Zürcher Hochschule
für Angewandte Wissenschaften

Deep Learning und Medien

IAM MediaLab, Winterthur, 06. Dezember 2018

Thilo Stadelmann

Swiss Alliance for
Data-Intensive Services

www.zhaw.ch/datalab

Prolog

Was ist künstliche Intelligenz?

thinking

"The exciting new effort to **make computers think**... machines with minds, in the full and literal sense." (Haugeland, 1985)

"[The automation of] **activities that we associate with human thinking**, activities such as decision-making, problem solving, learning..." (Bellman, 1978)

"The study of mental faculties through the use of **computational models**." (Charniak and McDermott, 1985)

"The study of the **computations that** make it possible to **perceive, reason, and act**." (Winston, 1992)

humanly ← standard measured by → rationally

← concerned with →

"The art **of creating machines that perform** functions that require **intelligence** when performed by people." (Kurzweil, 1990)

"The study of **how to make computers do things** at which, at the moment, **people are better**." (Rich and Knight, 1991)

"Computational Intelligence is **the study of the design of intelligent agents**." (Poole et al., 1998)

"AI... is concerned with **intelligent behaviour in artefacts**." (Nilsson, 1998)

acting

Was ist künstliche Intelligenz?

thinking

"The exciting new effort to **make computers think**... machines with minds, in the full and literal sense." (Haugeland, 1985)

"[The automation of] **activities that we associate with human thinking**, activities such as decision-making, problem solving, learning..." (Bellman, 1978)

"**The study of mental faculties through** the use of **computational models**." (Charniak and McDermott, 1985)

"The study of the **computations that** make it possible to **perceive, reason, and act**." (Winston, 1992)

humanly ← Standard measured by → rationally

← concerned with →

"The art **of creating machines that perform functions** that require **intelligence** when performed by people." (Kurzweil, 1990)

"The study of **how to make computers do things** at which, at the moment, **people are better**." (Rich and Knight, 1991)

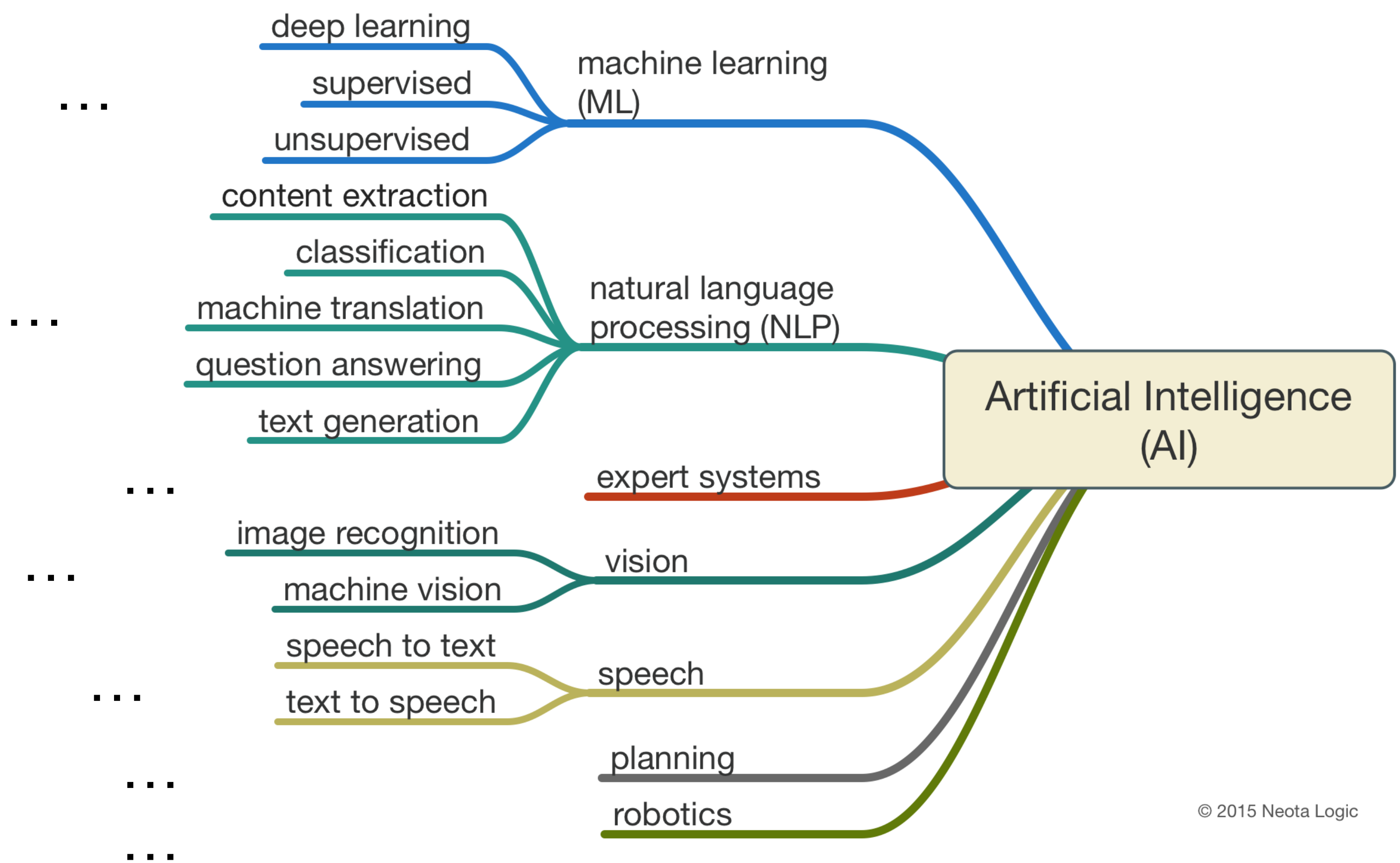
"Computational Intelligence is **the study of the design of intelligent agents**." (Poole et al., 1998)

"AI... is concerned with **intelligent behaviour in artefacts**." (Nilsson, 1998)

acting

The scientific field concerned with complex computer applications

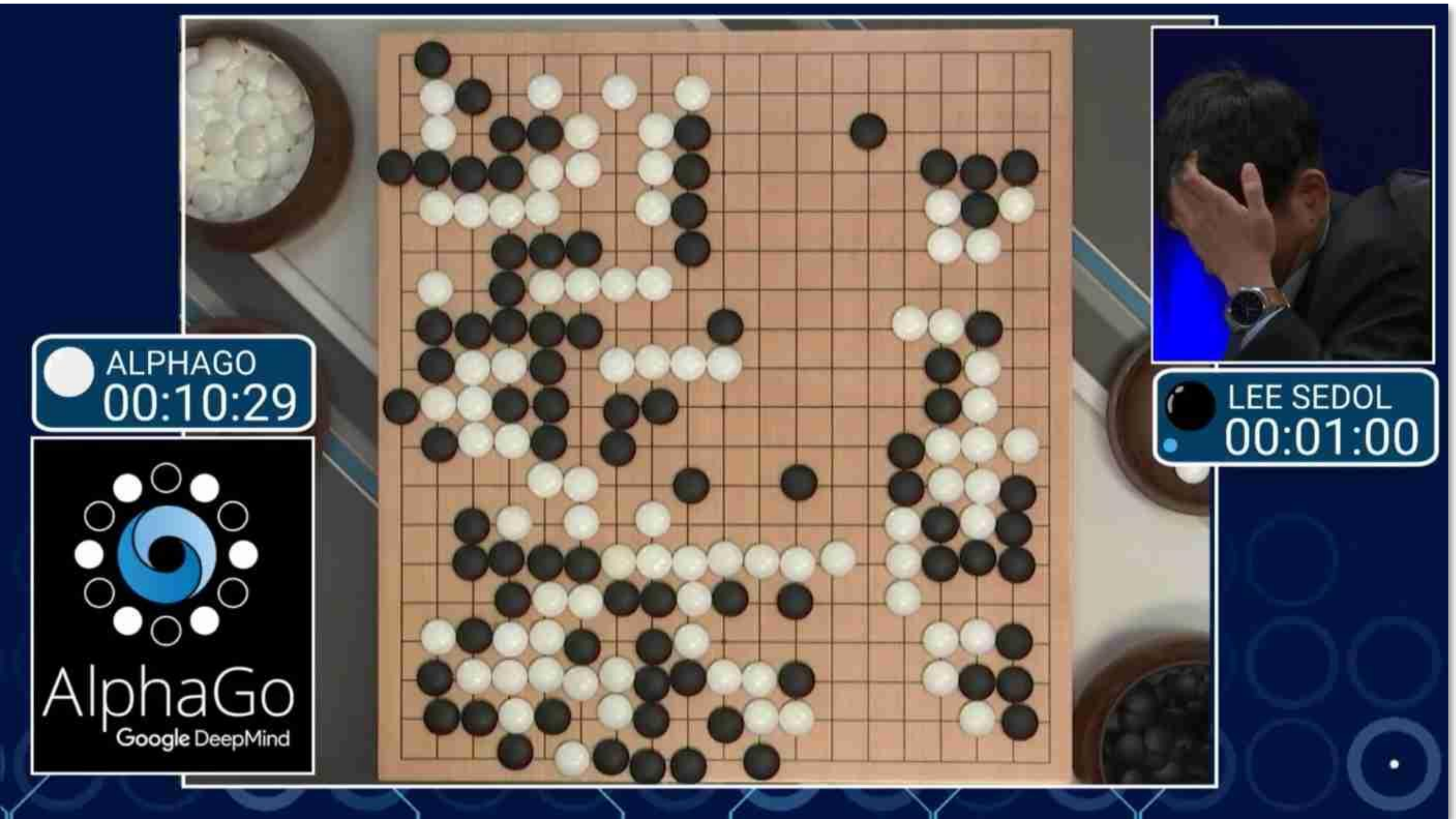
Was gehört zu künstlicher Intelligenz?

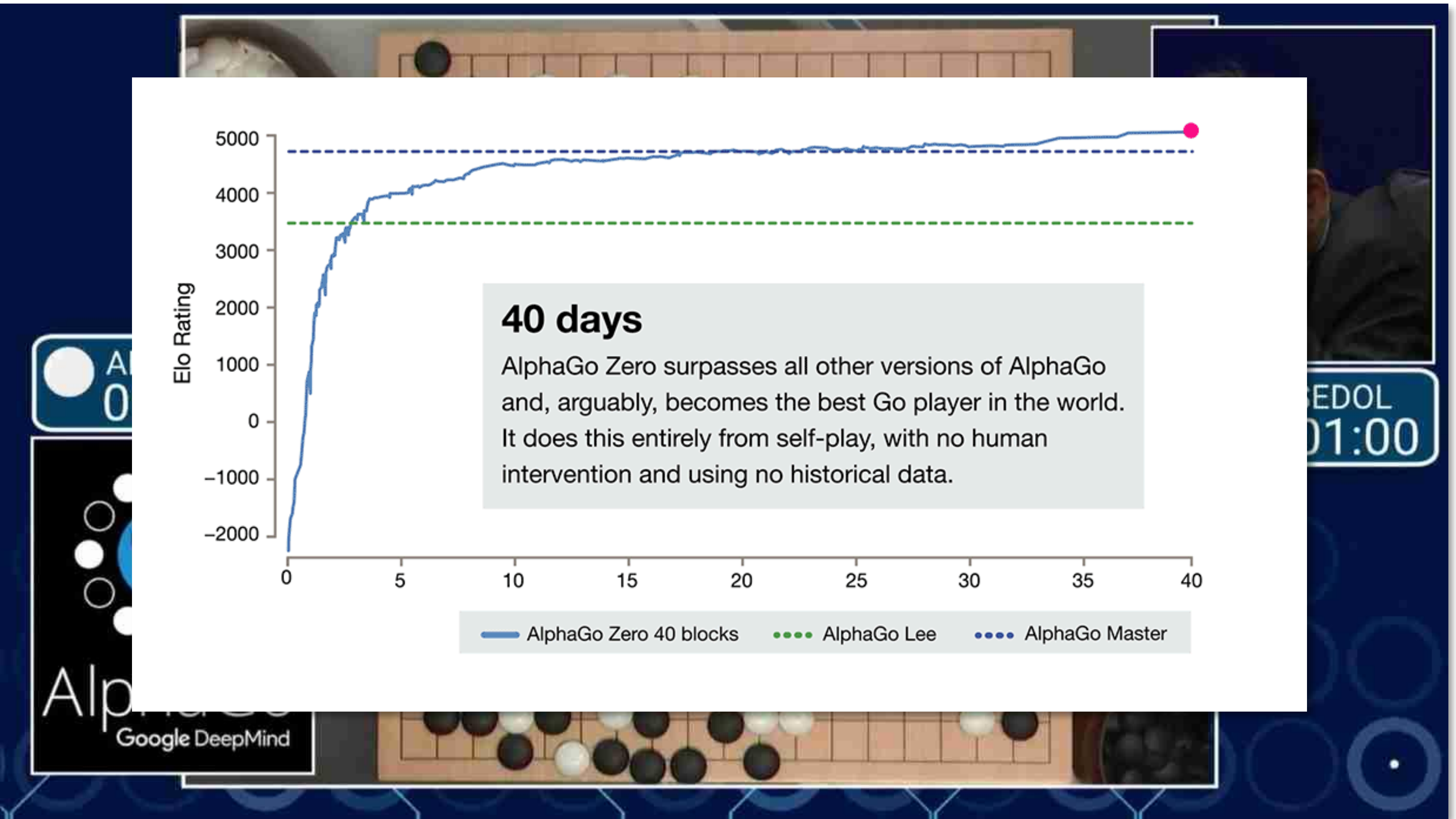


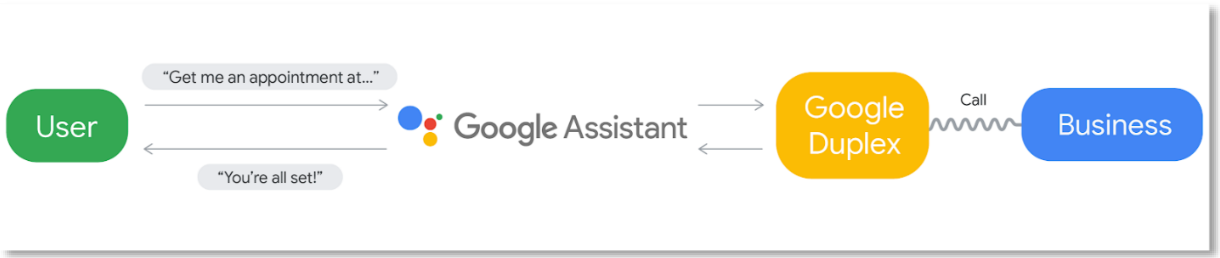
Zürcher Hochschule
für Angewandte Wissenschaften

1

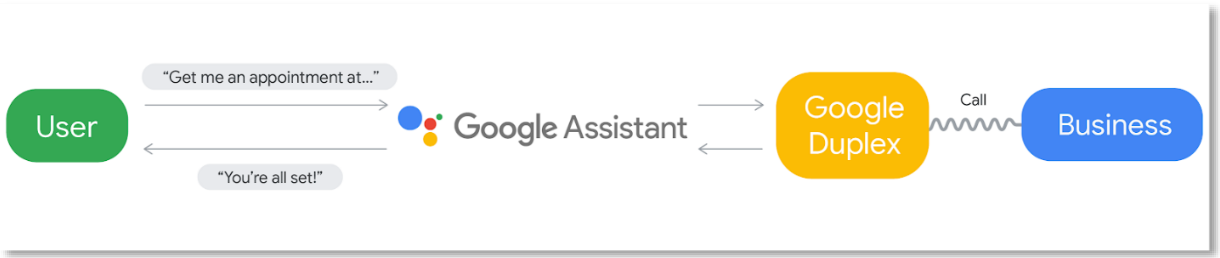
Was ist passiert?
(Eine kurze Geschichte der letzten Jahre)

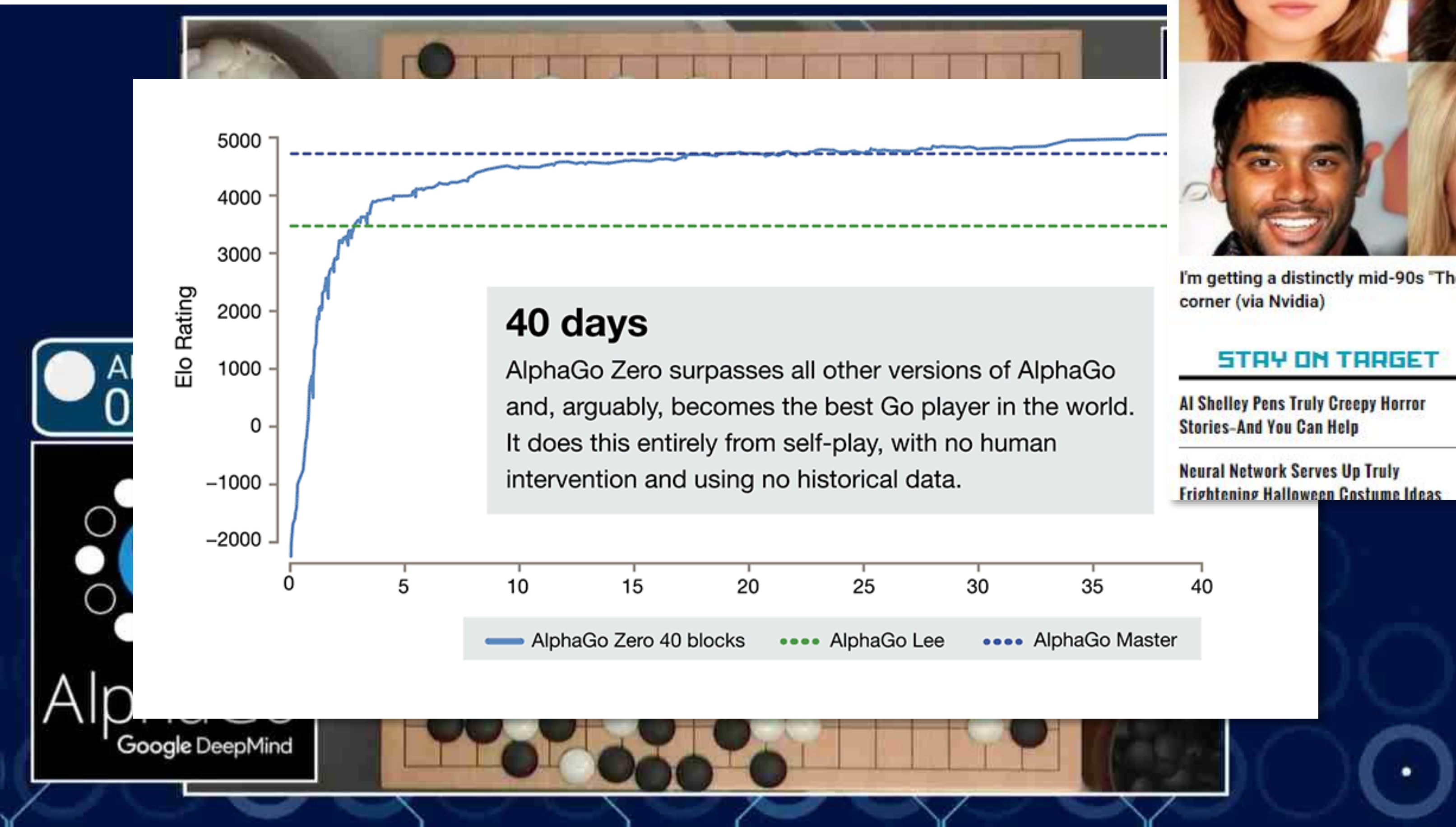

ALPHAGO
00:10:29
LEE SEDOL
00:01:00
AlphaGo
Google DeepMind


5000
4000
3000
2000
1000
0
−1000
−2000
Elo Rating
0
5
10
15
20
25
30
35
40
40 days
AlphaGo Zero surpasses all other versions of AlphaGo and, arguably, becomes the best Go player in the world. It does this entirely from self-play, with no human intervention and using no historical data.
AlphaGo Zero 40 blocks
AlphaGo Lee
AlphaGo Master
Google DeepMind
EDOL
01:00


User
"Get me an appointment at..."
"You're all set!"
Google Assistant
Google Duplex
Call
Business


Elo Rating
5000
4000
3000
2000
1000
0
−1000
−2000
0
5
10
15
20
25
30
35
40
40 days
AlphaGo Zero surpasses all other versions of AlphaGo and, arguably, becomes the best Go player in the world. It does this entirely from self-play, with no human intervention and using no historical data.
AlphaGo Zero 40 blocks
AlphaGo Lee
AlphaGo Master
Google DeepMind





GEEK.COM

TECH

Nvidia AI Generates Fake Faces Based On Real Celebs

BY STEPHANIE MLOT 10.31.2017 :: 10:00AM EST

32 SHARES

I'm getting a distinctly mid-90s "The Rachel" vibe from the woman in the top left corner (via Nvidia)

STAY ON TARGET

AI Shelley Pens Truly Creepy Horror Stories–And You Can Help

Neural Network Serves Up Truly Frightening Halloween Costume Ideas

Celebrity scandals are about to get a lot more complicated.

Nvidia has **developed** a way of producing photo-quality, AI-generated human profiles—by using famous faces.

Zürcher Hochschule
für Angewandte Wissenschaften

…und die Liste liesse sich fortsetzen!

the morning paper

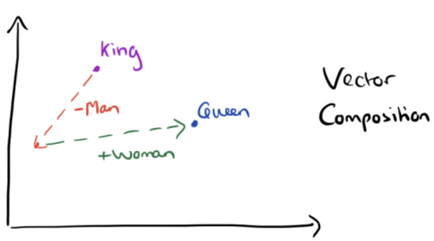
The amazing power of word vectors

APRIL 21, 2016

For today's post, I've drawn material not just from one paper, but from five! The subject matter is 'word2vec' – the work of Mikolov et al. at Google on efficient vector representations of words (and what you can do with them). The papers are:

- **Efficient Estimation of Word Representations in Vector Space** – Mikolov et al. 2013
- **Distributed Representations of Words and Phrases and their Compositionality** – Mikolov et al. 2013
- **Linguistic Regularities in Continuous Space Word Representations** – Mikolov et al. 2013
- **word2vec Parameter Learning Explained** – Rong 2014
- **word2vec Explained: Deriving Mikolov et al's Negative Sampling Word-Embedding Method** – Goldberg and Levy 2014

From the first of these papers ('Efficient estimation…') we get a description of the *Continuous Bag-of-Words* and *Continuous Skip-gram* models for learning word vectors (we'll talk about what a word vector is in a moment…). From the second paper we get more illustrations of the power of word vectors, some additional information on optimisations for the skip-gram model (hierarchical softmax and negative sampling), and a discussion



…und die Liste liesse sich fortsetzen!

the morning paper

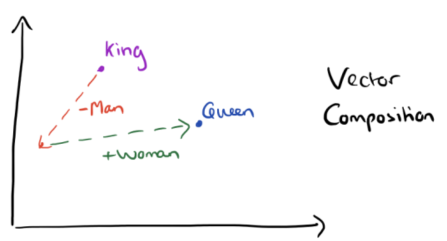
The amazing power of word vectors

APRIL 21, 2016

For today's post, I've drawn material not just from one paper, but from five! The subject matter is 'word2vec' – the work of Mikolov et al. at Google on efficient vector representations of words (and what you can do with them). The papers are:

- Efficient Estimation of Word Representations in Vector Space – Mikolov et al. 2013
- Distributed Representations of Words and Phrases and their Compositionality – Mikolov et al. 2013
- Linguistic Regularities in Continuous Space Word Representations – Mikolov et al. 2013
- word2vec Parameter Learning Explained – Rong 2014
- word2vec Explained: Deriving Mikolov et al's Negative Sampling Word-Embedding Method – Goldberg and Levy 2014

From the first of these papers ('Efficient estimation…') we get a description of the *Continuous Bag-of-Words* and *Continuous Skip-gram* models for learning word vectors (we'll talk about what a word vector is in a moment…). From the second paper we get more illustrations of the power of word vectors, some additional information on optimisations for the skip-gram model (hierarchical softmax and negative sampling), and a discussion



Andrej Karpathy blog — About — Hacker's guide to Neural Networks

The Unreasonable Effectiveness of Recurrent Neural Networks

May 21, 2015

There's something magical about Recurrent Neural Networks (RNNs). I still remember when I trained my first recurrent network for Image Captioning. Within a few dozen minutes of training my first baby model (with rather arbitrarily-chosen hyperparameters) started to generate very nice looking descriptions of images that were on the edge of making sense. Sometimes the ratio of how simple your model is to the quality of the results you get out of it blows past your expectations, and this was one of those times. What made this result so shocking at the time was that the common wisdom was that RNNs were supposed to be difficult to train (with more experience I've in fact reached the opposite conclusion). Fast forward about a year: I'm training RNNs all the time and I've witnessed their power and robustness many times, and yet their magical outputs still find ways of amusing me. This post is about sharing some of that magic with you.

> *We'll train RNNs to generate text character by character and ponder the question "how is that even possible?"*

By the way, together with this post I am also releasing code on Github that allows you to train character-level language models based on multi-layer LSTMs. You give it a large chunk of text and it will learn to generate text like it one character at a time. You can also use it to reproduce my experiments below. But we're getting ahead of ourselves; What are RNNs anyway?

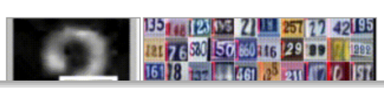
Recurrent Neural Networks

Sequences. Depending on your background you might be wondering: *What makes Recurrent Networks so special*? A glaring limitation of Vanilla Neural Networks (and also Convolutional Networks) is that their API is too constrained: they accept a fixed-sized vector as input (e.g. an image) and produce a fixed-sized vector as output (e.g. probabilities of different classes). Not only that: These models perform this mapping using a fixed amount of computational steps (e.g. the number of layers in the model). The core reason that recurrent nets are more exciting is that they allow us to operate over *sequences* of vectors: Sequences in the input, the output, or in the most general case both. A few examples may make this more concrete:

```
VIOLA:
Why, Salisbury must find his flesh and thought
That which I am not aps, not a man and in fire,
To show the reining of the raven and the wars
To grace my hand reproach within, and not a fair are hand,
That Caesar and my goodly father's world;
When I was heaven of presence and our fleets,
We spare with hours, but cut thy council I am great,
Murdered and by thy master's ready there
My power to give thee but so much as hell:
Some service in the noble bondman here,
Would show him to her wine.

KING LEAR:
O, if you were a feeble sight, the courtesy of your law,
Your sight and several breath, will wear the gods
With his heads, and my hands are wonder'd at the deeds,
So drop upon your lordship's head, and your opinion
Shall be against your honour.
```

the right, a recurrent network generates images of digits by learning to sequentially add color to a canvas (Gregor et al.):



…und die Liste liesse sich fortsetzen!

the morning paper

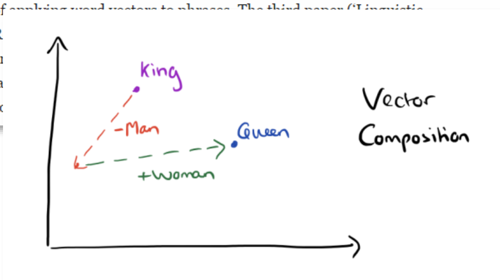
The amazing power of word vectors

APRIL 21, 2016

For today's post, I've drawn material not just from one paper, but from five! The subject matter is 'word2vec' – the work of Mikolov et al. at Google on efficient vector representations of words (and what you can do with them). The papers are:

- Efficient Estimation of Word Representations in Vector Space – Mikolov et al. 2013
- Distributed Representations of Words and Phrases and their Compositionality – Mikolov et al. 2013
- Linguistic Regularities in Continuous Space Word Representations – Mikolov et al. 2013
- word2vec Parameter Learning Explained – Rong 2014
- word2vec Explained: Deriving Mikolov et al's Negative Sampling Word-Embedding Method – Goldberg and Levy 2014

From the first of these papers ('Efficient estimation…') we get a description of the *Continuous Bag-of-Words* and *Continuous Skip-gram* models for learning word vectors (we'll talk about what a word vector is in a moment…). From the second paper we get more illustrations of the power of word vectors, some additional information on optimisations for the skip-gram model (hierarchical softmax and negative sampling), and a discussion of applying word vectors to phrases. The third paper ('Linguistic…



Andrej Karpathy blog — About — Hacker's guide to Neural Networks

The Unreasonable Effectiveness of Recurrent Neural Networks

May 21, 2015

There's something magical about Recurrent Neural Networks (RNNs). I still remember when I trained my first recurrent network for Image Captioning. Within a few dozen minutes of training my first baby model (with rather arbitrarily-chosen hyperparameters) started to generate very nice looking descriptions of images that were on the edge of making sense. Sometimes the ratio of how simple your model is to the quality of the results you get out of it blows past your expectations, and this was one of those times. What made this result so shocking at the time was that the common wisdom was that RNNs were supposed to be difficult to train (with more experience I've in fact reached the opposite conclusion). Fast forward about a year: I'm training RNNs all the time and I've witnessed their power and robustness many times, and yet their magical outputs still find ways of amusing me. This post is about sharing some of that magic with you.

> *We'll train RNNs to generate text character by character and ponder the question "how is that even possible?"*

By the way, together with this post I am also releasing code on Github that allows you to train character-level language models based on multi-layer LSTMs. You give it a large chunk of text and it will learn to generate text like it one character at a time. You can also use it to reproduce my experiments below. But we're getting ahead of ourselves; What are RNNs anyway?

Recurrent Neural Networks

Sequences. Depending on your background you might be wondering: *What makes Recurrent Networks so special?* A glaring limitation of Vanilla Neural Networks (and also Convolutional Networks) is that their API is too constrained: they accept a fixed-sized vector as input (e.g. an image) and produce a fixed-sized vector as output (e.g. probabilities of different classes). Not only that: These models perform this mapping using a fixed amount of computational steps (e.g. the number of layers in the model). The core reason that recurrent nets are more exciting is that they allow us to operate over *sequences* of vectors: Sequences in the input, the output, or in the most general case both. A few examples may make this more concrete:

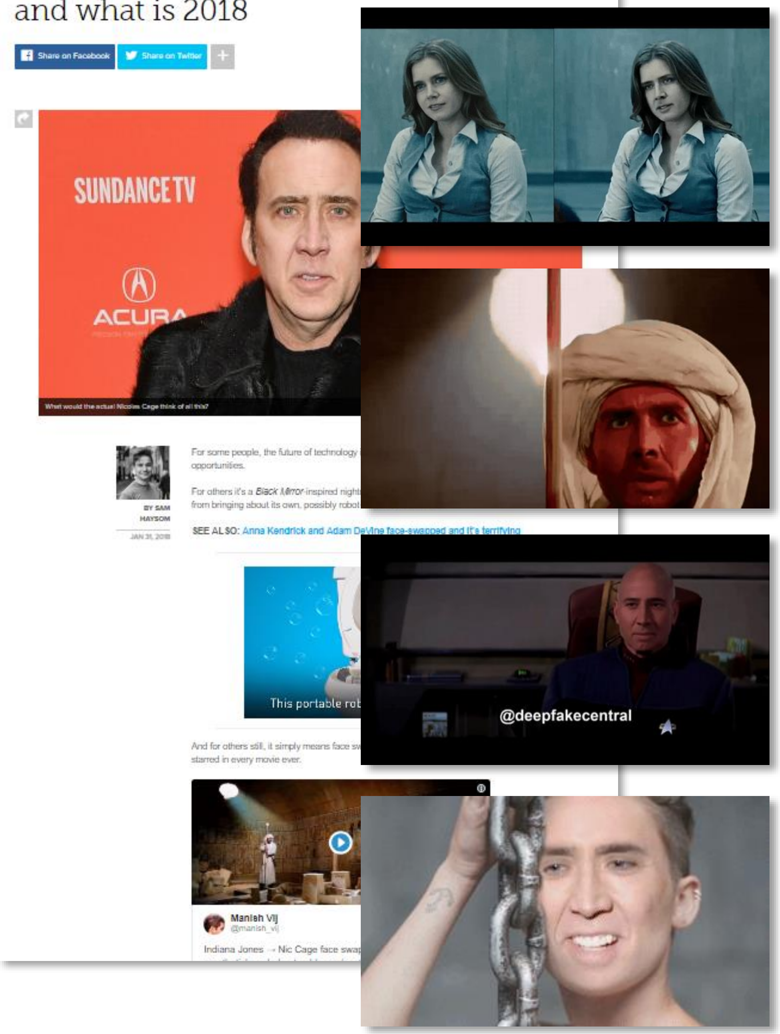
```
VIOLA:
Why, Salisbury must find his flesh and thought
That which I am not aps, not a man and in fire,
To show the reining of the raven and the wars
To grace my hand reproach within, and not a fair are hand,
That Caesar and my goodly father's world;
When I was heaven of presence and our fleets,
We spare with hours, but cut thy council I am great,
Murdered and by thy master's ready there
My power to give thee but so much as hell:
Some service in the noble bondman here,
Would show him to her wine.

KING LEAR:
O, if you were a feeble sight, the courtesy of your law,
Your sight and several breath, will wear the gods
With his heads, and my hands are wonder'd at the deeds,
So drop upon your lordship's head, and your opinion
Shall be against your honour.
```

the right, a recurrent network generates images of digits by learning to sequentially add color to a canvas (Gregor et al.):

People are using face-swapping tech to add Nicolas Cage to random movies and what is 2018

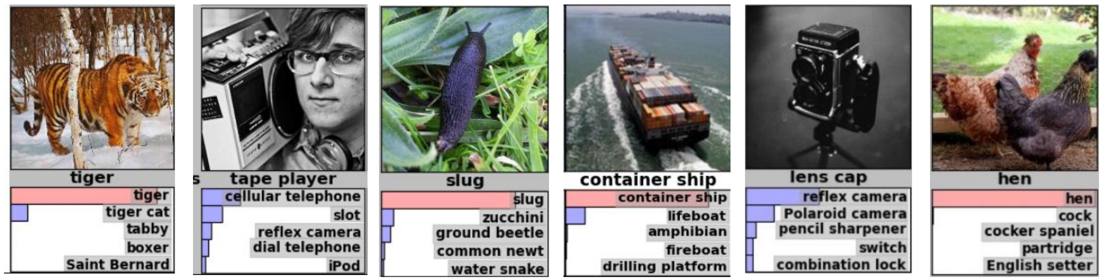
Share on Facebook — Share on Twitter



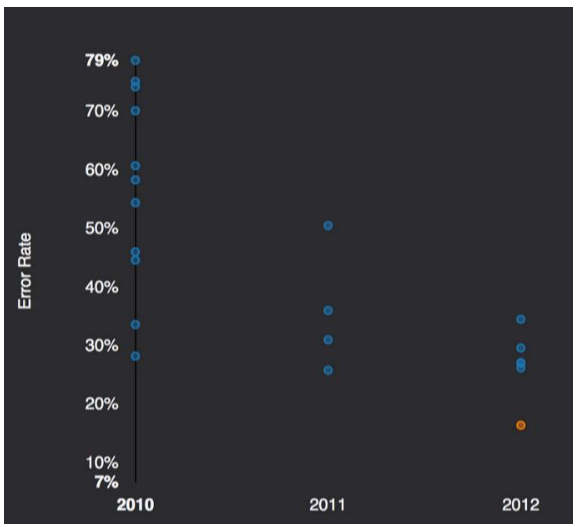
Was ist passiert?
Der ImageNet Wettbewerb

IMAGENET

1000 Kategorien
1 Mio. Beispiele



…

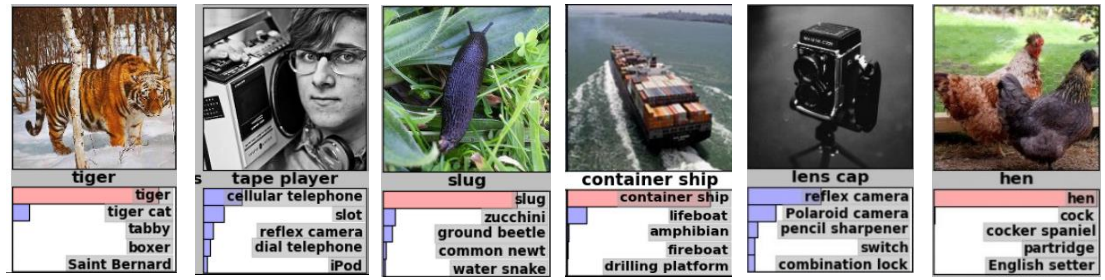


Zürcher Hochschule
für Angewandte Wissenschaften

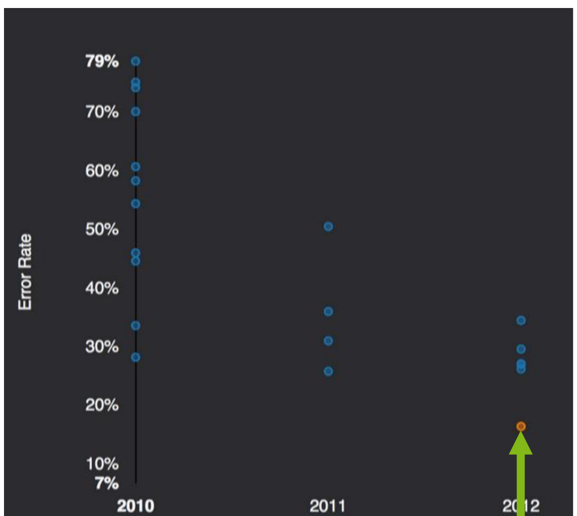
Was ist passiert?
Der ImageNet Wettbewerb

IMAGENET

1000 Kategorien
1 Mio. Beispiele



…

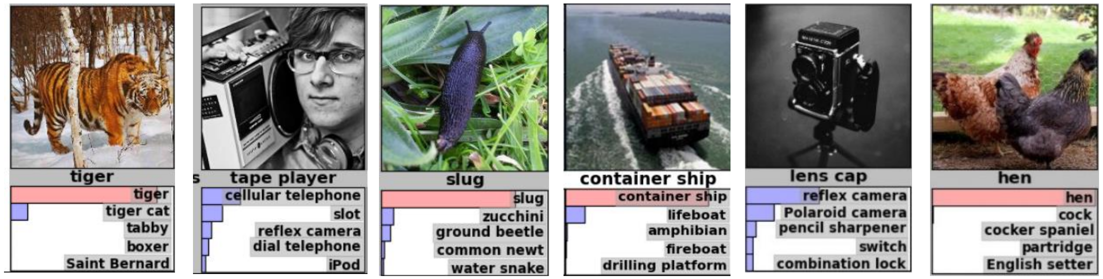


A. Krizhevsky verwendet als erster ein sog. «Deep Neural Network» (CNN)

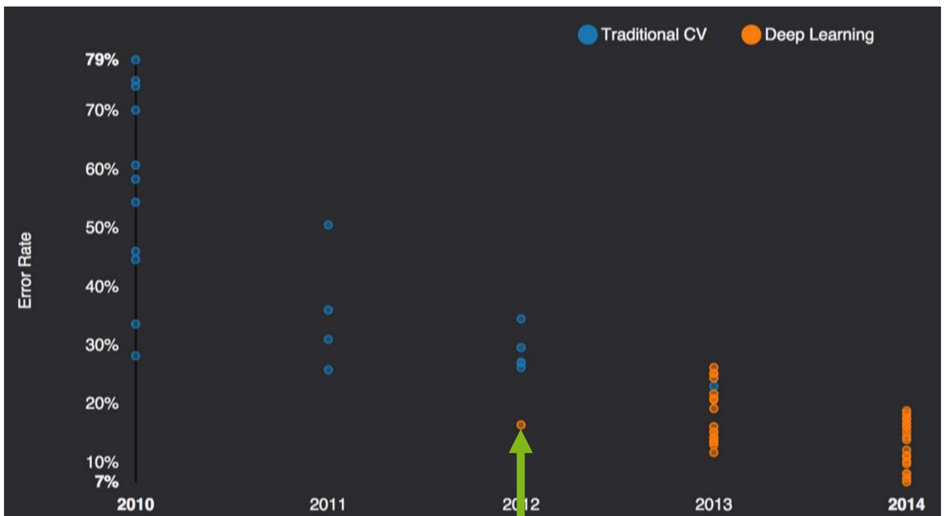
Was ist passiert?
Der ImageNet Wettbewerb

Zürcher Hochschule
für Angewandte Wissenschaften

1000 Kategorien
1 Mio. Beispiele



…

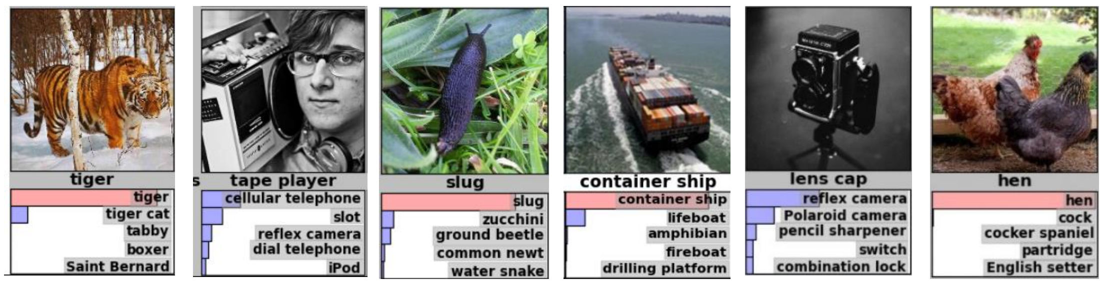


A. Krizhevsky verwendet als erster ein sog. «Deep Neural Network» (CNN)

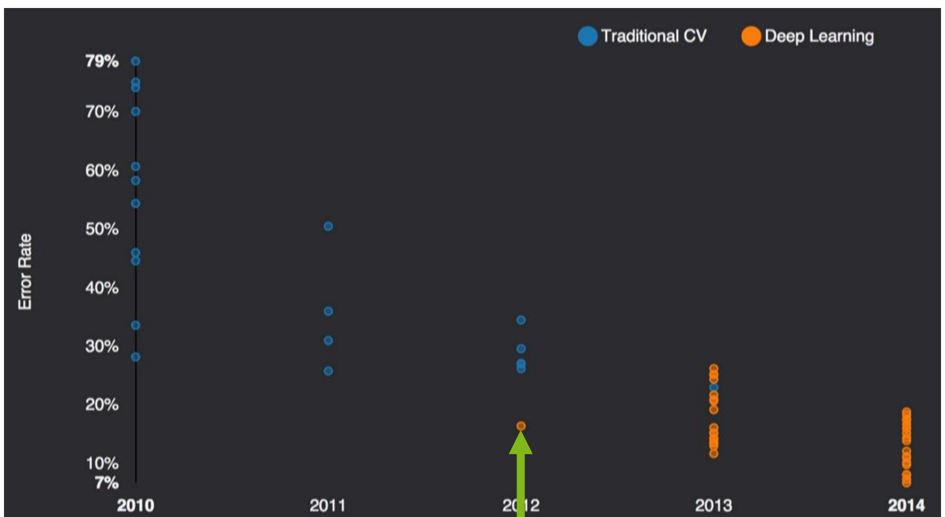
Was ist passiert?

Der ImageNet Wettbewerb

1000 Kategorien
1 Mio. Beispiele



…



A. Krizhevsky verwendet als erster ein sog. «Deep Neural Network» (CNN)

2015: Computer *haben "Sehen" gelernt*

4.95% Microsoft (06. Februar)
→ Besser als Menschen (5.10%)

4.80% Google (11. Februar)

4.58% Baidu (11. Mai)

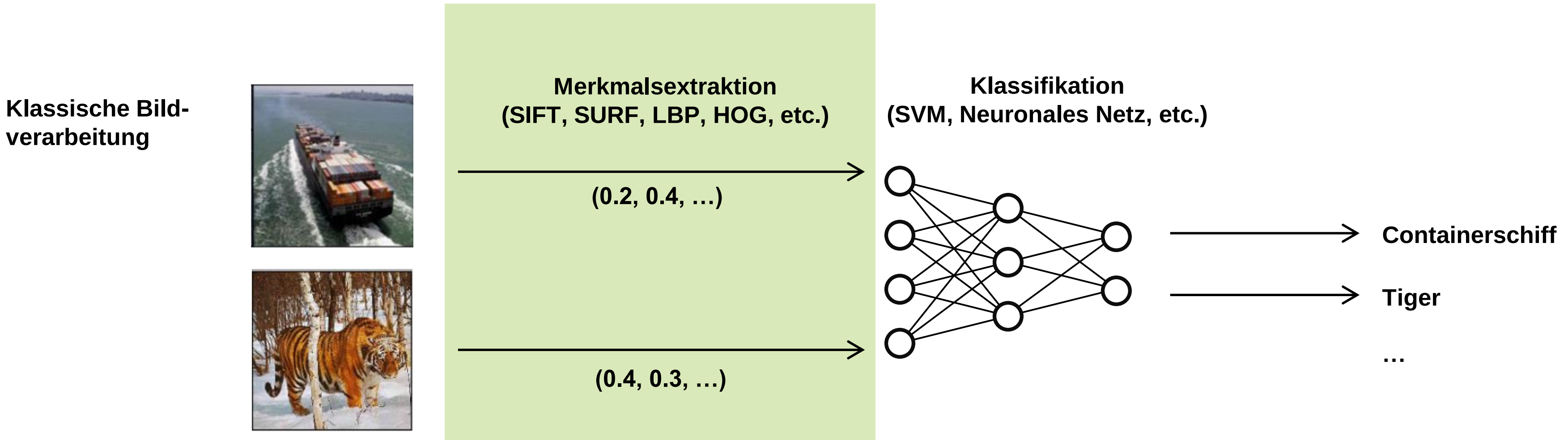
3.57% Microsoft (10. Dezember)

Was? → Wie? → Wow!

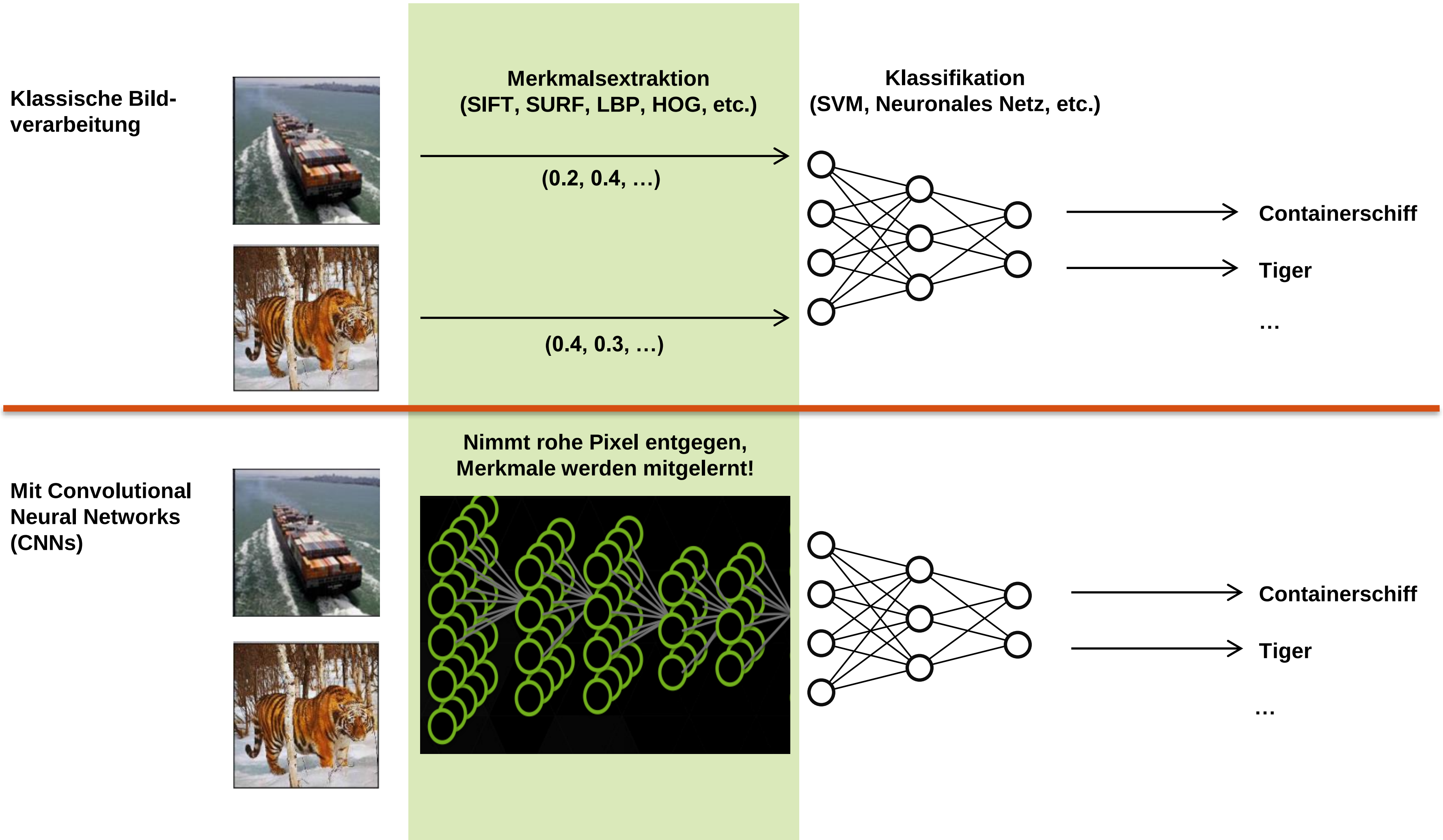
2

Wie geht das?

Idee: Mehr Tiefe zum Lernen von Merkmalen



Idee: Mehr Tiefe zum Lernen von Merkmalen

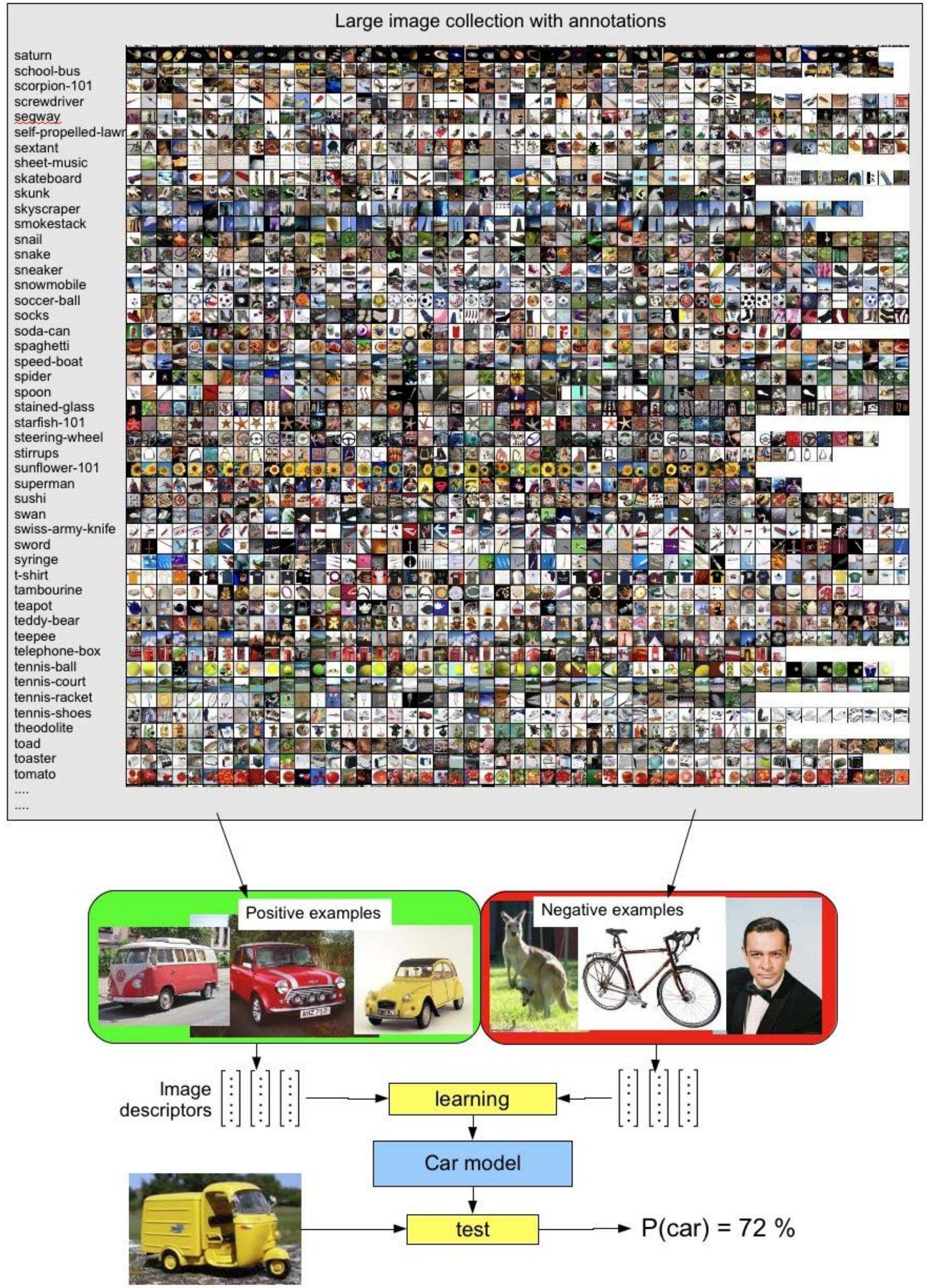


Grundlage

Induktives überwachtes Lernen

Annahme

- Ein an *genügend viele* Beispiele angepasstes Modell…
- …wird auch auf unbekannte Daten **generalisieren**



Quelle: http://lear.inrialpes.fr/job/postdoc-large-scale-classif-11-img/attribs_patchwork.jpg

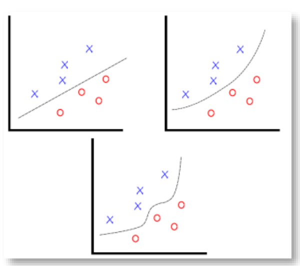
Grundlage
Induktives überwachtes Lernen

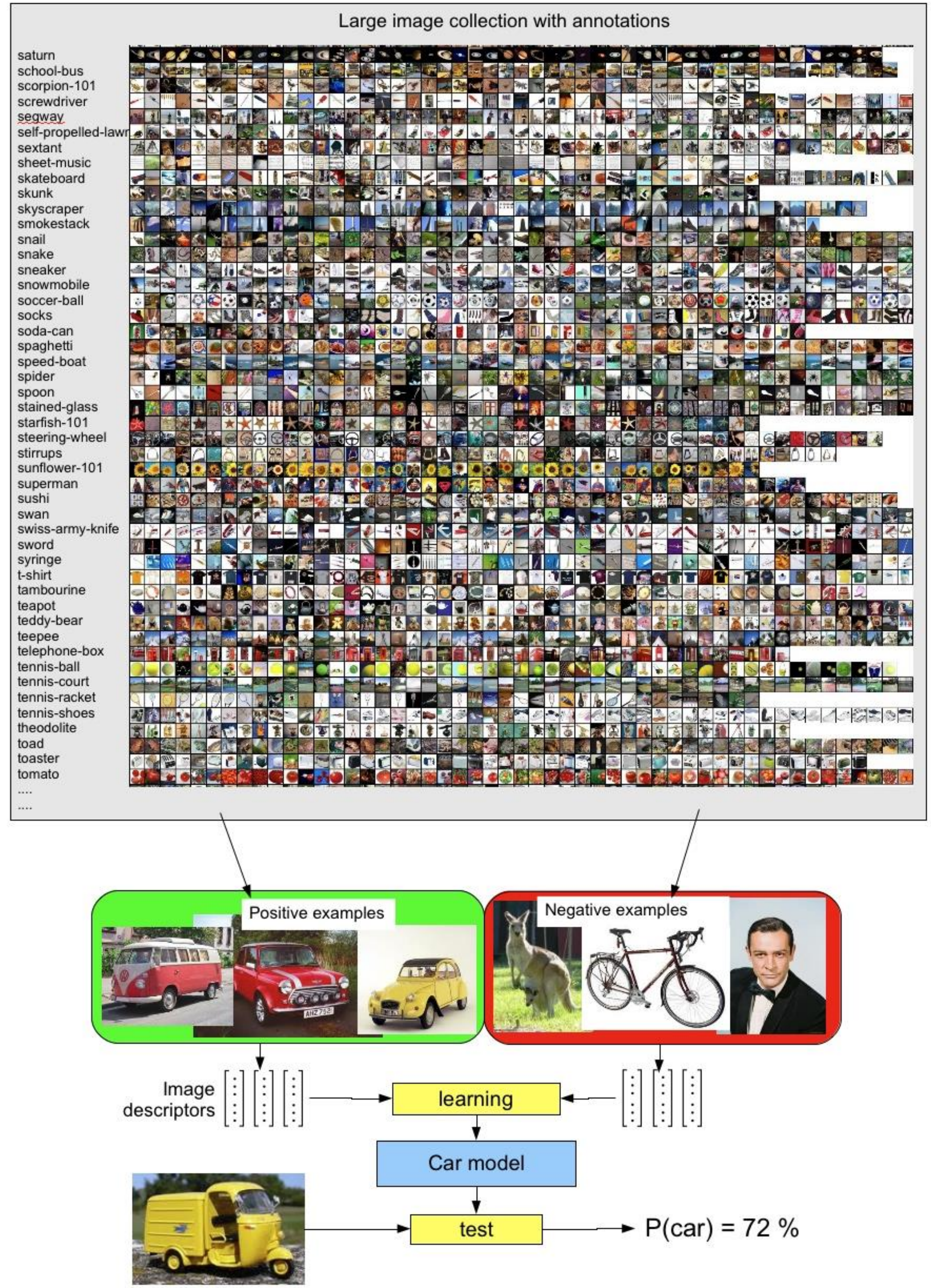
Annahme

- Ein an *genügend viele* Beispiele angepasstes Modell…
- …wird auch auf unbekannte Daten **generalisieren**

Methode

- **Suchen der Parameter einer gegebenen Funktion**…
- …so dass für alle Beispiele Eingabe (Bild) auf Ausgabe («Auto») abgebildet wird





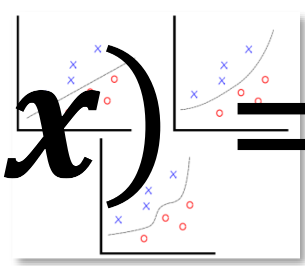
Quelle: http://lear.inrialpes.fr/job/postdoc-large-scale-classif-11-img/attribs_patchwork.jpg

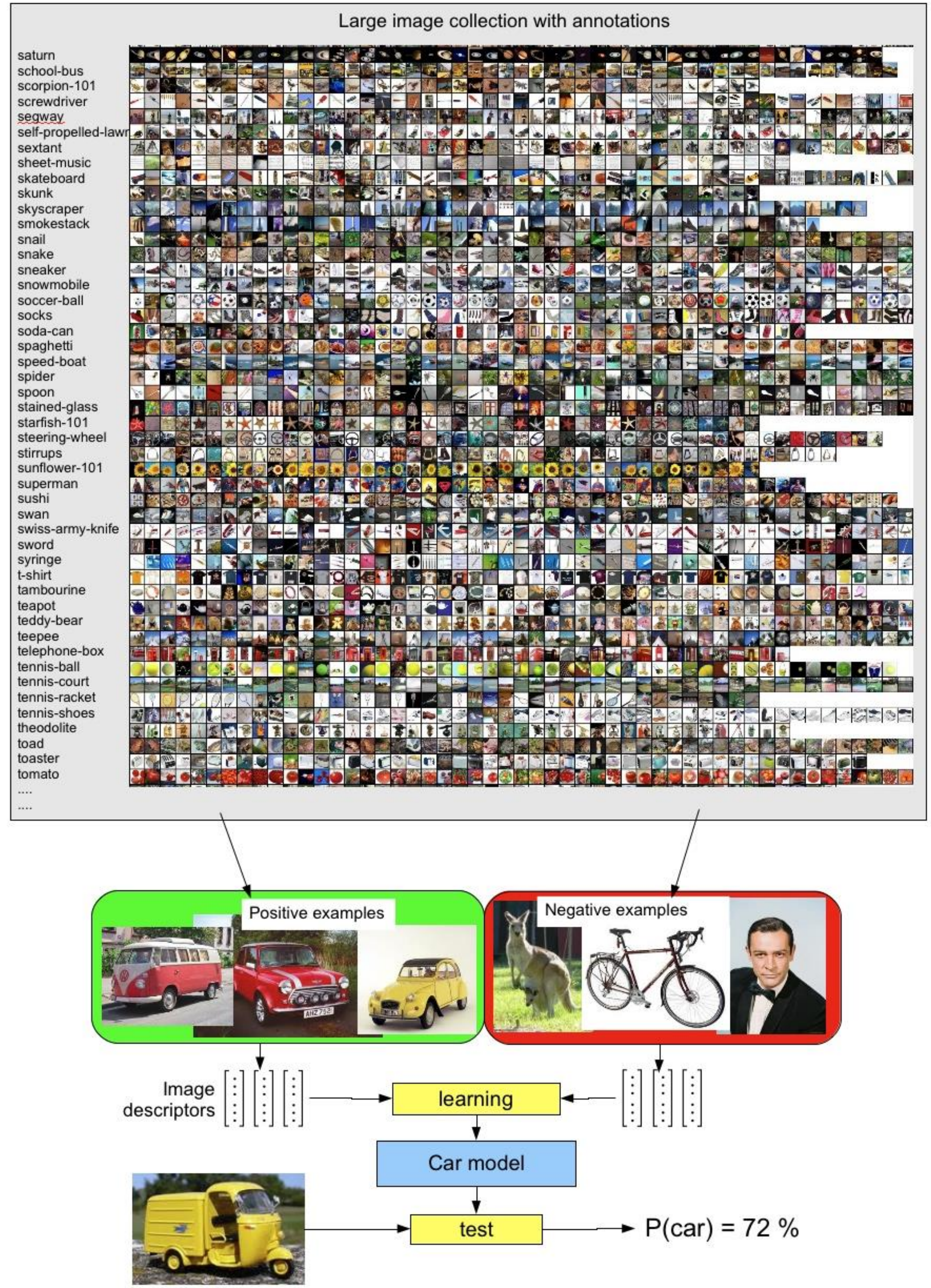
Grundlage
Induktives überwachtes Lernen

Annahme
- Ein an *genügend viele* Beispiele angepasstes Modell…
- …wird auch auf unbekannte Daten **generalisieren**

Methode
- **Suchen der Parameter einer gegebenen Funktion…**
- …so dass für alle Beispiele Eingabe (Bild) auf Ausgabe («Auto») abgebildet wird

$$f(x) = y$$

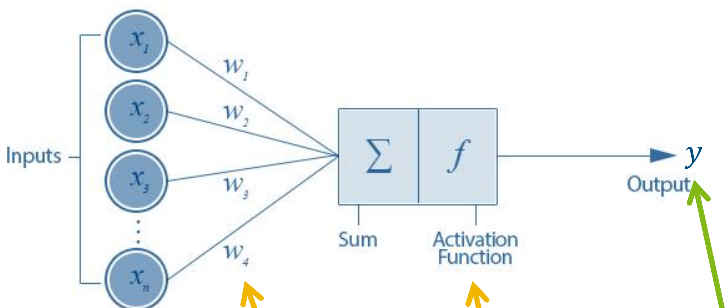




Quelle: http://lear.inrialpes.fr/job/postdoc-large-scale-classif-11-img/attribs_patchwork.jpg

Suche der Parameter *einer Funktion*?

Neuron



Merkmale (z.B. Pixel)

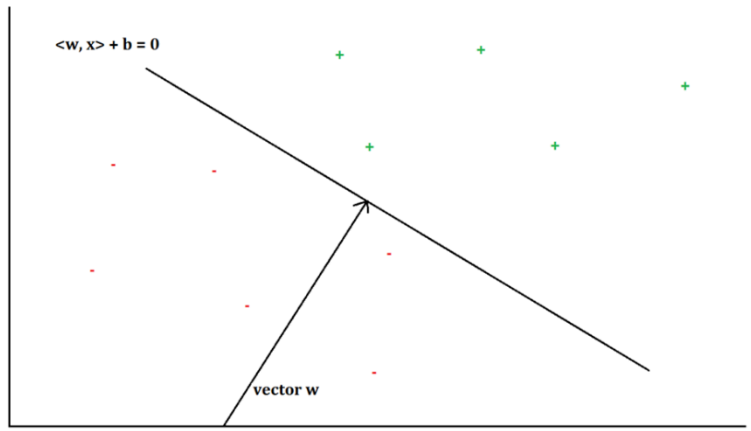
Anpassbare Parameter

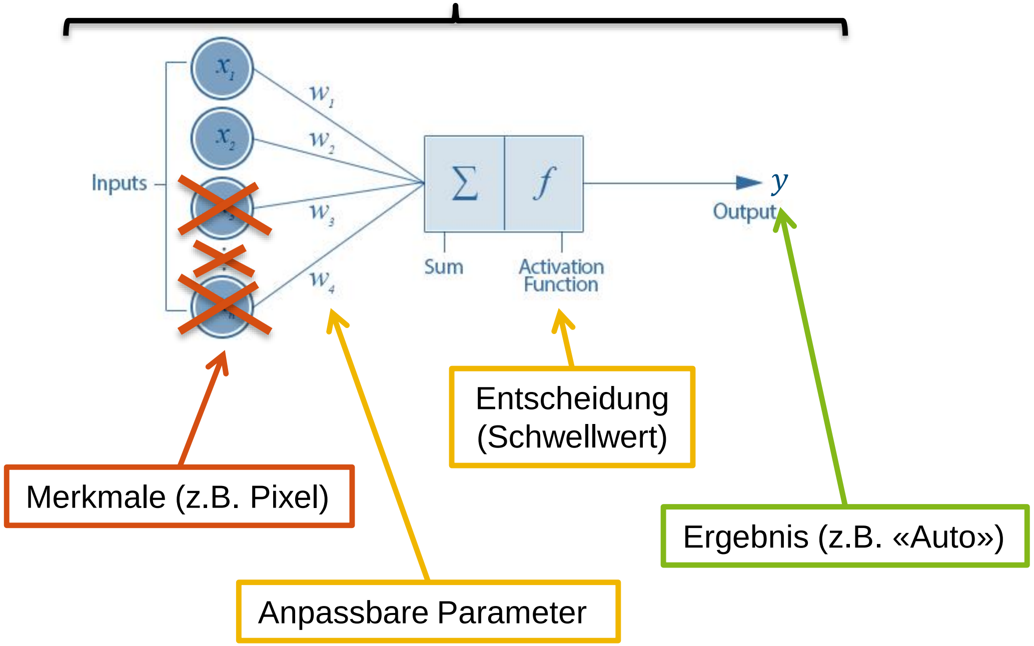
Entscheidung
(Schwellwert)

Ergebnis (z.B. «Auto»)

Suche der Parameter *einer Funktion*?

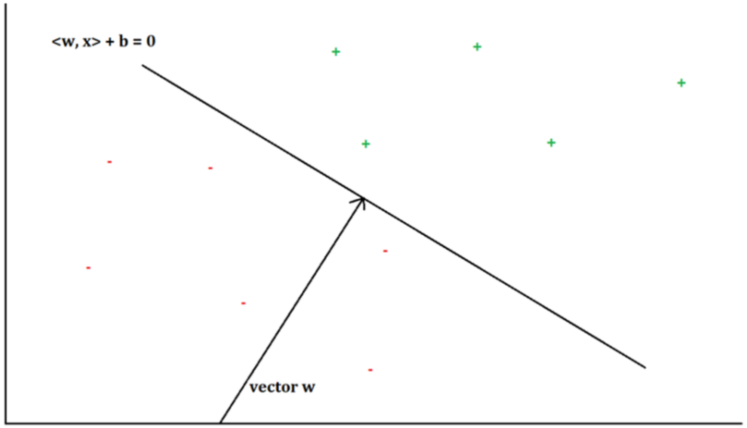
Neuron

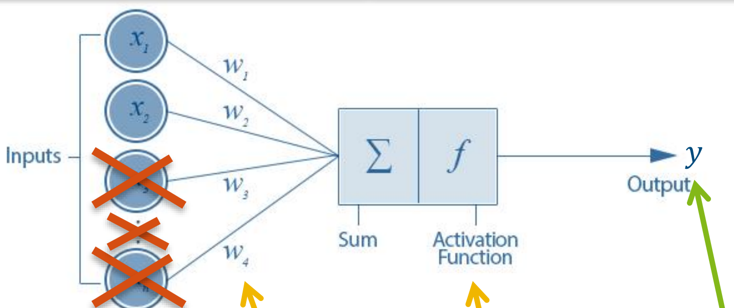




Suche der Parameter *einer Funktion*?

Neuron





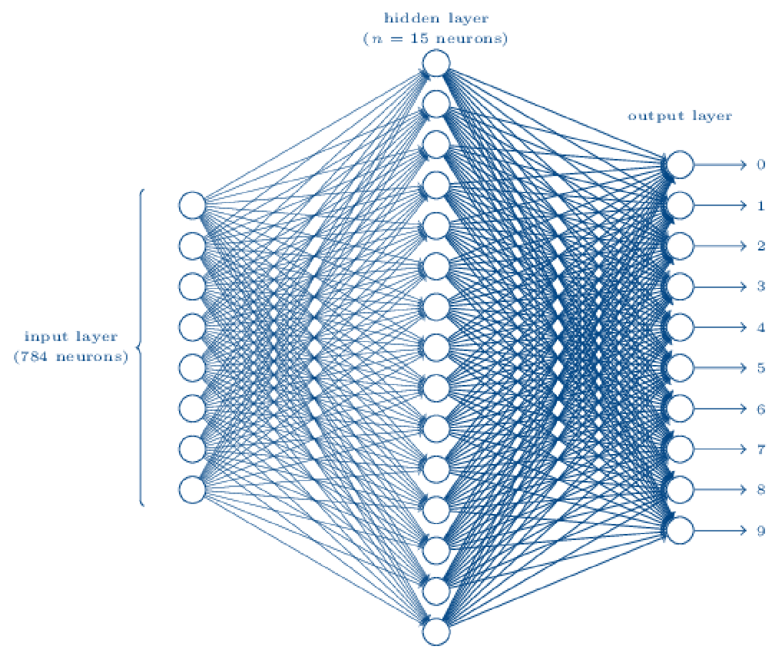
Merkmale (z.B. Pixel)

Anpassbare Parameter

Entscheidung (Schwellwert)

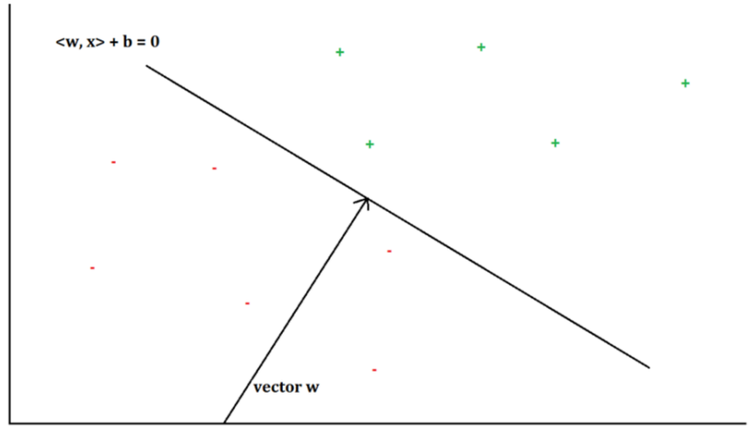
Ergebnis (z.B. «Auto»)

Neuronales Netz

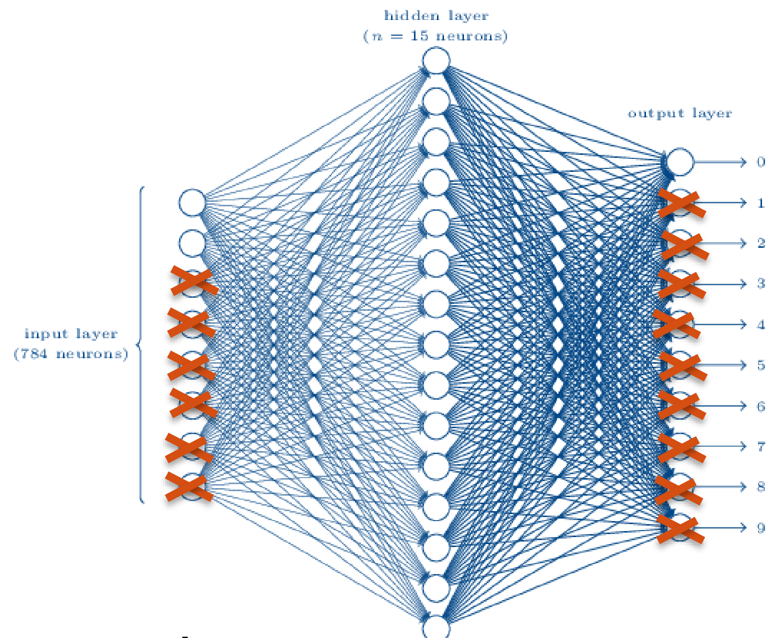


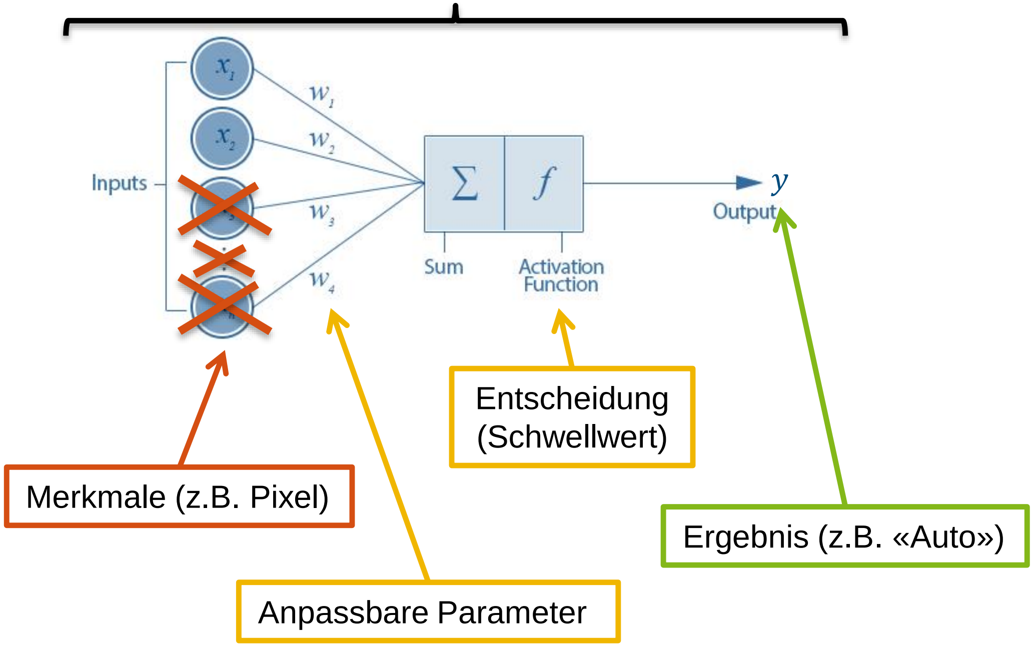
Suche der Parameter *einer Funktion*?

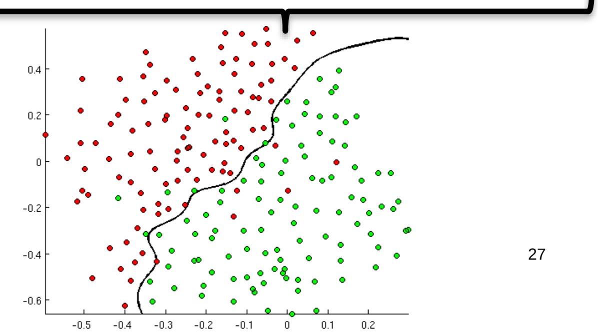
Neuron



Neuronales Netz







Zürcher Hochschule
für Angewandte Wissenschaften

3

Was machen wir damit?
(Wow, mit lokalen Unternehmen!)

Zürcher Hochschule
für Angewandte Wissenschaften

1. Print media monitoring

ARGUS DATA INSIGHTS
WISSEN ZUM ERFOLG

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

Task


International.

Spionage für den Erzfeind Iran

Israelischer Ex-Minister arbeitete als Agent für die Mullahs. Jetzt droht ihm lebenslänglich

Asylbewerber können bleiben

Nordkoreanischer Diktator zu Besuch in Peking

Challenge


«Steven hat sich alles selber beigebracht»

Unsere Nati ist gegen Serbien unter Druck

Verlieren verboten!

Transfer Ticker

Liverpool will Yann Sommer!

Nuisance


Das Tages-Horoskop

Liebling der Sterne

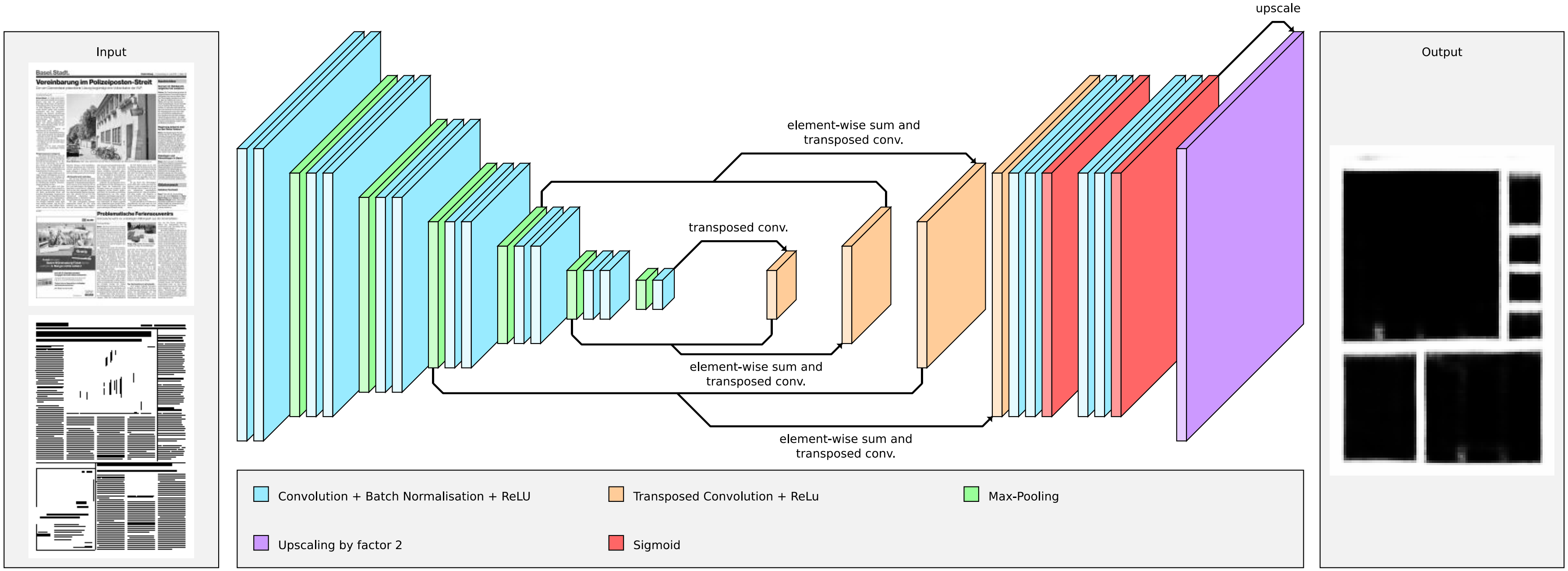
15,1 Millionen

Sind Sie der nächste Lotto-König?

Schweden-Rätsel

Sudoku

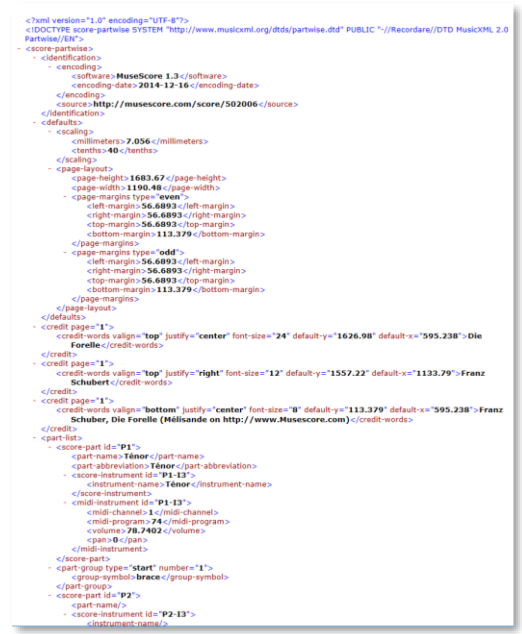
1. Print media monitoring – ML solution



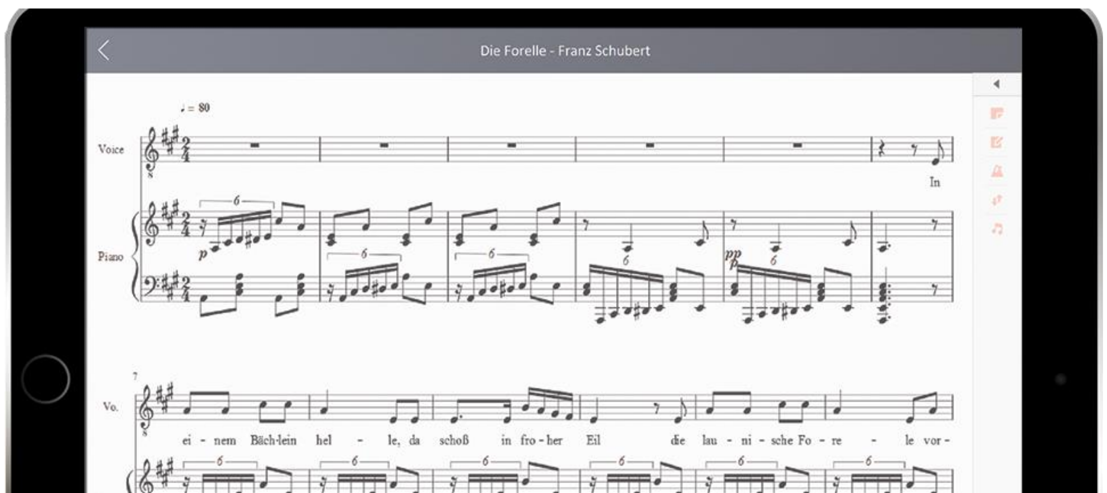
Meier, Stadelmann, Stampfli, Arnold & Cieliebak (2017). «*Fully Convolutional Neural Networks for Newspaper Article Segmentation*». ICDAR'2017.
Stadelmann, Tolkachev, Sick, Stampfli & Dürr (2018). «*Beyond ImageNet - Deep Learning in Industrial Practice*». In: Braschler et al., «Applied Data Science», Springer.

Zürcher Hochschule
für Angewandte Wissenschaften

2. Music scanning



```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE score-partwise SYSTEM "http://www.musicxml.org/dtds/partwise.dtd" PUBLIC "-//Recordare//DTD MusicXML 2.0 Partwise//EN">
- <score-partwise>
  - <identification>
    - <encoding>
        <software>MuseScore 1.3</software>
        <encoding-date>2014-12-16</encoding-date>
      </encoding>
      <source>http://musescore.com/score/502006</source>
    </identification>
  - <defaults>
    - <scaling>
        <millimeters>7.056</millimeters>
        <tenths>40</tenths>
      </scaling>
    - <page-layout>
        <page-height>1683.67</page-height>
        <page-width>1190.48</page-width>
      - <page-margins type="even">
          <left-margin>56.6893</left-margin>
          <right-margin>56.6893</right-margin>
          <top-margin>56.6893</top-margin>
          <bottom-margin>113.379</bottom-margin>
        </page-margins>
      - <page-margins type="odd">
          <left-margin>56.6893</left-margin>
          <right-margin>56.6893</right-margin>
          <top-margin>56.6893</top-margin>
          <bottom-margin>113.379</bottom-margin>
        </page-margins>
      </page-layout>
    </defaults>
  - <credit page="1">
      <credit-words valign="top" justify="center" font-size="24" default-y="1626.98" default-x="595.238">Die Forelle</credit-words>
    </credit>
  - <credit page="1">
      <credit-words valign="top" justify="right" font-size="12" default-y="1557.22" default-x="1133.79">Franz Schubert</credit-words>
    </credit>
  - <credit page="1">
      <credit-words valign="bottom" justify="center" font-size="8" default-y="113.379" default-x="595.238">Franz Schuber, Die Forelle (Mélisande on http://www.Musescore.com)</credit-words>
    </credit>
  - <part-list>
    - <score-part id="P1">
        <part-name>Ténor</part-name>
        <part-abbreviation>Ténor</part-abbreviation>
      - <score-instrument id="P1-I3">
          <instrument-name>Ténor</instrument-name>
        </score-instrument>
      - <midi-instrument id="P1-I3">
          <midi-channel>1</midi-channel>
          <midi-program>74</midi-program>
          <volume>78.7402</volume>
          <pan>0</pan>
        </midi-instrument>
      </score-part>
    - <part-group type="start" number="1">
        <group-symbol>brace</group-symbol>
      </part-group>
    - <score-part id="P2">
        <part-name/>
      - <score-instrument id="P2-I3">
          <instrument-name/>
```

SCOREPAD

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

2. Music scanning – challenges & solutions

Tuggener, Elezi, Schmidhuber, Pelillo & Stadelmann (2018). *«DeepScores – A Dataset for Segmentation, Detection and Classification of Tiny Objects»*. ICPR'2018.

Zürcher Hochschule
für Angewandte Wissenschaften

2. Music scanning – challenges & solutions

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

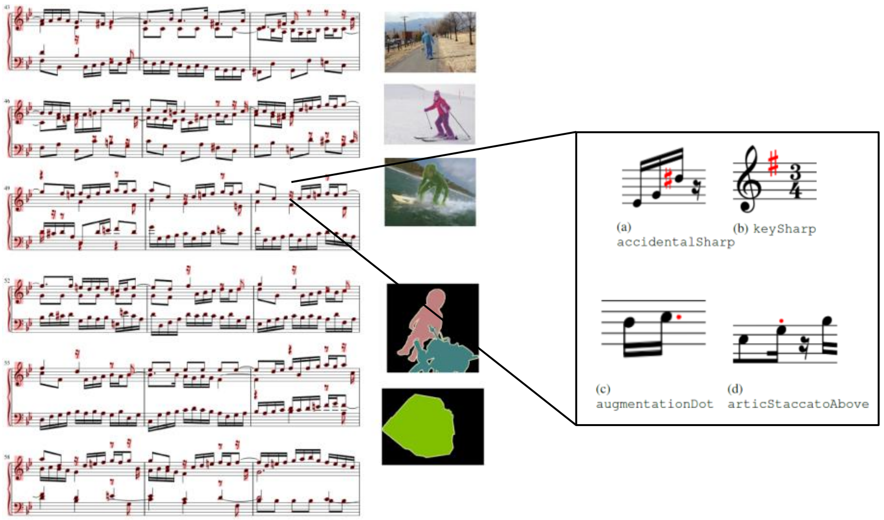
Swiss Confederation

Innosuisse – Swiss Innovation Agency

Tuggener, Elezi, Schmidhuber, Pelillo & Stadelmann (2018). *«DeepScores – A Dataset for Segmentation, Detection and Classification of Tiny Objects»*. ICPR'2018.

Zürcher Hochschule
für Angewandte Wissenschaften

2. Music scanning – challenges & solutions

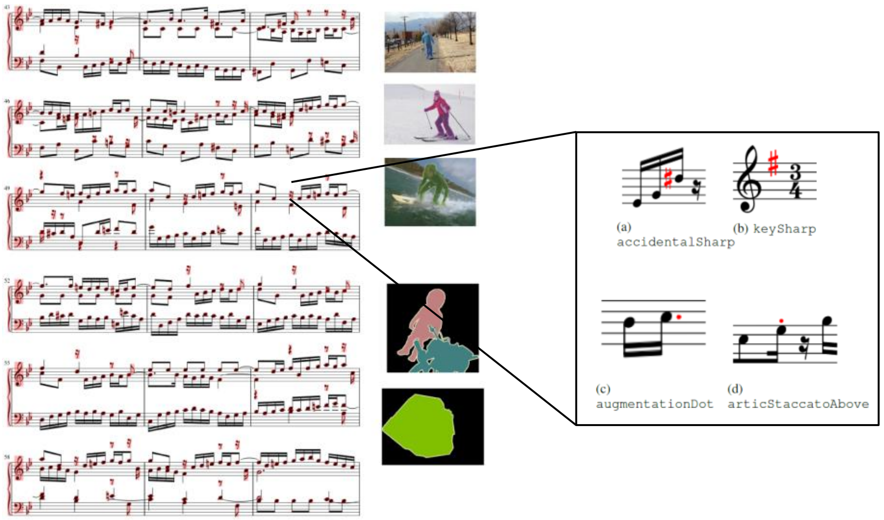


Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Innosuisse – Swiss Innovation Agency

Tuggener, Elezi, Schmidhuber, Pelillo & Stadelmann (2018). *«DeepScores – A Dataset for Segmentation, Detection and Classification of Tiny Objects»*. ICPR'2018.

Zürcher Hochschule
für Angewandte Wissenschaften

2. Music scanning – challenges & solutions



,

SCOREPAD

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

Tuggener, Elezi, Schmidhuber, Pelillo & Stadelmann (2018). *«DeepScores – A Dataset for Segmentation, Detection and Classification of Tiny Objects»*. ICPR'2018.

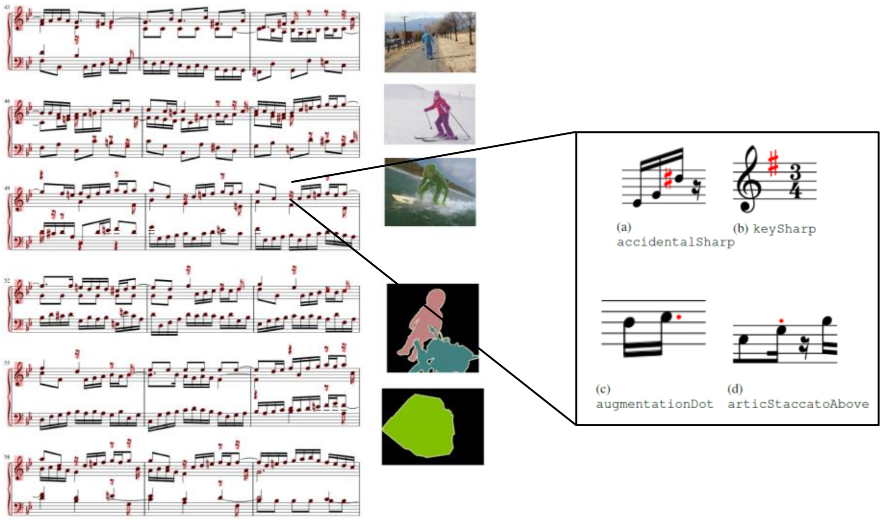
2. Music scanning – challenges & solutions

SCOREPAD

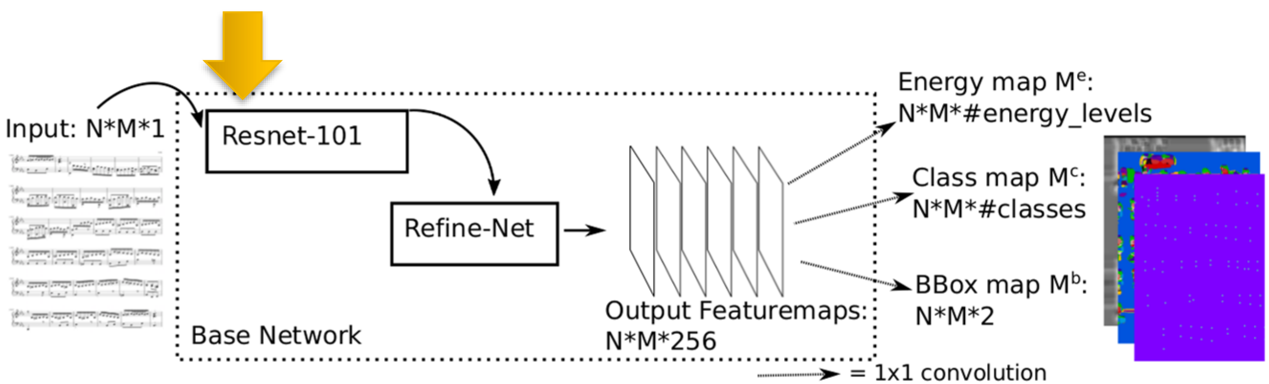
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

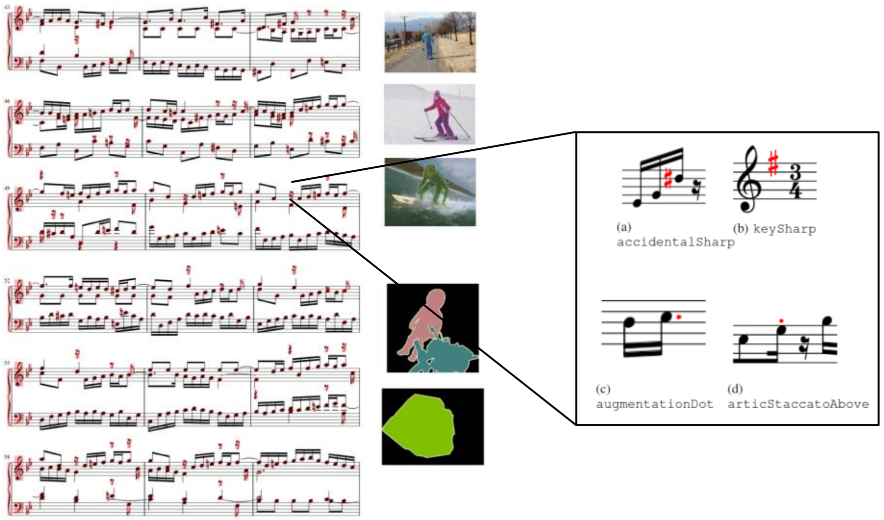


,



Tuggener, Elezi, Schmidhuber, Pelillo & Stadelmann (2018). *«DeepScores – A Dataset for Segmentation, Detection and Classification of Tiny Objects»*. ICPR'2018.
Tuggener, Elezi, Schmidhuber & Stadelmann (2018). *«Deep Watershed Detector for Music Object Recognition»*. ISMIR'2018.

2. Music scanning – challenges & solutions



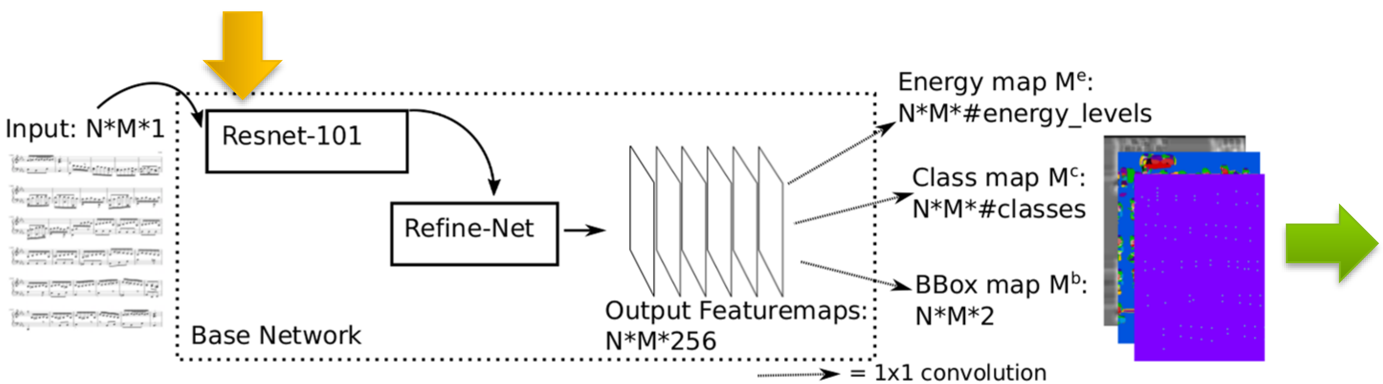
,

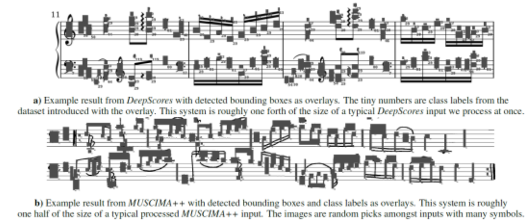
SCOREPAD

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

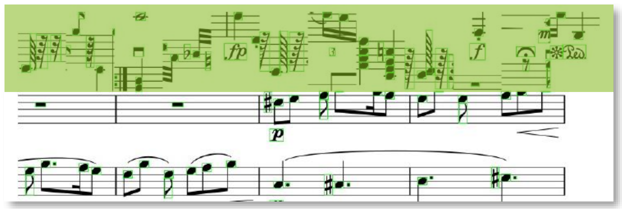




Tuggener, Elezi, Schmidhuber, Pelillo & Stadelmann (2018). *«DeepScores – A Dataset for Segmentation, Detection and Classification of Tiny Objects»*. ICPR'2018.
Tuggener, Elezi, Schmidhuber & Stadelmann (2018). *«Deep Watershed Detector for Music Object Recognition»*. ISMIR'2018.

2. Music scanning – industrialization
(Work in progress)

Recent results on class imbalance and robustness challenges
1. Added sophisticated **data augmentation** in every page's margins



2. Put additional effort (and compute) into hyperparameter **tuning** and **longer training**
3. Trained also on scanned (more **real-worldish**) scores

➔ **Improved** our **mAP** from 16% (on purely synthetic data) **to 73%** on more challenging real-world data set (additionally, using Pacha et al.'s evaluation method as a 2nd benchmark: from 24.8% to 47.5%)

Elezi, Tuggener, Pelillo & Stadelmann (2018). *«DeepScores and Deep Watershed Detection: current state and open issues»*. WoRMS @ ISMIR'2018.
Pacha, Hajic, Calvo-Zaragoza (2018). *«A Baseline for General Music Object Detection with Deep Learning»*. Appl. Sci. 2018, 8, 1488, MDPI.

Schlussfolgerungen

- *KI löst komplexe (einzelne) Probleme*; es geht nicht um «Intelligenz» in unserem Sinne
- Deep Learning hat zu Paradigmenwechsel in *Mustererkennungsaufgaben* geführt
- Deren Anwendung (in Unternehmen & Produkten) führt zu grossem Veränderungspotential in der Gesellschaft – ganz *ohne Science Fiction*
- Die Veränderung wird kommen – *gestalten wir* sie!

Swiss Alliance for
Data-Intensive Services

www.zhaw.ch/datalab

Zu mir:
- Leiter ZHAW Datalab, Board Data+Service
- thilo.stadelmann@zhaw.ch
- 058 934 72 08
- https://stdm.github.io/

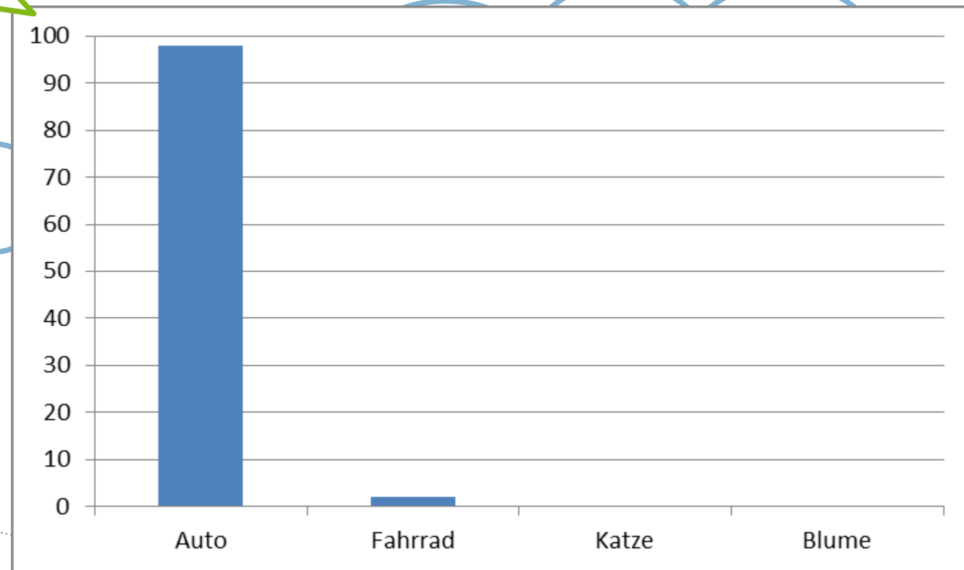
Mehr zum Thema:
- KI: https://sgaico.swissinformatics.org/
- Data+Service Alliance: www.data-service-alliance.ch
- Gemeinsame Projekte: datalab@zhaw.ch

➔ Fragen Sie gerne nach.

ANHANG

Suche der Parameter einer Funktion?

Wahrscheinlichkeit [%] für bestimmtes Ergebnis



- Unser Neuronales Netz: $f_W(x) = y$
  mit Bild x, echtem Resultat y und Parametern $\boldsymbol{W}$
  ($\boldsymbol{W} = \{w_1, w_2, \ldots\}$ anfangs zufällig gewählt)

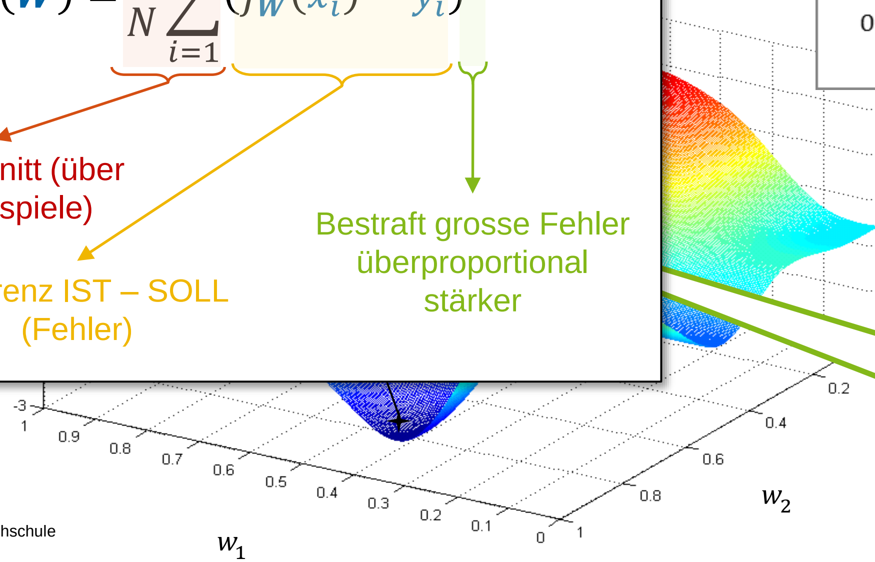
- Fehlermass: $l(\boldsymbol{W}) = \frac{1}{N}\sum_{i=1}^{N}(f_W(x_i) - y_i)^2$
  Durchschnitt der quadratischen Abweichungen über alle Bilder (Loss)

$$l(\boldsymbol{W}) = \underbrace{\frac{1}{N}\sum_{i=1}^{N}}_{\text{Durchschnitt (über alle Beispiele)}}\underbrace{(f_W(x_i) - y_i)}_{\text{Differenz IST – SOLL (Fehler)}}\underbrace{{}^2}_{\text{Bestraft grosse Fehler überproportional stärker}}$$

Durchschnitt (über alle Beispiele)

Differenz IST – SOLL (Fehler)

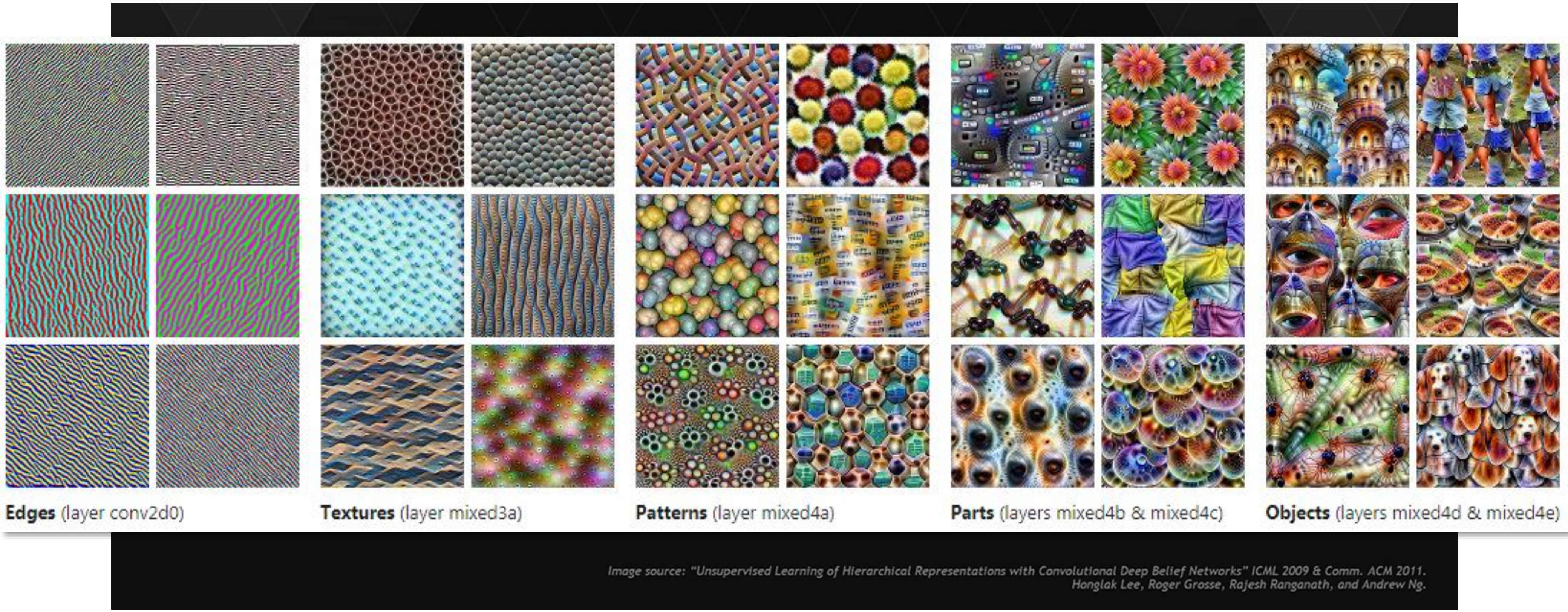
Bestraft grosse Fehler überproportional stärker



← Fehlerlandschaft

Methode: Anpassung der Gewichte von f in Richtung der steilsten Steigung (abwärts) von J

Was «sieht» das Neuronale Netz?
Hierarchien komplexer werdender Merkmale

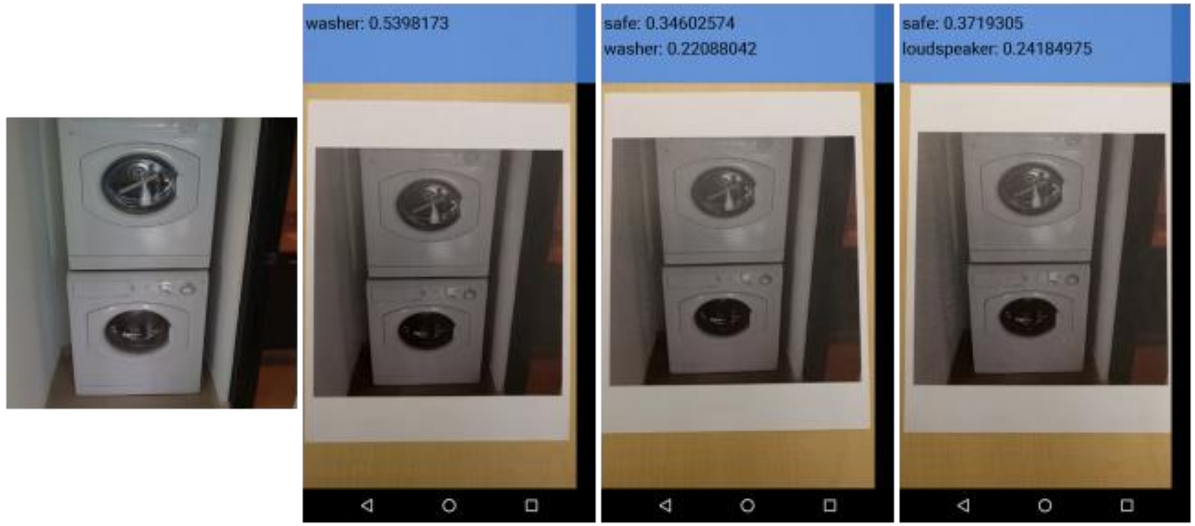


Quellen: https://www.pinterest.com/explore/artificial-neural-network/
Olah, et al., "Feature Visualization", Distill, 2017, https://distill.pub/2017/feature-visualization/.

Wie schlussfolgert die Maschine?
«Debugging» für Einblicke in die vermeintliche «Black Box»

Verdeutlichen ein Problem:

- Adversarial Examples

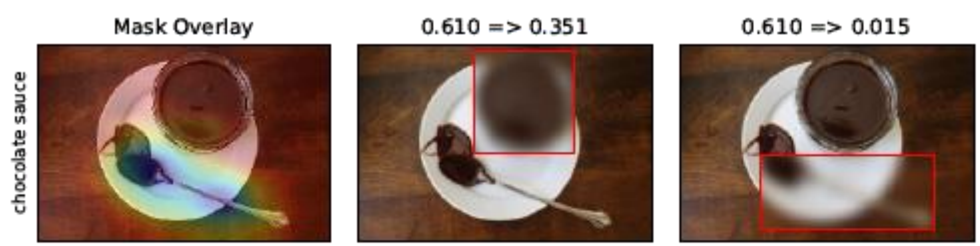


(a) Image from dataset (b) Clean image (c) Adv. image, $\epsilon = 4$ (d) Adv. image, $\epsilon = 8$

https://blog.openai.com/adversarial-example-research/

Bieten eine Lösung:

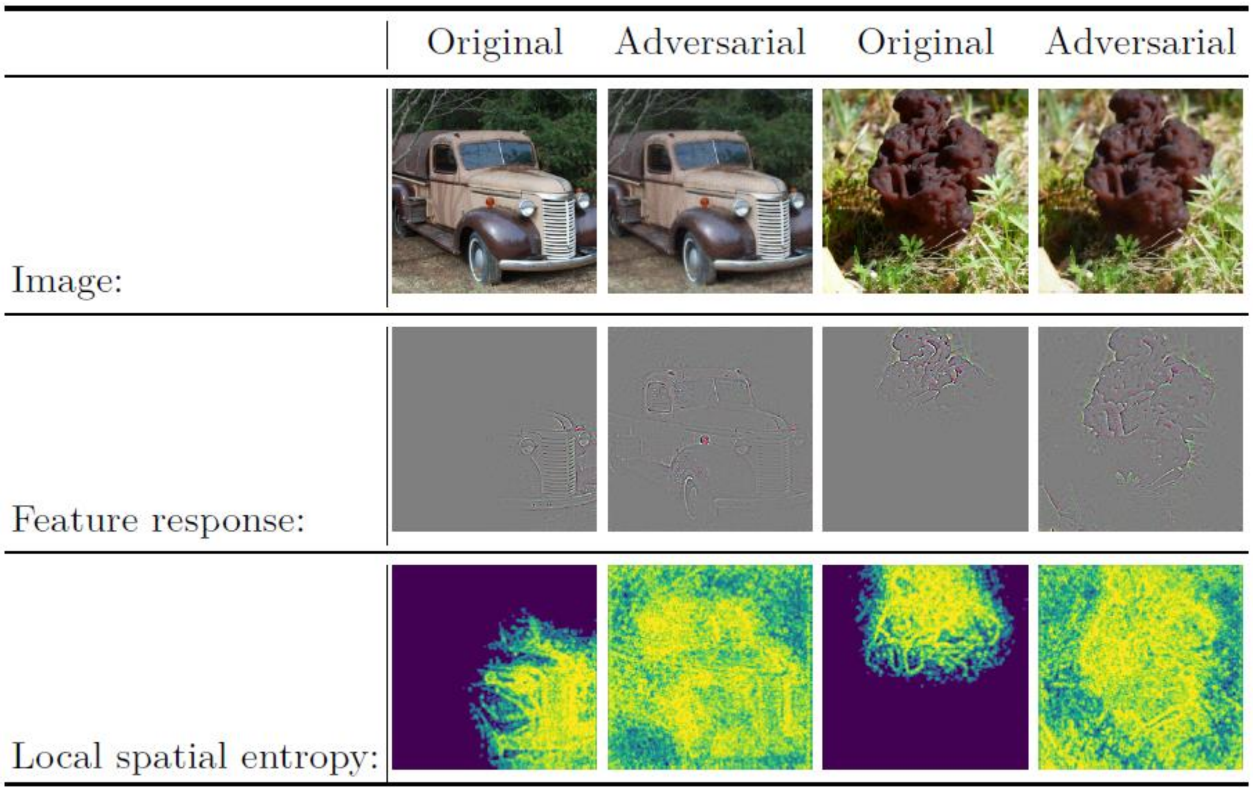
- Saliency Maps



Ruth C. Fong & Andrea Vedaldi, «Interpretable Explanations of Black Boxes by Meaningful Perturbation», 2017

Adversarial attacks erkennen

…mittels Local Spatial Entropy der Feature Responses



Amirian, Schwenker & Stadelmann (2018). *«Trace and Detect Adversarial Attacks on CNNs using Feature Response Maps»*. ANNPR’2018.

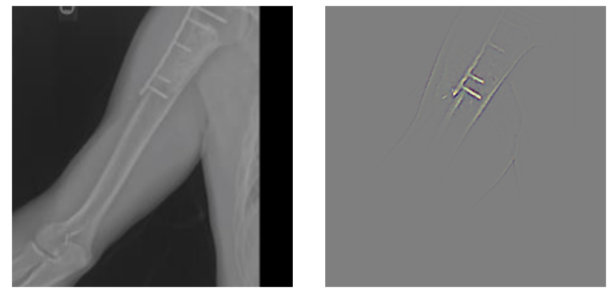
Lessons learned – model interpretability

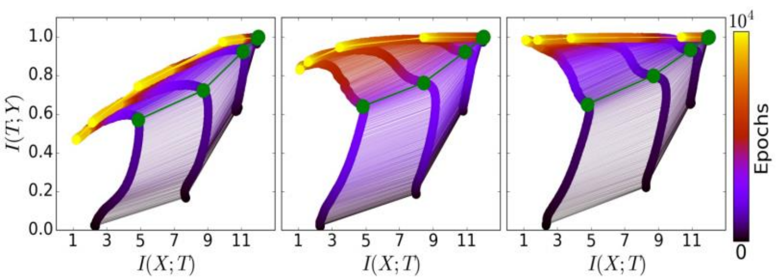
Interpretability is required.

- Helps the developer in «debugging», needed by the user to trust
  → visualizations of learned features, training process, learning curves etc. should be «always on»

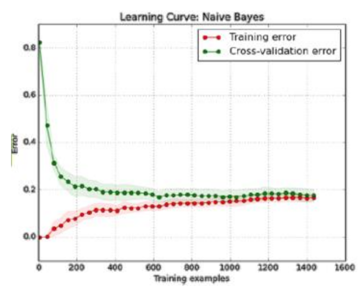
negative X-ray

positive X-ray

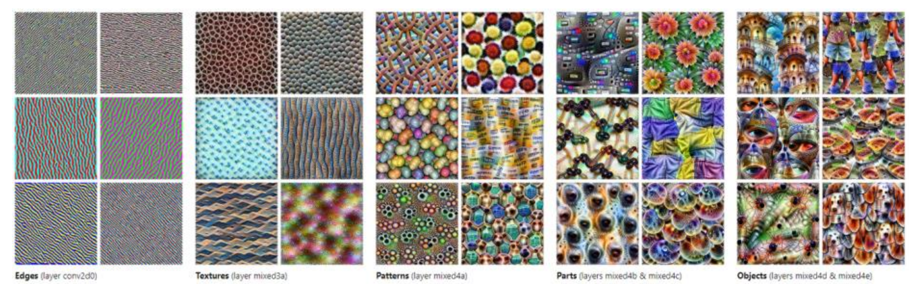




DNN training on the Information Plane



a learning curve



feature visualization

Stadelmann, Amirian, Arabaci, Arnold, Duivesteijn, Elezi, Geiger, Lörwald, Meier, Rombach & Tuggener (2018). *«Deep Learning in the Wild»*. ANNPR'2018.
Schwartz-Ziv & Tishby (2017). *«Opening the Black Box of Deep Neural Networks via Information»*.
https://distill.pub/2017/feature-visualization/, https://stanfordmlgroup.github.io/competitions/mura/

2. OMR deep dive
OMR vs state of the art object detectors

Zürcher Hochschule für Angewandte Wissenschaften

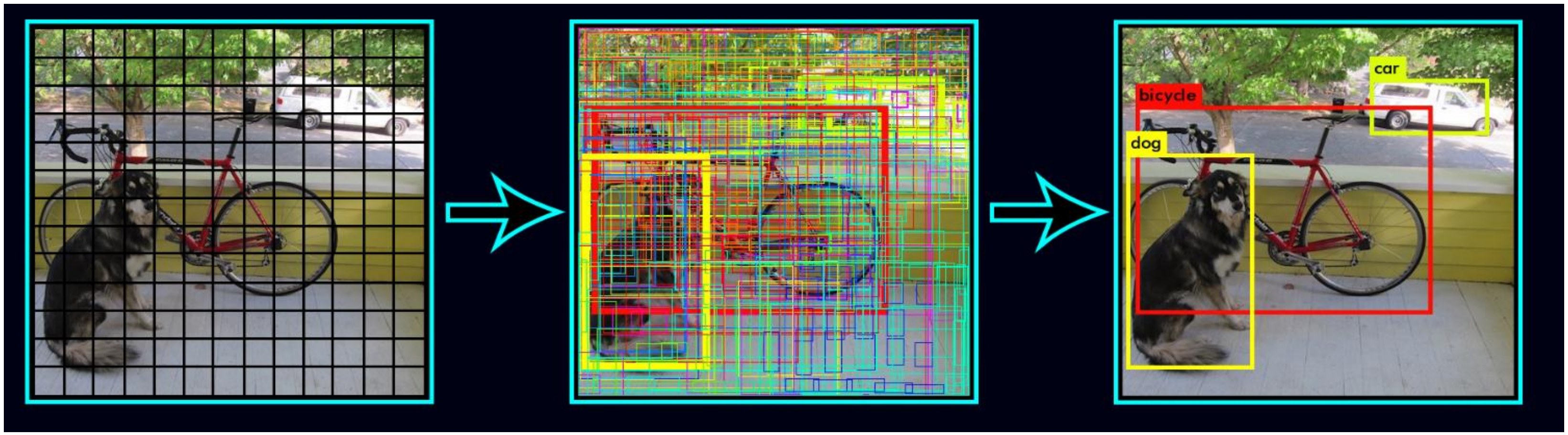
SCOREPAD

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

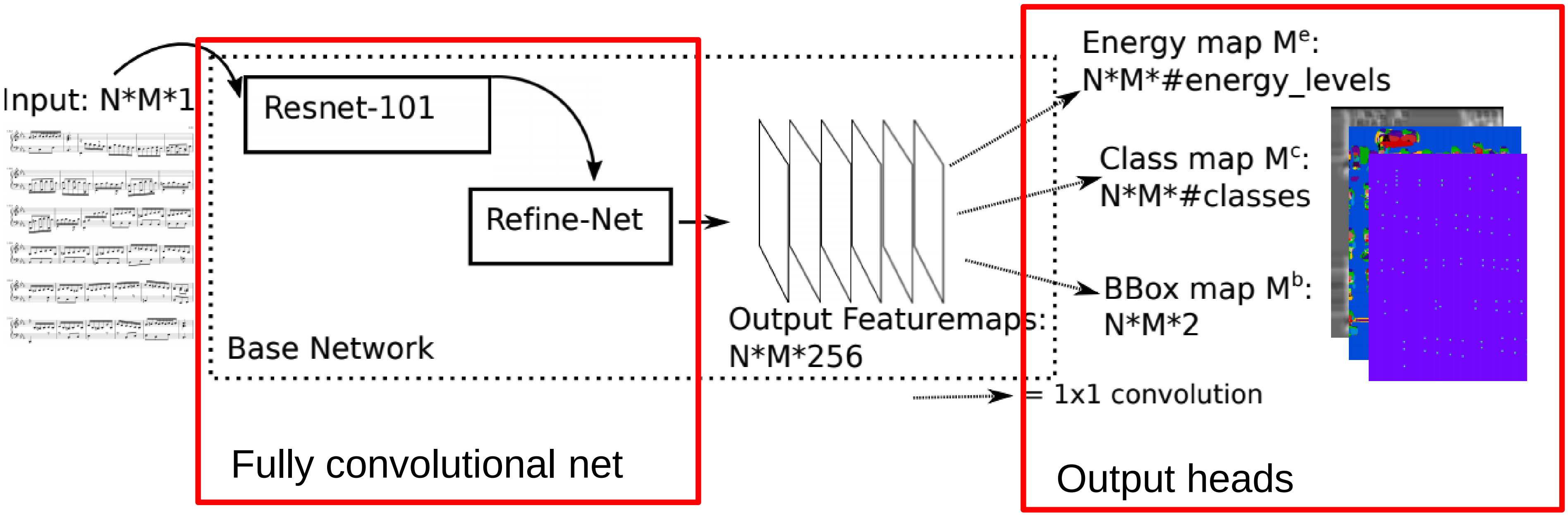
YOLO/SSD-type detectors



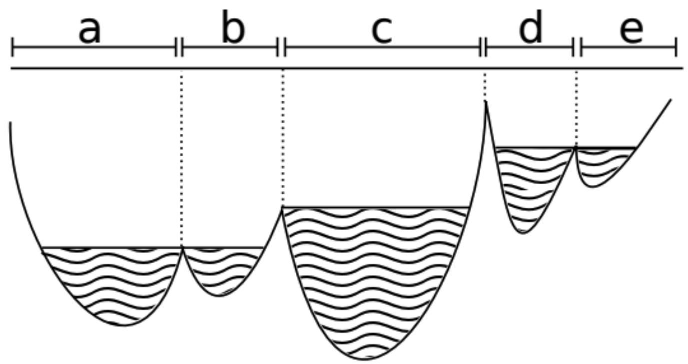
Source: https://pjreddie.com/darknet/yolov2/ (11.09.2018)

R-CNN
- Two-step proposal and refinement scheme
- Very large amount of proposals at high resolution needed

2. OMR deep dive (contd.)
The deep watershed detector



2. OMR deep dive (contd.)
The (deep) watershed transform



Zürcher Hochschule
für Angewandte Wissenschaften

2. OMR deep dive (contd.)

Output heads of the deep watershed detector