

(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
24. Oktober 2019 (24.10.2019)



(10) Internationale Veröffentlichungsnummer
WO 2019/201586 A1

(51) Internationale Patentklassifikation:
G05B 13/02 (2006.01) G06F 11/00 (2006.01)
G06N 3/063 (2006.01)

(21) Internationales Aktenzeichen: PCT/EP2019/058258

(22) Internationales Anmeldedatum:
02. April 2019 (02.04.2019)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
10 2018 205 869.5
18. April 2018 (18.04.2018) DE

(71) Anmelder: **ROBERT BOSCH GMBH** [DE/DE]; Post-
fach 30 02 20, 70442 Stuttgart (DE).

(72) Erfinder: **SCHORN, Christoph**; Roemerstrasse 175,
71229 Leonberg (DE). **TOPP, Jaroslaw**; Anna-Rauh-
Str. 26, 71272 Renningen (DE). **GEHRER, Stefan**; 2555
Smallman Street, Suite 300, Pittsburgh, Pennsylvania 15222
(US). **GAUERHOF, Lydia**; Jaegerstr. 25, 71272 Renning-
en (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, CZ, DE, DJ, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN,
KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD,
ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO,
NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW,

(54) Title: METHOD, COMPUTER SYSTEM AND COMPUTER PROGRAM FOR CONTROLLING AN ACTUATOR

(54) Bezeichnung: VERFAHREN, COMPUTERSYSTEM UND COMPUTERPROGRAMM ZUM STEUERN EINES AKTORS

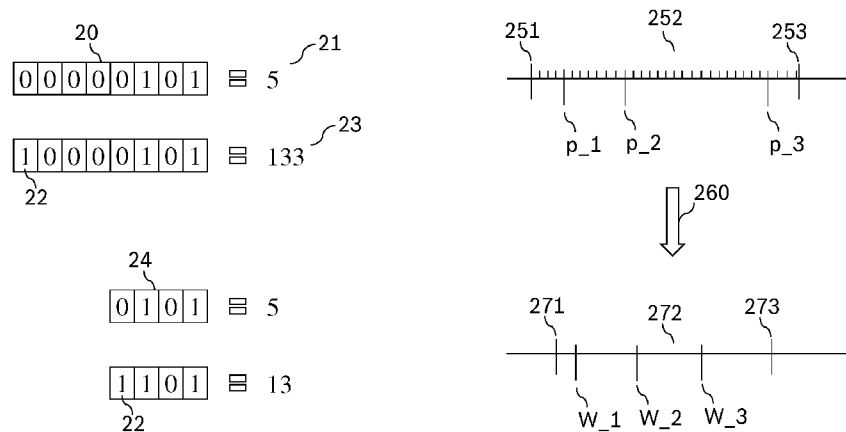


Fig. 2

(57) Abstract: The invention relates to a method (40) for controlling an actuator (10), comprising the steps: - mapping each parameter of a trained machine learning system (12), each parameter having a magnitude from a first set of different possible magnitudes (p_1, p_2, p_3), to a magnitude of at least one specifiable second set of different possible magnitudes (w_1, w_2, w_3); - storing the converted parameters in respective memory blocks; - determining an output variable of the machine learning system in dependence on an input variable and the stored parameters, the stored parameters each being read from the memory block in question by means of at least one mask; controlling the actuator (10) in dependence on the determined output variable. The invention further relates to a computer system, a computer program and a machine-readable memory element, in which the computer program is stored.



WO 2019/201586 A1

SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) Bestimmungsstaaten** (soweit nicht anders angegeben, für jede verfügbare regionale Schutzrechtsart): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, 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, 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)

(57) Zusammenfassung: Die Erfindung betrifft ein Verfahren (40) zum Steuern eines Aktors (10) mit den Schritten: Abbilden von Parametern eines angelernten maschinellen Lernsystems (12), die jeweils eine Größe aus einer ersten Menge unterschiedlicher möglicher Größen (p₁, p₂, p₃) aufweisen, jeweils auf eine Größe zumindest einer vorgebbaren zweiten Menge unterschiedlicher möglicher Größen (w₁, w₂, w₃). Hinterlegen der ungewandelten Parameter jeweils in einem Speicherblock. Ermitteln einer Ausgangsgröße des maschinellen Lernsystems abhängig von einer Eingangsgröße und den hinterlegten Parametern. Die hinterlegten Parameter werden jeweils mittels zumindest einer Maske aus dem jeweiligen Speicherblock ausgelesen. Ansteuern des Aktors (10) abhängig von der ermittelten Ausgangsgröße. Die Erfindung betrifft ferner ein Computersystem, Computerprogramm und ein maschinenlesbares Speicherelement, auf dem das Computerprogramm gespeichert ist.

5 Beschreibung

Titel

Verfahren, Computersystem und Computerprogramm zum Steuern eines Aktors

10 Technisches Gebiet

Die Erfindung betrifft ein Verfahren zum Steuern eines Aktors abhängig von einer Ausgangsgröße eines maschinellen Lernsystems. Ebenso betrifft die Erfindung ein Computersystem zum Steuern des Aktors und ein Computerprogramm, das
15 eingerichtet ist, das Verfahren auszuführen.

Stand der Technik

Die DE 10 2011 087 791 A1 offenbart ein Verfahren in einem
20 Manöverassistenzsystem in Fahrzeugen sowie entsprechende Manöverassistenzsysteme. Das Verfahren umfasst die Schritte: Erkennen einer Situation, gemäß der ein Fahrer des Fahrzeuges in Bezug auf ein Objekt eine manuelle Handlung vornehmen will, sowie Unterstützen bei der Ausrichtung des Fahrzeuges in Bezug auf das Objekt.

25

Vorteile der Erfindung

Während des Betriebes eines neuronalen Netzes können Signalstörungen innerhalb des neuronalen Netzes auftreten, die zu einer Falschklassifikation
30 führen können. Dies kann bei sicherheitskritischen Anwendungen schwerwiegende Folgen haben, beispielsweise wenn die Anwesenheit eines Objekts in der Verkehrsszene mittels des durch eine Signalstörung beeinträchtigten neuronalen Netzes nicht realitätsgemäß detektiert werden kann. Der Gegenstand der unabhängigen Ansprüche hat demgegenüber den Vorteil,
35 dass Parameter eines tiefen neuronalen Netzes mittels Masken aus einem Speicher ausgelesen werden, wodurch Störungen, insbesondere Bit-Kipper,

innerhalb des Speichers eine geringere Auswirkung auf den Betrieb des tiefen neuronalen Netzes haben können.

Offenbarung der Erfindung

5

In einem ersten Aspekt wird ein Verfahren zum Steuern eines Aktors gemäß des unabhängigen Anspruchs 1 vorgestellt. Das Verfahren umfasst unter anderem die Schritte mit folgenden Merkmalen:

10

- Abbilden von Parametern eines angelernten maschinellen Lernsystems. Die Parameter, die jeweils eine Größe einer ersten Menge unterschiedlicher möglicher Größen aufweisen, werden jeweils auf eine Größe zumindest einer vorgebbaren zweiten Menge unterschiedlicher möglicher Größen abgebildet. Die zweite Menge unterschiedlicher möglicher Größen weist weniger Größen als die erste Menge unterschiedlicher möglicher Größen auf.

15

- Hinterlegen der abgebildeten Parameter jeweils in einem Speicherblock.
- Ermitteln einer Ausgangsgröße des maschinellen Lernsystems abhängig von einer Eingangsgröße und den hinterlegten Parametern. Die hinterlegten Parameter werden mittels einer Maske aus dem jeweiligen Speicherblock ausgelesen.

20

Ein angelerntes maschinelles Lernsystem ist ein maschinelles Lernsystem, welches unter Verwendung vorgegebbarer Trainingsdaten angelernt wurde, eine vorgebbare Aufgabe zu lösen, beispielsweise ein Bild zu segmentieren oder eine Eingangsgröße zu klassifizieren.

25

30

35

Ein Vorteil dieses Aspektes ist, dass durch eine Reduzierung der Anzahl der möglichen unterschiedlichen Größen, z.B. ein Wert oder ein zumindest eindimensionaler Vektor, eine Reduktion des Speicheraufwands zum Speichern der Parameter erzielt werden kann. Ferner kann einhergehend mit der Reduktion der Anzahl unterschiedlicher Größen auch die Anzahl der benötigten Bits zur Darstellung der unterschiedlichen Größen reduziert werden. Daher werden nicht mehr alle Bits eines Registers benötigt, um die jeweilige Größe in dem Register zu hinterlegen. Nur die benötigten Bits zur Darstellung der hinterlegten Größe im Speicherblock werden mittels der Maske ausgelesen, wodurch z.B. Bit-Kipper, die bei den nicht benötigten Bits zur Darstellung der hinterlegten Größe aufgetreten sein könnten, durch das maskierte Auslesen des Registers nicht den

ursprünglich hinterlegten Parameter verändern. Dies hat den vorteilhaften Effekt, dass eine Erhöhung der Robustheit des maschinellen Lernsystems gegenüber hardwareseitigen Störungen erreicht werden kann, die beispielhaft durch hochenergetische Strahlung entstehen können.

5

Ein weiterer Vorteil ist, dass durch das Speichern eines Parameters jeweils in einem Speicherblock im Falle eines Kontrollflussfehlers nur ein Parameter falsch ausgelesen wird. Unter einem Kontrollflussfehler wird im Folgenden verstanden, dass ein falsches Register ausgelesen wird. Ferner können durch die reduzierte Anzahl an unterschiedlichen Größen weniger unterschiedliche Fehler auftreten, wodurch die mögliche Anzahl unterschiedlicher Fehler eingeschränkt werden kann.

10

Vorteilhaft ist, wenn eine maximale Größe der ersten Menge unterschiedlicher möglicher Größen größer als eine maximale Größe der zweiten Menge unterschiedlicher möglicher Größen ist. Der Vorteil hierbei ist, dass durch kleinere maximale Größen im Falle einer Bit-Kipper Störung einer der Bits die benötigt werden, die maximale Größe darzustellen, der absolute maximale Fehler kleiner wird und dass eine kleinere maximale Größe mit weniger Bit im Speicher hinterlegt werden kann.

15

20

Weiterhin vorteilhaft ist, wenn während des Ermitteln der Ausgangsgröße des maschinellen Lernsystems ermittelte Größen, insbesondere Zwischengrößen, daraufhin geprüft werden, ob diese größer als eine Maximalgröße sind. Wenn beim Ermitteln der Ausgangsgröße des maschinellen Lernsystems eine Größe des maschinellen Lernsystems größer als eine Maximalgröße ist, wird die Ausgangsgröße des maschinellen Lernsystems als fehlerhaft gekennzeichnet. Dadurch können Störungen oder Fehler frühzeitig detektiert werden.

25

30

Das Verfahren des ersten Aspektes kann in einem physikalischen/technischen System, z.B. einem Roboter oder einem zumindest teilautonomen Fahrzeug eingesetzt werden, welches einen Aktor aufweist. Die ermittelte Ausgangsgröße kann von einer Steuereinheit verwendet werden, sodass die Steuereinheit abhängig von dieser Ausgangsgröße den Aktor steuert. Die Steuereinheit kann hierfür eine Steuergröße abhängig von der ermittelten Ausgangsgröße bestimmen und diese zum Steuern des Aktors verwenden.

35

In einem weiteren Aspekt wird ein Computersystem, insbesondere zum Steuern des Aktors vorgestellt. Das Computersystem umfasst wenigstens einen Prozessor, der mit einem Speichermodul verbunden ist. Auf dem Speichermodul ist ein Computerprogramm hinterlegt und zumindest Parameter eines maschinellen Lernsystems sind in Speicherblöcken des Speichermoduls, insbesondere Registern, hinterlegt. Unter anderem ist das Computersystem eingerichtet, die Parameter jeweils mittels zumindest einer Maske aus dem jeweiligen Speicherblock auszulesen.

Der Vorteil ist, dass mittels der Maske ausschließlich die Bits zur Darstellung des Parameters aus dem Speicher ausgelesen werden, wodurch Bits, die nicht zur Darstellung des Parameters benötigt werden, durch die Maskierung nicht ausgelesen werden und damit auch keine Störungen dieser Bits auftreten.

Vorteilhaft ist, wenn eine vorgebbare Anzahl der Speicherblöcke jeweils in eine Mehrzahl von Abschnitten aufgeteilt sind und in einem jeden Abschnitt jeweils einer der Parameter hinterlegt ist. Die Maske berücksichtigt nur diejenigen Bits des jeweiligen Abschnitts, die benötigt werden, um eine maximale Größe der zweiten vorgebbaren Menge darzustellen oder die Maske berücksichtigt nur diejenigen Bits des jeweiligen Speicherblocks, die in jedem Abschnitt benötigt werden, um die maximale Größe darzustellen. Dies hat den Vorteil, dass unter Beibehalten des Vorteils der Maskierung der Speicher effizient genutzt wird.

Weiterhin vorteilhaft ist, wenn das maschinelle Lernsystem ein tiefes neuronales Netz umfasst. Die Größen derjenigen Parameter, die einer Schicht, insbesondere Neuron, des tiefen neuronalen Netz zugeordnet sind, wurden jeweils auf eine der Größen der jeweiligen zugeordneten Menge unterschiedlicher möglicher Größen dieser Schicht, insbesondere Neuron, abgebildet. Das Computersystem ist des Weiteren eingerichtet, jeweils unterschiedliche Masken zum Auslesen der Parameter der jeweiligen Schichten, insbesondere Neuronen, zu verwenden.

Der Vorteil hierbei ist, dass die Maske flexibel für die jeweils unterschiedlichen Schichten, bzw. Neuronen angepasst werden kann. Damit kann eine höhere Genauigkeit und/oder ein größerer Wertebereich durch eine entsprechende Maske für die jeweilige Schicht bzw. Neuron bereitgestellt werden.

In einem weiteren Aspekt wird ein Computerprogramm, welches eingerichtet ist, eines der vorherigen genannten Verfahren auszuführen vorgeschlagen. Das Computerprogramm umfasst Anweisungen, die einen Computer oder das Computersystem des weiteren Aspektes veranlassen, eines dieser genannten Verfahren mit all seinen Schritten auszuführen, wenn das Computerprogramm auf dem Computer/-system abläuft. Ferner wird ein maschinenlesbares Speichermodul vorgeschlagen, auf welchem das Computerprogramm gespeichert ist.

Ausführungsbeispiele der oben genannten Aspekte sind in den beiliegenden Zeichnungen dargestellt und in der nachfolgenden Beschreibung näher erläutert. Dabei zeigen:

Kurze Beschreibung der Zeichnungen

- Fig. 1 eine schematische Darstellung eines zumindest teilautonomen Fahrzeuges;
- Fig. 2 eine schematische Darstellung eines hardwarenahen Zahlendarstellungsformates ohne und mit einer Bit-Kipper Störung und einer Abbildung von Werten eines ersten Wertebereichs auf einen zweiten Wertebereich;
- Fig. 3 eine schematische Darstellung einer Ausführungsform eines Computersystems;
- Fig. 4 eine schematische Darstellung einer Ausführungsform des Verfahrens zum Steuern eines Aktors.

Figur 1 zeigt eine schematische Darstellung eines zumindest teilautonomen Roboters, der im Ausführungsbeispiel gegeben ist durch ein zumindest teilautonomes Fahrzeug (10). In einem weiteren Ausführungsbeispiel kann der zumindest teilautonome Roboter ein Service-, Montage- oder stationärer Produktionsroboter sein. Das zumindest teilautonome Fahrzeug (10) kann eine Erfassungseinheit (11) umfassen. Die Erfassungseinheit (11) kann zum Beispiel eine Kamera sein, welche eine Umgebung des Fahrzeugs (10) erfasst. Die Erfassungseinheit (11) kann mit einem maschinellen Lernsystem (12) verbunden

sein. Das maschinelle Lernsystem (12) ermittelt abhängig von einer bereitgestellten Eingangsgröße, z.B. mittels der Erfassungseinheit (11), und in Abhängigkeit einer Mehrzahl von das maschinelle Lernsystem charakterisierenden Größen eine Ausgangsgröße. Die Ausgangsgröße kann an eine Aktorsteuerungseinheit (13) weitergeleitet werden. Die Aktorsteuerungseinheit (13) kann in Abhängigkeit der Ausgangsgröße des maschinellen Lernsystems (12) einen Aktor ansteuern. Der Aktor kann in diesem Ausführungsbeispiel ein Motor des Fahrzeugs (10) sein.

In einem weiteren Ausführungsbeispiel umfasst die Aktorsteuerungseinheit (13) ein Freigabesystem. Das Freigabesystem entscheidet, ob ein Objekt, z.B. ein erfasster Roboter oder eine erfasste Person, Zugang zu einem Bereich hat, abhängig von der Ausgangsgröße des maschinellen Lernsystems (12). Vorzugsweise kann der Aktor abhängig von einer Entscheidung des Freigabesystems angesteuert werden.

Ferner umfasst das Fahrzeug (10) eine Recheneinheit (14) und ein maschinenlesbares Speicherelement (15). Auf dem Speicherelement (15) kann ein Computerprogramm gespeichert sein, welches Befehle umfasst, die beim Ausführen der Befehle auf der Recheneinheit (14) dazu führen, dass die Recheneinheit (14) das erfindungsgemäße Verfahren ausführt. Denkbar ist auch, dass ein Downloadprodukt oder ein künstlich generiertes Signal, die jeweils das Computerprogramm umfassen können, nach Empfangen an einem Empfänger des Fahrzeugs (10) die Recheneinheit (14) veranlassen, das erfindungsgemäße Verfahren auszuführen.

In einer weiteren Ausführungsform wird das maschinelle Lernsystem (12) in einem Messsystem verwendet, welches nicht in den Figuren dargestellt ist. Das Messsystem unterscheidet sich zu dem Roboter nach Figur 1 dahingehend, dass das Messsystem keine Aktorsteuerungseinheit (13) umfasst. Das Messsystem kann die Ausgangsgröße des maschinellen Lernsystems (12) statt sie an die Aktorsteuerungseinheit (13) zu weiterzuleiten, abspeichern oder darstellen, beispielsweise mittels visueller oder auditiver Darstellungen.

Es ist auch denkbar, dass in einer Weiterentwicklung des Messsystems die Erfassungseinheit (11) ein Bild eines menschlichen oder tierischen Körpers oder

eines Teils davon erfasst. Beispielsweise kann dies mittels eines optischen Signals, mittels eines Ultraschallsignals, oder mittels eines MRT/CT-Verfahrens erfolgen. Das Messsystem kann in dieser Weiterentwicklung ein maschinelle Lernsystem umfassen, das derart angelern ist, abhängig von der Eingangsgröße eine Klassifikation auszugeben, z.B. welches Krankheitsbild auf Basis diese Eingangsgröße möglicherweise vorliegt.

Das maschinelle Lernsystem (12) kann ein tiefes neuronales Netz umfassen.

Figur 2 zeigt eine schematische Darstellung eines hardwarenahen Zahlendarstellungsformats (20,24). Ein erstes hardwarenahes Zahlendarstellungsformat (20), ein Binärsystem mit 8 Bit, stellt einen Wert 5 (21) mittels acht unterschiedlichen Stellenwerten dar. Eine Störung, beispielsweise ein Bit-Kipper, kann in einem Register, in welchem der Wert 5 (21) mit dem ersten hardwarenahen Zahlendarstellungsformat (20) hinterlegt ist, dazu führen, dass einer der Stellenwerte (22) verändert wird. Diese Bit-Kipper-Störung ist beispielhaft in Figur 2 dargestellt, in dem im ersten hardwarenahen Zahlendarstellungsformat (20) der höchstwertige Stellenwert (22) invertiert wird. Diese Störung kann dazu führen, dass anstatt dem Wert 5 (21) nun der Wert 133 (23) in dem Register hinterlegt ist.

In einem zweiten hardwarenahen Zahlendarstellungsformat (24), beispielhaft ein Binärsystem mit 4 Bit, führt eine Störung des höchstwertigen Stellenwertes (22) dazu, dass anstatt dem Wert 5 nun der Wert 13 hinterlegt ist.

Anhand dieses Beispiels wird ersichtlich, dass eine Einschränkung des maximalen erlaubten Wertebereichs des hardwarenahen Zahlendarstellungsformats und die damit verbundene Reduktion der benötigten Bits, die zur Darstellung des eingeschränkten Wertebereichs benötigt werden, zu einer Reduktion des maximalen Fehlers einer Bit-Kipper-Störung führen kann. Daher ist es vorteilhaft, wenn Parameter des maschinellen Lernsystems mit wenig Bits dargestellt werden, wodurch Bit-Kipper zu einer geringeren Veränderung des Wertes, der mit diesen Bits dargestellt wird, führen.

In Figur 2 sind des Weiteren unterschiedliche Wertebereiche (252,272) von Parametern des maschinellen Lernsystems schematisch dargestellt. Die Parameter des angelerten tiefen neuronalen Netzes weisen jeweils Werte aus einem ersten

Wertebereich (252) auf. Der erste Wertebereich (252) umfasst eine erste Menge unterschiedlicher möglicher Werte (p_1, p_2, p_3), die einen minimalen Wert (251) und einen maximalen Wert (253) umfassen kann.

5 Die Werte der Parameter werden mittels einer vorgebbaren Abbildung (260) derart auf einen zweiten Wertebereich (272) abgebildet, dass die Werte mit weniger Bits dargestellt werden können. Der zweite Wertebereich (272) umfasst eine vorgebbare Menge unterschiedlicher möglicher Werte (w_1, w_2, w_3) und kann sich zwischen zwei vorgebbaren Werten (271, 273) erstrecken.

10

Bevorzugt ist ein maximaler Wert (p_3) der Menge unterschiedlicher möglicher Werte (p_1, p_2, p_3) des ersten Wertebereichs (252) größer als ein maximaler Wert (w_3) der Menge unterschiedlicher möglicher Werte Menge unterschiedlicher möglicher Werte (w_1, w_2, w_3) des zweiten Wertebereichs (272).

15

In einer weiteren Ausführungsform ist eine Differenz zwischen zwei, insbesondere unmittelbar, aufeinanderfolgenden Werten (w_2, w_3) des zweiten Wertebereichs (272) kleiner als eine Differenz zwischen zwei, insbesondere unmittelbar, aufeinanderfolgenden Werten (p_2, p_3) des ersten Wertebereichs (252), die auf diese zwei aufeinanderfolgende Werte (w_2, w_3) abgebildet werden. Durch die kleine Differenz zwischen zwei unterschiedlichen Werten kann beim Auftreten des Bit-Kippers der Unterschied zu dem ursprünglichen Wert klein sein, sodass der daraus resultierende Fehler einen kleinen Einfluss auf das maschinelle Lernsystem hat.

20

25 In einem alternativen Ausführungsbeispiel kann das maschinelle Lernsystem (12) ein faltendes neuronales Netz (engl. Convolutional Neural Network) umfassen. Die Parameter können dann auch Filter-Kernels sein. Koeffizienten der Filter-Kernels können in diesem Ausführungsbeispiel von einem ersten Wertebereich (252) auf einen zweiten Wertebereich (272) abgebildet werden. Alternativ oder zusätzlich können auch ganze Filter-Kernels auf vorgebbare Filter-Kernels abgebildet werden, die
30 beispielsweise mit einer reduzierten Anzahl von Bits darstellbar sind.

30

Figur 3 zeigt eine schematische Darstellung eines Computersystems (30), das bevorzugt eingerichtet ist, das maschinelle Lernsystem zu betreiben, insbesondere
35 eingerichtet ist, einen Aktor abhängig von einer Ausgangsgröße des maschinellen

35

Lernsystems (12) zu steuern. Das Computersystem (30) umfasst wenigstens eine Recheneinheit (31), welche mit einem Speicher (32) verbunden ist. In dem Speicher (32) sind eine Software und Parameter eines maschinellen Lernsystems hinterlegt. Zusätzlich können auch weitere Informationen über das maschinelle Lernsystem, beispielsweise dessen Architektur, in dem Speicher (32) hinterlegt sein.

Figur 4 zeigt eine schematische Darstellung eines Verfahrens (40) zum Steuern des Aktors.

Das Verfahren (40) beginnt mit Schritt 41. In Schritt 41 wird ein angeleitetes maschinellen Lernsystem (12) bereitgestellt. Die Parameter des angeleitetes maschinellen Lernsystems weisen Werte aus dem ersten Wertebereich (252) auf. Diese Werte werden anschließend in Schritt 42 auf einen zweiten Wertebereich (272) abgebildet. Das Abbilden der Werte der Parameter kann derart durchgeführt werden, dass die Werte jeweils auf einen Wert einer vorgebbaren Menge unterschiedlicher möglicher Werte (w_1, w_2, w_3) abgebildet werden. Beispielsweise kann ein Wert des Parameters des angeleitetes maschinellen Lernsystems, der einem der unterschiedlicher möglicher Werte (w_1, w_2, w_3) der vorgebbaren Menge nach einem mathematischen Distanzmaß am nächsten liegt, auf diesen abgebildet werden.

Die vorgebbare Menge unterschiedlicher möglicher Werte (w_1, w_2, w_3) kann mittels einer Clusteranalyse (z.B. k-Means-Algorithmus) der Werte der Parameter des bereitgestellten maschinellen Lernsystems (12) ermittelt werden.

Alternativ kann auch eine Mehrzahl von vorgebbaren Mengen unterschiedlicher möglicher Werte (w_1, w_2, w_3) bereitgestellt werden, z.B. mit jeweils unterschiedlichen Quantisierungsstufen. Mittels einer Kullback-Leibler (KL)-Divergenz, die jeweils zwischen einer der vorgebbaren Mengen und der Verteilung der Werte der Parameter des bereitgestellten maschinellen Lernsystems ermittelt wird, kann aus der Mehrzahl vorgebbarer Mengen unterschiedlicher möglicher Werte (w_1, w_2, w_3), die Menge mit der geringsten ermittelten KL-Divergenz ausgewählt werden. Die KL-Divergenz charakterisiert einen Informationsverlust, der beim Abbilden der Werte auf einen kleineren Wertebereich auftritt. Diejenige Menge, die die geringste KL-Distanz aufweist, kann anschließend verwendet werden, um die Werte der Parameter des bereitgestellten maschinellen Lernsystem auf Werte dieser Menge abzubilden.

In einem weiteren Ausführungsbeispiel, kann der maximale Wert (w_3) der unterschiedlichen möglichen Werte (w_1, w_2, w_3) vorgebar sein, bevorzugt kleiner als der maximale mögliche Wert der Parameter des bereitgestellten maschinellen Lernsystems, um den Wertebereich der vorgebbaren Menge unterschiedlichen möglichen Werte (w_1, w_2, w_3) weiter einzuschränken. Der maximale Wert (w_3) kann ebenfalls in der Ermittlung der KL-Divergenz berücksichtigt werden, um einen geeigneten maximalen Wert (w_3) zu bestimmen, der zu einem geringen Informationsverlust führt.

Nachdem die Parameter des angelernten maschinellen Lernsystems auf den zweiten Wertebereich (272) abgebildet wurden folgt Schritt 43. In Schritt 43 werden die abgebildeten Parameter jeweils in einem Speicherelement des Speichers (32) hinterlegt.

In dem nachfolgenden Schritt 44 wird dem maschinellen Lernsystem eine Eingangsgröße bereitgestellt und abhängig von der Eingangsgröße wird eine Ausgangsgröße des maschinellen Lernsystems ermittelt. Die Eingangsgröße kann beispielsweise eine Videoaufnahme oder ein Kamerabild der Umgebung des Aktors sein. Das maschinelle Lernsystem ermittelt abhängig von der bereitgestellten Eingangsgröße und den hinterlegten Parametern eine Ausgangsgröße. Dabei werden die hinterlegten Parameter aus dem Speicher mittels einer Maske ausgelesen. Die Maske liest dabei nur diejenigen Bits des Speicherelements aus, die benötigt werden um den maximalen Wert (w_3) der unterschiedlichen möglichen Werte (w_1, w_2, w_3) darzustellen.

Die Maske kann in Software implementiert sein, beispielsweise durch eine Bitfolge, die aus Nullen und Einsen besteht. Die Bitfolge wird mit den Bits des Speicherelements logisch AND verknüpft. Alternativ kann die Maske in Hardware implementiert sein.

Wenn beispielsweise das Register 8 Bit umfasst und der maximale Wert (w_3) der unterschiedlichen möglichen Werte (w_1, w_2, w_3) durch die hinteren 4 Bit dargestellt werden kann, können die nicht benötigten vorderen 4 Bit durch die Maske logisch AND mit einer 0 verknüpft werden, wodurch Bit-Kipper-Störungen dieser Bits nicht zu einem falschen ausgelesenen Parameter aus der dem Register führen. Damit könnte in Figur

2 dennoch die ursprünglich hinterlegte Zahl 5 ausgelesen werden, auch wenn in dem 8 Bit Register die Zahl 133 fälschlicherweise hinterlegt ist.

5 Nachdem die Ausgangsgröße des maschinellen Lernsystems ermittelt wurde, kann ein Aktor, zum Beispiel der Motor eines zumindest teilautonomen Fahrzeuges, mittels der Aktorsteuerungseinheit (13) angesteuert werden. Detektiert beispielsweise das maschinelle Lernsystem (12) eine Person, kann die Aktorsteuerungseinheit (13) den Aktor ansteuern, sodass dieser das Fahrzeug (10) abbremst.

10 Damit endet das Verfahren (40). Alternativ oder zusätzlich ist es denkbar, dass das Verfahren zyklisch mit Schritt 44 erneut durchgeführt wird, wenn zum Beispiel ein neues Kamerabild erfasst wurde oder ein Trigger diesbezüglich ausgelöst wurde.

15 Es versteht sich, dass das Verfahren nicht nur wie beschrieben vollständig in Software, sondern auch in Hardware oder in einer Mischform aus Software und Hardware implementiert sein kann.

5 Ansprüche

1. Verfahren (40) zum Steuern eines Aktors (10), umfassend die Schritte:
 - Abbilden von Parametern eines angelernten maschinellen Lernsystems (12), die
10 jeweils eine Größe aus einer ersten Menge unterschiedlicher möglicher Größen (p_1, p_2, p_3) aufweisen, jeweils auf eine Größe zumindest einer vorgebbaren zweiten Menge unterschiedlicher möglicher Größen (w_1, w_2, w_3),
wobei die zweite Menge unterschiedlicher möglicher Größen (w_1, w_2, w_3)
weniger Größen als die erste Menge unterschiedlicher möglicher Größen
15 (p_1, p_2, p_3) aufweist;
 - Hinterlegen der abgebildeten Parameter jeweils in einem Speicherblock;
 - Ermitteln eine Ausgangsgröße des maschinellen Lernsystems abhängig von einer
Eingangsgröße und den hinterlegten Parametern,
wobei die hinterlegten Parameter jeweils mittels zumindest einer Maske aus dem
20 jeweiligen Speicherblock ausgelesen werden und mittels der Maske nur diejenigen
Bits des jeweiligen Speicherblocks berücksichtigt werden, die benötigt werden, um
eine maximale Größe (w_3) der zweiten Menge unterschiedlicher möglicher Größen
(w_1, w_2, w_3) darzustellen; und
 - Ansteuern des Aktors (10) abhängig von der ermittelten Ausgangsgröße.
- 25 2. Verfahren nach Anspruch 1, wobei eine maximale Größe (p_3) der ersten Menge unterschiedlicher möglicher Größen (p_1, p_2, p_3) größer als die maximale Größe (w_3) der zweiten Menge unterschiedlicher möglicher Größen (w_1, w_2, w_3) ist.
- 30 3. Verfahren nach einem der vorherigen Ansprüche, wobei das maschinelle Lernsystem ein tiefes neuronales Netz umfasst,
wobei die Größen derjenigen Parameter, die einer Schicht, insbesondere Neuron,
des tiefen neuronalen Netz zugeordnet sind, jeweils auf eine der Größen der

jeweiligen zugeordneten Menge unterschiedlicher möglicher Größen (w_1, w_2, w_3) dieser Schicht, insbesondere Neuron, abgebildet werden,

wobei zum Auslesen der Parameter jeweils unterschiedliche Masken für die jeweiligen Schichten, insbesondere Neuronen, verwendet werden.

5

4. Verfahren nach einem der vorherigen Ansprüche, wobei eine vorgebbare Anzahl der Speicherblöcke jeweils in eine Mehrzahl von Abschnitten aufgeteilt sind und in einem jeden Abschnitt jeweils einer der Parameter hinterlegt ist,

10

wobei die Maske nur diejenigen Bits des jeweiligen Abschnitts berücksichtigt, die benötigt werden, um die zweite maximale Größe (w_3) darzustellen oder die Maske nur diejenigen Bits des jeweiligen Speicherblocks berücksichtigt, die in jedem Abschnitt benötigt werden, um die zweite maximale Größe (w_3) darzustellen.

15

5. Verfahren nach einem der vorherigen Ansprüche, wobei die Maske aus einer vorgebbaren Bitfolge besteht und diese Bitfolge die gleiche Anzahl von Bits wie das Speicherelement aufweist,

wobei beim Auslesen ein jedes Bit des Speicherelements mit wenigstens einem jeweiligen zugeordneten Bit der Maske logisch verknüpft wird.

20

6. Verfahren nach einem der vorherigen Ansprüche, wobei dann, wenn beim Ermitteln der Ausgangsgröße des maschinellen Lernsystems (12) eine Größe des maschinellen Lernsystems (12) größer als eine Maximalgröße ist, wird die Ausgangsgröße des maschinellen Lernsystems (12) als fehlerhaft gekennzeichnet.

25

7. Computersystem (30) umfassend wenigstens einen Prozessor (31), der mit einem Speichermodul (32) verbunden ist, wobei auf dem Speichermodul (32) ein Computerprogramm hinterlegt ist und zumindest Parameter eines maschinellen Lernsystems (12) in Speicherblöcken des Speichermoduls (32), insbesondere Registern, hinterlegt sind,

30

wobei das maschinelle Lernsystem (12) auf einem weiteren Computersystem angelernt wurde und nach dem Anlernen jeweils eine Größe der Parameter aus einer ersten Menge unterschiedlicher möglicher Größen (p_1, p_2, p_3) jeweils auf eine Größe zumindest einer vorgebbaren zweiten Menge unterschiedlicher möglicher Größen (w_1, w_2, w_3) abgebildet wurde,

35

wobei die zweite Menge unterschiedlicher möglicher Größen (w_1, w_2, w_3) weniger Größen als die erste Menge unterschiedlicher möglicher Größen (p_1, p_2, p_3) aufweist, und

wobei das Computersystem (30) eingerichtet ist, die Parameter jeweils mittels
5 zumindest einer Maske aus dem jeweiligen Speicherblock auszulesen und mittels der Maske nur diejenigen Bits des jeweiligen Speicherblocks zu berücksichtigen, die benötigt werden, um eine maximale Größe (w_3) der zweiten Menge unterschiedlicher möglicher Größen (w_1, w_2, w_3) darzustellen.

10 8. Computersystem nach Anspruch 7, wobei das maschinelle Lernsystem (12) ein tiefes neuronales Netz umfasst,

wobei die Größen derjenigen Parameter, die einer Schicht, insbesondere Neuron, des tiefen neuronalen Netz zugeordnet sind, jeweils auf eine der Größen der jeweiligen zugeordneten Menge unterschiedlicher möglicher Größen (w_1, w_2, w_3)
15 dieser Schicht, insbesondere Neuron, abgebildet wurden,

wobei das Computersystem des Weiteren eingerichtet ist, zum Auslesen der Parameter jeweils unterschiedliche Masken für die jeweiligen Schichten, insbesondere Neuronen, zu verwenden.

20 9. Computersystem nach Anspruch 7 oder 8, wobei eine maximale Größe (p_3) der ersten Menge unterschiedlicher möglicher Größen (p_1, p_2, p_3) größer als die maximale Größe (w_3) der zweiten Menge unterschiedlicher möglicher Größen (w_1, w_2, w_3) ist,

wobei eine vorgebbare Anzahl der Speicherblöcke jeweils in eine Mehrzahl
25 von Abschnitten aufgeteilt sind und in einem jeden Abschnitt jeweils einer der Parameter hinterlegt ist,

wobei die Maske nur diejenigen Bit des jeweiligen Abschnitts berücksichtigt, die benötigt werden, um die zweite maximale Größe (w_3) darzustellen oder die Maske nur diejenigen Bits des jeweiligen Speicherblocks berücksichtigt, die in jedem
30 Abschnitt benötigt werden, um die zweite maximale Größe (w_3) darzustellen.

10. Computerprogramm, welches Befehle umfasst, die beim Ausführen auf einem Computer diesen veranlassen, dass das Verfahren nach einem der Ansprüche 1 bis 6 auszuführen wird.

11. Maschinenlesbares Speicherelement (15), auf welchem das Computerprogramm nach Anspruch 10 gespeichert ist.

1/4

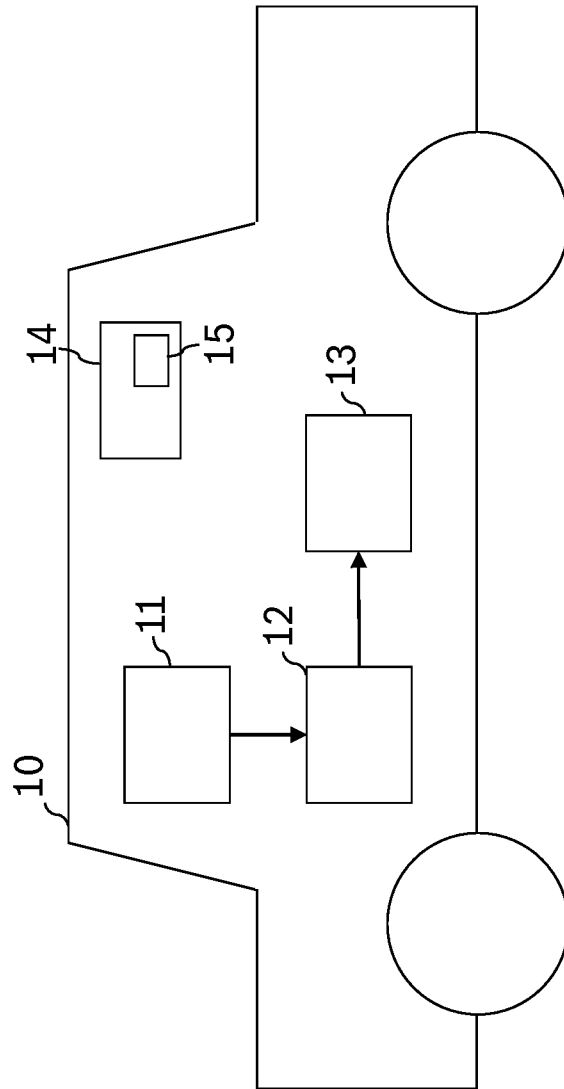


Fig. 1

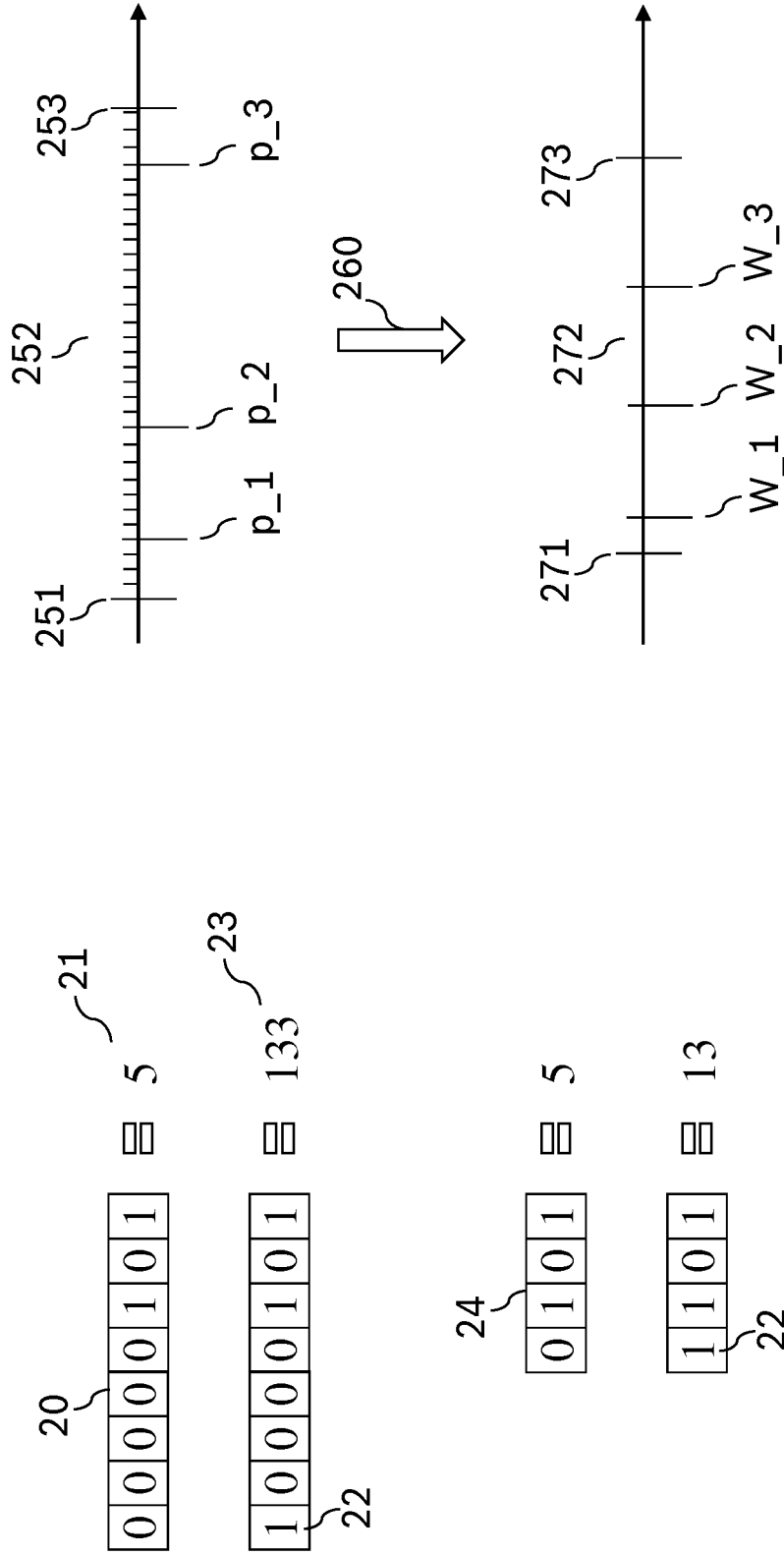


Fig. 2

3/4

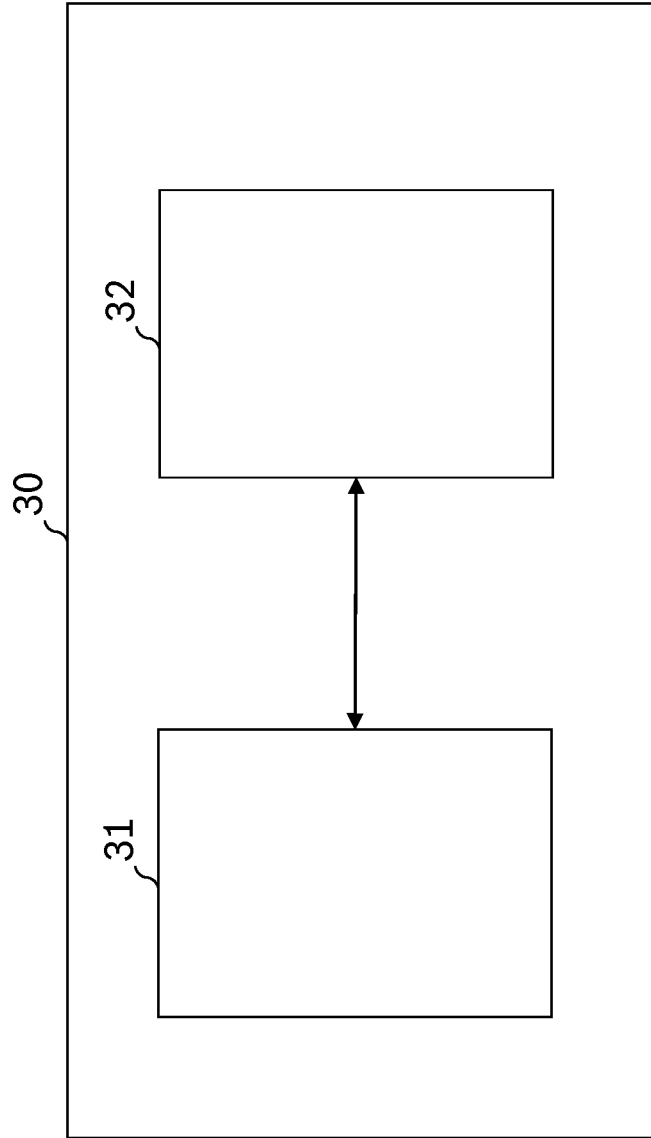


Fig. 3

4/4

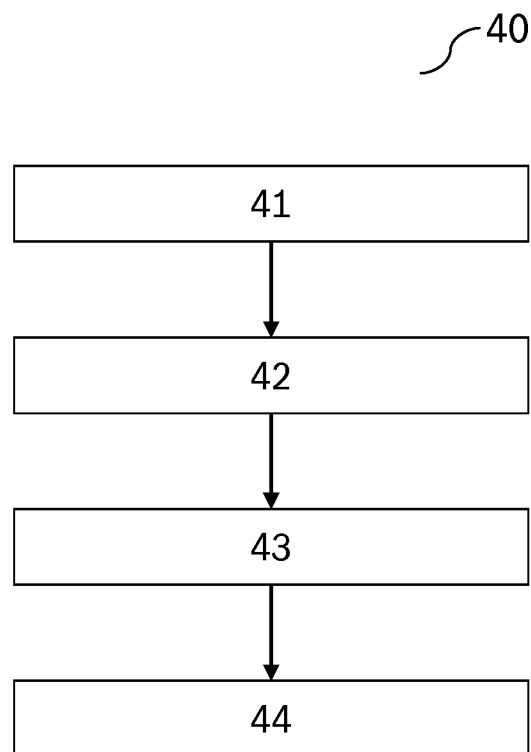


Fig. 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2019/058258

A. CLASSIFICATION OF SUBJECT MATTER		
<i>G05B 13/02</i> (2006.01)i; <i>G06N 3/063</i> (2006.01)i; <i>G06F 11/00</i> (2006.01)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) G05B; G06N; G06F		
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, WPI Data		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	VIVIENNE SZE ET AL. "Efficient Processing of Deep Neural Networks: A Tutorial and Survey" <i>ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 27 March 2017 (2017-03-27), DOI: 10.1109/JPROC.2017.2761740</i> XP080759895 abstract page 2316 - page 2318	1-11
X	SEBASTIAN VOGEL ET AL. "Efficient Stochastic Inference of Bitwise Deep Neural Networks" <i>ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 20 November 2016 (2016-11-20), XP080733355</i> pages 1-3	1-11
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> 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 01 July 2019		Date of mailing of the international search report 17 July 2019
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 Dörre, Thorsten Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2019/058258

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	REAGEN BRANDON ET AL. "Minerva: Enabling Low-Power, Highly-Accurate Deep Neural Network Accelerators" <i>2013 21ST INTERNATIONAL CONFERENCE ON PROGRAM COMPREHENSION (ICPC); [INTERNATIONAL SYMPOSIUM ON COMPUTER ARCHITECTURE.(ISCA)], IEEE, US,</i> 18 June 2016 (2016-06-18), pages 267-278, [retrieved on 2016-08-24] DOI: 10.1109/ISCA.2016.32 ISSN: 1063-6897, XP032950665 abstract page 271 page 274	1-11

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES INV. G05B13/02 G06N3/063 G06F11/00 ADD.		
Nach der Internationalen Patentklassifikation (IPC) oder nach der nationalen Klassifikation und der IPC		
B. RECHERCHIERTE GEBIETE Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole) G05B G06N G06F		
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, WPI Data		
C. ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	VIVienne SZE ET AL: "Efficient Processing of Deep Neural Networks: A Tutorial and Survey", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 27. März 2017 (2017-03-27), XP080759895, DOI: 10.1109/JPROC.2017.2761740 Zusammenfassung Seite 2316 - Seite 2318 ----- -/--	1-11
<input checked="" type="checkbox"/> Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen <input type="checkbox"/> 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	Absendedatum des internationalen Recherchenberichts	
1. Juli 2019	17/07/2019	
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 Dörre, Thorsten	

C. (Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	<p>SEBASTIAN VOGEL ET AL: "Efficient Stochastic Inference of Bitwise Deep Neural Networks", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 20. November 2016 (2016-11-20), XP080733355, Seiten 1-3</p> <p style="text-align: center;">-----</p>	1-11
X	<p>REAGEN BRANDON ET AL: "Minerva: Enabling Low-Power, Highly-Accurate Deep Neural Network Accelerators", 2013 21ST INTERNATIONAL CONFERENCE ON PROGRAM COMPREHENSION (ICPC); [INTERNATIONAL SYMPOSIUM ON COMPUTER ARCHITECTURE.(ISCA)], IEEE, US, 18. Juni 2016 (2016-06-18), Seiten 267-278, XP032950665, ISSN: 1063-6897, DOI: 10.1109/ISCA.2016.32 ISBN: 978-0-7695-3174-8 [gefunden am 2016-08-24] Zusammenfassung Seite 271 Seite 274</p> <p style="text-align: center;">-----</p>	1-11