3D Gaussian Splatting auf Basis von LiDAR Punktwolken

TIM SCHIMANSKY[1] & MONIKA SESTER[1]

Zusammenfassung: 3D Gaussian Splatting ist eine neuartige Methode zur impliziten Darstellung von 3D-Szenen mittels 3D Gaussverteilungen. Die implizite Darstellung wird aus einem Satz von Trainingsbildern erlernt und berücksichtigt dabei in der dreidimensionalen Darstellung den Betrachtungswinkel, was indirekt zu einer Modellierung von Reflexionen/Farbveränderung usw. führt. Um die Anwendung von 3D Gaussian Splatting in neuen Szenarien zu verbessern, schlagen wir vor, diesen Ansatz mit LiDAR-Daten aus einem Mobile Mapping System zu kombinieren, um den bisher nötigen Structure-from-Motion Schritt zu ersetzen. Dazu werden die bekannten Kamerapositionen sowie die gegebene Punktwolke verwendet. Im Beitrag wird die Grundidee vorgestellt und erste Experimente dazu beschrieben.

1 Einleitung

Die Bildsynthese von neuen Blickwinkeln in 3D Umgebungen ist in den letzten Jahren ein zentrales Thema der Forschung. Die Nutzung reicht von der reinen Visualisierung für Unterhaltungszwecke (z.B. in VR-Umgebungen oder für VFX) (VACHHA 2023) über professionelle Einsatzgebiete im Bereich der Digitalen Zwillinge bis hin zu Anwendungsfällen in der Lokalisierung (MATSUKI et al. 2023). 3D Gaussian Splatting (KERBL et al. 2023) ist eine neuartige Methode zur Blickfeldsynthese (Abb. 1). Die Szenerie wird im Gegensatz zur klassischen Computergrafik nicht durch ein vermaschte Oberfläche, sondern durch die Überlagerung von 3D Gaussverteilungen mit Transparenzanteil repräsentiert. Die Ausrichtung, blickwinkelabhängige Färbung und Transparenz wird dabei iterativ aus Trainingsbildern erlernt. Die Positionen der Bilder, sowie eine Punktwolke, die als Initialisierung für die 3D Gaussverteilungen genutzt wird, liefert in der originalen Implementierung die Structure-from-Motion (SfM) Pipeline COLMAP (SCHÖNBERGER et al. 2016a, SCHÖNBERGER et al. 2016b). Im Vergleich zu Neural Radiance Fields (NeRFs) (MILDENHALL et.al. 2020) zeichnet sich Gaussian Splatting durch deutlich verkürzte Trainingszeiten aus und eignet sich besonders für unbeschränkte Szenarien, wie beispielsweise die Außenfassaden gesamter Straßenzüge oder Gebäudekomplexe.
Der Einsatz von Mobile Mapping Systemen (MMS), welche mit Hilfe von LiDARs, Kameras und weiteren Sensoren die Umwelt abtasten, findet immer mehr Anwendung (etwa zum Monitoring von Straßen oder zur Erfassung der Infrastruktur), so dass diese Art von Datensätzen eine zunehmende Verbreitung findet. Die Idee des vorgestellten Ansatzes versucht, die in den MMS-Daten vorhandenen Daten direkt auszunutzen, um Teile der ursprünglichen Prozessschritte für das Gaussian Splatting zu umgehen.
Die mittels LiDAR erhobene Punktwolke liefert ein hoch genaues 3D Abbild der Messumgebung. Da die Transformationen zwischen den Sensoren und dem Navigationsrahmen in einem MMS sowie die befahrene Trajektorie genau bekannt sind, lassen sich auch die Kameraposen aus diesen Daten ohne den Umweg über SfM extrahieren. Durch diese Vorgehensweise werden die Ausgangsdaten für den Gaussian Splatting Prozess direkt aus den Messdaten des MMS abgeleitet.

[1] Leibniz Universität Hannover, Institut für Kartographie und Geoinformatik, Appelstraße 9a, D-30167 Hannover, E-Mail: [tim.schimansky, monika.sester]@ikg.uni-hannover.de

 DOI: 10.24407/KXP:1885292287

Abb. 1: Beispielszenario: Trainingsbild (1 von 26) (links) und synthetisiertes Bild (rechts)

Im Beitrag werden die Grundidee und Umsetzung des Ansatzes vorgestellt. Eine aktuelle Veröffentlichung beschreibt einen ähnlichen Ansatz, der auf die Visualisierung von Umgebungsinformation aus Umgebungskameras wie beispielsweise NuScenes und KITTI-360 abzielt (ZHOU et al. 2023).
Neben der Synthetisierung von Bildern bietet die Idee des vorliegenden Beitrags eine Grundlage, um perspektivisch Gaussian Splatting für die Positionierung von unbekannten Bilddaten verwenden zu können. Dabei wird davon ausgegangen, dass zunächst eine Repräsentation in Form der Gaussian Splats von einer Umgebung erzeugt wird. Aufnahmen, die später von der gleichen Umgebung gemacht werden, können in dieser Repräsentation zugeordnet werden.

2 Grundprinzip Gaussian Splatting

In der Ursprungsimplementierung von (KERBL et al. 2023) werden Punktwolken aus SfM Pipelines wie COLMAP verwendet (SCHÖNBERGER et al. 2016a; SCHÖNBERGER et al. 2016b). Initial wird für jeden Punkt eine isotrope 3D-Gaussverteilung angelegt. Die Achslänge entspricht dabei dem Mittel des Punktabstandes zu den drei nächsten Nachbarn. Gaussian Splatting verwendet ein Optimierungsverfahren, bei welchem Geometrie erstellt, bewegt und auch zerstört werden kann. Das bedeutet, dass das zugrundeliegende Verfahren auch auf Punktwolken mit zufallsverteilten Punkten angewendet werden kann. Das ist notwendig, um Mehrdeutigkeiten aus der 3D zu 2D Projektion zu eliminieren. Alle 100 Iterationen findet eine Verdichtung der Wolke aus 3D-Gaussverteilungen statt und es werden 3D-Gaussverteilungen entfernt, die einen Transparenzschwellwert unterschreiten.

In der Optimierung müssen Bereiche in denen die initiale Punktwolke keine Ausgangspositionen liefert ("under-reconstruction") oder mit wenigen, sehr großen 3D-Gaussverteilungen bestückt ist ("over-reconstruction") mit neuen 3D Gaussverteilungen geschlossen werden (KERBL et al. 2023).

2.1 Einfärbung über Kugelflächenfunktionen

Im Optimierungsprozess des Gaussian Splatting wird nicht nur die Geometrie, sondern auch die Farbgebung der 3D-Gaussverteilungen iterativ verbessert. Um Probleme zu reduzieren, wenn große Teile der Blickrichtungen auf eine 3D-Gaussverteilung fehlen wird zu Beginn des Optimierungsprozesses der Wert für die Komponente Nullter Ordnung der Kugelflächenfunktionen ermittelt um die Basisfarbe zu ermitteln. Erst später in der Optimierung werden die Parameter höherer Ordnung ermittelt. Ab der Iteration 1000 werden im Abstand von jeweils 1000 Iterationen die 4 Grade der Kugelflächenfunktionen in die Optimierung eingeführt (KERBL et al. 2023).

2.2 Differenzierbare Rasterisierung

Neben schneller Trainingszeit ist auch ein schnelles Rendering wichtig für die Nutzbarkeit. Im Gegensatz zu NeRFs (MILDENHALL et.al. 2020) wird beim Gaussian Splatting anstatt auf Ray Tracing auf klassische Rasterisierung gesetzt (KERBL et al. 2023). Das hat vor allem Geschwindigkeitsvorteile. Basierend auf der Pipeline zum Rendern der neuen Blickwinkel von (LASSNER 2020) sind Optimierungen, wie kachelbasiertes Rendering und die Vorsortierung von primitiven Objekten implementiert worden. Letzteres ist wichtig, um sinnvolle Transparenzüberblendung zu gewährleisten und ist in vorherigen Implementationen Pixelweise sortiert worden.
Essentiell für den Trainingsprozess ist die Differenzierbarkeit des rasterisierten Bildes. Die genutzte Pipeline ist unabhängig von der Anzahl überblendeter 3D-Gaussverteilungen differenzierbar mit einem fixen Zusatzaufwand je Pixel.

3 Datengrundlage & Methodologie

3.1 Datengrundlage

Die benötigten Eingangsdaten für das Gaussian Splatting sind Bilder, welche bereits von Verzeichnungen befreit sind. Zudem sind die intrinsischen und extrinsischen Parameter erforderlich. Darüber hinaus wird als Prior für die Position der 3D-Gaussverteilungen eine Punktwolke der beobachteten Umgebung benötigt. Diese Datengrundlage wird in der Basisimplementierung durch die Software COLMAP (SCHÖNBERGER et al. 2016a, SCHÖNBERGER et al. 2016b) geliefert. COLMAP implementiert dabei die SfM Pipeline in ein kompaktes Tool mit fester Ablagestruktur der Daten.
Im Rahmen dieser Arbeit liegen die Messdaten eines Riegel VMX-250 Mobile Mapping Systems (MMS) vor. Das System verfügt sowohl über ein Rad-Odometer, eine Inertial Measurement Unit (IMU), als auch über eine GNSS-Antenne zur Positionierung. Neben der Trajektorie sind auch die Transformationen bekannt, die die Beziehung der Kameras zum Referenzpunkt des MMS herstellen. Dies ermöglicht die Rekonstruktion der Kameraposen, um eine Georeferenzierung je Bild zu erhalten. Im System sind zwei seitlich blickende Nikon D850 Spiegelreflexkameras (siehe: Abb. 2), sowie zwei Industriekameras mit Blick nach hinten (leicht rechts/leicht links) integriert. Für beide Kameras liegen auch die Verzeichnungsparameter vor,

sodass die Bilder vor der Weiterverarbeitung entzerrt werden können. Die Kameras des Mobile Mapping Systems werden basierend auf der zurückgelegten Distanz ausgelöst. Die Auslösung findet dabei circa alle 2m bis 5m statt.

Abb. 2: Beispiel: Aufeinanderfolgende rohe Kamerabilder MMS (Nikon D850; nach rechts blickend)

Zentrale Messeinheit des MMS sind die zwei Zeilen-LiDARs, die in drei Dimensionen schräg jeweils einen Querschnitt des Straßenbilds aufzeichnen können. Dabei werden bis zu 600.000 Messungen mit jeweils einer Genauigkeit von 5mm ausgelöst. Die daraus hervorgehende Punktwolke ist ebenfalls georeferenziert. Das System erfasst somit Punktwolken (siehe Abb. 3) und Kamerabilder (siehe Abb. 2) im Straßenumfeld, wobei die nach rechts und links zeigenden Kameras jeweils überlappende Bildszenen erfassen können.

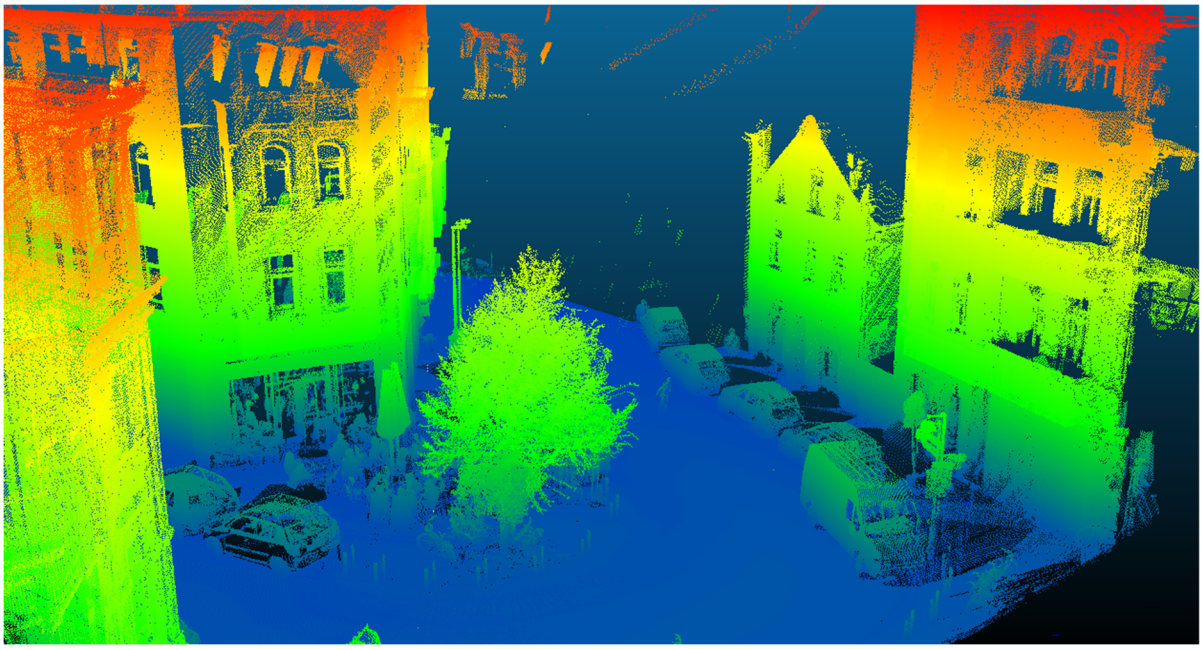

Abb. 3: LiDAR Punktwolke des MMS, eingefärbt nach Z-Komponente

3.2 Klassische Generierung von Gaussian Splats aus Mobile Mapping Bildern

Zunächst werden Gaussian Splats aus den Mobile Mapping Bilddaten erzeugt. Hierzu wird ein Bereich in der Hannoveraner Nordstadt nahe der Lutherkirche ausgewählt (Abb. 4 und Abb. 5). Die Szene enthält Gebäude, Menschen und auch Vegetation. Die Bilder durchlaufen in COLMAP die SfM Pipeline. Da es keine Überlappung der Bilder für die Kameras der beiden Fahrzeugseiten gibt, kann COLMAP ohne weitere Informationen keine Rekonstruktion für die Fassaden auf beiden Seiten der Straße produzieren (Abb. 5). Für die Vergleiche wird sich daher auf die unten gezeigte Fassade fokussiert. In Abbildung 4 ist zu sehen, dass die bekannte Perspektive optisch überzeugend vom Modell gelernt worden ist. Auch mit kleinen Versätzen zu bekannten Positionen (Abb. 4 rechts) sind die Objekte weiterhin sehr gut erkennbar. Jedoch zeigt sich in den feinen Strukturen des Baumes die Bildung von Artefakten.

Abb. 4: vlnr: Referenzbild Nr. 13, synthetisierte Sicht aus selber Perspektive, synthetisierte Sicht mit Versatz (jeweils SfM als Prior)

Abb. 5: Überblick über die gesamte Gaussian Splatting Szene mit SfM Prior. Rekonstruiert wurden nur die Fassaden in Nordrichtung (rechts im Bild), sowie die Stirnseite des Hauses im Westen (links im Bild)

3.3 Ansatz zum Adaptierten Gaussian Splatting

Um 3D Gaussian Splatting für Anwendungen tauglich zu machen, bei denen eine genaue Skalierung und Positionierung relevant ist, schlagen wir vor, diesen Ansatz mit LiDAR-Daten aus einem Mobile Mapping System als Prior zu verknüpfen, um den SfM-Schritt zu umgehen. Weiterhin ist die externe Kalibrierung zwischen Kamera(s) und LiDAR eine zu bewältigende Aufgabe, da diese für die virtuellen Kamerapositionen im Training des Gaussian Splatting benötigt wird.
Die korrekte Skalierung und Positionierung ist relevant, wenn beispielsweise eine Positionsbestimmung auf Basis der Gaussian Splatting Szenerie erfolgen soll. Weitere Vorteile des vorgeschlagenen Verfahrens werden im Verhalten bei Oberflächen mit wenig oder keiner Textur erwartet, da die Punktdichte beim Einsatz von LiDAR nicht von der Featuredichte abhängt, sondern eine direkte Folge der Sensorwahl und des Messaufbaues ist.
Die Messdaten liegen im ECEF vor und werden zunächst in UTM projiziert. Um numerische Probleme durch die Zahlenwerte der UTM Koordinaten zu vermeiden werden diese um ihren Mittelpunkt reduziert. Die Kameraposen sind durch eine Rotationsmatrix und eine Koordinate für das Projektionszentrum der Kamera gegeben. Die Kameradaten werden vor der weiteren Verarbeitung mittels der bekannten Verzeichnungsparameter entzerrt, um Bilder frei von Verzeichnungen für die Weiterverarbeitung nutzen zu können. Um die bestehenden Arbeitsabläufe weitestgehend wiederverwenden zu können, werden die Posen in die Repräsentationsform von COLMAP transformiert. Dazu wird die Kameraposition mittels der Transformation des Weltkoordinatensystems in das Kamerakoordinatensystem definiert.
Die beiden Riegl LiDAR des MMS können jeweils bis zu 300.000 Punkte pro Sekunde messen. Während der SfM Punktwolken Prior für die Szene in Abbildung 4 rund 3000 Punkte enthält, erzeugt das MMS in der gleichen Umgebung rund 11 Millionen Punkte. Bevor der Trainingsprozess starten kann, muss die LiDAR Punktwolke stark ausgedünnt werden, da der Grafikspeicher der genutzten GPU (Nvidia GTX 3060; 12GB VRAM) limitiert ist.

4 Erste Ergebnisse des adaptierten Gaussian Splatting

Um die Ergebnisse zu evaluieren, wird dieselbe Szenerie in drei verschiedenen Pipelines prozessiert. Zunächst wird die originale Implementierung vom (KERBL et al. 2023) genutzt, um eine Referenz für das Ergebnis zu gewinnen (siehe Abschnitt 3.2). Des Weiteren wird der in Abschnitt 3.3 dargestellte Ansatz evaluiert. Zusätzlich wird der Einfluss der Punktwolke für die Initialisierung auf das Ergebnis untersucht: dazu wird eine beliebige Punktwolke (ohne Geometriebezug zur Szene) vorgegeben. Somit können drei Experimente für die Initialisierung verglichen werden: Originalansatz mit SfM, Punktwolke, beliebige 3D-Geometrie.
Im Folgenden werden die Ergebnisse qualitativ dargestellt und diskutiert. Die auf den MMS Daten basierte synthetisierte Sicht in Abb. 6 zeigt die Fassaden, wie sie auch in Abb. 4 zu sehen sind. Die seitlichen Bildränder sind stark von Artefakten geprägt, da die Trainingsdaten im Hochformat vorliegen. Die Bildbereiche, die von den Trainingsdaten abgedeckt sind, verfügen über eine geringere Detaildichte, als es bei dem Modell mit SfM Prior der Fall ist.
Die Gründe für die Qualitätseinbußen können vielschichtig sein. Ein möglicher Faktor ist die Art der Ausdünnung der Punktwolke. Hinzu kommt, dass auch wenn die Kameraposen grundsätzlich bekannt sind, Unsicherheiten in der Positionierung sowie die Zeitsynchronisierung der Kamera eine Rolle für die Genauigkeit der Kameraposen spielen.

Abb. 6: Synthetisierte Sicht aus der Perspektive von Bild 13 mit Prior aus LiDAR Punktwolken

Während die visuelle Qualität des SfM basierten Ansatzes bisher nicht von dem LiDAR basierten Ansatz erreicht werden kann (siehe Abb. 6 im Vergleich mit Abb. 4 rechts), zeigt sich, dass LiDAR Daten als Prior eine Verbesserung der visuellen Qualität gegenüber einer Initialisierung mit rein zufällig verteilten Punkten erzielen (Abb. 7).

Abb. 7: Vergleich der Sichtqualität aus neuer Perspektive mit Prior LiDAR (links) und mit Prior zufallsverteilte Punktwolke (rechts)

Obwohl beide Ergebnisse sehr starke Artefakte aufweisen, sind im synthetisierten Modell mit dem Prior vom LiDAR mehr Details zu erkennen (Fensterstreben, Radkappen, Werbeposter). Die Artefaktbildung kann eine Folge der Unsicherheiten in der Bestimmung der Kameraposen sein.

Kameraposen, welche auf gegenüberliegende Straßenseiten zeigen, können mit SfM nicht in Relation zueinander gesetzt werden, da es keine Überlappungen in den Bilddaten gibt. Abb. 8 zeigt die fehlende Rekonstruktion der gegenüberliegenden Straßenseite aufgrund mangelnder SfM Verknüpfungsmöglichkeiten. Während die Rekonstruktion nicht frei von Artefakten ist, kann die Straßenseite aufgrund der absolut bekannten Posen für die Kamera rekonstruiert werden. Die bekannten Kameraposen des MMS machen somit von Überlappungen in sich geschlossener Areale unabhängiger. Überlappung für Bilder des gleichen Objektes/Areals sind

dennoch wichtig, um sogenannte Floater zu vermeiden. Diese bilden sich bei geringer Bildabdeckung in der Nähe der Kamera und versperren das Blickfeld. An dieser Stelle ist anzumerken, dass die Aufnahmefrequenz der Kameras für ein verbessertes Ergebnis höher sein sollte.

Abb. 8: Vergleich der Sichtqualität auf gegenüberliegende Fassaden. vlnr: Trainingsbild Nr. 21, synthetisierte Sicht mittels SfM Prior (keine Daten für diese Sicht), synthetisierte Sicht mittels MMS Prior

5 Fazit & Ausblick

Im Beitrag wurde zunächst Gaussian Splatting als Methode vorgestellt, um synthetische neue Blickwinkel auf eine 3D Repräsentation, gelernt aus Bilddaten, zu gewinnen. Als Datenquelle wurden Kamera-Daten eines Mobile Mapping Systems verwendet. Anschließend wurde ein Ansatz vorgestellt, um die im Originalansatz verwendete 3D-Informationen für den Prior direkt aus den 3D-LiDAR-Daten des Mobile Mapping Systems zu gewinnen. Zusammenfassend versprechen wir uns von dem Ansatz eine Grundlage für zukünftige Arbeiten, mit welcher Blicke auf 3D-Szenerien für nachfolgende Verarbeitungsschritte oder zur Visualisierung synthetisiert werden können, ohne dass sie den räumlichen Bezug verlieren, und auch für die Positionierung genutzt werden könnten.

Während die bisherigen Ergebnisse in ihrer optischen Qualität deutlich hinter denen des klassischen Ansatzes liegen, gibt es relevante Aspekte, die der vorgestellte Ansatz verbessert. Ein großer Vorteil ist die Verfügbarkeit von Kameraposen, trotz mangelhafter oder fehlender Bildüberlappung.

In zukünftigen Ansätzen kann es sinnvoll sein, die Vorteile eines LiDAR Punktwolken Priors mit der feineren Kamera Ausrichtung aus SfM zu einer Hybridlösung zu kombinieren. Auch eine Optimierung auf die Verarbeitung räumlicher Kacheln kann sinnvoll sein, um den zeitgleichen Berechnungsaufwand für eine längere Messfahrt im Rahmen zu halten.

6 Literaturverzeichnis

KERBL, B., KOPANAS, G., LEIMKÜHLER, T. & DRETTAKIS, G., 2023: 3D Gaussian Splatting for Real-Time Radiance Field Rendering. ACM Transactions on Graphics, **42**(4), https://doi.org/10.48550/arXiv.2308.04079.

LASSNER, C., 2020: Fast Differentiable Raycasting for Neural Rendering using Sphere-based Representations. CoRR, https://doi.org/10.48550/arXiv.2004.07484.

MATSUKI, H, MURAI, R., KELLY, P. & DAVISON, A., 2023: Gaussian Splatting SLAM. CoRR, https://doi.org/10.48550/arXiv.2312.06741.

MILDENHALL, B., SRINIVASAN, P., TANCIK, M., BARRON J., RAMAMOORTHI, R. & NG, R., 2020: NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. European Conference on Computer Vision (ECCV), https://doi.org/10.48550/arXiv.2003.08934.

SCHÖNBERGER, J. L. & FRAHM, J.-M., 2016a: Structure-from-Motion Revisited. IEEE conference on Computer Vision and Pattern Recognition (CVPR), 4104-411, http://dx.doi.org/10.1109/CVPR.2016.445.

SCHÖNBERGER, J. L., ZHENG, E., POLLEFEYS, M. & FRAHM, J.-M., 2016b: Pixelwise View Selection for Unstructured Multi-View Stereo. European Conference on Computer Vision (ECCV), http://dx.doi.org/10.1007/978-3-319-46487-9_31.

VACHHA, C., 2023: Creating Visual Effects with Neural Radiance Fields. arXiv preprint arXiv:2401.08633, https://doi.org/10.48550/arXiv.2308.04079.

ZHOU, X., LIN Z., SHAN X., WANG, Y., SUN, D. & YANG, M.-H.,2023: DrivingGaussian: Composite Gaussian Splatting for Surrounding Dynamic Autonomous Driving Scenes. arXiv preprint arXiv:2312.07920, https://doi.org/10.48550/arXiv.2312.07920, letzter Zugriff 03.20.24.