



(12) **Offenlegungsschrift**

(21) Aktenzeichen: **10 2022 129 575.3**

(22) Anmeldetag: **09.11.2022**

(43) Offenlegungstag: **16.05.2024**

(51) Int Cl.: **G06V 10/774 (2022.01)**

**G07C 3/14 (2006.01)**

**G06Q 50/04 (2012.01)**

(71) Anmelder:  
**Bayerische Motoren Werke Aktiengesellschaft,  
80809 München, DE**

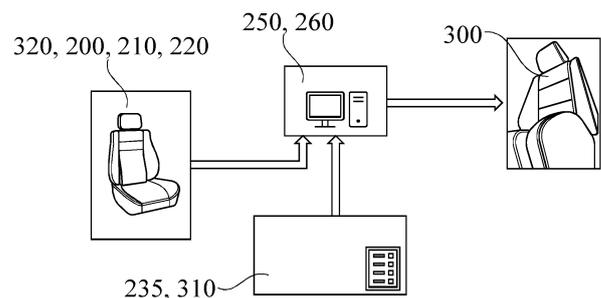
(72) Erfinder:  
**Schlather, Florian, Dr.-Ing., 80807 München, DE;  
Patron, Christian, Dr., 81667 München, DE**

Rechercheantrag gemäß § 43 PatG ist gestellt.

**Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen.**

(54) Bezeichnung: **Computerimplementiertes Verfahren zum Erzeugen von Trainingsbilddaten zum visuellen Überwachen eines Objektes bei einer Produktion eines Kraftfahrzeugs und/oder einer Kraftfahrzeugkomponente, Trainingsbilddaten, Modell, Verfahren zum visuellen Überwachen, Computerprogramm und/oder computerlesbares Medium, Datenverarbeitungsvorrichtung**

(57) Zusammenfassung: Bereitgestellt wird ein computerimplementiertes Verfahren (100) zum Erzeugen von Trainingsbilddaten (300) für ein Modell für maschinelles Lernen zum visuellen Überwachen (400) eines Objektes (200) bei einer Produktion eines Kraftfahrzeugs (205) und/oder einer Kraftfahrzeugkomponente (210), wobei das Verfahren (100) aufweist: Erfassen (110) von das Objekt (200) betreffenden Textinformation (310); Erfassen (120) von das Objekt (200) betreffenden Bilddaten (320); und Verarbeiten (130) der Bilddaten (320) in Kombination mit der Textinformation (310) zum Erzeugen der Trainingsbilddaten (300), wobei die Trainingsbilddaten (300) das Objekt (200) betreffende und gemäß der Textinformation (310) ausgeprägte Trainingsbilddaten (300) umfasst.



## Beschreibung

**[0001]** Die vorliegende Offenbarung betrifft ein computerimplementiertes Verfahren zum Erzeugen von Trainingsbilddaten für ein Modell für maschinelles Lernen zum visuellen Überwachen eines Objektes bei einer Produktion eines Kraftfahrzeugs und/oder einer Kraftfahrzeugkomponente. Die Offenbarung betrifft auch Trainingsbilddaten zur Verwendung zum Trainieren eines Modells für maschinelles Lernen zum visuellen Überwachen eines Objektes bei einer Produktion eines Kraftfahrzeugs und/oder einer Kraftfahrzeugkomponente, ein Modell für maschinelles Lernen zum visuellen Überwachen eines Objektes bei einer Produktion eines Kraftfahrzeugs und/oder einer Kraftfahrzeugkomponente, ein Verfahren zum visuellen Überwachen eines Objektes bei einer Produktion eines Kraftfahrzeugs und/oder einer Kraftfahrzeugkomponente, ein Computerprogramm und/oder computerlesbares Medium und eine Datenverarbeitungsvorrichtung.

**[0002]** In modernen Produktionsstätten für ein Kraftfahrzeug und/oder für eine Komponente eines Kraftfahrzeugs werden zunehmend Bildverarbeitungssysteme eingesetzt, um beispielsweise die Qualität in der Produktion sowie Eigenschaften des Objektes zu überwachen, also zur Qualitätssicherung. Das Objekt kann dabei eine Kraftfahrzeugkomponente sein, deren Produktion zu überwachen ist. Bei der Produktion des Objekts kann das Objekt beispielsweise auf mögliche Fehler und/oder auf eine korrekte Ausführung überwacht werden. Derartige Bildverarbeitungssysteme können auf künstlicher Intelligenz (KI) beziehungsweise maschinellem Lernen beruhen. Diese Systeme erfordern das Training von KI-Modellen, um zum Überwachen der Produktion mit hoher Qualität und Verlässlichkeit arbeiten zu können.

**[0003]** Nach dem Stand der Technik werden zum Trainieren derartiger Bildverarbeitungssysteme Bilder der entsprechend zu prüfenden Bereiche des Objektes beziehungsweise des Kraftfahrzeugs und/oder der Komponente des Kraftfahrzeugs aufgenommen. Die Bilder werden zum Trainieren beispielsweise mit einem Label versehen. Für ein KI-Modell sind daher Bilder beispielsweise mit einem Label des Zustandes „in Ordnung“ (i.O.), d.h. ohne Fehler, oder einem Label des Zustandes „nicht in Ordnung“ (n.i.O.), d.h., mit einem Fehler versehen, also einen Fehlerzustand, wobei eine weitere Differenzierung der Label möglich ist.

**[0004]** Das Training derartiger KI-Modelle ist nach dem Stand der Technik aufwendig, da eine Vielzahl von Bilddaten aufgenommen werden müssen. Zudem müssen Fehlerbilder für den Zustand „nicht in Ordnung“ aufwendig erzeugt werden, beispielsweise indem bewusst Fehlverbauten physisch aufge-

baut werden und diese dann als Bild aufgenommen werden. Im Betrieb liefern KI-Modelle nach dem Stand der Technik häufig ungenaue Aussagen, da sie nicht auf neue, in der Produktion möglicherweise erstmalig auftretende Fehler trainiert sind. So kann es beispielsweise vorkommen, dass ein beschädigter Spiegel verbaut wird, da ein KI-Modell nicht trainiert wurde, einen beschädigten Spiegel und/oder einen Spiegel mit einem spezifischen Fehler zu erkennen, sondern nur beispielsweise die Farbe der Spiegel kontrolliert.

**[0005]** Die Erzeugung synthetischer Daten für das Modelltraining der KI ist bekannt, allerdings entweder anhand manuell erzeugter Daten oder ohne einen steuernden Eingriff durch beispielsweise erzeugende generische Netzwerke (generative adversarial networks, GANs).

**[0006]** Die aus dem Stand der Technik bekannten Verfahren sind aufwendig und erfordern einen hohen Zeitaufwand beim Training der KI-Modelle. Zudem lernen KI-Systeme in der Produktion heute nicht mit und können daher nicht auf spezifische, neu auftretende Fehler reagieren.

**[0007]** Vor dem Hintergrund dieses Standes der Technik besteht die Aufgabe der vorliegenden Offenbarung darin, ein verbessertes Verfahren anzugeben, welche geeignet ist, den Stand der Technik zu bereichern. Eine konkrete Ausgestaltung der Offenbarung kann die Aufgabe lösen, eine effiziente und kontextsensitive Überwachung eines Objektes bei einer Produktion eines Kraftfahrzeugs und/oder einer Kraftfahrzeugkomponente zu ermöglichen.

**[0008]** Gelöst wird die Aufgabe durch die Merkmale des unabhängigen Anspruchs. Die Unteransprüche haben optionale Weiterbildungen der Offenbarung zum Inhalt.

**[0009]** Danach wird die Aufgabe durch ein computerimplementiertes Verfahren zum Erzeugen von Trainingsbilddaten für ein Modell für maschinelles Lernen zum visuellen Überwachen eines Objektes bei einer Produktion eines Kraftfahrzeugs und/oder einer Kraftfahrzeugkomponente gelöst. Dabei weist das Verfahren auf: Erfassen von das Objekt betreffenden Textinformation; Erfassen von das Objekt betreffenden Bilddaten; und Verarbeiten der Bilddaten in Kombination mit der Textinformation zum Erzeugen der Trainingsbilddaten, wobei die Trainingsbilddaten das Objekt betreffende und gemäß der Textinformation ausgeprägte Trainingsbilddaten umfasst.

**[0010]** Das Verfahren kann somit Trainingsbilddaten erzeugen beziehungsweise generieren. Die Trainingsbilddaten können zum Training für das Modell zum Überwachen des Objektes bei der Produktion

des Kraftfahrzeugs und/oder der Kraftfahrzeugkomponente verwendet werden. Dafür kann ein Erfassen beziehungsweise Eingaben von Bilddaten und Textinformation erfolgen, die jeweils das Objekt, also das Kraftfahrzeug und/oder die Kraftfahrzeugkomponente betreffen. Die Bilddaten können eine bildliche, also visuelle, Darstellung des Objektes umfassen. Die Textinformation kann eine Eigenschaft beziehungsweise ein Merkmal des Objektes beschreiben, insbesondere eine Eigenschaft, die von in den Bilddaten dargestellten Eigenschaften abweicht. Insbesondere ausgehend von den Bilddaten und anhand der Textinformation können die Trainingsbilddaten generiert werden. Dafür können die Bilddaten gemäß der Textinformation modifiziert, also die Bilddaten mit der Textinformation kombiniert werden, um die Trainingsbilddaten zu generieren. Die Trainingsbilddaten können dann eine Darstellung des Objektes gemäß der Bilddaten umfassen, wobei das Objekt die durch die Textinformation angegebene Eigenschaft aufweist.

**[0011]** Damit ist es entbehrlich, das Objekt mit der Eigenschaft beispielsweise herstellen zu müssen, um Trainingsbilddaten zu generieren. Die Trainingsbilddaten können effektiv anhand der Bilddaten und der Textinformation generiert werden. Damit kann das Modell für maschinelles Lernen zum visuellen Überwachen der Produktion effektiv trainiert werden. Es ist möglich, dass das Modell auch während der Produktion mitlernt und daher auf spezifische, neu auftretende Fehler reagieren kann. Ein effektives Erzeugen der Bilddaten ist insbesondere unter Verwendung eines generativen Transformers (generative pretrained transformer, GPT) als künstliche Intelligenz möglich, die die Bilddaten und die Textinformationen verarbeitet.

**[0012]** Die Bilddaten können das Objekt in einem Referenzzustand abbilden. Beispielsweise kann der Referenzzustand den Zustand „in Ordnung“ (i.O.), also einen fehlerfreien Zustand, des Objektes betreffen. Alternativ oder zusätzlich kann der Referenzzustand einen prototypischen Zustand betreffen, wobei die Bilddaten beispielsweise nur ein geometrisches Grundgerüst des Objekts umfassen.

**[0013]** Die Textinformation kann einen Fehlerzustand und/oder eine Variante des Objekts beschreiben. Beispielsweise kann die Textinformation als Fehlerzustand einen durch das Überwachen aufzufindenden Fehler angeben, beispielsweise einen Kratzer und/oder eine Delle, wobei optional Angaben zu Größe und genauer Anordnung von der Textinformation umfasst sein können. Die Textinformation kann als Variante beispielsweise eine Angabe zu einer Oberflächengegebenheit, einer Farbe und/oder eines Materials des Objektes machen. Damit können die Textinformation typisch zu Überwachende Eigen-

schaften angeben, um entsprechende Trainingsbilddaten erzeugen zu können.

**[0014]** Die Textinformation und/oder die Bilddaten können einen Abschnitt des Objekts definieren, wobei bei dem Verarbeiten der Bilddaten der Abschnitt anhand der Textinformation modifiziert wird. Damit können die Bilddaten gezielt gemäß der Textinformation effektiv zum Erzeugen der Trainingsbilddaten führen. Mit anderen Worten ermöglicht der Abschnitt eine in-Image-Bearbeitung, wobei die Bilddaten innerhalb des Abschnitts, einem sogenannten in-paint-Bereich, bearbeitet werden, während der Rest der Bilddaten außerhalb des Abschnitts optional unverändert bleibt.

**[0015]** Beim Verarbeiten der Bilddaten in Kombination mit der Textinformation pro Bilddaten und Textinformation eine Mehrzahl von Trainingsbilddaten erzeugt wird. Damit kann anhand eines Bildes und einer Textinformation eine Mehrzahl von Trainingsbilddaten erzeugt werden, um ein effektives Lernen des Modells durch eine vergleichsweise einfach bereitzustellende und dennoch große Datenbasis zu ermöglichen.

**[0016]** Die Bilddaten können Objektfotos und/oder Konstruktionsdaten umfassen. Damit ist es möglich, die Bilddaten als Darstellung des realen Objektes und/oder als Rendering als Grundlage für das Erzeugen der Trainingsbilddaten zu verwenden. Damit kann das tatsächliche und/oder ein zukünftiges Objekt durch die Bilddaten beschrieben werden.

**[0017]** Zum Erzeugen der Trainingsbilddaten kann ein generativer Transformer verwendet werden. Ein generativer Transformer (generative pre-trained transformer, GPT) kann Trainingsbilddaten effizient anhand der Textinformation und insbesondere in Kombination mit den Bilddaten erzeugen.

**[0018]** Das oben Beschriebene lässt sich mit anderen Worten und auf eine konkrete Ausgestaltung bezogen, die als für die vorliegende Offenbarung nicht limitierend beschrieben wird, wie folgt zusammenfassen: Die Offenbarung betrifft kontextsensitives Training von KI-Modellen für die Produktion auf Basis von GPTs, CAD- und Produktionsdaten beziehungsweise ein effizientes, kontextsensitives und adaptives Training von KI-Modellen für die Produktion. Der Kerngedanke ist es, dass bereits vorhandene Informationen aus der Objektentwicklung sowie neue KI-Methoden in Form von generative pretrained transformers (GPT) kombiniert genutzt werden. Konkret sind dies 1. Konstruktionsdaten (z.B. aus CAD) der Umfänge, die später geprüft werden müssen. Diese dienen als geometrische Grundstruktur bzw. Basis des zu erzeugenden Trainingsbildes. 2. Orderdaten aus der Programmplanung, die die einzelnen Umfänge genauer spezifizieren (z.B.

Material- bzw. Stücklisten von Komponenten). 3. KI-Methoden aus dem Feld der generative pretrained transformers (GPT). Diese erlauben die Erzeugung oder Veränderung eines Bildes aus Textinformationen. Dies kann beispielsweise wie Folgt umgesetzt werden: Ein Fahrzeugsitz muss im Objektivbetrieb auf die korrekte Variante geprüft werden, d.h. auf die korrekte Geometrie sowie Ausstattung, z.B. Sportsitz mit schwarzem Leder und roter Kontraststeppung. Um ein Bildverarbeitungssystem anzulernen, müssen Referenzbilder aus verschiedenen Ansichten erzeugt werden. Diese stehen ohne Vorhandensein des Sitzes selbst nicht zur Verfügung. Mit Hilfe der Offenbarung werden nun 1. die CAD Daten der Sitze und 2. die Orderdaten der Sitzvarianten (z.B. „Sportsitz mit schwarzem Leder und roter Kontraststeppung“) in ein 3. GPT Modell geladen. Dieses erzeugt daraufhin photorealistische Bilder der künftigen Komponente, ohne dass diese zu diesem Zeitpunkt real existiert und ohne manuelle Design-Aufwände. Gleichmaßen ist es Kern der Offenbarung, dass bereits vorhandene Textinformationen aus der Nacharbeit genutzt werden um KI Modelle für die eine visuelle Prüfung kontinuierlich zu verbessern. In der Nacharbeit von Fahrzeugen werden Fehler und Beschädigungen in Textform erfasst z.B. „Spiegel beschädigt“, „Kratzer in der Tür innen“ usw. Das System nimmt diese Textinformationen und erzeugt mit Hilfe eines Text-to-Image-KI-Modells ein Trainingsbild. Dieses wird wiederum benutzt um das KI-Modell des Qualitätssystems zu verbessern. Je nach Qualität der Textinformationen aus den (Nacharbeits-) Systemen kann mit Maskierung und gezielter Text-To-In-Image-KI gearbeitet werden oder ohne. Beispielsweise tritt bei der Türmontage ein neues Fehlerbild auf: „Türgriff fehlt“. Das KI-Modell, das die Qualitätssicherung durchführt, kennt dieses Fehlerbild nicht und wertet möglicherweise falsch aus. Parallel dazu werden mittels einer Text-To-Image-KI ein oder mehrere synthetische Fehlerbilder des Falles „Türgriff fehlt“ erzeugt. Das KI-Modell, das die Qualitätssicherung übernimmt, wird auf Basis dieser Bilder nachtrainiert und kennt künftig dieses Fehlerbild.

**[0019]** Gemäß einem Aspekt wird ein Verfahren zur frühzeitigen Bewertung von Änderungsauswirkungen auf KI-basierte Absicherungen vorgeschlagen. Beispielsweise kann mittels KI optisch überprüft werden, ob bestimmte Objekte im Fahrzeug in Ordnung sind. Dabei kann es jedoch vorkommen, dass bereits angelernte Objekte ausgetauscht werden, wodurch es eventuell zu Fehleinschätzungen der KI kommen kann. Bildverarbeitung auf Basis künstlicher Intelligenz wird in zahlreichen Branchen zur Bewertung von Qualitätsmerkmalen oder der Erfassung des Umfeldes eingesetzt. Dafür werden neuronale Netze mit verschiedenen Verfahren angelernt, eine spezifische Aufgabe zu erfüllen, bspw. das Erkennen von Gurtschlössern im Fahrzeuginnenraum. Das

Anlernen erfolgt auf Basis repräsentativer Bilder wie sie später auch in der realen Anwendung auftreten.

**[0020]** Durch das Anlernen auf repräsentative/bekannte Szenarien sind die KI-Modelle nicht immer in der Lage, mit veränderten Umfeldbedingungen zu funktionieren. Tritt eine Änderung auf (z.B. Gurtschlösser anderer Form, Farbe oder Position als zuvor üblich; beispielsweise aufgrund von neuen Fahrzeugmodell Anpassungen) müssen die Modelle auf dieser Basis nachtrainiert werden. Eine Bewertung vorab, wie veränderte Umfeldbedingungen die Güte des bestehenden Modells beeinflussen, ist nur möglich, wenn Bilddaten für die neuen Bedingungen vorhanden sind. Gerade in der frühen Phase der Produktentwicklung oder bei kurzfristigen Änderungen ist dies nicht immer möglich. Bekannte Verfahren zum synthetischen Erzeugen haben alle gemein, dass sie mit manuellem Aufwand verbunden sind und/oder nicht repräsentative Ergebnisse erzeugen. Die Daten müssen erzeugt, mit richtigen Lichteinstellungen und Texturen versehen werden und anschließend lagegerecht berechnet werden. Dies erfordert hohen Rechenaufwand und aufwendige Prozessabläufe die zudem fehleranfällig sind.

**[0021]** Daher wird gemäß einem Aspekt der Offenbarung vorgeschlagen, Produktänderungen rein verbal/textuell zu beschreiben und die notwendigen Bilddaten für das KI Training (z.B. Trainingsbilddaten) automatisiert aus dieser Verbal-/Textbeschreibung abzuleiten. Der Prozess der Modellgenerierung bei Änderungen kann hierdurch vereinfacht werden, da zahlreiche Zwischenschritte entfallen. Sind künftige Änderungen im Umfeld oder am Produkt selbst durch die Produkt- oder Prozessplanung beschrieben, z.B. textuelle Beschreibung „Gurtschlösser in neuem Derivat zukünftig rund statt eckig“ oder „schwarz statt silbern“, dient dies als Input für besondere, sogenannte generative adversarial networks (GAN). Damit ist es möglich, allein aus Textvorgaben Bilder zu erzeugen. Ebenso ist die Kombination mit vorhandenen Bildern möglich. Nun kann die geplante Änderung (Textinformation) zusammen mit dem Ausgangszustand (Bild; zum Beispiel zuvor zum Training einer Qualitätskontroll-KI verwendetes Bild ohne die vorgenommene Produktänderung oder Bild des Produkts ohne Änderung) von dem beschriebenen GAN so verarbeitet werden, dass gezielt die beschriebenen Änderungen im bestehenden Bild (sogenanntes GAN in-painting) erzeugt werden (z.B. kann im ursprünglich verwendeten Bild das silberne Gurtschloss mittels KI durch ein schwarzes Gurtschloss ersetzt werden, das dem neuen Produktdesign entspricht). Diese Bilder können nun verwendet werden, um: - zu prüfen, ob die bestehende KI Absicherung (z.B. Qualitätskontroll-KI) auch die künftigen Produktzustände bewerten kann; und - die bestehende KI anzulernen, falls die neuen Produktzustände (z.B. geändertes Design) nicht bewertet werden können.

**[0022]** Das oben Beschriebene lässt sich mit anderen Worten und auf eine konkrete Ausgestaltung bezogen, die als für die vorliegende Offenbarung nicht limitierend beschrieben wird, wie folgt zusammenfassen: In einem bestehenden Produktionsprozess werden Objekte, z.B. Gurtschlösser auf Vorhandensein im Fahrzeug durch eine Bildverarbeitung auf Basis von KI geprüft. Es ist geplant, eine Änderung an den Objekten, z.B. Gurtschlössern vorzunehmen (diese sollen rund statt eckig werden). Es werden gemäß dem offenbarten Verfahren nun: - Die Änderungsbeschreibung in Textform bereitgestellt; - Die aktuell angelernten, realen Bilder der Objekte, z.B. Gurtschlösser aus der Produktion bereitgestellt; und in-paintfähige GANs verwendet, um Bilder zu erzeugen, wie das künftige Produkt aussieht. Diese KI-generierten Bilder werden dann verfahrensgemäß verwendet, um: - Zu prüfen, ob die bestehende KI Absicherung (z.B. Qualitätskontroll-KI basierend auf den ursprünglichen Bildern der aktuellen Produktversion) auch die künftigen Produktzustände (geändertes Produkt) bewerten kann; und - Die bestehende KI anzulernen, falls mit der aktuell trainierten KI keine Produktbewertung des geplanten geänderten Produkts möglich ist.

**[0023]** Beim verfahrensgemäßen Erfassen der Bilddaten können also zum Beispiel aktuelle Bilddaten eines Produkts (z.B. Produkt mit aktuellem Gurtschloss) verwendet werden. Beim Erfassen der Textinformation kann z.B. eine geplante Produktänderung beschrieben werden (z.B. Beschreibung des geänderten Gurtschlösses). Die Textinformation kann also eine Variante des Objekts beschreiben (z.B. geplante Form- oder Farbänderung des Produkts oder zu prüfenden Objekts). Die textuelle/verbale Beschreibung kann z.B. mittels In-Paint durch eine KI in die aktuellen Bilddaten aufgenommen werden, sodass ein generiertes Bild mit dem geplanten Produkt entstehen kann. Vorteilhaft kann an dieser Ausführungsform sein, dass der Trainingsprozess bei Produktänderungen sehr einfach an einer bestehenden Hardware des Vorgängermodells stattfinden kann z.B. durch das textuelle/verbale Beschreiben der geplanten oder vorgesehenen Änderung.

**[0024]** Gemäß einem Aspekt der Offenbarung werden Trainingsbilddaten zur Verwendung zum Trainieren eines Modells für maschinelles Lernen zum visuellen Überwachen eines Objektes bei einer Produktion eines Kraftfahrzeugs und/oder einer Kraftfahrzeugkomponente bereitgestellt. Die Trainingsbilddaten wurden durch das oben beschriebene Verfahren erzeugt.

**[0025]** Gemäß einem Aspekt der Offenbarung wird ein Modell für maschinelles Lernen zum visuellen Überwachen eines Objekts bei einer Produktion eines Kraftfahrzeugs und/oder einer Kraftfahrzeugkomponente bereitgestellt. Das Modell wurde

anhand der oben beschriebenen Trainingsbilddaten trainiert.

**[0026]** Gemäß einem Aspekt der Offenbarung wird ein Verfahren zum visuellen Überwachen eines Objekts bei einer Produktion eines Kraftfahrzeugs und/oder einer Kraftfahrzeugkomponente unter Verwendung des oben beschriebenen Modells bereitgestellt.

**[0027]** Ferner wird ein Computerprogramm, umfassend Befehle, die bei der Ausführung des Programms durch einen Computer diesen veranlassen, das oben beschriebene Verfahren zumindest teilweise aus- bzw. durchzuführen, bereitgestellt.

**[0028]** Ein Programmcode des Computerprogramms kann in einem beliebigen Code vorliegen.

**[0029]** Das oben mit Bezug zum Verfahren Beschriebene gilt analog auch für das Computerprogramm und umgekehrt.

**[0030]** Ferner wird ein computerlesbares Medium, insbesondere ein computerlesbares Speichermedium, bereitgestellt. Das computerlesbare Medium umfasst Befehle, die bei der Ausführung des Programms durch einen Computer diesen veranlassen, das oben beschriebene Verfahren zumindest teilweise auszuführen.

**[0031]** Das heißt, es kann ein computerlesbares Medium bereitgestellt werden, das ein oben definiertes Computerprogramm umfasst. Bei dem computerlesbaren Medium kann es sich um ein beliebiges digitales Datenspeichergerät handeln, wie zum Beispiel einen USB-Stick, eine Festplatte, eine CD-ROM, eine SD-Karte oder eine SSD-Karte. Das Computerprogramm muss nicht zwingend auf einem solchen computerlesbarem Speichermedium gespeichert sein, sondern kann auch über das Internet oder anderweitig extern bezogen werden.

**[0032]** Ferner wird eine Datenverarbeitungsvorrichtung bereitgestellt, wobei die Datenverarbeitungsvorrichtung dazu eingerichtet ist, das oben beschriebene Verfahren zumindest teilweise aus- bzw. durchzuführen. Damit ist das Verfahren ein computerimplementiertes Verfahren.

**[0033]** Nachfolgend wird eine Ausführungsform mit Bezug zu den Figuren beschrieben.

**Fig. 1** zeigt schematisch ein Verfahren zum visuellen Überwachen eines Objekts bei einer Produktion eines Kraftfahrzeugs und/oder einer Kraftfahrzeugkomponente gemäß einem Aspekt der Offenbarung;

**Fig. 2** zeigt schematisch ein Erzeugen von Trainingsbilddaten nach einem Verfahren gemäß einem Aspekt der Offenbarung;

**Fig. 3** zeigt Bilddaten und gemäß einem Verfahren nach einem Aspekt der Offenbarung erzeugte Trainingsbilddaten; und

**Fig. 4** zeigt schematisch einen Ablauf eines Verfahrens zum Erzeugen von Trainingsbilddaten für ein Modell für maschinelles Lernen zum visuellen Überwachen eines Objektes bei einer Produktion eines Kraftfahrzeugs und/oder einer Kraftfahrzeugkomponente gemäß einem Aspekt der Offenbarung.

**[0034]** **Fig. 1** zeigt schematisch ein Verfahren zum visuellen Überwachen 400 eines Objekts 200 bei einer Produktion eines Kraftfahrzeugs 205 und/oder einer Kraftfahrzeugkomponente 210 gemäß einem Aspekt der Offenbarung.

**[0035]** Dabei zeigt **Fig. 1** oben rechts schematisch ein zu produzierendes Objekt 200, beispielsweise ein Kraftfahrzeug 205. Dabei wird ein computerimplementiertes und visuelles Überwachen 400 des Objekts 200 durchgeführt, wie schematisch durch eine Kameravorrichtung dargestellt. Die Kameravorrichtung erfasst das Objekt 200 betreffende Objektbilddaten. Die Objektbilddaten werden durch ein Modell 500 für maschinelles Lernen ausgewertet werden, um das Objekt 200 bei einer Produktion des Kraftfahrzeugs 205 und/oder einer Kraftfahrzeugkomponente 210 zu überwachen.

**[0036]** Dabei kann das Modell 500 wie folgt trainiert werden: Das Objekt 200 wird oder wurde in der Vergangenheit in einem Fehlerzustand 230 begutachtet, beispielsweise durch einen menschlichen Gutachter. Der Gutachter erstellt eine Textinformation 310, die den Fehlerzustand 230 in Worten beschreibt, beispielsweise „Spiegel zerkratzt“. Die Textinformation 310 kann in einer Datenbank gespeichert werden.

**[0037]** Die Textinformation 310 wird mit das Objekt 200 betreffenden Bilddaten 320 in eine Datenverarbeitungsvorrichtung 250 eingegeben. Die Datenverarbeitungsvorrichtung 250 ist dazu eingerichtet, einen generativen Transformer 260 auszuführen. Der generative Transformer 260 generiert anhand der Bilddaten 320 und der Textinformation 310 Trainingsbilddaten 300 (siehe **Fig. 2** und 3), die zum Trainieren des Modells 500 an eine Trainingsvorrichtung 510 übermittelt beziehungsweise ausgegeben werden. Die Trainingsvorrichtung 510 ist dazu eingerichtet, das Modell 500 zum Überwachen 400 des Objekts 200 zu trainieren. Dabei kann die Trainingsvorrichtung 510 beispielsweise die Trainingsbilddaten 300 zusammen mit einem sich aus der Textinformation 310 durch Klassifizierung der Textinformation

310 ergebenden Label verarbeiten, um das Modell 500 zu trainieren.

**[0038]** **Fig. 2** zeigt schematisch ein Erzeugen von Trainingsbilddaten 300 nach einem Verfahren 100 gemäß einem Aspekt der Offenbarung. Figur wird unter Bezugnahme auf **Fig. 1** beschrieben.

**[0039]** Gemäß **Fig. 2** liegen Bilddaten 320 eines Objekts 200 vor, beispielsweise einer Kraftfahrzeugkomponente 210, insbesondere eines Sitzes. Die Bilddaten 320 bilden das Objekt 200 in einem Referenzzustand 220 ab beziehungsweise stellen das Objekt 200 in dem Referenzzustand 220 dar. Die Bilddaten 320 sind beispielsweise Konstruktionsdaten, insbesondere CAD-Daten aus der Konstruktion, also roh beziehungsweise untexturiert, um die Geometrie des Objekts 200 zu definieren. Die Bilddaten 320 können somit allein als geometrische Maske vorliegen. Die Bilddaten 320 können auch Fotodaten des Objekts 200 sein (siehe **Fig. 3**). Die Bilddaten 320 werden in die Datenverarbeitungsvorrichtung 250 beziehungsweise den generativen Transformer 260 eingegeben, wie in **Fig. 2** schematisch gezeigt.

**[0040]** Es liegen Textinformationen 310 vor, die das Objekt 200 betreffen. Die Textinformationen 310 umfassen Informationen zu Varianten 235 des Objekts 200. Die Textinformationen 310 dienen als textuelle Beschreibung der in den Bilddaten 320 fehlenden und/oder gegenüber der Bilddaten 200 verschiedenen Merkmale des Objekts 200. Die Textinformationen 310 sind beispielsweise Produktions-Bestelldaten beziehungsweise Produktions-Orderdaten mit Informationen zu den Merkmalen des Objekts 200 in möglichen Varianten 235. Beispielsweise kann die Textinformation 310 umfassen: „Leder schwarz mit Kontraststeppung in rot“. Damit umfasst die Textinformation 310 beispielsweise drei Merkmale, die in den Bilddaten 320 nicht zu sehen sind, nämlich das Material (Leder) betreffend, die Grundfarbe (schwarz) betreffend und eine Naht (Kontraststeppung; rot) betreffend. Die Textinformationen 310 werden in die Datenverarbeitungsvorrichtung 250 beziehungsweise den generativen Transformer 260 eingegeben. Die Textinformation 310 können wir mit Bezug zu **Fig. 1** und 3 beschrieben auch einen Fehlerzustand 230 des Objekts 205 beschreiben und beispielsweise aus einer Datenbank abrufbar sein.

**[0041]** Die Datenverarbeitungsvorrichtung 250 beziehungsweise der generative Transformer 260 erzeugen wie schematisch in **Fig. 2** gezeigt anhand der Bilddaten 320 und der Textinformation 310 die Trainingsbilddaten 300. Dabei betreffen die Trainingsbilddaten 300 das Objekt 200, also die Kraftfahrzeugkomponente 210 mit gemäß der Textinformation 310 ausgeprägten Markmalen. Durch die Eingabe weiterer als der oben beispielhaft genannter

ten Textinformation 310 können als Ausgabe der Datenverarbeitungsvorrichtung 250 beziehungsweise des generativen Transformers 260 photorealistische Renderings möglicher Varianten 235 des Objekts 200 zur Verwendung als Trainingsbilddaten 300 erzeugt werden, die wie mit Bezug zu **Fig. 1** beschrieben in eine Trainingsvorrichtung 510 zum Trainieren des Modells 500 eingegeben werden können.

**[0042]** **Fig. 3** zeigt Bilddaten 320 und gemäß einem Verfahren 100 nach einem Aspekt der Offenbarung erzeugte Trainingsbilddaten 300. **Fig. 3** wird unter Bezugnahme auf **Fig. 1** und **2** beschrieben. Dabei zeigt **Fig. 3** vier Abbildungen (**Fig. 3(A)**, **3(B)**, **3(C)**, **3(D)**) mit jeweiligen Bilddaten 320 beziehungsweise Trainingsbilddaten 300.

**[0043]** **Fig. 3 (A)** zeigt Bilddaten 320 eines Objekts 200 als eine Kraftfahrzeugkomponente 210, die an einem Kraftfahrzeug 200 beispielsweise während der Produktion des Kraftfahrzeugs 200 montiert wurde. Das Objekt 200 wird überwacht, um ein hochqualitatives und möglichst fehlerfreies Kraftfahrzeug 200 produzieren zu können. **Fig. 3 (A)** zeigt dabei das Objekt 200 in einem Referenzzustand 220. Insbesondere zeigt **Fig. 3** einen Spiegel als Kraftfahrzeugkomponente 210, die fehlerfrei ist, also einem Zustand „in Ordnung“ zuordenbar ist beziehungsweise in dem Referenzzustand 220.

**[0044]** **Fig. 3 (B)**, **3 (C)** und **3 (D)** zeigen je ein Beispiel von anhand der Bilddaten 320 gemäß **Fig. 3 (A)** erzeugter Trainingsbilddaten 300. Dafür wurden die Bilddaten 320 gemäß **Fig. 3 (A)** zusammen mit einer Textinformation 310, beispielsweise „Spiegel beschädigt“ oder „Spiegel zerkratzt“ in eine Datenverarbeitungsvorrichtung 250 beziehungsweise einen generativen Transformer 260 eingegeben. Die Datenverarbeitungsvorrichtung 250 beziehungsweise der generative Transformer 260 liefern als Ausgabe die Trainingsbilddaten 300 gemäß **Fig. 3 (B)**, **3 (C)** und **3 (D)**. Dabei zeigen die Trainingsbilddaten 300 gemäß **Fig. 3 (B)**, **3 (C)** und **3 (D)** jeweils das Objekt 200 mit einem nicht in den Bilddaten 320 vorhandenen Merkmal, nämlich einer Beschädigung beziehungsweise einem Kratzer. Durch die Bilddaten 320 gemäß **Fig. 3 (A)** in Kombination mit der Textinformation 310 werden pro Bilddaten 320 und Textinformation 310 eine Mehrzahl von Trainingsbilddaten 300 erzeugt wie in **Fig. 3 (B)**, **3 (C)** und **3 (D)** gezeigt.

**[0045]** Die Trainingsbilddaten 300 werden, wie mit Bezug zu **Fig. 1** beschrieben, in eine Trainingsvorrichtung 510 zum Trainieren des Modells 500 eingegeben, um mögliche Fehler während der Produktion des Kraftfahrzeugs 200 an dem Objekt 200 erkennen zu können.

**[0046]** Die Textinformation 310 und/oder die Bilddaten 320 definieren optional einen Abschnitt des Objekts 200, wobei bei dem Verarbeiten 130 der Bilddaten 320 zum Erzeugen der Trainingsbilddaten 300 der Abschnitt anhand der Textinformation 310 modifiziert wird. Mit anderen Worten ist es möglich, ein aktuelles Foto als Bilddaten 320 bereitzustellen und in bestimmten Abschnitten beziehungsweise Teilen oder Zonen einen „inpaint“ Bereich zu definieren. Der „inpaint“ Bereich wird dann von der Text-to-Image-KI, also dem generativen Transformer 260, bearbeitet. D.h. es werden synthetische Bilder, z.B. mit Kratzer, als Trainingsbilddaten 300 erzeugt. Der Rest des Bildes, also die Bilddaten 320 außerhalb des Abschnitts, bleibt stabil, also bei einer Verarbeitung der Bilddaten 320 durch den generativen Transformer 260 unverändert. Die erzeugten Trainingsbilddaten 300 können nochmals eine KI, beispielsweise den generativen Transformer 260, durchlaufen, um Varianten zu erzeugen d.h. unterschiedliche Fehlerbilder.

**[0047]** **Fig. 4** zeigt schematisch einen Ablauf eines Verfahrens 100 zum Erzeugen von Trainingsbilddaten 300 für ein Modell für maschinelles Lernen zum visuellen Überwachen eines Objektes 200 bei einer Produktion eines Kraftfahrzeugs 205 und/oder einer Kraftfahrzeugkomponente 210 gemäß einem Aspekt der Offenbarung.

**[0048]** Das Verfahren 100 ist ein computerimplementiertes Verfahren 100 und kann wie mit Bezug zu **Fig. 1** und **2** beschrieben auf und/oder von einer Datenverarbeitungsvorrichtung 250 durchgeführt werden. Das Verfahren 100 wird unter Bezugnahme auf **Fig. 1** bis **3** beschrieben.

**[0049]** Gemäß **Fig. 4** erfolgt ein Erfassen 110 von das Objekt 200 betreffenden Textinformation 310. Die Textinformation 310 beschreiben einen Fehlerzustand 230 und/oder eine Variante 235 des Objekts 220.

**[0050]** Es erfolgt ein Erfassen 120 von das Objekt 200 betreffenden Bilddaten 320. Die Bilddaten 320 bilden das Objekt 200 in einem Referenzzustand 220 ab. Die Bilddaten 320 umfassen Objektfotos und/oder Konstruktionsdaten.

**[0051]** Es erfolgt ein Verarbeiten 130 der Bilddaten 320 in Kombination mit der Textinformation 310 zum Erzeugen der Trainingsbilddaten 300, wobei die Trainingsbilddaten 300 das Objekt 200 betreffende und gemäß der Textinformation 310 ausgeprägte Trainingsbilddaten 300 umfasst. Zum Erzeugen 130 der Trainingsbilddaten 200 wird ein generativer Transformer 260 verwendet. Bei dem Verarbeiten 130 der Bilddaten 320 in Kombination mit der Textinformation 310 werden pro Bilddaten 320 und Textinformation 310 eine Mehrzahl von Trainingsbilddaten 300

erzeugt. Die Textinformation 310 und/oder die Bilddaten 320 definieren einen Abschnitt des Objekts 200, wobei bei dem Verarbeiten 130 der Bilddaten 320 der Abschnitt anhand der Textinformation 310 modifiziert wird.

#### Bezugszeichenliste

100	Verfahren
110	Erfassen von Textinformation
120	Erfassen von Bilddaten
130	Verarbeiten
200	Objekt
205	Kraftfahrzeug
210	Kraftfahrzeugkomponente
220	Referenzzustand
230	Fehlerzustand
235	Variante
250	Datenverarbeitungsvorrichtung
260	generativer Transformer
300	Trainingsbilddaten
310	Textinformation
320	Bilddaten
400	Überwachen
500	Modell
510	Trainingsvorrichtung

#### Patentansprüche

1. Computerimplementiertes Verfahren (100) zum Erzeugen von Trainingsbilddaten (300) für ein Modell (500) für maschinelles Lernen zum visuellen Überwachen (400) eines Objektes (200) bei einer Produktion eines Kraftfahrzeugs (205) und/oder einer Kraftfahrzeugkomponente (210), wobei das Verfahren (100) aufweist:

- Erfassen (110) von das Objekt (200) betreffenden Textinformation (310);
- Erfassen (120) von das Objekt (200) betreffenden Bilddaten (320); und
- Verarbeiten (130) der Bilddaten (320) in Kombination mit der Textinformation (310) zum Erzeugen der Trainingsbilddaten (300), wobei die Trainingsbilddaten (300) das Objekt (200) betreffende und gemäß der Textinformation (310) ausgeprägte Trainingsbilddaten (300) umfasst.

2. Computerimplementiertes Verfahren (100) nach Anspruch 1, wobei die Bilddaten (320) das Objekt (200) in einem Referenzzustand (220) abbilden.

3. Computerimplementiertes Verfahren (100) nach Anspruch 1 oder 2, wobei die Textinformation (310) einen Fehlerzustand (230) und/oder eine Variante (235) des Objekts (220) beschreiben.

4. Computerimplementiertes Verfahren (100) nach einem der vorherigen Ansprüche, wobei die Textinformation (310) und/oder die Bilddaten (320) einen Abschnitt des Objekts (200) definieren, wobei bei dem Verarbeiten (130) der Bilddaten (320) der Abschnitt anhand der Textinformation (310) modifiziert wird.

5. Computerimplementiertes Verfahren (100) nach einem der vorherigen Ansprüche, wobei beim Verarbeiten (130) der Bilddaten (320) in Kombination mit der Textinformation (310) pro Bilddaten (320) und Textinformation (310) eine Mehrzahl von Trainingsbilddaten (300) erzeugt wird.

6. Computerimplementiertes Verfahren (100) nach einem der vorherigen Ansprüche, wobei die Bilddaten (320) Objektfotos und/oder Konstruktionsdaten umfassen.

7. Computerimplementiertes Verfahren (100) nach einem der vorherigen Ansprüche, wobei zum Erzeugen (130) der Trainingsbilddaten (200) ein generativer Transformer (260) verwendet wird.

8. Trainingsbilddaten (300) zur Verwendung zum Trainieren eines Modells (500) für maschinelles Lernen zum visuellen Überwachen (400) eines Objektes (200) bei einer Produktion eines Kraftfahrzeugs (205) und/oder einer Kraftfahrzeugkomponente (210), wobei die Trainingsbilddaten (200) durch ein Verfahren (100) nach einem der vorherigen Ansprüche erzeugt wurden.

9. Modell (500) für maschinelles Lernen zum visuellen Überwachen (400) eines Objektes (200) bei einer Produktion eines Kraftfahrzeugs (205) und/oder einer Kraftfahrzeugkomponente (210), wobei das Modell (500) anhand der Trainingsbilddaten nach Anspruch 8 trainiert wurde.

10. Verfahren zum visuellen Überwachen (400) eines Objektes (200) bei einer Produktion eines Kraftfahrzeugs (205) und/oder einer Kraftfahrzeugkomponente (210) unter Verwendung eines Modells (500) nach Anspruch 9.

11. Computerprogramm und/oder computerlesbares Medium, umfassend Befehle, die bei der Ausführung des Programms bzw. der Befehle durch einen Computer diesen veranlassen, das Verfahren (100) und/oder die Schritte des Verfahrens (100) nach einem der Ansprüche 1 bis 7 oder 10 durchzuführen.

12. Datenverarbeitungsvorrichtung (250), wobei die Datenverarbeitungsvorrichtung (250) dazu eingerichtet ist, das Verfahren (100) nach einem der Ansprüche 1 bis 7 oder 10 durchzuführen.

Es folgen 2 Seiten Zeichnungen

Anhängende Zeichnungen

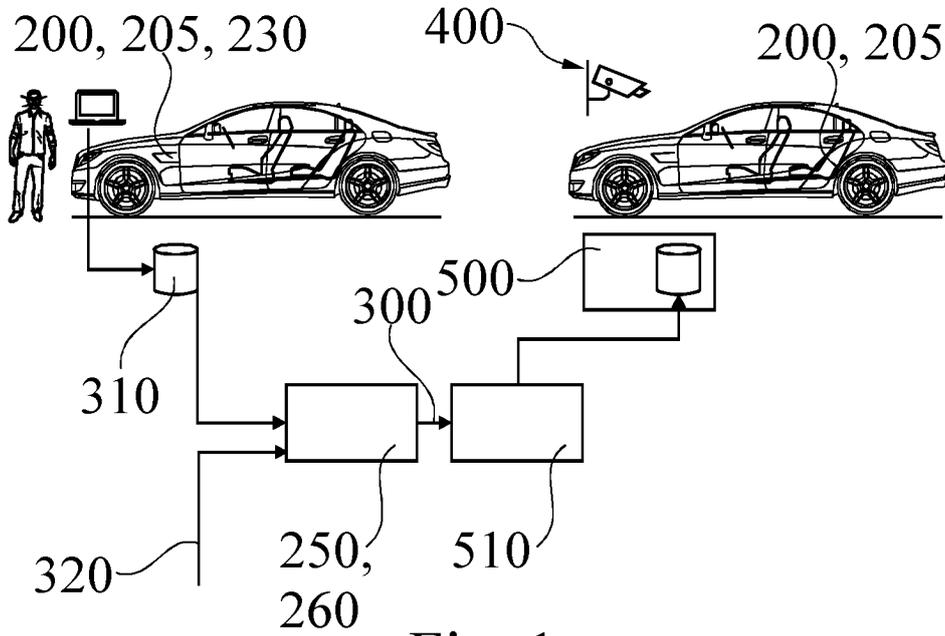


Fig. 1

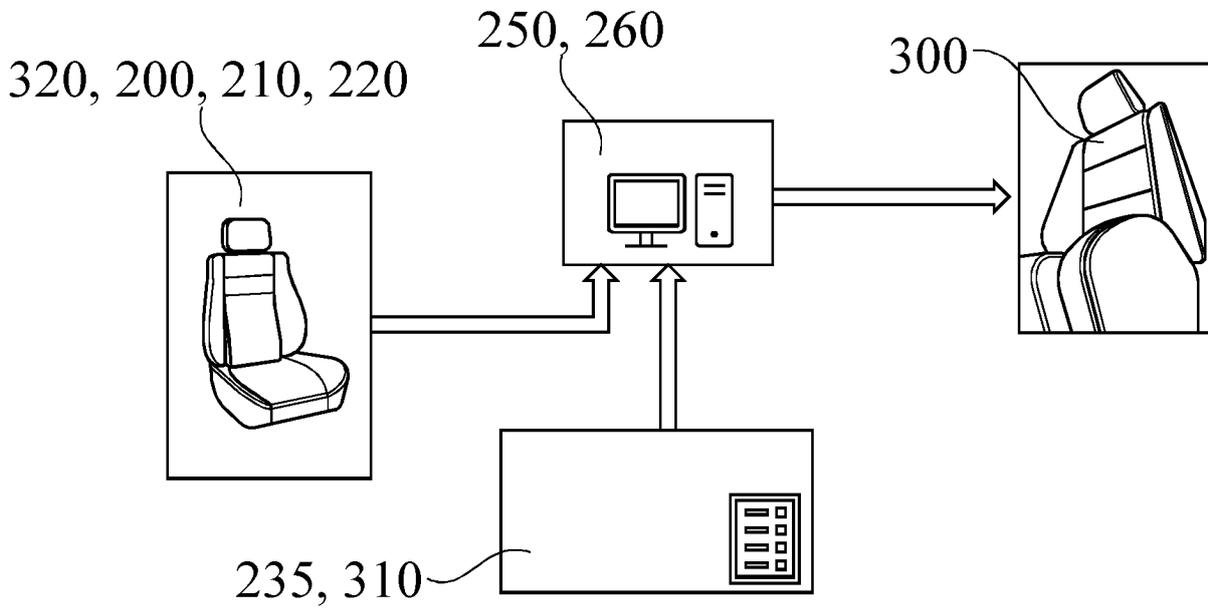


Fig. 2

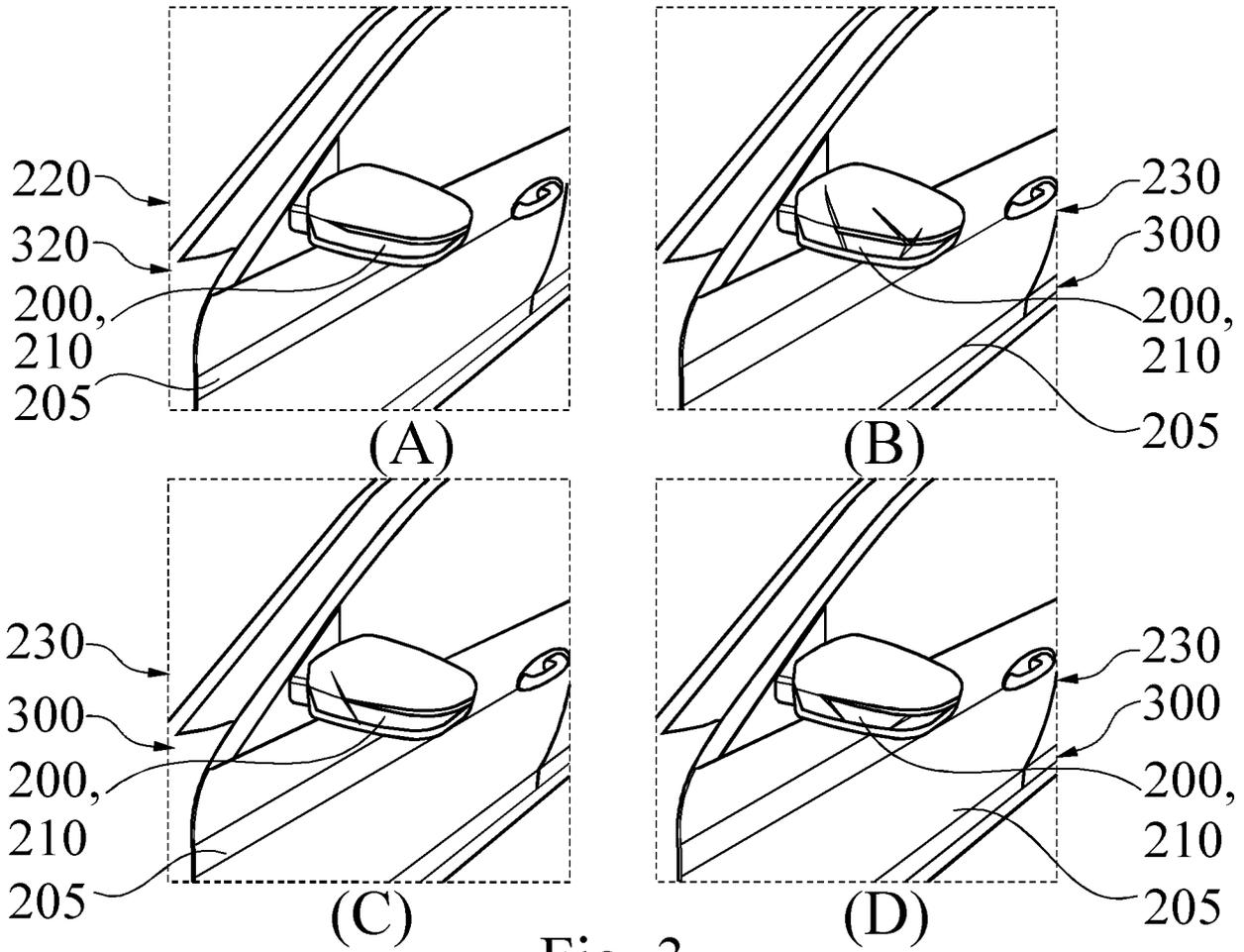


Fig. 3

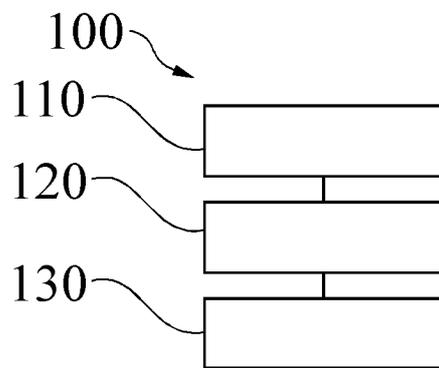


Fig. 4