

Exploration and systematic assessment of the resource efficiency of Machine Learning

Achim Guldner, Sandro Kreten, Stefan Naumann

Hochschule Trier
Umwelt-Campus Birkenfeld
Institut für Softwaresysteme

01. Oktober 2021

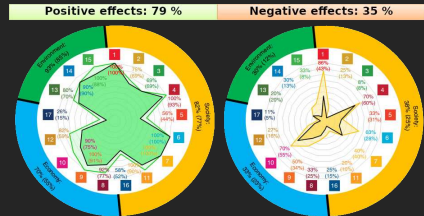
Sowohl positive als auch negative Effekte auf nachhaltige Entwicklung durch KI

Starke Zunahme notwendiger Rechenleistung, insbesondere im Bereich Deep Learning.

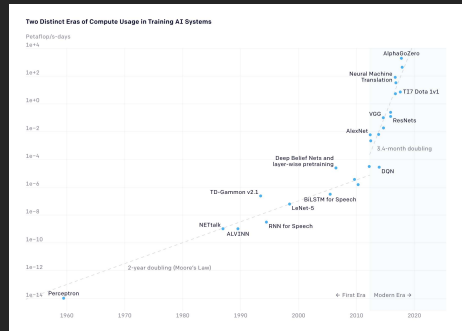
→ OpenAI (2019): Seit 2012 300.000-facher Anstieg der Rechenzeit für KIs

Dadurch z.T. extremer Energieverbrauch/CO₂-Ausstoß

→ Strubell (2019): Neural Architecture Search (NAS) – Transformer model (213M parameters) zur Übersetzung Englisch/Deutsch: ca. 656.000 kWh (280 Tonnen CO₂e)



Aus Vinuesa et al. (2020) *The role of artificial intelligence in achieving the Sustainable Development Goals*. Nature Communications 11/1.



<https://openai.com/blog/ai-and-compute/>

Bestehender Ansatz

- Software triggert Auslastung, aber steuert auch die Hardware.
 - Evaluation und Optimierung der Energie- und Ressourceneffizienz von Software, Programmiersprachen, Bibliotheken etc.
- Entwicklung von Kriterien und Modellen zur Erfassung der Nachhaltigkeit von Software¹
 - Energie- und Ressourceneffizienz
 - Einflussfaktoren auf die Hardware-Nutzungsdauer
 - Nutzungsautonomie
- Resultat: Seit 2020 Blauer Engel für Software²



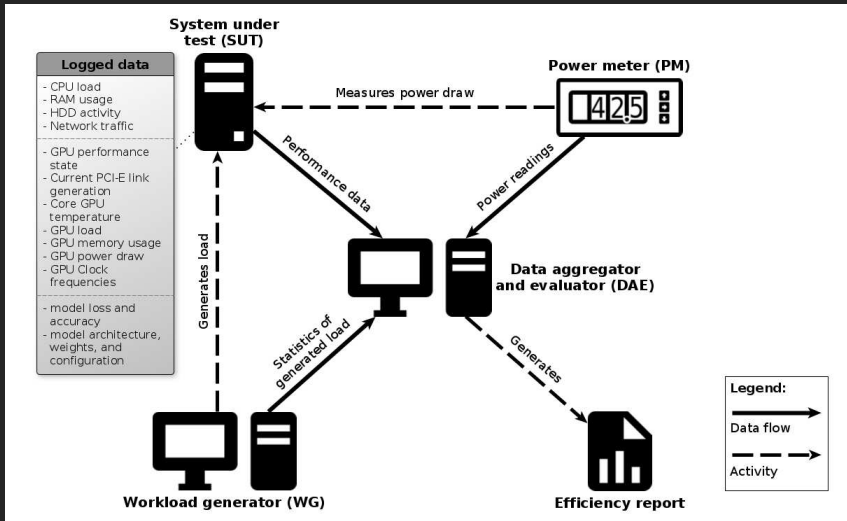
Ziele der Arbeit

- Übertragung des Ansatzes auf den Bereich KI und ML
- Fallbeispiel: Bestimmung des Ressourcen- und Energieverbrauchs von CNNs zur Bilderkennung
- Evaluation des Ansatzes im Bereich KI/ML und Ausblick auf weitere Forschung

¹<https://green-software-engineering.de/kriterienkatalog>

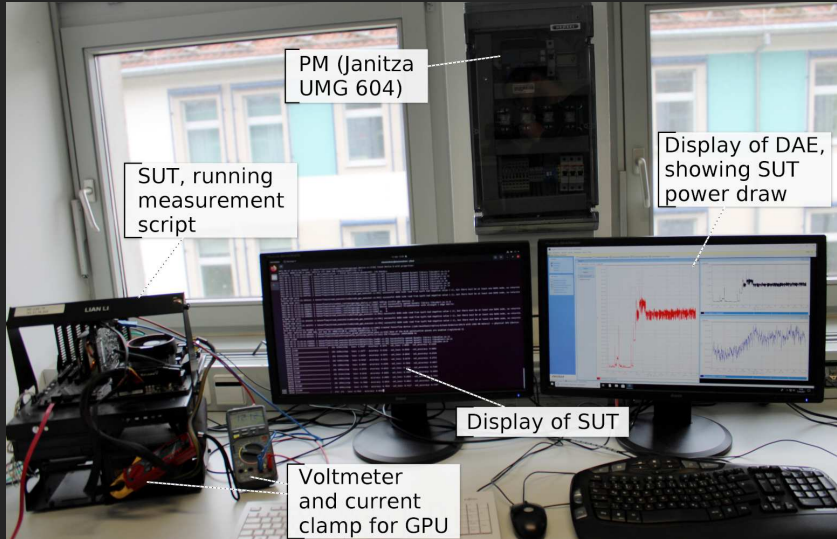
²<https://www.blauer-engel.de/de/produktwelt/elektrogeraete/ressourcen-und-energieeffiziente-softwareprodukte>

Messaufbau



Eigene Darstellung

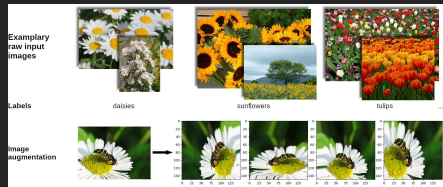
Messaufbau



Eigene Darstellung

Übertragung des Ansatzes und Experimente im Bereich CNN

- Messung des Trainings von Neuronalen Netzen mit TensorFlow
- Klassifikation von Blumen in Bildern durch Convolutional Neural Network (CNN)
- Vergleich eines selbst erstellten Netzes mit *MobileNet V2*



Beispielbilder aus https://www.tensorflow.org/datasets/catalog/tf_flowers

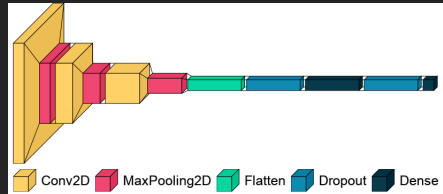
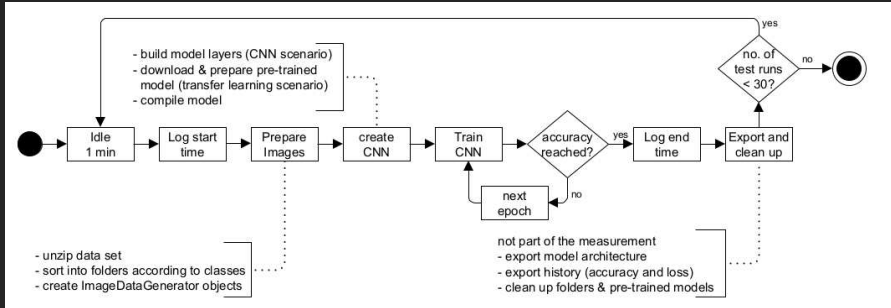


Abbildung des CNN. Erstellt mit visualkerkeras <https://github.com/paulgavrikov/visualkerkeras/>

Replication package:

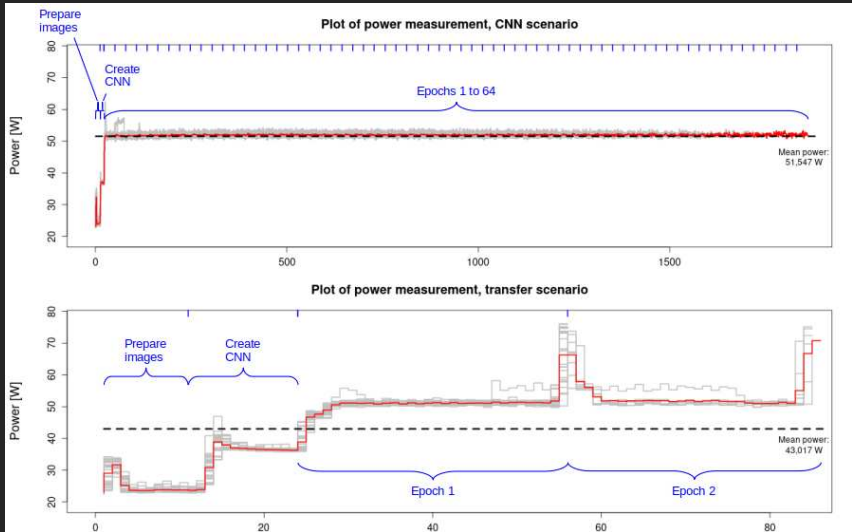
<https://gitlab.rlp.net/a.guldner/exploration-and-systematic-assessment-of-the-resource-efficiency-of-machine-learning>

Szenario-Ablauf



Eigene Darstellung

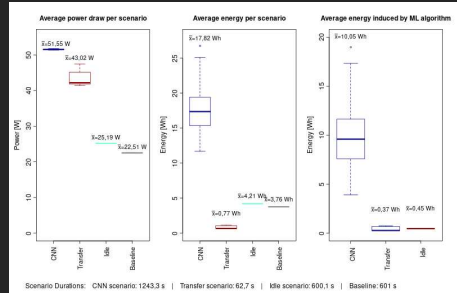
- Normalverteilte Stichprobe ($n = 30$)
- Reproduzierbarkeit
- Baseline-, Idle- und Szenariomessung (Training eines CNN, Transfer-learning)



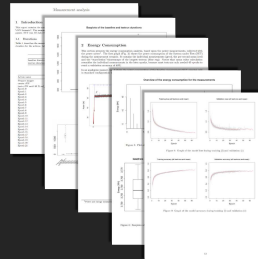
Eigene Darstellung

Analyse

- *Energieverbrauch und Hardwarenutzung insgesamt und nach Aktionen (Datenvorverarbeitung, Training bis epoch-Ebene, Klassifikation etc.)*
- *Software-induzierte Verbräuche*
- *Energieeffizienzfaktoren*
- *Verbrauchsintensive Bereiche und Konfigurationen*
- *Vergleich des Verbrauchs zwischen*
 - Bibliotheken
 - Netzen und Versionen von Netzen
 - Programmiersprachen
 - Algorithmen/Implementierungen
 - ...



Eigene Darstellung

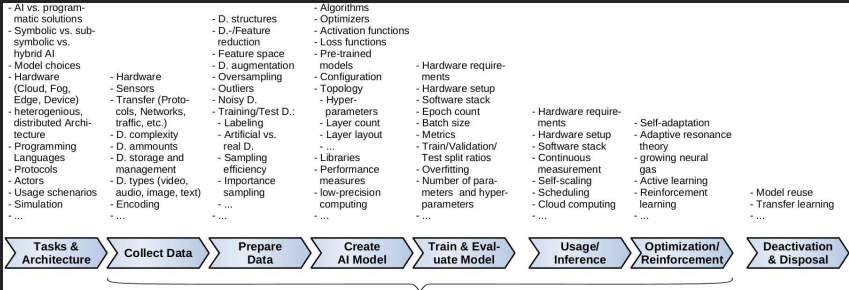


Eigene Darstellung

Besonderheiten bei der Messung von CNNs

- Synchronisierung & Reproduzierbarkeit der Messdurchläufe
 - Seed für RNG setzen
 - RAM und GRAM leeren
 - Modelle und Trainingsdaten zurücksetzen
 - BG-Prozesse deaktivieren
- Integration der Messung der Grafikkartenparameter mittels nvidia-smi
- Energie- und Ressourcenverbrauchs bis auf epoch-Ebene:
Erlaubt Vergleiche z.B. bis bestimmte epoch oder accuracy erreicht.
- Integration der Modell-Metriken
- Berechnung der Energieeffizienz der Algorithmen möglich, z.B.
 - "validation accuracy pro Wh"
 - "F1-score pro Wh"
 - "correct classifications pro Wh"

- Was beeinflusst den Verbrauch?
- Wie kann die Ressourceneffizienz gemessen werden?
- Ansätze zur Optimierung?



Measurement

- Induced energy consumption and Hardware usage
- Scenario runtime

Influencing factors

- Measurement hardware
- Measurement setup
 - PUE
 - AC-DC-conversion
 - SUT hardware components
 - SUT software stack
 - Distributed hard- and software
- Measuring errors/inaccuracies and interpolation
- Extrapolation
- Reporting
- ...

Eigene Darstellung



Arbeitsgruppe Green Software Engineering
Institut für Softwaresysteme
Hochschule Trier, Umwelt-Campus Birkenfeld

Forschungskolleg
"AI-based Cyber-Physical Process Systems"

E-Mail: greensoft@umwelt-campus.de
Internet: <https://www.green-software-engineering.de>

Vielen Dank für Ihre Aufmerksamkeit!



Achim Guldner, M. Sc.



Sandro Kreten, M. Ed.



Prof. Dr. Stefan Naumann