

Effiziente Segmentierung trachealer Strukturen in MRI-Aufnahmen

Philip Dietrich¹, Catherine Schmidt², Henry Völzke², Achim Beule^{2,3},
Florentin Wörgötter¹, Tatyana Ivanovska¹

¹Georg-August-Universität Göttingen, Deutschland

²Universitätsklinikum Greifswald, Deutschland

³Universitätsklinikum Münster, Deutschland

tiva@phys.uni-goettingen.de

Kurzfassung. Die Segmentierung verschiedener Strukturen im Körper ist eine der grundlegenden Operationen in der medizinischen Bildverarbeitung. In dieser Arbeit werden auf Machine Learning basierende Methoden zur Segmentierung medizinischer Bilder untersucht. Das Ziel ist es, in MRI-Scans die Trachea zu segmentieren. Jedoch soll in dieser Arbeit speziell die Effizienz der Algorithmen im Vordergrund stehen. Die verwendeten Ansätze basierten auf einer Deep Learning Architektur, welche zunächst individuell optimiert wird.

Es konnte ein maximaler DICE-Koeffizient von $(94.4 \pm 2.1)\%$ erzielt werden. Zusätzlich kann festgestellt werden, dass die Segmentierung sehr effizient geschieht. Die Segmentierung von einem Datensatz aus 40 Schichten dauert dabei weniger als eine Sekunde, wobei bei bisherigen Methoden es über eine Minute benötigte.

1 Einleitung

Die Segmentierung medizinischer Bilddaten ist ein zentrales Problem der medizinischen Bildanalyse. Sie ist für die computergestützte ärztliche Diagnostik und Therapie von besonderer Bedeutung, da sie die Grundlage für eine weitergehende Analyse, Vermessung und 3D-Visualisierung medizinischer Bildobjekte bildet [1].

In diesem Bereich konnten zuletzt mit Hilfe des Maschinellen Lernens große Fortschritte erzielt werden. In den letzten Jahren brachten Neuronale Netzwerke deutliche Verbesserungen im Bereich der Bildverarbeitung [2]. Speziell die sogenannten Convolutional Neural Networks konnten klassische Methoden deutlich überbieten (zum Beispiel [3]). Im Bereich der Segmentierung wurden auch signifikante Verbesserungen erzielt, wie z. B., durch die Methode von Ronneberger et al. [4].

Die Untersuchung der Trachea in MRI-Aufnahmen spielt speziell bei Patienten mit einem inspiratorischen Stridor eine große Rolle. Hierzu ist es nötig die Trachea in den entsprechenden Aufnahmen zu segmentieren. Es gibt einige Verfahren für automatische und semi-automatische Segmentierung der oberen Atemwege. Ivanovska et al. entwickelte hierfür einen Algorithmus, welcher auf

Clustering-Methoden und Graph-Cuts basierte [5]. Shahid et al. [6] präsentierte einen Ansatz für Segmentierung von Pharynx und Fat Pads. Seifert et al. [7] schlug eine kombinierte Methode vor, um die Trachea und Wirbelsäule zu extrahieren. Das Problem von allen klassischen Ansätze ist allerdings in der fehlenden Effizienz. Die Segmentierung von einem MRI Datensatz benötigte damit über eine Minute und die Genauigkeit lag ung. bei 90 – 91%. Zu unserem Wissen wurde noch keine Deep Learning Methode bei der Trachea Segmentierung angewendet.

Das Ziel dieser Arbeit ist einen effizienten Deep Learning Algorithmus zur Segmentierung der Trachea in MRI-Aufnahmen zu finden.

2 Material und Methoden

2.1 Der Datensatz

Der Datensatz ist als Teil der Study of Health in Pomerania (SHIP) [8] aufgenommen worden. Sie wurde im Raum Vorpommern durchgeführt und dient der allgemeinen Untersuchung von Gesundheitsfaktoren und Risiken. Bei den verwendeten Daten handelt es sich um Aufnahmen aus MRT-Scans mit einem 1.5 Tesla MRT-Gerät (Magnetom Avanto, Siemens Medical Solutions, Erlangen, Deutschland). Ein einzelner Scan besteht aus 40 Bildern, welche jeweils mit 256×256 Pixeln aufgelöst sind. Speziell im unteren Bereich der Scans sind aufgrund der Atemluft der Patienten starke Artefakte zu erkennen, welche die Segmentierung deutlich erschweren.

Insgesamt wurden 110 Datensätze, entsprechend 110 Probanden, von den HNO-Experten manuell segmentiert.

2.2 Vorverarbeitungsmethoden

Die Vorverarbeitung lässt sich in drei Schritte gliedern. Das Ziel des ersten Schrittes ist das Entfernen von Rauschen in den Bildern. Hierzu wurde ein anisotroper Diffusionsfilter verwendet [1]. Danach wird eine Histogrammäquilisation angewendet [1]. In Abb. 1 sind die einzelnen Schritte der Vorverarbeitung an einem Beispiel dargestellt.

Im Anschluss werden die Grauwerte der einzelnen Bilder noch normalisiert. Das heißt, es wird jeweils der Mittelwert subtrahiert und durch die Standardabweichung geteilt, sodass die Grauwerte einer Normalverteilung folgen.

2.3 Die U-Net Architektur

Die folgende Netzwerkarchitektur wurde von Ronneberger et al. vorgeschlagen [4]. Aufgrund seiner Form in dieser Darstellung wird es als U-Net bezeichnet. Hier haben wir die Architektur in der ursprünglichen Form [4] verwendet.

Ein U-Net lässt sich in die typische Encoder-Decoder-Struktur unterteilen. Im ersten Teil des Netzwerks wird die Größe der Aktivierungen immer weiter verkleinert, wobei die Anzahl der Filter der Convolutional Layer jeweils nach einem Verkleinerungsschritt verdoppelt wird.

Im zweiten Teil des Netzwerkes wird mittels Up-Sampling die Größe der Aktivierungen wieder erhöht, wobei durch Halbierung der Filter die Tiefe jeweils verringert wird. Um die entgültige Segmentierungsmasken zu erhalten wird die Ausgabe des Netzwerkes noch mittels eines Thresholds binarisiert, dieser Wert wurde auf 0.5 festgesetzt.

3 Ergebnisse

Die Segmentierungen werden mit Hilfe des DICE-Koeffizienten ausgewertet, wobei die manuellen Segmentierungen von HNO Experten bereitgestellt wurden. Das Training Set besteht aus 101 MRT Datensätzen, von welchen 20% randomisiert zur Validierung verwendet wurden, um die Hyperparameter des Modells zu optimieren. Das Test Set besteht aus neun Datensätzen, welche zufällig aus dem gesamten Datensatz gewählt wurden.

Dieser lässt sich mit Hilfe von True Positives TP , False Positives FP und False Negatives FN schreiben als

$$DICE = \frac{2TP}{2TP + FP + FN}. \quad (1)$$

In den durchgeführten Experimenten wurden neben den Parametern für Diffusionsfilter und Datenerweiterungsmethoden auch die Netzwerkarchitektur auf verschiedene Weisen variiert. Einerseits wurden die Anzahl und Größe der Filter optimiert und andererseits die Verwendung von Dropout und Batch Normalisierung getestet. Hierbei konnte festgestellt werden, dass die Anzahl von 64 Filtern in der ersten Ebene des Netzwerkes, wie sie von der originalen U-Net-Architektur [4] verwendet wird, hier zu einer zu komplexen Architektur führte. Dies zeigte sich in starkem overfitting. Speziell die Verwendung von Batch Normalisierung

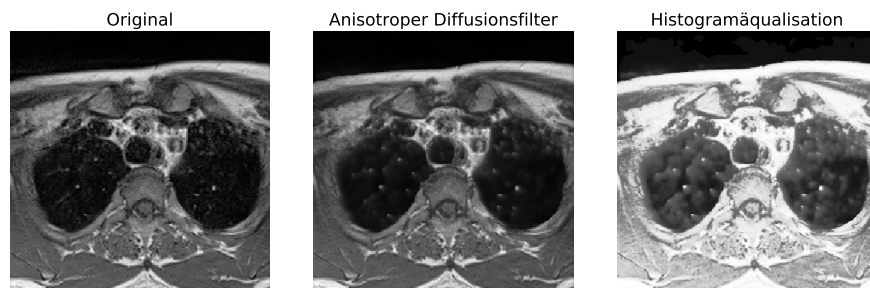


Abb. 1. Die einzelnen Vorverarbeitungsschritte am Beispiel: links das originale Bild, mittig das Bild nach der Anwendung des Diffusionsfilters, und rechts nach Histogramm-äqualisation.

Tabelle 1. Ergebnisse auf dem Testdatensatz für verschiedene Größen und Anzahlen an Filtern. Das Netzwerk wurde ohne Dropout trainiert.

Filteranzahl	8	8	16	16
Filtergröße	3	5	3	5
DICE [%]	93.3 ± 1.6	92.5 ± 2.8	93.9 ± 2.2	93.4 ± 2.0

war in diesem Fall nicht mehr sinnvoll, da die Batches aufgrund der gegebenen Speicherlimitierung nur noch zu wenige Beispiele enthielten [9].

Die Ergebnisse auf einem unabhängigen Testdatensatz aus 360 Bildern sind in Tab. 1 zusammengetragen. Der DICE-Koeffizient ist dabei jeweils als Mittelwert und Standardabweichung über die verschiedenen Scans angegeben. Die Verwendung von Dropout konnte nur in einem kleinen Bereich von Wahrscheinlichkeiten das overfitting vermindern (s. Abb. 3). Dieser Bereich lag bei etwa 0.2.

Zusätzlich sollen hier auch noch die Laufzeiten der Netzwerke untersucht werden. Diese teilen sich auf die Vorverarbeitung der Bilder und die eigentliche Segmentierung auf. Die Vorhersagen der Netzwerke wurden dabei auf einer Nvidia Titan X GPU gemacht. Als Zeit für die Vorverarbeitung ergaben sich auf einem Cluster mit $12 \times$ Intel(R) Core(TM) i7 CPU @ 3.33 GHz 23 Gb DDR4 durchschnittlich 0.38 s. Die Segmentierungen benötigten 0.24 s bei 8 Filtern beziehungsweise 0.39 s bei 16 Filtern in der ersten Schicht des Netzwerkes. Als gesamte Laufzeit des Algorithmus ergab sich also bei der Konfiguration, welche den höchsten DICE-Koeffizienten erzielte 0.77 s.

4 Diskussion

In Abb. 2 sind exemplarisch zwei Segmentierungen, wie sie vom U-Net erzeugt wurden gezeigt. Bei der Betrachtung aller Segmentierungen fällt auf, dass im oberen, sowie im unteren Bereich der Trachea, verglichen mit den mittleren Bereichen etwas schlechtere Ergebnisse erzielt werden. Dies lässt sich einerseits dadurch begründen, dass im unteren Bereich der Scans besonders starkes Rauschen vorkommt, was die Segmentierung deutlich erschwert. Das zweite Problem, welches mehrfach zu beobachten war, liegt darin, dass Flächen auf Bildern in denen die Trachea bereits nicht mehr zu sehen war, nicht als solche erkannt wurden. Speziell im Bereich knapp oberhalb der Trachea trat dieses Problem auf.

Zur Vermeidung von overfitting wurde außerdem die Verwendung von Batch Normalisierung und Dropout variiert. Hierbei konnte festgestellt werden, dass Batch Normalisierung sich dazu sehr gut eignet, sobald die Batches eine ausreichende Größe besitzen (in diesem Fall 16 Bilder). Hier wurde ein maximaler DICE-Koeffizient von $(94.4 \pm 2.1)\%$ auf den Testdaten erzielt. Jedoch ist zu erkennen, dass die Unterschiede zwischen einem Netzwerk mit Dropout (0.2) und einem Netzwerk ohne Dropout nur sehr gering sind. Zu erklären ist dies damit, dass bei der geringen Anzahl an verwendeten Filtern das overfitting nur eine sehr

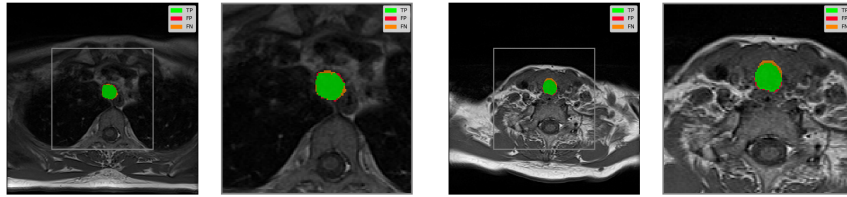


Abb. 2. Exemplarische Darstellung automatischer Segmentierungen, wobei jeweils True Positives, False Positives und False Negatives markiert wurden im unteren (links) und im oberen Bereich (rechts) der Trachea.

kleine Rolle spielt. Eine weitere Erhöhung der Dropout-Rate verschlechterte das Ergebnis wieder (s. Abb. 3).

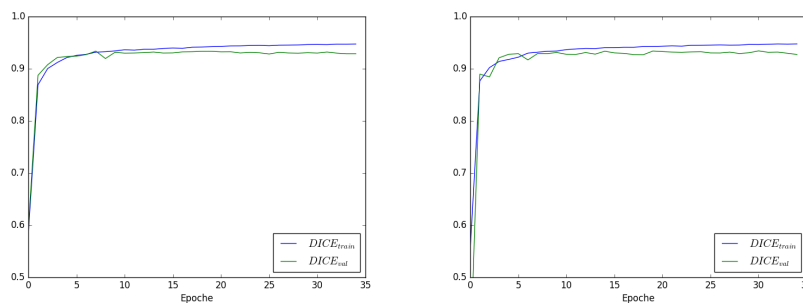


Abb. 3. Darstellung der Trainingskurven des DICE 2D -Koeffizienten (für die Trainings-Daten in blau, für die Validation-Daten in grün) in Abhängigkeit der verwendeten Dropoutrate für ein U-Net mit 16 Filtern in der ersten Ebene. Links: Dropout 0.0, Rechts: Dropout 0.2.

Es konnte gezeigt werden, dass die Segmentierung der Trachea in MRI-Aufnahmen mittels klassischer Methoden, wie z. B. von Ivanovska et al. [5], durch CNNs deutlich verbessert (94% vs 90%) werden konnte. Die Laufzeit für eine Segmentierung konnte dabei auf unter eine Sekunde reduziert werden.

Literaturverzeichnis

1. Handels H. Segmentierung medizinischer Bilddaten. Wiesbaden: Vieweg+Teubner; 2009.
2. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556. 2014;.

3. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*. 2012; p. 1097–1105.
4. Ronneberger O, Fischer P, Brox T; Springer. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2015; p. 234–241.
5. Ivanovska T, Buttke E, Laqua R, Völzke H, Beule A; IEEE. Automatic trachea segmentation and evaluation from MRI data using intensity pre-clustering and graph cuts. *Image and Signal Processing and Analysis (ISPA), 2011 7th International Symposium on*. 2011; p. 513–518.
6. Shahid MLUR, Chitiboi T, Ivanovska T, Molchanov V, Völzke H, Linsen L. Automatic MRI segmentation of para-pharyngeal fat pads using interactive visual feature space analysis for classification. *BMC medical imaging*. 2017;17(1):15.
7. Seifert S, Wachter I, Dillmann R. Segmentation of intervertebral discs, trachea and spinal cord from MRI images. *Proc Comput Assist Radiol Surg*. 2006;.
8. Voelzke H, Alte D, Schmidt CO, et al. Cohort Profile: The Study of Health in Pomerania. *Int J Epid*. 2011;40(2):294.
9. Ioffe S. Batch Renormalization: Towards Reducing Minibatch Dependence in Batch-Normalized Models. *arXiv preprint arXiv:170203275*. 2017;.