

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro

(43) Internationales Veröffentlichungsdatum
02. Mai 2024 (02.05.2024)



(10) Internationale Veröffentlichungsnummer
WO 2024/088682 A1

(51) Internationale Patentklassifikation:

G06N 3/048 (2023.01) G06N 3/084 (2023.01)
G06N 3/02 (2006.01)

(21) Internationales Aktenzeichen: PCT/EP2023/076715

(22) Internationales Anmeldedatum:
27. September 2023 (27.09.2023)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
10 2022 211 447.7
28. Oktober 2022 (28.10.2022) DE

(71) Anmelder: ZF FRIEDRICHSHAFEN AG [DE/DE]; Lö-
wentaler Straße 20, 88046 Friedrichshafen (DE).

(72) Erfinder: BIRK, Markus; Amselweg 5, 88441 Mittelbi-
berach (DE). BIEL, Steffen; Weidenring 75, 88046 Fried-
richshafen (DE).

(81) Bestimmungsstaaten (soweit nicht anders angegeben, für
jede verfügbare nationale Schutzrechtsart): AE, AG, AL,
AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY,
BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DJ, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG,
KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY,
MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA,
NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO,
RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH,

TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS,
ZA, ZM, ZW.

(84) Bestimmungsstaaten (soweit nicht anders angegeben, für
jede verfügbare regionale Schutzrechtsart): ARIPO (BW,
CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD,
SL, ST, SZ, TZ, UG, ZM, ZW), eurasisches (AM, AZ, BY,
KG, KZ, RU, TJ, TM), europäisches (AL, AT, BE, BG, CH,
CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS,
IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO,
RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM,
GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Veröffentlicht:

— mit internationalem Recherchenbericht (Artikel 21 Absatz
3)

(54) Title: RUNTIME OPTIMIZATION OF NEURAL NETWORKS WITH QUASI-STATIC INPUTS

(54) Bezeichnung: LAUFZEITOPTIMIERUNG VON NEURONALEN NETZEN BEI QUASISTATISCHEN INPUTS

(57) Abstract: The invention relates to a method for using an artificial neural network. A check is carried out in order to determine whether a first activation and a second activation of at least one neuron equals zero and whether a second network input of the at least one neuron comes closer to a threshold than a first network input of the at least one neuron. In the event of a difference between the first network input of the at least one neuron and the threshold, the period of time until the network input of the at least one neuron reaches the threshold is extrapolated from the difference between the second network input of the at least one neuron and the threshold and from the period of time which has elapsed between a first measurement and a second measurement, and a check is carried out in order to determine whether the third measurement was taken before the extrapolated period of time expired. If so, the value zero is output as the third activation of the at least one neuron without applying the activation function of the at least one neuron.

(57) Zusammenfassung: Die Erfindung betrifft ein Verfahren zur Anwendung eines künstlichen neuronalen Netzes. Es wird geprüft, ob eine erste Aktivierung und eine zweite Aktivierung mindestens eines Neurons Null betragen und eine zweite Netzeingabe des mindestens einen Neurons einem Schwellenwert näherkommt als eine erste Netzeingabe des mindestens einen Neurons, und in dem Fall in Abhängigkeit von einer Differenz zwischen der ersten Netzeingabe des mindestens einen Neurons und dem Schwellenwert, von einer Differenz zwischen der zweiten Netzeingabe des mindestens einen Neurons und dem Schwellenwert und von einer zwischen einer ersten Messung und einer zweiten Messung verstrichenen Zeit eine Zeit extrapoliert wird, bis die Netzeingabe des mindestens einen Neurons den Schwellenwert erreicht, und geprüft wird, ob die dritte Messung vor Ablauf der extrapolierten Zeit erfolgte, und in dem Fall als dritte Aktivierung des mindestens einen Neurons ohne Anwendung der Aktivierungsfunktion des mindestens einen Neurons der Wert Null ausgegeben wird.



WO 2024/088682 A1

Laufzeitoptimierung von neuronalen Netzen bei quasistatischen Inputs

Die Erfindung betrifft ein Verfahren nach dem Oberbegriff von Anspruch 1 und ein Computerprogramm nach Anspruch 8.

Für anspruchsvolle Steuerungsaufgaben werden in Fahrzeugsteuergeräten häufig neuronale Netze verwendet. Deren Bedarf an Rechenkapazität ist vergleichsweise hoch. Dies führt dazu, dass sich die Hardware der Steuergeräte verteuert.

Der Erfindung liegt die Aufgabe zugrunde, die Anwendung eines künstlichen neuronalen Netzes ressourceneffizienter zu gestalten. Diese Aufgabe wird gelöst durch ein Verfahren nach Anspruch 1 und ein Computerprogramm nach Anspruch 8. Bevorzugte Weiterbildungen sind in den Unteransprüchen enthalten und ergeben sich aus nachfolgender Beschreibung.

Das erfindungsgemäße Verfahren dient der Anwendung eines zuvor bereits trainierten künstlichen neuronalen Netzes. Ein künstliches neuronales Netz besteht aus miteinander verknüpften künstlichen Neuronen. Ein künstliches Neuron weist ein oder mehrere Eingänge und einen Ausgang auf. An den Eingängen anliegende Eingangswerte werden gewichtet. Dies geschieht in der Regel, indem die Werte mit jeweils einem Gewicht multipliziert werden. Aus den so gewichteten Eingangswerten wird mittels einer Übertragungsfunktion eine Netzeingabe ermittelt. Als Übertragungsfunktion wird gewöhnlich eine Summierungsfunktion verwendet, welche die Summe der gewichteten Eingangswerte berechnet. Auf die Netzeingabe wird eine Aktivierungsfunktion angewendet. Diese ordnet der Netzeingabe eine Aktivierung zu, welche über den Ausgang weitergeleitet wird. Der Ausgang des Neurons kann einen Ausgang des künstlichen neuronalen Netzes bilden oder mit einem Eingang eines weiteren Neurons verknüpft sein. Eingänge des neuronalen Netzes werden durch Eingänge von Neuronen gebildet. Über die Eingänge erhält das neuronale Netz Eingangssignale. Entsprechend werden Ausgangssignale des neuronalen Netzes über die Ausgänge ausgegeben.

Das erfindungsgemäße Verfahren setzt voraus, dass die Aktivierungsfunktion mindestens eines Neurons den Netzeingaben des Neurons in Abhängigkeit eines Schwellwerts den Wert Null zuordnet. Die Aktivierungsfunktion nimmt also bei allen Netzeingaben unterhalb des Schwellwerts oder alternativ bei allen Netzeingaben oberhalb des Schwellwerts den Wert Null an.

Bei Ausführung des erfindungsgemäßen Verfahrens werden drei Iterationen mit jeweils einer Messung ein oder mehrerer physikalischer Größen durchlaufen. Es werden also eine erste Messung, eine zweite Messung und eine dritte Messung der ein oder mehreren physikalischen Größen durchgeführt.

Durch die Messungen wird jeweils ein Satz eines oder mehrerer Eingangssignale des Netzes bestimmt. Ein erster Satz wird durch die erste Messung, ein zweiter Satz durch die zweite Messung und ein dritter Satz durch die dritte Messung bestimmt. Die Eingangssignale werden dabei jeweils direkt gemessen oder aus einer oder mehreren gemessenen physikalischen Größen berechnet.

Mittels des neuronalen Netzes werden jeweils in Abhängigkeit der ein oder mehreren Eingangssignale eines oder mehrerer Ausgangssignale ermittelt. In Abhängigkeit des ersten Satzes der ein oder mehreren Eingangssignale wird ein erster Satz eines oder mehrerer Ausgangssignale, in Abhängigkeit des zweiten Satzes der ein oder mehreren Eingangssignale ein zweites Satz der ein oder mehreren Ausgangssignale und in Abhängigkeit des dritten Satzes der ein oder mehreren Eingangssignale ein dritter Satz der ein oder mehreren Ausgangssignale ermittelt.

Die Ermittlung der Ausgangssignale in Abhängigkeit des ersten Satzes und des zweiten Satzes der ein oder mehreren Eingangssignale erfolgt dabei auf konventionellem Wege. Dies bedeutet, dass die Aktivierungen der einzelnen Neuronen des Netzes durch Anwendung einer jeweiligen Aktivierungsfunktion auf eine mittels einer jeweiligen Übertragungsfunktion berechnete jeweilige Netzeingabe ermittelt wird. Der erste Satz der ein oder mehreren Ausgangssignale wird also durch Ermittlung jeweils einer ersten Aktivierung der Neuronen des Netzes durch jeweilige Anwendung einer jeweiligen Aktivierungsfunktion auf eine mittels einer jeweiligen Übertragungsfunktion be-

rechnet jeweilige erste Netzeingabe ermittelt. Analog wird der zweite Satz der ein oder mehreren Ausgangssignale durch Ermittlung jeweils einer zweiten Aktivierung der Neuronen des Netzes durch jeweilige Anwendung der jeweiligen Aktivierungsfunktion auf eine mittels der jeweiligen Übertragungsfunktion berechnete jeweilige zweite Netzeingabe ermittelt. Für jedes Neuron des Netzes wird also durch Anwendung der Übertragungsfunktion des Neurons die erste beziehungsweise zweite Netzeingabe des Neurons berechnet. Die erste beziehungsweise zweite Netzeingabe des Neurons wird der Übertragungsfunktion des Neurons als Funktionswert übergeben. Diese ordnet der ersten beziehungsweise zweiten Netzeingabe eine Aktivierung als erstes beziehungsweise zweites Ausgangssignal zu.

Die Aktivierungsfunktion kann ebenso wie die Übertragungsfunktion für einzelne oder alle Neuronen des Netzes unterschiedlich sein. Vorzugsweise wird aber nur eine einzige, für jedes Neuron identische Aktivierungsfunktion verwendet.

Auch der dritte Satz der ein oder mehreren Ausgangssignale wird auf Basis der Aktivierungen der Neuronen des Netzes in Abhängigkeit der ein oder mehreren Eingangssignale ermittelt. In Abhängigkeit des dritten Satzes der ein oder mehreren Eingangssignale wird also der dritte Satz der ein oder mehreren Ausgangssignale durch Ermittlung jeweils einer dritten Aktivierung der Neuronen des Netzes ermittelt.

Die Erfindung sieht allerdings eine abweichende Ermittlung der Aktivierung des oben genannten mindestens einen Neurons vor, dessen Aktivierungsfunktion in Abhängigkeit von dem Schwellwert den Wert Null annimmt. Dem liegt die Idee zugrunde, auf eine Berechnung der Aktivierungsfunktion zu verzichten, wenn absehbar ist, dass die Aktivierungsfunktion den Wert Null annimmt.

Dazu wird geprüft, ob die erste Aktivierung und die zweite Aktivierung des mindestens einen Neurons Null betragen. Im Einzelnen werden die erste und zweite Netzeingabe des mindestens einen Neurons in Hinblick auf den Schwellwert geprüft. Es wird geprüft, ob die erste und zweite Netzeingabe des mindestens einen Neurons unterhalb bzw. oberhalb des Schwellwerts liegen, sodass die Aktivierungsfunktion

des mindestens einen Neurons der ersten und zweiten Netzeingabe den Wert Null zuordnet.

Zusätzlich wird geprüft, ob die zweite Netzeingabe des mindestens einen Neurons dem Schwellwert näherkommt als die erste Netzeingabe. Es wird also geprüft, ob ein Betrag einer Differenz zwischen der zweiten Netzeingabe und dem Schwellwert geringer ist als ein Betrag einer Differenz zwischen der ersten Netzeingabe und dem Schwellwert.

Liegen die erste und zweite Netzeingabe des mindestens einen Neurons unterhalb bzw. oberhalb des Schwellwerts, und kommt die zweite Netzeingabe des mindestens einen Neurons dem Schwellwert näherkommt als die erste Netzeingabe, so deutet dies darauf hin, dass sich die Netzeingabe des mindestens einen Neurons dem Schwellwert nähert und diesen über- bzw. unterschreiten wird, sodass die Aktivierungsfunktion des mindestens einen Neurons einen von Null verschiedenen Wert annehmen wird. Die Zeit bis dahin wird extrapoliert. Wenn also die zuvor durchgeführte Prüfung ergeben hat, dass die erste und zweite Netzeingabe des mindestens einen Neurons unterhalb bzw. oberhalb des Schwellwerts liegen, sodass die erste Aktivierung und die zweite Aktivierung des mindestens einen Neurons Null betragen, und die zweite Netzeingabe des mindestens einen Neurons dem Schwellwert näherkommt als die erste Netzeingabe des mindestens einen Neurons, wird eine Zeit extrapoliert, bis die Netzeingabe des mindestens einen Neurons den Schwellwert erreicht.

Die Zeit, bis die Netzeingabe des mindestens einen Neurons den Schwellwert erreicht, beginnt mit dem Zeitpunkt der zweiten Messung zu laufen und endet, sobald die Differenz zwischen der ersten Netzeingabe und der zweiten Netzeingabe des mindestens einen Neurons den Schwellwert erreicht. Die Zeit ist abhängig von einer Differenz zwischen der ersten Netzeingabe des mindestens einen Neurons und dem Schwellwert, von einer Differenz zwischen der zweiten Netzeingabe des mindestens einen Neurons und dem Schwellwert und von einer zwischen der ersten Messung und der zweiten Messung verstrichenen Zeit. Diese wird bei der Extrapolation der Zeit berücksichtigt.

Weiterhin wird geprüft, ob die dritte Messung vor oder nach Ablauf der extrapolierten Zeit erfolgte. Erfolgte die dritte Messung vor Ablauf der extrapolierten Zeit, deutet dies darauf hin, dass die Aktivierungsfunktion des mindestens einen Neurons der dritten Netzeingabe des mindestens einen Neurons den Wert Null zuordnet. In dem Fall kann auf eine Berechnung der Gewichtungen des mindestens einen Neurons, eine Anwendung der Übertragungsfunktion des mindestens einen Neurons und eine Anwendung der Aktivierungsfunktion des mindestens einen Neurons verzichtet werden, um die dritte Aktivierung des mindestens einen Neurons zu ermitteln. Stattdessen wird als dritte Aktivierung des mindestens einen Neurons ohne Anwendung der Aktivierungsfunktion des mindestens einen Neurons der Wert Null ausgegeben.

Andernfalls wird die dritte Aktivierung des mindestens einen Neurons auf herkömmlichem Weg ermittelt. Wenn also die dritte Messung nach Ablauf der interpolierten Zeit erfolgte und/oder wenn die erste Aktivierung und/oder die zweite Aktivierung des mindestens einen Neurons von Null verschieden sind und/oder wenn die zweite Netzeingabe des mindestens einen Neurons dem Schwellwert nicht näherkommt als die erste Netzeingabe des mindestens einen Neurons, wird die dritte Aktivierung des mindestens einen Neurons auf konventionelle Weise durch Anwendung der Aktivierungsfunktion des mindestens einen Neurons eine mittels der Übertragungsfunktion des mindestens einen Neurons berechnete dritte Netzeingabe ermittelt.

Das erfindungsgemäße Verfahren ist dann besonders vorteilhaft, wenn die gemessenen ein oder mehreren physikalischen Größen sich nur träge verändern. Dies trifft bei vielen in einem Fahrzeug gemessenen physikalischen Größen zu. Bei träge veränderlichen Größen ist das Fehlerrisiko durch die Extrapolation vergleichsweise gering. Zudem ist es möglich, große Zeiträume zu extrapolieren. Durch Verzicht auf die Anwendung der Aktivierungsfunktion lassen sich dabei erhebliche Mengen an Rechenzeit einzusparen.

In einer bevorzugten Weiterbildung wird geprüft, ob die erste Aktivierung und die zweite Aktivierung des mindestens einen Neurons Null betragen, bzw. ob die erste und zweite Netzeingabe des mindestens einen Neurons unterhalb bzw. oberhalb des

Schwellwerts liegen, und ob die zweite Netzeingabe des mindestens einen Neurons weiter von dem Schwellenwert entfernt ist als die erste Netzeingabe des mindestens einen Neurons, d.h. ob ein Betrag einer Differenz zwischen der zweiten Netzeingabe und dem Schwellwert größer ist als ein Betrag einer Differenz zwischen der ersten Netzeingabe und dem Schwellwert. Dies ist ein Indiz dafür, dass sich die Netzeingabe des mindestens einen Neurons nicht dem Schwellwert annähert, sodass die Aktivierung des mindestens einen Neurons auf absehbare Zeit weiterhin den Wert Null annehmen wird.

Weiterbildungsgemäß wird vor diesem Hintergrund eine vordefinierte, d.h. vor Ausführung des Verfahrens festgelegt Zeit herangezogen, bis zu deren Ablauf von einer Anwendung der Übertragungsfunktion des mindestens einen Neurons abgesehen wird. Es wird also geprüft, ob die dritte Messung vor Ablauf einer vordefinierten Zeit erfolgte. Erfolgte die dritte Messung vor Ablauf einer vordefinierten Zeit, wird weiterbildungsgemäß ohne Anwendung der Aktivierungsfunktion des mindestens einen Neurons der Wert Null als dritte Aktivierung des mindestens einen Neurons gewählt. Nach Ablauf der vordefinierten Zeit wird die Aktivierung des mindestens einen Neurons bevorzugt auf konventionellem Wege durch Anwendung der Aktivierungsfunktion des mindestens einen Neurons auf eine mittels der Übertragungsfunktion des mindestens einen Neurons berechnete Netzeingabe ermittelt.

Die Zeit, bis die Netzeingabe des mindestens einen Neurons den Schwellwert erreicht, wird in einer bevorzugten Weiterbildung durch lineare Extrapolation ermittelt. Insbesondere kann die Zeit ermittelt werden als Produkt aus der zwischen der ersten Messung und der zweiten Messung verstrichenen Zeit, aus der Differenz zwischen der zweiten Netzeingabe und dem Schwellwert und aus einem Kehrwert einer Differenz zwischen der ersten Netzeingabe und der zweiten Netzeingabe des mindestens einen Neurons. Die weiterbildungsgemäße Ermittlung der Zeit durch lineare Extrapolation gestaltet sich besonders einfach und schnell.

In einer bevorzugten Weiterbildung repräsentieren ein oder mehrere der Eingangssignale des Netzes physikalische Größen eines Fahrzeugs. In Betracht kommen etwa Größen wie Temperatur, Kraft, Drehmoment, Drehzahl, Position und eingelegte

Gangstufe beziehungsweise Übersetzung. Vorzugsweise handelt es sich um physikalische Größen eines Getriebes und/oder eines Motors des Fahrzeugs.

In einer bevorzugten Weiterbildung werden die Eingangssignale des Netzes mindestens teilweise von Sensoren eines Fahrzeugs bereitgestellt. Bei ein oder mehreren der Eingangssignale des Netzes handelt es sich weiterbildungsgemäß also um Signale von Sensoren eines Fahrzeugs. Bevorzugt sind die Sensoren an einem Getriebe und/oder Motor des Fahrzeugs angebracht. Insbesondere können mittels der Sensoren die oben genannten physikalischen Größen gemessen werden.

Analog repräsentieren ein oder mehrere der Ausgangssignale des Netzes in einer bevorzugten Weiterbildung physikalische Größen eines Fahrzeugs, beispielsweise Temperatur, Kraft, Drehmoment, Position, Geschwindigkeit, Drehzahl und/oder eingelegte Gangstufe beziehungsweise Übersetzung. Repräsentiert ein Ausgangssignal die eingelegte Gangstufe beziehungsweise Übersetzung, lässt sich mittels des erfindungsgemäßen Verfahrens eine Getriebesteuerung realisieren.

Auch lassen sich über das neuronale Netz Aktoren ansteuern. In einer entsprechenden bevorzugten Weiterbildung handelt es sich bei ein oder mehreren der Ausgangssignale des Netzes um Steuersignale für Aktoren eines Fahrzeugs. Insbesondere können Gangsteller eines automatisch schaltbaren Getriebes angesteuert werden.

Das erfindungsgemäße Computerprogramm implementiert das erfindungsgemäße Verfahren oder eine bevorzugte Weiterbildung des Verfahrens. Das Computerprogramm umfasst also Anweisungen zur Ausführung des erfindungsgemäßen Verfahrens oder einer bevorzugten Weiterbildung. Dies bedeutet, dass das Computerprogramm einen Computer zur Ausführung des erfindungsgemäßen Verfahrens oder einer bevorzugten Weiterbildung veranlasst, wenn das Computerprogramm von dem Computer ausgeführt wird. Das Computerprogramm kann auf einem Speichermedium gespeichert oder in einem oder mehreren übertragungsfähigen oder übertragenen Signalen codiert sein. Insbesondere kann das Computerprogramm als ein Computerprogrammprodukt vorliegen, das heißt als eine handelbare Einheit beziehungs-

weise eine Einheit, die dem Zweck dient, den Besitz an dem Computerprogramm zu übertragen. Bei dem genannten Computer handelt es sich bevorzugt um ein Fahrzeugsteuergerät, etwa ein Motor- oder Getriebesteuergerät.

Patentansprüche

1. Verfahren zur Anwendung eines künstlichen neuronalen Netzes; wobei die Aktivierungsfunktion mindestens eines Neuron den Netzeingaben des Neurons in Abhängigkeit eines Schwellwerts den Wert Null zuordnet; wobei durch eine erste Messung einer oder mehrerer physikalischer Größen ein erster Satz eines oder mehrere Eingangssignale des Netzes bestimmt wird; wobei in Abhängigkeit des ersten Satzes der ein oder mehreren Eingangssignale ein erster Satz eines oder mehrerer Ausgangssignale durch Ermittlung jeweils einer ersten Aktivierung der Neuronen des Netzes durch jeweilige Anwendung einer jeweiligen Aktivierungsfunktion auf eine mittels einer jeweiligen Übertragungsfunktion berechnete jeweilige erste Netzeingabe ermittelt wird; wobei durch eine zweite Messung der ein oder mehreren physikalischen Größen ein zweiter Satz der ein oder mehreren Eingangssignale bestimmt wird; wobei in Abhängigkeit des zweiten Satzes der ein oder mehreren Eingangssignale ein zweites Satz der ein oder mehreren Ausgangssignale durch Ermittlung jeweils einer zweiten Aktivierung der Neuronen des Netzes durch jeweilige Anwendung der jeweiligen Aktivierungsfunktion auf eine mittels der jeweiligen Übertragungsfunktion berechnete jeweilige zweite Netzeingabe ermittelt wird; wobei durch eine dritte Messung der ein oder mehreren physikalischen Größen ein dritter Satz der ein oder mehreren Eingangssignale bestimmt wird; und wobei in Abhängigkeit des dritten Satzes der ein oder mehreren Eingangssignale ein dritter Satz der ein oder mehreren Ausgangssignale durch Ermittlung jeweils einer dritten Aktivierung der Neuronen des Netzes ermittelt wird; dadurch gekennzeichnet, dass geprüft wird, ob die erste Aktivierung und die zweite Aktivierung des mindestens einen Neurons Null betragen und die zweite Netzeingabe des mindestens einen Neurons dem Schwellenwert näherkommt als die erste Netzeingabe des mindestens einen Neurons, und in dem Fall in Abhängigkeit von einer Differenz zwischen der ersten Netzeingabe des mindestens einen Neurons und dem Schwellwert, von einer Differenz zwischen der zweiten Netzeingabe des mindestens einen Neurons und dem Schwellwert und von einer zwischen der ersten Messung und der zweiten Messung verstrichenen Zeit eine Zeit extrapoliert wird, bis die Netzeingabe des mindestens einen Neurons den Schwellenwert erreicht, und

geprüft wird, ob die dritte Messung vor Ablauf der extrapolierten Zeit erfolgte, und in dem Fall

als dritte Aktivierung des mindestens einen Neurons ohne Anwendung der Aktivierungsfunktion des mindestens einen Neurons der Wert Null ausgegeben wird.

2. Verfahren nach dem vorhergehenden Anspruch; dadurch gekennzeichnet, dass geprüft wird, ob die erste Aktivierung und die zweite Aktivierung des mindestens einen Neurons Null betragen, ob die zweite Netzeingabe des mindestens einen Neurons weiter von dem Schwellenwert entfernt ist als die erste Netzeingabe des mindestens einen Neurons, und ob die dritte Messung vor Ablauf einer vordefinierten Zeit erfolgte, und in dem Fall

als dritte Aktivierung des mindestens einen Neurons ohne Anwendung der Aktivierungsfunktion des mindestens einen Neurons der Wert Null gewählt wird.

3. Verfahren nach einem der vorhergehenden Ansprüche; dadurch gekennzeichnet, dass

die Zeit, bis die Netzeingabe des mindestens einen Neurons den Schwellenwert erreicht, durch lineare Extrapolation ermittelt wird.

4. Verfahren nach einem der vorhergehenden Ansprüche; dadurch gekennzeichnet, dass

ein oder mehrere der Eingangssignale des Netzes Werte physikalischer Größen eines Fahrzeugs repräsentieren.

5. Verfahren nach einem der vorhergehenden Ansprüche; dadurch gekennzeichnet, dass

es sich bei ein oder mehreren der Eingangssignalen des Netzes um Signale von Sensoren eines/des Fahrzeugs handelt.

6. Verfahren nach einem der vorhergehenden Ansprüche; dadurch gekennzeichnet, dass

ein oder mehrere der Ausgangssignale Werte physikalischer Größen eines/des Fahrzeugs repräsentieren.

7. Verfahren nach einem der vorhergehenden Ansprüche; dadurch gekennzeichnet, dass
es sich bei ein oder mehreren der Ausgangssignale um Steuersignale für Aktoren eines/des Fahrzeugs handelt.
8. Computerprogram mit Anweisungen zur Ausführung eines Verfahrens einem der vorhergehenden Ansprüche.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2023/076715

A. CLASSIFICATION OF SUBJECT MATTER G06N 3/048 (2023.01)i; G06N 3/02 (2006.01)i; G06N 3/084 (2023.01)i According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G06N Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Pinto Dennis ET AL. "Mixture-of-Rookies: Saving DNN Computations by Predicting ReLU Outputs" <i>arXiv (Cornell University)</i> , Ithaca, 10 February 2022 (2022-02-10), Retrieved from the Internet: https://arxiv.org/pdf/2202.04990v1.pdf [retrieved on 2023-12-14] DOI: 10.48550/arxiv.2202.04990 XP093112461 title sections 3.2, 4.1;abstract	1-8
A	Bojarski Mariusz ET AL. "Explaining How a Deep Neural Network Trained with End-to-End Learning Steers a Car" <i>arXiv (Cornell University)</i> , Ithaca, 25 April 2017 (2017-04-25), Retrieved from the Internet: https://arxiv.org/pdf/1704.07911.pdf [retrieved on 2023-12-14] DOI: 10.48550/arxiv.1704.07911 XP093112460 abstract; figures 1,5	1-8
☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 15 December 2023		Date of mailing of the international search report 05 January 2024
Name and mailing address of the ISA/EP European Patent Office p.b. 5818, Patentlaan 2, 2280 HV Rijswijk Netherlands Telephone No. (+31-70)340-2040 Facsimile No. (+31-70)340-3016		Authorized officer Thielemann, Benedikt Telephone No.

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES INV. G06N3/048 G06N3/02 G06N3/084 ADD.		
Nach der Internationalen Patentklassifikation (IPC) oder nach der nationalen Klassifikation und der IPC		
B. RECHERCHIERTE GEBIETE		
Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole) G06N		
Recherchierte, aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen		
Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe) EPO-Internal		
C. ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	Pinto Dennis ET AL: "Mixture-of-Rookies: Saving DNN Computations by Predicting ReLU Outputs", arXiv (Cornell University), 10. Februar 2022 (2022-02-10), XP093112461, Ithaca DOI: 10.48550/arxiv.2202.04990 Gefunden im Internet: URL:https://arxiv.org/pdf/2202.04990v1.pdf [gefunden am 2023-12-14] Titel Abschnitte 3.2, 4.1; Zusammenfassung ----- -/-	1-8
☒ Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen ☐ Siehe Anhang Patentfamilie		
* Besondere Kategorien von angegebenen Veröffentlichungen : "A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist "E" frühere Anmeldung oder Patent, die bzw. das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist "L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt) "O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht "P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist "T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist "X" Veröffentlichung von besonderer Bedeutung:: die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden "Y" Veröffentlichung von besonderer Bedeutung:: die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist "&" Veröffentlichung, die Mitglied derselben Patentfamilie ist		
Datum des Abschlusses der internationalen Recherche 15. Dezember 2023		Absendedatum des internationalen Recherchenberichts 05/01/2024
Name und Postanschrift der Internationalen Recherchenbehörde Europäisches Patentamt, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Bevollmächtigter Bediensteter Thielemann, Benedikt

C. (Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	Bojarski Mariusz ET AL: "Explaining How a Deep Neural Network Trained with End-to-End Learning Steers a Car", arXiv (Cornell University), 25. April 2017 (2017-04-25), XP093112460, Ithaca DOI: 10.48550/arxiv.1704.07911 Gefunden im Internet: URL: https://arxiv.org/pdf/1704.07911.pdf [gefunden am 2023-12-14] Zusammenfassung; Abbildungen 1,5 -----	1-8