Schlussbericht

zum Teilvorhaben NeuroSys - Impulse durch Anwendungen (Projekt D) - C

im Verbundprojekt NeuroSys

Zuwendungsempfänger: Gremse-IT GmbH

Förderkennzeichen: 03ZU1106DC
Laufzeit des Vorhabens: 1.1.2022 – 31.12.2024
Berichtszeitraum: 01.1.2022 — 31.12.2024

Teil I: Zusammenfassung

Das Ziel unseres Teilprojektes war der Aufbau von Kompetenzen in der Softwareentwicklung zur Verwendung von effizienter neuromorpher bzw. GPU-beschleunigter Hardware in zukünftigen biomedizinischen Softwareanwendungen. Anhand verschiedener Anwendungsbeispiele in der medizinischen Bildverarbeitung untersuchten wir Möglichkeiten, um rechenintensive Algorithmen auf modernen energieeffizienten Computerarchitekturen ausführen zu können, die auf lokalen Computern, Laptops, Tablets oder gar Handys in der Breite zum Einsatz kommen können. Dadurch könnten z.B. Algorithmen für maschinelles Lernen zukünftig in tragbare medizinische Geräte eingebettet werden.
Konkret wurde in Zusammenarbeit mit Projektpartner MEDIT ein portabler Wundscanner als Prototyp entwickelt: ein KI-gestütztes, kompaktes Assistenz-Gerät inklusive Software zur umfassenden Analyse chronischer Wunden. Der Wundscanner nutzt einen Akku und einen leistungsstarken, GPU-beschleunigten Einplatinencomputer, sowie vier Kameras, um Farbbilder, 3D-Tiefenkarten und Temperaturverteilungen von Wunden aufnehmen, speichern, verarbeiten und darstellen zu können. Wir unterstützten Projektpartner MEDIT bei der Erstellung und Verwendung des Datenformates, um die Bilddaten im visuellen und nahinfraroten Bereich aufnehmen und strukturiert speichern zu können, so dass Projektpartner STAR diese Daten in den cloudbasierten Prototypen der digitalen Patientenakte zum Verwalten der Bilddaten und anderen Metadaten einspeisen und schließlich darstellen konnte.
Um weitere Erfahrungen mit den Infrarotdaten sammeln zu können, beschafften wir dieselbe Stereokamera mit Infrarotsensoren und generierten Beispieldaten von Händen und Markern in verschiedenen Szenarien. Wir nutzten diese Daten und entwickelten Algorithmen zum Erkennen und Verfolgen von Markern und zur Erstellung von Tiefenkarten. Wir passten unsere Bildverarbeitungssoftware an, um die akquirierten mehrkanaligen 2D Bilddaten interaktiv visualisieren und auswerten zu können. Die Funktionalität wurde erweitert, um medizinische DICOM-Daten von und zu Servern im Netzwerk transportieren zu können. Unser DICOM-Import und -Export wurde erweitert, um 4D und 5D Daten systematisch lesen und schreiben zu können. Wir untersuchten verlustlose Kompressionsverfahren für Bilddaten, die sich mittels Hardwarebeschleunigung implementieren lassen, um die Bilddaten in Echtzeit für kompakteren Datentransfer und Speicherung komprimieren zu können.

Zur automatischen Segmentierung der Bilddaten untersuchten wir KI-Methoden zum inkrementellen Training anhand von partiell gelabelten Daten, um den Entwicklungsaufwand neuer KI-Anwendungen zu reduzieren. Diese Trainingsdaten wurden verwendet, um faltungsbasierte neuronale Netze verschiedener Topologien zu fitten, z.B., um Lungenfibrose in µCT-Bildern von Mäusen und Ratten systematisch und effizient auswerten zu können. Wir beschleunigten mehrere Bildverarbeitungsmethoden unter Verwendung hardwarenaher GPU-Operationen, zur Kompression und Dekompression von Segmentierungsdaten, der iterativen Rekonstruktion von µCT-Daten, zum Echtzeittracking von Markern in Stereobildern, sowie zum Schätzen des Wärmeeintrages beim hochintensiven fokussierten Ultraschall in heterogenem Gewebe.
Wir passten unsere Bildverarbeitungssoftware an, um 2D, 3D, 4D und 5D Bilddaten in Voxeltypen von reduzierter Genauigkeit darstellen und verarbeiten zu können, da diese Datentypen auf Hardwarebeschleunigern üblich sind und damit Berechnungen schneller und stromsparender durchgeführt werden können.
Unsere iterative Kegelstrahlrekonstruktion wurde hinsichtlich der Nutzung von modernen GPU-Operationen angepasst und wir untersuchten Methoden zur automatischen Geometriekalibrierung und der Korrektur von Bildartefakten, als rechenaufwändige und komplexe numerisches Beispielprobleme.
Weiterhin entwickelten wir Softwaremethoden zur interaktiven Segmentierung und Visualisierung von Bilddaten und Zwischenergebnissen, wie 3D Segmentierungen, um Algorithmen zur medizinischen Bildverarbeitung effizienter entwickeln und einsetzen zu können. Zur automatischen Segmentierung von Gewebetypen untersuchten wir faltungsbasierte neuronale Netze verschiedener Topologien und verwendeten GPU-basierte Trainingsmethoden unter Verwendung von interaktiv und inkrementell segmentierten Trainingsdaten. Ein Fokus lag hierbei auf der Untersuchung von µCT-Daten zur Erforschung der Erfassung und Therapie von Tumorerkrankungen und Lungenfibrose in verschiedenen Stadien. Im Kontext der interaktiven und automatischen Therapieplanung untersuchten wir den Einsatz verfügbarer Operationen zur Multiplikation kleiner Matrizen mittels Tensorcores, um den Wärmeeintrag von fokussiertem hochintensivem Ultraschall in heterogenem Gewebe effizient und in Echtzeit simulieren zu können.
Weiterhin untersuchten wir die Nutzung von Cloudangeboten, die leistungsfähige GPU-beschleunigte Computer als individuell nutzbare virtuelle Maschine skalierbar bereitstellen können. Mit diesen Diensten können besonders rechenintensive Anwendungen lokal nutzbar gemacht werden, sofern eine Netzwerkverbindung verfügbar ist.
Insgesamt konnten wir im Verlauf des Projektes erhebliche Kompetenzen zur Nutzung von GPU-beschleunigter Hardware für Softwareanwendungen aufbauen und unsere Infrastruktur zur systematischen Softwareentwicklung sinnvoll ausbauen.

Teil II: Ausführlicher Sachbericht

AP D.1.1.3: Gesamtmatrix, Speicherung und Darstellung (M13-M24)

In diesem Arbeitspaket passten wir unsere Imalytics Preclinical Software an, um die mit dem Prototyp generierten Bilddaten speichern, öffnen, darstellen und verarbeiten zu können. Diese Bilddaten umfassen reguläre RGB-Farbbilder, sowie Tiefenkarten und Wärmebilder, welche in unterschiedlicher Auflösung und Genauigkeit gespeichert werden können. Die Arbeiten erfolgen in enger Abstimmung mit den Projektpartnern und es wurde ein geeignetes Datenformat auf Basis unseres offenen GFF Dateiformates (Yamoah et al. 2019) entwickelt und genutzt. Weiterhin implementierten wir Übertragungsprotokolle, um Bilddaten an und aus DICOM-Servern laden zu können.
Weiterhin entwickelte Gremse-IT Methoden zur visuellen Darstellung der Tiefenkarten und Wärmebilder, sowie zur geometrischen Korrektur derselben. Um die Bilddaten effizient speichern zu können, untersuchte Gremse-IT Methoden zur verlustfreien Echtzeit-geeigneten Bildkompression, die mit dem DICOM-Format kompatibel sind. Unsere Software wurde den Projektpartnern zur Verfügung gestellt, um die mit dem Prototyp erfasste Test- und Studiendaten sichten und auswerten zu können.
Weiterhin sammelt Gremse-IT in diesem AP biomedizinische Bilddaten für das KI-Training. Neben zweidimensionalen Daten wurden von unseren Kunden und auch 3D CT-Daten von Mäusen angefragt, da sich hierzu ein hoher Bedarf an automatischer Auswertung zeigte. Hierzu wurden über 300 µCT-Datensätze verschiedener Geräte (MILabs, Mediso, Molecubes, Inviscan, Inveon) gesammelt, um diese für das Training von KI-Modellen zur automatischen Segmentierung zu nutzen.

AP D.1.2.1: Merkmale und Algorithmen zur klinischen Diagnose (M1-M24)

In diesem AP wurden Algorithmen erforscht, um multimodale Bilddaten mit neuromorphen bzw. GPU-beschleunigten Operationen zu analysieren und so neue diagnostische medizinische Anwendungen zu beschleunigen oder zu ermöglichen. Moderne bildgebende Geräte wie CT, MRT, Ultraschall oder multispektrale optische Kameras liefern große Datenmengen, die von Menschen oft nicht mehr vollständig erfasst oder interpretiert werden können, aber mit Computerprogrammen verarbeitet werden müssen, um eine für den Menschen brauchbare Essenz zu extrahieren. Oft sind in dieser Verarbeitung die verfügbare Rechenleistung bzw. die sich daraus ergebende Berechnungsdauer ein limitierender Faktor. Wir suchten in diesem AP nach konkreten biomedizinischen diagnostischen Anwendungen, bei denen umfangreiche Berechnungen erforderlich sind und mit neuromorphen Operationen essenziell beschleunigt werden können.
Ein wichtiger Anwendungsfall in der biomedizinischen Forschung aber auch der klinischen Praxis ist die Rekonstruktion volumetrischer CT-Bilddaten aus akquirierten Kegelstrahlprojektionen. Hierbei setzten wir ein iteratives modell-basiertes Verfahren ein, welches mittels GPU-beschleunigter Gradientenberechnungen das von den Projektionen erfasste Volumen rekonstruiert. Solche Rekonstruktionsverfahren liefern im Vergleich zu direkten Verfahren oft ein Ergebnis höherer Qualität oder mit quantitativen Eigenschaften, woraus sich messbare und diagnostisch verwendbare Parameter ergeben können, die neue diagnostische Anwendungen ermöglichen können. Hierfür wird ein möglichst genaues Model des Gerätes verwendet, welches mittels aufwändiger Kalibrierungsexperimente generiert werden kann. Im praktischen Forschungsbetrieb wird

diese Kalibrierung jedoch leider oft vernachlässigt, weil für die Wartung kein Budget vorgesehen ist, oder weil für solche Arbeiten kein Personal eingeplant oder zuständig ist. Oftmals fallen die durch fehlerhafte Kalibrierung entstehenden Bildartefakte auch erst während der Datenauswertung auf und die Objekte können nicht nochmal gescannt werden, weil sie bereits weiterverarbeitet wurden. Glücklicherweise kann man diese Kalibrierung prinzipiell auch retrospektiv anhand der Bilddaten durchführen, und so die fehlerhaften Scans „retten“. Hierzu kann man mehrere Rekonstruktionen unter Variierung der Kalibrierungsparameter (z.B. Abstand von Quelle und Detektor) erstellen, bis man eine gute, z.B. scharfe, Rekonstruktion erhält. Dieser Vorgang kann jedoch mehrere Stunden oder Tage dauern, da man etliche Rekonstruktionen unter Eingabe neuer Parameter durchführen muss, wovon jede einzelne bereits mehrere Minuten oder Stunden dauern kann. Daher automatisierten wir die Suche nach diesen Kalibrierungsparametern. Wir implementierten eine Optimierungsmethode (Nelder-Mead), welche die Kalibrierungsparameter unter Verwendung eines Gütekriteriums (das Residuum der Rekonstruktion) sucht. Diese Art von Optimierung eignet sich besonders für solche Minimierungsprobleme mit einer geringen Anzahl von gesuchten Parametern (üblicherweise 3 bis 7). Um die Gesamtdauer zu verringern, implementierten und optimierten wir mehrere verwendete GPU-Kernel, z.B. für die Vorwärtsprojektion, Rückprojektion und Fouriertransformation. Anhand mehrerer Beispiele konnten eine deutliche Verbesserung der Bildqualität zeigen (Abbildung 1).

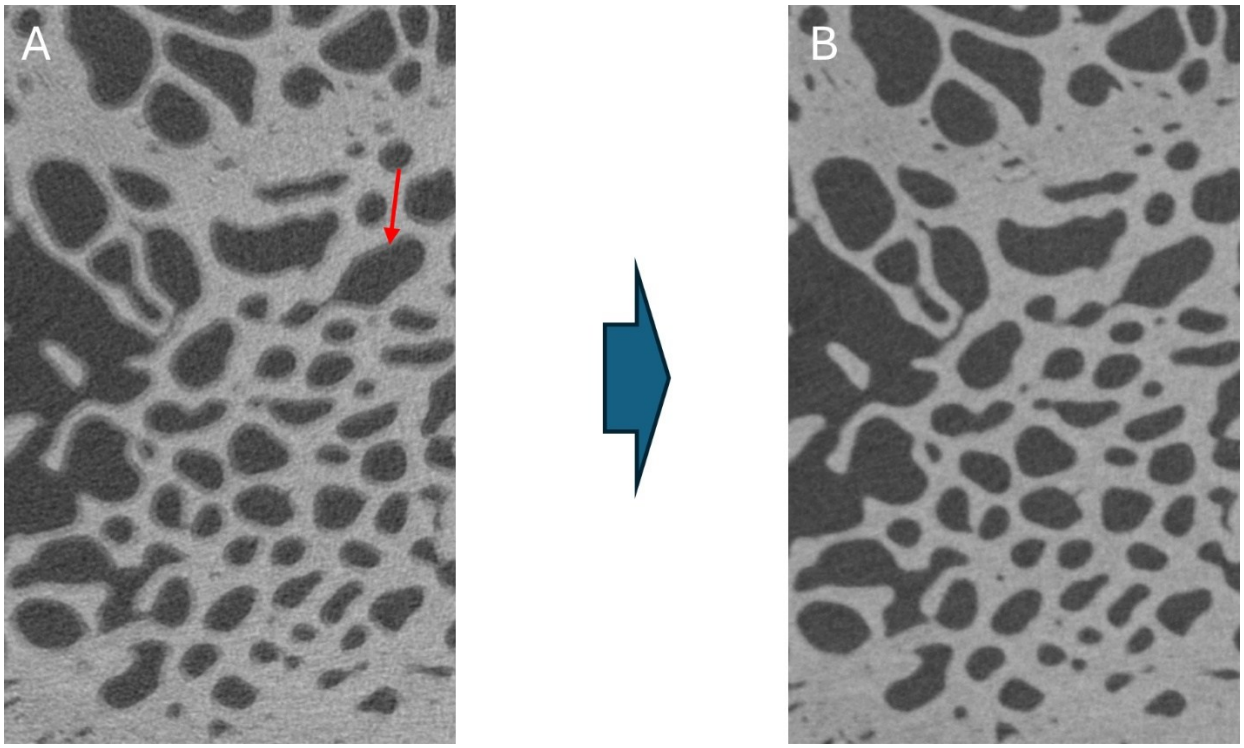


Abbildung 1: A) Fehlerhafte Kalibrierung führt zu verschwommenen Artefakten an trabekulären Strukturen. B) Nach iterativer retrospektiver Korrektur erscheinen die Knochenstrukturen schärfer.

Die Auswertung klinischer volumetrischer Bilddaten erfordert oft den Zwischenschritt der Segmentierung, d.h. der Zuordnung der Bildpunkte an benannte Regionen, wie Organe oder Läsionen. KI-basierte Methoden konnten hierbei in der letzten Zeit deutliche Fortschritte erzielen, weshalb wir diese Ansätze untersucht haben (siehe AP D.1.3.2). Ein wesentlicher Flaschenhals ist hierbei jedoch die Verfügbarkeit von manuell segmentierten Bilddaten, um die neuronalen Netze überhaupt trainieren zu können. Um diese Trainingsdaten effizient für neue Anwendungsfälle generieren zu können, haben wir Methoden zur interaktiven Segmentierung von 3D, 4D und 5D Bilddaten implementiert. Der Benutzer kann die Segmentierungen für jede µCT-Schicht einzeln erstellen, was jedoch mit erheblichem Aufwand verbunden ist. Daher haben wir die Methode so angepasst, so dass nur wenige Schichten, z.B. drei pro Organ, interaktiv segmentiert werden, was auch anschaulich in 3D visualisiert wird (Abbildung 2). Diese „dünnen“ Segmentierungen haben wir dann für das KI-Training der neuronalen Netze verwendet, wofür der Kostenterm angepasst wurde, um sich nur auf die bekannten Regionen zu beziehen. Das KI-Modell wird hierbei jedoch weiterhin auf das ganze Volumen angewendet. Nach dem Training wird das Segmentierungsergebnis dem Benutzer angezeigt, der daraufhin weitere Annotationen tätigen kann. So wird iterativ mit minimalem Benutzeraufwand eine KI trainiert, bis sie den entsprechenden Ansprüchen der Anwendung genügt. Um die Wartezeiten des Benutzers während der Annotationen zu minimieren, wurden die Algorithmen zur interaktiven Segmentierung und Visualisierung mittels GPU-Programmierung beschleunigt.

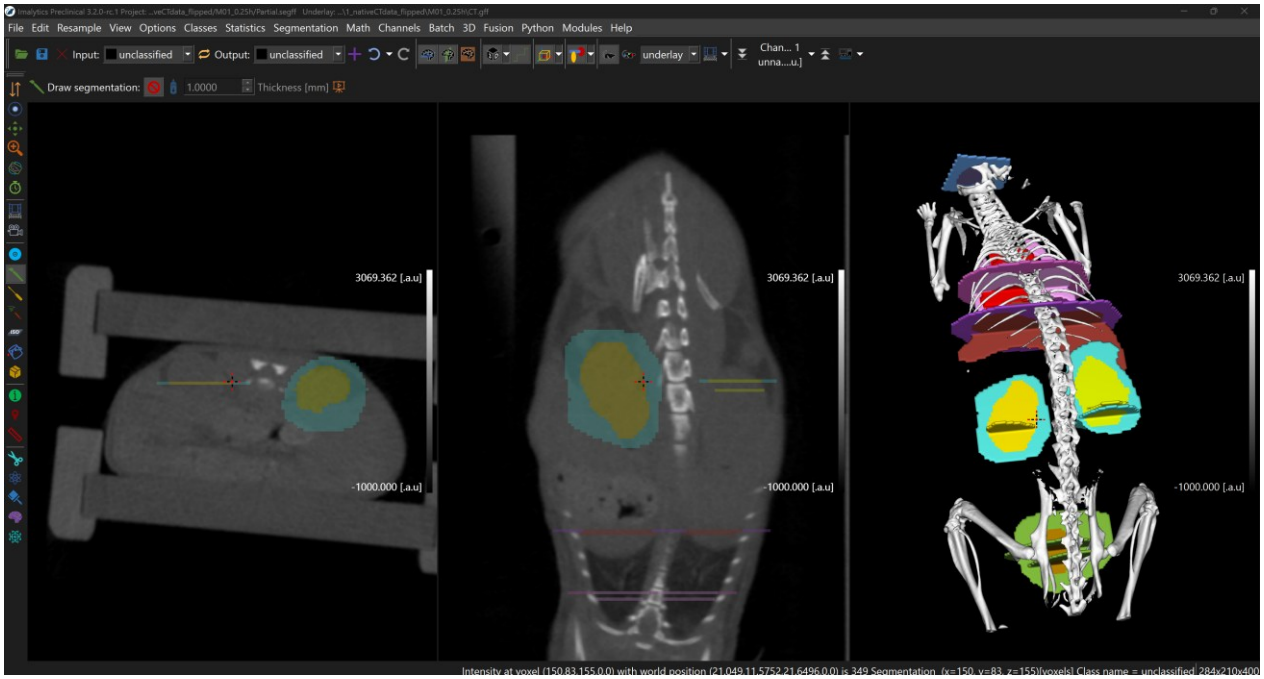


Abbildung 2: Interaktive Segmentierung von wenigen repräsentativen Schichten pro Organ.

AP D.1.2.4: Adaption der MBV auf neuromorphe Hardware (M13-M36)

Die interaktiven Softwaremethoden der anderen Arbeitspakete benötigen eine kurze Reaktionszeit zur besseren Benutzbarkeit. In diesem Arbeitspaket haben wir viele zugrundeliegenden Algorithmen der Bildverarbeitung isoliert und mit Operationen

implementiert, die auf aktuellen GPUs vorhanden sind, um eine entsprechende Beschleunigung zu erreichen. Wir haben dazu vorhandene Softwareentwicklungstools (NVIDIA Cuda Toolkit) verwendet. Hierbei haben wir neue Technologien wie Tensor-Cores, dynamische Texturen, dynamische Kernels und Datentypen mit reduzierter Genauigkeit untersucht und verwendet. Die beschleunigten Algorithmen beinhalten Methoden für die Fourier-Transformationen, das 3D-Rendering, das Lösen dünnbesetzter Gleichungen, Filter, und weitere, die wir durch Geschwindigkeitsmessungen bestimmt haben.
Da GPUs spezielle Speicherstrukturen mit begrenzter Größe verwenden, haben wir Strategien zur Reduzierung des Speicherbedarfes und der Bündelung von Speicherzugriffen entwickelt. Die 3D Segmentierung wird beispielsweise parallel mittels Run-Length-Encoding entlang der x-Achse komprimiert. Die sich ergebenden Vektoren werden mittels parallel implementierter Parallel-Prefix-Sum kompaktiert. Der komprimierte Datensatz wird für typische Anwendungsfälle oft auf wenige Prozent der Gesamtgröße reduziert und kann somit auch deutlich schneller auf die GPU und zurück kopiert werden. Die Dekompression ist innerhalb weniger Millisekunden möglich und kann nach Bedarf auf der GPU durchgeführt werden, auch für Teilbereiche des Volumens. Diese Technologie ermöglichte uns, ein allgemeines Undo/Redo System zu implementieren, um praktisch alle Segmentierungsoperationen damit abzubilden und wesentlich mehr Zustände vorzuhalten, so dass der Benutzer deutlich mehr Schritte zurückgehen kann (Undo).

Ein wesentlicher und oftmals erforderlicher Schritt in der biomedizinischen Bildverarbeitung ist die automatische Erkennung von Markierungen (Markern). Diese werden an mehreren bekannten Stellen angebracht und können dann genutzt werden, um das Objekt registrieren (ausrichten) zu können, um weitere Verarbeitungsschritte zu ermöglichen. Da diese Marker recht klein sein müssen, eignen sich vorhandene und robuste Methoden wie QR-Codes leider nicht, welche ohnehin nicht für 3D Daten gedacht sind. Wir haben daher ein Konzept implementiert, was solche Marker in 2D oder 3D mit einer Kaskade von ggf. GPU-beschleunigten Klassifikatoren detektiert. Dieses Konzept haben wir erfolgreich auf drei Anwendungsfälle angewendet (Abbildung 3).

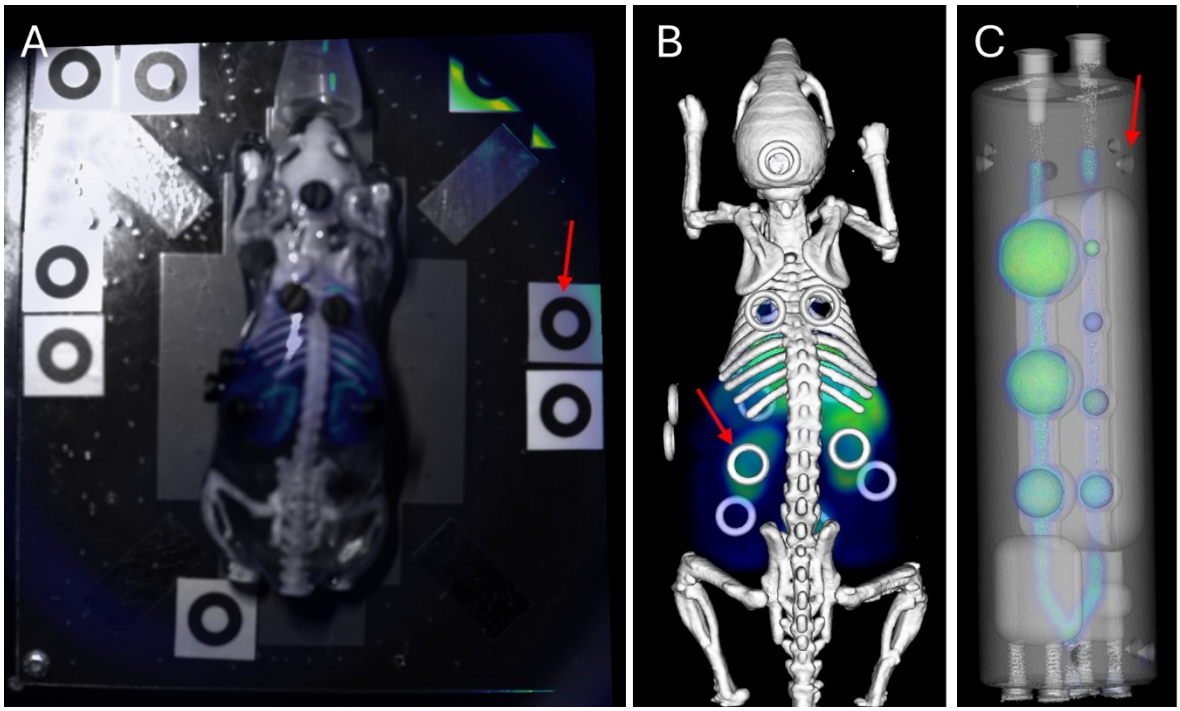


Abbildung 3: Die GPU-beschleunigte Erkennung von Markern wurde für 2D und 3D Beispiele implementiert. A) Ringmarker in einem 2D Bild. B) Ringmarker in 3D µCT Bildern. C) 3D Kegelmarker in µCT Bild. Zur effizienten Erkennung der Marker wurde eine Kaskade von lokalen Klassifikatoren trainiert, wobei die frühen Klassifikatoren die einfachen Kandidaten mit günstigen Operationen verwerfen, so dass die teuren und starken Klassifikatoren nur für wenige verbleibende Kandidaten genutzt werden.

AP D.1.2.6: Algorithmen zur Therapieplanung (M1-M36)

Die bildbasierte Therapieplanung für medizinische Interventionen basiert auf volumetrischen Bilddaten, die für den individuellen Patienten vor dem Eingriff z.B. mit CT, MRT, PET, oder SPECT erhoben werden. Anhand der Bilddaten werden Regionen wie Knochen, gefährdete Organe oder Läsionen segmentiert und der Eingriff wird simuliert und visualisiert. So können z.B. die Risiken und Belastungen eines Eingriffs vorab eingeschätzt oder auch minimiert werden. Dies erfordert oft Simulationen von biophysikalischen Prozessen: Bei der internen Strahlentherapie ist z.B. die regionale Absorption von Alpha-, Beta-, oder Gammastrahlen im Körper von Interesse, um ggf. das dosislimitierende Organ zu finden. Diese neue Art von Therapien hat in den letzten Jahren viel Aufmerksamkeit in der Forschung bekommen, da sie sich in der Behandlung von metastasiertem Prostatakrebs als sehr erfolgreich zeigte (Pluvicto von Novartis). Um eine effiziente interaktive Therapieplanung zu ermöglichen, sind schnelle Algorithmen zur Simulation solcher Prozesse erforderlich, was wir in diesem AP mit neuromorphen Operationen umgesetzt haben. Konkret wird aus dem kalibrierten und quantitativen PET- oder SPECT-Bild eine Dosiskarte berechnet, indem ein Isotop-spezifischer Kernel als Filter auf das Bild angewendet wird. Dieser Kernel ergibt sich aus dem Strahlungsprofil des radioaktiven Isotops und beschreibt die Intensität und den Grad der Ausbreitung der Strahlung. Alpha- und Betastrahler haben eine sehr lokale Ausbreitung, während

Gammastrahler ihre Energie sehr weiträumig verteilen. Viele Isotope haben jedoch eine Mischung, weshalb mehrere Kernels angewendet werden müssen. Diese Kernel werden als Faltung in 3D auf den Datensatz angewendet, um die Dosiskarte zu berechnen. Anschließend wird mit Hilfe der idealerweise automatisch generierten Segmentierung die Dosiskurve und das Integral pro Organ und Tumor berechnet. Mit diesen Informationen kann eine Therapiestudie besser geplant werden, indem die optimale zu applizierende Dosis geschätzt werden kann.
Um die berechneten Ergebnisse dem Benutzer effizient und genau visualisieren zu können, haben wir entsprechende Methoden zur 3D-Visualisierung untersucht und mit Hilfe neuromorpher Operationen beschleunigt.
Während unsere Implementierung zunächst für Wissenschaftler und der Auswertung ihrer Tierstudien gedacht ist, ist mittelfristig auch eine Überführung in die klinische Anwendung denkbar und interessant.

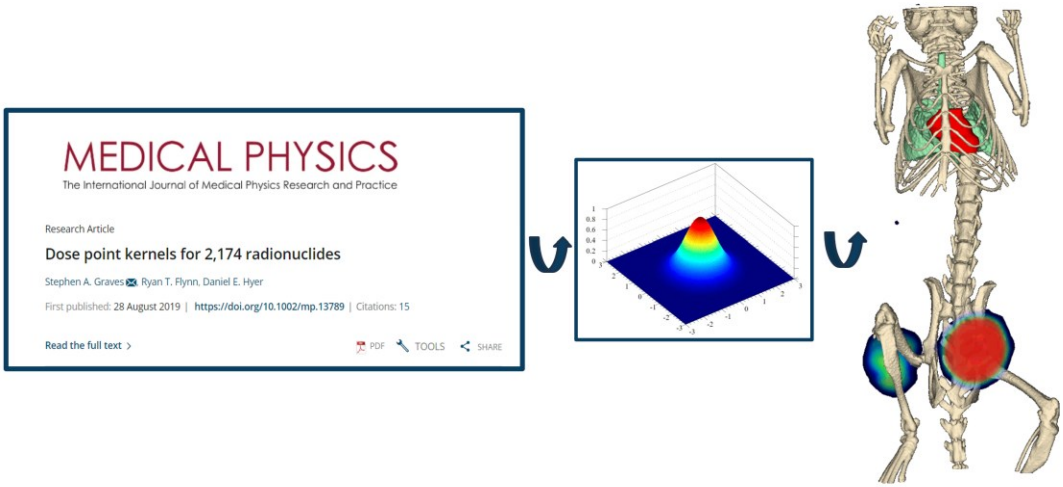
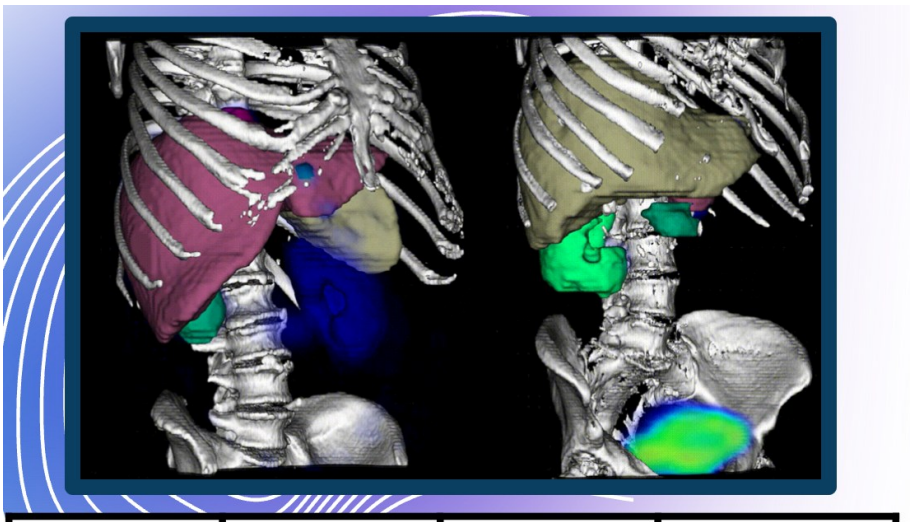


Abbildung 4: Bildbasierte Dosimetrieberechnung anhand von etablierten Dosiskerneln.



Region	PARaDIM	Voxel MC	Imalytics Preclinical
Left Kidney	3.44 Gy	3.65 Gy	3.67 Gy
Right Kidney	4.68 Gy	4.97 Gy	5.26 Gy
Liver	1.78 Gy	1.84 Gy	1.55 Gy
Lesion 1	3.45 Gy	3.18 Gy	3.36 Gy

Abbildung 5: Vergleich unserer Dosisberechnung mit vorhandenen Ansätzen.

Da diese Simulationen während der Entwicklung wiederholt und auf mehreren Datensätzen durchgeführt wurden, haben wir Cloud-basierte Server mit mehreren GPUs genutzt, die wir von etablierten Anbietern (Microsoft Azure) stundenweise mieten konnten.

AP D.1.3.2: Neuronale Netzwerktypen (M7-M30)

In diesem Arbeitspaket haben wir verschiedene neuronale Netzwerktypen auf ihre Eignung für die biomedizinische Bildverarbeitung untersucht. Da wir mehrdimensionale Bilddaten verarbeiten, sind faltungsbasierte Netze mit mehreren verbundenen Schichten interessant, bei denen ein hoher Rechenaufwand für das Training, aber auch das Inferencing entsteht, welches, wenn möglich, unter Verwendung von GPUs erfolgt. Für die neuronalen Netze haben wir verschiedene nichtlineare Funktionen an der Synapsenverbindung untersucht. Weiterhin haben wir die Verwendung verschiedener Darstellungen von Zahlenwerten in den Netzen (float, half, BFloat16) untersucht, welche sich effizient mit modernen neuromorphen Operationen berechnen lassen. Für die interaktive Segmentierung sind iterativ trainierte Netze relevant, die wiederholt auf Basis der Benutzereingaben adaptiert werden können und die auch für unvollständig segmentierte Trainingsdaten geeignet sind. Da wir für viele Anwendungen eine schnelle

Reaktionszeit benötigen, in Minuten oder gar Sekunden, untersuchten wir auch kompakte Netze, die eine schnelleres Inferencing ermöglichen.
Die trainierten neuronalen Netze benötigen spezifische, für die Fragestellungen repräsentative Trainingsdaten. Diese haben wir aus sowohl aus öffentlichen Datenbanken oder Publikationen extrahiert (Rosenhain et al. 2018), aber auch durch Anfragen an unsere bisherigen Nutzer gesammelt. Da wir genügend Datensätze sammeln konnten (über 300), war die Generierung künstlicher Daten mittels GANs nicht mehr erforderlich. Die Datensätze haben unsere Anwendungsexperten aufwändig manuell annotiert, indem sie mit unseren interaktiven Werkzeugen verschiedene Organe und Läsionen segmentierten. Unter Verwendung dieser Datenbanken konnten wir KI-Modelle für die Segmentierung von Organen, Tumoren und Knochen in präklinischen µCT-Daten trainieren, die sich in der Anwendung als brauchbar zeigten (Abbildung 6).

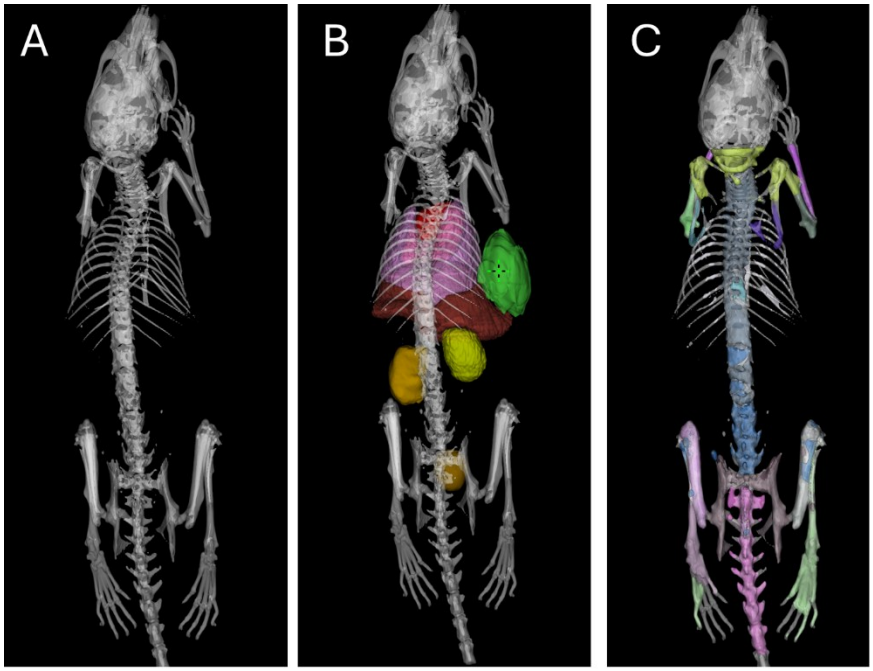


Abbildung 6: CT Bilder und KI-basierte Segmentierungen: A) 3D Rendering eines µCT Scans einer Maus, bereitgestellt von der Wisconsin University. B) Ergebnis der KI-basierten Organ- und Tumorsegmentierung. C) Ergebnis der KI-basierten Knochensegmentierung.

Teil III: Ergänzende Darstellung

Die wichtigsten Positionen des zahlenmäßigen Nachweises

Die mit Abstand meisten Kosten fielen in diesem Projekt für Personal, wie es auch geplant war. Ausschließlich fest angestellte Mitarbeiter wurden eingesetzt, um die umfangreichen Programmierarbeiten und das Testen der Software durchzuführen.

Notwendigkeit und Angemessenheit der geleisteten Projektarbeiten

Die durchgeführten Arbeiten zur Softwareprogrammierung und Bilddatenverarbeitung dienten dem Zwecke des Projektes entlang der geplanten Projektdurchführung. Neben der Erstellung des Prototyps im Verbundprojekt, konnten durch Umsetzung dieser Softwareentwicklungsarbeiten wesentliche Kompetenzen bei Gremse-IT zur Verwendung moderner KI-Methoden für die Bilddatensegmentierung aufgebaut werden. Weiterhin konnten wir systematisch Bilddaten für das KI-Training sammeln und dabei lernen, wie diese sinnvoll und effizient annotiert werden können. Auch sind die im Projektverlauf entwickelten GPU-beschleunigten Algorithmen sind für Folgeprojekte und Anwendungen sehr gut brauchbar. Die Projektförderung hat es uns ermöglicht, unser Softwareteam in Größe und Kompetenz deutlich voranzubringen.

Verwertbarkeit der Ergebnisse

Die im Projekt gesammelten und annotierten Bilddaten, z.B. zur Organ- und Tumorsegmentierung eignen sich, um sinnvoll nutzbare KI-Modelle zu trainieren. Die hierfür entwickelten Methoden zur interaktiven Annotation erlauben uns, weitere solche Datenbanken zu generieren. Nach Abschluss der Projektlaufzeit haben wir bereits begonnen, erste Prototypen verschiedenen potentiellen Anwendern in der akademischen und industriellen Forschung zu zeigen und konnten deren Interesse wecken. Zur kommerziellen Verwertbarkeit wird noch die Erstellung von Dokumentation und Tutorials erfolgen. Wir hoffen, hiermit eine erhebliche Anzahl (100 oder mehr) von Neukunden in den drei Jahren nach Projektende gewinnen zu können.

Während der Durchführung des Vorhabens dem Zuwendungsempfänger bekannt gewordenen Fortschritt auf dem Gebiet des Vorhabens bei anderen Stellen

Im Verlauf der Projektdurchführung ergab sich ein gewisser Durchbruch von LLMs (Large Language Models), insbesondere ChatGPT. Diese KI-Sprachmodelle sind nicht direkt für die Bildverarbeitung bzw. -Auswertung vorgesehen, eignen sich aber auf höherer Ebene für die semantische Auswertung von Zwischenergebnissen der Bildverarbeitung sowie für die Generierung von Texten, die das Ergebnis sinnvoll beschreiben können. Diese LLLMs und ihr Potential waren während der Projektplanung noch nicht verfügbar bzw. bekannt LLMs und haben im Verlauf des Projektes deutlich an wirtschaftlicher Bedeutung gewonnen.

Erfolgte oder geplante Veröffentlichungen des Ergebnisses

Wir planen in den nächsten sechs Monaten eine Veröffentlichung zu unserer KI-unterstützten Anwendung zur Organ- und Tumorsegmentierung.

02.09.2025, Aachen

Datum, Ort Unterschrift: PD Dr. rer. medic. Felix Gremse