

Deep Learning-basierte Oberflächenrekonstruktion aus Binärmasken

Carina Tschigor, Grzegorz Chlebus¹, Christian Schumann¹

¹Fraunhofer-Institut für Digitale Medizin MEVIS
`christian.schumann@mevis.fraunhofer.de`

Kurzfassung. Die Darstellung anatomischer Strukturen auf Basis von Segmentierungsergebnissen in Form von Binärmasken ist eine grundlegende Aufgabe im Bereich der medizinischen Visualisierung. Hierfür werden meist polygonale Oberflächen genutzt. Bei Binärmasken fehlt jedoch die Information über die tatsächliche Oberfläche, wodurch die erzeugten Repräsentationen unter Artefakten leiden. Eine Glättung der Maske oder der polygonalen Repräsentation kann dies reduzieren, führt jedoch auch zu Ungenauigkeiten. Diese Arbeit beschreibt einen Ansatz zur Berechnung einer vorzeichenbehafteten Distanzfunktion für eine gegebene Binärmaske auf Basis eines neuronalen Netzwerkes. Diese lässt sich anschließend mit klassischen Methoden in eine Oberfläche überführen, deren Visualisierung glatt und artefaktfrei ist und nur minimale Abweichungen einführt.

1 Einleitung

Für Applikationen im Bereich der Diagnostik und vor allem der Therapieplanung ist eine qualitativ hochwertige Darstellung von segmentierten anatomischen Strukturen wünschenswert. Hierbei ist vor allem eine realistische Darstellung der Morphologie ohne störende Artefakte wichtig. Gleichzeitig sollte die Visualisierung nur minimale Ungenauigkeiten einführen. Durch die Segmentierung anatomischer Strukturen aus CT- oder MRT-Volumen werden typischerweise Binärmasken gewonnen, die jeden Voxel entweder als Teil der Struktur oder des Hintergrundes klassifizieren. Hierdurch geht der Partialvolumeneffekt verloren, wodurch die Erzeugung einer glatten Oberfläche durch gängige Polygonalisierungsverfahren wie den Marching Cubes Algorithmus [1] erschwert wird. Die Ergebnisse lassen sich teilweise durch Glättung verbessern, entweder angewendet auf die Binärmaske oder das Polygonalisierungsergebnis [2]. Die Reduktion aller Artefakte erfordert jedoch viele Iterationen oder große Nachbarschaften, worunter die Genauigkeit der Rekonstruktion leidet. Im Fall der nachträglichen Glättung der polygonalen Oberfläche kann das Ergebnis jedoch durch eine explizite Analyse der eingeführten Artefakte weiter verbessert werden [3,4]. Ein naheliegender Ansatz für die Glättung der Binärmaske vor der Oberflächenextraktion ist die Nutzung eines Gauß-Filters. Dies kann in Einzelfällen zu akzeptablen Ergebnissen führen. Je nach Morphologie werden aber unterschiedlich große

Kernel benötigt. Somit ist eine Anwendung unter Erhaltung kleiner Strukturen bei gleichzeitig ausreichender Glättung sehr weitläufiger Treppenartefakte nicht möglich. Vielversprechender ist die explizite Rekonstruktion bzw. Optimierung einer vorzeichenbehafteten Distanzfunktion auf Basis der Binärdaten, welche die Extraktion einer glatten Oberfläche erlaubt. Einige Verfahren leiten hierbei auf Basis der Voxeldaten Punkte und assoziierte Normalenvektoren an der Oberfläche ab [5,6], um gängige Punktwolken-basierte Oberflächenrekonstruktionsverfahren nutzen zu können, welche vorzeichenbehaftete Distanzfunktionen auf Basis der Punktwolken erzeugen, die in einem weiteren Schritt polygonalisiert werden können. Weitere Verfahren zur Erzeugung von Distanzfunktionen auf Basis von Punktwolken, wie KinectFusion [7] oder gar mit Hilfe von Deep Learning [8], wären hier denkbar, wurden aber noch nicht auf Voxeldaten angewendet. Eine direkte Optimierung einer Distanzfunktion auf Basis der Binärmaske wird zum Beispiel von Lempitzky vorgeschlagen [9]. Der hier präsentierte Ansatz, *Neural Network Narrow Distance Function* (N^3DF), verfolgt ein ähnliches Ziel, nutzt jedoch ein neuronales Netz, um aus der Binämaske eine Distanzfunktion abzuleiten, welche die Extraktion einer glatten Oberfläche ermöglicht.

2 Material und Methoden

Die grundlegende Idee des Ansatzes ist die Überführung der Binämaske in eine vorzeichenbehaftete Distanzfunktion mit Hilfe eines trainierten neuronalen Netzes. Hierbei markiert die Eingabemaske Hintergrundvoxel mit -1 und Voxel innerhalb der Struktur mit 1. Analog zu der in der Bildverarbeitung verbreiteten *Distanztransformation* (*DTF*) nennen wir diese Umwandlung *Neural Network Narrow Distance Function* (N^3DF). Im Unterschied zur klassischen Distanztransformation, die lediglich den Abstand zum nächstgelegenen Objekt-Voxel ausgibt, approximiert N^3DF die Entfernung zur ursprünglichen Oberfläche innerhalb einer kleinen Nachbarschaft der Oberfläche. Voxel innerhalb der Struktur erhalten den Wert 1, Voxel außerhalb den Wert -1. Voxel mit einem Abstand kleiner als 1 mm zur Oberfläche erhalten einen entsprechenden Gleitkommawert. Das Ergebnis der N^3DF ist somit eine beschnittene Distanzfunktion [7]. Sie kann anschließend mit einem Polygonalisierungsverfahren wie Marching Cubes in eine Oberfläche überführt werden unter Nutzung des Iso-Wertes 0 (Abb. 1).

Das Training des Netzes wird nicht mit realen Daten durchgeführt, da die Ground Truth für eine gegebene Maske nicht ermittelt werden kann. Stattdes-

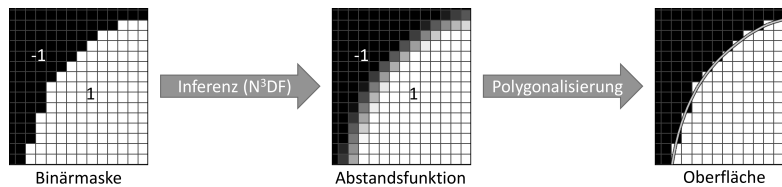


Abb. 1. Pipeline zur Erzeugung einer glatten Oberfläche aus einer binären Maske verdeutlicht an einem Ausschnitt einer 2D-Schicht.

sen werden künstliche Daten genutzt. Die grundlegende Idee ist hierbei, aus den glatten Oberflächen einfacher generischer Grundformen sowohl Binärmasken als auch Abstandsfunktionen abzuleiten und für das Training zu nutzen. Da hierbei die echte Geometrie bekannt ist, werden in den Trainingsdaten die echten Abstände gespeichert und für das Training genutzt. Ziel des Trainings ist somit, ein neuronales Netz zu erzeugen, das aus einer Binärmaske eine beschnittene Distanzfunktion ableiten kann, die den Abstand zu einer Oberfläche, die zu der Eingabemaske passen würde, darstellt. Es wird somit nicht direkt die Distanz zu der Grenze zwischen Vorder- und Hintergrund in der Maske wiedergegeben, wie dies bei gängigen Distanztransformationen der Fall ist, sondern der zu einer möglichen, glatten Oberfläche, die nahe dieser Grenze liegt.

Das trainierte Netz wird anschließend auf Binärmasken angewendet, die verschiedene Strukturen aus einem CT-Datensatz repräsentieren. Die resultierenden Distanzfelder werden mit Hilfe des Marching Cubes Algorithmus [1] in polygonale Oberflächen überführt. Die Experimente wurden in *MeVisLab*¹, einer integrierten Entwicklungsumgebung für medizinische Bildverarbeitung und Visualisierung durchgeführt. Hierbei wurde das integrierte *Remote Deep Learning Framework (RedLeaf)*² genutzt.

2.1 Trainingsdaten

Abb. 2 zeigt die verwendeten Grundformen für Training und Validierung des neuronalen Netzes. Um möglichst vielfältige Daten zu erhalten, werden die Grundformen in den drei Hauptachsen anisotrop skaliert und rotiert. Zusätzlich werden Kugeloberflächen mit kleinen Vertiefungen oder Erhebungen erzeugt, um auch kleine Details im Voxelbereich zu trainieren. Die Oberflächen werden unter Nutzung von *OpenVDB*³ voxelisiert bei gleichzeitiger Berechnung der Distanzfunktion. Daraus wird mit Thresholding die Binämaske erzeugt, um so die benötigten Datenpaare zu erhalten. Die Binärmasken müssen hierbei analog zur Inferenz ebenfalls Hintergrundvoxel mit -1 und Voxel innerhalb der Struktur mit 1 markieren. Insgesamt werden so 1600 Datenpaare erzeugt, die in 70% Training, 10% Validierung und 20% Testen gesplittet werden.

¹ <https://mevislab.de>

² <https://www.mevis.fraunhofer.de/en/solutionpages/deep-learning-in-medical-imaging.html>

³ <https://www.openvdb.org>

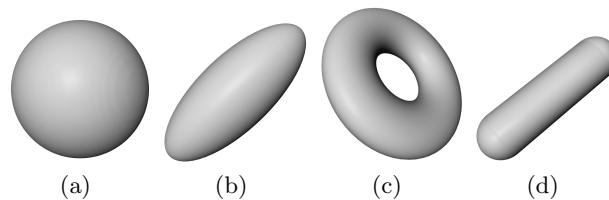


Abb. 2. Für das Training verwendete Grundformen: Kugel, Ellipsoid, Torus und Zylinder.

2.2 Netzwerk Architektur

Die hier verwendete Architektur ist ein 3D U-net [10]. Dies findet Anwendung sowohl in der Klassifikation, als auch in der Regression von dreidimensionalen medizinischen Bilddaten. Es werden drei Auflösungsstufen und eine Filtergröße von $3 \times 3 \times 3$ genutzt. Somit ergibt sich ein rezeptives Feld von 44^3 . Die Outputfunktion $\tanh$ schränkt den Output auf den gewünschten Wertebereich von $(-1,1)$ ein. Das Downsampling besteht aus je zwei aufeinanderfolgenden Faltungen und anschließenden Max Pooling ($2 \times 2 \times 2$). Zudem wird Batch Normalization verwendet. Beim Upsampling folgen auf eine transponierte Faltung ($2 \times 2 \times 2$) jeweils zwei Faltungen.

2.3 Training

Für das Training wird eine Patchsize von 64^3 sowie eine Batchsize von 2 verwendet. Eine Validierung erfolgt alle 250 Iterationen. Als Lossfunktion wird die mittlere quadratische Abweichung (MSE) gewählt. Die Lernrate ist auf 0.0001 gesetzt. Das Training hat eine Dauer von 200 Epochen. Anschließend wird das beste Modell verwendet, d.h. jenes, welches auf den Validierungsdaten den geringsten Fehler hat.

3 Ergebnisse

Aus einem CT-Datensatz eines Torsoausschnittes wurden fünf Binärmasken verschiedener Strukturen (Rippen, Leber, Lebergefäßbaum, Lunge und Haut) erzeugt und die Distanzfunktionen durch Inferenz des trainierten neuronalen Netzes erlangt (NVIDIA GeForce GTX 1050 Ti GPU mit 4GB Videospeicher). Hierbei kam eine Tilesizesize von 32^3 zum Einsatz. Als Beispiel werden hier die Ergebnisse für den Datensatz mit knöchernen Gewebe wie Rippen und Wirbelsäule gezeigt. Die Berechnung der N^3DF für die $333 \times 267 \times 169$ große Binärmaske dauerte 91s. Abb. 3 zeigt das Ergebnis der Oberflächenrekonstruktion mit

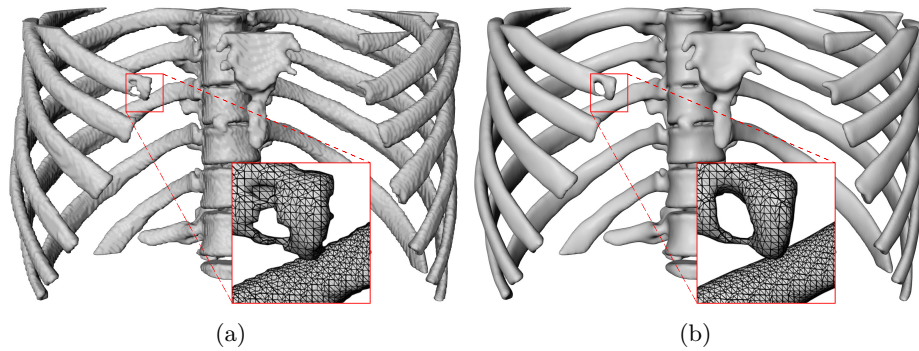


Abb. 3. Vergleich von Marching Cubes angewendet auf die Binärmaske (a) und auf das Ergebnis der N^3DF (b). Die Vergrößerung (rotes Quadrat) zeigt die Polygondichte sowie Abweichungen am Beispiel eines kleinen Loches.

Marching Cubes auf Basis der N^3DF im Vergleich zur Rekonstruktion auf Basis der Binärmaske. Zur quantitativen Bewertung wurde die Abweichung der N^3DF -Oberfläche von der Binärmaske berechnet. Die mittlere Abweichung beträgt 0,28 Voxeldiagonalen und ist somit sehr gering. Es wurde jedoch eine maximale Abweichung von 3,6 Voxeldiagonalen ermittelt. Solch starke Abweichungen treten vor allem bei sehr kleinen Strukturen oder Vertiefungen auf (siehe Vergrößerung in Abb. 3).

Um die Glattheit der erzeugten Oberflächen objektiv beurteilen zu können, wurden Krümmungswerte in Form des Dihedralwinkels für die Oberflächenrekonstruktion mit Marching Cubes auf Basis der Binärmaske und der N^3DF berechnet und auf der Oberfläche visualisiert (Abb. 4). Die Polygongitter unterscheiden sich bzgl. ihrer Komplexität kaum voneinander. Die N^3DF -Oberfläche besteht aus 212902 Knoten und 425876 Polygonen, die Oberfläche basierend auf der Binärmaske aus 214865 Knoten und 429802 Polygonen.

4 Diskussion

Wir konnten zeigen, dass mit Hilfe eines neuronalen Netzes aus einer dreidimensionalen Binärmaske eine Distanzfunktion berechnet werden kann, die die Erzeugung einer glatt wirkenden Oberfläche erlaubt. Diese Glattheit resultiert daraus, dass hier nicht direkt die Distanz zu der Grenze zwischen Vorder- und Hintergrund berechnet wird, wie dies bei gängigen Distanztransformationen der Fall ist, sondern der Abstand zu einer möglichen, glatten Oberfläche, die nahe dieser Grenze liegt. Unseres Wissens nach ist N^3DF der erste Ansatz zur Berechnung einer solchen Distanzfunktion mit Hilfe von Deep Learning. Die Berechnung der Distanzfunktion nimmt dabei je nach Hardware eine nicht unerhebliche Zeit in Anspruch. Im Vergleich zu anderen Verfahren ist die Berechnungsdauer jedoch durchaus praktikabel und wir hoffen die Geschwindigkeit in Zukunft durch die

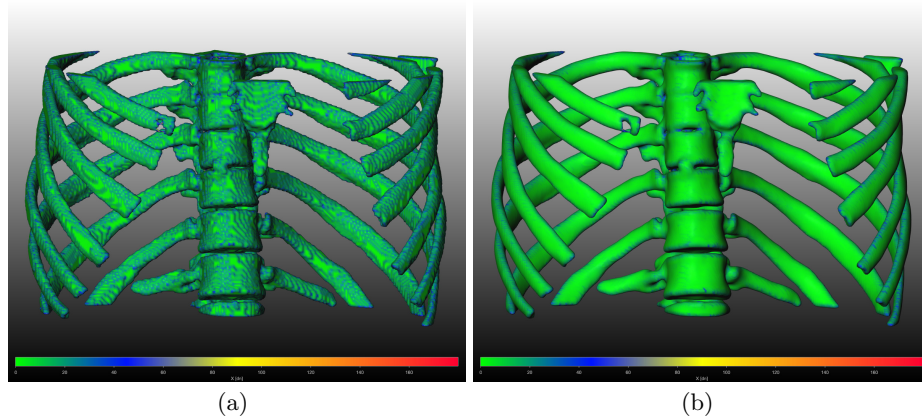


Abb. 4. Visualisierung des Dihedralwinkels für die Rekonstruktion auf Basis der Binärmaske (a) sowie auf Basis von N^3DF (b).

Optimierung der Netzwerkarchitektur weiter steigern zu können. Desweiteren profitiert das Verfahren von der rasanten Entwicklung von Hardware, die speziell auf Deep Learning ausgelegt ist, wie zum Beispiel die hier genutzten GPUs von NVIDIA. Die Komplexität der entstandenen Oberflächen ist in Hinsicht auf die Polygonzahl vergleichbar mit der Anwendung von Marching Cubes auf Binärmasken. Die erzeugten Oberflächen sind jedoch weitaus glatter und geben die Morphologie der anatomischen Strukturen ohne störende Artefakte wieder. Allerdings können noch vor allem bei kleinen Strukturen und Vertiefungen nicht unerhebliche Fehler auftreten. Eine Überabtastung könnte hier zu einer Verbesserung führen und so auch die kleinsten Strukturen korrekt darstellen.

In zukünftigen Experimenten sollten auch Trainingsdaten genutzt werden, die aus anatomischen Datensätzen gewonnen wurden, um das neuronale Netz noch besser an solche Strukturen anzupassen. Desweiteren soll die Nutzung anderer Polygonalisierungsverfahren getestet werden um z.B. eine adaptive Polygongröße zu gewährleisten und somit die Komplexität der erzeugten polygonalen Oberflächen zu reduzieren.

Literaturverzeichnis

1. Lorensen WE, Cline HE. Marching cubes: A high resolution 3D surface construction algorithm. *Comput Graph (ACM)*. 1987;21(4):163–169.
2. Taubin G. A signal processing approach to fair surface design. *Proc Conf Comp Graph Interact Tech*. 1995; p. 351–358.
3. Moench T, Gasteiger R, Janiga G, et al. Context-aware mesh smoothing for biomedical applications. *Comput Graph*. 2011;35(4):755–767.
4. Wei M, Wang J, Guo X, et al. Learning-based 3D surface optimization from medical image reconstruction. *Opt Lasers Eng*. 2018;103:110–118.
5. Schumann C, Oeltze S, Bade R, et al. Model-free surface visualization of vascular trees. *Eurographics/ IEEE-VGTC Sym Vis*. 2007; p. 283–290.
6. Wu J, Wei M, Li Y, et al. Scale-adaptive surface modeling of vascular structures. *Biomed Eng Online*. 2010;9(1):75.
7. Newcombe RA, Izadi S, Hilliges O, et al. KinectFusion: Real-time dense surface mapping and tracking. *Proc IEEE Int Sym Mix Augm Real*. 2011; p. 127–136.
8. Park JJ, Florence P, Straub J, et al. DeepSDF: Learning continuous signed distance functions for shape representation. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2019; p. 165–174.
9. Lempitsky V. Surface extraction from binary volumes with higher-order smoothness. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2010; p. 1197–1204.
10. Çiçek Ö, Abdulkadir A, Lienkamp SS, et al. 3D U-Net: learning dense volumetric segmentation from sparse annotation. *Med Image Comput Comput Assist Interv*. 2016; p. 424–432.