HINREICHENDE SPEZIFIKATION EINER KI-FUNKTION AM BEISPIEL FUßGÄNGERSCHUTZ

Bernhard Knie

Fachbereich Industrial Engineering

Università degli studi di Roma Tor Vergata
Via Cracovia n.50
00133 Roma
E-Mail: b.knie@aol.com

Prof. Stefan Kubica

Fachbereich Wirtschaftsinformatik

Technische Hochschule Wildau
Hochschulring 1
15745 Wildau
E-Mail: skubica@th-wildau.de

ABSTRAKT

Transparenz und Erklärbarkeit sind essenzielle Faktoren bei der Entwicklung sicherer KI-Modelle für den Fußgängerschutz. Dies ermöglicht nicht nur eine verbesserte Akzeptanz, sondern fördert auch das Vertrauen der Nutzer in die präventiven Fähigkeiten dieser Systeme.

Die Integration von Sensoren, maschinellem Lernen und fortschrittlichen Algorithmen ermöglicht eine präzise Analyse von Fußgängerverhalten und Umgebungsfaktoren. Dabei spielt die Echtzeitverarbeitung eine entscheidende Rolle, um schnell auf sich ändernde Situationen zu reagieren.

Zusätzlich zur Technologieentwicklung betont diese Analyse die Notwendigkeit strikter Sicherheitsmaßnahmen, um potenzielle Risiken und Sicherheitslücken zu minimieren. Dazu gehören Sicherheitsprotokolle für die Datenspeicherung, den Modellschutz und den Zugriff auf KI-Systeme.

Die Einhaltung gesetzlicher und ethischer Richtlinien ist ein weiterer wichtiger Aspekt. Regulatorische Vorgaben und ethische Standards sollen sicherstellen, dass KI-Technologien im Automobilsektor nicht nur sicher, sondern auch im Einklang mit gesellschaftlichen Werten und Normen eingesetzt werden.

Die Bedeutung von sicherer KI im Kontext des Fußgängerschutzes im Automobilsektor wird weiter wachsen. Durch die richtige Balance zwischen Technologieentwicklung, Sicherheitsmaßnahmen und regulatorischer Compliance kann Künstliche Intelligenz dazu beitragen, die Verkehrssicherheit zu verbessern und Unfälle zu minimieren.

SCHLÜSSELWÖRTER

KI-Modelle, Automotive, Fußgängerschutz

1 HINTERGRUND UND MOTIVATION

Die Entwicklung von KI-Funktionen, abgebildet durch tiefe neuronale Netze für den Fußgängerschutz ist eine Reaktion auf die zunehmende Komplexität und Vielfalt von Verkehrssituationen, insbesondere in städtischen Gebieten. Traditionelle Bilderkennungsalgorithmen haben Schwierigkeiten, die Vielfalt menschlicher Bewegungen und die komplexe Umgebung angemessen zu erfassen. Tiefe neuronale Netze, insbesondere Convolutional Neural Networks (CNNs), haben sich als leistungsfähige Werkzeuge erwiesen, um die Herausforderungen im Fußgängerschutz zu bewältigen. Tiefe neuronale Netze können hierarchisch abstrahierte Merkmale aus Bildern extrahieren. In Bezug auf den Fußgängerschutz ermöglicht dies eine automatische Lernfähigkeit, um komplexe Muster und Merkmale, die für die Erkennung von Fußgängern relevant sind, zu identifizieren. Zudem ermöglichen sie eine bessere Bewältigung von Varianzen. Fußgänger können in unterschiedlichen Umgebungen, bei verschiedenen Lichtverhältnissen und in unterschiedlichen Positionen auftreten. Tiefe neuronale Netze können lernen, mit dieser Varianz umzugehen und robuste Modelle zu entwickeln, die in verschiedenen Szenarien gut funktionieren. [1]

Ein weiterer Aspekt ist das End-to-End-Lernen, bei dem das Modell direkt aus den Rohdaten lernt, ohne manuell definierte Merkmale. Dies ist besonders vorteilhaft, wenn komplexe Beziehungen zwischen den Eingabeinformationen und der Fußgängererkennung schwer vorhersehbar sind. Schlüssel zum Erfolg sind jedoch vor allem große, zusammenhängende Datensätze. Tiefe neuronale Netze profitieren von großen Trainingsdatensätzen. Mit der Zunahme verfügbarer Daten können tiefe Modelle besser

generalisieren und Muster erkennen, was zu einer verbesserten Leistung in der Fußgängererkennung führt. Die rasante Entwicklung im Bereich der tiefen neuronalen Netze wird durch kontinuierliche Forschung vorangetrieben. Neue Architekturen, Optimierungstechniken und Trainingsstrategien ermöglichen eine ständige Verbesserung der Leistung von Modellen im Fußgängerschutz.

2 PROBLEMSTELLUNG

Der Einsatz von Künstlicher Intelligenz im Bereich des Fußgängerschutzes bringt zwar vielversprechende Fortschritte hinsichtlich der Verkehrssicherheit, wirft jedoch auch einige Herausforderungen und Problemstellungen auf, die sorgfältig berücksichtigt werden müssen:

Unvorhersehbare menschliche Interaktionen:
Fußgänger verhalten sich oft weniger vorhersehbar als andere Verkehrsteilnehmer. Ihre Bewegungen können durch verschiedene Faktoren wie Ablenkung, unerwartete Handlungen oder das Überqueren außerhalb von Fußgängerüberwegen geprägt sein. KI-Funktionen müssen in der Lage sein, diese unvorhersehbaren Verhaltensweisen zu erkennen und darauf angemessen zu reagieren.

Vielfalt der Umgebungsfaktoren:
Fußgänger bewegen sich in unterschiedlichsten Umgebungen, sei es in städtischen Gebieten, ländlichen Regionen oder auf stark frequentierten Straßen. Die Vielfalt der Umgebungsfaktoren, wie unterschiedliche Lichtverhältnisse, Wetterbedingungen und Straßentypen, stellt eine Herausforderung für die KI dar, die Modelle müssen robust gegenüber solchen Variablen sein.

Datenschutz und Ethik:
Die Verarbeitung von Informationen über Fußgänger, insbesondere durch KI-Funktionen in vernetzten Fahrzeugen, wirft Datenschutz- und Ethikfragen auf. Wie werden die Daten gesammelt, gespeichert und genutzt? Wie wird die Privatsphäre der Fußgänger geschützt? Es ist wichtig, einen ausgewogenen Ansatz zu finden, der sowohl die Sicherheit als auch die Privatsphäre respektiert.

Fehlende Normen und Standards:
Im Bereich der KI für den Fußgängerschutz fehlen oft einheitliche Normen und Standards. Die Entwicklung von Richtlinien und Normen ist entscheidend, um eine konsistente und sichere Implementierung von KI-Systemen zu gewährleisten und das Vertrauen der Öffentlichkeit zu stärken.

Adversarial Attacks:
KI-Modelle im Fußgängerschutz könnten anfällig für Adversarial Attacks sein, bei denen absichtlich gestaltete Eingaben die KI dazu verleiten, falsche Entscheidungen zu treffen. Der Schutz vor solchen Angriffen erfordert fortschrittliche Sicherheitsmaßnahmen.

Haftungsfragen und rechtliche Aspekte:
Im Falle von Unfällen oder Fehlfunktionen von KI-gesteuerten Fahrzeugen stellt sich die Frage der Haftung. Klare rechtliche Rahmenbedingungen und Verantwortlichkeiten müssen definiert werden, um Rechtsunsicherheiten zu minimieren.

Die Lösung dieser Problemstellungen erfordert eine multidisziplinäre Herangehensweise, die technologische, ethische, rechtliche und soziale Aspekte gleichermaßen berücksichtigt. Es ist entscheidend, dass Entwickler, Forscher, Gesetzgeber und die Gesellschaft gemeinsam daran arbeiten, sichere und verantwortungsbewusste KI-Systeme im Bereich des Fußgängerschutzes zu gewährleisten.

3 ZIELSETZUNG

Das übergeordnete Ziel besteht darin, eine robuste und verlässliche KI-Funktion zu entwickeln, die den Fußgängerschutz in Verkehrsszenarien optimiert. Die Zielsetzung umfasst mehrere spezifische Aspekte. Elementar ist zunächst eine präzise Fußgängererkennung, bei der die KI-Funktion in der Lage ist, Fußgänger unabhängig von deren Position, Bewegung oder Umgebungsfaktoren zu erkennen. Dies gewährleistet eine zuverlässige Grundlage für weitere Sicherheitsmaßnahmen. Desweiteren muss eine verlässliche Verhaltensvorhersage sichergestellt werden. Die KI soll in der Lage sein, das Verhalten von Fußgängern akkurat zu modellieren und vorherzusagen. Dies ermöglicht es, potenziell gefährliche Situationen frühzeitig zu identifizieren und angemessen darauf zu reagieren. Ein weiterer Aspekt ist die Echtzeit-Reaktionsfähigkeit.

Die KI-Funktion muss in Echtzeit auf sich ändernde Verkehrssituationen zu reagieren, insbesondere wenn Fußgänger unvorhersehbares Verhalten zeigen. Eine kurze Reaktionszeit und hohe Präzision ist entscheidend, um Kollisionen zu vermeiden. Zudem muss die Bilderkennung Fehlerresistenz und Robustheit aufweisen. Insbesondere die Objekterkennung soll robust gegenüber Fehlern und Unsicherheiten sein. Dies beinhaltet die Fähigkeit, auch in herausfordernden Umgebungen wie schlechtem Wetter oder ungünstigen Lichtverhältnissen zuverlässig zu funktionieren. Erreicht werden kann dies nur mithilfe einer Sensorfusion für umfassende Wahrnehmung. Daher müssen verschiedene Sensorinformationen, wie Kamera, Radar und Lidar, effektiv fusionieren, um eine umfassende Wahrnehmung der Verkehrsumgebung und der Fußgänger zu gewährleisten.

Schlussendlich muss die KI-Funktion automatisch und präzise Notbremsungen oder andere Kollisionsvermeidungsmaßnahmen einleiten, wenn eine potenzielle Gefahr für Fußgänger erkannt wird. Ein weiterer Aspekt ist die Anpassungsfähigkeit an verschiedene Verkehrsszenarien. Die entwickelte KI-Funktion soll sich nahtlos an verschiedene Verkehrsumgebungen anpassen können, darunter städtische Gebiete, ländliche Straßen oder komplexe Kreuzungen unter Einhaltung ethischer und rechtlicher Standards.

Die Zielsetzung umfasst auch die Entwicklung der KI-Funktion unter Berücksichtigung ethischer Prinzipien und geltender rechtlicher Standards. Datenschutz und Privatsphäre sollen dabei ebenso berücksichtigt werden wie die Einhaltung gesetzlicher Vorgaben. Durch die Verfolgung dieser Zielsetzung streben wir die Schaffung einer innovativen und sicheren KI-Funktion im Fußgängerschutz an, die die Verkehrssicherheit erhöht und gleichzeitig ethische und rechtliche Anforderungen erfüllt.

4 EXEMPLARISCHE UMSETZUNG DER FUNKTION

Die Fußgängererkennung durch Künstliche Intelligenz ist ein komplexer Prozess, der Bildverarbeitung, maschinelles Lernen und neuronale Netze integriert. Zunächst muss die Funktion mit der Erfassung von Bilddaten beginnen. Dies kann von verschiedenen Sensoren wie Kameras aufgenommen werden, die am Fahrzeug angebracht sind. Die aufgenommenen Bilddaten durchlaufen anschließend eine Datenpräverarbeitung. Hierzu gehören Schritte wie Rauschunterdrückung, Kontrastanpassung und möglicherweise die Normalisierung der Helligkeit. Ein Objekterkennungsalgorithmus wird auf die vorverarbeiteten Bilddaten angewendet. Dieser Algorithmus identifiziert potenzielle Objekte im Bild, einschließlich Fußgänger. Basierend auf den erkannten Objekten wird eine Region-of-Interest (ROI) um jeden potenziellen Fußgänger definiert. Dies hilft, den Fokus auf relevante Bereiche des Bildes zu legen und die Verarbeitung zu optimieren. Die Merkmale der erkannten Objekte, insbesondere der Fußgänger, werden extrahiert. Dies kann das Aussehen, die Größe, die Form und andere relevante visuelle Merkmale umfassen. Darauffolgend werden die extrahierten Merkmale einem trainierten neuronalen Netzwerk zugeführt.

Ein tiefes neuronales Netzwerk (Deep Neural Network, DNN) ist hierbei typisch. Das Netzwerk wurde zuvor mit großen Datensätzen von Fußgänger- und Nicht-Fußgängerbildern trainiert. Das neuronale Netzwerk verwendet Aktivierungsfunktionen und gewichtete Verbindungen zwischen Neuronen, um Entscheidungen darüber zu treffen, ob das erkannte Objekt tatsächlich ein Fußgänger ist. Entscheidend ist hier, dass die Ausgabe des neuronalen Netzwerks auf einen Schwellenwert angewendet wird, um zu bestimmen, ob ein erkanntes Objekt als Fußgänger klassifiziert wird. Dies hilft, Fehlklassifikationen zu minimieren. In einigen Anwendungen kann die Fußgängererkennung auch mit Daten anderer Sensoren kombiniert werden, um die Position der Fußgänger im dreidimensionalen Raum zu bestimmen. Dies ist insbesondere für autonome Fahrzeuge wichtig. Basierend auf den Ergebnissen der Fußgängererkennung trifft das Fahrzeug Entscheidungen und reagiert entsprechend. Dies kann das Auslösen eines Bremsvorgangs, eine Anpassung der Geschwindigkeit oder das Einleiten von Ausweichmanövern umfassen. Abbildung 1 skizziert den Funktionsablauf zusammenfassend und gibt einen Überblick über die grundlegenden Schritte in der Anwendung einer KI-Fußgängererkennungsfunktion.

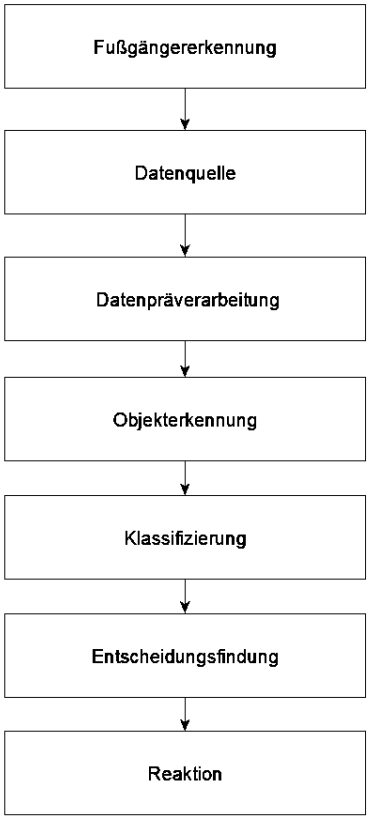

Abbildung 1: Hierarchischer Ablauf der KI-Funktion zur Fußgängererkennung.

Die spezifische Implementierung kann je nach den Anforderungen der Anwendung und den verfügbaren Sensordaten variieren.

5 GOAL STRUCTURING NOTATION

Die Goal Structuring Notation (GSN) ist eine Methode, die in der Entwicklung sicherheitskritischer Systeme verwendet wird, um die rationale Struktur von Argumentationen für sicherheitsrelevante Ziele und Entscheidungen zu modellieren. Die GSN bietet eine klare und strukturierte Darstellung von Sicherheitsargumentationen. Durch die Verwendung von Diagrammen und Notationen wird die Komplexität reduziert, was zu einer leichteren Verständlichkeit für alle Beteiligten führt. Dabei hilft sie den Argumentationsprozess transparent zu machen. Durch die sichtbare Darstellung von Argumenten und den Beziehungen zwischen verschiedenen Sicherheitszielen und -entscheidungen wird es einfacher, die Sicherheitsannahmen und -überlegungen nachzuvollziehen. Hierbei ermöglicht sie eine klare Verbindung zwischen den Sicherheitszielen und den Argumenten, die diese Ziele unterstützen.

Dies erleichtert die Überprüfung der Konsistenz und Relevanz der Argumentation im Hinblick auf die gesetzten Sicherheitsziele. Darüber hinaus erlaubt die Notation die Wiederverwendung von Argumentationsstrukturen. In sicherheitskritischen Systemen gibt es oft ähnliche Sicherheitsziele oder -entscheidungen, wodurch es sinnvoll ist, bewährte Argumentationsstrukturen zu speichern und in verschiedenen Kontexten wiederzuverwenden. Ein weiterer Faktor ist die Darstellung von klaren Strukturen und Beziehungen. So können Sicherheitsexperten und Prüfer die Argumentation leicht überprüfen. Dies trägt zur Qualitätssicherung und zur Identifizierung von Schwachstellen im Sicherheitskonzept bei. Die GSN ist somit auch ein effektives Mittel zur Dokumentation von Sicherheitsargumentationen. Diese Dokumentation ist nicht nur für Entwickler und Ingenieure, sondern auch für Aufsichtsbehörden, Prüfer und andere Interessengruppen von großer Bedeutung. Sie wird in vielen sicherheitskritischen Branchen, wie der Luft- und Raumfahrt, der Medizintechnik und der Automobilindustrie, eingesetzt. Sie unterstützt die Einhaltung von Sicherheitsstandards, indem sie eine strukturierte Darstellung der Sicherheitsargumentation ermöglicht. Insgesamt trägt die Goal Structuring Notation dazu bei, die Entwicklung sicherheitskritischer Systeme durch eine klare, transparente und überprüfbare Argumentationsstruktur zu unterstützen. Dies ist besonders wichtig in Bereichen, in denen die Einhaltung von Sicherheitsstandards und die nachvollziehbare Dokumentation von Sicherheitsüberlegungen von höchster Bedeutung sind.

6 ASSURANCE CASE

Ein Assurance Case ist eine formale Argumentationsstruktur, die verwendet wird, um die Sicherheit, Verlässlichkeit oder Konformität eines Systems zu dokumentieren und zu demonstrieren. Diese Argumentationsstruktur besteht aus einer Hierarchie von Claims (Behauptungen), die mit Argumenten und Evidenzen gestützt werden. Der Zweck eines Assurance Cases besteht darin, das Vertrauen in die Fähigkeiten eines Systems zu stärken und die Überzeugung zu vermitteln, dass das System sicher und zuverlässig ist.

Die Behauptung ist die Hauptaussage, die im Assurance Case vertreten wird. Sie könnte eine Sicherheitsaussage, eine Leistungsaussage oder eine Konformitätsaussage sein. Zum Beispiel: "Das autonome Fahrzeug gewährleistet die Sicherheit von Fußgängern“. Anschließend werden Ziele und falls erforderlich Unterziele definiert. Die Annahme ist die rationale Randbedingung, die die Behauptung unterstützt. Sie kann aus mehreren Subannahmen oder Argumenten bestehen, die aufeinander aufbauen, um die Hauptbehauptung zu stützen. Fälle die sich innerhalb und außerhalb der Operational Design Domain (ODD) befinden, werden zuverlässig erkannt. Die Evidenzen sind konkrete Daten, Fakten oder Ergebnisse, die dazu dienen, die Argumente zu unterstützen. Das können Forschungsergebnisse, Testergebnisse, Simulationen oder andere nachprüfbare Informationen sein. Beispielsweise könnten Evidenzen für eine präzise Fußgängererkennung Testergebnisse von umfangreichen Datensätzen von Fußgängererkennungsalgorithmen sein. Hierbei verdeutlich Abbildung 2 die Struktur mit der letztendlich die Argumentation unter Einbindung der GSN realisiert werden kann.

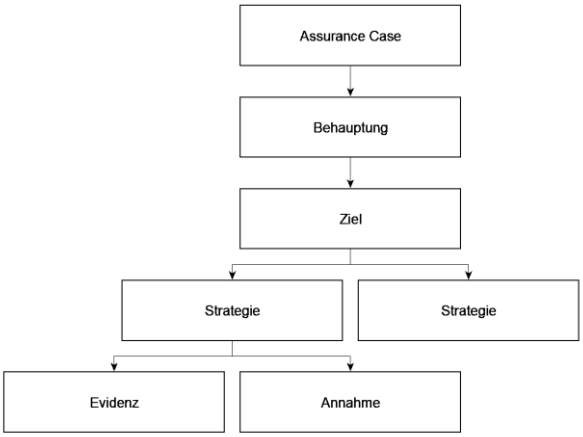

Abbildung 2: Generelle Struktur eines Assurance Case.

Darüber hinaus müssen Gegenmaßnahmen definiert werden, die ergriffen wurden, um Risiken zu mitigieren und die Sicherheit zu gewährleisten. Wenn beispielsweise identifizierte Risiken eine potenzielle Fehlfunktion der Fußgängererkennung beinhalten, könnten Gegenmaßnahmen redundante Sensoren oder kontinuierliche Überwachungssysteme sein. Eine Risikobewertung bewertet potenzielle Risiken und deren Auswirkungen auf das System. Dies hilft bei der Identifizierung kritischer Bereiche, die besondere Aufmerksamkeit erfordern. Auch hier muss Bezug auf einschlägige Standards und Normen genommen werden, um sicherzustellen, dass das System den allgemein anerkannten Anforderungen entspricht.

7 KOMBINATION BEIDER KONZEPTE

Die Goal Structuring Notation und der Assurance Case sind eng miteinander verbundene Konzepte und können effektiv miteinander kombiniert werden, um komplexe Sicherheitsargumentationen und -strukturen zu modellieren. Grundgerüst ist eine Gesamtstruktur in der GSN, die die zentrale Sicherheitsbehauptung (Claim) und die unterstützenden Ziele (Goals), Strategien (Strategies), Beweise (Evidence) und Annahmen (Assumptions) enthält. Jedes GSN-Element (Claim, Goal, Strategy, Evidence, Assumption) ist als eine spezifische Sicherheitsbehauptung im Assurance Case zu betrachten. Jedes Element trägt zur Gesamtargumentation bei. Anschließend erfolgt die Goal Decomposition in der GSN, um übergeordnete Sicherheitsziele in untergeordnete Ziele zu unterteilen. Jedes untergeordnete Ziel wird zu einer spezifischen Behauptung im Assurance Case. Strategien in der GSN mit konkreten Maßnahmen oder Nachweisen im Assurance

Case werden verbunden. Jede Strategie sollte durch geeignete Beweise gestützt werden, um ihre Wirksamkeit zu zeigen.

Diese Evidenz-Elemente in der GSN sind notwendig um die Sammlung von Beweisen für jede Sicherheitsbehauptung darzustellen. Dies können Daten aus Tests, Simulationen, Analysen oder anderen Quellen sein. Alle Annahmen in der GSN können grundsätzlich als Risikofaktoren betrachtet werden, die im Assurance Case transparent gemacht und gemanagt werden müssen. Diese Annahmen sollten klar dokumentiert und überwacht werden.

Jederzeit sind klare Verbindungen und Beziehungen zwischen den GSN-Elementen im Assurance Case notwendig. Zum Beispiel sollte jedes Goal direkt zur Unterstützung eines übergeordneten Claims beitragen. Sollten in der GSN besonders sicherheitskritische Punkte oder Aspekte vorliegen, müssen sie in einem Assurance Case besonders gründlich überprüft und dokumentiert werden. Standardisierte Diagramme und Grafiken existieren in beiden Notationsformen, um die Argumentationsstrukturen visuell darzustellen. Dies kann die Verständlichkeit und Klarheit der Sicherheitsargumentation verbessern. Da sowohl GSN als auch Assurance Case lebende Dokumente sind, sollten sie kontinuierlich aktualisiert und angepasst werden, um Änderungen im System, in den Anforderungen oder im Sicherheitskontext widerzuspiegeln.

Die Kombination von GSN und Assurance Case bietet eine strukturierte und systematische Möglichkeit, Sicherheitsargumentationen zu modellieren und zu dokumentieren, was insbesondere in sicherheitskritischen Anwendungen wie autonomen Fahrzeugen von entscheidender Bedeutung ist. [6]

Abbildung 3 zeigt einen exemplarischen GSN-Baum mit Fokus auf die Softwarequalität einer KI-Funktion. Zu Beginn ist das übergeordnete Ziel G1 Wartbarkeit (Maintainability) definiert. Die Strategie S1 verweist auf die entsprechenden Unterziele G1.1 bis G1.5, sowie auf einen Lösungsansatz E.1. Entscheidend ist der stringente Aufbau des Baumes entlang der Argumentationskette. Die Qualitätskriterien Wiederverwendbarkeit, Testbarkeit, Modifizierbarkeit und Analysierbarkeit bilden hier die Unterziele des Hauptziels Wartbarkeit. Der Lösungsansatz ist in dem Beispiel ein Verweis auf einschlägige Standards zur Softwarequalität wie z.B. die ISO 25010 Norm für Qualitätskriterien von Software, IT-Systemen und Software-Engineering.

Die Norm bietet ein Modell zur Evaluierung und Spezifikation der Qualitätsmerkmale für Softwareprodukte. Die ISO/IEC 25010 definiert acht Hauptqualitätsmerkmale, die weiter in Untermerkmale unterteilt sind. Wartbarkeit ist eines der Untermerkmale und ermöglicht Entwicklern und Qualitätsmanagern sicherzustellen, dass Softwareprodukte die erforderlichen Qualitätsstandards erfüllen.

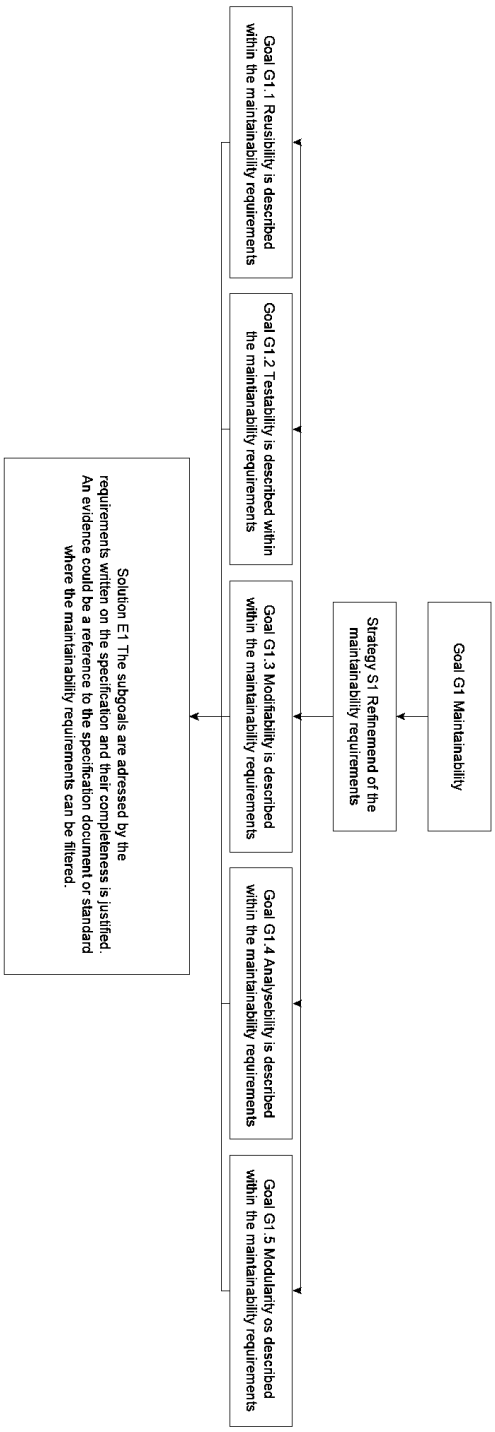

Abbildung 3: Beispiel eines GSN-Baums zur Fußgängererkennung mit Fokus auf die Softwarequalität des Algorithmus.

8 ZUSAMMENFASSUNG

Abschließend kann mit den zuvor genannten Methoden und Werkzeugen eine hinreichende Spezifikation einer KI-Funktion im Fußgängerschutz realisiert werden. Entscheidend ist hierbei vor allem die genaue Überprüfung der Randbedingungen unter Einbindung aller Aspekte der Entwicklung, Implementierung und Bewertung sowie die Hinzunahme von Standards und Normen. Die größte Herausforderung bleibt, dass viele KI-Modelle

weiterhin als "Black Box" betrachtet werden, da es schwierig sein kann, genau nachzuvollziehen, wie sie zu bestimmten Entscheidungen kommen. Die Undurchsichtigkeit von KI-Systemen erschwert die Identifizierung von Fehlern, unerwartetem Verhalten oder Diskriminierung. Es ist wichtig zu betonen, dass die Sicherheit von KI ein laufender Prozess ist, der sich mit der Entwicklung neuer Technologien und Angriffsmethoden weiterentwickeln wird. Die oben genannten Maßnahmen dienen als Ausgangspunkt.

LITERATUR

[1] Harald Rueß; Simon Burton (2022): Safe AI - How is this Possible? In: CoRR abs/2201.10436.

[2] Gauerhof, Lydia; Munk, Peter; Burton, Simon (2018): Structuring Validation Targets of a Machine Learning Function Applied to Automated Driving. In: Computer Safety, Reliability, and Security: Springer International Publishing, S. 45–58.

[3] Assion, Felix; Schlicht, Peter; Greßner, Florens; Günther, Wiebke; Hüger, Fabian; Schmidt, Nico; Rasheed, Umair (2019): The Attack Generator: A Systematic Approach towards Constructing Adversarial Attacks. In: Proc. 2019 IEEE Conf. Comput. Vision and Pattern Recognition Workshops.

[4] Burton, Simon; Gauerhof, Lydia; Sethy, Bibhuti Bhusan; Habli, Ibrahim; Hawkins, Richard (2019): Confidence Arguments for Evidence of Performance in Machine Learning for Highly Automated Driving Functions. In: Computer Safety, Reliability, and Security, Bd. 11699: Springer International Publishing, S. 365–377.

[5] Geirhos, Robert; Rubisch, Patricia; Michaelis, Claudio; Bethge, Matthias; Wichmann, Felix A.; Brendel, Wieland (2018): ImageNet-Trained CNNs Are Biased towards Texture; Increasing Shape Bias Improves Accuracy and Robustness. In: Proc. 7th Int. Conf. Learning Representations.

[6] Schwalbe, Gesina; Knie, Bernhard; Sämann, Timo; Dobberphul, Timo; Gauerhof, Lydia; Raafatnia, Shervin; Rocco, Vittorio (2020): Structuring the Safety Argumentation for Deep Neural Network Based Perception in Automotive Applications. In: António Casimiro, Frank Ortmeier, Erwin Schoitsch, Friedemann Bitsch und Pedro Ferreira (Hg.): Computer Safety, Reliability, and Security. SAFECOMP 2020 Workshops. Cham: Springer International Publishing, S. 383–394.

[7] ISO (2017): ISO/IEC/IEEE 12207:2017: Systems and Software Engineering Software Life Cycle Processes. 1. Aufl.

[8] ISO (2019): ISO/PAS 21448:2019(En): Road Vehicles Safety of the Intended Functionality.

[9] 16th ACM/IEEE Int. Conf. Formal Methods and Models for System Design (2018): IEEE.

[10] Burton, Simon; Herd, Benjamin (2023): Addressing uncertainty in the safety assurance of machine-learning. In: Frontiers in Computer Science 5. DOI: 10.3389/fcomp.2023.1132580.

[11] Carlini, Nicholas; Wagner, David (2017): Adversarial Examples Are Not Easily Detected: Bypassing Ten Detection Methods. In: Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security: Association for Computing Machinery (AISec '17), S. 3–14.

[12] Schleiss, P., Carella, F., & Kurzidem, I. (2022, November). Towards continuous safety assurance for autonomous systems. In 2022 6th International Conference on System Reliability and Safety (ICSRS) (pp. 457-462). IEEE.

[13] Kindermans, Pieter-Jan; Schütt, Kristof T.; Alber, Maximilian; Müller, Klaus-Robert; Erhan, Dumitru; Kim, Been; Dähne, Sven (2018): Learning How to Explain Neural Networks: PatternNet and PatternAttribution. In: Proc. 6th Int. Conf. on Learning Representations.

[14] Bagschik, G.; Menzel, T.; Maurer, M. (2018): Ontology Based Scene Creation for the Development of Automated Vehicles. In: Proc. 2018 IEEE Intelligent Vehicles Symp: IEEE, S. 1813–1820.

[15] Computer Safety, Reliability, and Security (2018): Springer International Publishing.

[16] Bojarski, Mariusz, et al. "End to end learning for self-driving cars." arXiv preprint arXiv:1604.07316 (2016).

[17] Proc. 10th European Congress Embedded Real Time Systems (2020).

[18] Proc. 2018 IEEE Intelligent Vehicles Symp (2018): IEEE.

[19] Bernhard, Julian, et al. "Risk-based safety envelopes for autonomous vehicles under perception uncertainty." 2022 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2022.

[20] Leveson, Nancy (2012): Engineering a Safer World: Systems Thinking Applied to Safety: MIT Press (Engineering Systems).