Zürcher Hochschule
für Angewandte Wissenschaften

Was Sie von KI erwarten können

Studerus Technology Forum, Regensdorf, 22. November 2018

Thilo Stadelmann

Swiss Alliance for
Data-Intensive Services

www.zhaw.ch/datalab

Prolog

Zürcher Hochschule
für Angewandte Wissenschaften

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

"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)

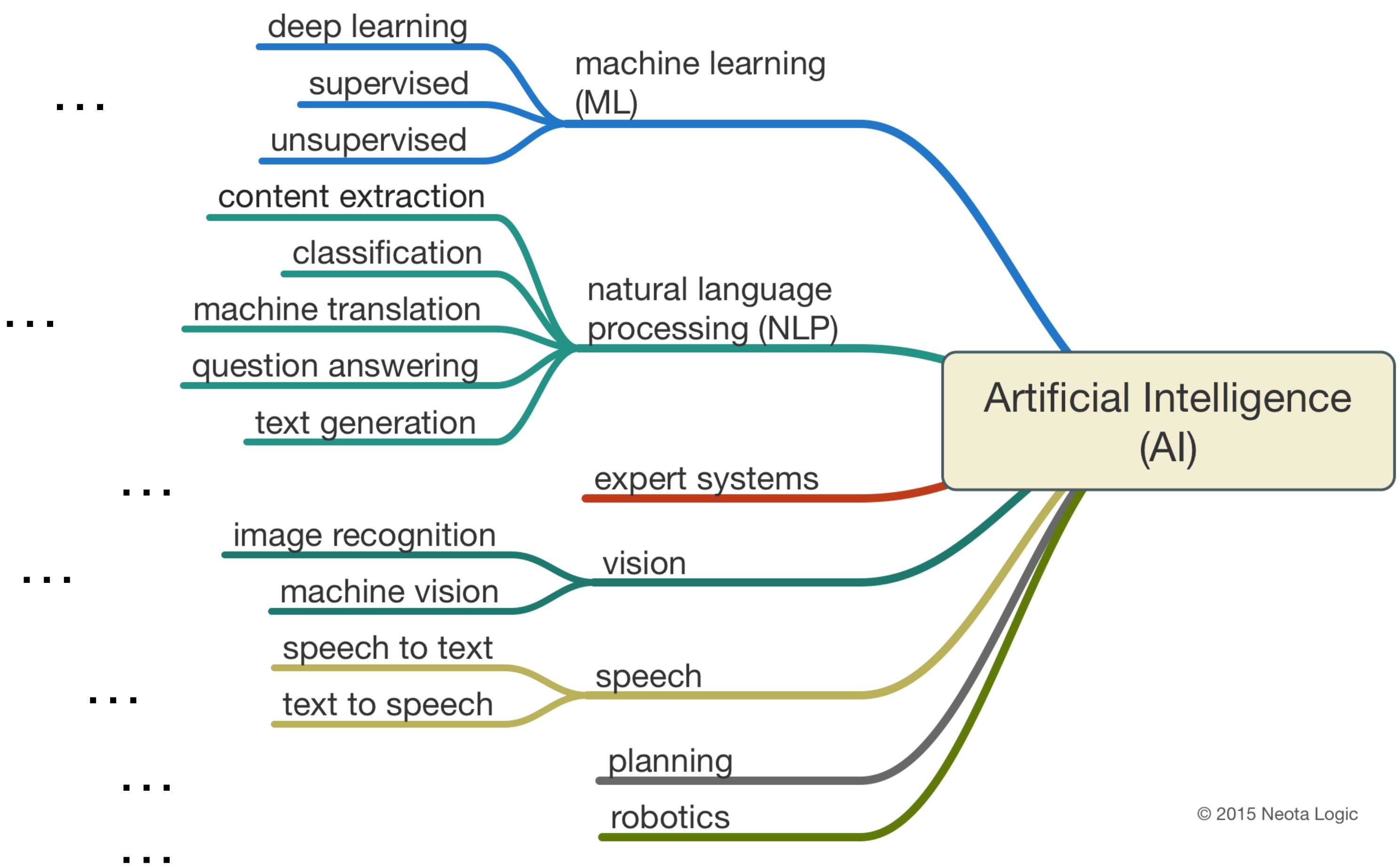
"AI... is concerned with **intelligent behaviour in artefacts**." (Nilsson, 1998)

← concerned with →

acting

The scientific field concerned with complex computer applications

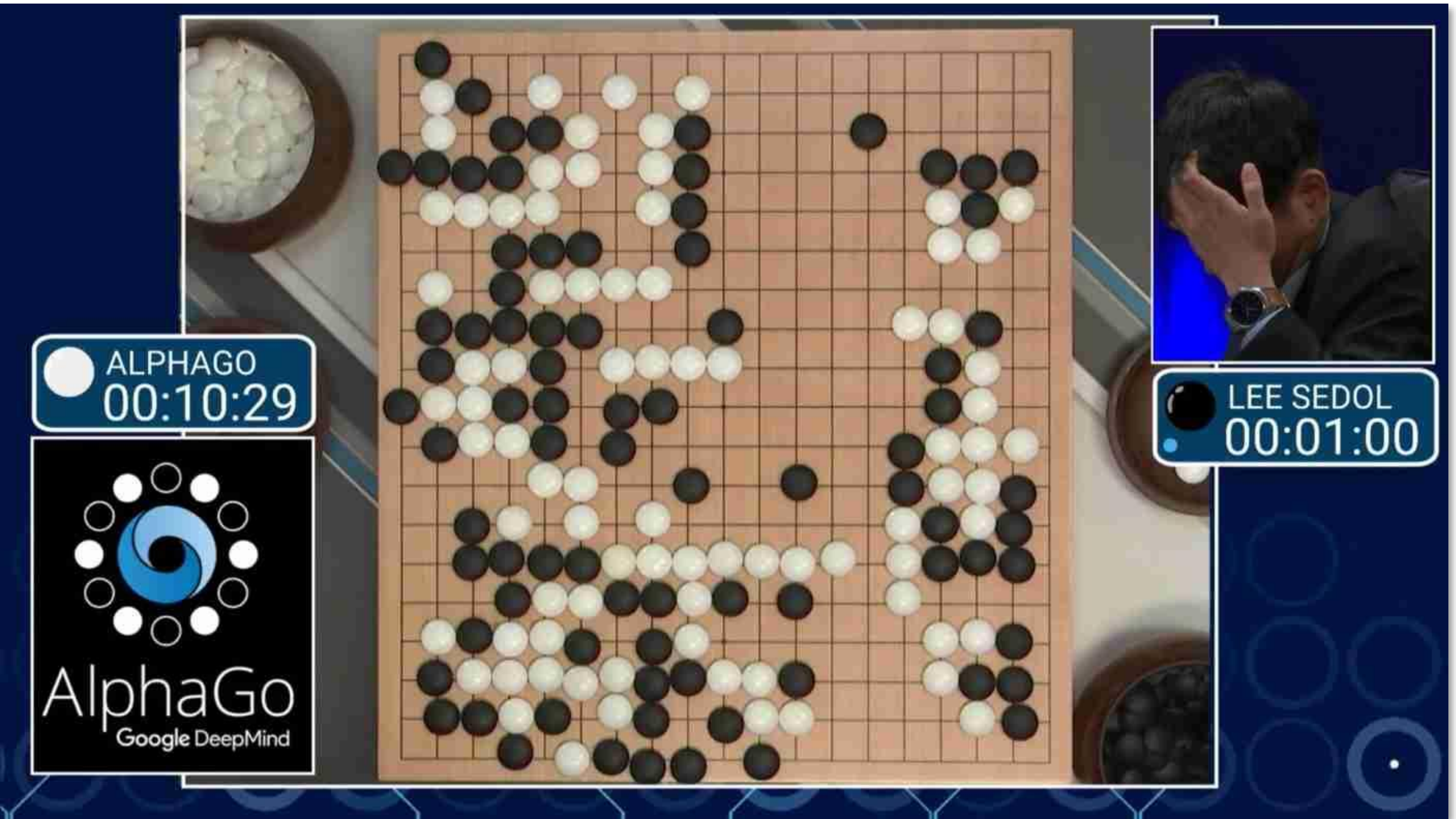
Was gehört zu künstlicher Intelligenz?



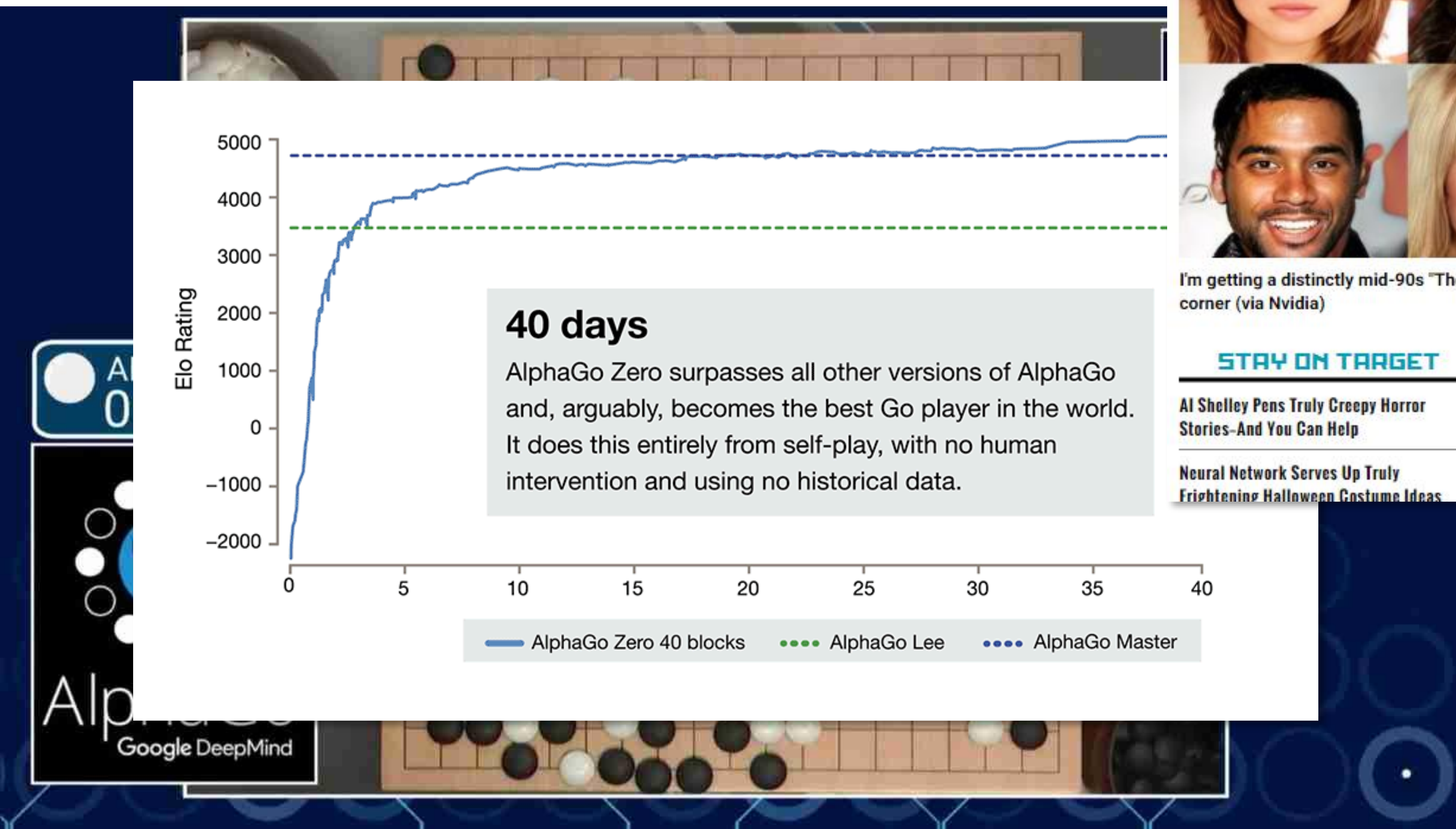
Was?→ Wie?

1

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


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

Zürcher Hochschule
für Angewandte Wissenschaften




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.

Deep neural networks can now transfer the style of one photo onto another

And the results are impressive

by James Vincent | @jjvincent | Mar 30, 2017, 1:53pm EDT

SHARE TWEET LINKEDIN

Original photo | Reference photo | Result

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.

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.

Deep neural networks can now transfer the style of one photo onto another

And the results are impressive

by James Vincent | @jjvincent | Mar 30, 2017, 1:53pm EDT

SHARE TWEET LINKEDIN

Original photo Reference photo Result

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)

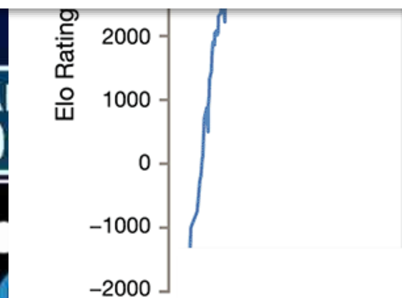
STAY ON TARGET

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.



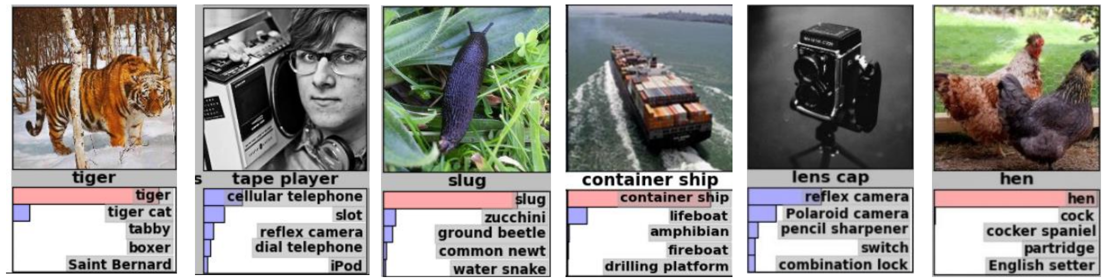
40 days

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.

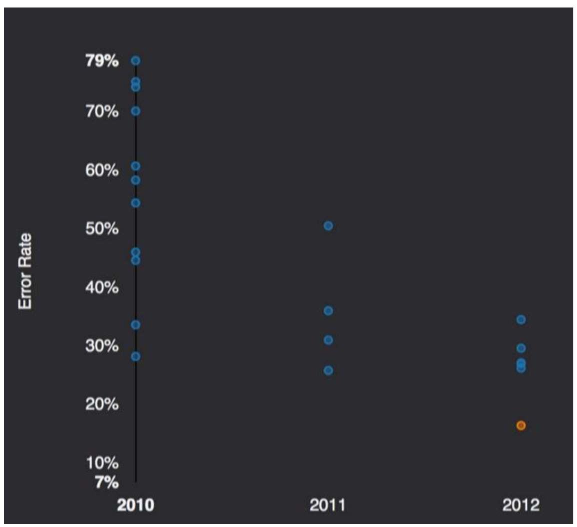
Zürcher Hochschule
für Angewandte Wissenschaften

Was ist passiert?
Der ImageNet Wettbewerb

1000 Kategorien
1 Mio. Beispiele

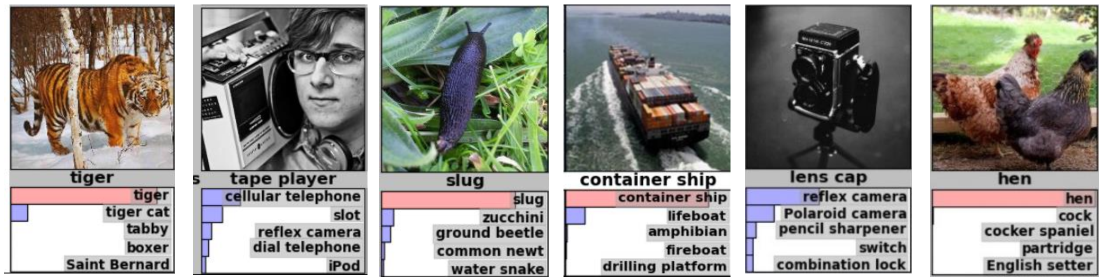


…

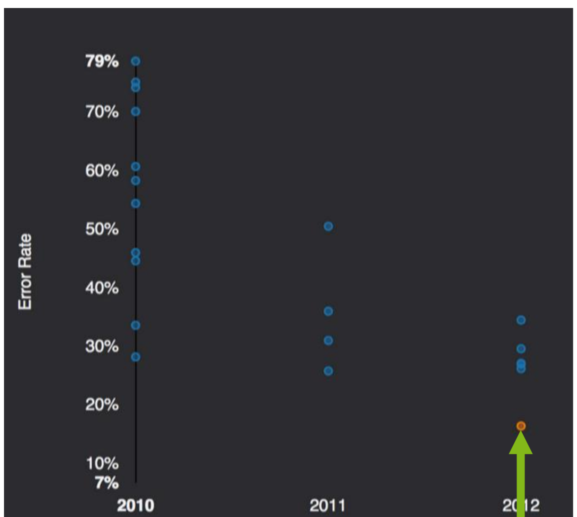


Was ist passiert?
Der ImageNet Wettbewerb

1000 Kategorien
1 Mio. Beispiele



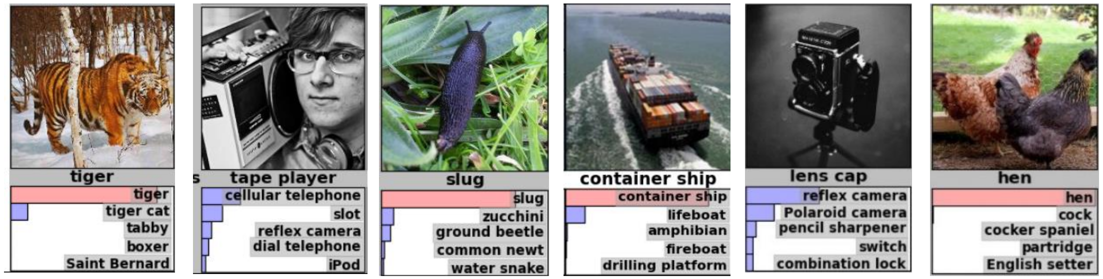
…



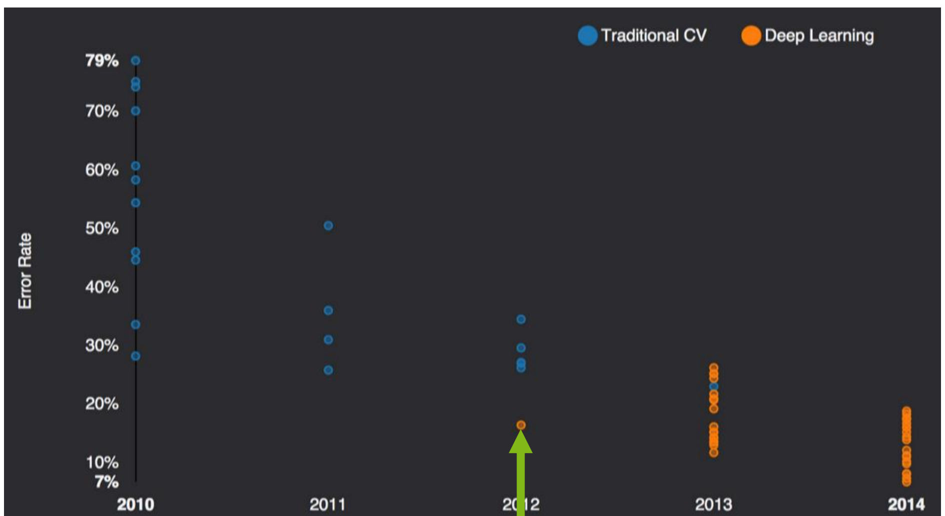
A. Krizhevsky verwendet als erster ein sog. «Deep Neural Network» (CNN)

Was ist passiert?
Der ImageNet Wettbewerb

1000 Kategorien
1 Mio. Beispiele



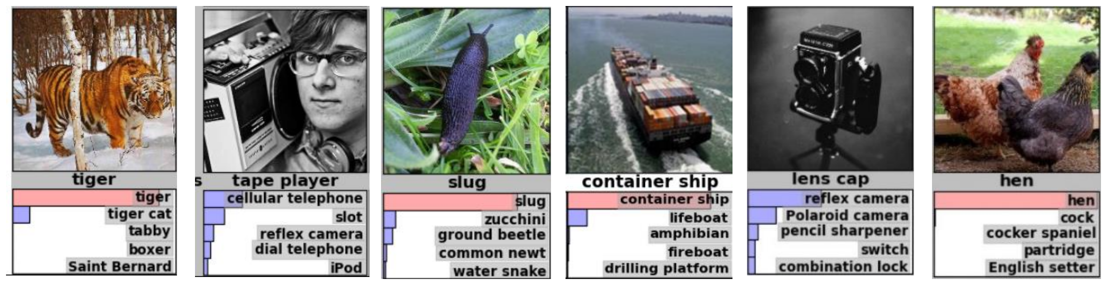
…



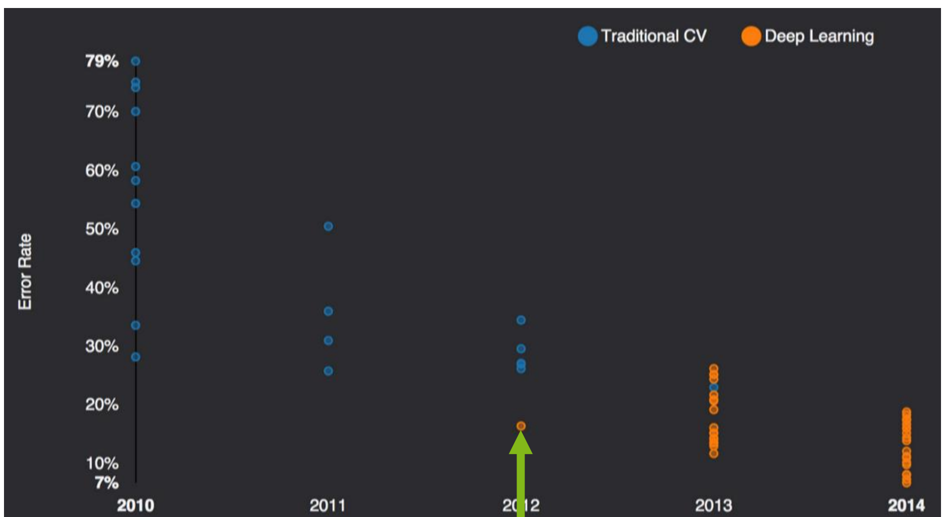
A. Krizhevsky verwendet als erster ein sog. «Deep Neural Network» (CNN)

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)

Zürcher Hochschule
für Angewandte Wissenschaften

2

Wie geht das?

Idee: Mehr Tiefe zum Lernen von Merkmalen

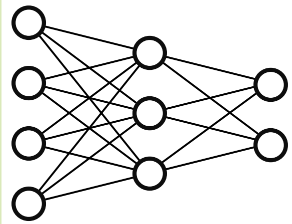
Klassische Bild-verarbeitung

Merkmalsextraktion
(SIFT, SURF, LBP, HOG, etc.)

(0.2, 0.4, …)

(0.4, 0.3, …)

Klassifikation
(SVM, Neuronales Netz, etc.)



Containerschiff

Tiger

…

Idee: Mehr Tiefe zum Lernen von Merkmalen

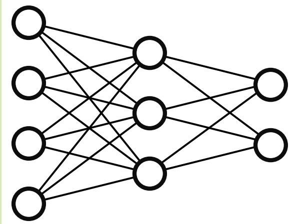
Klassische Bild-verarbeitung

**Merkmalsextraktion
(SIFT, SURF, LBP, HOG, etc.)**

(0.2, 0.4, …)

(0.4, 0.3, …)

**Klassifikation
(SVM, Neuronales Netz, etc.)**



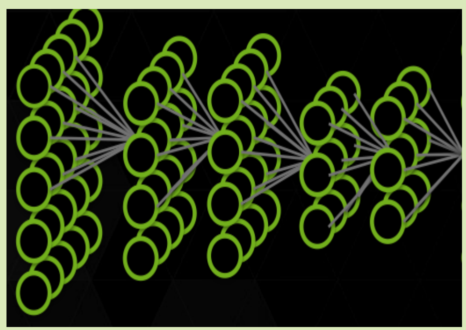
Containerschiff

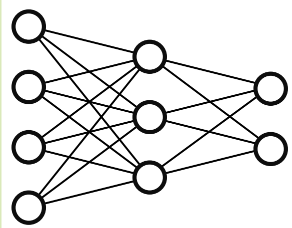
Tiger

…

Mit Convolutional Neural Networks (CNNs)

**Nimmt rohe Pixel entgegen,
Merkmale werden mitgelernt!**





Containerschiff

Tiger

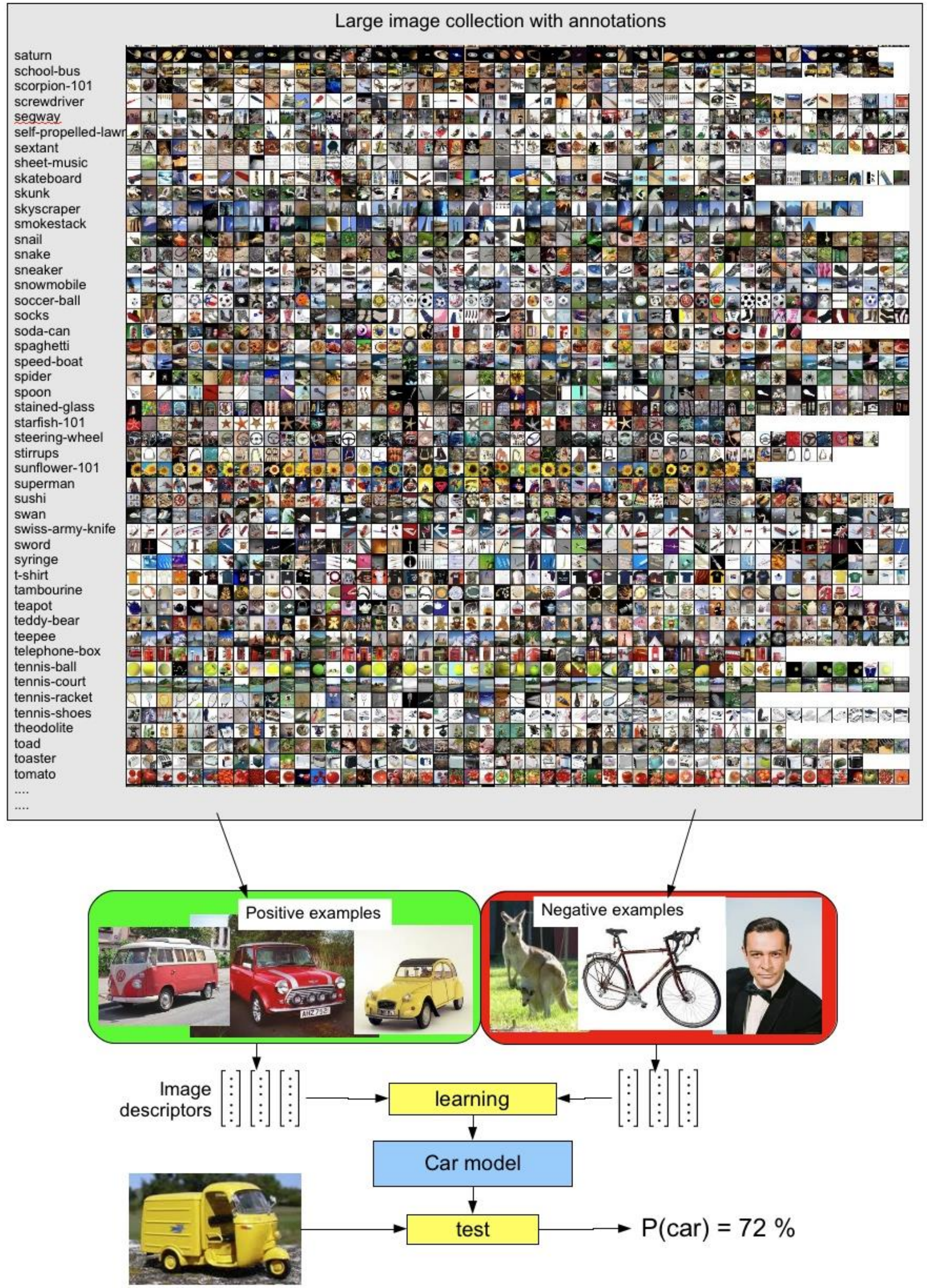
…

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

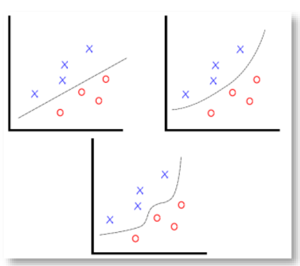
Grundlage
Induktives überwachtes Lernen

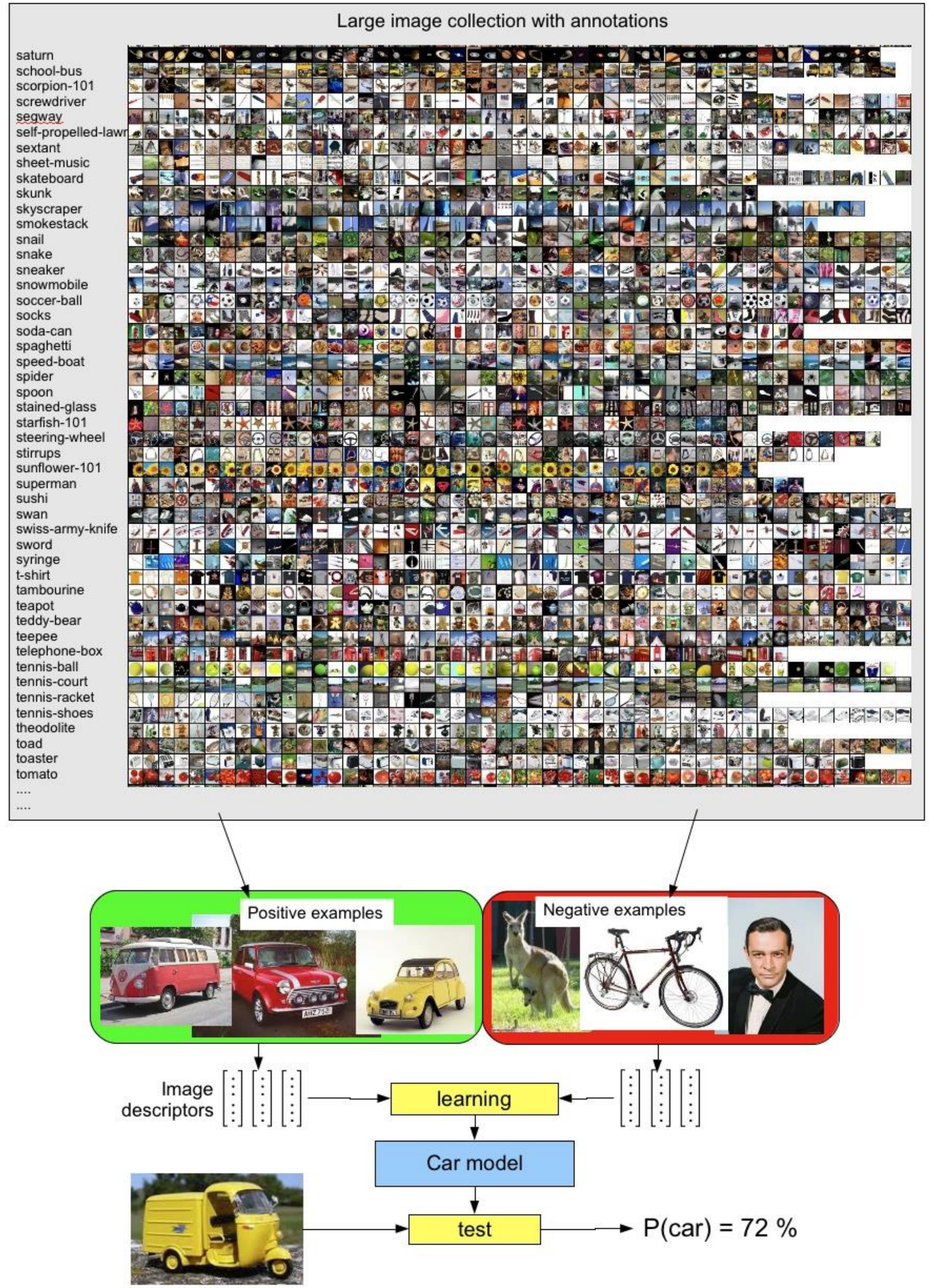
Annahme

- 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





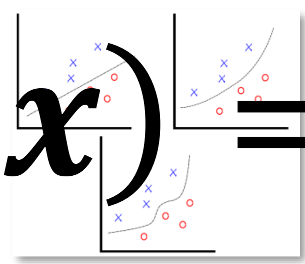
Quelle: http://lear.inrialpes.fr/job/postdoc-large-scale-classif-11-img/attribs_patchwork.jpg

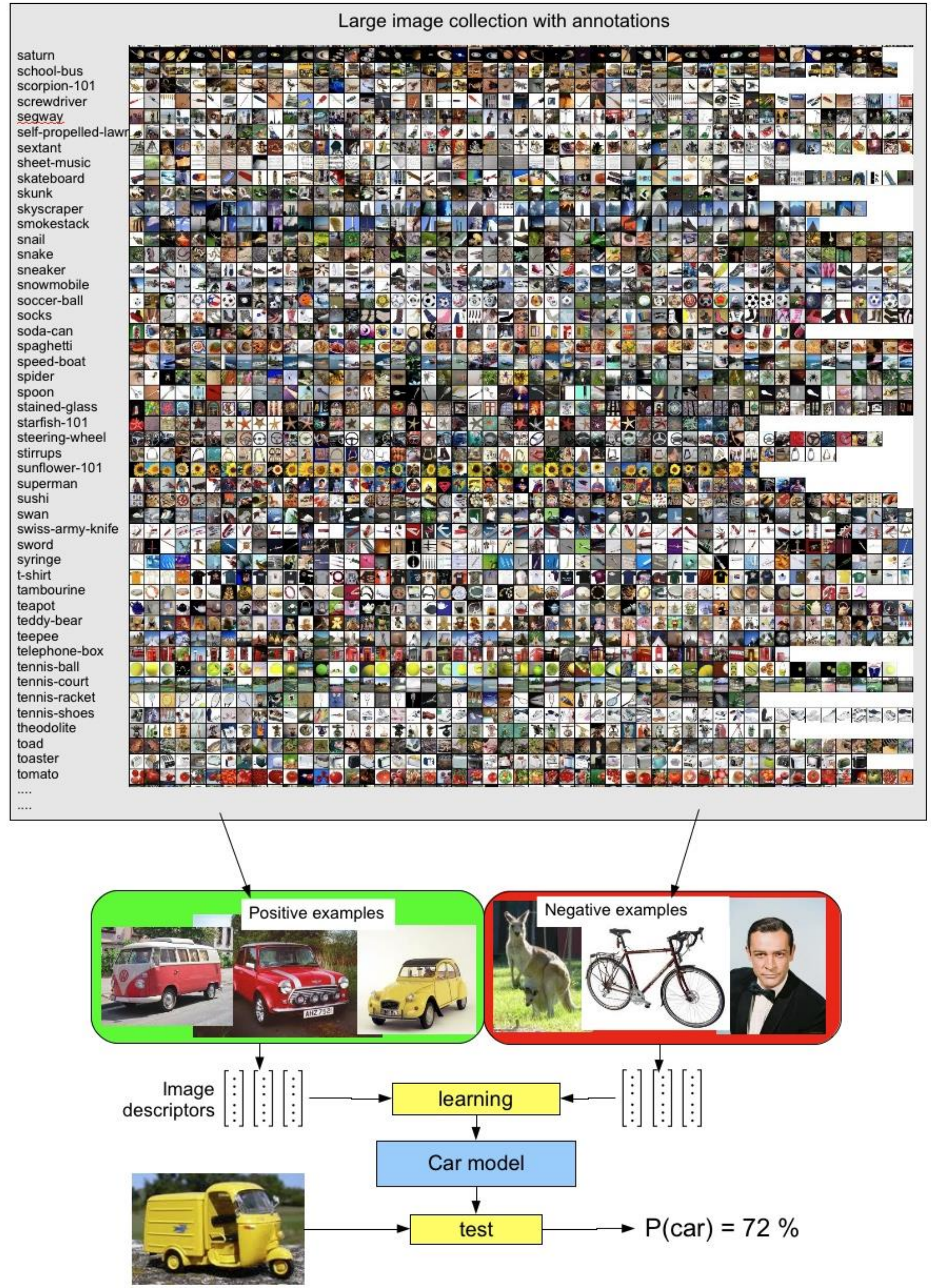
Grundlage
Induktives überwachtes Lernen

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

$$\boldsymbol{f(x) = y}$$

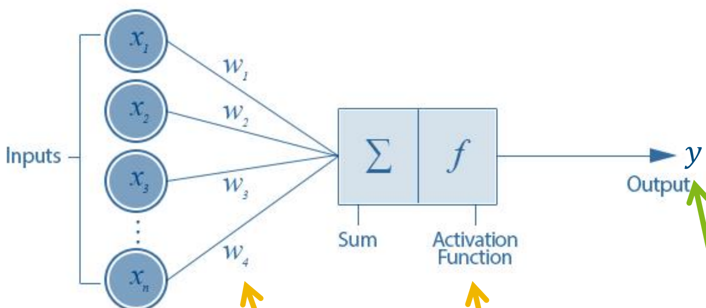




Quelle: http://lear.inrialpes.fr/job/postdoc-large-scale-classif-11-img/attribs_patchwork.jpg

Suche der Parameter *einer Funktion*?

Neuron



Merkmale (z.B. Pixel)

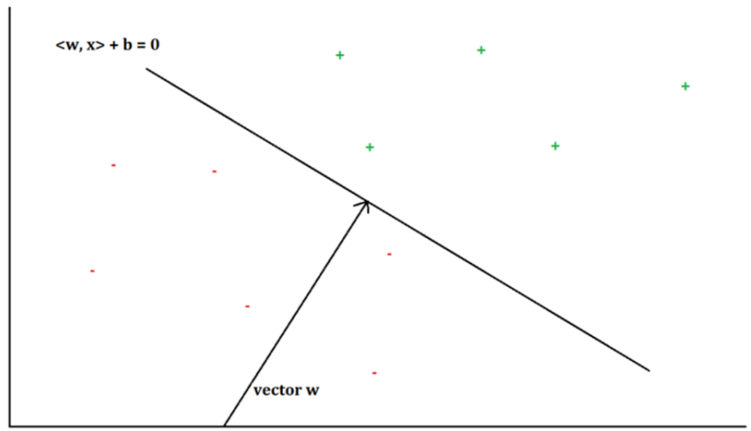
Anpassbare Parameter

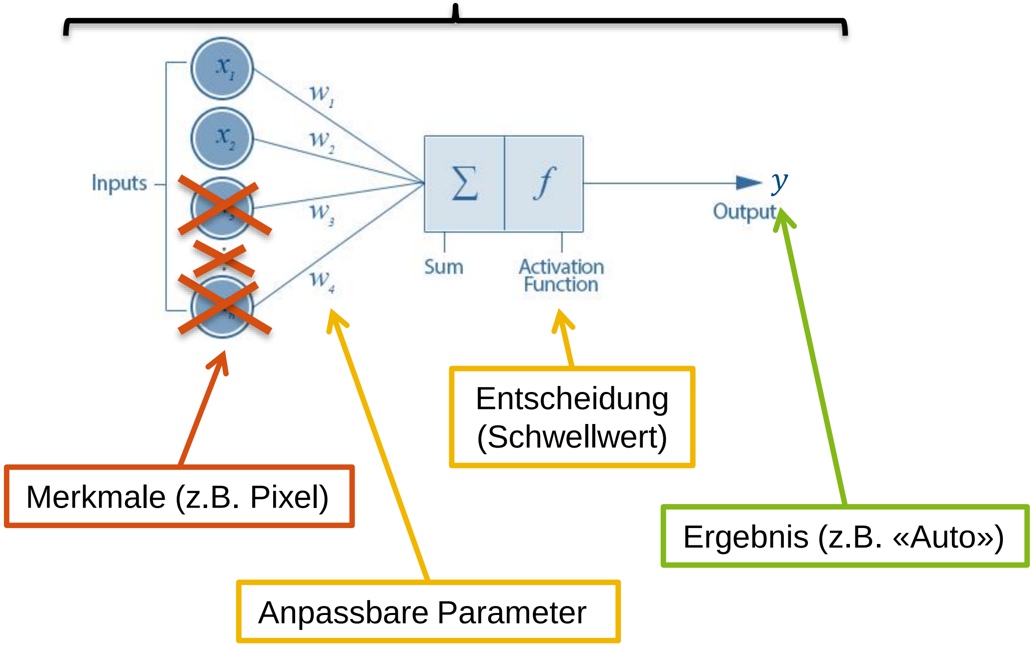
Entscheidung (Schwellwert)

Ergebnis (z.B. «Auto»)

Suche der Parameter *einer Funktion*?

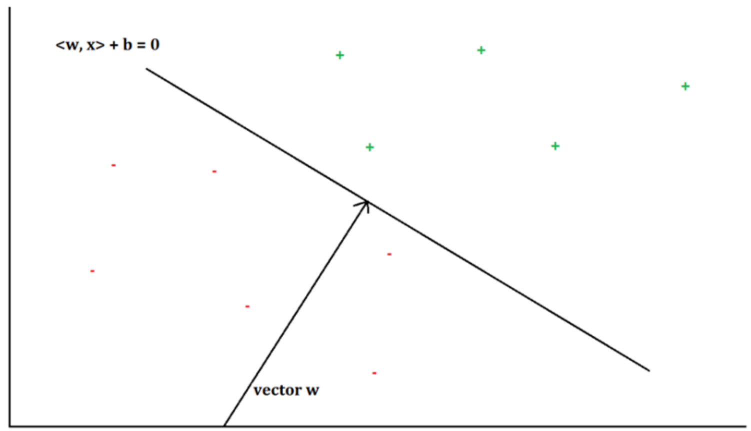
Neuron

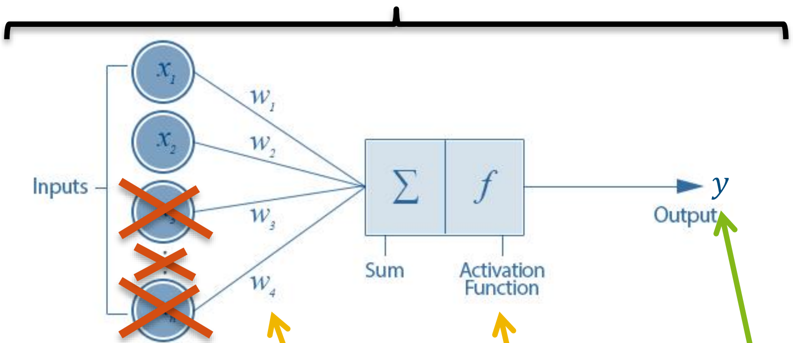




Suche der Parameter *einer Funktion*?

Neuron





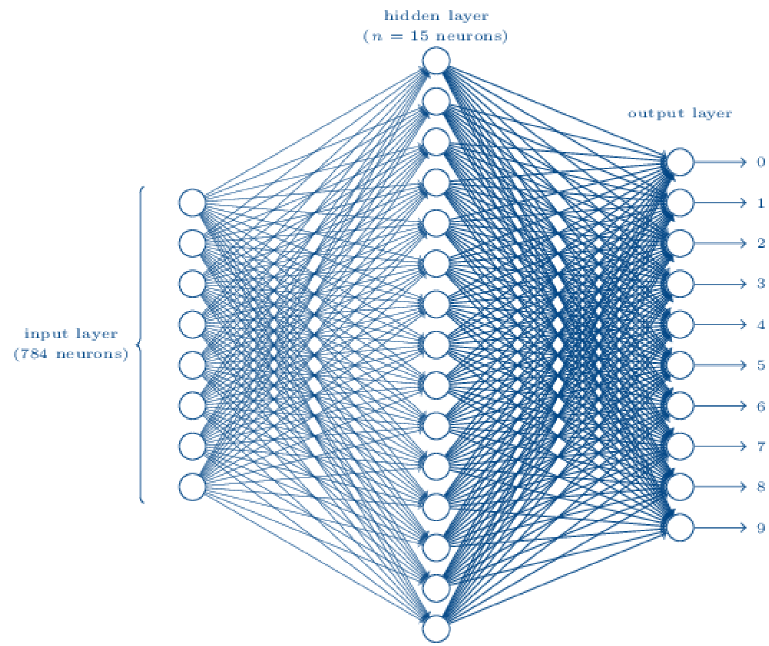
Merkmale (z.B. Pixel)

Anpassbare Parameter

Entscheidung (Schwellwert)

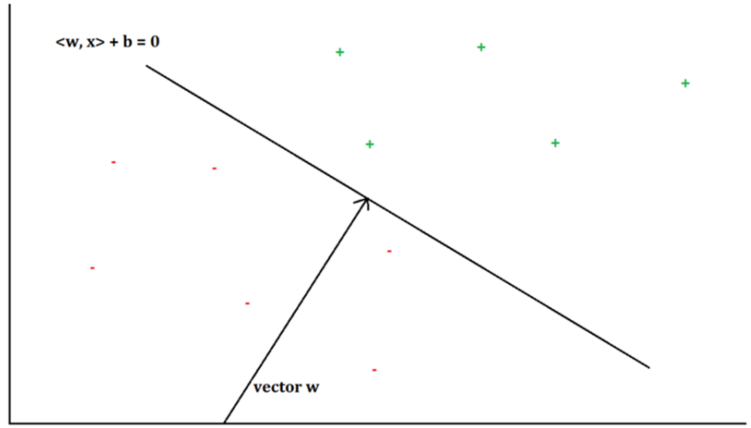
Ergebnis (z.B. «Auto»)

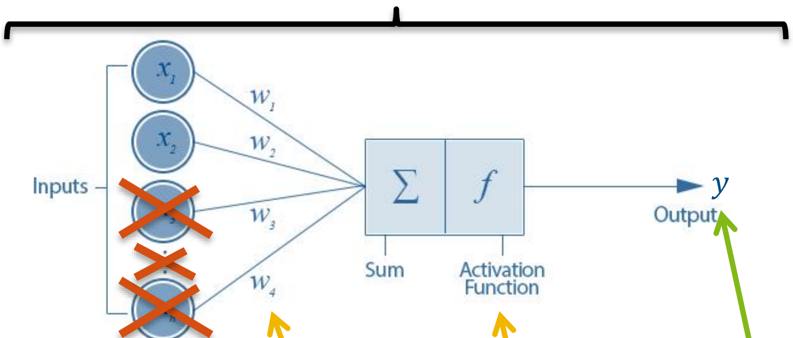
Neuronales Netz



Suche der Parameter *einer Funktion*?

Neuron





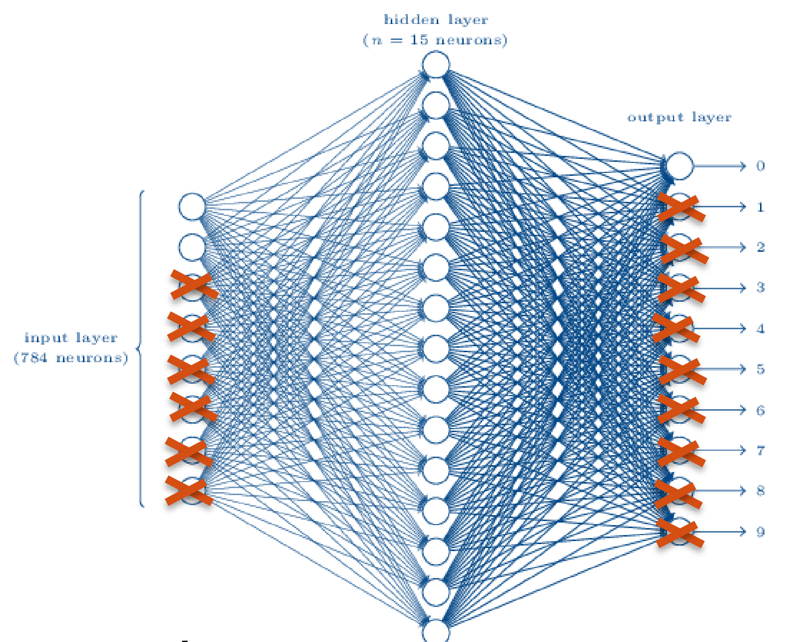
Entscheidung (Schwellwert)

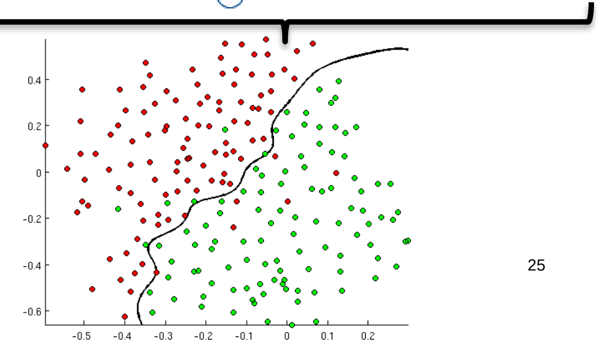
Merkmale (z.B. Pixel)

Ergebnis (z.B. «Auto»)

Anpassbare Parameter

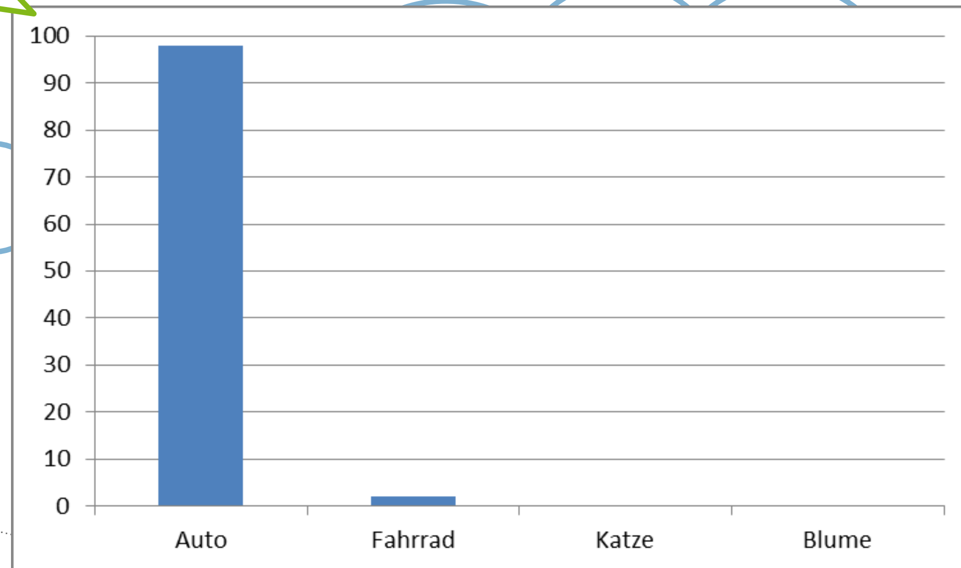
Neuronales Netz





Suche der Parameter einer Funktion?

Wahrscheinlichkeit [%] für bestimmtes Ergebnis



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

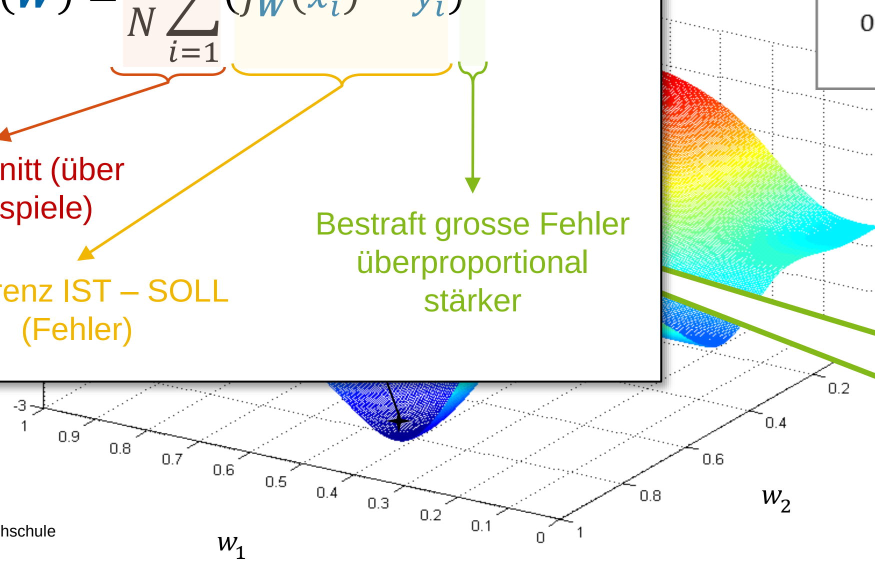
- Fehlermass: $l(\boldsymbol{W}) = \frac{1}{N}\sum_{i=1}^{N}(f_W(x_i) - y_i)^2$
  Durchschnitt der quadratischen Abweichungen
  über alle Bilder (Loss)

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

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

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!

datalab
www.zhaw.ch/datalab

swiss group for artificial intelligence and cognitive science
SGAICO

data+service
Swiss Alliance for Data-Intensive Services

Zu mir:

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

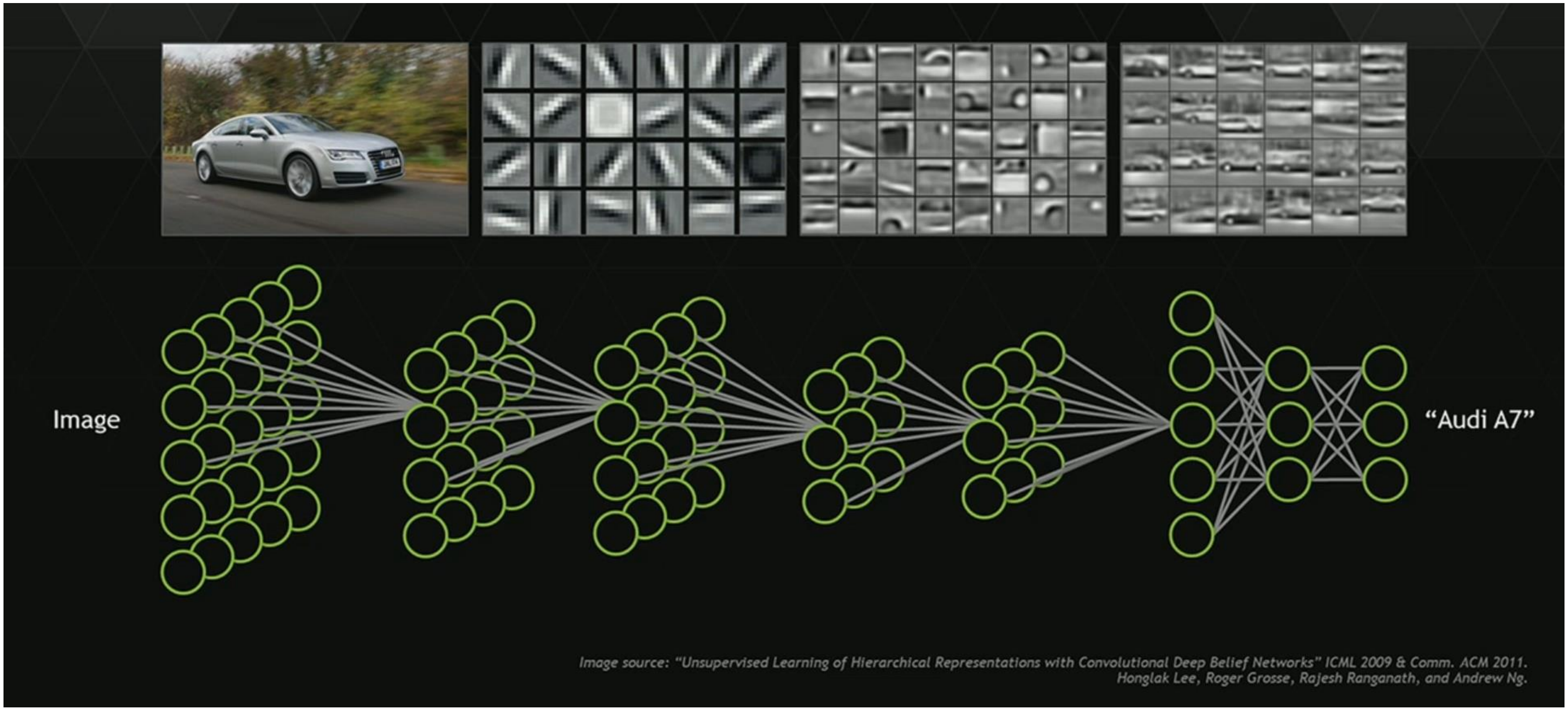
Mehr zum Thema:

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

➔ Fragen Sie gerne nach.

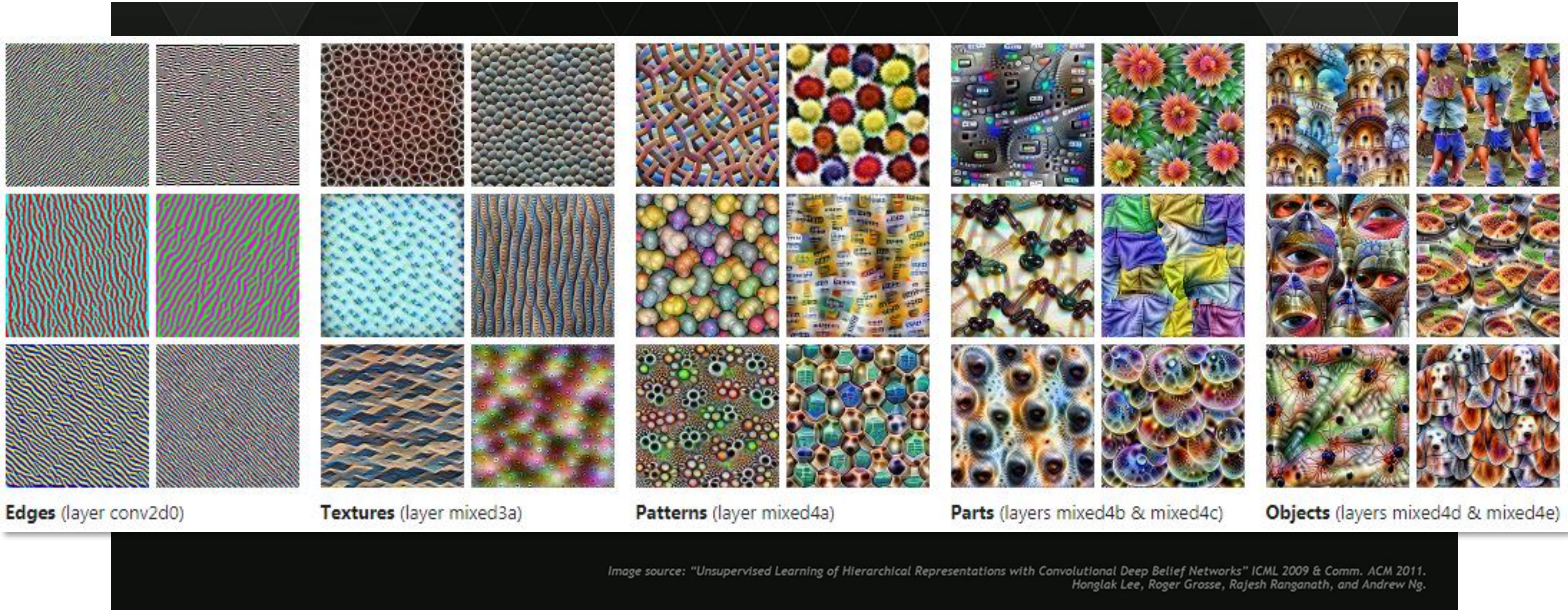
ANHANG

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/.

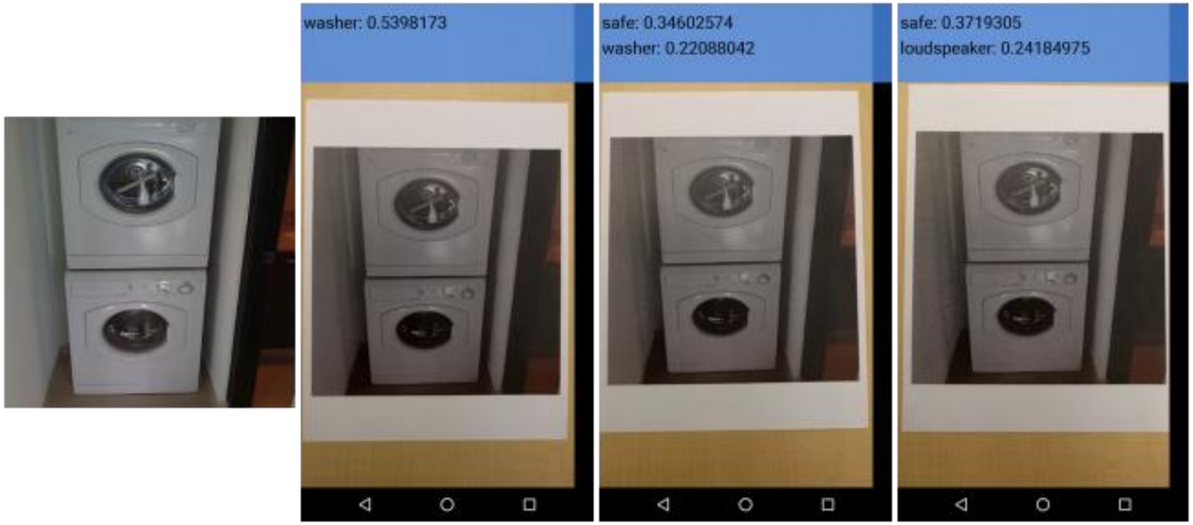
Was «sieht» das Neuronale Netz?
Hierarchien komplexer werdender Merkmale



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

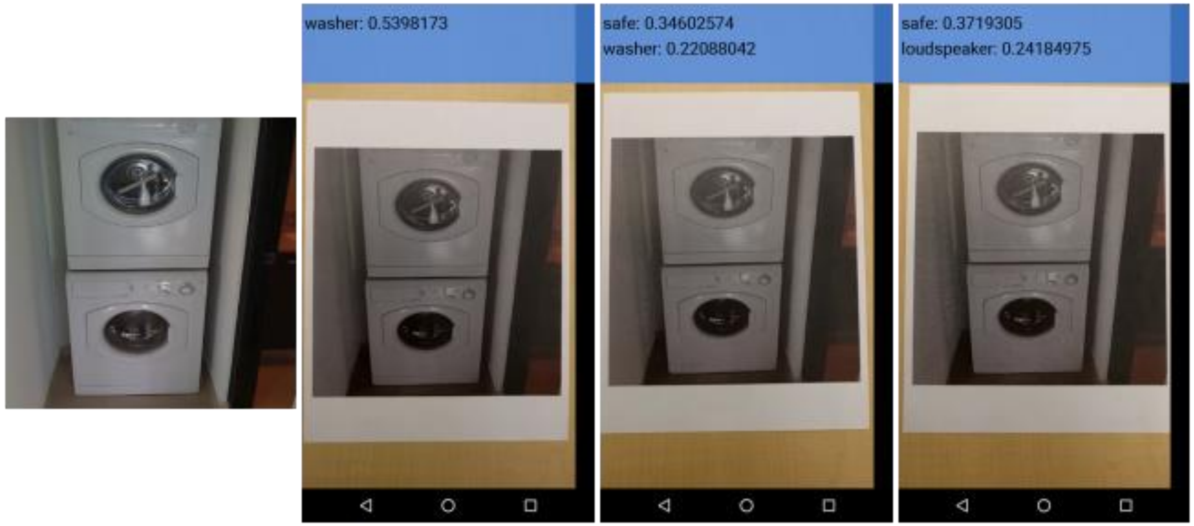


(a) Image from dataset (b) Clean image (c) Adv. image, $\epsilon = 4$ (d) Adv. image, $\epsilon = 8$

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

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

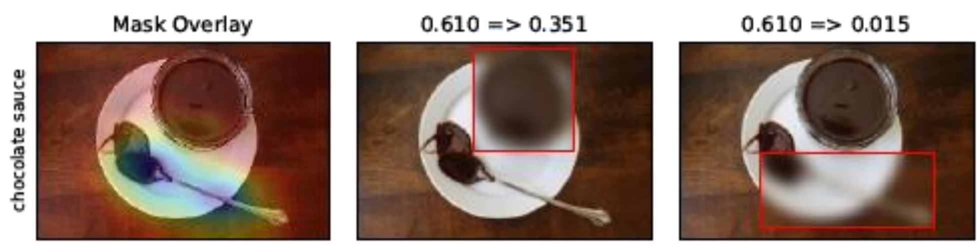
Verdeutlichen ein Problem:
- Adversarial Examples



(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

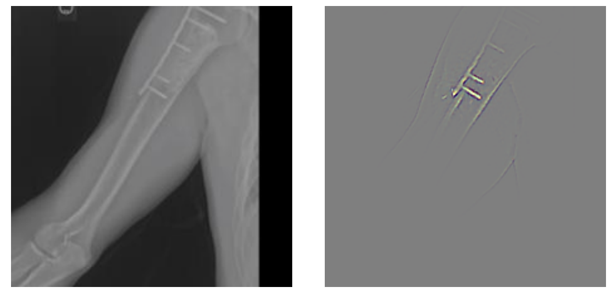
Lessons learned – model interpretability

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



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/

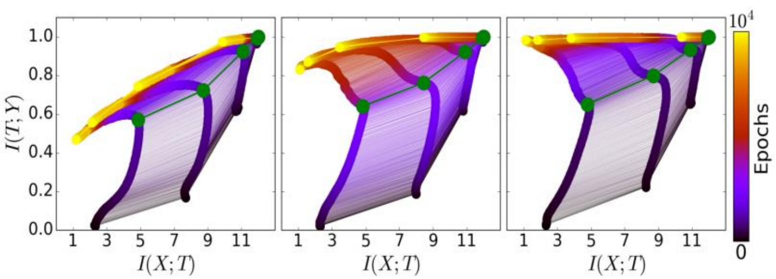
Lessons learned – model interpretability

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

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/

Zürcher Hochschule
für Angewandte Wissenschaften

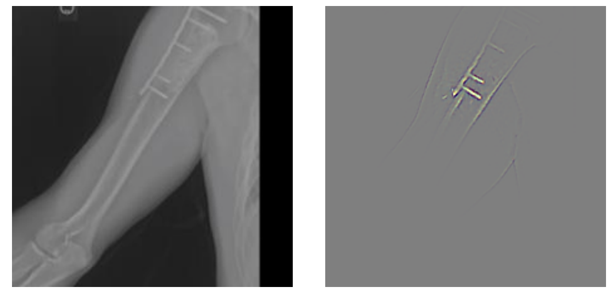
Lessons learned – model interpretability

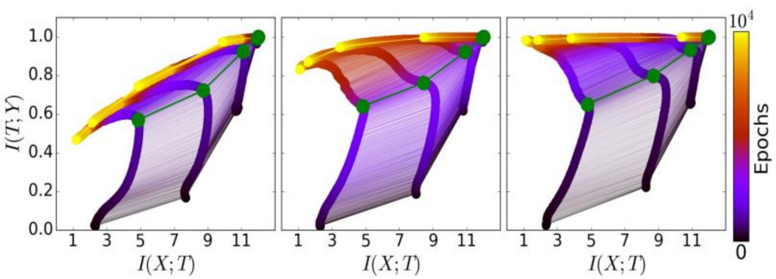
Interpretability is required.

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

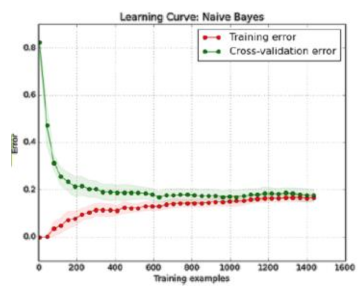
negative X-ray

positive X-ray





DNN training on the Information Plane



a learning curve

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/

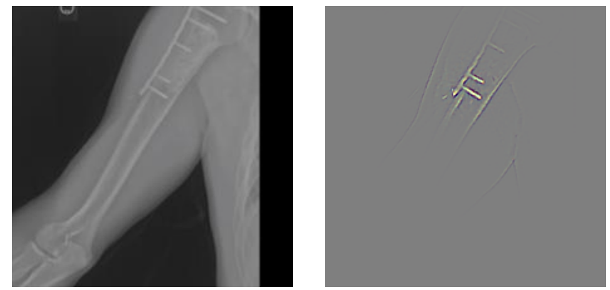
Lessons learned – model interpretability

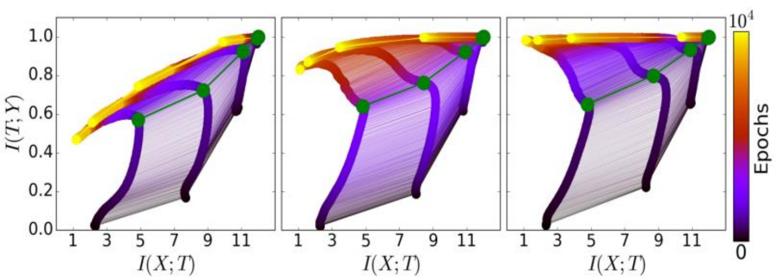
Interpretability is required.

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

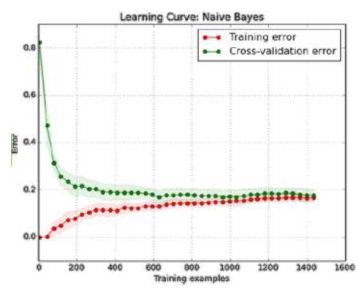
negative X-ray

positive X-ray

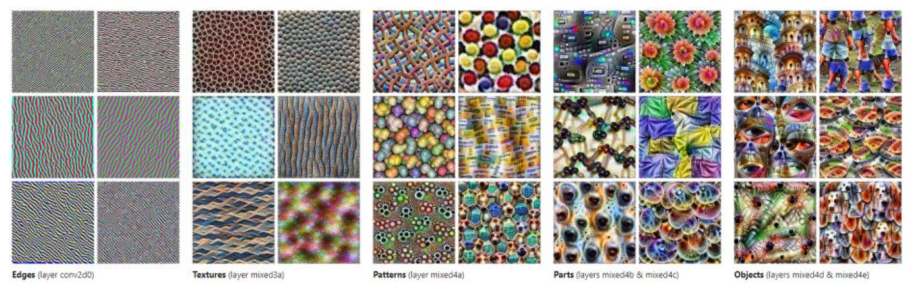




DNN training on the Information Plane



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/

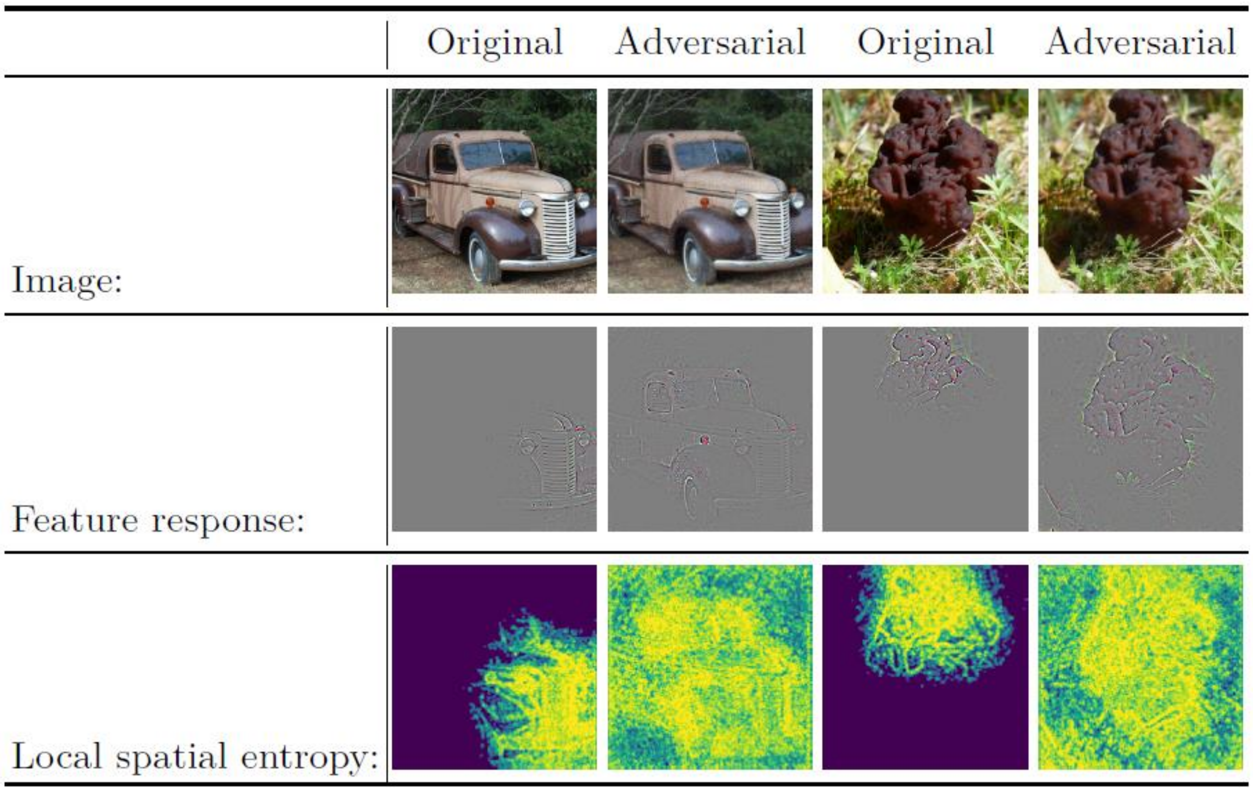
Zürcher Hochschule
für Angewandte Wissenschaften

Goody – trace & detect adversarial attacks
…using average local spatial entropy of feature response maps

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency



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