging …
- Question answering

3. AdaNet: AutoML with Ensembles

Alessio Ciullo

Goal

- Automate as many aspects as possible of a ML project

- The user only drops in data and the system sets the rest (e.g. architecture, hyperparameters, etc.) as well as performing training and validation

Attempts

- Google AutoML has been trying to do so using reinforcement learning and Recurrent Neural Networks (RNN) with the goal of optimizing the performance of the **overall system**

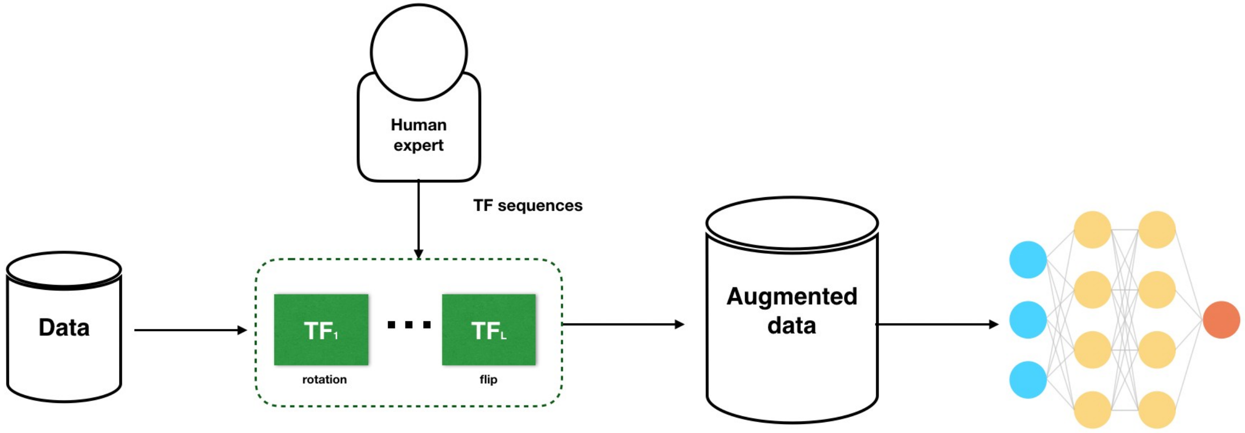
- AdaNet: it uses reinforcement learning to build **ensembles** of the networks, i.e. it makes use of various network architectures

- 4. AutoAugment:
- Deep RL Data Augmentation

- Desilvestro Valentino

- **Why data augmentation?**

- Not enough labeled data: need to learn a lot from a little!



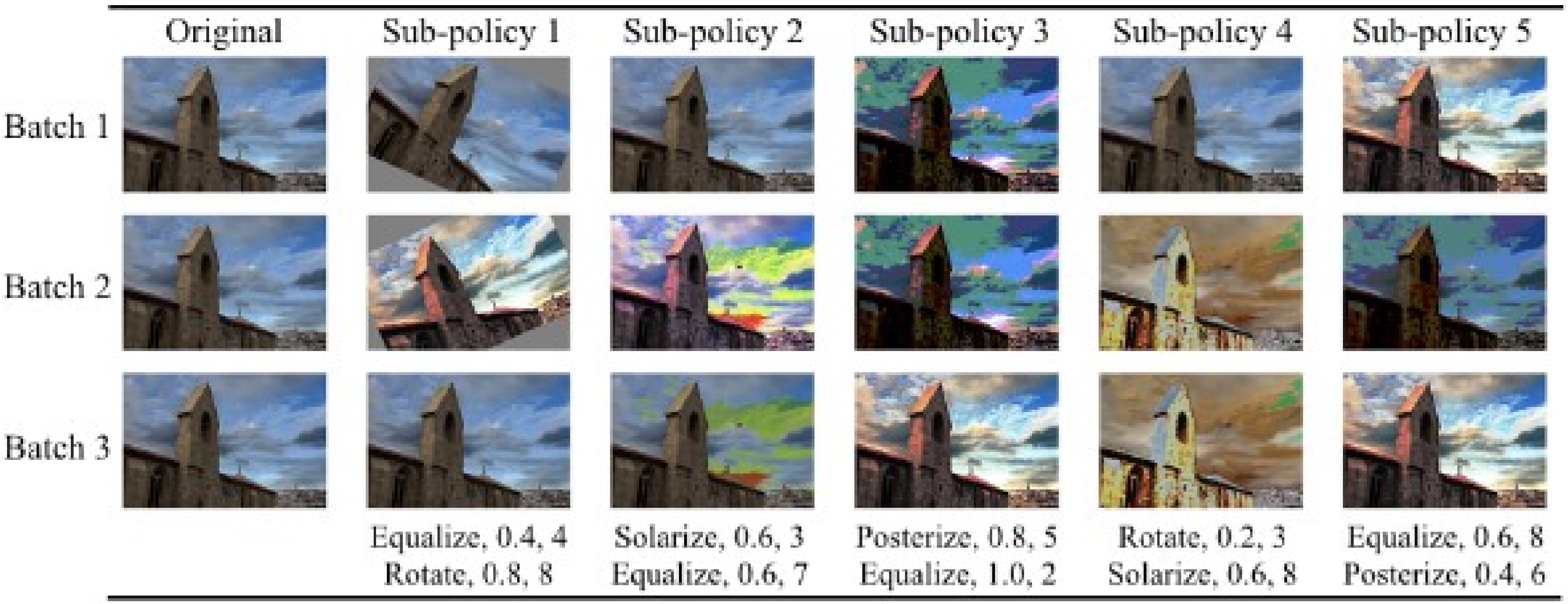
- **AutoAugment innovations:**
- Use RNN to find out which augmentation policies improve performance
- Transfer learning: can apply optimal policies to new data/problems

- **How does AutoAugment work?**

1) Determine a search space, with one policy consisting of many sub-policies (2 image processing functions, each with 2 hyperpar (prob, magnitude))

2) For each image in each mini-batch, randomly select one policy

3) Algor



Useful links:

https://arxiv.org/pdf/1805.09501.pdf

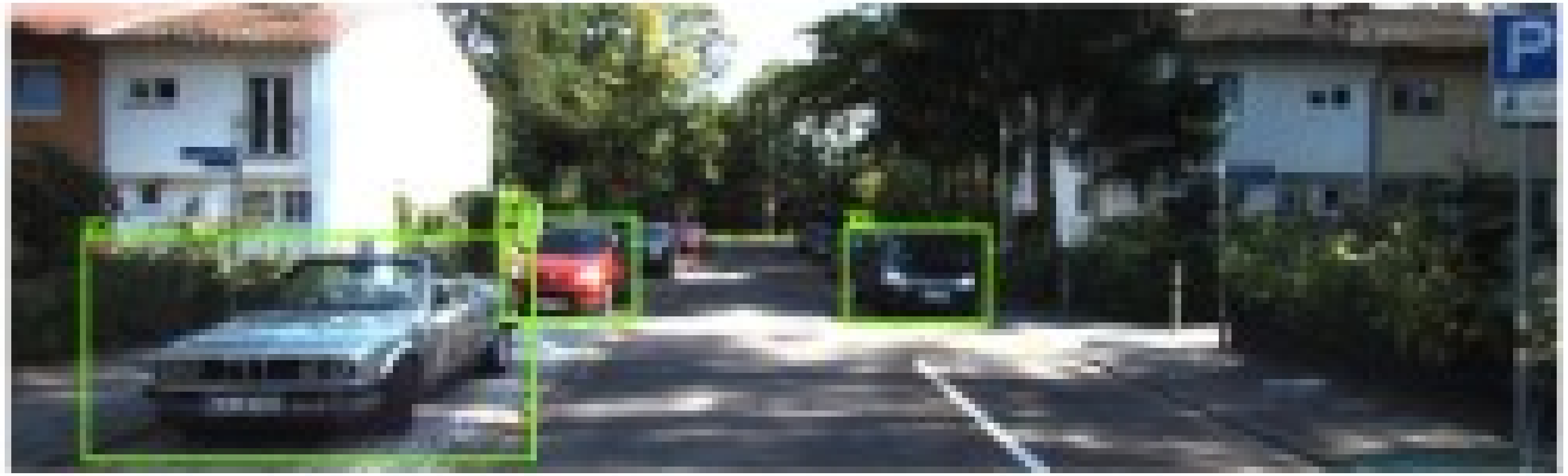
http://ai.stanford.edu/blog/data-augmentation/

5. Training Deep Networks with **Synthetic** Data

CAS AML Uni Bern

Michael Freunek

Goal: Object detection

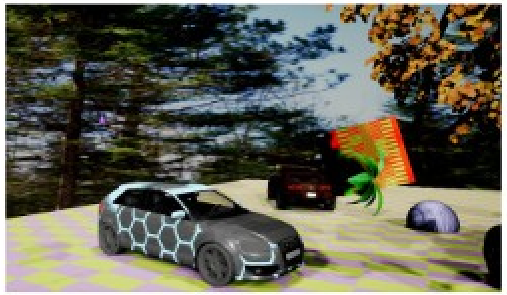


Why training with („messed“) synthetic data?

- Real data are „expensive“ (hard to collect)
- Also: High-fidelity synthetic data are „expensive“ to generate
- Training with (not so much) real data get poor results

How are the (messed) synthetic data generated?

1. Domain randomization ☾ lighting, pose, textures… of the environment is randomized to non-realistic values
2. Rotation, scaling, lighting etc. of the objects of interest
3. Flying distractors added to the scenes (flying cubes etc.)
4. Synthetic fine tuning/augmenting/modification of real da



Results / Benefits and Reference

- Model learns „to focus“ on relevant features in complex environments ☾ Outperformance of models trained only with real data
- Cheaper way of creating trained models
- Applicant: Nvidia

Reference: Training Deep Networks with Synthetic Data: Bridging the Reality Gap by Domain Randomization, Jonathan Tremblay, Aayush Prakash, David Acuna, Mark Brophy, Varun Jampani, Cem Anil, Thang To, Eric Cameracci, Shaad Boochoon, Stan Birchfield, 2018, eprint1804.06516, arXiv

6. Segmentation Annotation with Polygon-RNN++

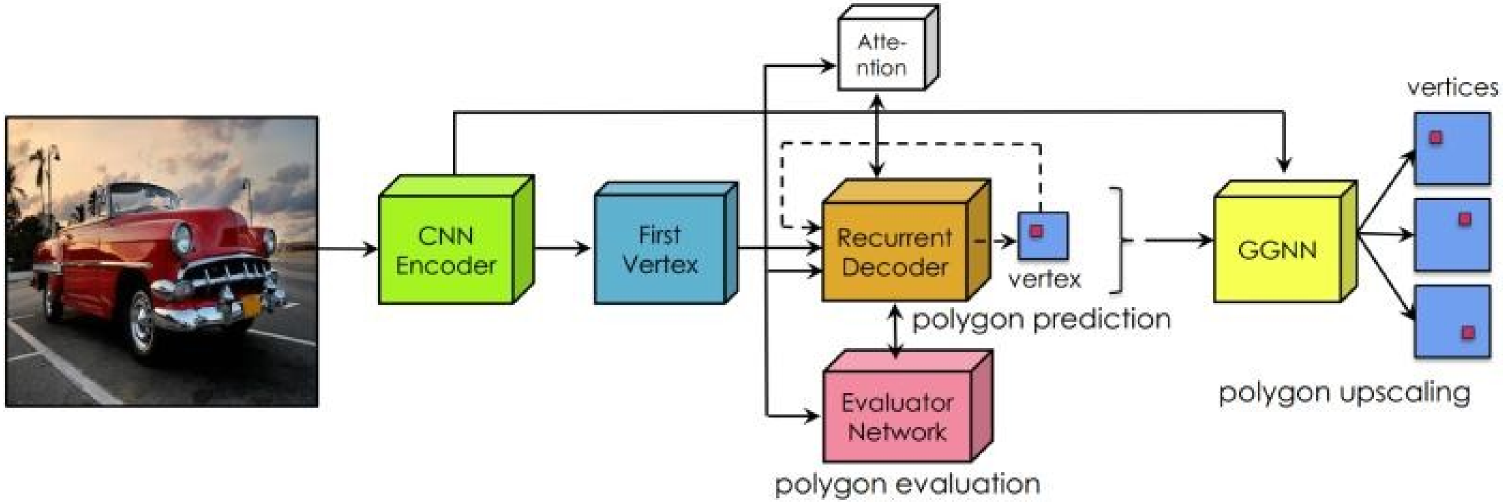
Andres Anobile Perez

What does it do

Used as an aid tool for object segmentation in digital images

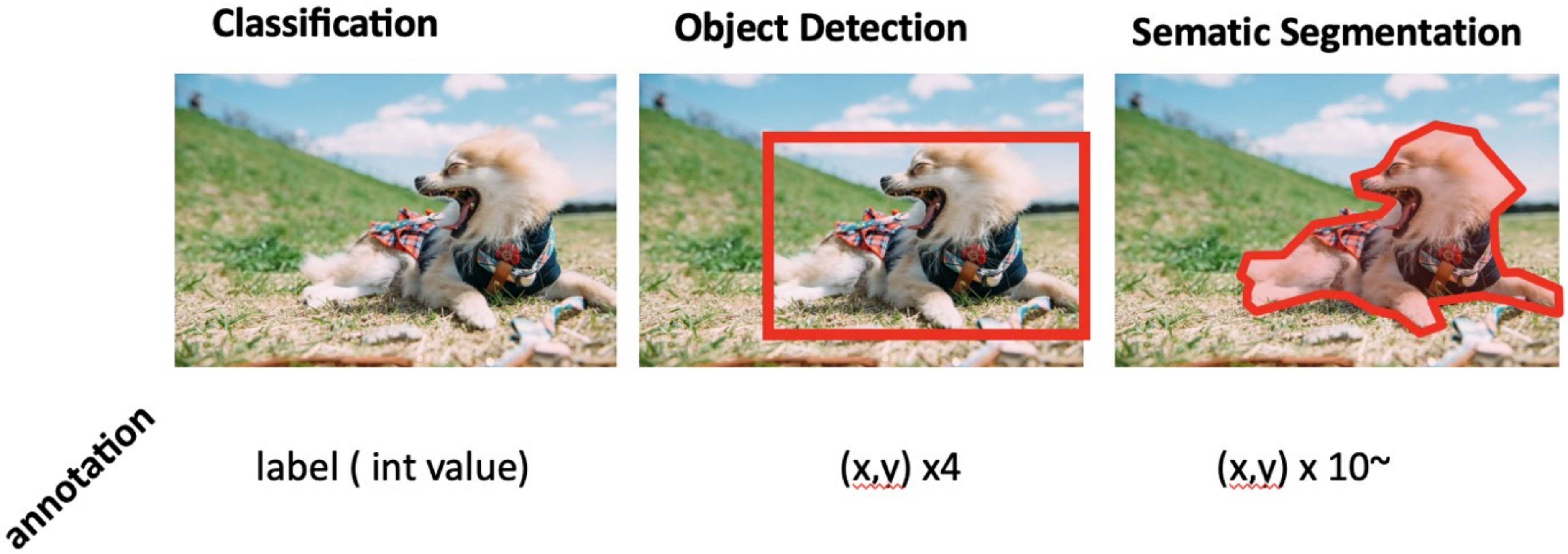
How does it do it

Uses RNN to suggest polygons to segment an object inside a boundary box.



Process workflow

- Label the image
- Draw a boundary box
- RNN++ suggest a segmentation
- Fine adjust manually



Classification

Object Detection

Sematic Segmentation

annotation

label (int value)

(x,y) x4

(x,y) x 10~

References

Semi-automated Annotation model, Polygon RNN, Polygon RNN++
https://medium.com/@akichan_f/semi-automated-annotation-model-polygon-rnn-polygon-rnn-fa3801019a29

Efficient Annotation of Segmentation Datasets with Polygon-RNN++
http://www.cs.toronto.edu/polyrnn/

7. DAWNBench : Training Fast and Cheap

Janicek Radoslav

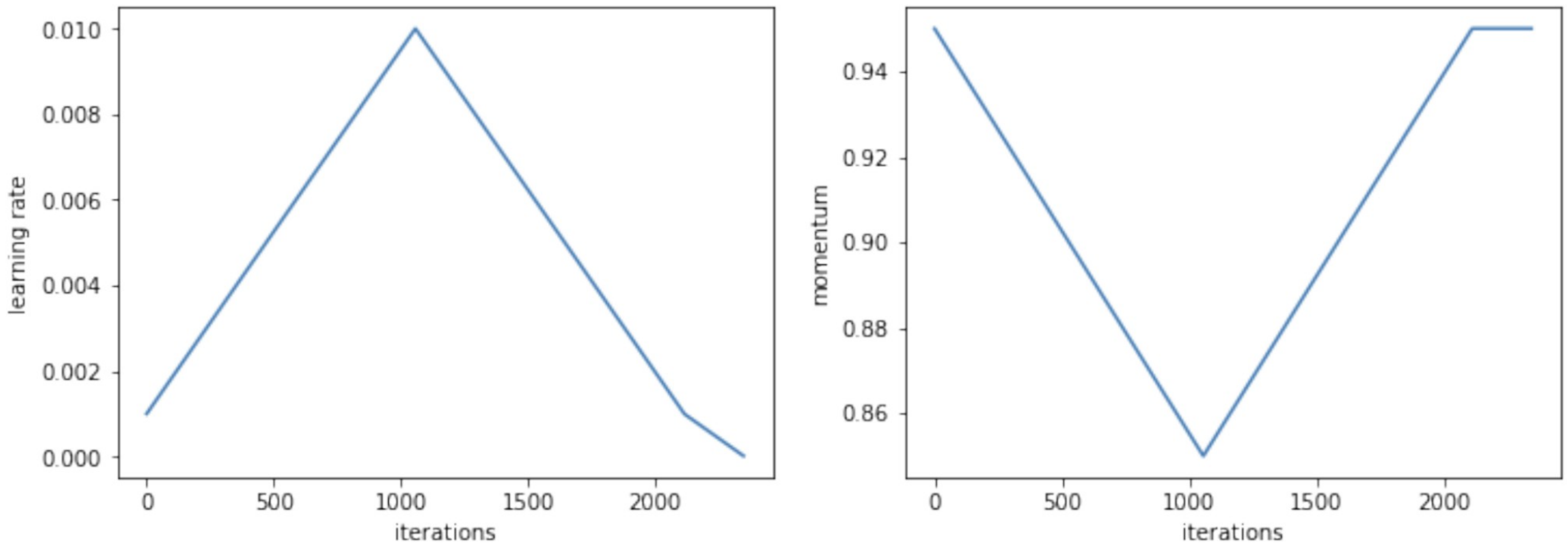
DAWNBench: Training fast and cheap

DAWNBench:

- Stanford University project designed to allow different deep learning methods to be compared by running a number of competitions.
- fastest and cheapest image classifier with defined accuracy (94 % for CIFAR 10 and 93% for Imagenet).

research group **fast.ai**:

- ImageNet in 3 hours for $25
- CIFAR10 for $0.26
- **key idea:** During training you very slowly increase learning rate while decreasing momentum (train at extremely high learning rates, thus avoiding over-fitting, and training in far fewer epochs).



Learning rate and momentum schedules for super-convergence

[redacted]

[redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]

(a) 128×128 (b) 256×256 (c) 512×512

[redacted]

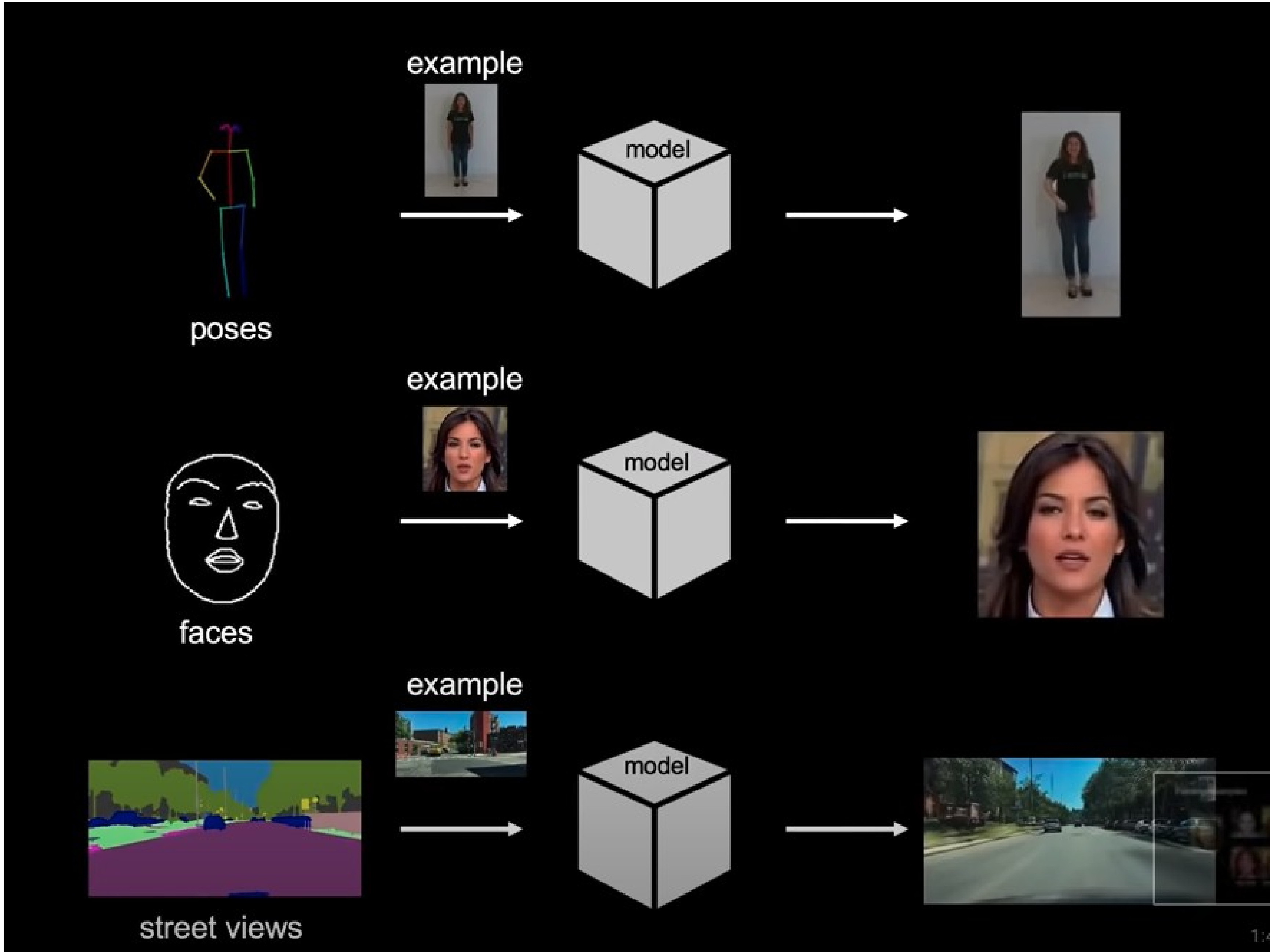
- Two minute papers
- [redacted]
- [redacted]
- [redacted]
- [redacted]

9. Video-to-video synthesis

Malte Sandow

Video to Video Synthesis

- From sketch to video
- Temporally smooth (not jumpy)
- Replace actors / speakers



Deep Learning State of the Art

10. Semantic segmentation

Christa Schneider

Semantic segmentation (object detection)

- Definition:
  - Describes the process of associating each pixel of an image with a certain label (mountain, lake, tree, person, car, …)
- Goal:
  - Describes what’s going on in an image
- Interaction with human ability
  - Human error rate = 5.1%
  - Surpassed in 2015!

Semantic segmentation (object detection

- How to use it:
  - Autonomous driving
  - Face recognition
  - …
- What’s behind:
  - Different CNNs with an increasing number of layers (AlexNet 2012 = 8 layers, ResNet 2015 = 152 layers
    - Region proposal
    - One-shot method

11. Symmetric Semantic Segmentation

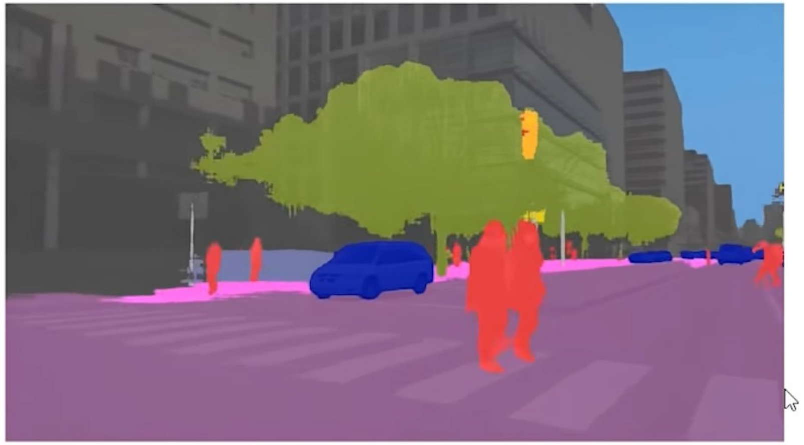
CAS Advanced Machine Learning – Module 2

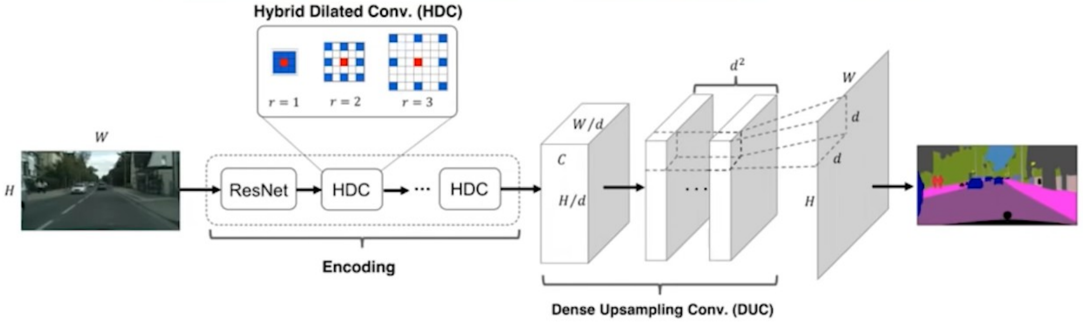
Christine Sigrist

Symmetric Semantic Segmentation

Semantic Segmentation:

- Perception problems
- Input = Image(s)
- Output = high resolution image in which each pixel is classified to a particular class
- Highest level of perception, high-level understanding from digital images, videos
- Automate tasks that the human visual system can do
- **Dense prediction**: label each pixel of an image with a corresponding class of what is being represented





Symmetric Semantic Segmentation

Started in 2014:

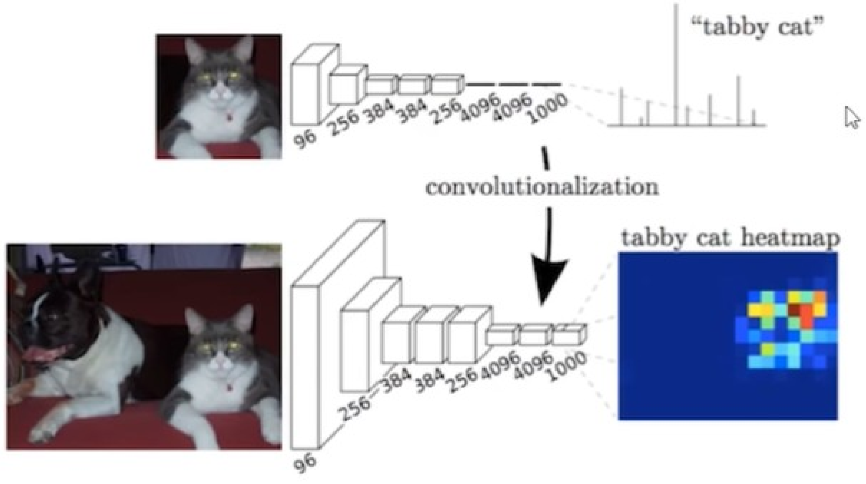
- Fully Convolutional Neural Networks (FCN) used for Semantic Segmentation
  - Outputting low resolution heatmaps

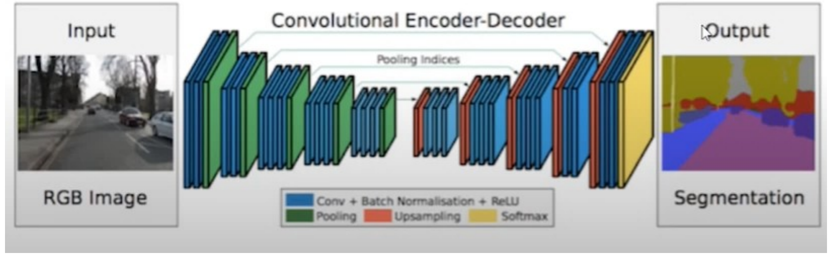
Improvements in 2015:

- SegNet (= Deep Convolutional Encoder-Decoder Architecture)
  - Maxpooling indices transferred to the decoder to improve the segmentation resolution

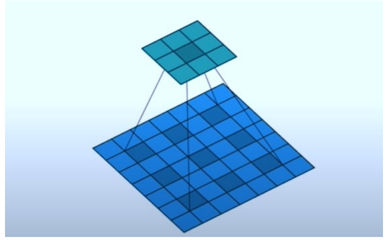
Breakthrough idea: Dilated Convolutions

  - Pooling decreases resolution – a dilated convolution layer was added
  - Interpolate up from 1/8 of original image size





SegNet



Dilated Convolutions

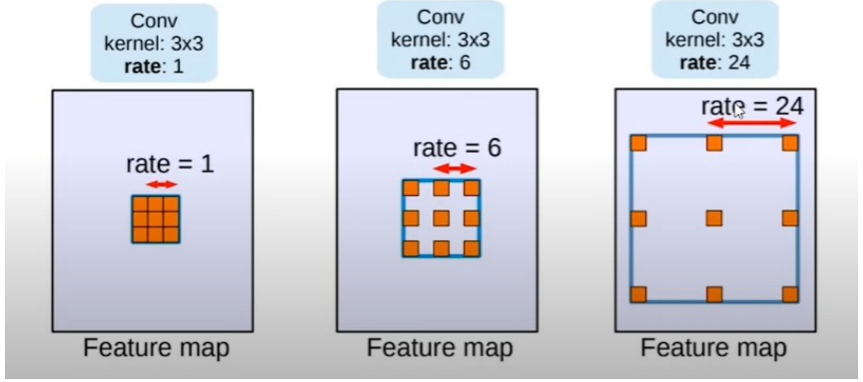
Symmetric Semantic Segmentation

DeepLab v3 (State-of-the-Art)

- **Produces "state-of-the-art performance"**

Key ideas:

- Multiscale processing by increasing "atrous rate" (**spacing**), without increasing parameters
- Enlarging the model's "field-of-view"
- Allows to be able to grasp the bigger context
- Applied e.g. for driving scene segmentation / autonomous driving systems



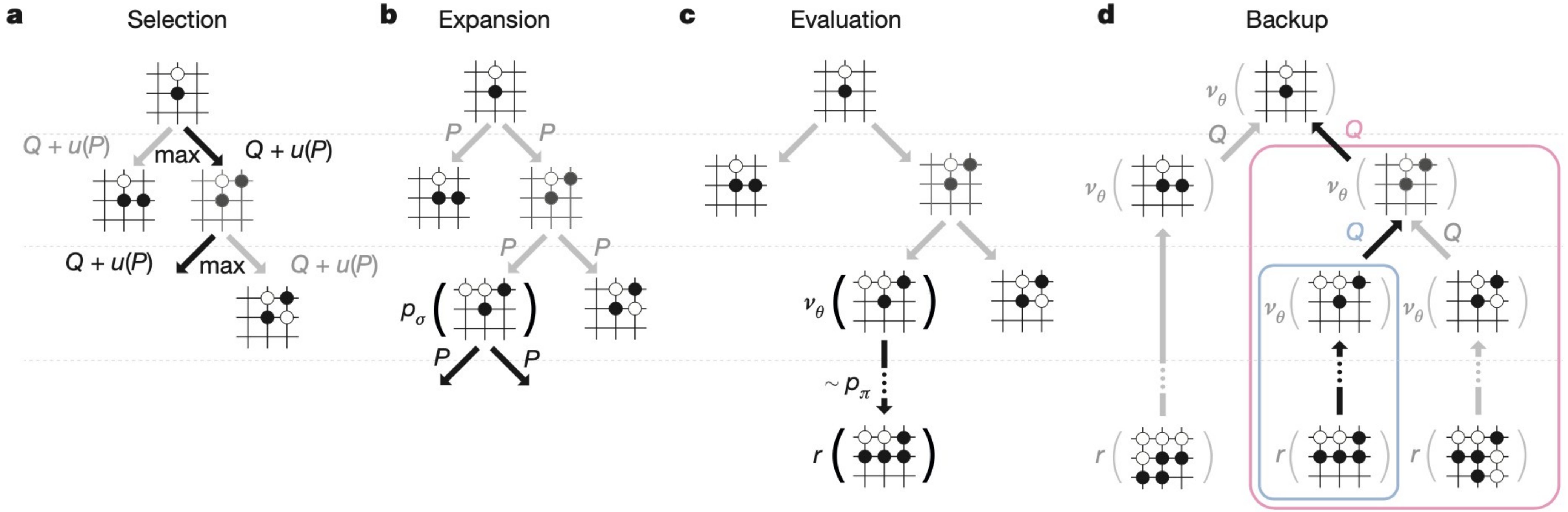
DeepLab v3 trained on CityScapes

12. Google’s AlphaGo

Ivo Vasconcelos

The AlphaGo Approach

- Value Networks -> Evaluate board positions
- Policy Networks -> Select moves
- Combination of Supervised Learning from human expert games and Reinforcement Learning from games of self-play
- Monte Carlo Three Search simulate thousands of random games of self-play
- Best leaf is selected for self-play games

a
Selection
Q + u(P)
max
Q + u(P)
Q + u(P)
max
Q + u(P)
b
Expansion
P
P
P
P
p_σ
P
P
c
Evaluation
v_θ
$\sim p_\pi$
r
d
Backup
v_θ
Q
Q
v_θ
v_θ
Q
Q
v_θ
v_θ
r
r
r

Reference

- *Deep Learning State of the Art (2019) - MIT*. 2019. [video] Boston: Lex Friedman on Youtube.
- Silver, D., Huang, A., Maddison, C., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T. and Hassabis, D., 2016. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), pp.484-489.

13. ALPHA ZERO

DEEP MIND TEAM

SCIENCE

VOL 362, ISSUE 6419
07 DECEMBER 2018

TAMARA VAUDROZ

ALPHA GO ZERO (2017)

- Beat AlphaGo Lee in 3 days of training

  Without any supervision, or any previous knowledge of the game.

  Only basic rules as input.

- In 40 days it beat every existing version of AlphaGo.

ALPHA ZERO (2017)

- Used 5000 first-generation TPUs to generate the games, and 64 second-generation TPUs to train the Neural Network (all in parallel) with no access to opening books or endgame tables.
- The trained algorithm played on a single machine with 4 TPUs.
- Alpha Zero beat StockFish in Chess with only 4 hours of training and Elmo in Shogi.

In Chess, there is always a tree of possible future moves. So, the furthest you can look down the tree, the better your moves will be.

StockFish is doing exactly this, billions of calculations…

Human Grand Masters are not playing like that. They can see the patterns in the board. They rely on a kind of intuition/feeling.

Alpha Zero works in this direction, it doesn’t need to make as many calculations. It estimates the quality of the board and moves, and learns the fundamental information.

In this sense, it is more human in its internal workings…

14. OpenAI vs. Humans on Dota 2

Fluri Wieland

Difference to Chess, Go etc

- The game is much less structured, its messier, since there are much more possibilities
- Not unlike the real world, the player can move around and so does the enemy
- Teamwork of players is now of big importance and emphasis on different skilltrees
- Uncertainty and hidden information are of importance now

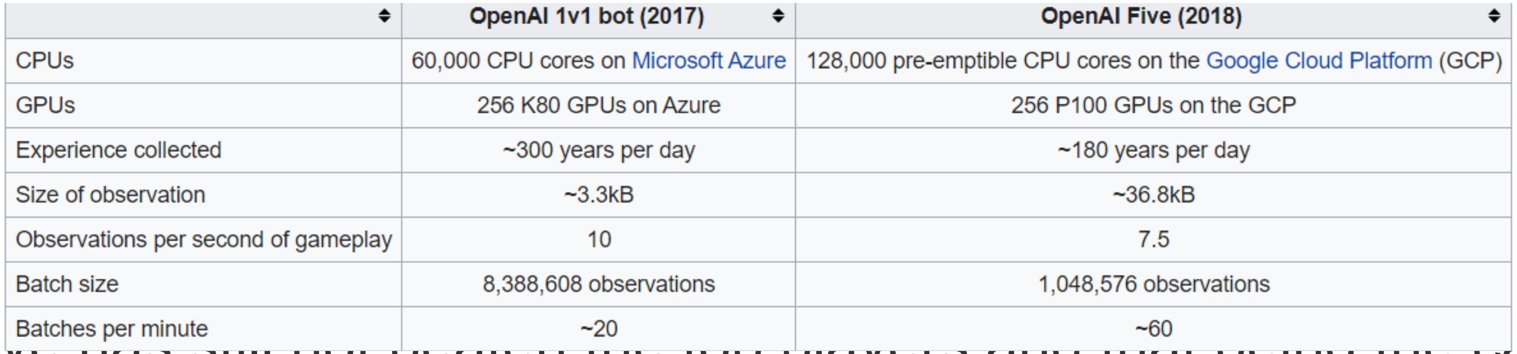
Dota 2 Championships

- Winning team gets 11 million dollars
- Very active scene with many professional players being fulltime players

Milestones

> 2017: 1v1, a OpenAI bot beats the top Dota2 Player in the world

> 2018: However, OpenAI five lost against the top Dota 2 Players at the international Championship in 5v5, twice



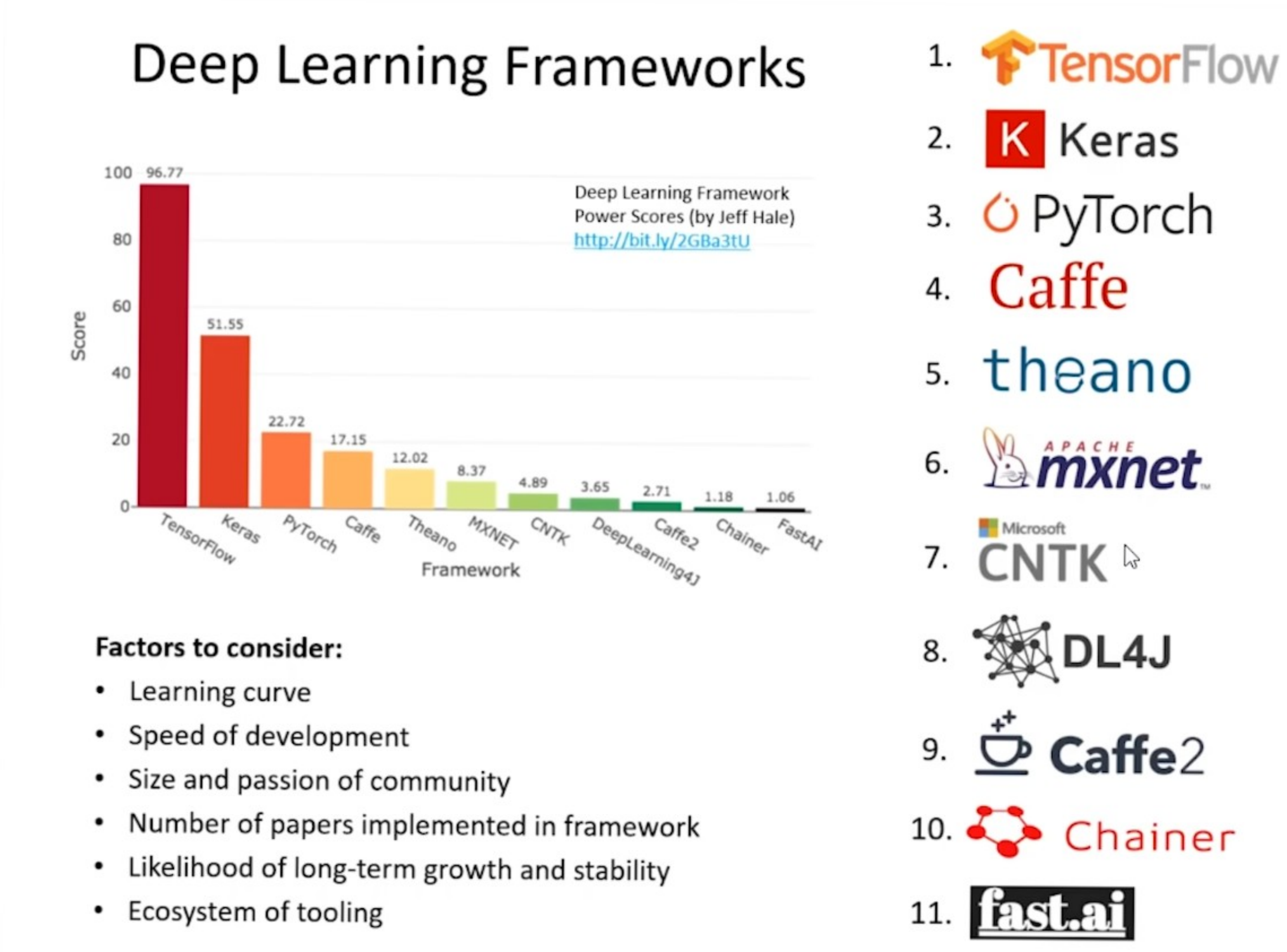
	OpenAI 1v1 bot (2017)	OpenAI Five (2018)
CPUs	60,000 CPU cores on Microsoft Azure	128,000 pre-emptible CPU cores on the Google Cloud Platform (GCP)
GPUs	256 K80 GPUs on Azure	256 P100 GPUs on the GCP
Experience collected	~300 years per day	~180 years per day
Size of observation	~3.3kB	~36.8kB
Observations per second of gameplay	10	7.5
Batch size	8,388,608 observations	1,048,576 observations
Batches per minute	~20	~60

> OpenAI five has still not beaten the top players and that being the case, Dota 2 is the current benchmark for AIs

15. Deep Learning Frameworks

Marcel Zeller

Deep learning frameworks



Deep Learning Frameworks

Deep Learning Framework Power Scores (by Jeff Hale)
http://bit.ly/2GBa3tU

Factors to consider:

- Learning curve
- Speed of development
- Size and passion of community
- Number of papers implemented in framework
- Likelihood of long-term growth and stability
- Ecosystem of tooling

1. TensorFlow
2. Keras
3. PyTorch
4. Caffe
5. theano
6. APACHE mxnet
7. Microsoft CNTK
8. DL4J
9. Caffe2
10. Chainer
11. fast.ai

Marcel Zeller

Deep learning

Deep learning frameworks

- Most of them are open-source
- Standardization of deep learning
- Accessibility and user friendly
- Big community, continuous development

Future of deep learning

- Ideas based on back propagation and stochastic gradient decent
- According to Geoffrey Hinton (Inventor of Deep learning) new ideas needed!!