

# Segmentierung von Brustvolumina in Magnetresonanztomographiedaten unter der Verwendung von Deep Learning

Thomas G. Jentschke<sup>1</sup>, Katrin Hegenscheid<sup>2,3</sup>, Henry Völzke<sup>3</sup>,  
Florentin Wörgötter<sup>1</sup>, Tatyana Ivanovska<sup>1</sup>

<sup>1</sup>Institut für Biophysik, Georg-August Universität Göttingen

<sup>2</sup>Sana Klinikum Lichtenberg Berlin

<sup>3</sup>Universitätsmedizin Greifswald

tiva@phys.uni-goettingen.de

**Kurzfassung.** Die Segmentierung von Hintergrund und Brustgewebe ist ein wichtiger Teil der Auswertung von Magnetresonanztomographie-Daten der Brust. Normalerweise wird diese von Ärzten manuell durchgeführt. In dieser Arbeit wurde die Segmentierung hingegen mit einer U-net Architektur realisiert. Dabei wurden zwei Netzwerke trainiert und anschließend auf ein unbekanntes Testset, bestehend aus 8 Probandinnen, angewendet. Die so berechneten Segmentierungen wurden dann mit von Ärzten manuell vorgenommenen verglichen.

Das erste U-net nutzt keine weitere Vorverarbeitungsmethode und erreicht einen DSC von  $0.91 \pm 0.09$  (Mittelwert  $\pm$  Standardabweichung). Beim zweiten Netzwerk wurde der N4ITK Bias Correction Algorithmus als Vorverarbeitungsmethode verwendet. Die Masken für N4ITK können sehr grob sein und daher in einer späteren Anwendung von einem Arzt schnell erstellt werden. In dieser Konstellation wurde bei der Segmentierung des Testsets ein DSC von  $0.98 \pm 0.05$  erreicht. Die Segmentierungen benötigen darüber hinaus nach Anfertigung der Masken für den Vorverarbeitungsalgorithmus 14s. Die Methode hat somit das Potential, Anwendung in der medizinischen Diagnostik zu finden.

## 1 Einleitung

Die Bedeutung der bildgebenden Diagnostik im Bereich der Medizin nimmt immer weiter zu. Eines der wichtigsten Verfahren ist dabei neben der Computertomographie (CT) die Magnetresonanztomographie (MRT). Diese ist sowohl zeit- als auch kostenintensiv, da die erzeugten Bilder heutzutage noch überwiegend manuell von Spezialisten ausgewertet werden müssen. Seit geraumer Zeit wird daher versucht dies mittels spezieller Algorithmen durchzuführen.

Bei der Auswertung von MRT-Bildern der Brust ist einer der wichtigsten Schritte die Segmentierung von Brustgewebe. Die so erhaltenen Daten können dann von anderen Algorithmen weiter ausgewertet werden. Darüber hinaus lässt sich direkt aus den Segmentierungen von Brustvolumina und Parenchym die Brustdichte bestimmen. Diese ist ein wichtiger Indikator zur Einschätzung des Risikos

einer Brustkrebserkrankung.

Es gibt viele vergleichbare Arbeiten zu diesem Thema, in denen jeweils unterschiedliche Algorithmen verwendet wurden. Die von Ivanovska et al. [1] verwendete Level-Set-Methode mit simultaner Bias-Korrektur erreichte einen durchschnittlichen Dice Koeffizient von  $0.960 \pm 0.017$  von 37 MRT-Scans. Ein anderes Beispiel ist die Methode von Gubern-Mérida et al. [2], bei der ein DSC von 0.94 erreicht wurde.

Alle diese Algorithmen sind mehr oder weniger stark abhängig von den verschiedenen Charakteristika der verwendeten MRT-Bilder.

In dieser Arbeit sollen Deep Learning Methoden zur Segmentierung der Brust verwendet werden. Diese sind in der Lage, die entsprechenden Charakteristika der jeweiligen MRT-Bilder zu lernen, und somit unabhängiger von den erhobenen Daten. Wie in der Arbeit von Dalmış et al. [3] soll hierfür eine sogenannte U-net Architektur [4] verwendet werden. Diese wurde schon in anderen Arbeiten sehr erfolgreich zur Lösung von biomedizinischen Segmentierungsproblemen verwendet. Wie in der medizinischen Diagnostik notwendig, ist das fertig trainierte U-net darüber hinaus in der Lage, die MRT-Bilder sehr schnell zu verarbeiten.

## 2 Material und Methoden

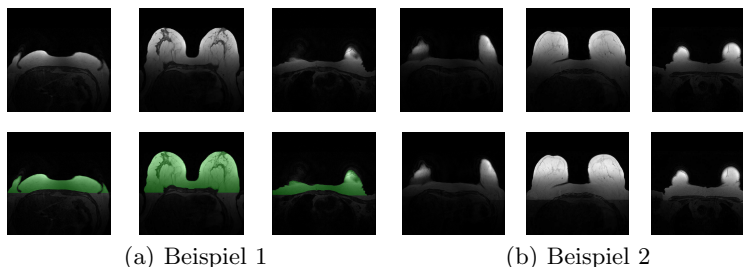
### 2.1 Eigenschaften der Daten

Das Datenset umfasst Brust-MRT Untersuchungen von 40 weiblichen Probanden aus der Study of Health in Pomerania [5] mit jeweils 128 Bildschichten, die im Abstand von 1.5 mm aufgenommen wurden. Je 10 Probanden gehören den 4 verschiedenen ACR Klassifikationen ACR0 - ACR3 an. Diese Klassifikation gibt an, wie groß der Anteil des Drüsengewebes jeweils ist, und wurde von einem Arzt vorgenommen. Die Bilder sind in axialer Richtung aufgenommen worden und verfügen über eine Auflösung von  $512 \times 512$  Pixeln. Die Aufnahmen wurden mittels eines 1.5 Tesla MRT-Geräts erstellt. Zu jedem Bild existiert außerdem eine von Ärzten mit mehr als 8-jähriger Berufserfahrung erstellte Maske mit korrekter Segmentierung von Brust und Hintergrund. Ein solches MRT-Bild mit der Maske ist beispielhaft in Abb. 2(a) dargestellt. Wie man sehen kann, weisen die Daten einen starken Helligkeitsgradienten in y-Richtung, sowie Helligkeitsinhomogenitäten an den Brusträndern auf. Dies ist problematisch, da der Algorithmus schneller lernt, die hellen Bereiche richtig zu segmentieren und die dunklen unter Umständen gar nicht. Auch der Kontrast zwischen Drüsen- und Fettgewebe ist in den dunklen Bereichen deutlich verringert. Ein Ziel der Datenvorbereitung wird es daher sein, diese Helligkeitsdifferenzen auszugleichen und den Kontrast zwischen den zu segmentierenden Klassen zu vergrößern.

### 2.2 Datenvorbereitung und Netzwerkarchitektur

Ein wichtiger Faktor für die Qualität des trainierten Netzwerkes ist die Menge der Daten. Um diese zu erhöhen, ist es sinnvoll beide Bildhälften einzeln für

**Abb. 1.** Obere Reihe: Exemplarische Darstellung der Daten aus verschiedenen Bereichen der Brust. Untere Reihe: Links: die manuell erstellten Masken wurden in grüner Farbe über das Bild gelegt. Rechts: Ergebnisse von mit N4ITK Bias Correction Algorithmus.



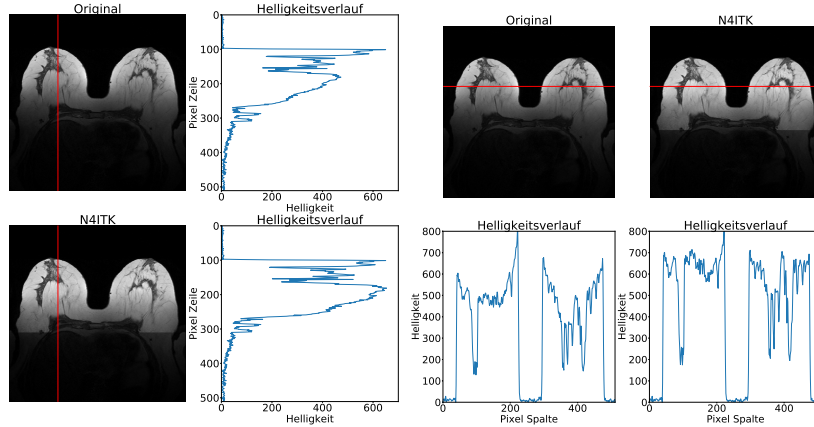
das Training zu verwenden und so die Anzahl der Bilder zu verdoppeln. Hierfür können die MRT-Bilder in der Mitte zerschnitten und die eine Hälfte, in diesem Fall die Rechte, gespiegelt werden. Da die MRT-Bilder bereits symmetrisch sind, ist dafür keine weitere Vorbereitung notwendig.

Darüber hinaus wurde für die Vorverarbeitung der Daten der sogenannte N4ITK Bias Correction Algorithmus [6] verwendet, um die im vorherigen Abschnitt beschriebenen Helligkeitsinhomogenitäten auszugleichen. Hierbei wurde der Algorithmus nur auf eine Maske angewendet, die durch Vergrößern und Verzerren der manuell erzeugten Segmentierungsmasken erzeugt wurde. Der Umriss der Brust ist dabei sehr grob und könnte leicht und unter geringem Zeitaufwand von einem Arzt erstellt werden.

Wie in Abb. 2 zu sehen ist, konnte auf diese Weise die Helligkeit im Vergleich zum Original im gesamten Bereich der Brust deutlich erhöht werden und damit auch der Kontrast zwischen Brust und Hintergrund. Auch die Helligkeitsdifferenz zwischen den Maxima und Minima und damit der Kontrast zwischen Drüsen- und Fettgewebe konnte durch die Verwendung der Maskierung erhöht werden. Die Helligkeitsinhomogenitäten an den Brusträndern konnten ebenfalls behoben werden. Betrachtet man die in Abb. 2 dargestellten Helligkeitsverläufe, so ist zu erkennen, dass der Helligkeitsgradient in y-Richtung nach wie vor vorhanden jedoch gegenüber dem Original verringert ist. Der Algorithmus funktioniert am besten in den mittleren Schichten, bei denen die Brust wie in Abb. 2 sehr großflächig ist. In den äußeren Schichten wird die Brust jedoch schmaler. Wie in Abb. 2(b) zu sehen, gelingt es dem Algorithmus in diesem Bereich nicht mehr, den Gradienten in y-Richtung zu korrigieren.

Wir haben die 2 Klassen-Netzwerkarchitektur vom Dalmis et al. [3] übernommen und für unsere Daten adaptiert. Als Lossfunktion  $E$  wird der Inverse Dice Koeffizient verwendet:  $E = 1 - DSC$ . Der Dice Koeffizient ist wie folgt definiert:  $DSC = 2|S_1 \cap S_2| / (|S_1| + |S_2|)$ , wobei  $S_1$  und  $S_2$  die manuelle Segmentierung und das vom Netzwerk gelernte Ergebnis sind. Aus jeder ACR-Gruppe mit jeweils 10 Datensätzen wurden randomisiert 2 Probandinnen für das Testing ausgewählt. Die restlichen 32 Datensätze wurden als Trainingsset genommen. Das Validie-

**Abb. 2.** Vergleich eines mit dem N4ITK Bias Correction Algorithmus mit Maskierung vorverarbeiteten MRT-Bildes mit dessen Original. Auf der linken Seite sind dabei die Helligkeitsverläufe in y-Richtung abgebildet, auf der rechten die Helligkeitsverläufe in x-Richtung.

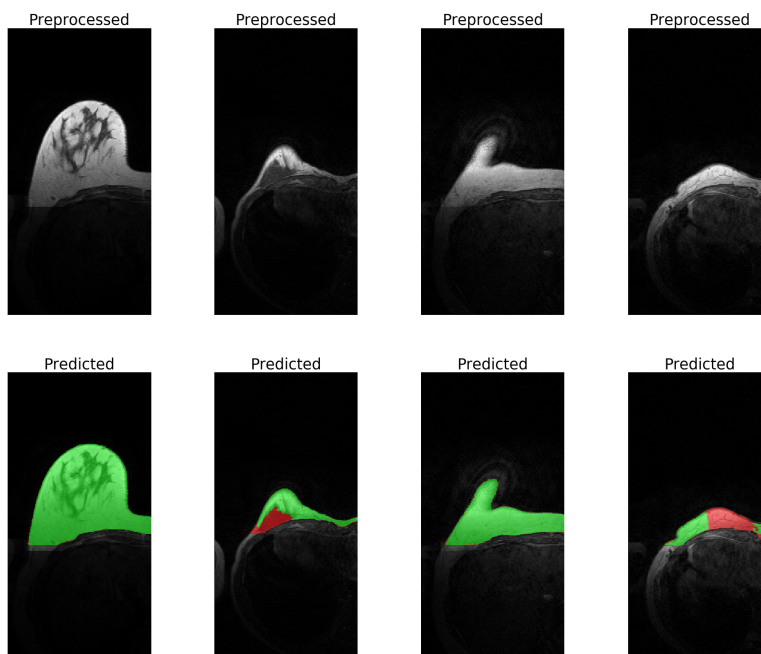


rungsset besteht aus ca. 20% der verbliebenen Daten, die vor jedem Trainingsdurchlauf zufällig ausgewählt werden. Dabei werden die Daten der Probandinnen so gruppiert, dass sie jeweils vollständig in einem der Sets lagen.

### 3 Ergebnisse

In einem ersten Versuch wurde ein U-net ohne Vorverarbeitung der Daten mit dem N4ITK Bias Correction Algorithmus trainiert. Dieses erreichte jedoch nur einen DSC von  $0.91 \pm 0.09$ . Im zweiten Versuch wurden die Daten mit dem N4ITK Bias Correction Algorithmus vorverarbeitet und bei der Segmentierung der Daten aus dem Testset ein DSC von  $0.98 \pm 0.06$  erreicht. In Abb. 3 sind exemplarisch 4 vorverarbeitete MRT-Bilder aus dem Testset und darunter die jeweils berechneten Segmentierungen dargestellt. Auf der linken Seite ist ein Bild aus einer der mittleren Schichten eines MRT-Scans dargestellt. Die berechnete Segmentierung stimmt bis auf einige Pixel am Rand genau mit der vom Arzt erstellten Maskierung überein. Die „Zebrastrifen“, wie sie im dritten Bild von links zu sehen sind, werden ebenso korrekt als Hintergrund klassifiziert. Problematisch sind nach wie vor MRT-Scans von Probandinnen mit viel Drüsengewebe und geringer Abgrenzung zwischen diesem und dem Hintergrund, wie beispielhaft auf dem zweiten Bild von links zu sehen ist. In diesem Fall wird nahezu das komplette Drüsengewebe dem Hintergrund zugeordnet. Der DSC der Segmentierungen der MRT-Bilder dieser Probandin liegt nur bei 0.93 und ist damit der niedrigste im gesamten Testset. In den äußeren Schichten werden anders als im ersten Netzwerk nicht mehr zu viel Pixel der Brust zugeordnet sondern zu wenige. Dabei haben die Segmentierungen oft unrealistische Formen und es lassen sich keine Bildfehler erkennen, die dieses Verhalten erklären könnten.

**Abb. 3.** Exemplarische Darstellung einiger Segmentierungen die mit dem U-net erstellt wurden. Dabei sind TP grün, FP rot und FN orange eingefärbt. Über den Segmentierungen ist jeweils zum Vergleich das vorverarbeitete Bild dargestellt.



Der Zeitaufwand pro Proband, also in diesem Fall 128 MRT-Bilder, bei der Segmentierung mit dem betrachteten U-net beträgt auf einer Graphikkarte mit 12 GB durchschnittlich ca. 13.6 s.

## 4 Diskussion

Dalmış et al. [3] zeigten, dass die U-Nets die klassischen Ansätze übertrafen, und beobachteten, dass die U-Net-basierten Methoden minimal von Intensitätsinhomogenitäten betroffen wurden. Wir haben eine ähnliche Netzwerkarchitektur auf Daten mit sehr starker Inhomogenität angewendet und gezeigt, dass das Netzwerk in unserem Fall Probleme hatte, die starken Inhomogenitäten zu lernen. Wir haben auch gezeigt, dass die Intensitätsinhomogenitätskorrektur mit N4ITK [6] die Leistung des Netzwerks deutlich verbessert hat. Bei der Segmentierung wurde bei dem ohne weitere Vorverarbeitungsmethoden trainierten U-net ein DSC von  $0.91 \pm 0.09$  erreicht. Dieses Ergebnis bleibt etwas zurück hinter dem von Dalmış et al. [3] erreichten DSC von 0.944 beim 2C-U-net zur selben Problemstellung, allerdings mit einem anderen Datensatz. In der vorliegenden Arbeit war dabei, der einschränkendste Faktor der starke Helligkeitsgradient in y-Richtung, der zu deutlich schlechteren Segmentierungen in den dunkleren Bereichen der Brust

führte. Unter Verwendung der mit dem N4ITK Bias Correction Algorithmus vorverarbeiteten Daten wurde ein U-net trainiert, das einen DSC von  $0.98 \pm 0.05$  bei der Segmentierung des Testsets erreichte.

Problematisch ist die Segmentierung von MRT-Scans mit sehr viel Drüsengewebe und geringer Abgrenzung zwischen diesem und dem Hintergrund, wie es in Abb. 3 zu sehen ist. Möglicherweise ließe sich dies durch Hinzufügen von Informationen der benachbarten Schichten verbessern, da diese zum Teil besser segmentiert werden konnten als im genannten Beispiel. In jedem Fall sinnvoll wäre das Hinzufügen der 3D-Informationen in den äußeren Schichten. In diesem Bereich ist es auch für Ärzte aufgrund eines einzelnen MRT-Bildes kaum möglich, dieses richtig zu segmentieren. Die manuelle Segmentierung erfolgt daher unter starker Berücksichtigung der umliegenden Schichten.

Sinnvoll wäre eine Verifizierung der Ergebnisse durch weitere Datensets. Ein Nachteil ist, dass die notwendigen groben Masken für N4ITK erstellt werden müssen. Zur Erstellung der Masken könnte beispielsweise ein Schwellenwert-Verfahren verwendet werden.

Abschließend lässt sich festhalten, dass die in dieser Arbeit erreichte Genauigkeit viel versprechend ist für einen Einsatz in der medizinischen Diagnostik. Darüber hinaus benötigt der Algorithmus im Durchschnitt lediglich eine Rechenzeit von 13.6 s pro Proband auf einer Graphikkarte mit 12 GB und ist somit für eine Anwendung in der Praxis geeignet.

## Literaturverzeichnis

1. Ivanovska T, Laqua R, Wang L, et al. A Level Set Based Framework for Quantitative Evaluation of Breast Tissue Density from MRI Data. PLOS ONE. 2014 11;9(11):1–19.
2. Gubern-Merida A, Kallenberg M, Mann RM, et al. Breast segmentation and density estimation in breast MRI: a fully automatic framework. IEEE J Biomed and Health Informatics. 2015;19(1):349–357.
3. Dalmis MU, Litjens G, Holland K, et al. Using deep learning to segment breast and fibroglandular tissue in MRI volumes. Medical Physics. 2017;44(2):533–546.
4. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2015;9351:234–241.
5. John U, Hensel E, Lüdemann J, Piek M, Sauer S, Adam C, et al. Study of Health In Pomerania (SHIP): a health examination survey in an east German region: objectives and design. Sozial-und Präventivmedizin. 2001;46(3):186–194.
6. Tustison NJ, Avants BB, Cook PA, et al. N4ITK: Improved N3 Bias Correction. IEEE Trans Med Imaging. 2010;29(6):1310–1320.