

Was Sie von KI erwarten können

Studerus Technology Forum, Regensdorf, 22. November 2018

Thilo Stadelmann



Swiss Alliance for
Data-Intensive Services



datalab

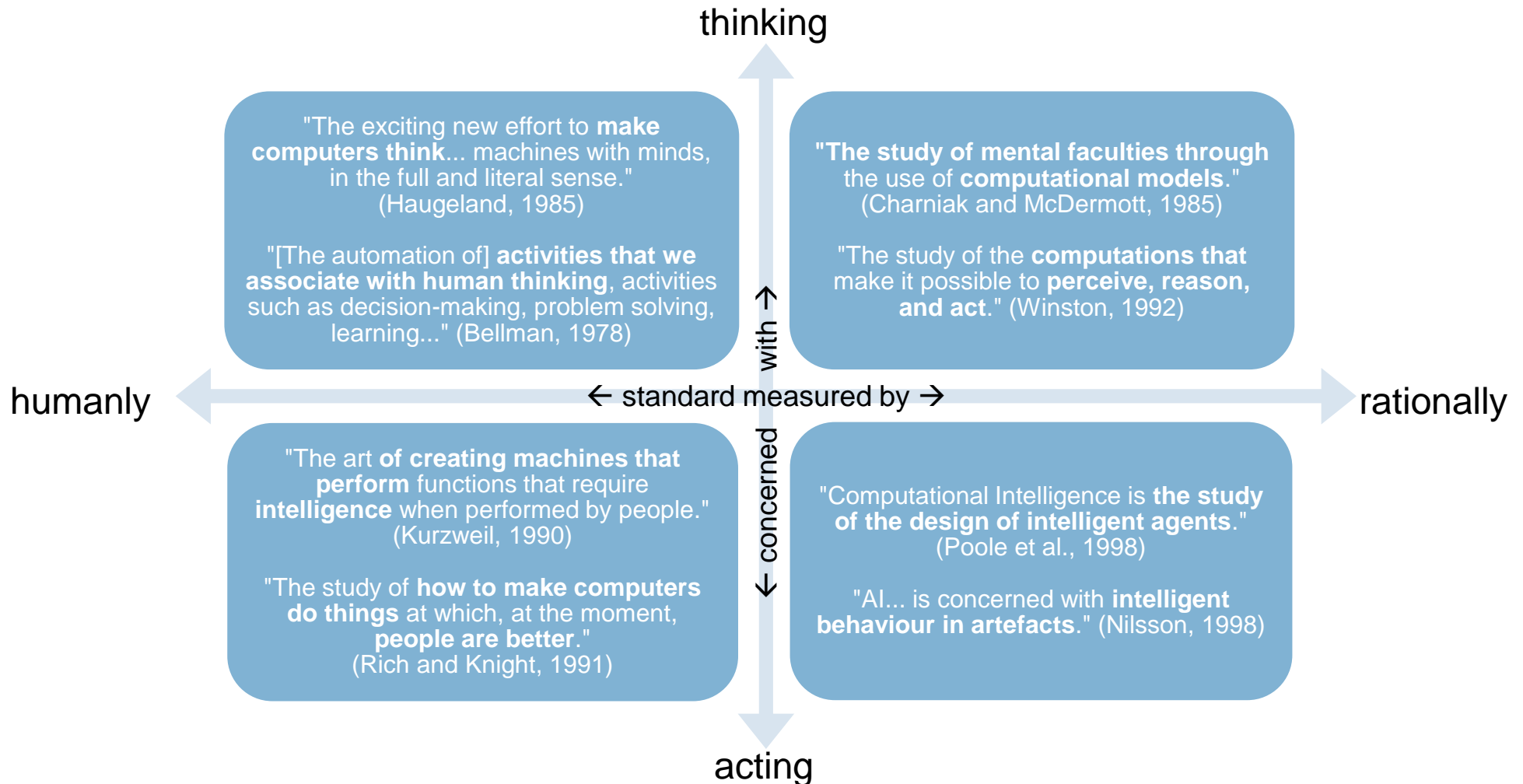
www.zhaw.ch/datalab

Prolog

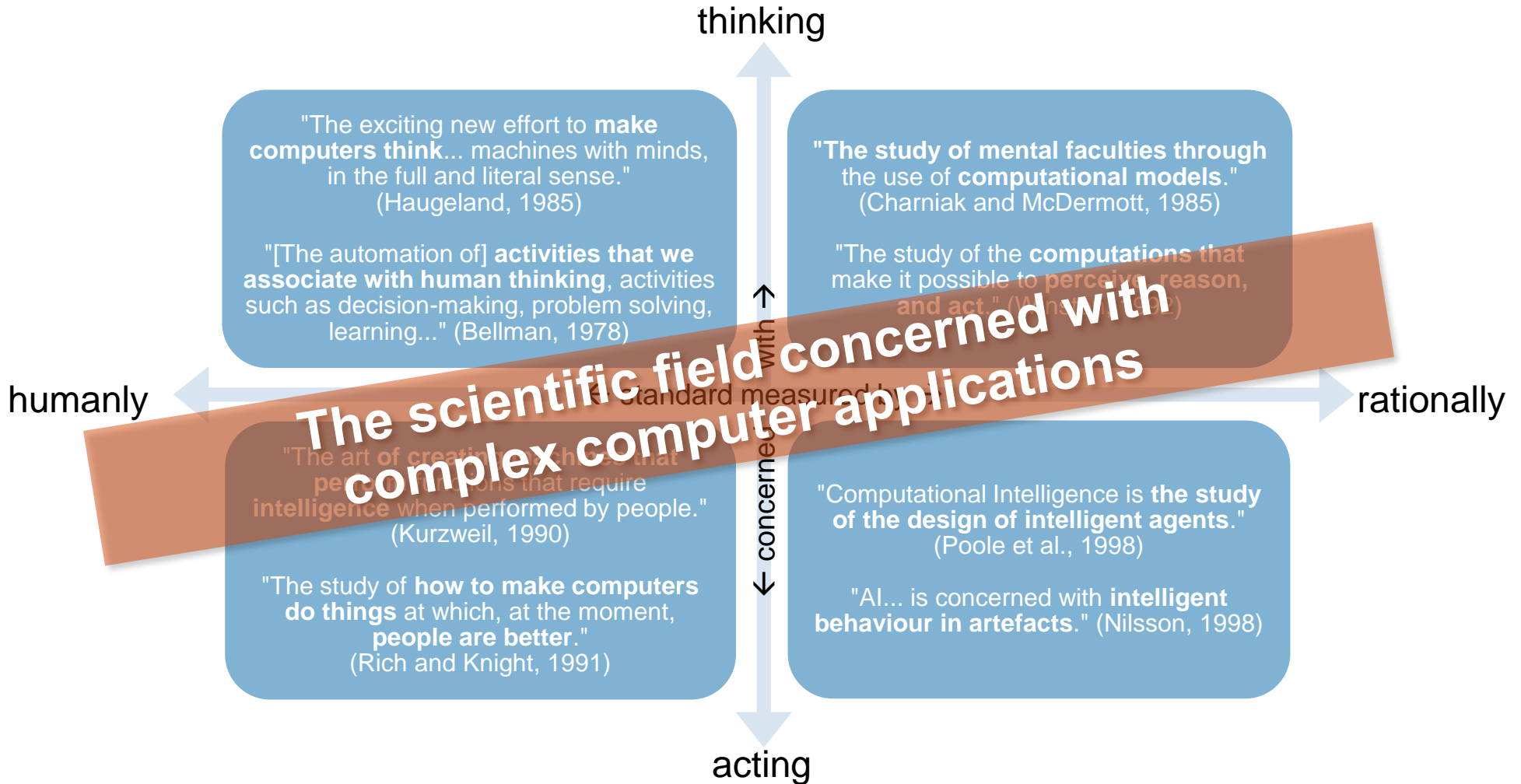


WHAT IS A.I.?

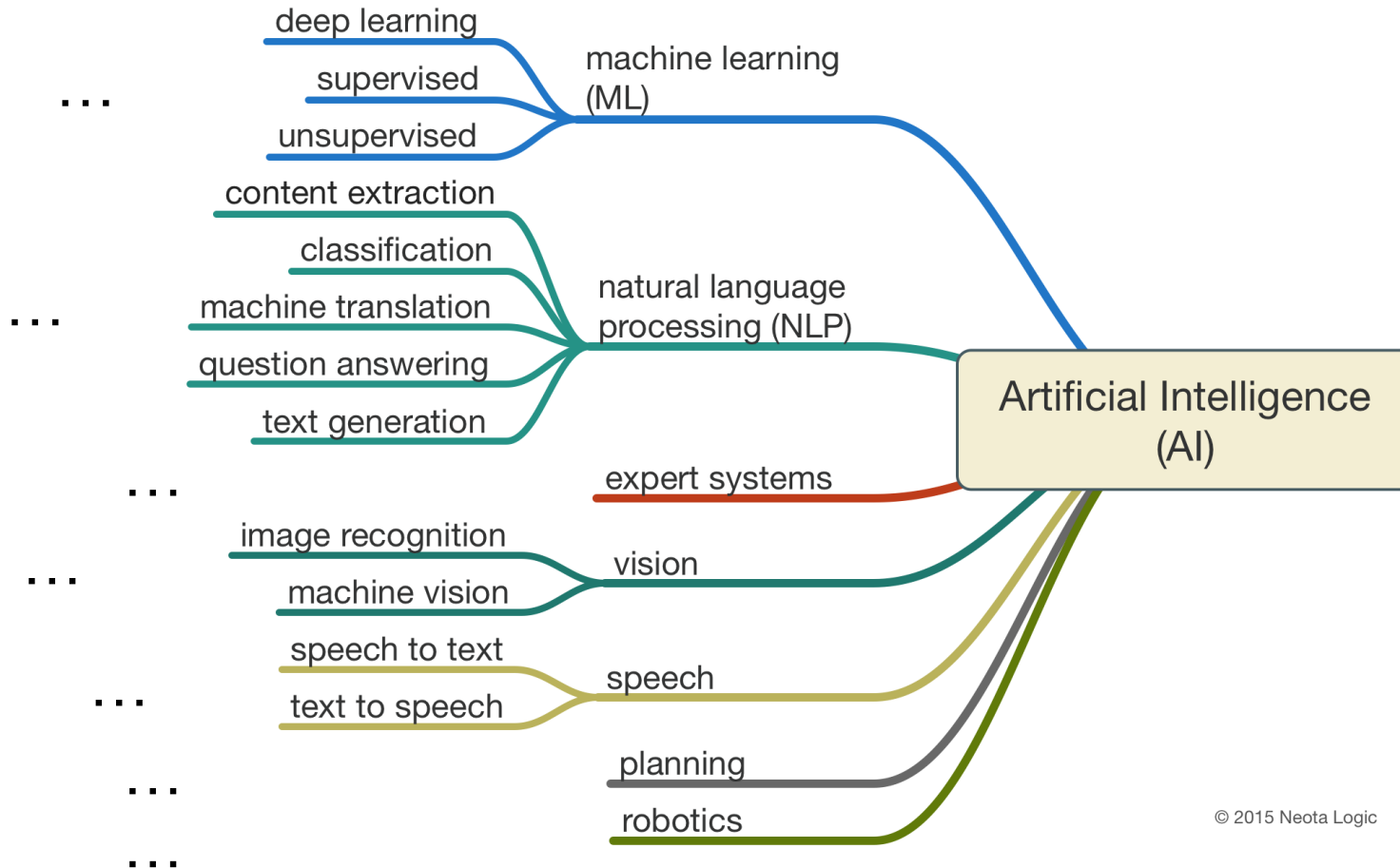
Was ist künstliche Intelligenz?



Was ist künstliche Intelligenz?



Was gehört zu künstlicher Intelligenz?

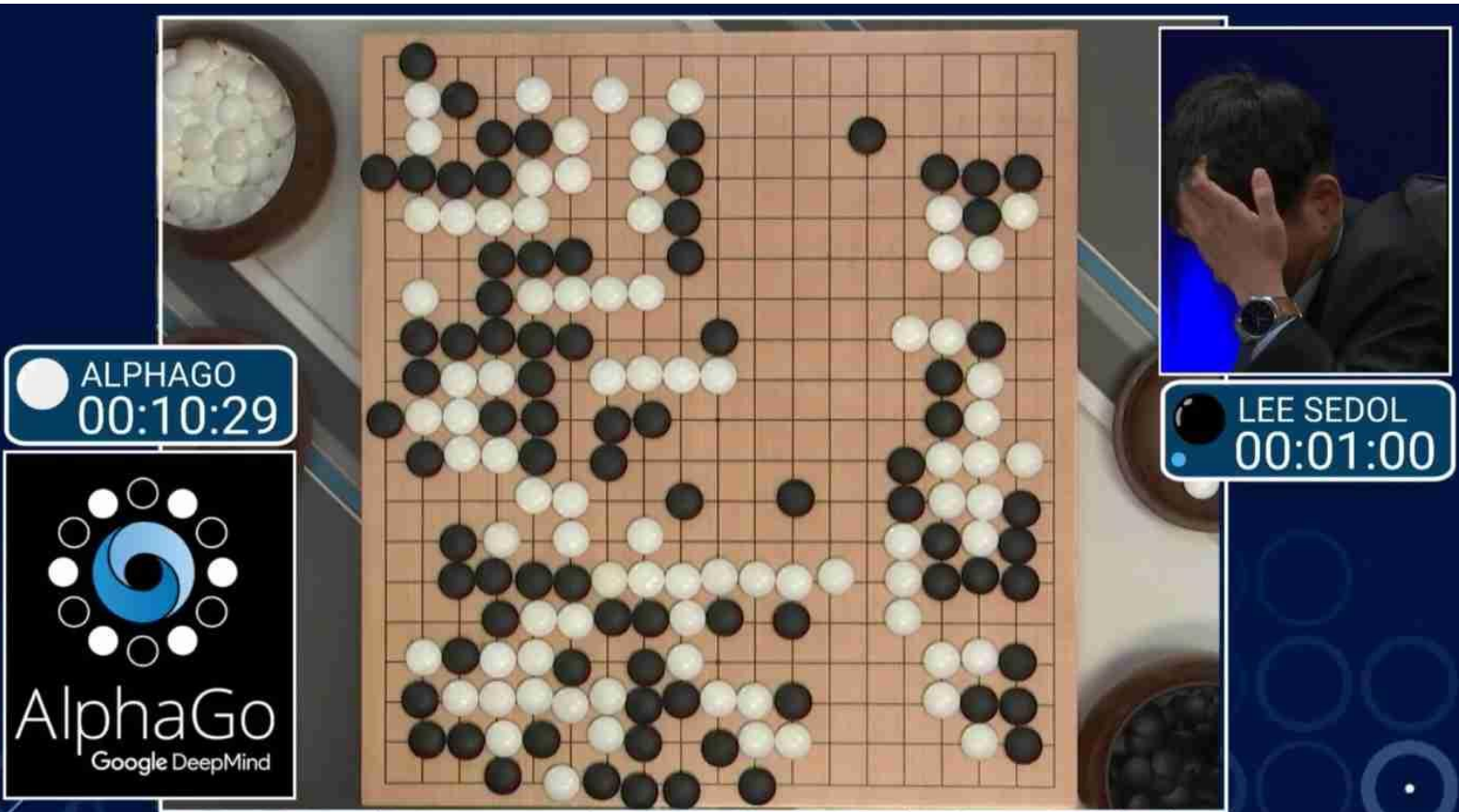


© 2015 Neota Logic

Was? → Wie?

1

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

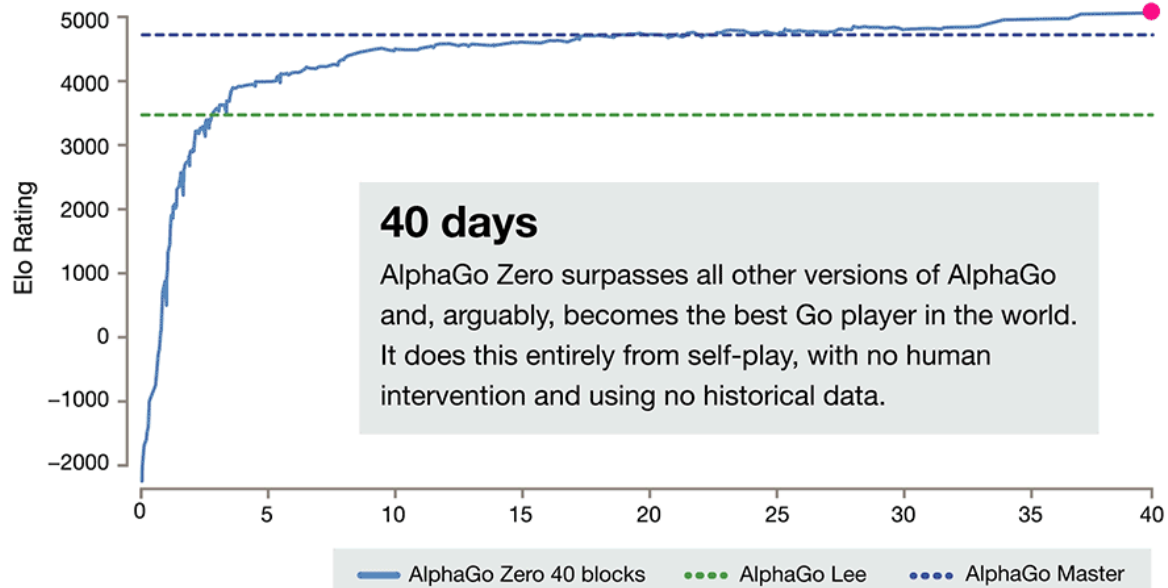


ALPHAGO
00:10:29

LEE SEDOL
00:01:00

AlphaGo
Google DeepMind

The image shows a Go board with black and white stones. On the left, a bowl contains white stones, and on the right, a bowl contains black stones. The board is partially filled with stones, indicating a game in progress. The player information overlays show AlphaGo's time at 00:10:29 and Lee Sedol's time at 00:01:00. The AlphaGo logo is also visible in the bottom left corner.



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
Google DeepMind

EDOL
01:00

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)

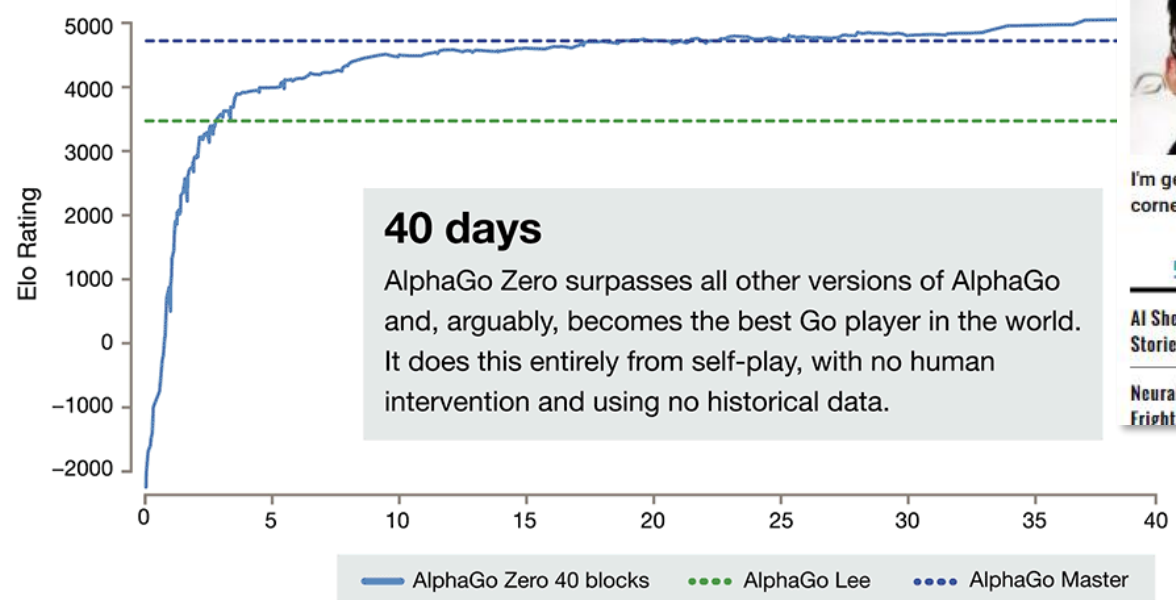
STAY ON TARGET

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.



AlphaGo
Google DeepMind



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

And the results are impressive

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



Original photo Reference photo Result

Ad closed by Google

Report this ad

AdChoices

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.



AlphaGo
Google DeepMind



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

And the results are impressive

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



Original photo

Reference photo

Result

Ad closed by Google

Report this ad

AdChoices

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)

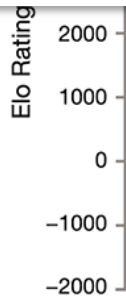
STAY ON TARGET

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.

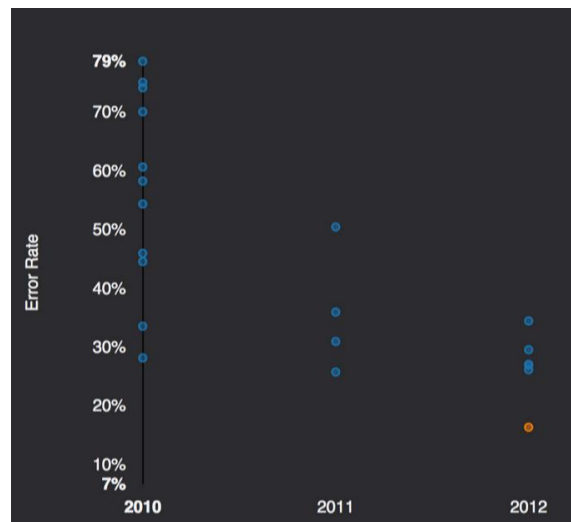
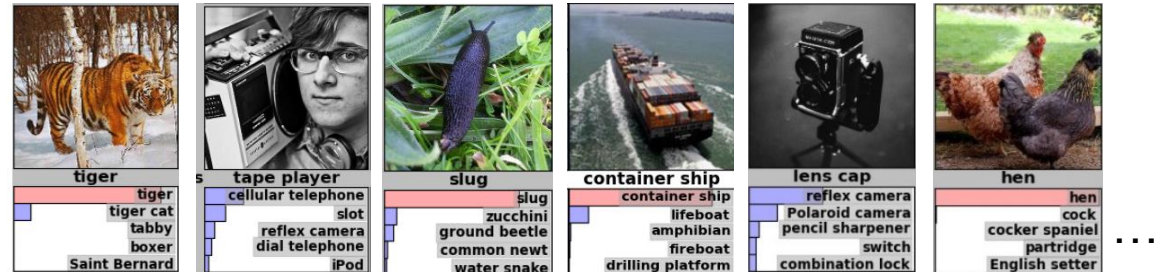


Was ist passiert?

Der ImageNet Wettbewerb



1000 Kategorien
1 Mio. Beispiele

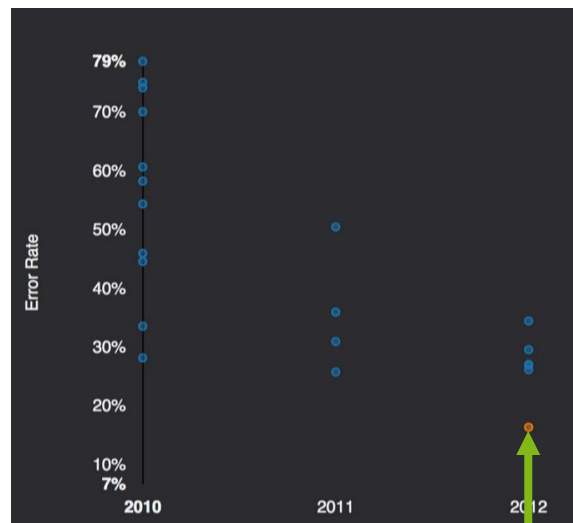
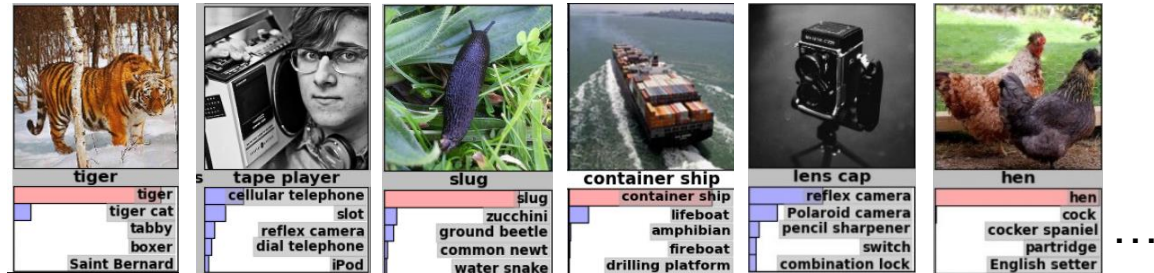


Was ist passiert?

Der ImageNet Wettbewerb



1000 Kategorien
1 Mio. Beispiele



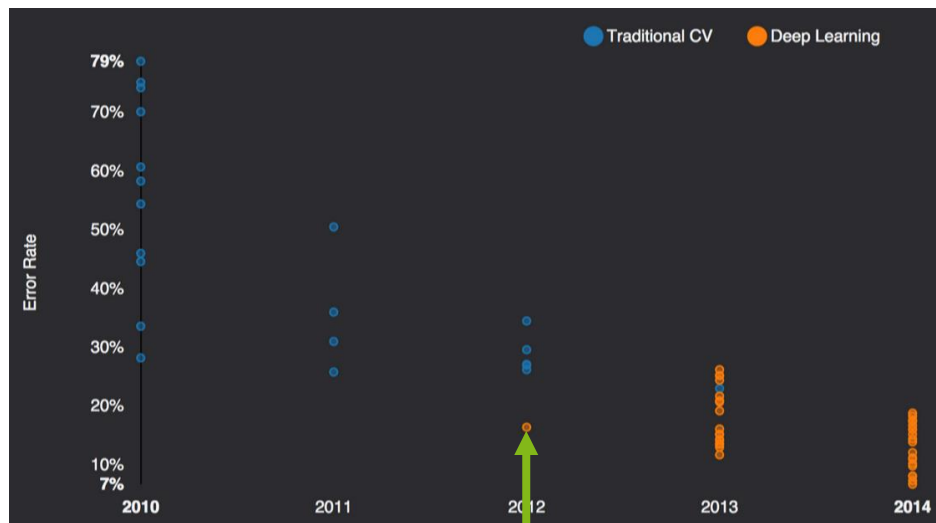
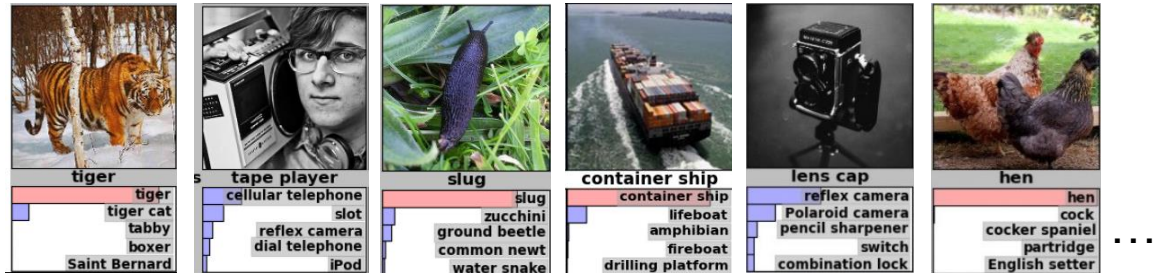
A. Krizhevsky verwendet als erster ein sog. «Deep Neural Network» (CNN)

Was ist passiert?

Der ImageNet Wettbewerb



1000 Kategorien
1 Mio. Beispiele



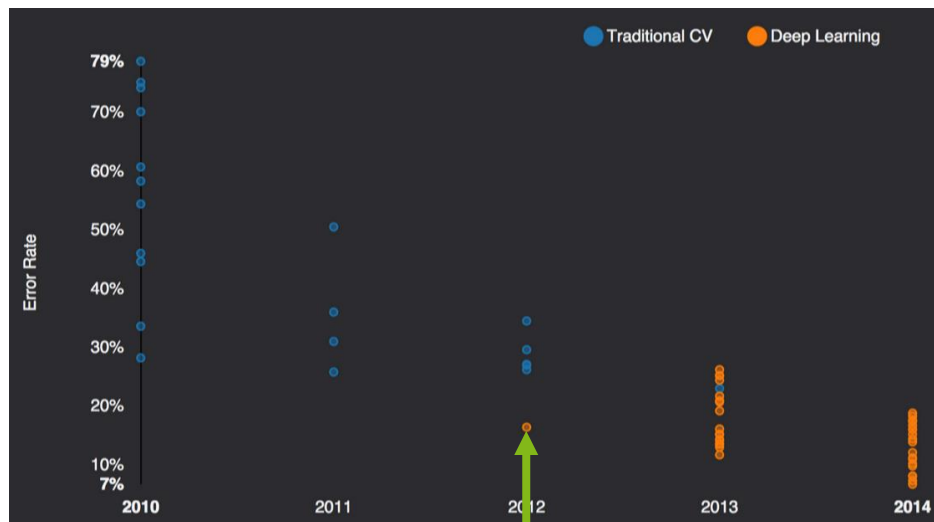
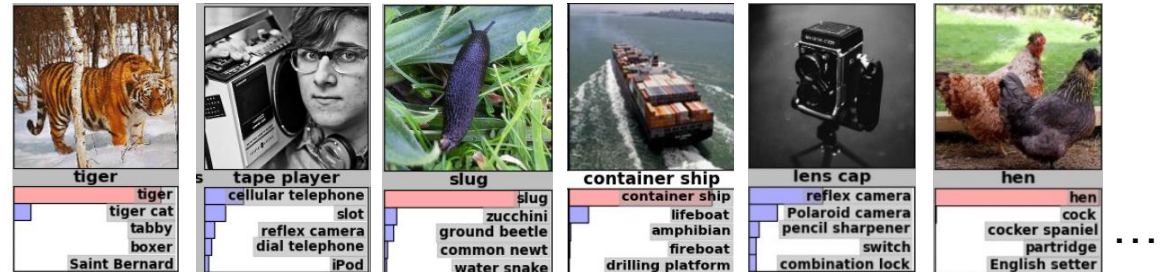
A. Krizhevsky verwendet als erster ein sog. «Deep Neural Network» (CNN)

Was ist passiert?

Der ImageNet Wettbewerb



1000 Kategorien
1 Mio. Beispiele



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)

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

Was? → Wie?

2

Wie geht das?

Idee: Mehr Tiefe zum Lernen von Merkmalen

Klassische Bild-
verarbeitung

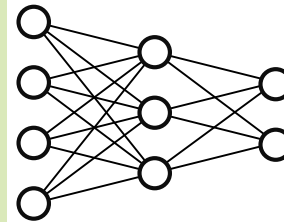


Merkmalsextraktion
(SIFT, SURF, LBP, HOG, etc.)

(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

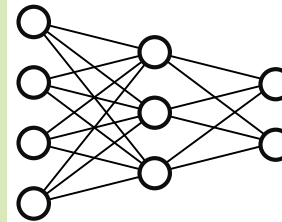


Merkmalsextraktion
(SIFT, SURF, LBP, HOG, etc.)

(0.2, 0.4, ...)

(0.4, 0.3, ...)

Klassifikation
(SVM, Neuronales Netz, etc.)



Containerschiff

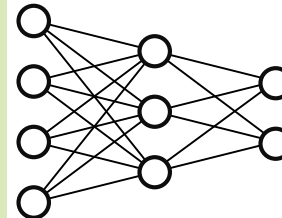
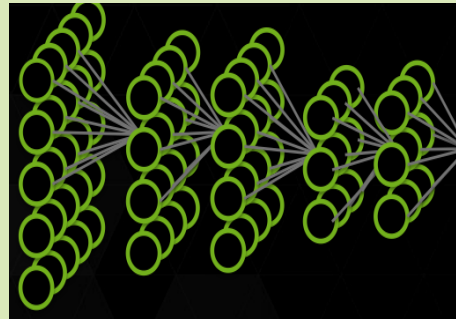
Tiger

...

Mit Convolutional
Neural Networks
(CNNs)



Nimmt rohe Pixel entgegen,
Merkmale werden mitgelernt!



Containerschiff

Tiger

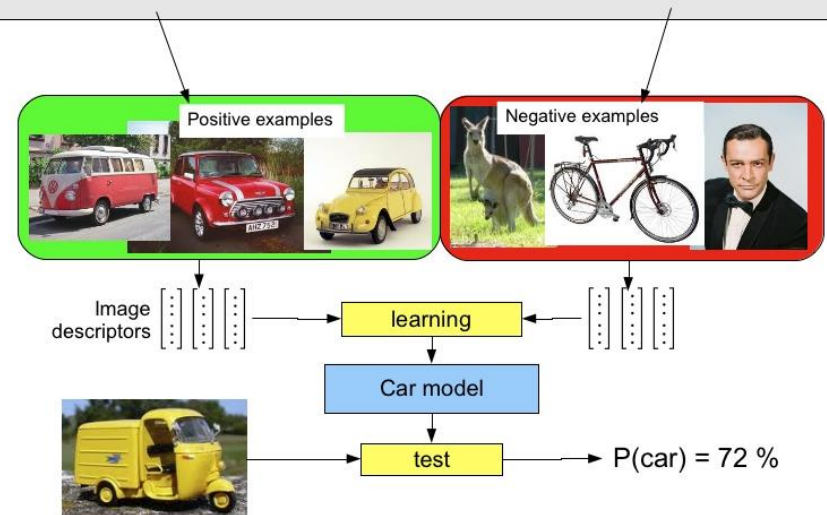
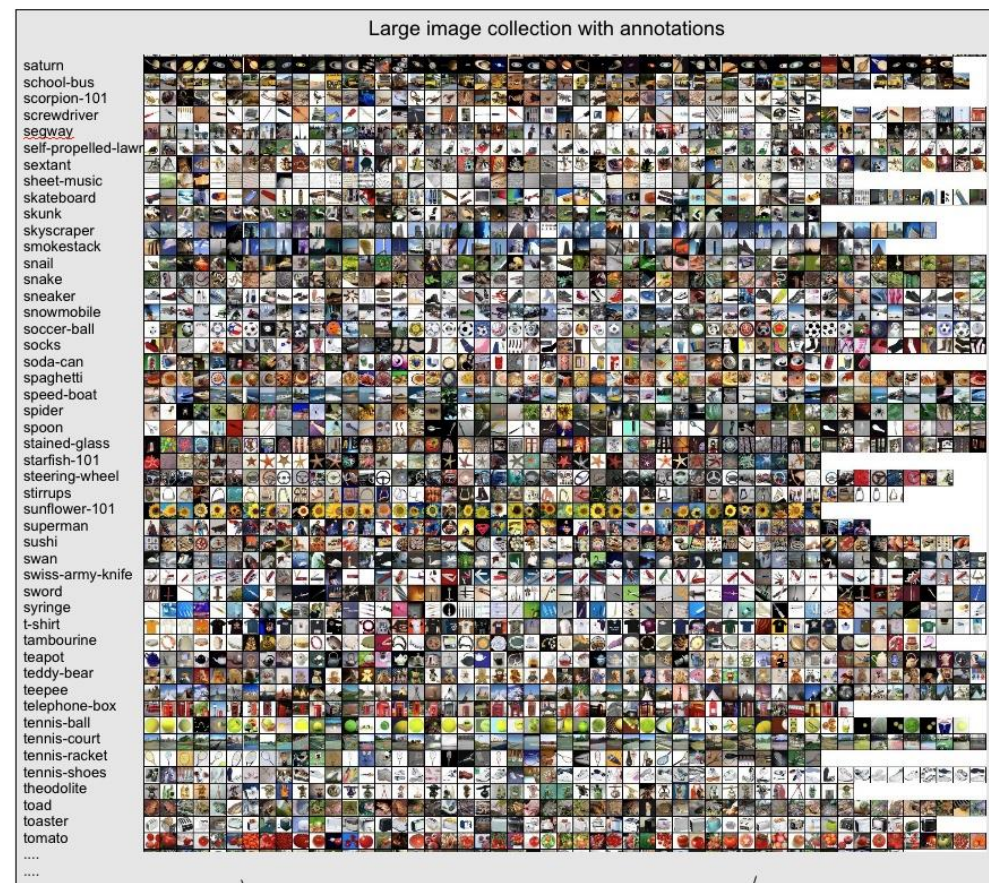
...

Grundlage

Induktives überwachtetes Lernen

Annahme

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



Grundlage

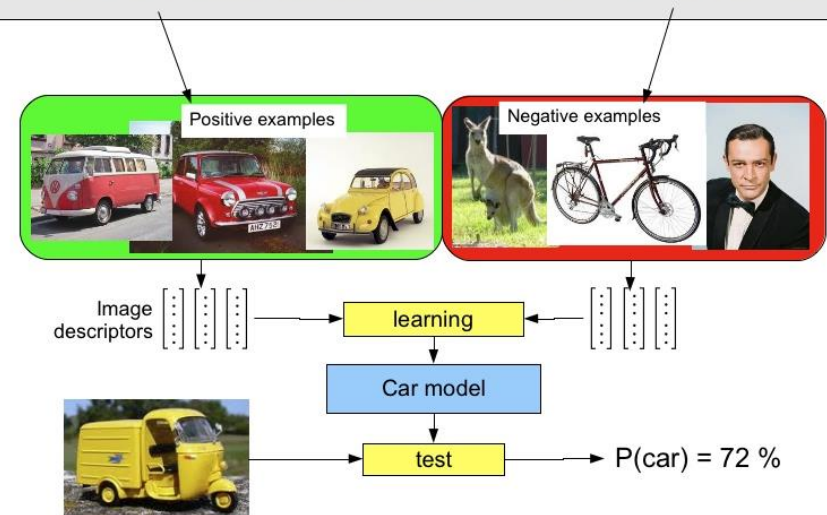
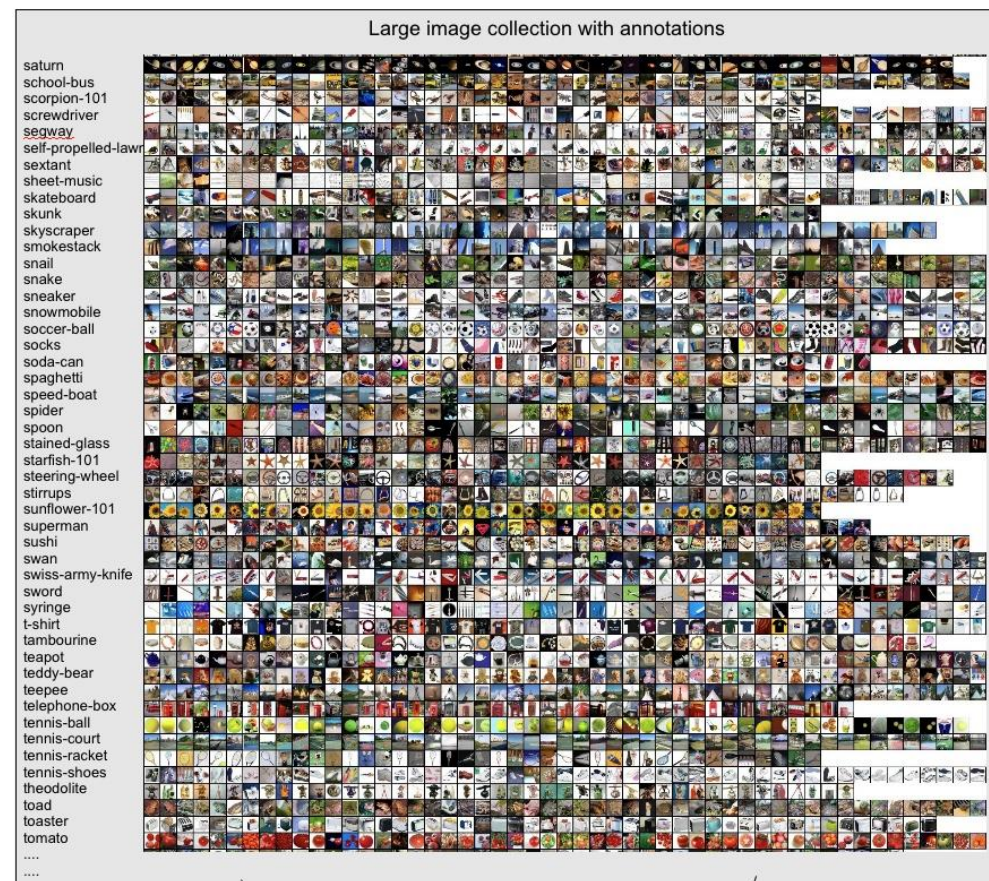
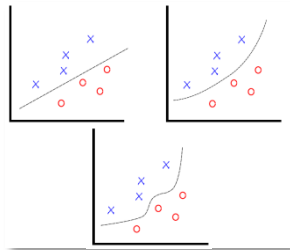
Induktives überwachtetes Lernen

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



Grundlage

Induktives überwachtetes Lernen

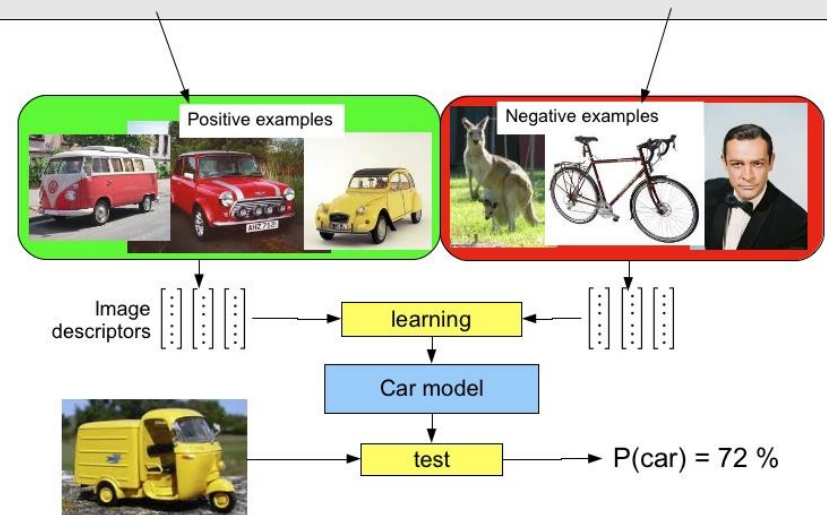
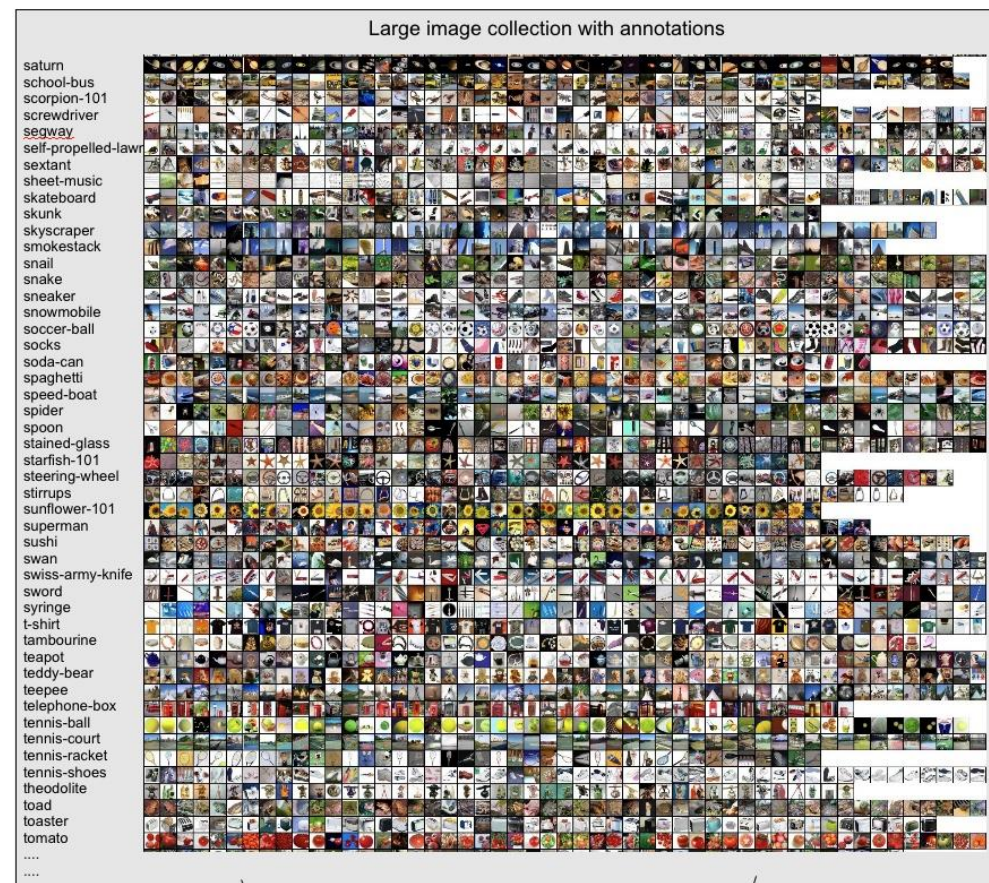
Annahme

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

Methode

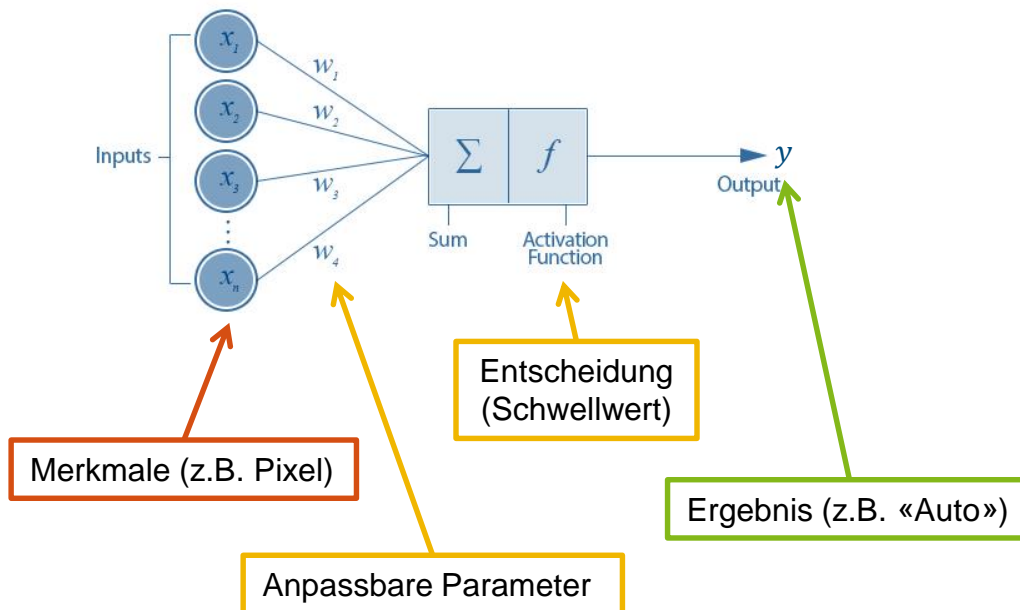
- **Suchen der Parameter einer gegebenen Funktion...**
- ...so dass für alle Beispiele Eingabe (Bild) auf Ausgabe («Auto») abgebildet wird

$$f(x) = y$$

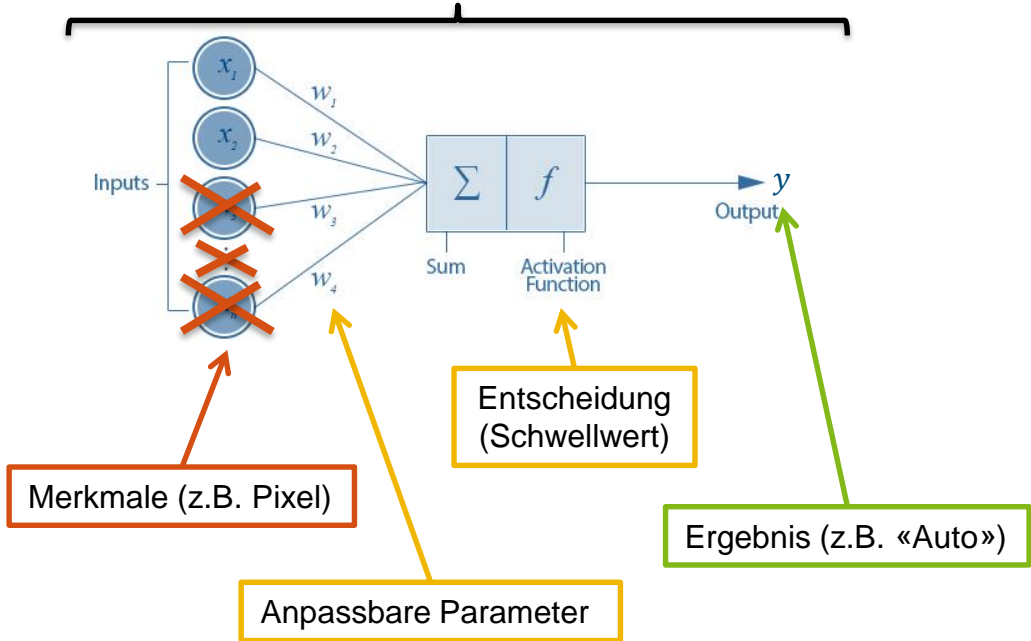
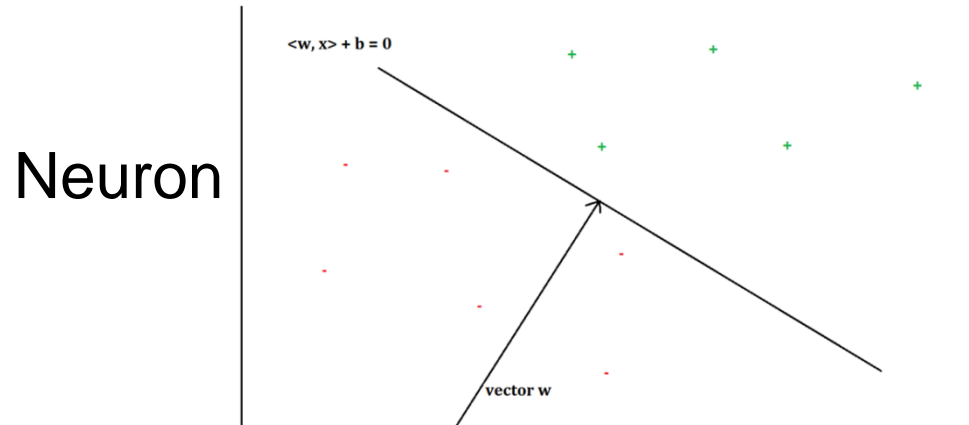


Suche der Parameter *einer Funktion*?

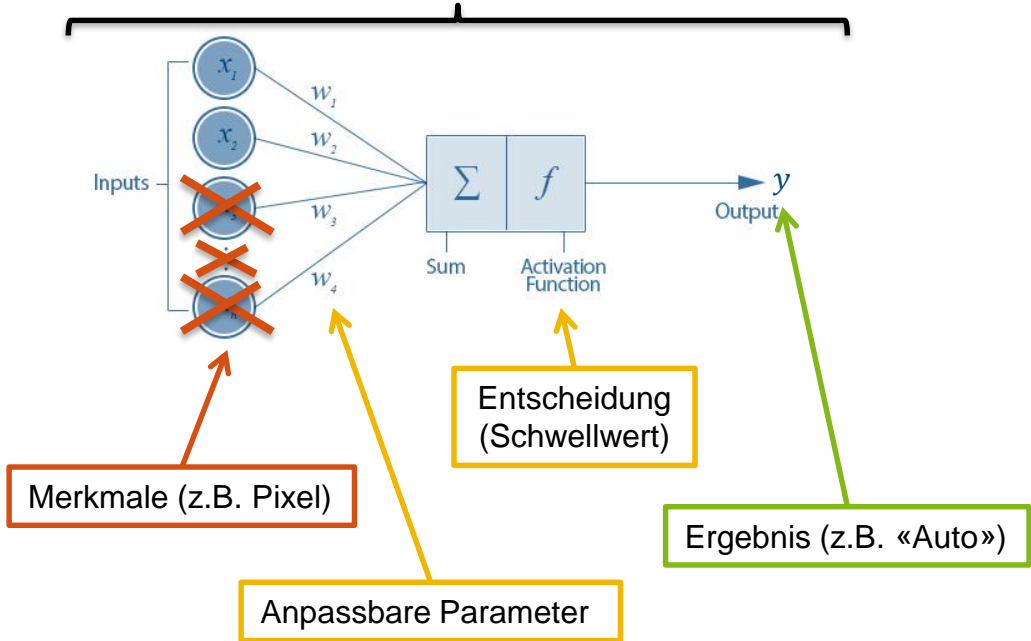
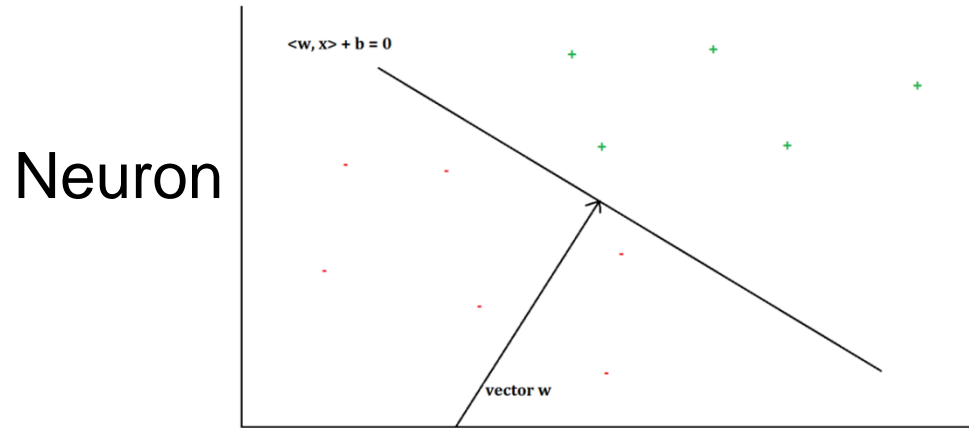
Neuron



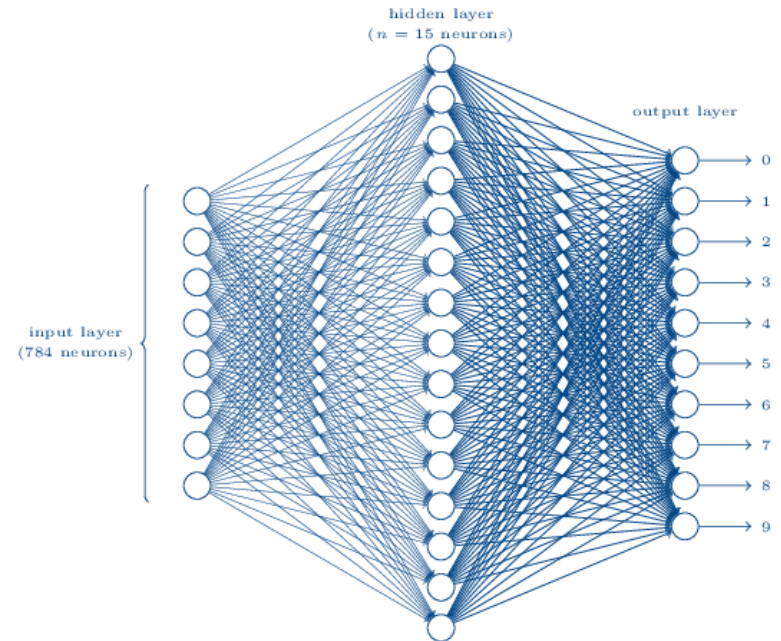
Suche der Parameter *einer Funktion*?



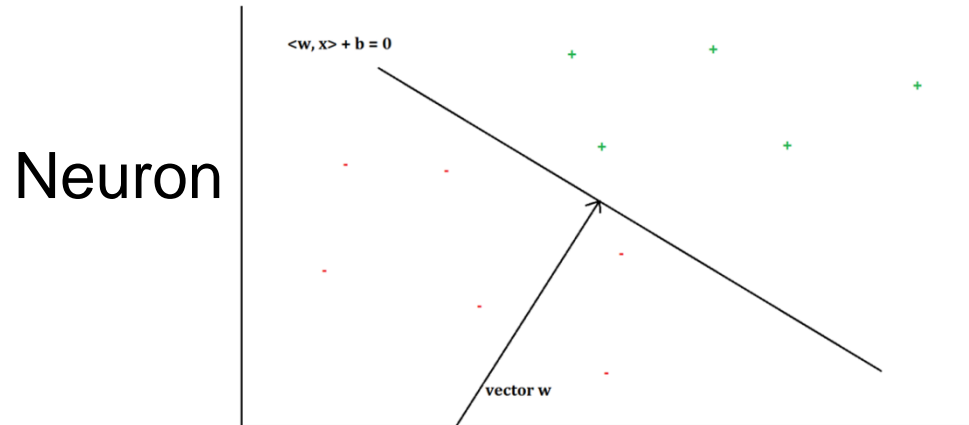
Suche der Parameter *einer Funktion?*



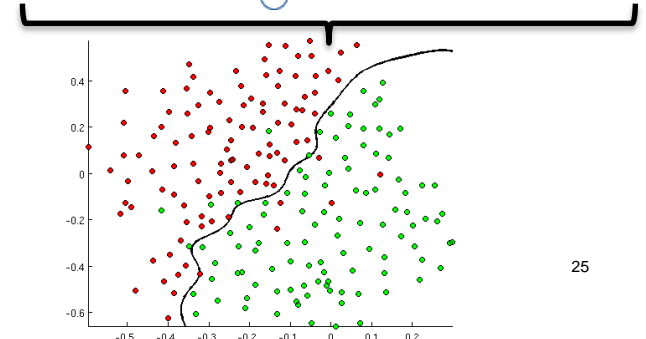
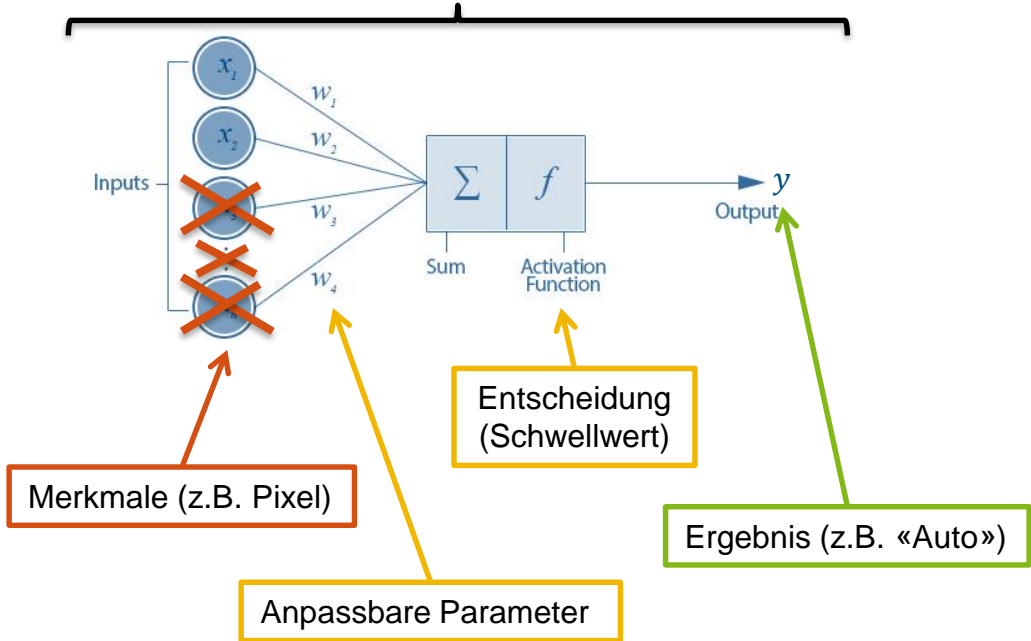
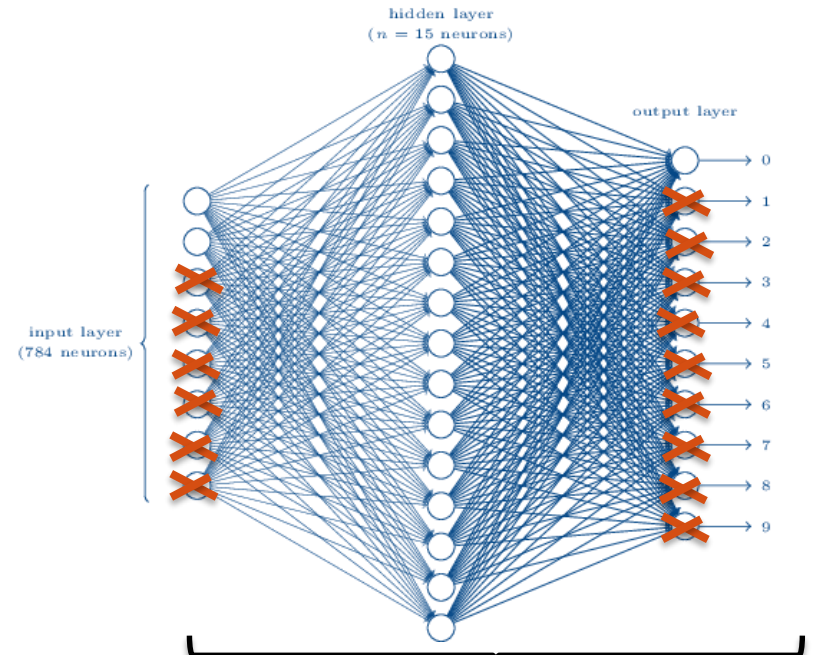
Neuronales Netz



Suche der Parameter *einer Funktion?*



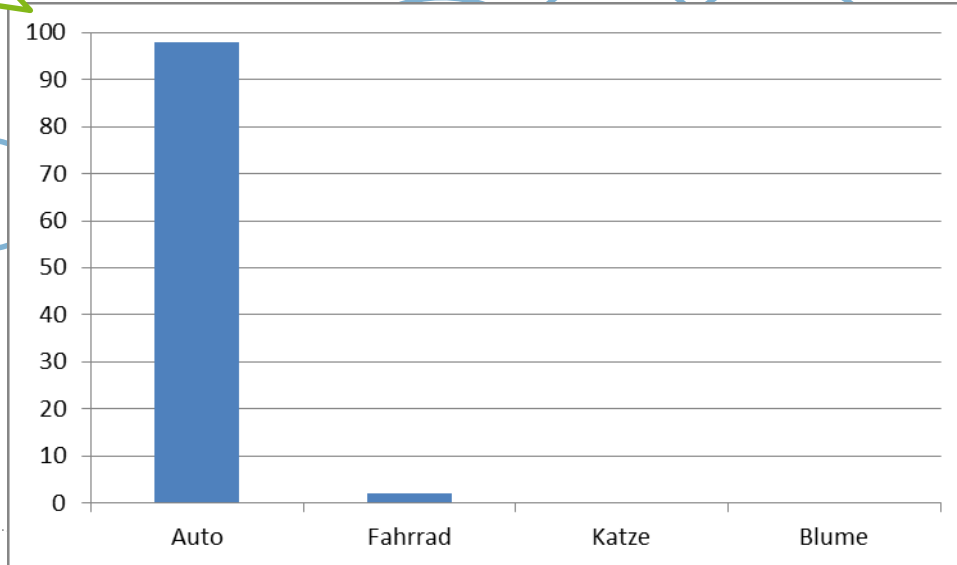
Neuronales Netz



Suche der Parameter einer Funktion?

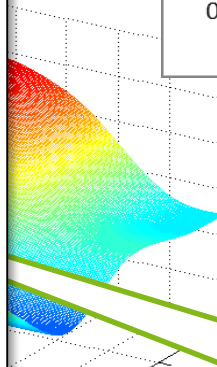
Wahrscheinlichkeit [%] für bestimmtes Ergebnis

- Unser Neuronales Netz: $f_W(x) = y$
mit Bild x , echtem Resultat y und Parametern W
($W = \{w_1, w_2, \dots\}$ anfangs zufällig gewählt)
- Fehlermass: $l(W) = \frac{1}{N} \sum_{i=1}^N (f_W(x_i) - y_i)^2$
Durchschnitt der quadratischen Abweichungen
über alle Bilder (Loss)



$$l(W) = \frac{1}{N} \sum_{i=1}^N (f_W(x_i) - y_i)^2$$

↙ Durchschnitt (über alle Beispiele)
↘ Differenz IST – SOLL (Fehler)
↓ Bestraft grosse Fehler überproportional stärker

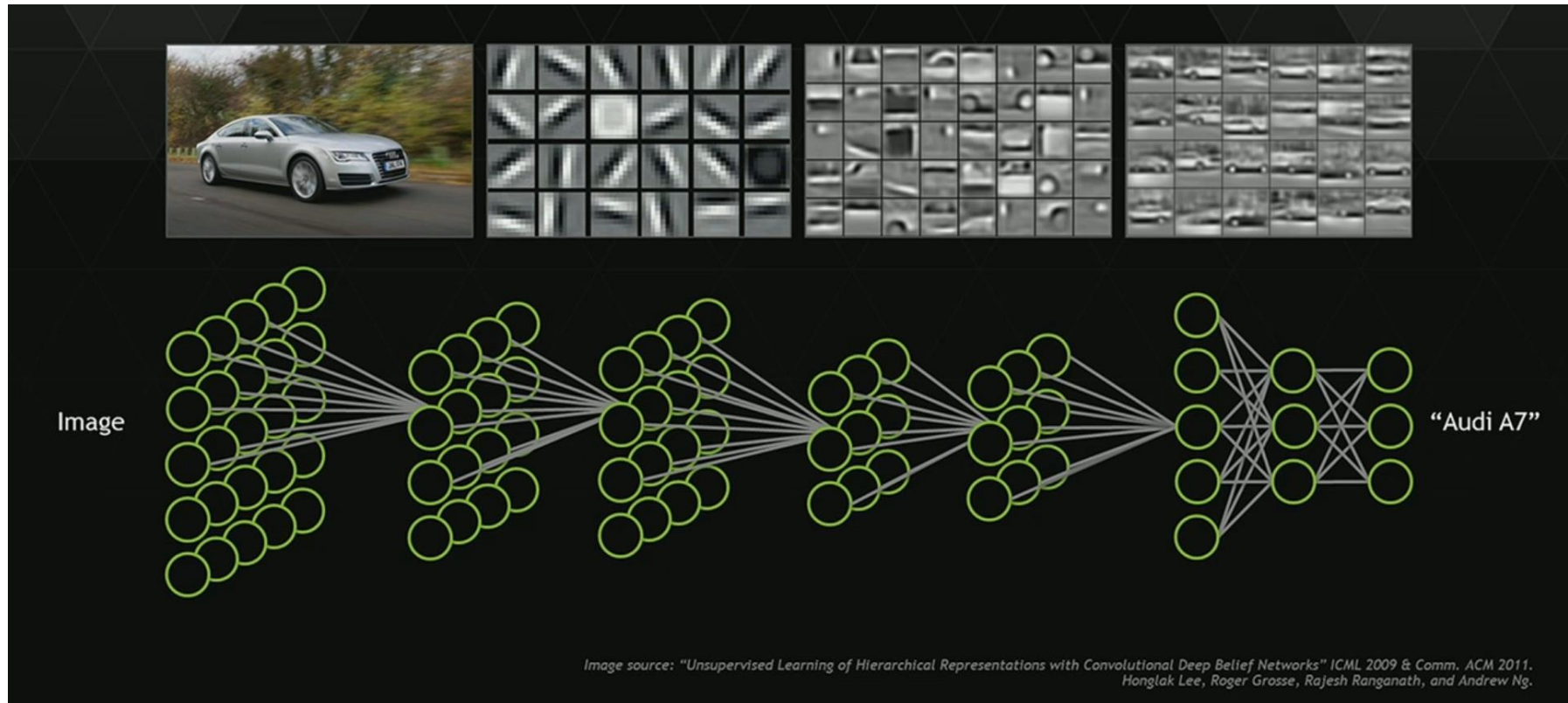


← Fehlerlandschaft

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/>.

Was «sieht» das Neuronale Netz?

Hierarchien komplexer werdender Merkmale

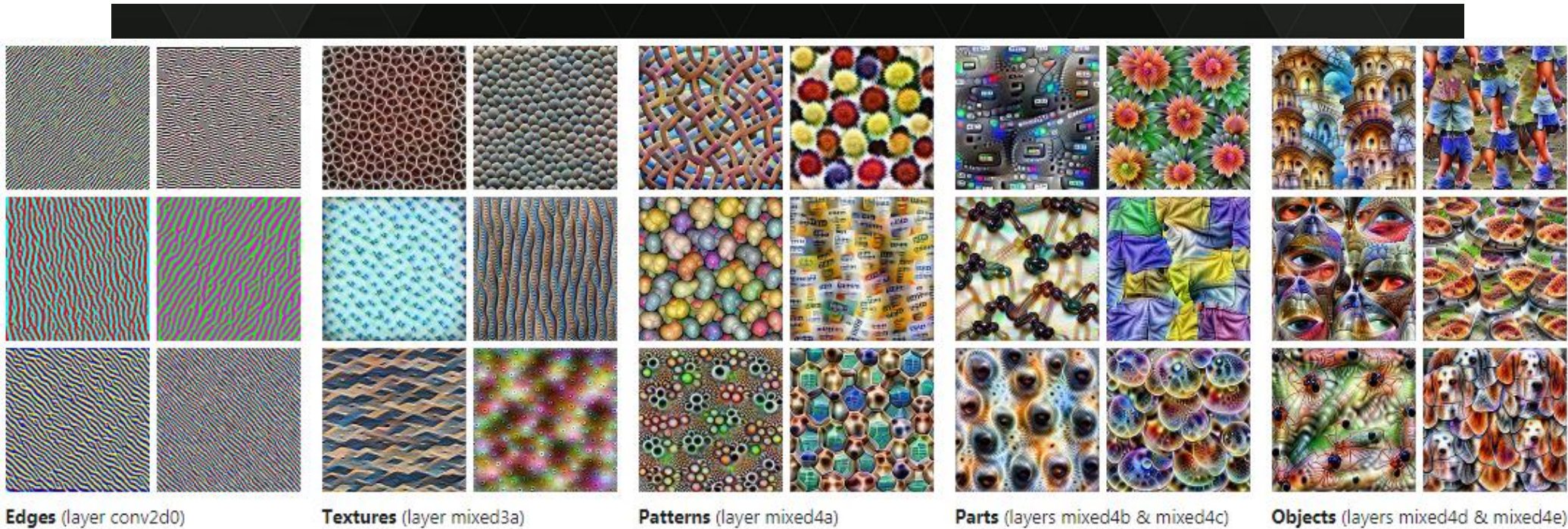


Image source: "Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks" ICML 2009 & Comm. ACM 2011.
Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Ng.

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?

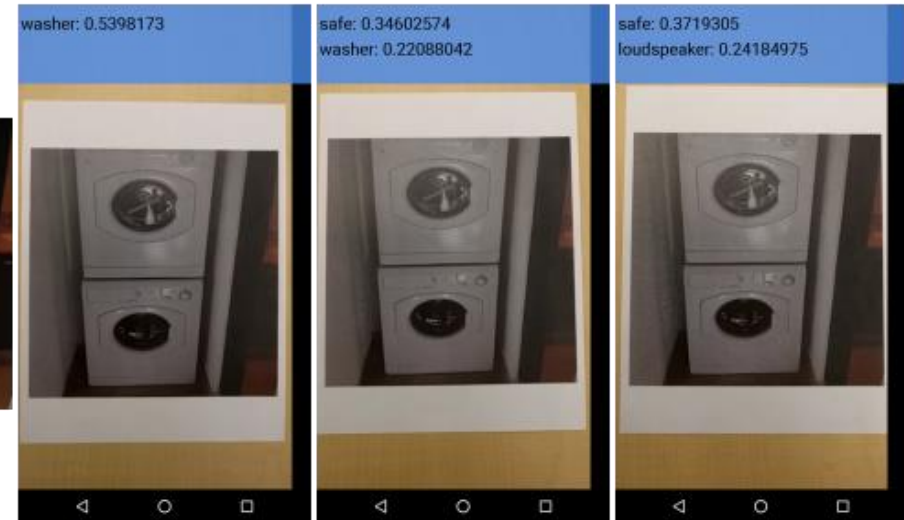
«Debugging» für Einblicke in die vermeintliche «Black Box»

Verdeutlichen ein Problem:

- Adversarial Examples



(a) Image from dataset



(b) Clean image

(c) Adv. image, $\epsilon = 4$

(d) Adv. image, $\epsilon = 8$

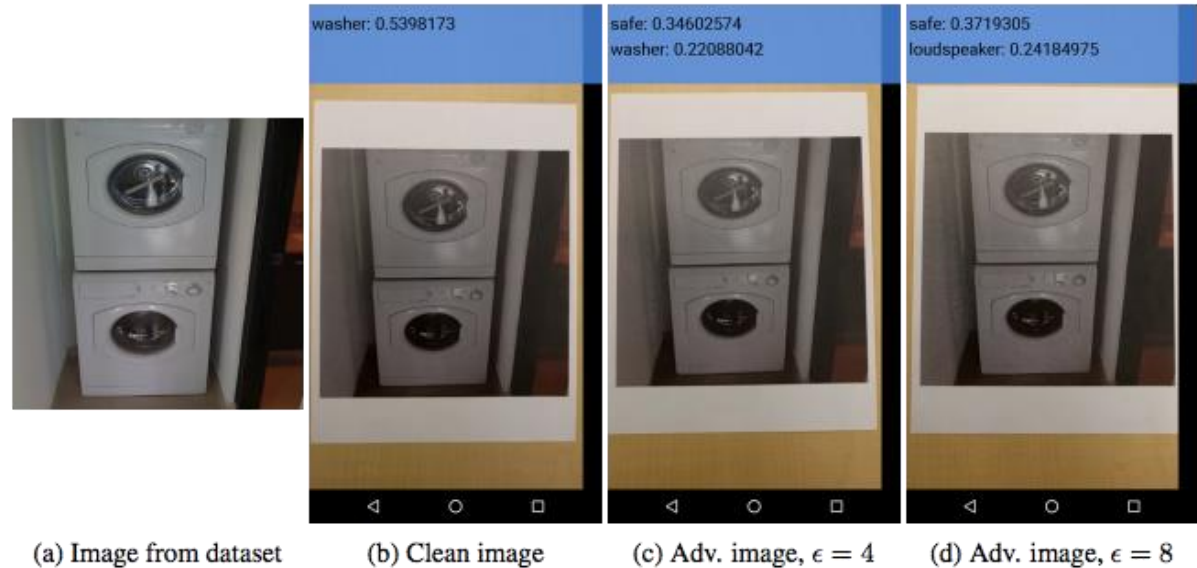
<https://blog.openai.com/adversarial-example-research/>

Wie schlussfolgert die Maschine?

«Debugging» für Einblicke in die vermeintliche «Black Box»

Verdeutlichen ein Problem:

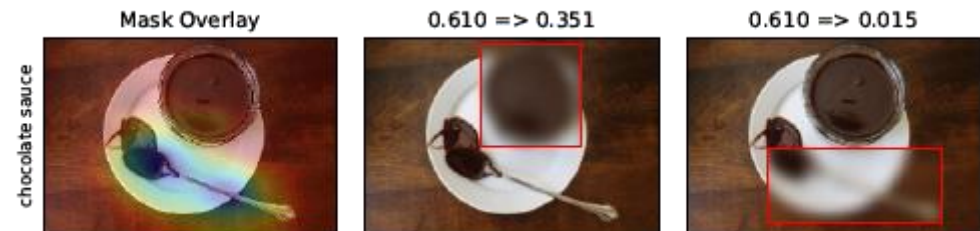
- Adversarial Examples



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

Bieten eine Lösung:







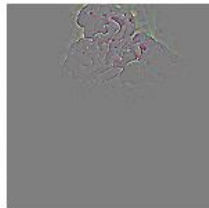
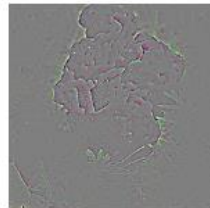
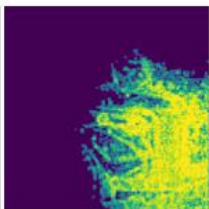
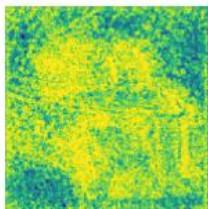
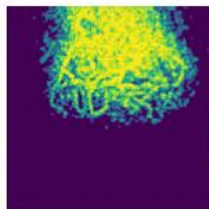
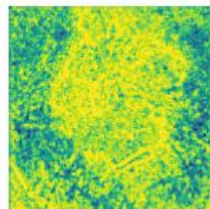
- Saliency Maps



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

Adversarial attacks erkennen

...mittels Local Spatial Entropy der Feature Responses

	Original	Adversarial	Original	Adversarial
Image:				
Feature response:				
Local spatial entropy:				

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!*



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

Lessons learned – model interpretability

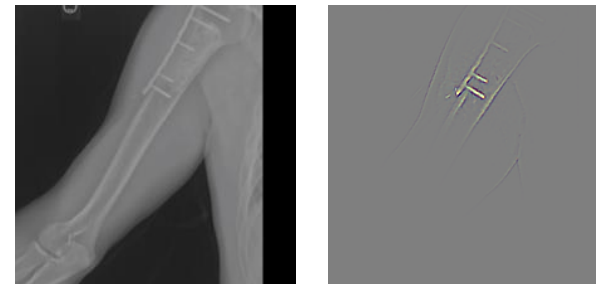
Interpretability is required.

- 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, Duivesteyn, 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

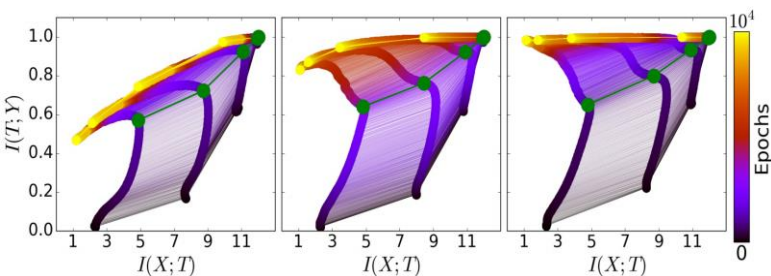
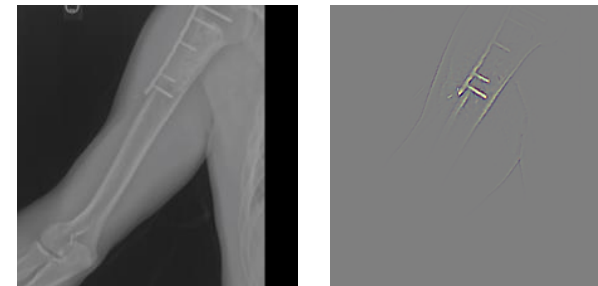
Interpretability is required.

- 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, Duivesteyn, 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

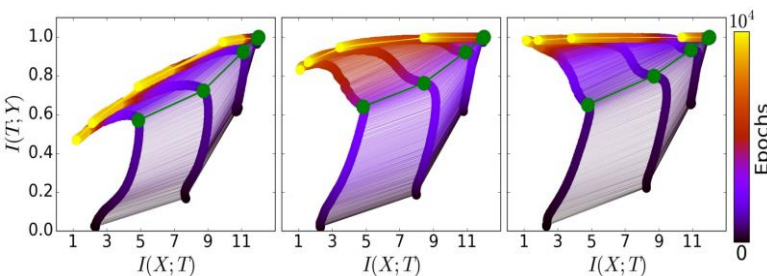
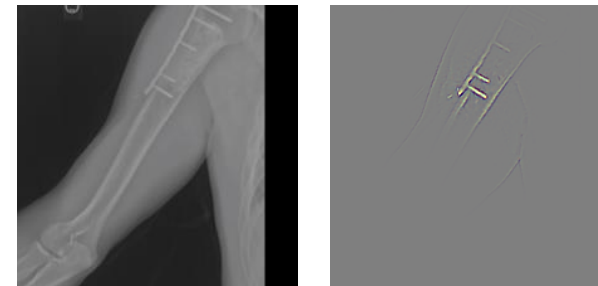
Interpretability is required.

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

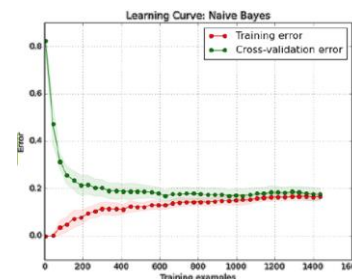
negative X-ray



positive X-ray



DNN training on the Information Plane



a learning curve

Stadelmann, Amirian, Arabaci, Arnold, Duivesteyn, 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

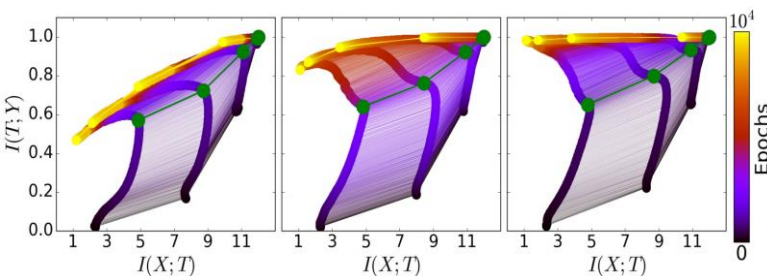
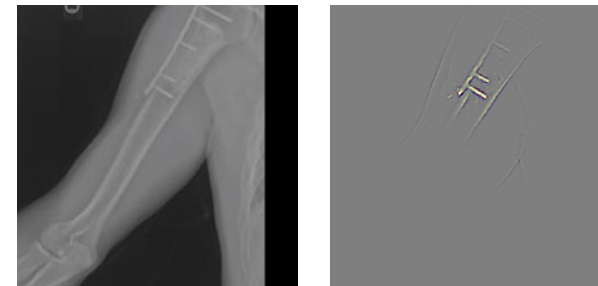
Interpretability is required.

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

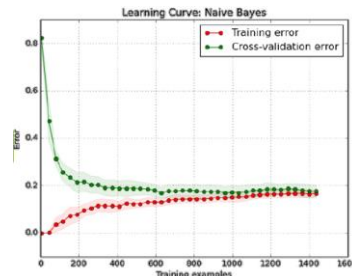
negative X-ray



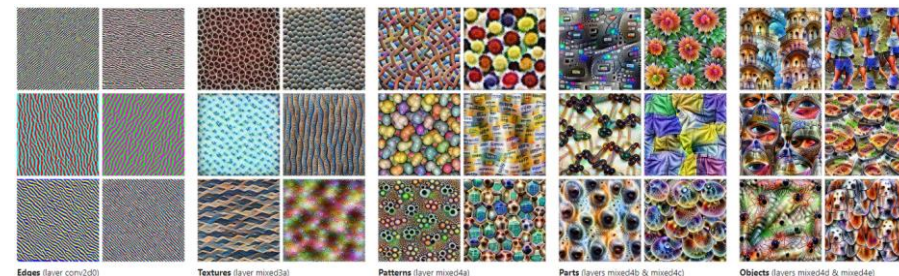
positive X-ray



DNN training on the Information Plane



a learning curve



feature visualization

Stadelmann, Amirian, Arabaci, Arnold, Duivesteyn, 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/>