Ist Neuron-Coverage ein sinnvolles Maß für das Testen tiefer neuronaler Netze?

Zeynep-Kadriye Epping, Consultant (QA)

Gummersbach, 16.02.2024

S&N INVENT

Value Driven Performance.

Agenda

1. Kurze Einführung - Neuronale Netze
2. Was ist die Neuron-Coverage?
3. Experimente von Harel-Canada et al.
4. Eigenen Experimente
5. Zusammenfassung
6. Ausblick

Steckbrief

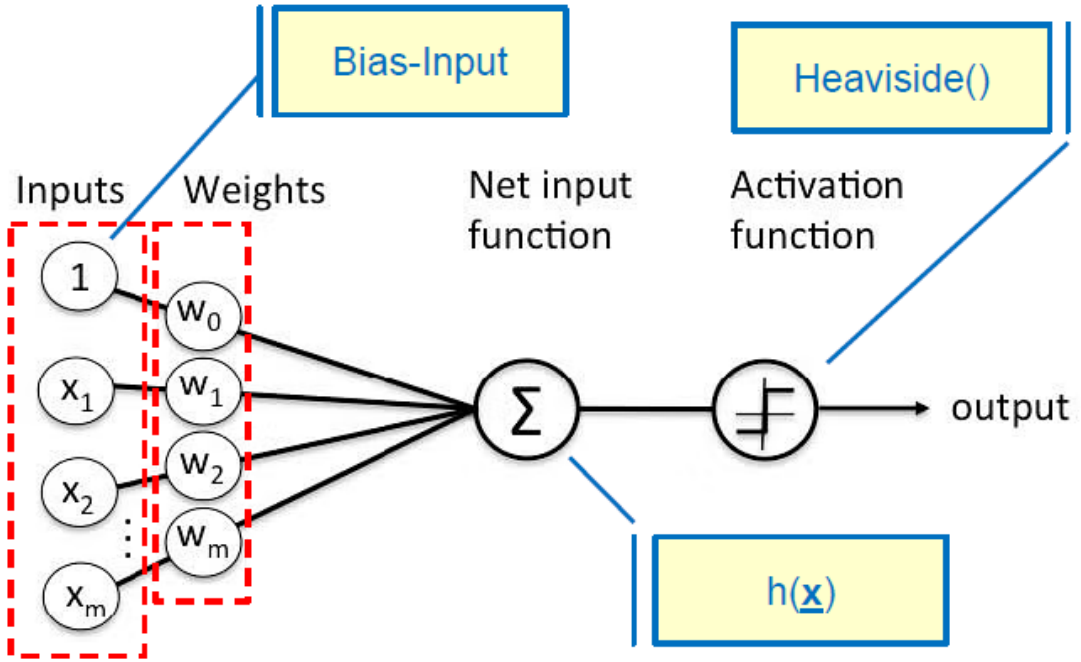
- Zeynep-Kadriye Epping
- B.Sc. WI an der FHDW in Paderborn
- Seit 2017 Testerin in der QA bei S&N Invent GmbH
- Berufsbegleitend M.Sc. WI im Verbundstudium an der FH Dortmund

Agenda

1. Kurze Einführung - Neuronale Netze
2. Was ist die Neuron-Coverage?
3. Experimente von Harel-Canada et al.
4. Eigenen Experimente
5. Zusammenfassung
6. Ausblick

Kurze Einführung – Neuronale Netze (Einlagiges Perzeptron)

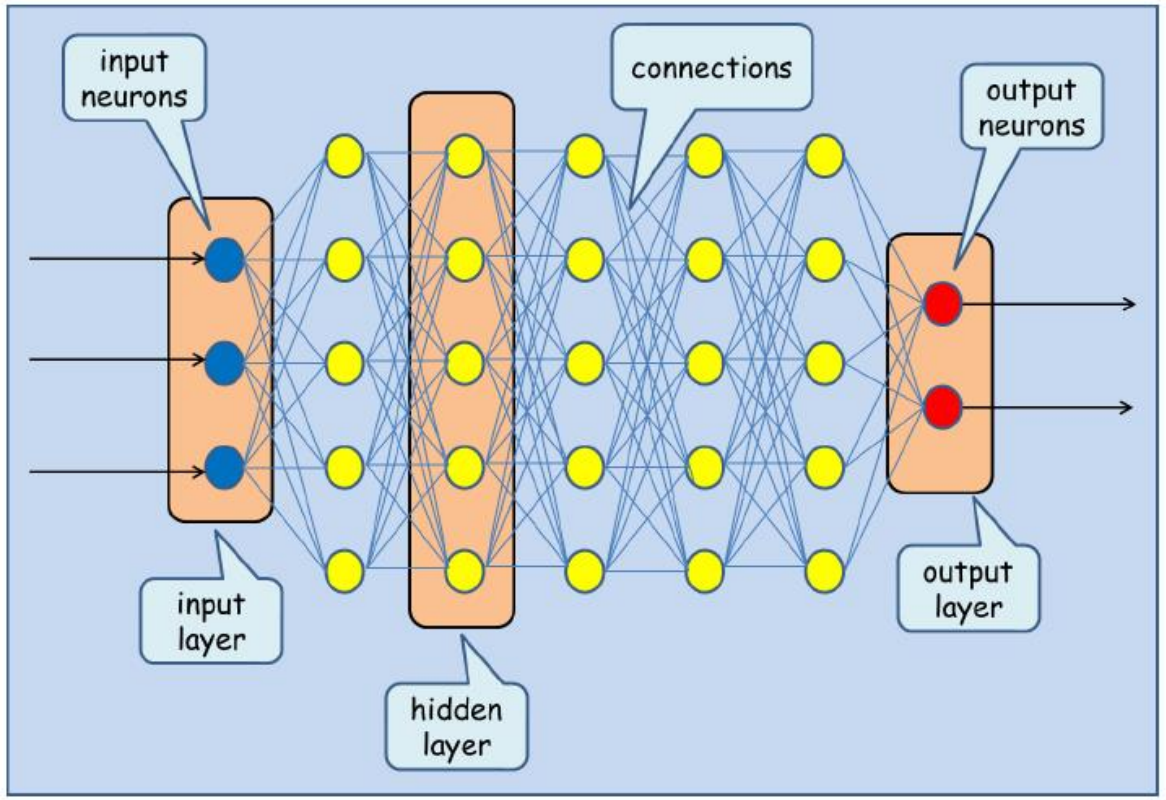
- Mathematische Funktion basierend auf dem Modell eines biologischen Neurons
- Eingangsvektor $\underline{x}$, Vorhersage $y = f(\underline{x})$
- Gewichtsvektor $\underline{w}$
- Netz Eingangsfunktion
  $h(\boldsymbol{x}) = \sum_{i=1}^{m} x_i w_i$
- Aktivierungsfunktion: Heaviside
- $\text{Heaviside}(h(\underline{x})) = \begin{cases} 1, wenn\ h(x) > 0 \\ sonst\ 0 \end{cases}$



Quelle: Mario Winter

Kurze Einführung – Neuronale Netze (Mehrlagiges Perzeptron)

- Netzwerk aus mehreren Schichten
- Eingangsschicht mit Eingangsneuronen
- Hidden Layer
- Verbindungen zwischen den Neuronen, welche Gewichte enthalten
- Ausgangsschicht mit Ausgangsneuronen

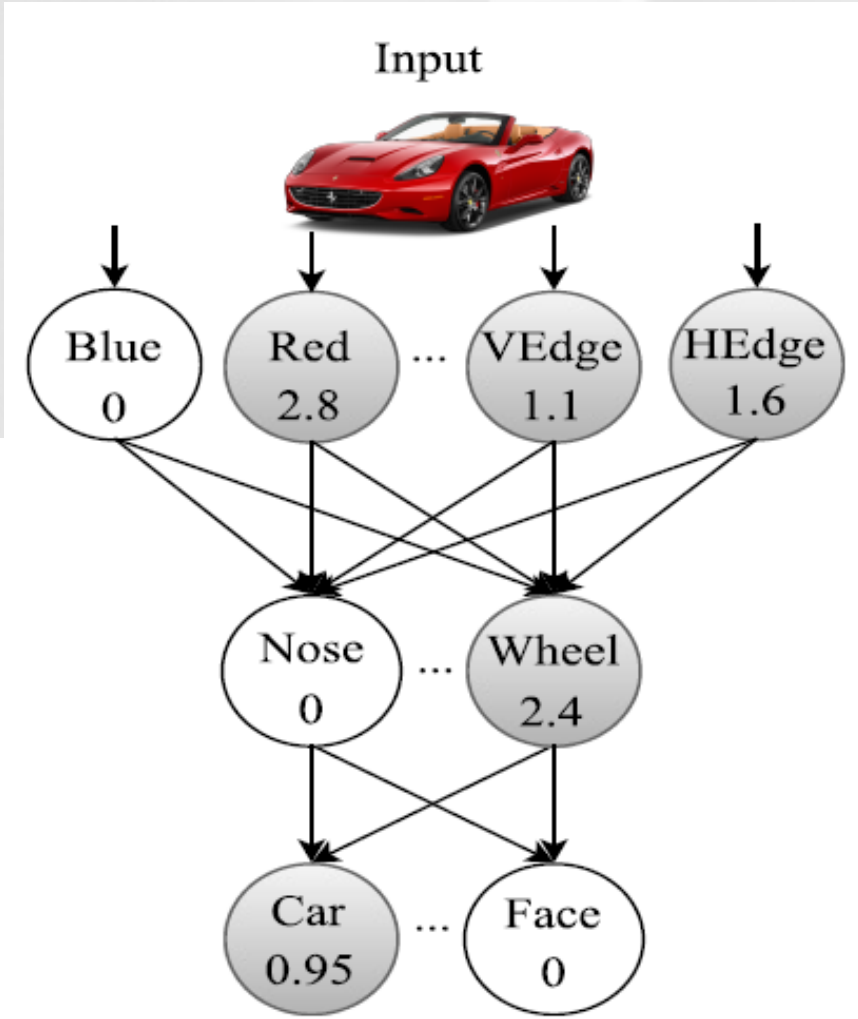


Quelle: Mario Winter

Agenda

1. Kurze Einführung - Neuronale Netze
2. Was ist die Neuron-Coverage?
3. Experimente von Harel-Canada et al.
4. Eigenen Experimente
5. Zusammenfassung
6. Ausblick

Was ist die Neuron-Coverage?



Quelle: Pei et al.

- Abgeleitet aus der klassischen Softwareentwicklung Code-Coverage (Testabdeckung)
- Schwellenwert t = 0
- Insgesamt 8 Neuronen im DNN
- 5 Neuronen überscheiten Schwellenwert
- NC = 5/8 = 0,625

Was ist die Neuron-Coverage (Neuronenabdeckung)?

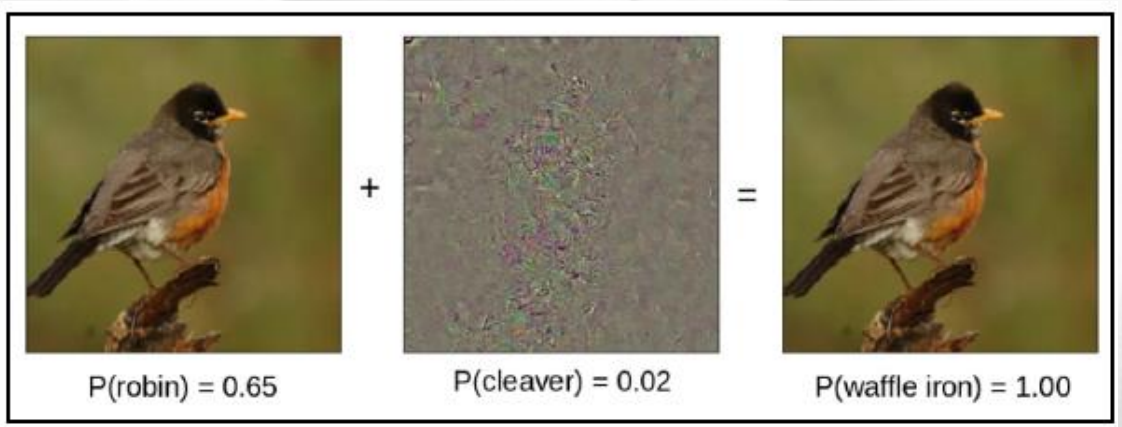
- $NCov(T,\ \boldsymbol{x}) = \frac{|\{n|\forall \boldsymbol{x} \in T, out(n, \boldsymbol{x}) > t\}|}{|N|}$ (Definition nach Kexin Pei et al.)
  - N repräsentiert alle Neuronen im DNN
  - T repräsentiert alle Test Eingaben
  - out(x,n) ist eine Funktion, welche den Outputwert des Neuron n im DNN für Testeingabe x zurückliefert
  - t repräsentiert den Schwellenwert, ab dem ein Neuron als aktiviert gilt

Agenda

1. Kurze Einführung - Neuronale Netze
2. Was ist die Neuron-Coverage?
3. **Experimente von Harel-Canada et al.**
4. Eigenen Experimente
5. Zusammenfassung
6. Ausblick

Experimente von Harel-Canada et al.

- Implizite Annahme: Erhöhung der NC bringt eine Verbesserung der Qualität von Testsuiten hervor
- Experimentaufbau
  - 3 Datensätze (MNIST, CIFAR-10 und Udacity Self-Driving Car)
  - 8 vortrainierte neuronale Netze (FCNet5, FCNet10, Conv1DNet, Conv2DNet, ResNET56, DenseNet121, DAVE2 und DAVE2-Norminit)
  - 2 feindliche Angriffe (Carlini-Wagner (CW) und Projected-Gradient-Descent (PGD))



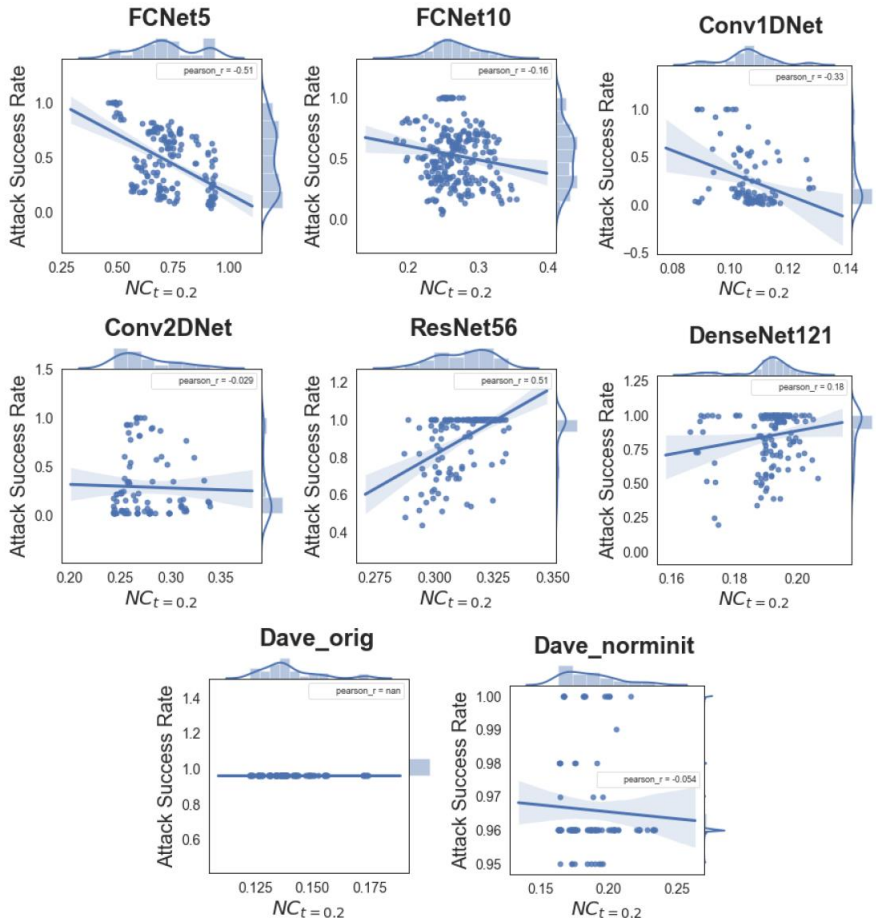
Experimente von Harel-Canada et al.

- Feindlichen Angriffe erweitert um eine neue Regularisierung, welche die neuronale Vielfalt fördert
- Testsuiten bewertet unter den Punkten:
  - Fehlerentdeckungsrate
  - Natürlichkeit der Testeingaben
  - Unparteilichkeit der Ausgaben

Experimente von Harel-Canada et al. (Ergebnisse)

- Fehlerentdeckungsrate
  - Begriffe Angriffserfolgsrate (Attack Success Rate – ASR) und Fehlerentdeckungsrate (Defect Detection Rate DDR) werden gleichgestzt und gleichwertig verwendet
  - Zur Ermittlung wird Klassifizierungsgenauigkeit pert_acc verwendet
  - Berechung ASR(T) = 1 – pert_acc

→ NC sehr schwankend; korreliert nicht deutlich mit Fehlerentdeckungsrate

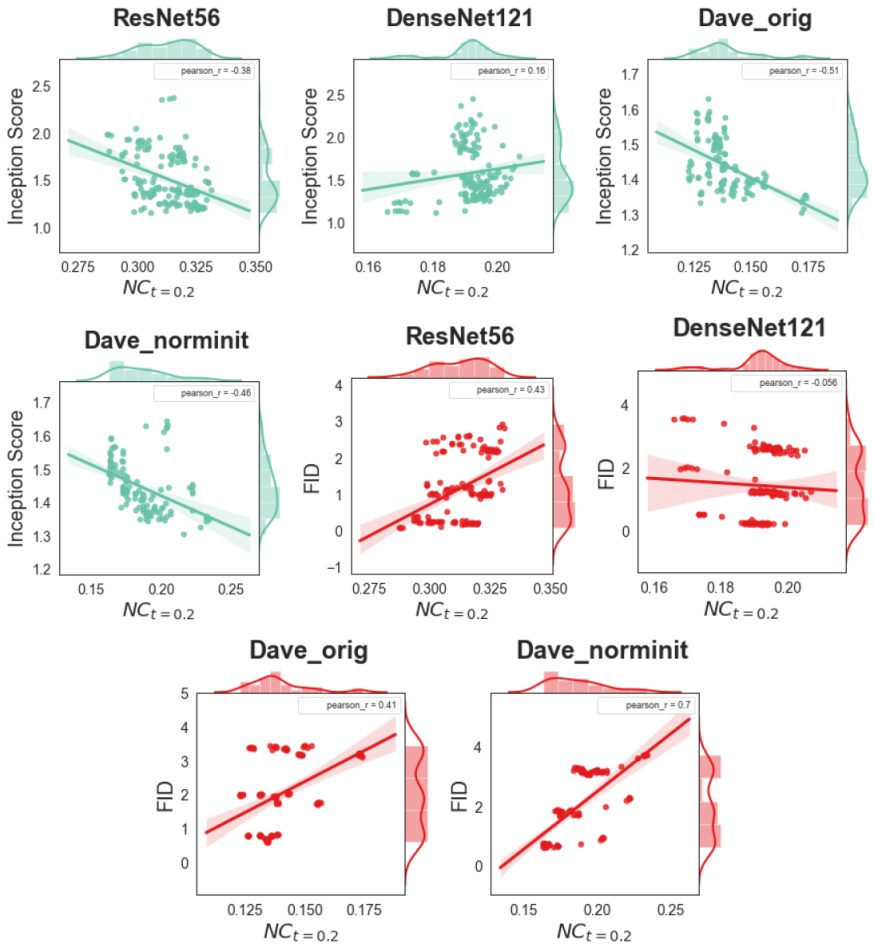


Quelle: Heral-Canada et al.

Experimente von Harel-Canada et al. (Ergebnisse)

- Natürlichkeit
  - Messung der Natürlichkeit mit Hilfe Inception Score (IS) und Fréchet Inception Distance (FID)
  - Je höher IS, umso realistischer und wiedererkennbarer ist das Bild
  - Je niedriger FID, desto ähnlicher ist manipulierte Bild zum Originalbild

→ Maximierung der NC untergräbt mit hoher Wahrscheinlichkeit die Natürlichkeit der Testeingaben

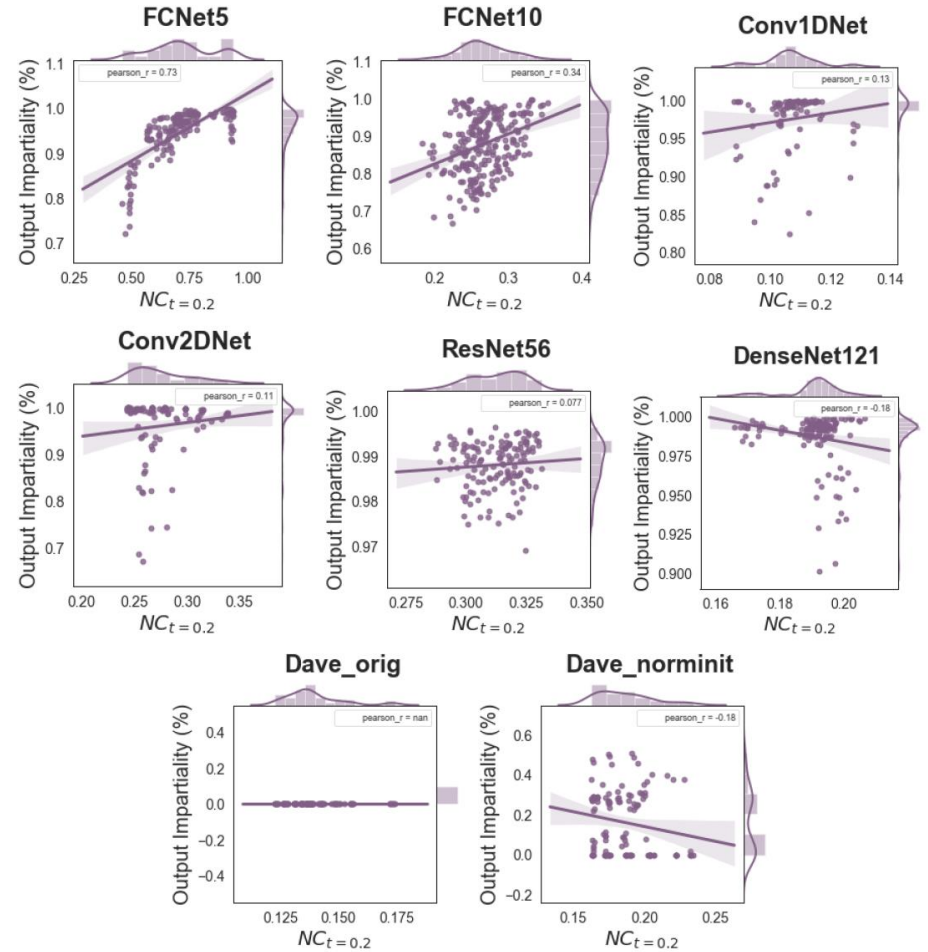


Quelle: Heral-Canada et al.

Experimente von Harel-Canada et al. (Ergebnisse)

- Unparteilichkeit der Ausgabe
  - Idee stammt aus der traditionellen Softwarentwicklung
  - Unparteilichkeit wird definiert als Maß für die Verteilung der Klassenvorhersagen unter einer einheitlichen Verteilung der Eingaben
  - Nutzung von Pielous Gleichmäßigkeitskennzahl
  - Hohe Gleichmäßigkeitskennzahl bedeutet hohe Unparteilichkeit der Ausgabe

→ Erhöhung des NC führt zu Verzerrung des Ausgabeverhaltens



Quelle: Heral-Canada et al.

Agenda

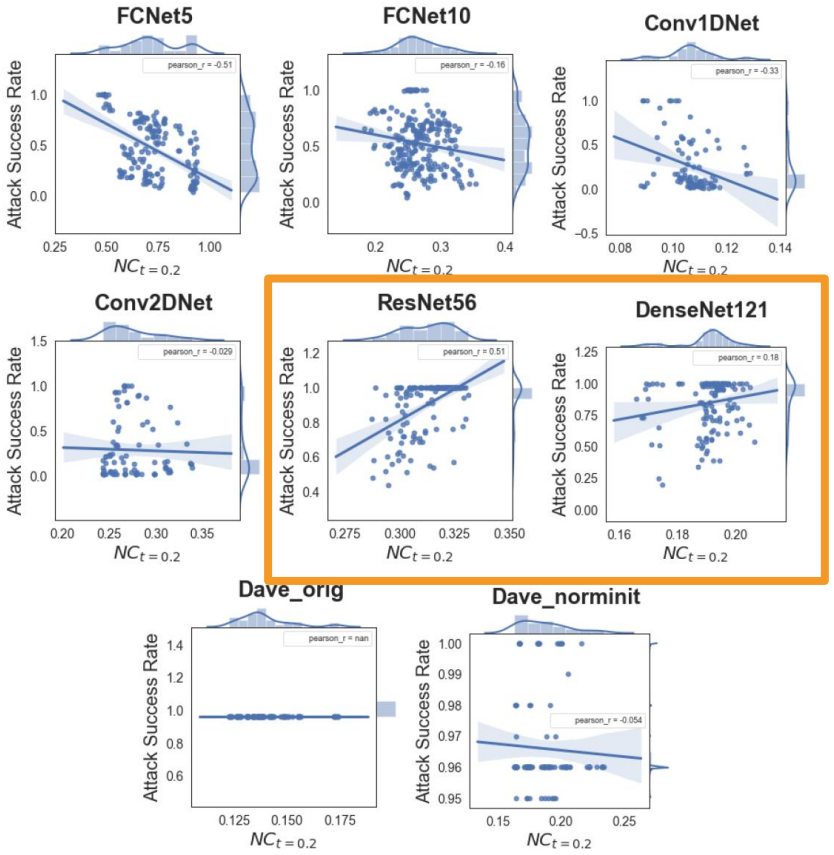
1. Kurze Einführung - Neuronale Netze
2. Was ist die Neuron-Coverage?
3. Experimente von Harel-Canada et al.
4. **Eigenen Experimente**
5. Zusammenfassung
6. Ausblick

Eigenen Experimente

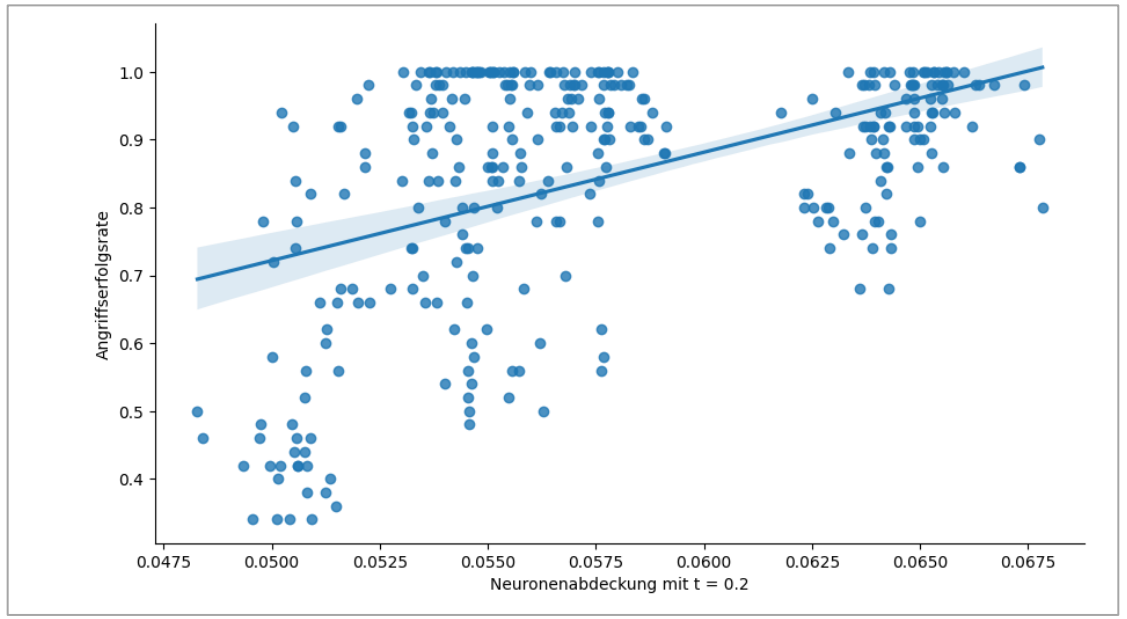
- Anwendung der Experimente auf ConvNeXt-Tiny und ImageNet
  - ConvNext-Tiny (ca. 28,6 Millionen Parameter)
  - Nur ein minimaler Ausschnitt von ImageNet
- Erweiterten CW- und PGD-Angriffe
  - Betrachtung Ergebnisse nur PGD-Angriff
  - Nicht ausreichend Ressourcen für den CW-Angriff
- Gleiche Parameter aus dem Artikel von Harel-Canada et al. für die Regularisierung der feindlichen Angriffe genutzt

Eigenen Experimente (Beispiele erzeugter feindlicher Angriffe)

Eigenen Experimente (Ergebnisse)

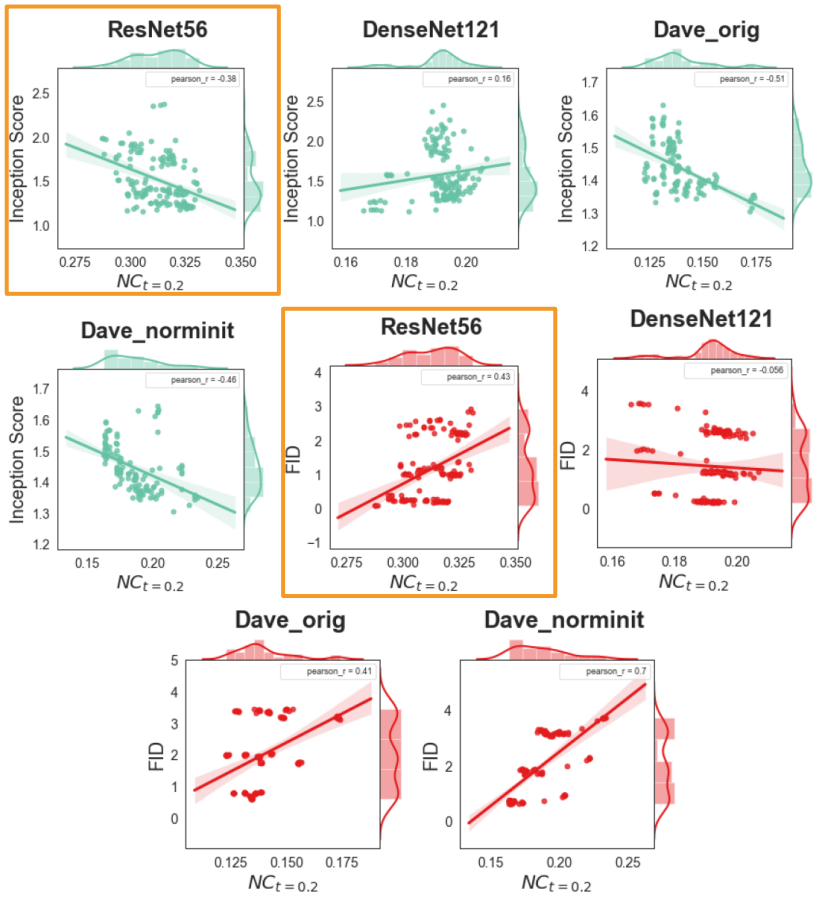


Quelle: Heral-Canada et al.

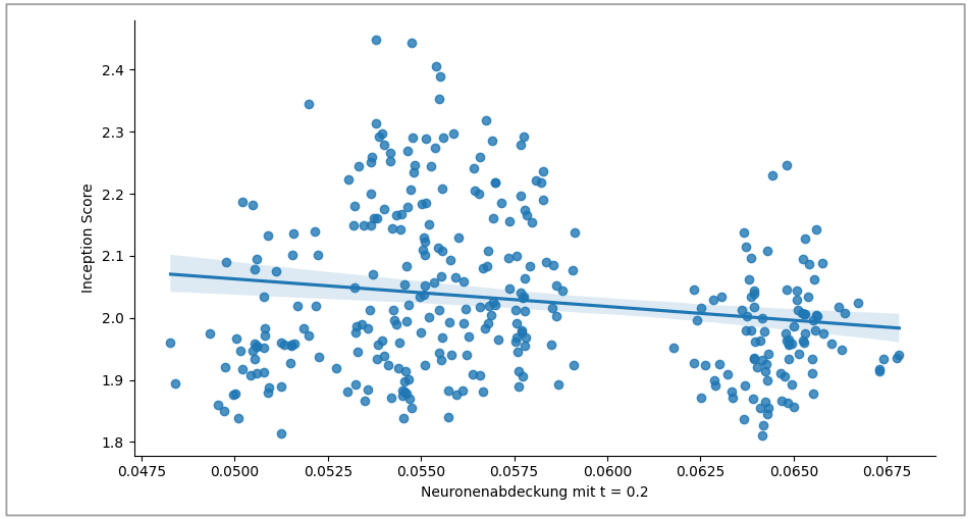


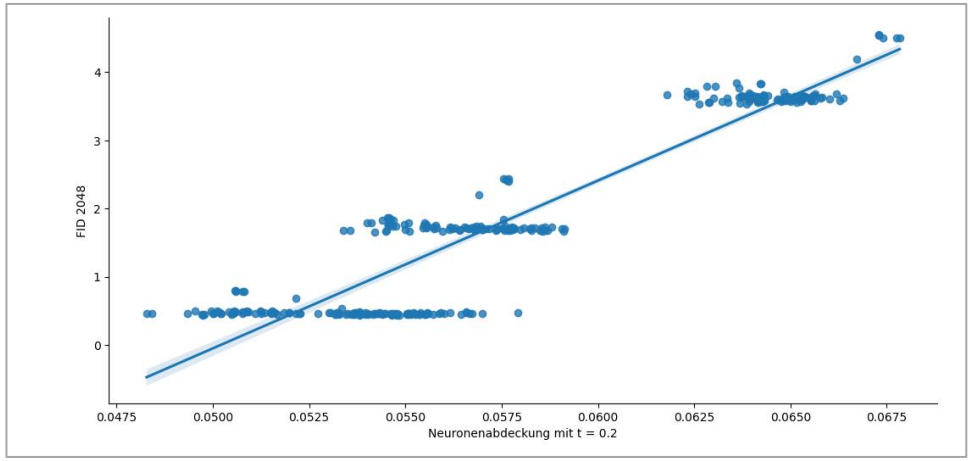
Quelle: Eigene Darstellung

Eigenen Experimente (Ergebnisse)



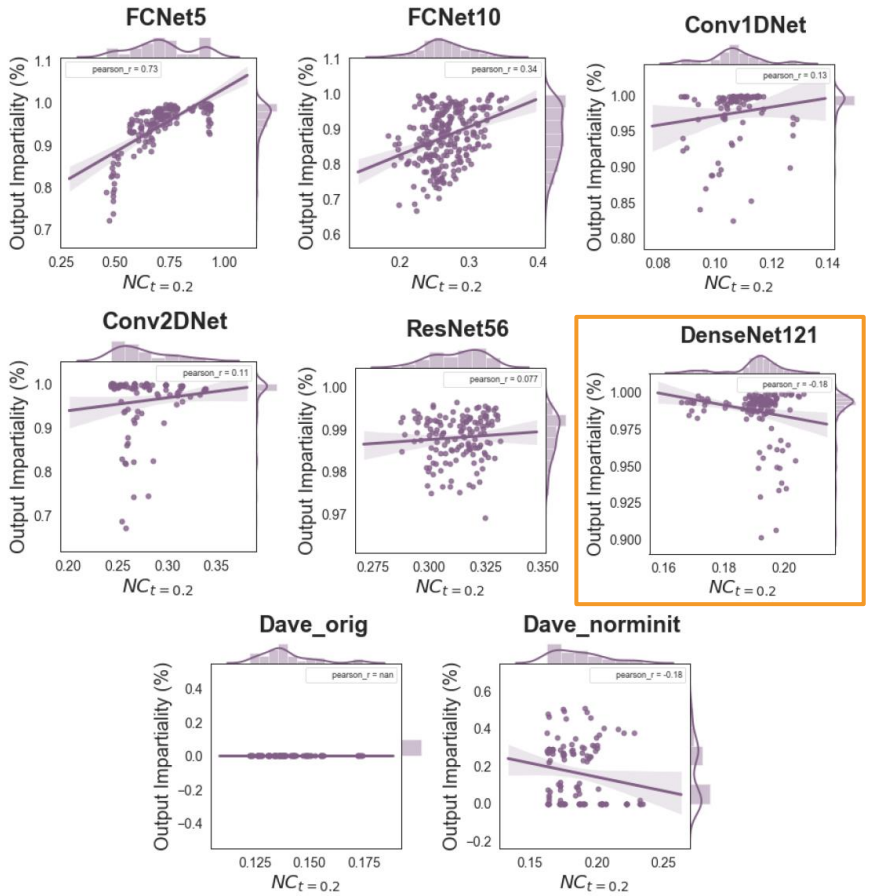
Quelle: Heral-Canada et al.



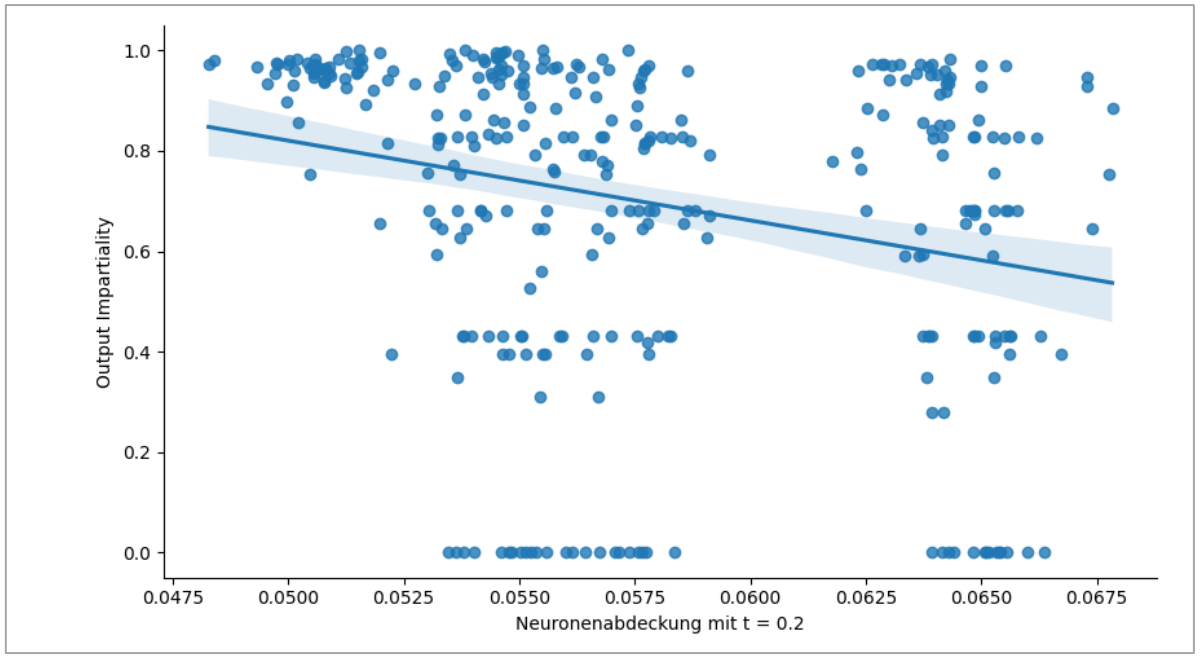


Quelle: Eigene Darstellung

Eigenen Experimente (Ergebnisse)



Quelle: Heral-Canada et al.



Quelle: Eigene Darstellung

Agenda

1. Kurze Einführung - Neuronale Netze
2. Was ist die Neuron-Coverage?
3. Experimente von Harel-Canada et al.
4. Eigenen Experimente
5. Zusammenfassung
6. Ausblick

Zusammenfassung

- Anzeichen, dass höhere NC nicht deutlich mit Fehlerentdeckungsrate korreliert
- Anzeichen, dass Maximierung der NC die Natürlichkeit der Testeingabe untergräbt
- Anzeichen, dass höherer NC zur Verzerrung des Ausgabeverhaltens führt

- Vermutung: Abhängigkeit von den erzeugten Testdaten
  - 100% Code-Coverage bedeutet nicht, dass der Code 100% fehlerfrei ist

Agenda

1. Kurze Einführung - Neuronale Netze
2. Was ist die Neuron-Coverage?
3. Experimente von Harel-Canada et al.
4. Eigenen Experimente
5. Zusammenfassung
6. Ausblick

Ausblick

- Weitere Artikel, welche die Aussagen von Heral-Canda et al. unterstützen
- Artikel, welche neue Testmethoden vorstellen, auf Grund der Ergebnisse von Heral-Canada et al.
  - RobOT
  - Operationale Adversarial Attacks
  - DialTest
  - u. v. m.
- Artikel, welche neue Testmetriken vorstellen
  - Independence Neuron Coverage
  - Neural Coverage (NLC) – schichtweises und verteilungsorientiertes Kriterium
  - u. v. m.

Fragen und Diskussion

Zeynep-Kadriye Epping
Kadriye.Epping@sn-invent.de
+49 5251 1581 193

Quellen

- Harel-Canada, Fabrice ; Wang, Lingxiao ; Gulzar, Muhammad A. ; Gu, Quanquan ; Kim, Miryung: Is Neuron Coverage a Meaningful Measure for Testing Deep Neural Networks? Erhältlich unter https://dl.acm.org/doi/pdf/10.1145/3368089.3409754
- J. Wang *et al*., "RobOT: Robustness-Oriented Testing for Deep Learning Systems," *2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE)*, Madrid, ES, 2021, pp. 300-311, doi: 10.1109/ICSE43902.2021.00038; Erhältlich unter https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9402039
- X. Zhao, W. Huang, S. Schewe, Y. Dong and X. Huang, "Detecting Operational Adversarial Examples for Reliable Deep Learning," *2021 51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks - Supplemental Volume (DSN-S)*, Taipei, Taiwan, 2021, pp. 5-6, doi: 10.1109/DSN-S52858.2021.00013; Erhältlich unter https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9525540
- Zixi Liu, Yang Feng, and Zhenyu Chen. 2021. DialTest: automated testing for recurrent-neural-network-driven dialogue systems. In Proceedings of the 30th ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA 2021). Association for Computing Machinery, New York, NY, USA, 115–126. https://doi.org/10.1145/3460319.3464829; Erhältlich unter DialTest: Automated Testing for Recurrent-Neural-Network-Driven Dialogue Systems (acm.org)
- W. Sun, Y. Lu and M. Sun, "Are Coverage Criteria Meaningful Metrics for DNNs?," *2021 International Joint Conference on Neural Networks (IJCNN)*, Shenzhen, China, 2021, pp. 1-8, doi: 10.1109/IJCNN52387.2021.9533987; Erhältlich unter https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9533987
- Y. Yuan, Q. Pang and S. Wang, "Revisiting Neuron Coverage for DNN Testing: A Layer-Wise and Distribution-Aware Criterion," *2023 IEEE/ACM 45th International Conference on Software Engineering (ICSE)*, Melbourne, Australia, 2023, pp. 1200-1212, doi: 10.1109/ICSE48619.2023.00107; Erhältlich unter https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10172683
- Pei, Kexin ; Cao, Yinzhi ; Yang, Junfeng ; Jana, Suman: DeepXplore: Automated Whitebox Testing of Deep Learning Systems. (2017), Oktober; Erhältlich unter http://dx.doi.org/10.1145/3132747.3132785
- Mario Winter: Testing AI – Where are we now?; 05.10.2022; Software QS-Tag