

Automatische Segmentierung des fibroglandulären Gewebes mittels U-Nets in der diffusionsgewichteten Bildgebung

Astrid Müller¹, Frederik B. Laun¹, Mona Pistel^{1,3}, Michael Uder¹, Rolf Janka¹, Andreas Maier², Sabine Ohlmeyer¹, Evelyn Wenkel¹, Sebastian Bickelhaupt¹

¹ Radiologisches Institut, Universitätsklinikum Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany.

² Department of Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany.

³ Siemens Healthcare, Erlangen, Germany.

Zusammenfassung: Es wurden zwei U-Nets für die Segmentierung des fibroglandulären Gewebes der weiblichen Brust in der diffusionsgewichteten Mammabildgebung trainiert und verglichen. Das FGT wird mit beiden U-Nets zum Großteil korrekt segmentiert.

Motivation

Die diffusionsgewichtete Bildgebung (DWI) ermöglicht in der Magnetresonanztomographie (MRT)¹ die Kartierung des Diffusionsverhaltens von Wassermolekülen unter anderem im Brustgewebe. Hier ist neben der Detektion von Läsionen auch die Beurteilung des fibroglandulären Drüsengewebes (fibroglandular tissue, FGT) von Bedeutung². Es lassen sich durch Verwendung unterschiedlicher Diffusionswichtungen (b-Werte) quantitative Parameter wie der ADC (Apparent Diffusion Coefficient) im FGT ableiten¹. Hierzu ist eine standardisierte und valide Segmentierung des FGTs notwendig. Ein Ansatz für die automatisierte Segmentierung des FGTs ist die semantische Segmentierung mittels des U-Nets³.

Material und Methoden

Es wurden zwei U-Nets auf ihre Fähigkeit das FGT auf DWI-Daten ($b = 800 \text{ s/mm}^2$) zu segmentieren untersucht. U-Net₁ wurde mit 250 DWI-Einzelschichten von 19 Probandinnen, akquiriert mit drei unterschiedlichen EPI-DWI-Sequenzen trainiert. Die zugehörigen Referenzmasken wurden durch Clustern mit dem k-Means-Algorithmus und Gaussian-Mixture-Modellen generiert. Für das Training des zweiten U-Net, U-Net₂, wurden 127 Referenzmasken ersetzt durch Segmentierungen, generiert durch U-Net₁ von 29 Probandinnen, akquiriert mit acht unterschiedlichen EPI-DWI-Sequenzen. Beide Datensätze wurden jeweils durch Datenaugmentation auf einen Datensatz von 3250 DWI-Bildern erweitert. Die Segmentierungsleistung der U-Netze wurde an einem Testdatensatz (30 Einzelschicht-Bilder, $b = 800 \text{ s/mm}^2$) mittels zweier Likert-Skalen (Tab. 1) von einem Reader (A.M.) visuell für alle Einzelschichten bewertet. Die Ergebnisse der Evaluierung wurden mittels Wilcoxon-Vorzeichen-Rang-Test evaluiert. Das Signifikanzniveau wurde mit $\alpha = 0.05$ gewählt.

Tab. 1: Likert-Skalen

Skala	Vollständigkeit der FGT Segmentierung	Inkorrekte Segmentierung
1	Keine	Großflächig
2	Vereinzelte Voxel	Gehäuft kleinflächig
3	Teilweise	Kleinflächig
4	Größtenteils	Vereinzelte Voxel
5	Nahezu vollständig, falls vorhanden	Keine

Ergebnisse

Die Vollständigkeit der FGT-Segmentierung der beiden U-Netze war weitgehend identisch. Der mediane Likert-Wert für die FGT-Segmentierung für U-Net₁ war 4.13, für U-Net₂ 4.18 ($p = 0.6811$). Allerdings waren inkorrekte Segmentierungen bei U-Net₂ seltener zu finden als bei U-Net₁ ($p = 0.0005$). Abbildung 1ab zeigt eine Schicht, auf der beide U-Netze das FGT bis auf den kleinen durch den weißen Pfeil markierten Bereich sehr gut segmentierten (Likert-Skalen-Werte: 5 für U-Net₁ und 5 für U-Net₂). Bei Abbildung 1cd wurde Fett- und Muskelgewebe von U-Net₁, im Gegensatz zu U-Net₂, inkorrekt als FGT segmentiert (Likert-Skalen-Werte: 1 für U-Net₁ und 5 für U-Net₂).

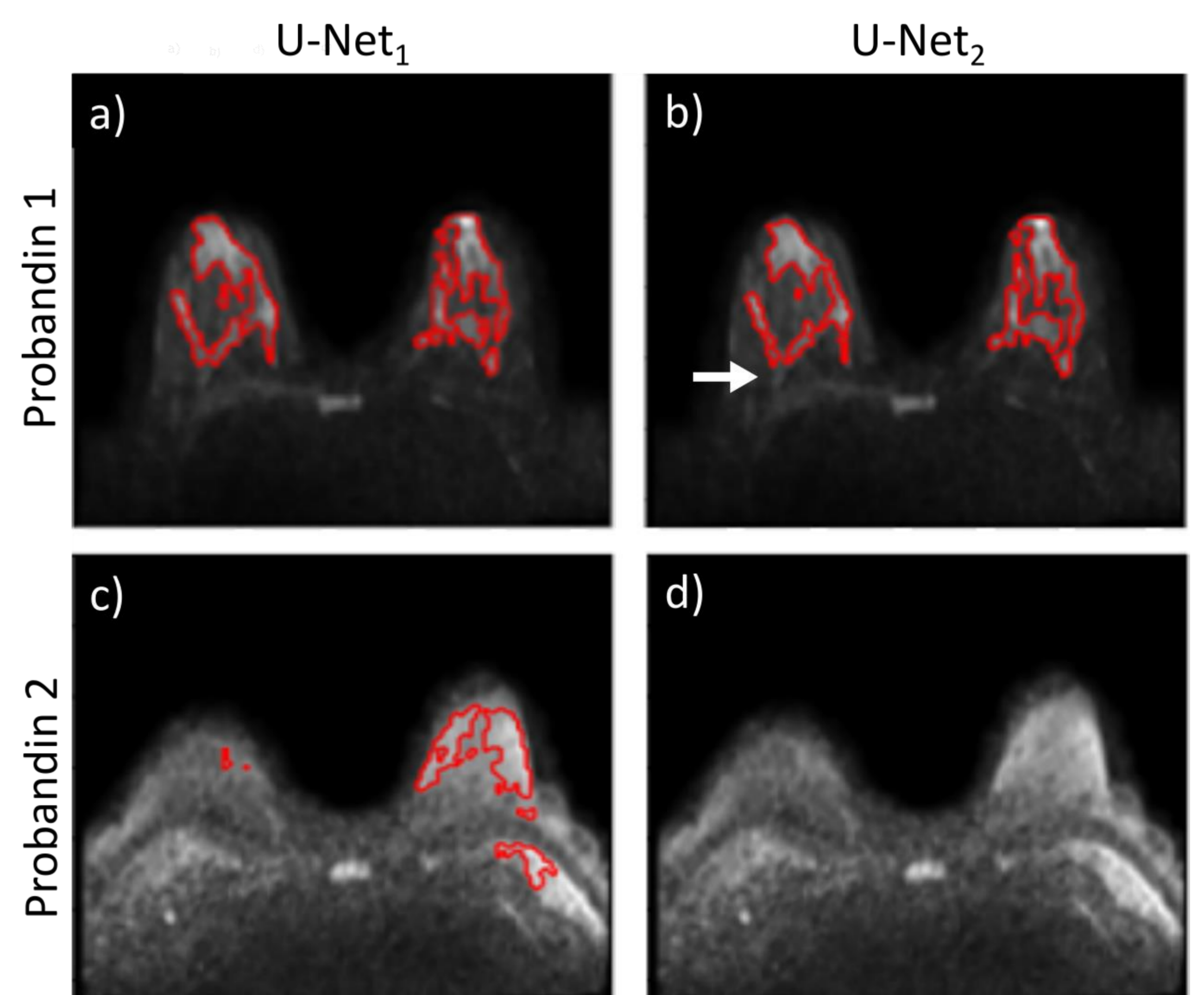


Abb. 1: FGT-Segmentierung durch U-Net₁ und U-Net₂

Diskussion

Beide U-Netze konnten das FGT größtenteils korrekt segmentieren, jedoch war keiner der Ansätze in der Lage, eine ausnahmslos perfekte Segmentierung zu erreichen. U-Net₁ zeigte hierbei eine größere Anfälligkeit für inkorrekte Segmentierungen. Ursächlich dafür könnten die komplexe Struktur und die hohen Unterschiede in der Signalintensität des FGTs bei hohen b-Werten sein. Ebenso war die vorhandene Menge an Trainingsdaten limitiert. Diese soll in weiterführenden Ansätzen deutlich erhöht werden. Auch sollen manuell eingezeichnete FGT-Segmentierungen unterschiedlicher DWI-Sequenzen erstellt werden, um eine Evaluation der Segmentierungsleistung mit verschiedenen Referenzmetriken zu ermöglichen.