(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
17. Februar 2022 (17.02.2022)

WIPO | PCT

(10) Internationale Veröffentlichungsnummer
WO 2022/033746 A1

(51) Internationale Patentklassifikation:
G06N 3/00 (2006.01) *B60W 30/18* (2012.01)
G06N 3/08 (2006.01)

(21) Internationales Aktenzeichen: PCT/EP2021/065919

(22) Internationales Anmeldedatum:
14. Juni 2021 (14.06.2021)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
10 2020 121 150.3
11. August 2020 (11.08.2020) DE

(71) Anmelder: **BAYERISCHE MOTOREN WERKE AKTIENGESELLSCHAFT** [DE/DE]; Petuelring 130, 80809 München (DE).

(72) Erfinder: **WERLING, Moritz**; Am Oberwiesenfeld 43, 80809 München (DE). **KALWEIT, Gabriel**; Blauenstraße 18, 79115 Freiburg (DE). **HÜGLE, Maria**; Gauchstraße 19, 79098 Freiburg (DE). **BÖDECKER, Joschka**; Schlehenrain 5, 79108 Freiburg (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, IT, 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, 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.

(54) **Title:** TRAINING A REINFORCEMENT LEARNING AGENT TO CONTROL AN AUTONOMOUS SYSTEM

(54) **Bezeichnung:** TRAINIEREN EINES REINFORCEMENT LEARNING AGENTEN ZUR STEUERUNG EINES AUTONOMEN SYSTEMS

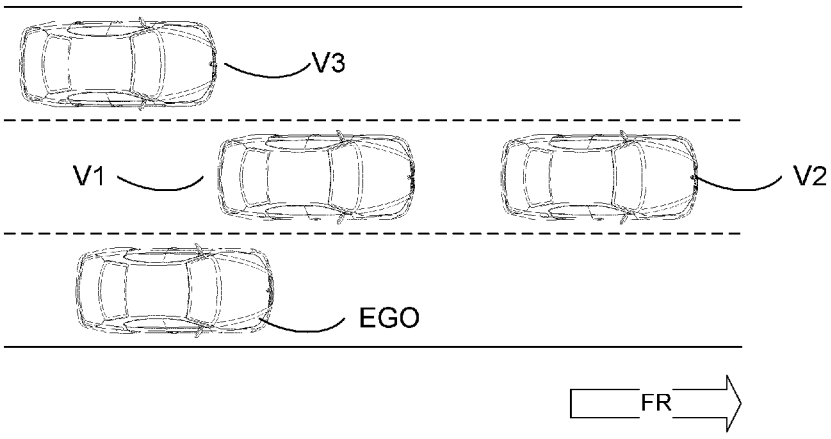


Fig. 1

FR = Direction of travel

(57) **Abstract:** One aspect of the invention relates to a device for training a reinforcement learning agent to control an autonomous system, wherein the device is designed to detect the environment of the autonomous system, to detect at least one object in the environment of the autonomous system that can be compared to the autonomous system, to detect a behaviour of the at least one object that can be compared to the autonomous system, and to train the reinforcement learning agent in accordance with the detected behaviour of the at least one object that can be compared to the autonomous system.

(57) **Zusammenfassung:** Ein Aspekt der Erfindung betrifft eine Vorrichtung zum Trainieren eines Reinforcement Learning Agenten zur Steuerung eines autonomen Systems, wobei die Vorrichtung eingerichtet ist, das Umfeld des autonomen Systems zu erfassen, zumindest ein zu dem autonomen System vergleichbares Objekt im Umfeld des autonomen Systems zu erkennen, ein Verhalten des zumindest einen zu dem autonomen System vergleichbaren Objekts zu erfassen, und den Reinforcement Learning Agenten in Abhängigkeit von dem erfassten Verhalten des zumindest einen zu dem autonomen System vergleichbaren Objekts zu trainieren.

WO 2022/033746 A1

[Fortsetzung auf der nächsten Seite]

(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)*

Trainieren eines Reinforcement Learning Agenten zur Steuerung eines autonomen Systems

Die Erfindung betrifft eine Vorrichtung und ein Verfahren zum Trainieren eines Reinforcement Learning Agenten zur Steuerung eines autonomen Systems.

Unter dem Begriff „automatisiertes Fahren“ kann im Rahmen des Dokuments ein Fahren mit automatisierter Längs- oder Querführung oder ein autonomes Fahren mit automatisierter Längs- und Querführung verstanden werden. Der Begriff „automatisiertes Fahren“ umfasst ein automatisiertes Fahren mit einem beliebigen Automatisierungsgrad. Beispielhafte Automatisierungsgrade sind ein assistiertes, teilautomatisiertes, hochautomatisiertes oder vollautomatisiertes Fahren. Diese Automatisierungsgrade wurden von der Bundesanstalt für Straßenwesen (BASt) definiert (siehe BASt-Publikation „Forschung kompakt“, Ausgabe

11/2012). Beim assistierten Fahren führt der Fahrer dauerhaft die Längs- oder Querführung aus, während das System die jeweils andere Funktion in gewissen Grenzen übernimmt. Beim teilautomatisierten Fahren (TAF) übernimmt das System die Längs- und Querführung für einen gewissen Zeitraum und/oder in spezifischen Situationen, wobei der Fahrer das System wie beim assistierten Fahren dauerhaft überwachen muss. Beim hochautomatisierten Fahren (HAF) übernimmt das System die Längs- und Querführung für einen gewissen Zeitraum, ohne dass der Fahrer das System dauerhaft überwachen muss; der Fahrer muss aber in einer gewissen Zeit in der Lage sein, die Fahrzeugführung zu übernehmen. Beim vollautomatisierten Fahren (VAF) kann das System für einen spezifischen Anwendungsfall das Fahren in allen Situationen automatisch bewältigen; für diesen Anwendungsfall ist kein Fahrer mehr erforderlich. Die vorstehend genannten vier Automatisierungsgrade gemäß der Definition der BASt entsprechen den SAE-Level 1 bis 4 der Norm SAE J3016 (SAE - Society of Automotive Engineering). Beispielsweise entspricht das hochautomatisierte Fahren (HAF) gemäß der BASt dem Level 3 der Norm SAE J3016. Ferner ist in der SAE J3016 noch der SAE-Level 5 als höchster Automatisierungsgrad vorgesehen, der in der Definition der BASt nicht enthalten ist. Der SAE-Level 5 entspricht einem fahrerlosen Fahren, bei dem das System während der ganzen Fahrt alle Situationen wie ein menschlicher Fahrer automatisch bewältigen kann; ein Fahrer ist generell nicht mehr erforderlich.

Reinforcement Learning (deutsch: bestärkendes oder verstärkendes Lernen) ist bekannt als eine Reihe von Methoden des maschinellen Lernens, bei denen ein Agent selbstständig eine Strategie erlernt, um erhaltene Belohnungen zu maximieren. Dabei wird dem Agenten nicht vorgezeigt, welche Aktion in welcher Situation die beste ist, sondern er erhält zu bestimmten Zeitpunkten eine Belohnung, die auch negativ sein kann. Anhand dieser Belohnungen approximiert er eine Nutzenfunktion, die beschreibt, welchen Wert ein bestimmter Zustand oder Aktion hat.

Zum Trainieren des Agenten ist eine große Anzahl von Trainingsdatensätzen notwendig. Das Sammeln der Trainingsdatensätze ist insbesondere im Umfeld des Einsatzes von Reinforcement Learning für automatisiertes Fahren sehr aufwändig, da ein Testfahrzeug, das zum Sammeln von Trainingsdatensätzen eingerichtet ist, nur eine sehr begrenzte Anzahl von relevanten Verkehrssituationen in einer vorgegebenen Zeit erlebt.

Es ist Aufgabe der Erfindung, die Anzahl der zum Training des Reinforcement Learning Agenten verfügbaren Trainingsdatensätze zu erhöhen.

Die Aufgabe wird durch die Merkmale der unabhängigen Patentansprüche gelöst. Vorteilhafte Ausführungsformen sind in den abhängigen Ansprüchen beschrieben. Es wird darauf hingewiesen, dass zusätzliche Merkmale eines von einem unabhängigen Patentanspruch abhängigen Patentanspruchs ohne die Merkmale des unabhängigen Patentanspruchs oder nur in Kombination mit einer Teilmenge der Merkmale des unabhängigen Patentanspruchs eine eigene und von der Kombination sämtlicher Merkmale des unabhängigen Patentanspruchs unabhängige Erfindung bilden können, die zum Gegenstand eines unabhängigen Anspruchs, einer Teilungsanmeldung oder einer Nachanmeldung gemacht werden kann. Dies gilt in gleicher Weise für in der Beschreibung beschriebene technische Lehren, die eine von den Merkmalen der unabhängigen Patentansprüche unabhängige Erfindung bilden können.

Ein erster Aspekt der Erfindung betrifft Vorrichtung zum Trainieren eines Reinforcement Learning Agenten zur Steuerung eines autonomen Systems.

Das autonome System, das oft auch als autonomer Roboter bezeichnet wird, ist ein technisches System, das sein Verhalten oder seine Aufgaben mit einem hohen Grad an Autonomie, also ohne externe Steuerung, ausführt. Insbesondere handelt es sich bei dem autonomen System um ein

technisches System, das sich autonom bewegt, also beispielsweise ein autonomes Luft-, Land- oder Wasserfahrzeug.

Die Vorrichtung ist eingerichtet, das Umfeld des autonomen Systems zu erfassen, insbesondere mittels Sensoren wie beispielsweise Kameras, Radar und/oder Lidar. Sensordaten dieser Sensoren können beispielsweise fusioniert werden, um ein Umfeldmodell des Umfelds des autonomen Systems zu erzeugen, um das Umfeld des autonomen Systems gesamthaft zu erfassen.

Außerdem ist die Vorrichtung eingerichtet, zumindest ein zu dem autonomen System vergleichbares Objekt im Umfeld des autonomen Systems zu erkennen.

Das zumindest eine zu dem autonomen System vergleichbare Objekt ist insbesondere dann zu dem autonomen System vergleichbar, wenn es vergleichbare Eigenschaften und/oder einen vergleichbaren Einfluss auf sein Umfeld hat.

Das zumindest eine zu dem autonomen System vergleichbare Objekt kann selbst ein autonomes System oder aber auch ein manuelles, bzw. nicht-autonomes, also ein von einem Menschen gesteuertes System sein.

Außerdem ist die Vorrichtung eingerichtet, ein Verhalten des zumindest einen zu dem autonomen System vergleichbaren Objekts zu erfassen, und den Reinforcement Learning Agenten in Abhängigkeit von dem erfassten Verhalten des zumindest einen zu dem autonomen System vergleichbaren Objekts zu trainieren.

Hierbei liegt der Erfindung der Gedanke zugrunde, dass der Reinforcement Learning Agent nicht, wie im Stand der Technik bekannt, an Hand des Verhaltens des autonomen Systems, das der Reinforcement Learning Agent

selbst steuert oder steuern soll, lernt, sondern an Hand des Verhaltens des zumindest einem zu dem autonomen System vergleichbaren Objekts. Mit anderen Worten können somit zum Training des Reinforcement Learning Agenten nicht nur Trainingsdatensätze verwendet werden, die von dem autonomen System selbst stammen, sondern auch Trainingsdatensätze von dem zumindest einen zu dem autonomen System vergleichbaren Objekt im Umfeld des autonomen Systems.

In einer vorteilhaften Ausführungsform der Erfindung ist das autonome System ein automatisiertes Kraftfahrzeug.

Das zumindest eine zu dem automatisierten Kraftfahrzeug vergleichbares Objekt im Umfeld des automatisierten Kraftfahrzeugs hat vergleichbare Eigenschaften und/oder einen vergleichbaren Einfluss auf sein Umfeld wie das automatisierte Kraftfahrzeug. Die Eigenschaften des Kraftfahrzeugs sind dabei beispielsweise die Abmessungen, das Beschleunigungspotential, das Geschwindigkeitspotential und/oder die Umfeldkommunikationsmöglichkeiten des automatisierten Kraftfahrzeugs. Der Einfluss des automatisierten Kraftfahrtzeugs auf sein Umfeld zeigt sich beispielsweise darin, wie viel Abstand andere Verkehrsteilnehmer zu dem automatisierten Kraftfahrzeug halten.

Das zumindest eine zu dem automatisierten Kraftfahrzeugs vergleichbare Objekt im Umfeld des automatisierten Kraftfahrzeugs ist insbesondere ein anderes Kraftfahrzeug, beispielsweise ein anderes Kraftfahrzeug aus der gleichen Kraftfahrzeugklasse.

In einer weiteren vorteilhaften Ausführungsform der Erfindung ist die Vorrichtung eingerichtet, den Reinforcement Learning in Abhängigkeit von dem erfassten Verhalten des zumindest einen zu dem autonomen System vergleichbaren Objekts, und einer Zielfunktion (englisch: reward function) für die Steuerung des autonomen Systems zu trainieren.

Hierbei liegt der Erfindung die Erkenntnis zugrunde, dass es nicht notwendig ist, eine mögliche Zielfunktion des zumindest einen zu dem autonomen System vergleichbaren Objekts zu ermitteln, sondern dass stattdessen die bekannte Zielfunktion für die Steuerung des autonomen Systems zum Trainieren des Reinforcement Learning Agenten genutzt werden kann.

Die Zielfunktion für die Steuerung des autonomen System kann, wenn es sich bei dem autonomen System um ein automatisiertes Kraftfahrzeug handelt, beispielsweise das Einhalten einer vorgegebenen Geschwindigkeit des automatisierten Kraftfahrzeugs. Alternativ oder zusätzlich kann die Zielfunktion des auch Komfortaspekte umfassen wie z.B. eine Begrenzung der maximalen Beschleunigung des automatisierten Kraftfahrzeugs in Längs- und/oder Querrichtung. Außerdem kann die Zielfunktion auch Sicherheitsaspekte umfassen wie z.B. das Einhalten von Sicherheitsabständen zu anderen Verkehrsteilnehmern.

Es ist zwar beispielsweise möglich, dass die Zielfunktion des automatisierten Kraftfahrzeugs auf ein Einhalten einer Geschwindigkeit von 50 km/h abzielt und die tatsächliche Zielfunktion eines weiteren Verkehrsteilnehmers als zu dem automatisierten Kraftfahrzeug vergleichbaren Objekts das Einhalten einer anderen Geschwindigkeit, z.B. 70 km/h ist, dennoch ist es nicht notwendig diese Zielfunktion des weiteren Verkehrsteilnehmers zu ermitteln oder anzunähern.

In einer weiteren vorteilhaften Ausführungsform der Erfindung ist die Vorrichtung eingerichtet, einen Zustand des zumindest einen zu dem autonomen System vergleichbaren Objekts zu erfassen.

Der Zustand des zumindest einen Objekts wird beispielsweise durch dessen Position, Geschwindigkeit und/oder Bewegungsrichtung beschrieben. Wenn es sich bei dem zumindest einen Objekt um ein Kraftfahrzeug handelt, so

wird der Zustand alternativ oder zusätzlich auch durch dessen Fahrspur und/oder den Abstand des zumindest einen Objekts zu möglichen weiteren Verkehrsteilnehmern beschrieben.

Außerdem ist die Vorrichtung eingerichtet, eine den Zustand des zumindest einen zu dem autonomen System vergleichbaren Objekts verändernde Aktion des zumindest einen zu dem autonomen System vergleichbaren Objekts zu erfassen.

Die Aktion ist beispielsweise eine Änderung der Position, der Geschwindigkeit und/oder der Bewegungsrichtung des Objekts. Wenn es sich bei dem Objekt um ein Kraftfahrzeug handelt, so ist die Aktion beispielsweise auch ein Spurwechsel, ein Einfädeln oder ein Abbiegen des Objekts.

Außerdem ist die Vorrichtung eingerichtet, einen durch die Aktion des zumindest einen zu dem autonomen System vergleichbaren Objekts ausgelösten Folgezustand des zumindest einen zu dem autonomen System vergleichbaren Objekts zu erfassen, und den Reinforcement Learning Agenten in Abhängigkeit von dem Zustand des zumindest einen zu dem autonomen System vergleichbaren Objekts, der Aktion des zumindest einen zu dem autonomen System vergleichbaren Objekts, dem Folgezustand des zumindest einen zu dem autonomen System vergleichbaren Objekts und einer Zielfunktion für die Steuerung des autonomen Systems zu trainieren.

In einer weiteren vorteilhaften Ausführungsform der Erfindung ist die Vorrichtung eingerichtet, eine Repräsentation des zumindest einen zu dem autonomen System vergleichbaren Objekts im Umfeld des autonomen Systems derart zu transformieren, dass die transformierte Repräsentation des zumindest einen zu dem autonomen System vergleichbaren Objekts einer möglichen Repräsentation des autonomen Systems entspricht, und den Reinforcement Learning Agenten in Abhängigkeit von dem erfassten

Verhalten des zumindest einen zu dem autonomen System vergleichbaren Objekts zu trainieren.

Diese Transformation ist beispielsweise eine Koordinatentransformation, bei der die Repräsentation des zumindest einen zu dem autonomen System vergleichbare Objekt in einem Koordinatensystem an die Stelle des autonomen Systems gesetzt wird.

In einer weiteren vorteilhaften Ausführungsform der Erfindung ist die Vorrichtung eingerichtet, zumindest zwei zu dem autonomen System vergleichbare Objekte im Umfeld des autonomen Systems zu erkennen, jeweils ein Verhalten der zumindest zwei zu dem autonomen System vergleichbaren Objekte zu erfassen, und den Reinforcement Learning Agenten in Abhängigkeit von dem jeweiligen erfassten Verhalten der zumindest zwei zu dem autonomen System vergleichbaren Objekte zu trainieren.

In einer weiteren vorteilhaften Ausführungsform der Erfindung ist die Vorrichtung eingerichtet, den Reinforcement Learning Agenten gleichzeitig in Abhängigkeit von dem jeweiligen erfassten Verhalten der zumindest zwei zu dem autonomen System vergleichbaren Objekte zu trainieren.

Hierbei liegt der Erfindung die Erkenntnis zugrunde, dass bei einem gleichzeitigen Trainieren des Reinforcement Learning Agenten in Abhängigkeit von dem jeweiligen erfassten Verhalten der zumindest zwei zu dem autonomen System vergleichbaren Objekte eine möglicherweise aufwändige Vorverarbeitung einer Repräsentation der zumindest zwei zu dem autonomen System vergleichbaren Objekte entfallen kann.

In einer weiteren vorteilhaften Ausführungsform der Erfindung ist die Vorrichtung eingerichtet, jeweils eine Repräsentation der zumindest zwei zu dem autonomen System vergleichbaren Objekte im Umfeld des autonomen

Systems mittels einer Permutations-invarianten oder mittels einer Permutations-äquivarianten Abbildung zu einer gemeinsamen Repräsentation zu fusionieren, und den Reinforcement Learning Agenten in Abhängigkeit von der gemeinsamen Repräsentation zu trainieren.

Ein zweiter Aspekt der Erfindung ist ein Verfahren zum Trainieren eines Reinforcement Learning Agenten zur Steuerung eines autonomen Systems.

Ein Schritt des Verfahrens ist das Erfassen des Umfelds des autonomen Systems.

Ein weiterer Schritt des Verfahrens ist das Erkennen von zumindest einem zu dem autonomen System vergleichbaren Objekt im Umfeld des autonomen Systems.

Ein weiterer Schritt des Verfahrens ist das Erfassen eines Verhaltens des zumindest einem zu dem autonomen System vergleichbaren Objekts.

Ein weiterer Schritt des Verfahrens ist das Trainieren des Reinforcement Learning Agenten in Abhängigkeit von dem erfassten Verhalten des zumindest einen zu dem autonomen System vergleichbaren Objekts.

Die vorstehenden Ausführungen zur erfindungsgemäßen Vorrichtung nach dem ersten Aspekt der Erfindung gelten in entsprechender Weise auch für das erfindungsgemäße Verfahren nach dem zweiten Aspekt der Erfindung. An dieser Stelle und in den Patentansprüchen nicht explizit beschriebene vorteilhafte Ausführungsbeispiele des erfindungsgemäßen Verfahrens entsprechen den vorstehend beschriebenen oder in den Patentansprüchen beschriebenen vorteilhaften Ausführungsbeispielen der erfindungsgemäßen Vorrichtung.

Die Erfindung wird nachfolgend anhand eines Ausführungsbeispiels unter Zuhilfenahme der beigefügten Zeichnungen beschrieben. In diesen zeigen:

Fig. 1 eine Fahrsituation zur Illustration der Erfindung,
Fig. 2 eine schematische Ausführungsform der Erfindung, und
Fig. 3 eine weitere schematische Ausführungsform der Erfindung.

Fig. 1 zeigt ein mit einer Vorrichtung zum Trainieren eines Reinforcement Learning Agenten zur Steuerung eines autonomen Systems EGO ausgestattetes autonomes System EGO in Form eines Kraftfahrzeugs.

Die Vorrichtung ist eingerichtet das Umfeld des autonomen Systems EGO zu erfassen und zumindest ein zu dem autonomen System EGO vergleichbares Objekt V1, V2, V3 im Umfeld des autonomen Systems EGO zu erkennen.

Bei dem zumindest einem zu dem autonomen System EGO vergleichbaren Objekt V1, V2, V3 handelt es sich insbesondere um weitere Verkehrsteilnehmer, beispielsweise um weitere Kraftfahrzeuge im Umfeld des autonomen Systems EGO.

Des Weiteren ist die Vorrichtung eingerichtet ein Verhalten des zumindest einen zu dem autonomen System EGO vergleichbaren Objekts V1, V2, V3 zu erfassen, und den Reinforcement Learning Agenten in Abhängigkeit von dem erfassten Verhalten des zumindest einen zu dem autonomen System EGO vergleichbaren Objekts V1, V2, V3 zu trainieren.

Im vorliegenden Beispiel kann das Kraftfahrzeug V1 das Kraftfahrzeug V2 nur unter Berücksichtigung des Kraftfahrzeugs V3 überholen. Diesen nicht trivialen Überholvorgang kann die Vorrichtung im Kraftfahrzeug EGO als Verhalten des Kraftfahrzeugs V1 erfassen und zum Trainieren des Reinforcement Learning Agenten verwenden.

Die Vorrichtung ist dabei insbesondere eingerichtet, den Reinforcement Learning in Abhängigkeit von dem erfassten Verhalten des zumindest einen zu dem autonomen System EGO vergleichbaren Objekts V1, V2, V3 und einer Zielfunktion für die Steuerung des autonomen Systems EGO zu trainieren.

Das Verhalten des zumindest einen zu dem autonomen System EGO vergleichbaren Objekts V1, V2, V3 ist beispielsweise durch einen Zustand des zumindest einen zu dem autonomen System EGO vergleichbaren Objekts V1, V2, V3, eine den Zustand des zumindest einen zu dem autonomen System EGO vergleichbaren Objekts V1, V2, V3 verändernde Aktion des zumindest einen zu dem autonomen System EGO vergleichbaren Objekts V1, V2, V3 und/oder einen durch die Aktion des zumindest einen zu dem autonomen System EGO vergleichbaren Objekts V1, V2, V3 ausgelösten Folgezustand des zumindest einen zu dem autonomen System EGO vergleichbaren Objekts V1, V2, V3 definiert.

Fig. 2 zeigt eine schematische Ausführungsform der Erfindung.

Die darin gezeigte Vorrichtung ist eingerichtet, zumindest zwei zu dem autonomen System EGO vergleichbare Objekte V1, V2, V3 im Umfeld des autonomen Systems EGO zu erkennen, jeweils ein Verhalten der zumindest zwei zu dem autonomen System EGO vergleichbaren Objekte V1, V2, V3 zu erfassen, und den Reinforcement Learning Agenten in Abhängigkeit von dem jeweiligen erfassten Verhalten der zumindest zwei zu dem autonomen System EGO vergleichbaren Objekte V1, V2, V3 zu trainieren.

Dabei ist die Vorrichtung eingerichtet, den Reinforcement Learning Agenten gleichzeitig in Abhängigkeit von dem jeweiligen erfassten Verhalten der zumindest zwei zu dem autonomen System EGO vergleichbaren Objekte V1, V2, V3 zu trainieren.

Hierzu ist die Vorrichtung beispielsweise eingerichtet, eine Repräsentation des zumindest einen zu dem autonomen System EGO vergleichbaren Objekts V1, V2, V3 im Umfeld des autonomen Systems EGO derart zu transformieren, dass die transformierte Repräsentation des zumindest einen zu dem autonomen System EGO vergleichbaren Objekts V1, V2, V3 einer möglichen Repräsentation des autonomen Systems EGO entspricht, und den Reinforcement Learning Agenten in Abhängigkeit von dem erfassten Verhalten des zumindest einen zu dem autonomen System EGO vergleichbaren Objekts V1, V2, V3 zu trainieren.

Fig. 3 zeigt eine weitere schematische Ausführungsform der Erfindung.

Die darin gezeigte Vorrichtung ist eingerichtet, zumindest zwei zu dem autonomen System EGO vergleichbare Objekte V1, V2, V3 im Umfeld des autonomen Systems EGO zu erkennen, jeweils ein Verhalten der zumindest zwei zu dem autonomen System EGO vergleichbaren Objekte V1, V2, V3 zu erfassen, und den Reinforcement Learning Agenten in Abhängigkeit von dem jeweiligen erfassten Verhalten der zumindest zwei zu dem autonomen System EGO vergleichbaren Objekte V1, V2, V3 zu trainieren.

Außerdem ist die Vorrichtung eingerichtet, jeweils eine Repräsentation der zumindest zwei zu dem autonomen System EGO vergleichbaren Objekte V1, V2, V3 im Umfeld des autonomen Systems EGO mittels einer Permutations-invarianten oder mittels einer Permutations-äquivarianten Abbildung MAP zu einer gemeinsamen Repräsentation REP zu fusionieren, und den Reinforcement Learning Agenten in Abhängigkeit von der gemeinsamen Repräsentation REP zu trainieren.

Patentansprüche

1. Vorrichtung zum Trainieren eines Reinforcement Learning Agenten zur Steuerung eines autonomen Systems (EGO), wobei die Vorrichtung eingerichtet ist,
   - das Umfeld des autonomen Systems (EGO) zu erfassen,
   - zumindest ein zu dem autonomen System (EGO) vergleichbares Objekt (V1, V2, V3) im Umfeld des autonomen Systems (EGO) zu erkennen,
   - ein Verhalten des zumindest einen zu dem autonomen System (EGO) vergleichbaren Objekts (V1, V2, V3) zu erfassen, und
   - den Reinforcement Learning Agenten in Abhängigkeit von dem erfassten Verhalten des zumindest einen zu dem autonomen System (EGO) vergleichbaren Objekts (V1, V2, V3) zu trainieren.

2. Vorrichtung nach Anspruch 1, wobei das autonome System (EGO) ein automatisiertes Kraftfahrzeug ist.

3. Vorrichtung nach einem der vorherigen Ansprüche, wobei die Vorrichtung eingerichtet ist, den Reinforcement Learning in Abhängigkeit von
   - dem erfassten Verhalten des zumindest einen zu dem autonomen System (EGO) vergleichbaren Objekts (V1, V2, V3), und
   - einer Zielfunktion für die Steuerung des autonomen Systems (EGO) zu trainieren.

4. Vorrichtung nach einem der vorherigen Ansprüche, wobei die Vorrichtung eingerichtet ist,
    - einen Zustand des zumindest einen zu dem autonomen System (EGO) vergleichbaren Objekts (V1, V2, V3) zu erfassen,
    - eine den Zustand des zumindest einen zu dem autonomen System (EGO) vergleichbaren Objekts (V1, V2, V3) verändernde Aktion des zumindest einen zu dem autonomen System (EGO) vergleichbaren Objekts (V1, V2, V3) zu erfassen,
    - einen durch die Aktion des zumindest einen zu dem autonomen System (EGO) vergleichbaren Objekts (V1, V2, V3) ausgelösten Folgezustand des zumindest einen zu dem autonomen System (EGO) vergleichbaren Objekts (V1, V2, V3) zu erfassen, und
    - den Reinforcement Learning Agenten in Abhängigkeit von
        - dem Zustand des zumindest einen zu dem autonomen System (EGO) vergleichbaren Objekts (V1, V2, V3), der Aktion des zumindest einen zu dem autonomen System (EGO) vergleichbaren Objekts (V1, V2, V3), dem Folgezustand des zumindest einen zu dem autonomen System (EGO) vergleichbaren Objekts (V1, V2, V3), und
        - einer Zielfunktion für die Steuerung des autonomen Systems (EGO) zu trainieren.

5. Vorrichtung nach einem der vorherigen Ansprüche, wobei die Vorrichtung eingerichtet ist,
    - eine Repräsentation des zumindest einen zu dem autonomen System (EGO) vergleichbaren Objekts (V1, V2, V3) im Umfeld des autonomen Systems (EGO) derart zu transformieren, dass die transformierte Repräsentation des zumindest einen zu dem autonomen System (EGO) vergleichbaren Objekts (V1, V2, V3)

einer möglichen Repräsentation des autonomen Systems (EGO) entspricht, und

- den Reinforcement Learning Agenten in Abhängigkeit von dem erfassten Verhalten des zumindest einen zu dem autonomen System (EGO) vergleichbaren Objekts (V1, V2, V3) zu trainieren.

6. Vorrichtung nach einem der vorherigen Ansprüche, wobei die Vorrichtung eingerichtet ist,
   - zumindest zwei zu dem autonomen System (EGO) vergleichbare Objekte (V1, V2, V3) im Umfeld des autonomen Systems (EGO) zu erkennen,
   - jeweils ein Verhalten der zumindest zwei zu dem autonomen System (EGO) vergleichbaren Objekte (V1, V2, V3) zu erfassen, und
   - den Reinforcement Learning Agenten in Abhängigkeit von dem jeweiligen erfassten Verhalten der zumindest zwei zu dem autonomen System (EGO) vergleichbaren Objekte (V1, V2, V3) zu trainieren.

7. Vorrichtung nach Anspruch 5, wobei die Vorrichtung eingerichtet ist, den Reinforcement Learning Agenten gleichzeitig in Abhängigkeit von dem jeweiligen erfassten Verhalten der zumindest zwei zu dem autonomen System (EGO) vergleichbaren Objekte (V1, V2, V3) zu trainieren.

8. Vorrichtung nach Anspruch 5 oder 6, wobei die Vorrichtung eingerichtet ist,
   - jeweils eine Repräsentation der zumindest zwei zu dem autonomen System (EGO) vergleichbaren Objekte (V1, V2, V3) im Umfeld des autonomen Systems (EGO) mittels einer

Permutations-invarianten oder mittels einer Permutations-äquivarianten Abbildung (MAP) zu einer gemeinsamen Repräsentation (REP) zu fusionieren, und

- den Reinforcement Learning Agenten in Abhängigkeit von der gemeinsamen Repräsentation (REP) zu trainieren.

9. Verfahren zum Trainieren eines Reinforcement Learning Agenten zur Steuerung eines autonomen Systems (EGO), wobei das Verfahren die folgenden Schritte umfasst:
   - Erfassen des Umfelds des autonomen Systems (EGO),
   - Erkennen von zumindest einem zu dem autonomen System (EGO) vergleichbaren Objekt (V1, V2, V3) im Umfeld des autonomen Systems (EGO),
   - Erfassen eines Verhaltens des zumindest einem zu dem autonomen System (EGO) vergleichbaren Objekts (V1, V2, V3), und
   - Trainieren des Reinforcement Learning Agenten in Abhängigkeit von dem erfassten Verhalten des zumindest einen zu dem autonomen System (EGO) vergleichbaren Objekts (V1, V2, V3).

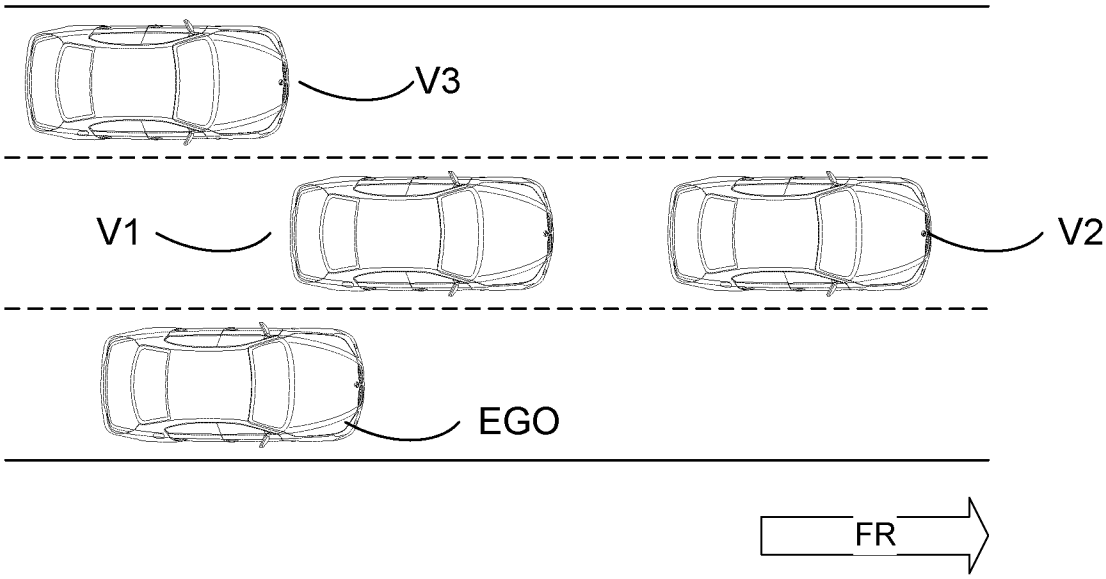


Fig. 1

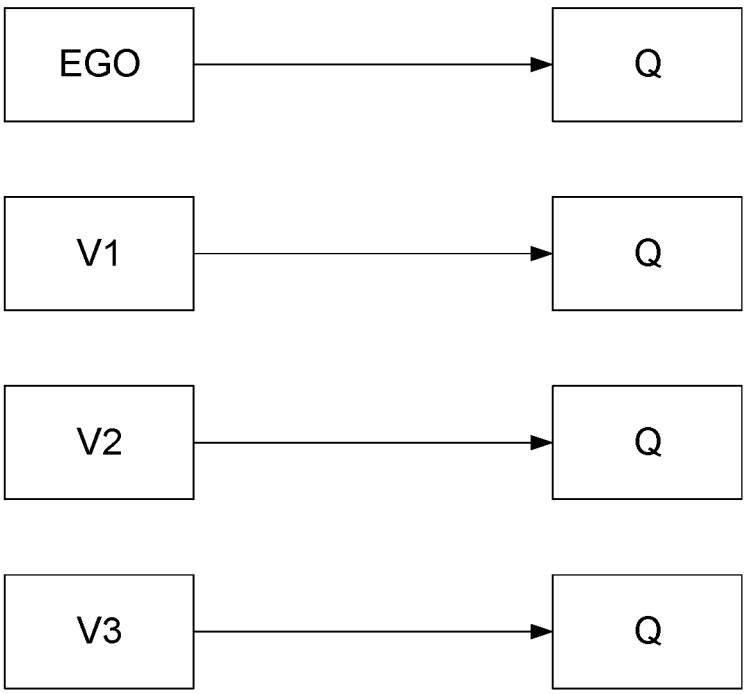


Fig. 2

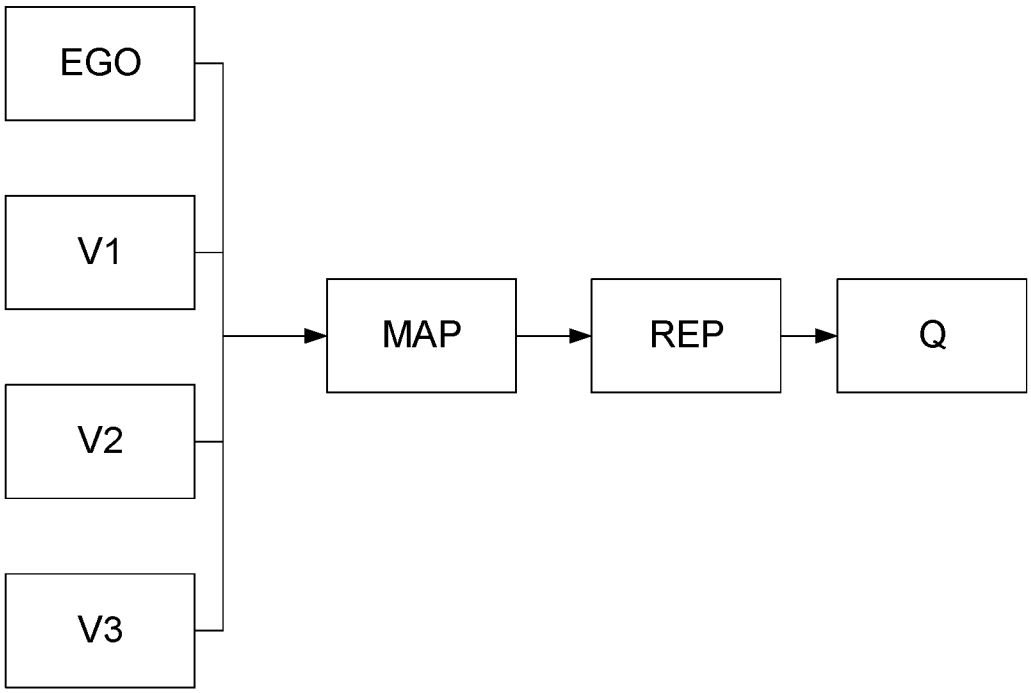


Fig. 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2021/065919

A. CLASSIFICATION OF SUBJECT MATTER

G06N 3/00(2006.01)i; ***G06N 3/08***(2006.01)i; ***B60W 30/18***(2012.01)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; B60W

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	US 10627823 B1 (KIM KYE-HYEON [KR] ET AL) 21 April 2020 (2020-04-21) column 1, line 15 - line 20 column 11, line 60 - line 65 column 15, line 20 - line 46 figure 2	1-9
X	US 2020189597 A1 (VERONESE LUCAS [DE] ET AL) 18 June 2020 (2020-06-18) paragraph [0008] - paragraph [0067] figures 1-5	1-9
X	WANG PIN ET AL. "A Reinforcement Learning Based Approach for Automated Lane Change Maneuvers" *2018 IEEE INTELLIGENT VEHICLES SYMPOSIUM (IV), IEEE,* 26 June 2018 (2018-06-26), pages 1379-1384 DOI: 10.1109/IVS.2018.8500556 XP033423455 page 1379 - page 1383	1-9

☑ 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	Date of mailing of the international search report
09 September 2021	**17 September 2021**
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 **Bohn, Patrice** Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2021/065919

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MICHEL DEUDON ET AL. "HighRes-net: Recursive Fusion for Multi-Frame Super-Resolution of Satellite Imagery" *ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853,* 15 February 2020 (2020-02-15), XP081600852 page 13, paragraph 1.3	8

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/EP2021/065919

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
US 10627823 B1	21 April 2020	CN 111507372 A	07 August 2020
		EP 3690711 A1	05 August 2020
		JP 6913969 B2	04 August 2021
		JP 2020123335 A	13 August 2020
		KR 20200094641 A	07 August 2020
		US 10627823 B1	21 April 2020
US 2020189597 A1	18 June 2020	CN 111301419 A	19 June 2020
		EP 3667556 A1	17 June 2020
		US 2020189597 A1	18 June 2020

INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen
PCT/EP2021/065919

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES
INV. G06N3/00 G06N3/08 B60W30/18
ADD.

Nach der Internationalen Patentklassifikation (IPC) oder nach der nationalen Klassifikation und der IPC

B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)
G06N B60W

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	US 10 627 823 B1 (KIM KYE-HYEON [KR] ET AL) 21. April 2020 (2020-04-21) Spalte 1, Zeile 15 - Zeile 20 Spalte 11, Zeile 60 - Zeile 65 Spalte 15, Zeile 20 - Zeile 46 Abbildung 2	1-9
X	US 2020/189597 A1 (VERONESE LUCAS [DE] ET AL) 18. Juni 2020 (2020-06-18) Absatz [0008] - Absatz [0067] Abbildungen 1-5	1-9

-/--

[X] Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen [X] 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
9. September 2021	17/09/2021

Name und Postanschrift der Internationalen Recherchenbehörde	Bevollmächtigter Bediensteter
Europäisches Patentamt, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Bohn, Patrice

1

C. (Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	WANG PIN ET AL: "A Reinforcement Learning Based Approach for Automated Lane Change Maneuvers", 2018 IEEE INTELLIGENT VEHICLES SYMPOSIUM (IV), IEEE, 26. Juni 2018 (2018-06-26), Seiten 1379-1384, XP033423455, DOI: 10.1109/IVS.2018.8500556 [gefunden am 2018-10-18] Seite 1379 - Seite 1383 -----	1-9
A	MICHEL DEUDON ET AL: "HighRes-net: Recursive Fusion for Multi-Frame Super-Resolution of Satellite Imagery", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 15. Februar 2020 (2020-02-15), XP081600852, Seite 13, Absatz 1.3 -----	8

INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen
PCT/EP2021/065919

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
US 10627823 B1	21-04-2020	CN 111507372 A	07-08-2020
		EP 3690711 A1	05-08-2020
		JP 6913969 B2	04-08-2021
		JP 2020123335 A	13-08-2020
		KR 20200094641 A	07-08-2020
		US 10627823 B1	21-04-2020
US 2020189597 A1	18-06-2020	CN 111301419 A	19-06-2020
		EP 3667556 A1	17-06-2020
		US 2020189597 A1	18-06-2020