

FRAUNHOFER-INSTITUT FÜR TECHNO- UND WIRTSCHAFTSMATHEMATIK ITWM

DAV/DGVFM
Jahrestagung

2022

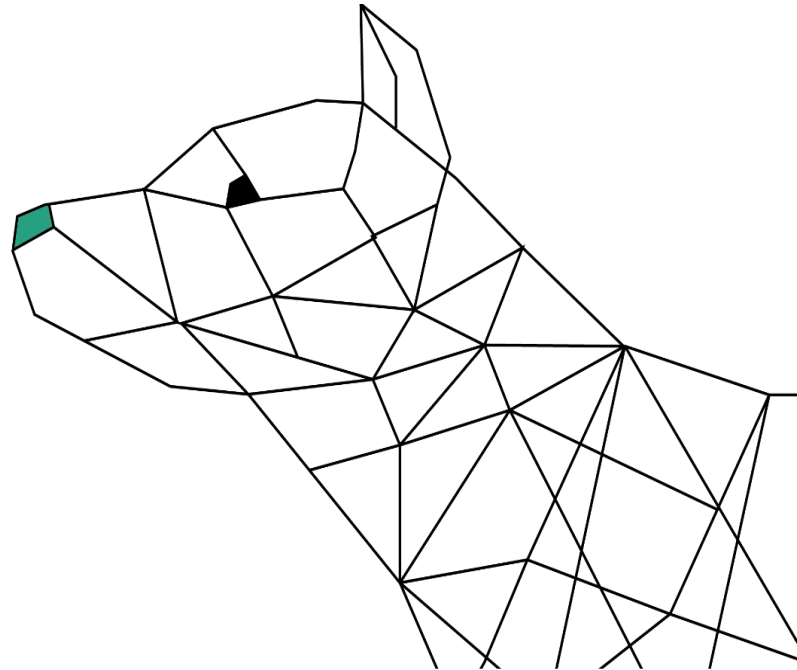
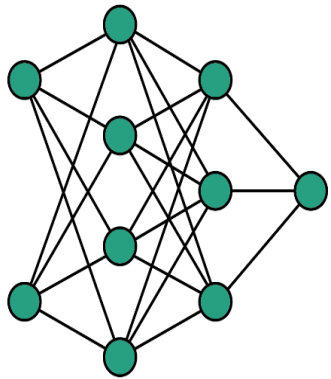
DAV/DGVFM Jahrestagung 2022

Dr. Stefan Mai
Dominik Lorocho

29.04.2022



Optimierung Neuronaler Netze mit der Software *Neural Architecture Search Engine* (NASE)



Wettbewerbssieger
BMBF 2021
Preisgeld: 1 Mio. Euro

Ihre Referenten stellen sich vor



Diplomökonom Univ.
Dr. Stefan Mai

Kundenmanagement
+49 151 19783934
stefan.mai@itwm.fraunhofer.de

Heutige Rolle:
Einführung und
Ansprechpartner für NASE



Diplom Elektroingenieur
Dominik Lorocho

Mitarbeiter High
Performance Computing
Deep Learning Gruppe

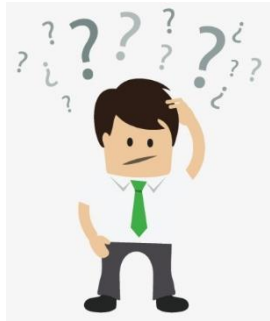
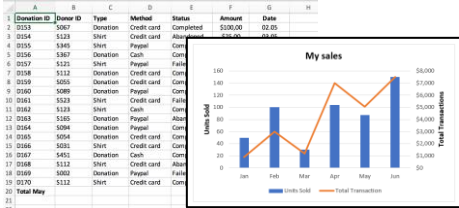
Heutige Rolle:
Fachexperte NASE

AGENDA

- Wer sind wir?
- Wie funktioniert die Innovation NASE?
- Welche Ergebnisse erbrachte ein Leistungstest beim Testkunden?
- Welche wirtschaftlichen Vorteile ergeben sich für den Testkunden?
- Wie funktioniert ein kostenloser Leistungstest mit NASE in Ihrem Unternehmen?

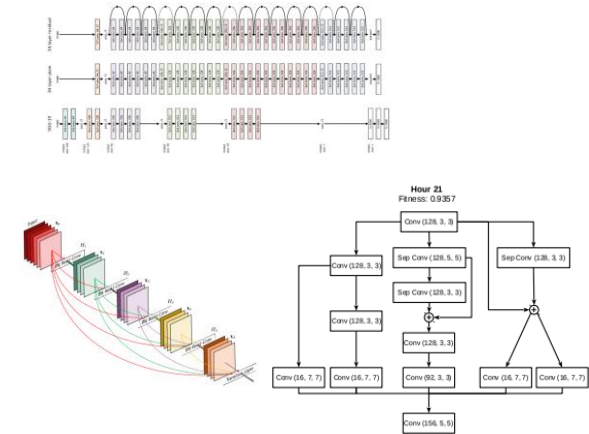
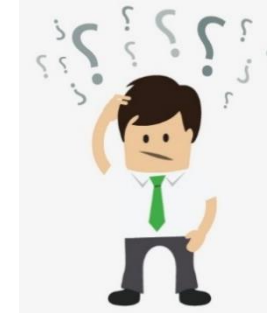
Wie finde ich ein neuronales Netzwerk für mein Problem?

Problem: Deep Learning Entwurf braucht viel Know-how!



Anforderungen:

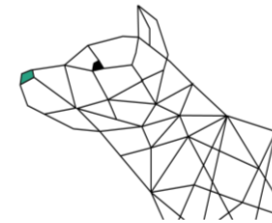
- Modellgröße
- Laufzeit
- Speicherverbrauch
- ...



- Wie entwerfe ich ein DL Modell?
- Wie trainiere und evaluiere ich das Modell?

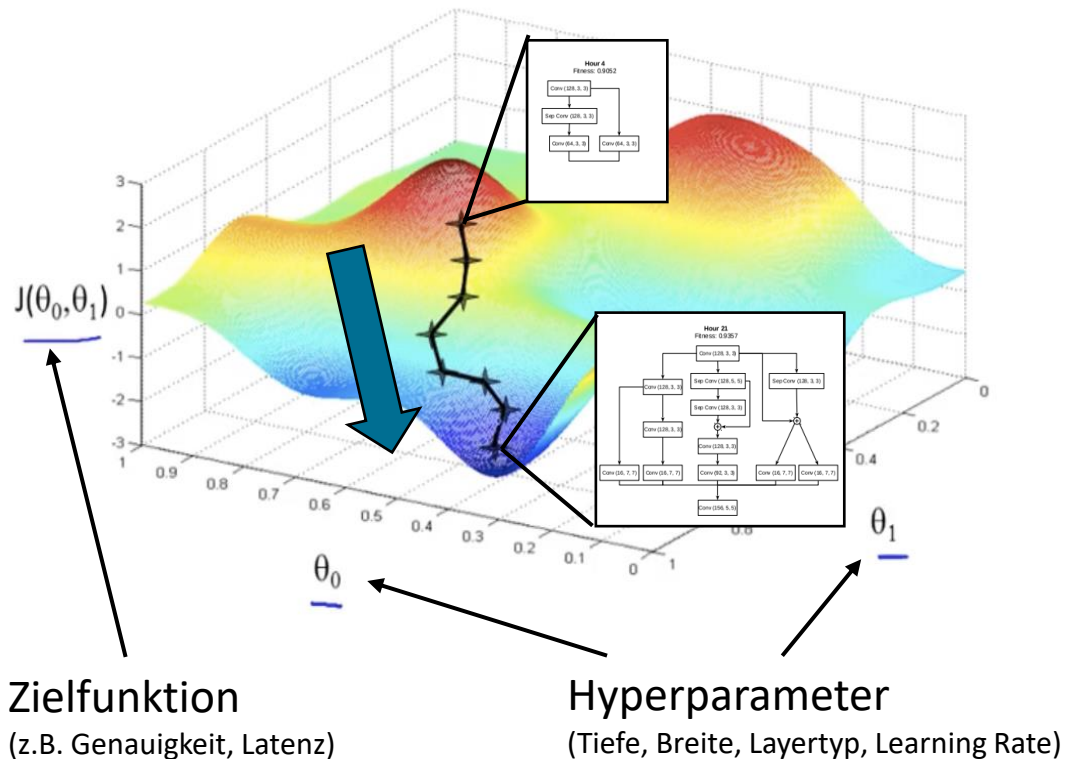
- Was für ein Modell eignet sich für das Problem?
- Was sind die Kriterien für eine gute Lösung?

Lösung: Automatisierter Modellentwurf mit NASE!



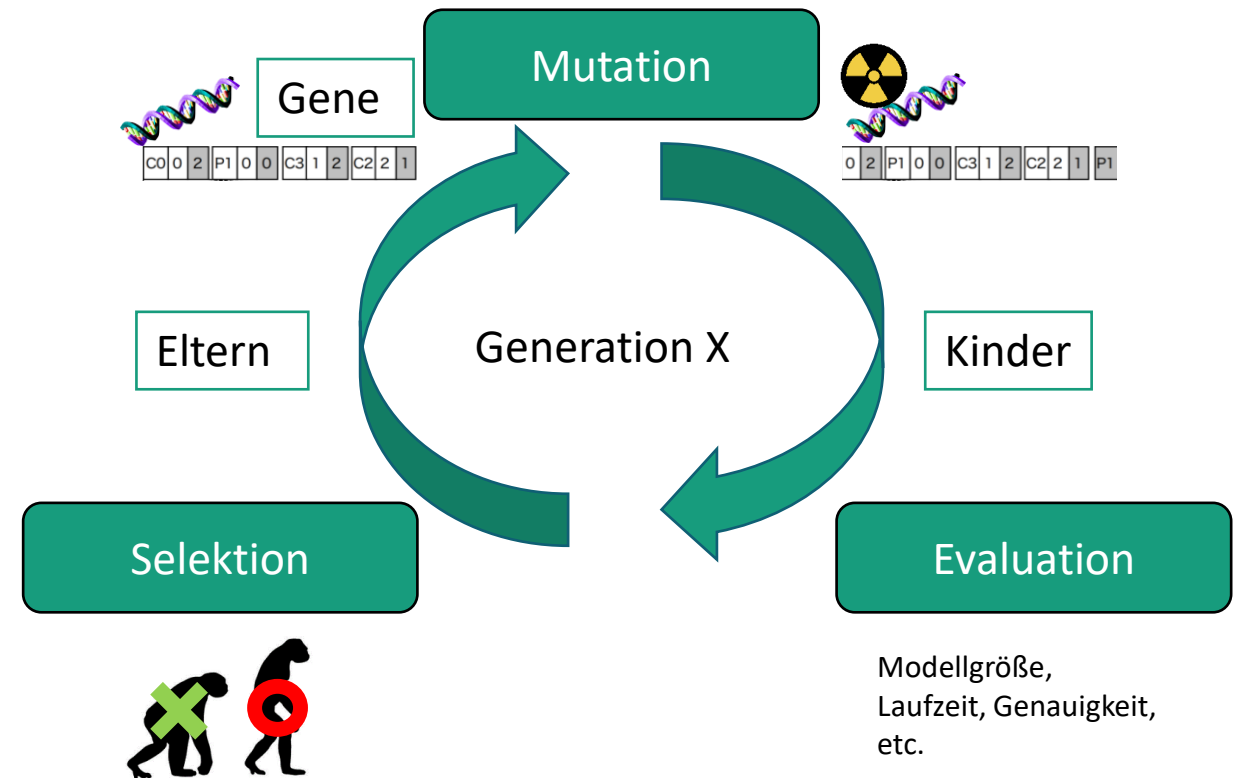
Schauen Sie mit uns in das Innere der Innovation NASE

Suchraum besteht aus neuronalen Netzen



Problem: Kein formalisierbarer Zusammenhang!

Evolutionärer, genetischer Algorithmus



Mutation wird auf Grundlage der Fitness gesteuert

Spezifische Zielfunktionen

$$f_{\text{energy}}(N) = \sum_{l \in L_N} (n_{\text{inputs}}^{(l)} + n_{\text{outputs}}^{(l)} + n_{\text{params}}^{(l)}),$$

Lookup Tables, Simulator, ...

Extrem flexibel!

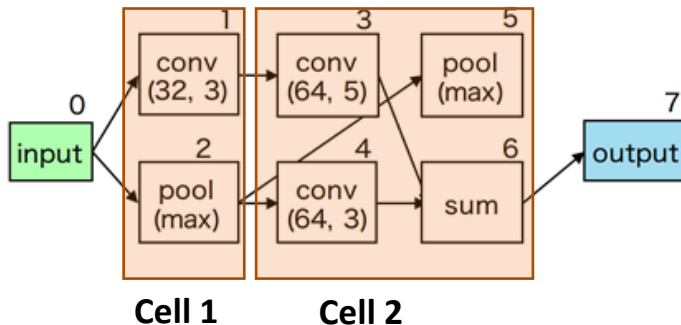
Bedingungen und Gewichtung

```
[{"name": "number_parameters", "hard_constraint": [0, 100000], "soft_constraint": [0, 10000], "weight": 0.05}, {"name": "non_detection_rate", "hard_constraint": [0, 0.9], "soft_constraint": [0, 0.1], "weight": 0.45}, {"name": "false_alarm_rate", "hard_constraint": [0, 1], "soft_constraint": [0, 0.2], "weight": 0.45}, {"name": "floating_point_operations", "hard_constraint": [0, 200000000], "soft_constraint": [0, 100000000], "weight": 0.05}]
```

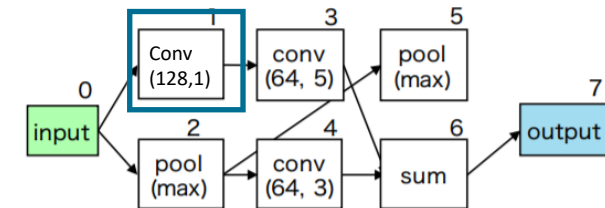
Fitness = gewichtete Summe aller Zielfunktionen

Fundamentaler Trade-off

Exploitation vs. Exploration



Crossover, inaktive Gene



Hyperparameter, Zellstruktur

Evolutionäre Algorithmen brauchen eine effiziente Implementierung

Problem: Stochastisches Optimierungsverfahren mit vielen Zielfunktionen kostet Zeit!

accuracy

+

latency

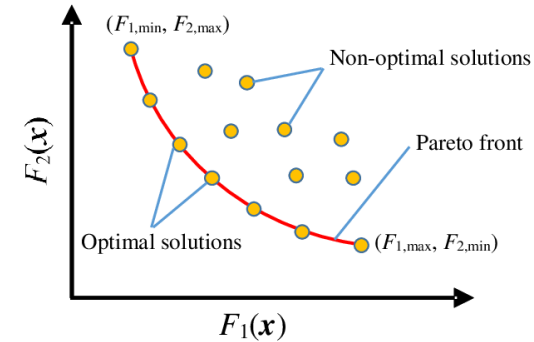
power

...

NAS



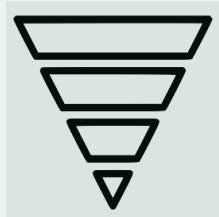
time!



**Fluch der Dimensionalität!
Hoher Zeitaufwand!**

Lösung: Algorithmisches und High-Performance Computing Know-how!

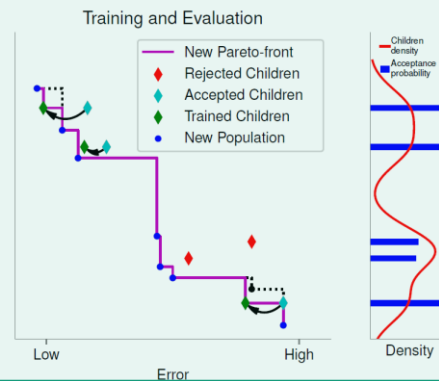
Hierarchische Evaluation



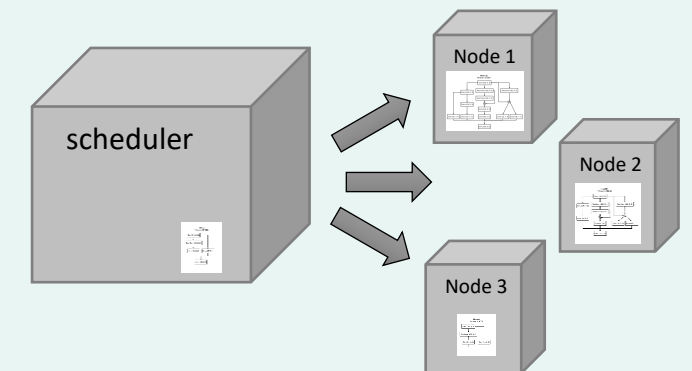
viele DNNs
→ # parameters
→ approx. latency

wenige DNNs
→ accuracy
→ quantization

Smarte Pareto Front Exploration



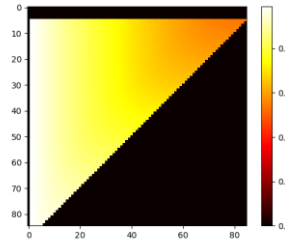
HPC für schnelle Auswertung



Der Test-User beauftragte einen kostenlosen Leistungstest mit folgenden Anforderungen:

Drei Inputgrößen und eine Outputgröße:

Geschlecht	Beitrittsalt	Laufzeit	Wert
M	5	1	0.99108
W	5	1	0.982242
M	5	2	0.973483
W	5	2	0.964804
M	5	3	0.956204
W	5	3	0.947681



Leistungsanforderungen:

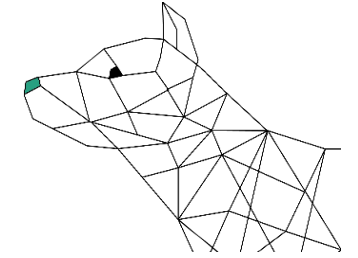
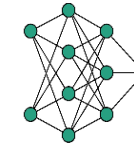
- Max. Absolute Abweichung $< 1e-4$
- Anteil (Abweichung $< 1e-6$) größer als 10%
- Trainingsdauer $< 24h$
- Modellgröße $< 30MB$

Anzahl Sample: ca. 7000.

Regressionsproblem: Finde ein neuronales Netzwerk, welches diese Funktion abbildet!

Wichtigste Leistungsanforderung: Genauigkeit!

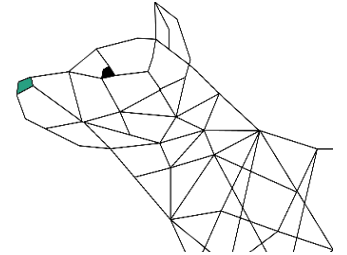
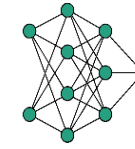
NASE erreichte folgende Leistungsergebnisse:



Testergebnisse nach
Vorgaben des Test-
Users

	Leistungs- vorgaben	<u>NASE</u> Size S	<u>NASE</u> Size M	<u>NASE</u> Size L	<u>NASE</u> Size XL
Max. Abweichung	1,00E-04	5,65E-05	5,79E-05	2,79E-05	2,96E-05
In Toleranz (<1e-6)	6,00%	14,50%	23,70%	44,05%	75,40%
Größe (#Param)	752.000	1.400.000	7.400.000	37.800.000	400.000.000
Speicher (MB)	3,3	5,4	29,6	145	1500
Trainingszeit (Min)	1800	9	13	37	400

NASE bietet dem Test-User folgende unternehmerischen Mehrwerte:



- ✓ Kosteneinsparung (weniger personeller Aufwand)
- ✓ Bessere Qualität der Modelle
- ✓ Mehr Kundenprojekte in gleicher Zeit möglich
- ✓ Geringere Kosten für die Hardware (Anschaffung bzw. Cloud-Miete)
- ✓ Schnellere Anpassungsmöglichkeiten an andere Probleme

Wir bieten Ihnen als Test-User einen kostenlosen Test mit Ihren Daten!

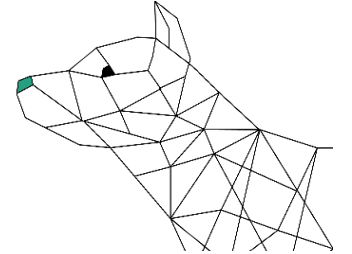
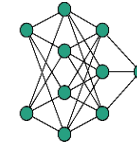
Wir benötigen:

- Einen (gelabelten) Datensatz (auch synthetisch!)
- Gewünschte Anforderungen bezüglich Vorhersagegenauigkeit, Modellgröße, etc.
- Vergleichsdaten Ihres Netzes, um die Optimierungsergebnisse von NASE darzustellen

Wir liefern:

- Eine Aussage über das Spektrum an erreichbaren DNN Lösungen für Ihr Problem
- In der Regel binnen 14 Tagen
- Mit der Auslieferung des optimalen Netzes betreten Sie den kostenpflichtigen Bereich

Interesse? Bitte Mail an stefan.mai@fraunhofer.itwm.de oder +49 151-19783934



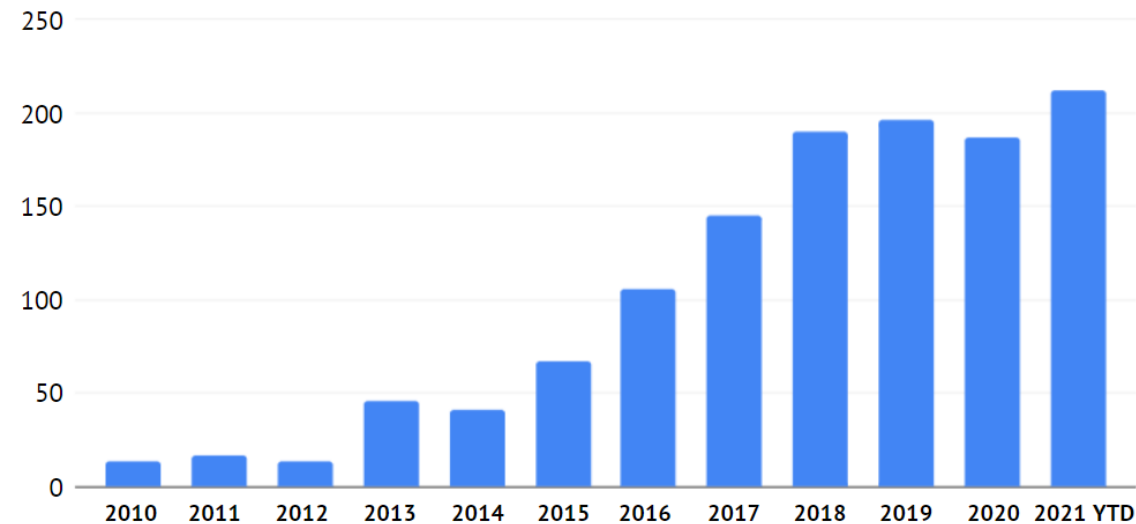
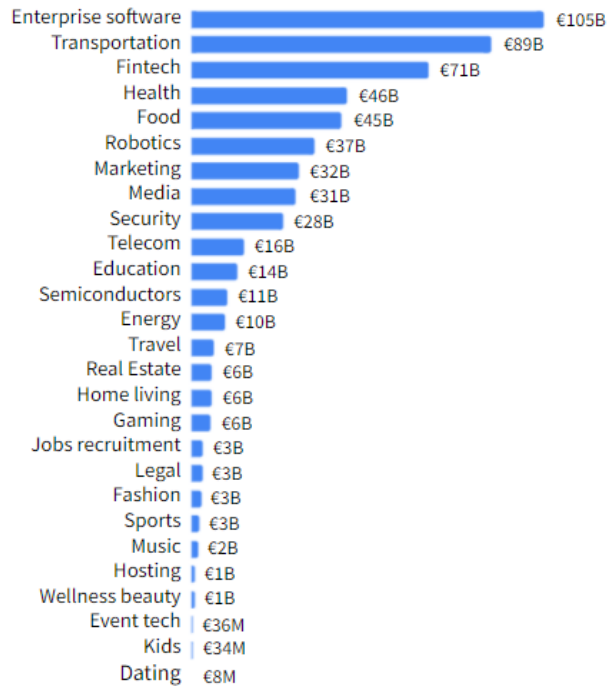
Dominik Loroach und Stefan Mai freuen sich jetzt über Ihre Fragen



Industrie und Dienstleister investieren in KI.

Investment in KI Firmen, Finanztechnologien (Fintech)
auf Platz 3

Rekord an Übernahmen von AI Firmen zeigt ein
ansteigendes Interesse an KI Firmen



KI findet mehr und mehr konkrete Anwendung in diversen Bereichen!

<https://www.stateof.ai/>