Fakultät für Informatik
Technische Universität München

Convolutional Neural Networks for Classification and Segmentation of Medical Images

Patrick Ferdinand Christ

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor- Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender:
Prof. Dr. rer. nat. Nils Thuerey

Prüfende der Dissertation:
Prof. Dr. rer. nat. Bjoern Menze
Prof. Dr.-Ing. Klaus Diepold

Die Dissertation wurde am 06.07.2017 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 02.11.2017 angenommen.

Abstract

Over 700.000 people die every year due to liver cancer. Tumors in the liver are according to WHO the fifth most common cancer type. To stage the therapy response of tumor diseases, radiologists and oncologists analyze tumors of the liver over time. Changes in tumor size and texture give experts information about therapy success. Detection and segmentation of tumor diseases as well as classification of tumor maligncy plays an important role in the development of computer-aided diagnosis systems (CADs).

This thesis investigates the application of Convolutional and Fully Convolutional Neural Networks for automatic detection und segmentation of medical image data. Cascaded Fully Convolutional Networks have been developed to tackle the segmentation of liver and liver tumor in Computed Tomography and Magnetic Resonance Imaging. In addition, the survival time of HCC patient could be predicted using a 3D Convolutional Neural Network. Furthermore, Convolutional and Fully Convolutional Neural Networks have been succesfully applied to estimate bread units for diabetes patients from food images.

The automatic segmentation of liver with Fully Convolutional Neural Networks achieved a DICE score of 94% for CT and 88% of MRI. Liver tumor was segmented with a Dice Score of 58% for CT and 69% for MRI. An automatic CAD system to stratify HCC patients according to their predicted survival time achieved an accuracy of 68%. The automatic estimation of bread units for diabetes patients achieved an RMSE of 1.53 bread units in comparision to RMSE of 0.89 bread units rated by expierenced diabetes patients.

All in all, this thesis showed that Convolutional and Fully Convolutional Neural Netwoks have a large potential for developing computer-aided diagnosis systems (CADs) for tumor diseases. This technology could also be applied in the field of computer-aided nutrition estimation for diabetes patients.

Zusammenfassung

Über 700.000 Menschen sterben jedes Jahr an den Folgen einer Tumorerkrankung der Leber. Lebertumor ist nach WHO die fünfthäufigste Krebserkankung. Radiologen und Onkologen untersuchen Tumore der Leber zur Feststellung des Therapieverlaufs. Änderungen der Tumorgrößen und -texturen geben Experten Aufschluss über den Therapieerfolg der durchgeführten Behandlung und Medikation. In der medizinischen Bildverarbeitung spielen die automatische Detektion, Segmentierung und Klassifizierung von Tumorerkankungen eine wichtige Rolle bei der Entwicklung von computer-gestützten Diagnosesystemen (CADs).

Diese Arbeit untersucht den Einsatz von Convolutional und Fully Convolutional Neural Networks für automatische Detektion und Segmentierung von medizinischen Bilddaten. Cascaded Fully Convolutional Neural Networks wurden entwickelt, um die Leber und Tumore der Leber automatisch auf Computertomographie- und Magnetresonanztomographie-Aufnahmen detektieren und segmentieren zu können. Des Weiteren konnte die Überlebenszeit von HCC-Patienten mit Hilfe von 3D Convolutional Neural Networks vorhergesagt werden. Weitere Anwendung fanden Convolutional and Fully Convolutional Neural Networks bei der automatischen Schätzung von Broteinheiten für Diabetiker aus Bildern von Gerichten.

Die automatische Segmentierung der Leber mit Hilfe von Fully Convolutional Neural Networks erreichte einen DICE Score von 94% bei CT und 88% bei MRT. Lebertumore konnten mit einem DICE Score von 58% für CT und 69% für MRT segmentiert werden. Ein automatisches CAD-System zur Stratifizierung von HCC-Patienten hinsichtlich ihrer erwarteten Überlebenszeit aus DW-MRI Bilddaten erzielte eine Genauigkeit von 68%. Die automatische Broteinheitenschätzung für Diabetiker erreichte einen RMSE von 1.53 Broteinheiten, während Diabetiker selbst Broteinheiten mit einem RMSE von 0.89 schätzten.

Zusammenfassend konnte in dieser Arbeit gezeigt werden, dass Convolutional und Fully Convolutional Neural Networks großes Potential zur Entwicklung computergestützter Diagnosesysteme (CADs) haben. Weitere Anwendungsfelder dieser Technologien können im Bereich der computergestützen Nährstoffermittlung für Diabetespatienten liegen.

Danksagungen

Zu Beginn möchte ich mich bei all jenen Menschen bedanken, die mich auf dem Weg zur Fertigstellung dieser Dissertation begleitet haben.

Zuerst danke ich Prof. Dr. Bjoern Menze dafür, mir die Möglichkeit gegeben zu haben, an diesem interessanten und fordernden Thema zu arbeiten. Dank Bjoern gelang mir der Wechsel von der Physik zur Medizininformatik, in der ich viele spannende Themen der künstlichen Intelligenz kennenlernen konnte. Durch sein Vertrauen und seiner gezielten Förderung konnten viele spannende Forschungsprojekte in dieser Arbeit realisiert werden. Vielen, vielen Dank.

Ich freue mich sehr darüber mit Prof. Dr. Klaus Diepold einen zweiten Doktorvater und Mentor gefunden zu haben. Vielen Dank für die tolle Unterstützung und die vielen interessanten Diskussionen bei der ein oder anderen Tasse Kaffee. Die gemeinsame Arbeit am CDTM und die spannenden Kurs- und Lehrprojekte werden mir immer in Erinnerung bleiben.

Natürlich sei auch Dr. Seyed-Ahmad Ahmadi für die vielen Diskussionen und die kurzweiligen Exkursionen in die Welt der VR und CB gedankt. Nur mit ihm und seiner Hilfe und Erfahrung konnten viele Publikationen geplant und durchgeführt werden.

Ich freue mich ganz besonders, dass mir die Möglichkeiten gegeben wurden viele spannende Projekte und Arbeiten zu betreuen. Die Arbeit hat mich persönlich sehr bereichert. Ich freue mich, dass in diesem Zusammenhang zahlreiche Freundschaften entstanden sind. Vielen Dank an Marc Bickel, Patrick Bilic, Mohamed Ezz, Florian Ettlinger, Robert Weindl, Sebastian Schlecht, Felix Grün, Timmy Smith und Christoph Heinle.

Des Weiteren danke ich dem gesamten CA Team und allen Studenten und Mitarbeiter des CDTMs für die offene Aufnahme, die herzliche Atmosphäre und die spannende Zeit in meinem Leben. Mein besonderer Dank gilt Stefan Nothelfer, Florian Lachner und Patrick Bilic für die tolle Unterstützung bei unseren gemeinsamen Projekten.

Ich danke allen Forschungs- und Kollaborationspartnern sowie dem Lehrstuhl für die gute und erfolgreiche Zusammenarbeit. Insbesondere möchte ich meinen Dank an Sunil Tatavatry, Markus Rempfler, Jana Lipkova, Georg Kaissis, Rickmer Braren, Julian Holch und Wieland Sommer aussprechen. Vielen Dank.

Schließlich danke ich Franziska und Roland Fresz sowie meinen Eltern Monika und Ferdinand Christ für das Korrekturlesen und die moralische Unterstützung.

Publikationen

Diese kumulative Dissertation enthält die folgenden Veröffentlichungen und unveröffentlichten Manuskripte in ihrer originalen Fassung:

Veröffentlichte Artikel

1. P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D'Anastasi, W. H. Sommer, S.-A. Ahmadi, and B. H. Menze. Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields. In *Medical Image Computing and Computer-Assisted Intervention*, pages 415–423. Springer International Publishing, 2016

2. P. F. Christ, F. Ettlinger, G. Kaissis, S. Schlecht, F. Ahmaddy, F. Grün, A. Valentinitsch, S.-A. Ahmadi, R. Braren, and B. H. Menze. SurvivalNet: Predicting patient survival from diffusion weighted magnetic resonance images using cascaded fully convolutional and 3D convolutional neural networks. In *IEEE International Symposium on Biomedical Imaging*. IEEE, 2017

3. P. F. Christ, F. Lachner, A. Hösl, B. H. Menze, K. Diepold, and A. Butz. Human-Drone-Interaction: A Case Study to Investigate the Relation Between Autonomy and User Experience. In *European Conference on Computer Vision Workshops*, pages 238–253. Springer International Publishing, 2016

Unveröffentlichte Manuskripte

1. P. F. Christ, S. Schlecht, F. Ettlinger, S.-A. Ahmadi, K. Diepold, and B. H. Menze. Diabetes60 - Inferring Bread Units From Food Images Using Fully Convolutional Neural Networks. *Unveröffentlichtes Manuskript*, 2017

2. P. F. Christ, F. Ettlinger, F. Grün, M. E. A. Elshaera, J. Lipkova, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, F. Hofmann, M. D. Anastasi, S.-A. Ahmadi, G. Kaissis, J. Holch, W. Sommer, R. Braren, V. Heinemann, and B. H. Menze. Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks. *Unveröffentlichtes Manuskript*, 2017

3. J. Lipková, M. Rempfler, P. F. Christ, J. Lowengrub, and B. H. Menze. Automated Unsupervised Segmentation of Liver Lesions in CT scans via Cahn-Hilliard Phase Separation. *Unveröffentlichtes Manuskript*, 2017

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

Akronyme

ADC	Apparent Diffusion Coefficent.
BRATS	Brain Tumor Segmentation.
CFCN	Cascaded Fully Convolutional Neural Network.
CNN	Convolutional Neural Networks.
CPU	Central Processing Unit.
CT	Computertomographie.
DW-MRI	Diffusion weighted Magnetic Resonance Imaging.
FCN	Fully Convolutional Neural Networks.
GAN	Generative Adversarial Network.
GLCM	Grey-Level Co-Occurrence Matrix.
GPU	Graphical Processing Unit.
HCC	Hepatocellular Carcinoma.
HCI	Human Computer Interaction.
HDI	Human Drone Interaction.
LITS	Liver Tumor Segmentation.
LSTM	Long Short Term Memory.
MLP	Multi Layer Perceptron.
MRI	Magnetic Resonance Imaging.
MRT	Magnetresonanztomographie.
NMR	Nuclear Magnetic Resonance.
PDAC	Pancreatic Ductal Adenocarcinoma.
RECIST	Response Evaluation Criteria in Solid Tumors.
RNN	Recurrent Neural Network.
ROI	Region of Interest.

Einleitung

1 Medizinische Bildgebung

1.1 Geschichtliche Entwicklung medizinischer Bildgebung

Der englische Physiker Robert Hooke veröffentlichte im Jahr 1665 seinen Buchband *Micrographia* über lichtmikroskopische Aufnahmen. Hooke, der der breiten Öffentlichkeit für die Entdeckung der elastischen Verformung von Festkörpern (Hooksches Gesetz der Physik) bekannt ist, baute eines der ersten Lichtmikroskope. In seinem Werk *Micrographia* untersuchte er mit seinem Lichtmikroskop verschiedene Objekte und Pflanzen. Seine Beobachtungen hielt er mit detailtreuen Zeichnungen fest. In einer seiner Beobachtungen von natürlichem Kork erkannte er die zelluläre Struktur von Pflanzen und führte den Begriff der Zelle ein [7, 8]. Abbildung 1 (a) zeigt die erste Darstellung der zellulären Struktur von Pflanzen.

Willhelm Conrad Röntgen entdeckte 220 Jahre nach Hooke in seiner Arbeit *Ueber eine neue Art von Strahlung* die nach ihm benannten Röntgenstrahlen. Mit dieser Entdeckung, die ihm 1901 auch den ersten Physiknobelpreis einbrachte, war Röntgen als erster Mensch in der Lage den menschlichen Körper ohne äußeren Eingriff (nicht-invasiv) zu untersuchen und begründete damit die moderne Radiologie [9, 8, 10]. In Abbildung 1 (b) ist eine der ersten nicht-invasiven Aufnahmen der menschlichen Hand abgebildet.

Ein wichtiger Schritt hin zu modernen Bildgebungsmodalitäten ist die Entwicklung der Computertomographie. Die Computertomographie ermöglicht, im Gegensatz zur klassischen zweidimensionalen Röntgenaufnahme, eine dreidimensionale Aufnahme und Untersuchung des menschlichen Körpers. Grundlage hierfür liefert das von dem Mathematiker Johann Radon entwickelte Konzept der Radontransformation. Die Radontransformation bildet die mathematische Basis zur Rekonstruktion von dreidimensionalen Objekten aus zweidimensionalen Röntgenaufnahmen [11, 12, 13].

Durch Allan M. Cormack und Godfrey Hounsfield wurde die Computertomographie zur praktischen Anwendung gebracht. Cormacks theoretische Beiträge halfen Hounsfield dabei den ersten Computertomographen zur Untersuchung von Menschen zu bauen. Am 01.10.1971 wurde am Atkinson Morley Hospital in Wimbledon der erste Mensch mit einem Computertomographen untersucht. Cormack und Hounsfield erhielten für ihre Beiträge zur Computertomographie 1979 gemeinsam den Nobelpreis für Medizin [12, 13, 14]. Seitdem hat sich die Qualität und Leistungsfähigkeit der Computertomographie stark verbessert. Im Jahr 2009 wurden in Deutschland über 9 Millio-

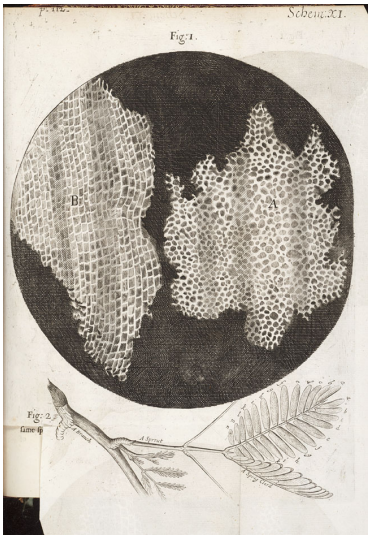
nen CT-Untersuchungen durchgeführt, Tendenz steigend [15].

Neben der Computertomographie ist die Magnetresonanztomographie mit über 7,9 Millionen Untersuchungen im Jahr 2009 eines der wichtigsten medizinischen Bildgebungsverfahren [15]. Die Magnetresonanztomographie nahm ihren Anfang durch die Entdeckung des magnetischen Kernspins (engl. Nuclear Magnetic Resonance, NMR) durch Felix Bloch und Edward Purcell [16, 17]. In einem konstanten Magnetfeld können Atomkerne mit einem Kernspin ungleich 0 elektromagnetische Wechselfelder absorbieren oder emittieren. Die emittierten elektromagnetischen Wechselfelder können aufgezeichnet werden und geben Aufschluss über die chemische Zusammensetzung der untersuchten Probe. 1971 gelang es Raymond Damadian mit Hilfe der Kernspinresonanz bösartigen Tumor von gesundem Gewebe zu unterscheiden. Damadian konnte Unterschiede in den Relaxationszeiten von Tumor- und Normalgewebe nachweisen und legte damit den Grundstein für die nicht-invasive Diagnostik von Tumorerkrankungen [18, 8]. Paul Lauterbur gelang durch Verwendung eines zusätzlichen ortsabhängigen Magnetfeldes die erste, zweidimensionale Aufnahme einer biologischen Probe [19]. Die erste Aufnahme des menschlichen Körpers fertigte Damadian im Jahr 1977 an. Er entwickelte den ersten MRT-Scanner, der in der Lage war den menschlichen Körper nichtinvasiv mit Hilfe von Magnetresonanz und ohne Strahlenbelastung zu untersuchen. Abbildung 1 (c) zeigt Damadian bei der ersten MRT-Untersuchung, welche mehrere Stunden andauerte. Die lange Untersuchungszeit von mehreren Stunden und die geringe örtliche Auflösung waren in der Anfangszeit der MR Bildgebungen die größten Hürden für den praktischen Einsatz im Krankenhaus. Peter Mansfield und Axel Haase arbeiteten beide an schnellen Bildgebungsverfahren und konnten die Aufnahmezeit von mehreren Stunden hin zu Minuten senken. Diese Errungenschaften ermöglichten den klinischen Einsatz von Magnetresonanztomographie zur Untersuchung von Patienten. Lauterbur und Mansfield wurden für ihre Beiträge zur Magnetresonanztomographie 2003 mit dem Nobelpreis für Medizin ausgezeichnet [19, 20, 8, 21].

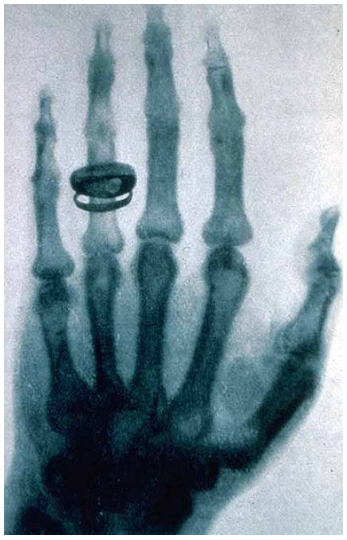
Moderne Bildgebungsmodalitäten wie Computertomographie und Magnetresonanztomographie erlauben Radiologen und Onkologen die nicht-invasive Untersuchung von Tumorerkrankungen. Mit Hilfe von Kontrastmitteln verstärkten MR- oder CT-Aufnahmen lassen sich Tumorerkankungen in frühen Stadien diagnostizieren. Eine frühe Diagnose ermöglicht einen früheren Therapiebeginn und führt schließlich zu einem höheren Therapieerfolg.

1.2 Bedeutung der Leber für die Diagnostik bei Tumorerkrankungen

Die Leber ist eines der wichtigsten Organe bei der Diagnose von Tumorerkrankungen [23, 1]. Sie übernimmt im menschlichen Körper wichtige Aufgaben im Stoffwechsel und sorgt für den Abbau von Nähr- und Giftstoffen. Aus diesem Grund streuen zahlreiche primäre Tumorerkankungen wie z.B. Prostata-, Brust-, Darm- und Pancreastumor im zeitlichen Krankheitsverlauf Metastasen in die Leber. Im Krankheitsverlauf lassen sich somit strukturelle Änderungen an der Leber sowie die Entstehung und



(a) Erste Darstellung einer Zelle



(b) Frühe Aufnahme der menschlichen Hand aufgenommen von Röntgen

(c) Erster MRI Scan eines Menschen von Damadian

Abbildung 1: Überblick über wichtige Entdeckungen der modernen medizinischen Bildgebung. (a) Der Physiker Robert Hooke entdeckte 1665 in seinem Werk *Micrographia* als erster Mensch mit Hilfe eines selbstgebauten Lichtmikroskops die zelluläre Struktur von Pflanzen. Die Abbildung zeigt eine detailtreue Skizze der zellulären Struktur von Kork, die Hooke mit seinem Lichtmikroskop untersuchte. (b) Willhelm Conrad Röntgen entdeckte 1898 die nach ihm benannten Röntgenstrahlen, die es ermöglichten den menschlichen Körper nicht-invasiv zu untersuchen. In der Abbildung ist eine der ersten Aufnahmen der menschlichen Hand, die von Röntgen aufgenommen wurde, zu sehen. (c) Raymond Damadian entwickelte den ersten MRI-Scanner, der in der Lage war, Aufnahmen des menschlichen Körpers anzufertigen. Die Abbildung zeigt den ersten MRI Scan eines Menschen aus dem Jahr 1977. Literaturquellen: [7, 10, 18, 22]

Veränderung von Leberläsionen beobachten [24, 25, 5]. Die Leber selbst kann ebenfalls vom primären Leberkarzinom (Hepatocellulare Carcinoma HCC) befallen werden. Hepatocellulare Carcinoma (HCC) ist nach Untersuchung von GloboCAN aus dem Jahr 2010 die sechsthäufigste Tumorerkrankung und die dritthäufigste Todesursache bei Tumorerkrankungen. HCC entsteht üblicherweise in chronisch geschädigten Lebern. Ursachen für die chronischen Leberschädigungen können von Viruserkrankungen der Leber, z.B. Hepatitis B, übermäßigem Alkoholmissbrauch oder krankhaften Leberverfettungen stammen. Im Krankheitsverlauf werden gesunde Leberzellen schrittweise zu HCC umgewandelt. Bei dieser molekularen Transformation werden auch makroskopische Veränderungen des Gewebes sichtbar. HCC führt zu einer höheren Zelldichte sowie einer Arterialisierung der Gefäßversorgung. Diese makroskopischen Veränderungen erlauben die Diagnose von HCC mittels nicht-invasiver Bildgebungsverfahren wie Computertomographie oder Magnetresonanztomographie [26, 5, 27, 28].

Abbildung 2 zeigt zwei typische Kontrastmittel-verstärkte CT-Aufnahmen der Leber mit Leberläsionen. Die Leberkarzinome lassen sich in der CT-Aufnahme mit Hilfe von Kontrastmittel durch einen niedrigeren Hounsfield Wert als den von gesunden Gewebe beobachten. Die Form, Größe, Anzahl und der Kontrast der Leberläsionen unterscheiden sich stark von Patient zu Patient und erschweren die automatische Mustererkennung. Weitere Strukturen innerhalb der Leber, wie Blutgefäße, Leberfalten oder Zysten können eine automatische Detektion und Segmentierung von Tumorgewebe behindern.

Unter der Therapie von primären Tumorerkankungen der Leber wie HCC werden die Leberläsionen über den zeitlichen Verlauf untersucht. Sofern sich Metastasen in der Leber gebildet haben beobachten Radiologen ebenfalls bei sekundären Karzinomen wie Prostata-, Brust-, Darm- und Pankreastumor die zeitliche Veränderung der Leberläsionen. Die Veränderung der Leberläsionen hinsichtlich der Größe, Anzahl und Textur unter Therapie geben dem Radiologen und Onkologen Rückschlüsse über das Ansprechen des Patienten auf seine Therapie und Medikation [24, 25].

Im klinischen Alltag hat sich das Response Evaluation Criteria in Solid Tumors (RECIST) Verfahren zur Untersuchung des Behandlungserfolg von Tumorerkrankungen durchgesetzt. Im RECIST Verfahren soll der Radiologe pro Organ die zwei größten Läsionen (Zielläsion) pro Organ detektieren und für diese zwei Läsionen (Zielläsion) den größten Durchmesser bestimmen. In einer Follow-Up-Untersuchung soll der Radiologe das Prozedere wiederholen. Der Vergleich zur Erstuntersuchung bestimmt den Behandlungserfolg. Falls der Durchmesser der Zielläsionen um mehr als 30% gefallen ist spricht man von einer partiellen Remission/Rückbildung. Ist der Durchmesser der Zielläsion um mehr als 20% gestiegen wird von einer Krankheitsprogession gesprochen. Wenn die Durchmesser der Zielläsionen keine Veränderung aufweisen wird der Krankheitsverlauf als stabil angesehen [29, 24, 25].

Durch die Entwicklung neuer Algorithmen zur automatischen Segmentierung von Organen und Läsionen wird das RECIST-Verfahren von mehr und mehr Radiologen kritisch gesehen. Da im RECIST-Verfahren nur zwei Zielläsionen und von diesen nur die größten Durchmesser zur Bewertung des Behandlungserfolg berücksichtigt werden, erhoffen sich Radiologen und Onkologen durch eine vollständige Volumetrierung aller Läsionen eine genauere Bestimmung des Behandlungserfolgs und somit bessere Therapien. Anschaulich lässt sich die Kritik an RECIST in Abbildung 2 ableiten. Nach RECIST würden nur jeweils zwei der sieben Läsionen für die Therapiebewertung in Betracht gezogen. Rothe et al. (2013) haben in ihrer Studie bereits den Vergleich zwischen vollständiger Volumetrierung und RECIST zur Bestimmung des Behandlungserfolgs gezogen. Sie kamen zu dem Ergebnis, dass die vielversprechenden Ergebnisse der vollständigen Volumetrierung den Anstoß liefern sollten für neue Kriterien zur Bewertung von Tumorerkrankungen basierend auf vollständiger Volumetrierung. Gründe

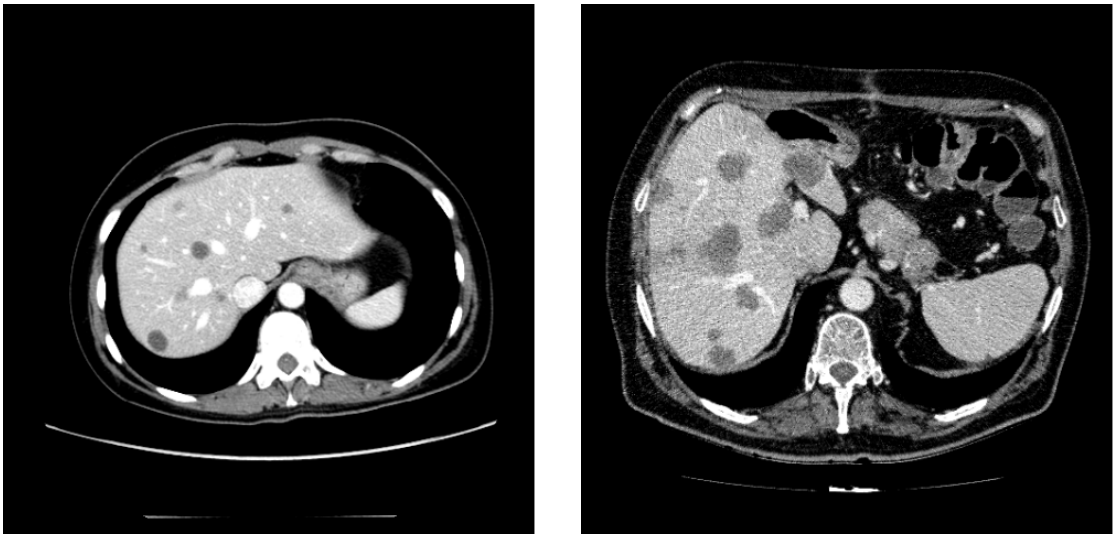


Abbildung 2: Kontrastmittel verstärkte Computertomographieaufnahmen der Leber und Leberläsionen. Die Form, Größe, Anzahl und der Kontrast der Leberläsionen unterscheidet sich in beiden Aufnahmen. Die hohe Heterogenität der Leberläsionen erschwert die automatische Detektion und Segmentierung. Literaturquelle: [5]

hierfür seien die geringere Subjektivität der vollständigen Volumetrierung im Vergleich zu RECIST [29, 30].

2 Medizinische Bildanalyse

2.1 Geschichtliche Entwicklung der medizinischen Bildanalyse

Die medizinische Bildanalyse (engl. Medical Image Analysis) ist aus der klassischen Bildverarbeitung und -analyse (engl. Computer Vision) entstanden. Zu Beginn wurden neue Methoden und Analysetechniken in Rahmen von Workshops an den etablierten Computer Vision Konferenzen wie Computer Vision Pattern Recognition (CVPR) und International Conference for Computer Vision (ICCV) diskutiert und publiziert. Seit 1998 findet jährlich die Konferenz Medical Image Computing and Computer Aided Interventions (MICCAI) mit mittlerweile über 1000 Forschern und Ärzten aus dem Bereich der medizinischen Bildanalyse statt [31, 32].

Die ersten Ideen zur Nutzung des Computers zur Analyse von medizinischen Bilddaten stammen aus Mitte der 1950er Jahre [31, 33]. In den 1960er und 1970er Jahren arbeiteten die ersten Wissenschaftler daran, medizinische Bilddaten zu digitalisieren und computergestützt auszuwerten. In den Anfängen hatte man sich der Detektion und Klassifizierung von Auffälligkeiten und Ungewöhnlichem gewidmet [33, 34, 35, 36, 37]. Die damals vorherrschende geringe Rechenleistung von Computern und die Problemen bei der Digitalisierung von medizinischen Aufnahmen, wie z.B.

Röngtenaufnahmen und Radiogrammen, erschwerten die Entwicklungen in dieser Zeit. Mit Beginn der 1980er Jahre entstanden die ersten digitalisierten Bildaufnahmeverfahren und neuartige Technologien wie z.B. Magnetresonanztomographie und Computertomographie. Diese Verfahren waren mit ihren aufwendigen Datenrekonstruktionsverfahren auf computergestützte Signalverarbeitung angewiesen [21, 38, 33, 14]. Mit dem Fortschritt dieser Techniken konnte sich das Gebiet stark weiterentwickeln.

In der medizinischen Bildanalyse werden verschiedene Problemstellungen behandelt. Die Klassifizierung und Segmentierung von medizinischen Bilddaten, die auch als punktweise Klassifizierung betrachtet werden kann, nehmen dabei eine zentrale Rolle ein. Die Entwicklung von computergestützter Diagnosesoftware (engl. computer-aided diagnosis CAD) ist eines der Hauptziele der medizinischen Bildanalyse. Um einen Radiologen bei seiner Diagnose unterstützen zu können, muss ein Algorithmus in der Lage sein, das zu untersuchende Organ zu lokalisieren (Segmentierung) und basierend auf der Lokalisierung und Segmentierung es hinsichtlich der medizinischen Fragestellung zu analysieren und zu bewerten (Klassifizierung). Ein typischer Anwendungsfall ist die Detektion und Klassifizierung von Tumorgewebe.

Neue Forschungsarbeiten, basierend auf künstlichen neuronalen Netzwerken, sind in der Lage Krankheiten mit gleicher oder höherer Genauigkeit klassifizieren zu können als erfahrene Ärzte [39, 40]. Der Fortschritt in diesem Gebiet wird unter anderem durch öffentliche Wettbewerbe angetrieben. Bei diesen Wettbewerben (engl. Challenges) formuliert der Organisator eine Problemstellung und stellt einen Datensatz zur Lösung der Problemstellung zur Verfügung. Die Auswertung des Wettbewerbs hinsichtlich der Problemstellung erfolgt nach objektiven Regeln und Metriken, welche eine Vergleichbarkeit und Bewertung von Methoden hinsichtlich ihrer Performance ermöglichen. Die Standardisierung der Datensätze und Performanzmetriken erlauben eine kontinuierliche Weiterentwicklung von Algorithmen. Kritiker dieser Wettbewerbe weisen auf die Spezialisierung der Methoden hinsichtlich des Datensatzes hin, die zur Folge hat, dass Algorithmen nicht in der Lage sind die eigentliche Problemstellung auf anderen Datensätze zu lösen.

Wichtige Wettbewerbe im Gebiet der medizinischen Bildanalyse waren die Grandchallenges zu Lebersegmentierung 2007 [23] und Lebertumorsegmentierung 2008 [41]. Diese beiden Wettbewerbe haben den Startschuss zur Entwicklung von neuartigen Segmentierungsalgorithmen geliefert. Über 500 Zitierungen (Stand 25.05.2017) hat die Zusammenfassung über die Lebersegmentierung Challenge aus 2007 von Heimann et al (2009) erhalten. Heimann et al. (2009) und Deng et al. (2008) haben in ihren Wettbewerben einen Datensatz mit 20 kontrastverstärkten CT-Volumen des Abdomens mit Leber- bzw. Lebertumorsegmentierung bereitgestellt [23, 41]. Neben der Leber- und Lebertumorsegmentierung Challenge von Heimann et al. (2009) und Deng et al. (2008) hat die Brain Tumor Segmentation Challenge (Brats) von Menze et al. (2015) einen großen Einfluss auf das Forschungsgebiet in jüngerer Zeit gehabt [42]. Neu entwickelte Algorithmen werden meist zur objektiven Performanzbewertung auf Challen-

gedatensätze angewandt. Im Bereich der Computer Vision haben zwei Wettbewerbe den wissenschaftlichen Fortschritt in diesem Bereich begünstigt. Der Klassifizierungswettbewerb IMAGENET wurde im Jahr 2012 von Krizhevsky et al. (2012) mit ihrem Algorithmus, basierend auf Convolutional Neural Networks, gewonnen [43, 44]. Dieser Sieg gilt als Geburtsstunde der neuen Forschungswelle im Bereich der künstlichen neuronalen Netzwerke. Die Entwicklung von Segmentierungsalgorithmen basierend auf künstlichen neuronalen Netzwerken wurde durch den Segmentierungswettbewerb PascalVOC befördert [45, 46].

2.2 Computergestützte Segmentierung in der Medizin

In der Vergangenheit wurden zahlreiche Methoden entwickelt, um die Leber und die Tumore innerhalb der Leber zu segmentieren. Die entwickelten Algorithmen lassen sich in automatische und semi-automatische sowie überwachte (engl. supervised) und unüberwachte (engl. unsupervised) Methoden untergliedern. Von semi-automatischen oder auch interaktiven Methoden spricht man, wenn der Algorithmus oder die Methode Interaktion von einer geschulten Person, z.B. Radiologen oder Onkologen, voraussetzt. In diesem Szenario würde beispielshalber der Arzt das Objekt, das es zu segmentieren gilt, markieren. Der Algorithmus würde dann basierend auf der Markierung des Objektes die Segmentierung erstellen. Die Auswertung von großen medizinischen Studien, wie z.B. nationale Kohorten, ist mit semi-automatischen Methoden kaum durchführbar. Personalkosten und geringere Objektivität im Vergleich zu automatischen Segmentierungsalgorithmen sind hierfür die Hauptgründe [1, 47, 48].

Der Unterschied zwischen überwacht/supervised und unüberwacht/unsupervised liegt in der Verwendung einer Grundwahrheit (engl. ground truth) zur Lösung des Problems. In einem überwachten Lernszenario würde ein Algorithmus zur Lösung des Problems in einer Lernphase mit Experten-annotierten Beispielen (Grundwahrheit) des zu lernenden Problems konfrontiert werden. Der Algorithmus erkennt Muster in der Grundwahrheit, die es ihm ermöglichen sein erlerntes Wissen auf unbekannte Beispiele anzuwenden und zu generalisieren. Zur Segmentierung der Leber würde ein überwachtes/supervised Lernverfahren darin bestehen, dass zum Trainieren des Algorithmus neben dem medizinischen Bildvolumen ebenfalls ein Volumen mit einer manuellen Segmentierung der Leber vorhanden ist. Der Algorithmus ist nach einer Trainingsphase in der Lage, selbständig in einem unbekannten medizinischen Bildvolumen eine Segmentierung der Leber zu erzeugen. Ein unüberwachtes/unsupervised Lernverfahren kann ohne Bespiele/Grundwahrheit und somit ohne Training die Problemstellung lösen [49, 50].

Methoden zur Segmentierung von Leber und Lebertumor lassen sich in folgende Klassen gruppieren [23, 48, 51]:

- Intensitätsbasierte Methoden (engl. grey level methods) [52, 53]
- Flächenbasierte Methoden (engl. region based methods) [54, 55, 56, 57]

- Kontur- und Formbasierte Methoden (engl. contour and shape based methods) [58, 59, 60, 61]
- Graphenbasierte Methoden (engl. graph-cut based methods) [62, 63, 64]
- Maschinelles-Lernen-basierte Methoden (engl. machine learning based methods) [65, 66]

Die Probleme der Leber- und Lebertumorsegmentierung haben größere Popularität durch die von Heimann et al. (2009) und Deng et al. (2008) organisierten Wettbewerbe im Rahmen der MICCAI Konferenz 2007 und 2008 [5, 1, 27, 23, 41] erlangt. Bei intensitätsbasierten Methoden werden die Intensitäten des Bildvolumens verwendet, um eine Segmentierung zu ermöglichen. In der Computertomographie entsprechen die Intensitäten den physikalischen Abschwächungskoeffizienten (Hounsfield Einheit) und liegen für die Leber im Bereich von 65 ± 5 HU [25]. Durch globales oder adaptives Thresholding um den Hounsfield Bereich der Leber kann eine Segmentierung der Leber erfolgen [53].

Bei den flächenbasierten oder region growing Verfahren handelt es sich um ein rekursives Segmentierungsverfahren, das einen Startpunkt (engl. seed) im zu segmentierenden Objekt benötigt. Dieser Startpunkt kann entweder manuell (interaktiv) oder automatisch bestimmt werden. Ausgehend vom Startpunkt untersucht der Algorithmus alle direkt benachbarten Bildpunkte der Startfläche und bestimmt deren Ähnlichkeit zur bereits segmentierten Startfläche. Ist ein Bildpunkt ähnlich zum aktuellen Stand der Segmentierung wird er hinzugefügt, bei Unterschieden geschieht dies nicht. Die Ähnlichkeitbewertung des Bildpunkte kann über Intensitäts-, Form- oder Texturdiskriptoren erfolgen. Diese Prozessschritte werden für alle Bildpunkte wiederholt [57, 23, 55].

Kontur- und Formbasierte Methoden verwenden die Eigenschaft, dass die zu segmentierenden Objekte ähnliche Konturen oder Formen besitzen. Insbesondere stellen diese Algorithmen die Annahme auf, dass das zu segmentierende Objekt eine Repräsentation der mittleren Objektform und der wichtigsten Objektformvariationen darstellt. Zu Beginn dieses Segmentierungsverfahrens werden mit Hilfe der Grundwahrheit die mittlere Objektform und die wichtigsten Objektformvariationen ermittelt. Diese werden dann in das zu segmentierende Objekt gelegt und die am besten passende Objektvariation gesucht, die dann die finale Segmentierung darstellt [58, 59, 60, 61].

Graphenbasierte Segmentierungsverfahren interpretieren das zugrundeliegende medizinische Volumen als einen verbundenen Graphen. Diese Algorithmenklasse benötigt einen Startpunkt im Objekt. Die zugrundeliegende Annahme dieser Algorithmen liegt darin, dass Objekte der gleichen Klassen ähnliche Eigenschaften besitzen und Objekte zusammenhängend (koherent) sind. Diese Eigenschaften werden in einem Energie-Minimierungsproblem modelliert, für welches ein graphenbasiertes Lösungsverfahren existiert. In diesem Lösungsverfahren stellt die bestmöglichste Segmentierung des Objektes die Minimierung der Energie, welche impliziert, dass Objekte der gleichen Klas-

sen ähnliche Eigenschaften besitzen und Objekte kohärent sind, dar [62, 63, 64].

Maschinelles-Lernen-basierte Methoden gehören zu der Klasse der überwachten/supervised Lernmethoden. Bei diesem zweistufigen Verfahren werden in einem ersten Schritt aus dem medizinischen Volumen Bilddeskriptoren (engl. features) extrahiert. Bilddeskriptoren stellen eine Repräsentation der Daten dar, die zur Lösung der Problemstellung hilfreich sein kann. Beispiele hierfür sind in folgenden Arbeiten zu finden [67, 68, 69, 70, 65, 66]. Die Auswahl und Entwicklung dieser Bilddeskriptoren für das entsprechende Segmentierungsproblem ist von besonderer Bedeutung. Im zweiten und letzten Schritt wird anhand der Bilddeskriptoren ein Klassifizierungsalgorithmus verwendet, um bei gegebenen Merkmalsausprägungen der Bilddeskriptoren auf die Problemlösung zu schließen. Typische Klassifizierungsalgorithmen sind die Logistische Regression, Support Vector Machine, künstliche Neuronale Netzwerke oder Random Forest Algorithmen [49, 67, 65].

Akkurate Leber- und Lebertumorsegmentierungen stellen die Basis für die quantitative Untersuchung von Tumorgewebe dar. Eine vollständige Segmentierung der Leber- und des Lebertumors kann prinzipiell von Radiologen durchgeführt werden, findet aber wegen Kosten- und Zeitgründen im klinischen Alltag nicht statt. Bei der retrospektiven Analyse von medizinischen Studien in der klinischen Forschung sprengen manuelle Segmentierungen meist den Forschungsetat, obwohl die bereits erhobenen Bilddaten und klinischen Daten großes Potential zur Analyse böten. Die im Rahmen dieser Arbeit entwickelte Methode zur automatischen Segmentierung von Leber und Lebertumorgewebe ist, wie bereits in Christ et al. (2016) gezeigt, in der Lage große medizinische Studien, wie z.B. Fire 3 Studie von Heinemann et al. (2014) mit über 3000 CT-Aufnahmen, zu segmentieren. [47, 1, 5, 2].

2.3 Überlebensvorhersage in medizinischen Bilddaten

Ein aktueller Trend in der klinischen Forschung ist die Suche nach quantitativen Biomarkern in radiologischen Bilddaten. Ein quantitativer Biomarker ist in der Lage einen Krankheitszustand zu beschreiben und somit eine Diagnose zu ermöglichen. Er kann dem Radiologen oder Onkologen Rückschlüsse über den aktuellen Krankheitszustand, z.B. Wirksamkeit einer Therapie, und den zukünftigen Zustand, z.B. Heilungschance, liefern. Anders als bei histologischen Untersuchungen, bei denen eine Gewebeprobe vom Patienten entnommen und untersucht werden muss, können quantitative Biomarker Rückschlüsse liefern ohne den Patienten einer Operation zu unterziehen [25, 2].

Aktuelle Forschung von Heid et al. (2017) konnte für den Pankreastumor einen quantitativen Biomarker finden [71]. Bei dem quantitativen Biomarker handelt es sich um die Verteilung des Apparent Diffusion Coefficent (ADC) in einer diffusions-gewichteten MR-Sequenz. Heid et al. (2017) konnten signifikant nachweisen, dass niedrige Tumorzellularität, die sich in der ADC-Sequenz durch hohe Werte ausweisen, auf verhältnismäßig langes Überleben der Patienten deutet. Solche Rückschlüsse und Analysen

erlauben eine Einteilung von Patienten (Stratifizierung) in Gruppen mit hohem Risiko und niedrigem Risiko und schließlich eine personalisierte Behandlung von Hochrisikopatientengruppen.

Nach aktuellem Stand der Forschung werden quantitative Biomarker in aufwendigen medizinischen Studien gesucht, indem großzahlig Bilddeskriptoren auf die medizinischen Bilddaten angewendet und getestet werden. Diese Bilddeskriptoren stammen meistens aus dem Bereich der Computer Vision und wurden ursprünglich für die Analyse von Bild- und Videodaten konzipiert. Diese explorativen Suchverfahren sind zeit- und kostenintensiv und führen nicht zwangsläufig zu einem Erfolg. In der Vergangenheit kamen folgende Bilddeskriptoren zum Einsatz:

- Histogrammbasierte Bilddeskriptoren [72]
- Texturbasierte Bilddeskriptoren [73, 74]
- Ensemble aus Histogramm- und Textur-Bilddeskriptoren [75]

Unter histogrammbasierten Bilddeskriptoren versteht man die Extraktion von statistischen Größen, wie Mittelwert, Varianz, Schiefheit (Skewness), Wölbung (Kurtosis) und Quantile einer Verteilung. Die Verteilung stellt im Zusammenhang mit der Überlebensvorhersage meistens die Verteilung der Bildintensitätswerte in einer Region-of-Interest (ROI), z.B. einer Tumorregion, dar. Das Ensemble an histogrammbasierten Bilddeskriptoren kann komplexere Konzepte wie Textur beschreiben [2, 71, 72].

Basierend auf den Arbeiten von Haralick (1979) zur Beschreibung von Texturen in Bilddaten versuchen diese Ansätze die Heterogenität von Flächen zu beschreiben. Im Gegensatz zu den histogramm basierten Bilddeskriptoren aus dem vorherigen Abschnitt wird die Textur nicht direkt aus den Intensitätswerten abgeleitet, sondern aus deren räumlicher Verteilung. In einem zweistufigen Verfahren werden zuerst die sogenannten Gray Level Cooccurrence Matrix (GLCM) bestimmt. Diese Matrix gibt an, wie oft ein bestimmtes Intensitätsniveau (z.B. Bereich 3 50-100 HU) eines Pixels neben einem Pixel mit gleichem (z.B. Bereich 3 50-100 HU) oder anderem Intensitätsniveau (z.B. Bereich 5 150-200 HU) existiert. Die Homogenität einer Textur lässt sich durch eine Verteilung der GLCM Matrix Einträge hin zu einer Diagonalmatrix (Alle Pixel, die räumliche Nachbarn sind liegen auch im gleichen Intensitätsniveau) beschreiben. Neben der Homogenität lassen sich auch Entropie und Energie der GLCM Matrix als Maße für Textureigenschaften berechnen [76, 27, 73, 74].

Die jüngsten Arbeiten auf diesem Gebiet von Zhao et al. (2016) verwendeten ein Ensemble aus verschiedenen Bilddeskriptoren. In ihrer Arbeiten konnten sie den prädikativen Wert ihrer Ensemble-Bilddeskriptoren zur Klassifizierung der Tumoragressivität bei HCC Lebertumor nachweisen. In ihrer Studie untersuchten sie 46 arterielle MR-Aufnahmen von Patienten mit HCC. Ihr Ensemble aus Histogramm und Haralick Bilddeskriptoren war in der Lage die Tumoragressivität mit einer Sensitivität von 76% bei

einer Spezifität von 100% [75] zu bestimmen.

Durch die Entwicklung von Convolutional Neural Networks (CNN) und schnellen Implementierungen von dreidimensionalen Faltungen können spezifische und auf die Problematik angepasste Bilddeskriptoren erlernt werden [2, 71]. Im Bereich der Computer Vision, wie in Kapitel 3.1 noch näher beschrieben wird, konnten Algorithmen basierend auf Convolutional Neural Networks für Aufgaben und Problemstellungen, die ein genaues Verständnis der Textur, Struktur und Semantik einer Bildkomposition vorraussetzen, große Erfolge feiern. Im Bereich der Bildklassifizierung von nichtmedizinischen Bildern können moderne Algorithmen, wie z.B. He et al. (2015), die menschliche Leistungsfähigkeit erreichen oder übertreffen [77, 43, 44]. Diese Leistungsfähigkeit kann nicht von Algorithmen und Bilddeskriptoren, die momentan noch zur Überlebensvorhersage in der medizinsichen Bildanalyse verwendet werden, erreicht werden. Folglich verspricht die Anwendungen und Adaption von Convolutional Neural Networks (CNN) im Bereich der Überlebensvorhersage zur Generierung von quantitativen Biomarkern großes Potential, wie im Rahmen dieser Arbeit und von Christ, Ettlinger und Kaissis et al. (2017) gezeigt wurde [2].

3 Künstliche neuronale Netzwerke

3.1 Geschichtliche Entwicklung von neuronalen Netzwerken

In den 1940er Jahren verwendeten die Wissenschaflter Mulloch und Pitts (1943) aktuelle Erkenntnisse aus der Nerven- und Hirnforschung, um das erste mathematische Modell eines Neurons aufzustellen [78, 50]. Die Motivation hinter der Modellierung des Gehirns rührt daher, dass das Gehirn die Schlüsselstelle der Intelligenz bei Tieren und Menschen ist und eine künstliche Nachbildung des Gehirns mit mathematischen Modellen zur Entwicklung von intelligenten Systemen führen kann. Neben dieser praktisch getriebenen Motivation, würde ein funktionierendes mathematisches Modell Aufschluss über grundlegende Prinzipien und Mechanismen des Gehirns liefern, die Erkenntnisse auf unsere Psyche und unser Verhalten geben könnten.

Die weitere Entwicklung trieb der Psychologe Frank Rosenblatt 1958 durch die Modellierung des Perzeptrons voran [79]. Das Perzeptron wird als das erstes künstliche neuronale Netzwerk gesehen. Sein mathematisches Modell wurde später auch in der Mark I Perzeptron Maschine umgesetzt. Sein Modell war in der Lage einen gegebenen Input in zwei Klassen zu klassifizieren (Binärklassifizierung) [79, 50, 49, 80]. Nachfolgend wird die mathematische Herleitung des Perzeptrons nach [49, 80] verwendet. Für ein Eingangssignal oder einen -vektor x der Länge n lässt sich das Perzeptron wie folgt definieren:

$$y(x) = \sigma\left(\sum_{i=1}^{n} x_i \cdot w_i + b\right) \tag{1}$$

Die Funktion $\sigma(\cdot)$ wird als Aktivierungsfunktion bezeichnet und ist typischerweise nicht-linear. Im Fall des Perzeptrons wird $\sigma(\cdot)$ als Stufenfunktion formuliert:

$$\sigma(a) = \begin{cases} +1, & \text{falls } a \geq 0 \\ -1, & \text{sonst} \end{cases} \tag{2}$$

Die Fehlerfunktion des Perzeptrons zum Zeitpunkt t wird wie folgt definiert:

$$E(t) = \frac{1}{s} \sum_{j=1}^{s} |d_j - y_j(t)| \tag{3}$$

Dabei bezeichnet $d_j \in D$ die richtige Klasse/Sollwert für den Eingangsvektor x_j. Das Perzeptron wird folgendermaßen mit Hilfe der Trainingsdaten D und X der Mächtigkeit s trainiert. Für alle s Paare $d_j \in D$ und $x_j \in X$ wird zunächst Gleichung 1 ausgewertet. Im Anschluss wird der aktuelle Fehler für jedes Paar d_j und x_j mit Hilfe von Gleichung 3 berechnet. Für die nächste Iteration $t + 1$ lassen sich die aktuellen Gewichte $w_i(t)$ mit der folgenden Gleichung und der Lernrate η korrigieren.

$$w_i(t + 1) = w_i(t) + \eta(d_j - y_j(t))x_{i,j} \tag{4}$$

Diese mathematische Formulierung des Perzeptrons stellt die Basis für künstliche neuronale Netzwerke dar. Abbildung 3 zeigt ein Schaubild eines Perzeptrons. Das Perzeptron hatte einen zentralen Nachteil, der weitere Entwicklungen nötig machte. Da es sich bei dem Perzeptron um einen Linearenklassifikator handelt, kann nur eine gewisse Klasse an Problemen gelöst werden. Diese Limitierung wurde von Minkey et al. (1988) durch die Entdeckung, dass das Perzeptron die XOR-Funktion nicht modellieren konnte, aufgezeigt [81, 50, 80]. Diese Problematik konnte von Paul Werbos im Jahr 1974 gelöst werden. Er erfand die Methode der Backprogation, die es ermöglichte mehrschichtige Neuronale Netzwerke (engl. multi-layer perceptrons MLP) zu trainieren. Mit einem zweischichtigen MLP ist es möglich die XOR-Funktion zu modellieren. Insbesondere die Entwicklung von tiefen neuronalen Netzwerken (engl. Deep Learning) und Convolutional Neural Networks wurde erst durch die Entwicklung des Backpropagation Algorithmus möglich. Abbildung 4 zeigt ein Schaubild eines Multi-Layer Perzeptrons. Der Ausgang eines Perzeptrons ist mit Eingang eines neues Perzeptrons in der nächsten Schicht verbunden. Der Mathematiker Kurt Hornik konnte 1991 in seiner Arbeit beweisen, dass sich mit einem Multi-Layer Perzeptron kontinuierliche Funktionen auf einer kompakten Teilmenge des $\mathbb{R}^n$ approximieren lassen. Diese Entdeckung wird als Universal Approximation Theorem bezeichnet [82, 50, 49, 80, 79, 81, 83].

3.2 Convolutional Neural Networks

Convolutional Neural Networks sind eine spezielle Klasse von neuronalen Netzwerken. Die Besonderheit im Vergleich zu MLPs ist die Verwendung der Faltungsoperation (engl. Convolution) anstelle der gewichteten Summe in Gleichung 1 [50]. Die folgende

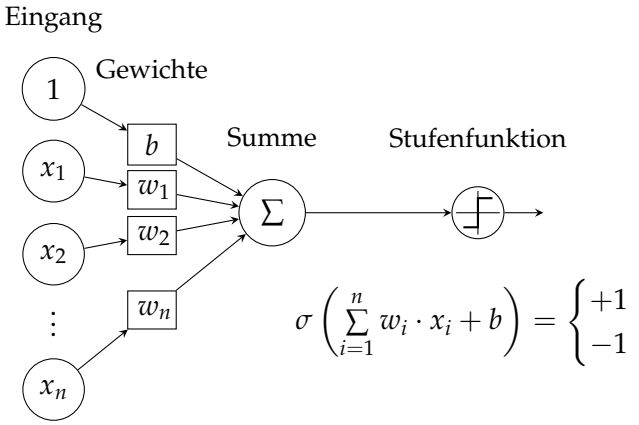


Abbildung 3: Schaubild des Perzeptrons entwickelt von Frank Rosenblatt im Jahr 1958. Das Perzeptron ist ein linearer Binärklassifikator und kann entscheiden, ob ein Eingangssignal X einer Klasse zugehörig ist oder nicht. Ein n-dimensionales Eingangssignal wird über eine gewichtete Summe komprimiert. Eine Aktivierungs- oder Stufenfunktion $\sigma(\cdot)$ legt fest, ob es sich bei der gewichteten Summe > 0 um Klasse 1 oder < 0 um Klasse -1 handelt [79].

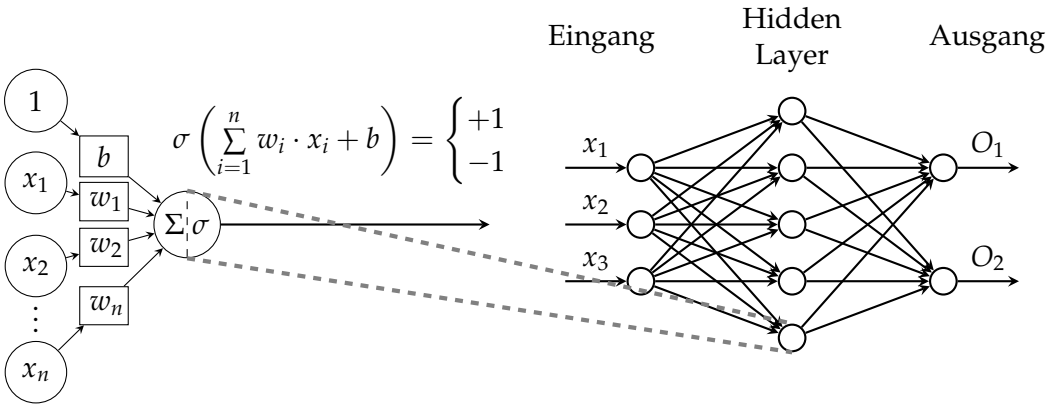


Abbildung 4: Schaubild des Multilayer Perzeptrons MLP. Das Multilayer Perzeptron besteht aus mehreren Schichten von Perzeptronen aus Abbildung 3. Die Ausgänge des Perzeptrons der Eingangsschicht sind mit den Eingängen der verdeckten Schicht (engl. hidden layer) verbunden. Die Ausgänge der verdeckten Schicht sind schließlich mit den Eingängen der Ausgangsschicht verbunden. Das Training des Multilayer Perzeptrons war erst durch die Entdeckung des Backpropagation Algorithmus durch Paul Werbos 1974 möglich [82].

mathematische Operation zwischen den diskreten Vektoren x und w wird als Faltung s bezeichnet:

$$s(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(a)w(t-a) \tag{5}$$

Im Spezialfall von zweidimensionalen Bilddaten I lässt sich die zweidimensionale Faltung wie folgt formulieren:

$$S(i,j) = (I * K)(i,j) = \sum_{m}\sum_{n} I(m,n)K(i-m, j-n) \tag{6}$$

$S(i,j)$ wird als Featuremap und K als zweidimensionaler Faltungsfilter oder Kernel bezeichnet. Die Einführung der Faltung anstelle der gewichteten Summe bzw. der Matrixmultiplikation bietet drei Vorteile. Ein typisches Problem von klassischen neuronalen Netzwerken ist die große Anzahl an Neuronen-Verbindungen in tiefen neuronalen Netzwerken oder bei der Anwendung von zweidimensionale Bilddaten. Würde man versuchen, mit Hilfe eines Perzeptrons eine medizinische CT Aufnahme (512x512 Pixel) zu klassifizieren, z.B. zu entscheiden ob die Leber auf den Bild zu sehen ist oder nicht, würde das Perzeptron bereits $(512 \cdot 512) + 1 = 262.145$ freie Parameter in Form der Gewichte w aufweisen. Die Anzahl der freien Parameter und somit auch die Komplexität des Systems lassen sich deutlich durch die Einführung der Faltungsoperation und Verwendung des Faltungsfilters K, der eine deutlich kleinere Größe besitzt als das Bild I, verringern. Der zweite Vorteil bezieht sich auf die Mehrfachverwendung der Gewichte w (engl. parameter sharing). Im Unterschied zum klassischen Perzeptron werden die im Faltungsfilter enthaltenen Gewichte auf das komplette Bild I angewendet und es existiert nicht für jedes Pixel ein einzelnes Gewicht. Dies führt zu einer höheren Performanz bei der Berechnung und einer höheren statistischen Aussagekraft und folglich einer größeren Klassifizierungsgenauigkeit. Eine weitere besondere Eigenschaft der Faltung ist die Translationsinvarianz. Die Faltungsoperation ist gegen Translationen/Verschiebungen des Bildes invariant. Diese Eigenschaft ist sehr hilfreich bei der Analyse von Bildern oder medizinischen Aufnahmen, da Objekte im Generellen unabhängig von ihrer momentanen Position im Bild erkannt werden können [49, 50, 82, 84].

Entwickelt wurde das Konzept der Convolutional Neural Networks von LeCun et al. (1989) zur Erkennung der menschlichen Handschrift [85]. Abbildung 5 zeigt den Aufbau des Convolutional Neural Networks von LeCun et al. (1989). Nach der Faltungsoperation werden die Featuremaps S mit einer nicht-linearen Aktivierungsfunktion aktiviert. Im Anschluss findet ein Sub-Sampling statt. Im Bereich der Convolutional Neural Networks sind die Max-Pooling und Average-Pooling Operationen verbreitet. Bei dem Max- oder Average Pooling wird das Ursprungsbild entweder durch Verwendung des Maximums oder des Durchschnittswertes aus einer lokalen Nachbarschaft des Pixels verkleinert. Das Subsampling reduziert die Anzahl an freien Parametern und hilft somit bei der Reduktion der Komplexität des Systems. Nach mehreren Faltungs- und Subsamplingblöcken sinken die Dimensionen der Featuremaps bei gleichzeitiger Erhöhung der Anzahl der Featuremaps immer weiter. Bei Convolutional Neural Net-

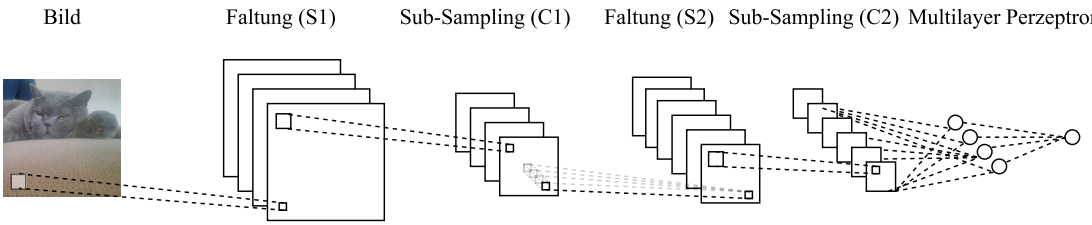


Abbildung 5: Erste Convolutional Neural Network Architektur LeNET von LeCun et al. (1989). Ein Eingangsbild wird mit Hilfe von trainierbaren Filterkernels gefaltet. Nach der Faltungsoperation werden die entstandenen Featuremaps in einer Subsampling Schicht komprimiert. Die komprimierten Featuremaps (C1) werden mit neuen trainierbaren Filterkernels erneut gefaltet und die entstandenen Featuremaps mit einer Subsampling Operation komprimiert. Die nach zwei Faltungs- und Subsamplingoperationen entstanden Featuremaps dienen als Eingang für ein Multilayer Perzeptron, welches die finale Klassifizierungsentscheidung trifft [89, 85].

works, die zur Klassifizierung von Objekten eingesetzt werden, bestehen die letzten Schichten des Netzwerkes aus einem Multi-Layer-Perzeptron. Das Multi-Layer Perzeptron schließt von der Repräsentation des Bildes, ausgedrückt in den Featuremaps der Schicht C2 in Abbildung 5, auf die finale Klasse [84, 50].

Neue Convolutional Neural Networks Architekturen, die zur erneuten Popularität dieser Methoden beigetragen haben, unterscheiden sich hauptsächlich in der Tiefe d.h. in der Anzahl der Schichten und der Operationen innerhalb der Schichten. Die Netzwerkarchitektur von Krizhevsky et al. (2012), die den Imagenet Wettbewerb 2012 gewann, besaß acht Schichten, die VGG Architektur von Simonyan et al. (2014) bereits 19 Schichten und die neusten Architekturen die ebenfalls Eingang in diese Arbeit fanden ResNet50 von He et al. (2016) und InceptionV3 von Szegedy et al (2016) über 50 Schichten. Die Performanz zur Erkennung von Objekten steigt mit größerer Netzwerktiefe, wird aber mit einer höherer Komplexität des Modells erkauft. Die Netzwerkarchitekturen von He und Szegedy können nur mit einer Vielzahl an Bildern und enormen Rechenkapazitäten, wie sie die beiden Arbeitgeber von He (Facebook) und Szegedy (Google) zur Verfügung haben, trainiert werden [84, 43, 86, 87, 88].

3.3 Fully Convolutional Neural Networks

Die ersten Ansätze zur Segmentierung mit Hilfe von Deep Learning Algorithmen versuchten Bildauschnitte mit Hilfe von Convolutional Neural Networks zu klassifizieren. Diese Methoden lassen sich wie folgt beschreiben. Aus einer CT-Aufnahme der Größe 512x512 Pixel werden kleine, sich überlappende Bildausschnitte extrahiert. Die extrahierten Bildausschnitte werden mit Hilfe eines CNN klassifiziert, wobei einem Bildausschnitt eine globale Klasse zugeordnet wird. Da sich die gezogenen Bildausschnitte überlappen, werden pro Bildpunkt mehrere Klassifizierungsvorhersagen gemittelt. Die Größe der Bildausschnitte ist aus diesem Grund von großer Bedeutung, da zum

einen bei einer zu großen Wahl kleine Objekte nicht erkannt werden können und zum anderen bei einer zu kleinen Wahl ein regionaler, semantischer Bildkontext nicht für die Klassifizierungsentscheidung im Bildausschnitt vorhanden ist. Wichtige Arbeiten in diesem Gebiet stammen von Wolf et al. (1994), Prasoon et al. (2013), Roth et al. (2014), Milletari et al. (2017) und Havaei et al. (2017) [90, 91, 92, 93, 94].

Long et al. (2014) begründeten mit ihrer Arbeit *Fully Convolutional Neural Networks for Semantic Segmentation* einen neuen Meilenstein im Bereich der Bildsegmentierung. Anders als die eben beschriebenen Ansätze, modifizierten Long et al. (2014) ihre Netzwerkarchitektur dahingehend, dass sie in der Lage waren eine Segmentierung in voller Auflösung zu erlernen. Sie haben dazu die letzten Schichten ihrer Netzwerkarchitektur, die auf der Arbeit von Krizhevsky et al. (2012) basiert und dort durch ein MLP dargestellt wird, durch Faltungs- und Upsamplingschichten ersetzt. Abbildung 6 zeigt ein Schaubild, das diese Modifikation erläutert. Des Weiteren führten sie verkürzte Verbindungen (engl. skip connections) von den vorderen Schichten zu den hinteren Schichten ein, um Informationen über die Lokalität der Objekte zu behalten. Man geht davon aus, dass in den ersten Schichten des Fully Convolutional Neural Networks (FCN) Information über die Lokalität der Objekte (Wo befindet sich das Objekt?) gespeichert wird. In den hinteren Schichten, bedingt durch sub-sampling und zahlreichen Faltungsoperationen sollen semantische Informationen über die Objekte (Um welches Objekt handelt es sich?) gespeichert sein. Beide Informationen über Lokalität und Semantik sind wichtig, um eine akkurate Segmentierung gewährleisten zu können. Aus diesem Grund führten sowohl Long et al. (2014), als auch Ronneberger et al. (2015) verkürzte Verbindungen von den ersten zu den letzten Schichten ein [45, 95].

Ronneberger et al. (2015) wandte das Konzept der Fully Convolutional Neural Networks erstmalig auf medizinische Daten an und verbesserte mehrere Aspekte an der Arbeit von Long et al. (2014). Ein inhärentes Problem bei medizinischen Bilddaten ist die ungleiche Verteilung der Klassen. Für die in dieser Arbeit untersuchten Datensätze ergeben sich folgende Zahlen. Ein typisches medizinisches Volumen eines Tumorpatienten besteht zu 93% aus Hintergrundpixeln, zu 7% aus Leberpixeln und zu 0.25% aus Tumorpixeln. Dieses Problem führt dazu, dass normale Lernverfahren (u.a. Long et al. (2014), wie im Rahmen dieser Arbeit gezeigt wurde [5]), die Klassen der Leber und des Tumors nur schwer detektieren können. Erst durch Ronneberger et al. (2015) Beitrag wurde die Tumordetektion mit Hilfe von Fully Convolutional Neural Networks möglich. Sein Beitrag bestand darin, dass er die Fehlerfunktion (engl. loss function) für ungleiche Klassenverteilungen anpasste. Des Weiteren konnte er das Konzept der verkürzten Verbindungen (engl. skip connections) noch weiter verbessern. Anders als Long et al. (2014), bei denen die verkürzten Verbindungen am Ende summiert werden, werden bei Ronneberger et al. (2015) die Featuremaps der vorherigen Schichten mit den hinteren Schichten konkateniert. Dies hat zur Folge, dass die Lokalitätsinformationen aus den früheren Schichten dem Netzwerk zu einem späteren Zeitpunkt direkt verfügbar sind und über weitere Faltungsoperationen einen direkteren Beitrag zur finalen Segmentierung beitragen können. Abbildung 7 zeigt die von

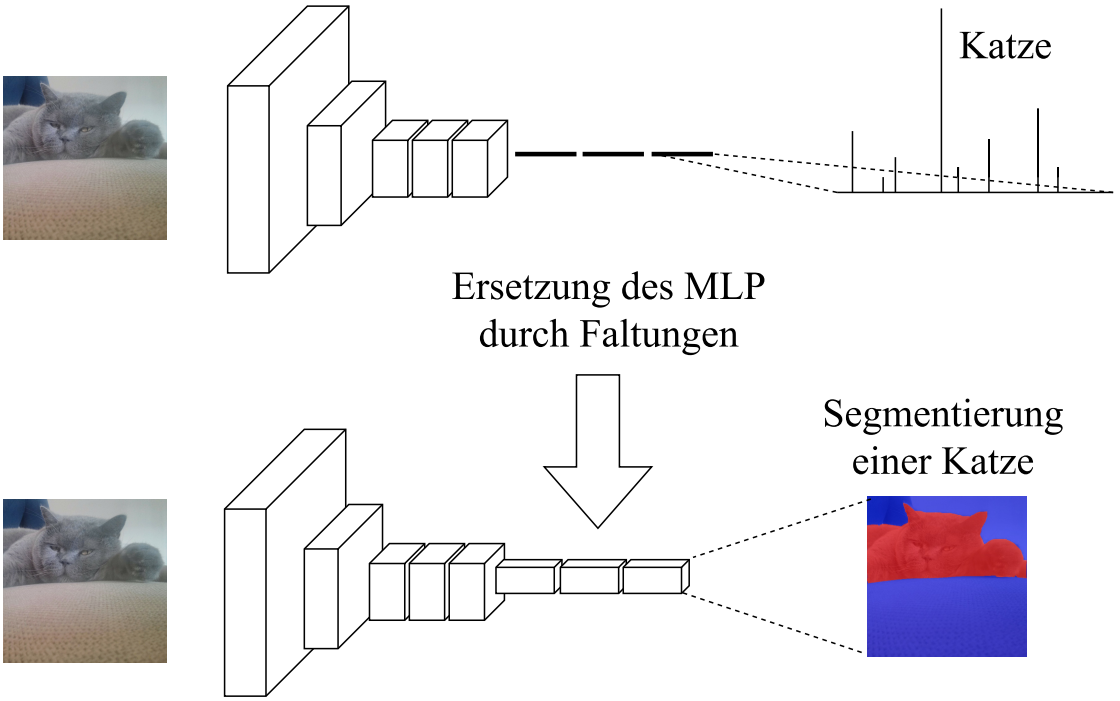


Abbildung 6: Schaubild zu Fully Convolutional Neural Networks nach Long et al. (2014). Long et al. (2014) wandelte die bis dato vorherrschenden Convolutional Neural Networks CNN, die in den letzten Schichten ein MLP zur Klassifizierung besitzen, dahingehend um, indem er die letzten Schichten des CNNs durch Faltungs- und Upsamplingschichten ersetzte [45].

Ronneberger et al. (2015) vorgestellte Fully Convolutional Netzwerkarchitektur UNet. Drozdal et al. (2016) haben in ihrer Arbeit den Einfluss von verkürzten Verbindungen ausführlich untersucht [96, 95, 45].

In dieser Arbeit wurde aufbauend auf den Werken von Long et al. (2014) und Ronneberger et al. (2015) das Konzept der Cascaded Fully Convolutional Neural Networks (CFCN) entwickelt. FCNs und insbesondere das UNet haben die semantische Bildsegmentierung vorangetrieben. In der medizinischen Bildverarbeitung erschweren zusätzliche Hürden die Leistungsfähigkeit von Segmentierungsverfahren. Ronneberger et al. (2015) konnten bereits Beiträge zur Überwindung des Problems der ungleichen Klassenverteilung leisten. Das Konzept Cascaded Fully Convolutional Neural Networks (CFCN) führt seine Beiträge weiter, in dem es eine kaskadierten Einsatz von Fully Convolutional Networks vorschlägt. In einem mehrstufigen Verfahren werden mehrere FCNs verwendet, um zuerst eine Region of Interest (ROI) zu segmentieren und in einem zweiten Schritt einen vergrößerten Ausschnitt der ROI hinsichtlich des gesuchten Objektes zu untersuchen. Im Rahmen dieser Arbeit konnte gezeigt werden, dass mit Hilfe dieses Konzeptes die Leber und der Lebertumor in CT und MRI-Aufnahmen segmentiert werden können. In CT konnte experimentell gezeigt werden, dass die Methode eine

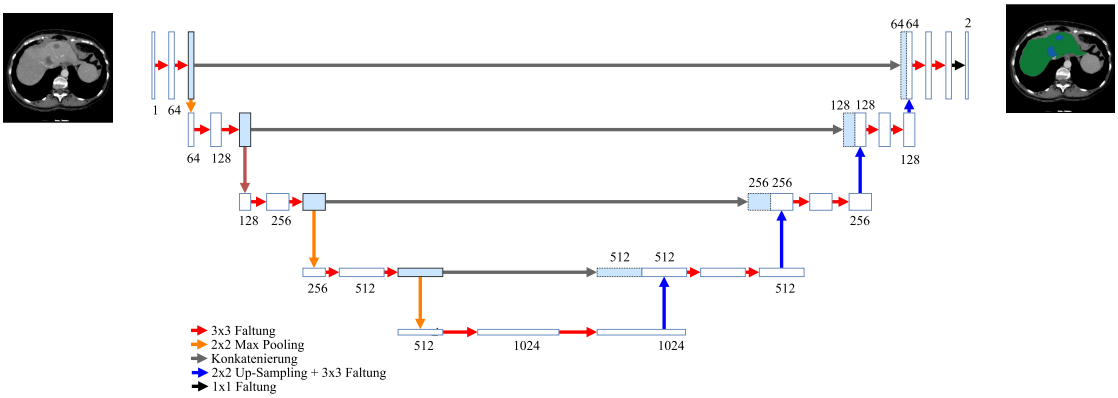


Abbildung 7: UNet Architektur nach Ronneberger et al. (2015). Das UNet besitzt 28 Schichten und verfügt über verkürzte Verbindungen (graue Pfeile). Mit Hilfe dieser verkürzten Verbindungen können Informationen über Lokalität von Objekten aus den frühen Schichten direkt zu späteren Schichten propagieren. Dies führt zur einer höheren Segmentierungsgenauigkeit und schnelleren Konvergenzzeiten [95, 96].

höhere Segmentierungsgenauigkeit als die UNet Architektur von Ronneberger et al. (2015) aufweist [95, 1, 2, 5].

Zusammenfassung und Diskussion der eigenen Forschungsarbeit

1 Segmentierung der Leber in CT und MRI

Im Rahmen dieser Arbeit wurde ein Verfahren entwickelt mit dessen Hilfe es möglich ist die Leber in medizinischen Aufnahmen der Computertomographie oder der Magnetresonanztomographie automatisch zu segmentieren. Automatische Segmentierungsmethoden haben noch nicht Einzug in den klinischen Alltag gehalten. Einige Arbeiten wie z.B. Chartrand et al. (2014) benötigen nach wie vor eine Interaktion des Menschen. Im Gegensatz zu Chartrand et al. (2014) ist die entwickelte Methode automatisch und benötigt keine Interaktion mit dem Menschen. Diese Arbeit setzt erstmalig Fully Convolutional Neural Networks ein, um die Leber in CT und MRI zu segmentieren. Tabelle 1 zeigt die quantitativen Ergebnisse der entwickelten Methode zur automatischen Segmentierung der Leber in CT [1]. Für medizinische MRT-Volumen schafft die entwickelte Methode einen DICE Score von 88% [2, 27, 5]. Die Ergebnisse der entwickelten Methode sind sehr vielversprechend und befinden sich nahe dem Bereich der Interrater Variabilität, die bei einer CT Lebersegmentierung ca. 95% DICE beträgt.

Verbesserungen an der Methode zur Lebersegmentierung könnten durch eine größere Anzahl an Daten im Falle der MR Lebersegmentierung erreicht werden. Die Daten- und Segmentierungsqualität im Bereich der Interrater Variabilität ist sehr entscheidend. Stammen Datensätze von unterschiedlichen Ratern können Feinheiten in den Segmentierungsprotokollen, wie der Ein- oder Ausschluss von Blutgefäßen innerhalb der Leber, zu Abweichungen von 5% DICE führen. Die Verwendung von dreidimensionalen Faltungen (engl. Convolution), wie vorgeschlagen in [97, 98], könnten weiteres Verbesserungspotential auf methodischer Ebene liefern, da ein dreidimensionaler Kontext berücksichtigt werden kann.

Die Beiträge zur Lebersegmentierung werden ausführlich in den veröffentlichen Artikeln *Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields* Seite 31 und *SurvivalNet: Predicting patient survival from diffusion weighted magnetic resonance images using cascaded fully convolutional and 3D convolutional neural networks* auf Seite 41 sowie dem unveröffentlichen Manuskript *Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks* im Anhang auf Seite 89 beschrieben.

Methode	VOE [%]	RVD [%]	ASD [mm]	MSD [mm]	DICE [%]
UNet wie in [95]	39	87	19,4	119	72,9
Eigene Methode: Cascaded UNet	**12,8**	**-3,3**	**2,3**	**46,7**	**93,1**
Eigene Methode: Cascaded UNet + 3D CRF	**10,7**	**-1,4**	**1,5**	**24,0**	**94,3**
Li et al. [99] (nur Leber)	9,2	−11,2	1,6	28,2	
Chartrand et al. [100] (semi-automatisch)	6,8	1,7	1,6	24	
Li et al. [101] (nur Leber)					94,5

Tabelle 1: Quantitative Segmentierungsergebnisse der Leber im CT Datensatz 3DIRCADb. Literaturquelle: [1]

2 Segmentierung von Lebertumor in CT und MRI

Die automatische Segmentierung von Lebertumor stellt für viele moderne Segmentierungsverfahren eine große Herausforderung dar. Die hohe Variabilität in Form, Kontrast und Größe sind nur einige Gründe für die Schwierigkeit dieses Lernproblems. Selbst erfahrene Radiologen nutzen für die Bestimmung von schwierigen Fällen weitere Informationsquellen, wie histologische Untersuchungen oder Aufnahmen mit anderen Bildgebungsmodalitäten. Diese zusätzlichen Informationensquellen stehen dem Segmentierungsalgorithmus nicht zur Verfügung. Im Laufe dieser Arbeit wurden die neusten methodischen Erkenntnisse aus dem Bereich der Segmentierung mit neuronalen Netzwerken angewandt und verbessert. Wie bereits für die Leber beschrieben, konnte im Rahmen dieser Arbeit zum ersten Mal die von Long et al. (2014) vorgeschlagene Methode der *Fully Convolutional Neural Networks* auf Lebertumore in CT und MRI angewandt werden. Eine ausführliche Beschreibung der Experimente mit der von Long et al. (2014) vorgestellten Architektur findet sich in dem unveröffentlichten Manuskript *Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks* im Anhang Seite 89. Mit der Arbeit von Ronneberger et al. (2015) konnten bereits starke Verbesserungen im Bereich der Lebertumorsegmentierung erzielt werden [1, 5].

Durch die Einführung der *Cascaded Fully Convolutional Networks*, einer Kaskade der von Ronneberger et al. (2015) eingeführten UNet Architektur, konnten weitere Verbesserungen erzielt werden. Die Verbesserungen können damit begründet werden, dass das Problem der Lebertumorsegmentierung in CT oder MRI durch Kenntnis der Leber als Region of Interest (ROI) vereinfacht werden kann. Wie bereits im vorherigen Kapitel beschrieben, erreicht die Methode der automatischen Lebersegmentierung bereits Segmentierungen im Bereich der Interrater Variabilität. Durch Einschränkung eines CT oder MRT-Volumens auf eine Lebermaske kann die prozentuelle Anzahl an Lebertu-

Datensatz	ASD [mm]	MSD [mm]	VOE [%]	RVD [%]	DICE [%]
CT LITS	13,7	63,5	53,0	2,1	58.0
MRI	13,1	111,4	46,3	37,2	69,4

Tabelle 2: Quantitative Segmentierungsergebnisse von Lebertumor im CT Datensatz LITS und MRT Datensatz. Literaturquelle: [5, 2, 27]

morpixel im Vergleich zu der Hintergrundklasse stark gesteigert werden. Der zweite Effekt der eine Verbesserung der Segmentierungsgenauigkeit hervorruft, ist die Spezialisierung der kaskadierten Netzwerke. In der ersten Stufe der Kaskade kann das UNet eine Spezialisierung erlernen, die es ihm ermöglicht die Leber in CT oder MRI vom Hintergrund zu unterscheiden. Die zweite Stufe der Kaskade spezialisiert sich auf die Unterscheidung von Lebertumor zu Lebergewebe und erreicht somit einen höheren Spezialisierungsgrad als bei einer direkten Segmentierung aus einem CT- oder MRI-Volumen.

Tabelle 2 zeigt die quantitativen Segmentierungsergebnisse von Lebertumor der entwickelten Methode in CT und MRI. Die Ergebnisse liegen deutlich unter den Werten für die Leber und weisen somit die Schwierigkeit des Segmentierungsproblems aus. Es bedarf noch weiterer Forschungsarbeit, um das Problem der Lebertumorsegmentierung vollständig zu lösen. Ansätze zur Verbesserung könnten in der Datenvorverarbeitung und der Gestaltung der Netzwerkarchitektur liegen. Die Variation von Lebertumoren ist so vielfältig, dass selbst neueste Methoden Schwierigkeiten besitzen alle Variationen zu erkennen. Durch Simulation und künstlicher Modellierung von Trainingsdaten mit Hilfe von Generative Adversarial Networks könnten den Variationen besser begegnet werden. Dreidimensionale Architekturen könnten ebenfalls Verbesserungen bringen, da auch für die Detektion und Segmentierung von Lebertumor der dreidimensionale Kontext eine Rolle spielen kann. Eigene Experimente mit den Methoden von Milletari et al. (2016) und Çiçek et al. (2016) brachten kein Ergebnis [50, 102, 1, 98, 97].

Eine ausführliche Beschreibung der Forschungsleistung im Bereich der Lebertumorsegmentierung in CT und MRI finden sich in den veröffentlichen Artikeln *Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields* Seite 31 und *SurvivalNet: Predicting patient survival from diffusion weighted magnetic resonance images using cascaded fully convolutional and 3D convolutional neural networks* auf Seite 41 sowie dem unveröffentlichen Manuskript *Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks* im Anhang auf Seite 89.

Ranking	Name	Institution	DICE	VOE	RVD	ASD	MSD
1	X. Han	Elekta Inc.	0,67	0,45	0,04	6,66	57,93
2	E. Vorontsov et al.	MILA	0,65	0,47	-0,21	7,12	51,96
2	G. Chlebus et al.	Fraunhofer	0,65	0,46	17,41	17,75	57,64
3	L. Bi et al.	Uni Sydney	0,64	0,46	1,9	21,19	72,8
4	**Eigene Methode**	**TU Munich**	**0,58**	**0,53**	**2,09**	**13,76**	**63,59**
4	C. Wang et al.	KTH Sweden	0,58	0,54	3,93	26,02	85,38
6	J. Lipkova et al.	TU Munich	0,48	0,63	103,31	32,32	105,82
7	J. Ma et al.	NJUST	0,47	0,65	-0,35	11,49	64,31
9	T. Konopczynski et al.	Uni Heidelberg	0,42	0,69	103,74	32,54	116,6
10	M. Beliver	ETH Zurich	0,41	0,69	3,6	36,29	130,46
11	K. Maninis	ETH Zurich	0,38	0,71	1,23	19,74	86,83
12	J. Qi et al.	UESTC	0,19	0,87	1985,2	40,61	95,63

Tabelle 3: Ergebnisse der Liver Tumor Segmentation Challenge IEEE ISBI Konferenz 2017 zur Lebertumorsegmentierung. Die Liver Tumor Segmentation wurde im Rahmen dieser Arbeit organisiert und durchgeführt. Die eigene Methode zur Lebertumorsegmentierung aus [5] erzielte den vierten Platz.

3 Liver Tumor Segmentation Challenge

Im Bereich der Leber- und Lebertumorsegmentierung wurden große Beiträge durch die Wettbewerbe von Heimann et a.l (2007) und Deng et al. (2008) geleistet [23, 41]. Im Rahmen dieser Arbeit wurden zwei Wettbewerbe im Rahmen der Konferenzen IEEE International Symposium of Biomedical Imaging 2017 (IEEE ISBI 2017) und Medical Image Computing and Computer Aided Interventions 2017 (MICCAI 2017) organisiert und veranstaltet. Im Rahmen dieser Tätigkeiten wurden Kooperationen mit sieben internationalen Institutionen und Forschungseinrichtungen geschlossen, die Trainingsdaten für den Wettbewerb zur Verfügung gestellt haben. Trainingsdaten wurden den Teilnehmer über die Plattform www.codalab.com zur Verfügung gestellt. Eine automatische Auswertungssoftware wurde programmiert, um verschiedene Methoden standardisiert miteinander vergleichen zu können. Die Liver Tumor Segmentation Challenge (LITS) 2017 umfasst 200 CT-Aufnahmen mit verschiedenen Lebertumorerkrankungen. Die Segmentierungen der Leber und der Lebertumore wurde von erfahrenen Radiologen durchgeführt. Der LITS Datensatz bietet eine hohe Variation der Lebertumore hinsichtlich des Subtypes des Tumors, der Anzahl an Tumoren und dem Kontrast der Aufnahme. Die LITS Challenge erfuhr mehr als 520 Teilnehmer (Stand Mai 2017) und wurde zur populärsten Challenge auf der IEEE ISBI Konferenz und der Codalab Plattform. Am Workshop zur LITS Challenge an der IEEE ISBI Konferenz nahmen über 80 Teilnehmer teil.

In Tabelle 3 werden die Ergebnisse der LITS Challenge auf der IEEE ISBI Konferenz

2017 aufgezeigt. Die entwickelte Methode der Lebertumorsegmentierung basierend auf Cascaded Fully Convolutional Neural Networks [1] erzielte den vierten Platz.

4 Vorhersage von Patientenüberleben in HCC-Tumor

Akkurate Segmentierungen von Organen und Gewebestrukturen liefern die Basis für die quantitative und computergestützte Analyse von medizinischen Bilddaten. Mit Hilfe von Segmentierungen, wie sie im Rahmen dieser Arbeit für Leber und Lebertumor entwickelt worden sind, können quantitative Biomarker gefunden und erforscht werden. Die genutzte Modalität spielt bei der Erforschung von quantitativen Biomarkern eine entscheidende Rolle. Während in der Computertomographie Unterschiede in den Abschwächungskoeffizenten der untersuchten Gewebe den Kontrast im Bild erzeugen, können spezielle MRT-Sequenzen andere physikalische Eigenschaften messen. Wie bereits in Kapitel 2.3 beschrieben, konnte Heid et al. (2017) einen Zusammenhang zwischen der Verteilung des Apparent Diffusion Coefficent und dem Überleben von Patienten mit Pankreastumor feststellen. Heid et al. (2017) nutzten für ihre Analyse manuelle Segmentierungen der Pancreastumore [71].
Im Rahmen dieser Arbeit wurde eine Methode entwickelt, die es erlaubt automatisch die Überlebenszeiten eines Lebertumorpatienten vorherzusagen. Im Gegensatz zu früheren Arbeiten, die einen Zusammenhang zwischen einem quantitativen Biomarker (wie z.B. ADC) und der Überlebenszeit herstellen konnten, benötigt die entwickelte Methode keine manuellen Segmentierungen [73, 103, 104, 74, 105]. Die benötigten Segmentierungen werden automatisch mit den im Rahmen dieser Arbeit entwickelten Methoden zur automatischen Segmentierung der Leber und der Lebertumore erzeugt.

Als weitere Neuerung kommen, anders als in früheren Arbeiten, wie z.B. [73, 103, 104, 74, 105], keine Bilddeskriptoren zum Einsatz, sondern ein 3D Convolutional Neural Network (3D CNN). Der Vorteil der Verwendung eines 3D CNN im Vergleich zu klassischen Bilddeskriptoren liegt darin, dass das künstliche neuronale Netzwerk spezifisch auf den jeweiligen Anwendungsfall (hier: Überlebensvorhersage von HCC Tumor in MR-DWI ADC) angepasst und trainiert wird. Die im Rahmen dieser Arbeit durchgeführten Experimente konnten zeigen, dass das trainierte 3DCNN eine höhere Genauigkeit von 65% bei der Überlebensvorhersage erreicht als die klassischen Methoden der Literatur, wie z.B. Histogramm-Bilddeskriptoren 61% [71] und Haralick-Bilddeskriptoren 61% [76, 74, 73].

Abschließend war es möglich zu zeigen, dass die vollständige automatische Überlebensvorhersage, bestehend aus automatischer Lebertumorsegmentierung mit Cascaded Fully Convolutional Neural Networks und Überlebensvorhersage mit 3D Convolutional Neural Networks, zu den gleichen Ergebnissen kommt ($p > 0.953$) wie eine Überlebensvorhersage, die auf manuellen Segmentierungen von Radiologen beruht.

Die Ergebnisse der Überlebensvorhersage bei HCC-Tumor finden sich in der Arbeit

SurvivalNet: Predicting patient survival from Diffusion Weighted Magnetic Resonance images using Cascaded Fully Convolutional and 3D Convolutional Neural Networks auf Seite 41 sowie dem unveröffentlichen Manuskript *Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks* im Anhang auf Seite 89.

5 Regression von Broteinheiten für Diabetes Patienten

Neben der Analyse von medizinischen Bilddaten kann die Untersuchung von Essensbildern von medizinischer Relevanz sein. Über 200 Millionen Patienten leiden weltweit unter Diabetes. Spezielle Formen und Ausprägungen von Diabetes, wie Schwangschaftsdiabetes oder Diabetes Typ 1 bei Kleinkindern, können plötzlich auftreten und stellen Betroffene vor große Herausforderungen. Betroffene müssen die Menge an aufgenommenen Kohlenhydraten oder Broteinheiten exakt bestimmen, um den Blutzuckerspiegel entsprechend mit künstlichem Insulin einstellen zu können. Diabetes Patienten mit längerem Krankheitsverlauf lernen über einen längeren Zeitraum eine Einschätzung von Broteinheiten ihrer Nahrung. Neuerkrankten und Angehörigen fehlt dieses Wissen. Dieses Nichtwissen kann zu Gefahren bei der falschen Dosierung und Einstellung des Blutzuckerspiegels führen [80].

Die automatische Bestimmung von Broteinheiten aus Fotos von Mahlzeiten würde die Gefahren der Fehldosierung reduzieren und könnte den Betroffenen helfen ein Expertenwissen aufzubauen. Im Rahmen dieser Arbeit wurde versucht dem Problem der automatischen Broteinheitenschätzung mit Hilfe von Algorithmen der medizinischen Bildverarbeitung zu begegnen. Die Problematik wird in zwei Schritten angegangen. Im ersten Schritt wird versucht die Menge/Volumen des Gerichts, welches der Patient zu sich nehmen möchte, in einem Foto zu erkennen. Im zweiten Schritt wird versucht mit Hilfe der erkannten Menge eine Abschätzung der Broteinheiten abzugeben. Zur Validierung des Konzeptes wurde ein Datensatz, bestehend aus 60 Gerichten, mit einer Kinect v2 Tiefenkamera erhoben. Die 60 RGB-Aufnahmen der Gerichte wurden 20 erfahrenen Diabetikern gezeigt, die für jedes Gericht die entsprechenden Broteinheiten schätzen sollten.

Die Bestimmung der Menge/Volumen eines Gerichts wurde durch Schätzung der Tiefe aus einer zweidimensionalen RGB-Projektion eines Gerichts modelliert. Ein Fully Convolutional Neural Network schätzt für jeden Bildpunkt dessen Abstand zum Kamerasensor/Nullpunkt. Dieser Lösungsweg weist große Ähnlichkeiten zu den Methoden der automatischen Segmentierung der Leber und des Lebertumors auf. Im Gegensatz zur Segmentierung, bei dem eine diskrete Klasse für jeden Bildpunkt gesucht wird (0: Hintergrund, 1: Leber, 2:Lebertumor), besteht die Regression aus einer Schätzung von kontinuierlichen Werten (Abstand zum Kamerasensor von 0-10m) für jeden Bildpunkt. Der Algorithmus wird mit Hilfe des erhobenen Datensatzes, bestehend aus einer RGB- und einer korrespondierenden Tiefenaufnahme, trainiert.

Für die finale Abschätzung der Broteinheiten wird eine RGB-Aufnahme gemeinsam mit ihrer geschätzten Tiefenkarte verarbeitet. Die RGB-Aufnahme und geschätzte Tiefenkarte werden verwendet, um mit Hilfe eines Convolutional Neural Network die Broteinheit (0-10BE) zu schätzen. Das Convolutional Neural Network kann somit auf die Semantik aus RGB-Aufnahme (Was für ein Gericht?) und Menge aus Tiefenkarte (Wieviel von einem Gericht?) zur Abschätzung der Broteinheiten zurückgreifen.

Diese vorgeschlagene Methode erreichte bei der Schätzung der Broteinheiten einen RMSE von 1.53 und liegt somit in der Nähe der Schätzungen der Diabetiker mit einem RMSE von 0.89. Frühere Arbeiten auf diesem Gebiet beschäftigten sich mit der Klassifizierung von Gerichten [106, 107] oder der Schätzung von Kalorien [108]. Diese Arbeit unterscheidet sich dahingehend, dass sie Bilddaten von Gerichten verwendet, um ein medizinisch relevantes Problem zu lösen und zukünftig die Lebensqualität von Diabetikern zu erhöhen.

Eine Zusammenfassung der Ergebnisse zur Regression von Broteinheiten für Diabetespatienten finden sich in dem unveröffentlichen Manuskript *Diabetes60 - Inferring Bread Units From Food Images Using Fully Convolutional Neural Networks* im Anhang Seite 77.

6 Untersuchung der User Experience bei autonom fliegenden Systemen

Computer und Maschinen übernehmen selbstständig immer wichtigere Aufgaben des Menschen. Von der automatischen Detektion von Tumoren, über die Überlebensvorhersage bis hin zur Schätzung von Broteinheiten wurden im Rahmen dieser Arbeit Methoden und Algorithmen entwickelt, die essentielle und existentielle Aufgaben des Menschen für den Menschen übernehmen. Maschinen agieren zunehmend autonom, angetrieben von den Entwicklungen im Bereich der künstlichen Intelligenz und der Bildverarbeitung. Gleichzeitig wächst die Komplexität und Verantwortung der Aufgaben sowie die Abhängigkeit des Menschen von der Maschine, denn ein höherer Grad an Autonomie der Maschine geht mit einem Verlust der Kontrolle des Menschen einher. Dieses wachsende Spannungsfeld zwischen Mensch und Maschine wurde im Rahmen dieser Arbeit in einer Fallstudie untersucht.

Im Rahmen einer Fallstudie wurde der Einfluss des Grades der Systemautonomie auf die Nutzererfahrung (engl. User Expierence) untersucht. Vier Drohnen mit unterschiedlichen Bilderkennungsalgorithmen und Systemautonomieleveln wurden prototypisch entwickelt. Die Autonomielevel der Drohnen ließen sich nach Sherdan et al. (1978) in halb-autonom und voll-autonom untergliedern [109]. Die Nutzererfahrung beim Umgang mit den autonomen Drohnen wurde qualitativ mit Hilfe von strukturierten Interviews erhoben. Für diese Fallstudie wurden 24 strukturierte Interviews mit Studenten im Alter von 22 bis 26 ($\mu = 24$) Jahren geführt.

Die Fallstudie zeigte, dass der Grad der Systemautonomie einen Einfluss auf die Nutzererfahrung hat. Ausgehend von den qualitativen Ergebnissen der strukturierten Interviews konnten Empfehlungen für die Entwicklung von autonomen System hinsichtlich der Nutzererfahrung gegeben werden.

Eine ausführliche Beschreibung der Ergebnisse finden sich in dem veröffentlichten Artikel *Human-Drone-Interaction: A Case Study to Investigate the Relation Between Autonomy and User Experience* auf Seite 47.

Ausblick

„I would rather have questions that can't be answered than answers that can't be questioned."

Richard Feynman (1918-1988)

In dieser Arbeit wurde der Einsatz von Convolutional und Fully Convolutional Neural Networks zur automatischen Detektion und Segmentierung von medizinischen Bilddaten untersucht. Durch die Entwicklung von Cascaded Fully Convolutional Neural Networks zur Segmentierung von Leber und Lebertumor und die Konzeption und Umsetzung von 3D Convolutional Neural Networks zur Stratifizierung von Tumorpatienten konnte ein wichtiger Beitrag geleistet werden. Zusammenfassend lassen sich die Beiträge dieser Arbeit wie folgt zusammenfassen:

- Anwendung von Fully Convolutional Neural Networks zur Segmentierung der Leber in Computertomographie- und Magnetresonanztomographie-Aufnahmen
- Anwendung von Fully Convolutional Neural Networks zur Segmentierung von Lebertumor in Computertomographie- und Magnetresonanztomographie-Aufnahmen
- Anwendung von Convolutional Neural Networks zur Stratifizierung von HCC Tumorpatienten aus Magnetresonanztomographie-Aufnahmen
- Organisation und Durchführung der Liver Tumor Segmentation Challenge im Rahmen der IEEE ISBI und MICCAI 2017 Konferenz
- Anwendung von Fully Convolutional Neural Networks zur Regression von Broteinheiten aus Bildern von Mahlzeiten
- Untersuchung der User Experience bei autonom fliegenden Systemen im Rahmen einer Fallstudie

Die Beiträge und Projekte dieser Arbeit werden in vielseitiger Form weitergeführt. Der Quellcode der entwickelten Methoden zur Segmentierung von Leber und Lebertumor wurde auf der Open-Source Software Plattform Github veröffentlicht [110]. Dieser Quellcode zu Cascaded Fully Convolutional Neural Networks fand bereits Anwendung in über 50 Projekten (Stand 24.06.2017). Erste Publikationen, basierend auf dieser Methode und diesem Quellcode, wurden bereits veröffentlicht und konnten Anwendung in der Liver Tumor Segmentation Challenge finden [111, 112, 113]. Aktuell ermöglicht der veröffentlichte Quellcode die automatische Segmentierung der Leber

und Lebertumore mit Hilfe einer webbasierten Benutzeroberfläche, die jedoch Vorwissen in der Programmiersprache Python voraussetzt. Geplante Projekte sollen die Methode in den klinischen Alltag überführen. Dazu soll eine Erweiterung für die weitverbreitete medizinische Bildanalyse-Software Osirix entwickelt werden. Diese Erweiterung soll in der Lage sein ein medizinisches Volumen automatisch mit der Methode der Cascaded Fully Convolutional Neural Networks zu segmentieren.

Die vielversprechenden Ergebnisse der Stratifizierung von HCC Tumorpatienten aus Magnetresonanztomographie-Aufnahmen sollen in weiteren medizinischen Studien tiefgehender untersucht werden. Dazu wurden bereits weitere Patienten rekrutiert und zusätzliche Datensätze gesichtet. Eine medizinische Publikation ist mit den Kollegen des Universitätskrankenhauses rechts der Isar (MRI) für Herbst-Winter 2017 geplant. Weiteres Potential bietet die Anwendung der Technologie bei Patienten mit Pankreastumor. Frühere Arbeiten konnten bereits Ergebnisse mit klassischen Methoden erzielen. Das *SurvivalNet* böte auch bei der Untersuchung von Patienten mit Pankreastumor die gleichen Vorteile gegenüber klassischen Methoden, wie bei HCC Patienten gezeigt wurde. Vorläufige Ergebnisse konnten eine Genauigkeit bei der Vorhersage von Überlebenszeiten bei Pankreastumorpatienten von über 75% erzielen.

Die Liver Tumor Segmentation Challenge konnte großen Zuspruch in der Wissenschaftsgemeinschaft feiern. Weiteres Interesse konnte die LITS Challenge bei Unternehmen aus dem Bereich der künstliche Intelligenz und Bildverarbeitung wecken. Der GPU-Hersteller NVIDIA förderte die LITS Challenge mit der Spende einer Titan X GPU. Weitere Unternehmen aus der medizinischen Bildverarbeitung, wie z.B. Fraunhofer Mevis und Predible Health, aber auch Technologieunternehmen, wie z.B. Google, nahmen an der LITS Challenge teil. Es ist zu erwarten, dass die LITS Challenge die zukünftige Forschung im Bereich der medizinischen Segmentierung beeinflussen wird. Neu entwickelte Segmentierungsalgorithmen können einfach mit Hilfe des LITS Datensatzes auf der Codalab Plattform hinsichtlich ihrer Segmentierungsgenauigkeit überprüft werden. Es ist geplant die besten eingereichten Methoden im Rahmen einer Journal Publikation zu untersuchen und zusammenzufassen. Eine Veröffentlichung des Quellcodes aller Methoden ist angedacht.

Bossard et al. (2014) haben mit ihrer Arbeit *Food-101: Mining discriminative components with random forests* und der Veröffenlichung ihres Food101 Datensatzes einen wichtigen Meilenstein im Bereich der computergestützten Nährstoffermittlung gelegt [106]. Anknüpfend an Bossard et al. (2014) ist auch geplant, den erhobenen RGB-D Datensatz der Wissenschaftsgemeinschaft nach Veröffentlichung des Manuskripts zur automatischen Regression von Broteinheiten zur Verfügung zu stellen. Der RGB-D Datensatz soll über die Plattform Codalab im Rahmen einer Challenge veröffentlicht werden. Weitere Forschungsarbeiten zur Verbesserung der Klassifizierungsgenauigkeit mit Hilfe von Generative Adversarial Networks (GANs) befinden sich in Planung.

Die zukünftige Forschung im Bereich autonom fliegender Systeme kann auf die im

Rahmen dieser Arbeit entwickelten Algorithmen zurückgreifen. Die Algorithmen wurden unter einer Open-Source Lizenz auf der Plattform Github veröffentlicht [114]. Der Quellcode enthält Unterrichtsmaterialien und mehrere prototypische Bilderkennungsalgorithmen, die im Rahmen des Seminars Autonomous Drones am Center for Digital and Technology Management entstanden sind. Quellcode und Bibliotheken zur Steuerung von Parrot ARDrones sind ebenfalls vorhanden und erleichtern den Einstieg in diese Thematik.

Zusammenfassend liefert diese Arbeit die Grundlage für viele neue und spannende Forschungsrichtungen. Neue Erkenntnisse im Bereich der künstlichen Intelligenz werden starke Auswirkungen auf die Medizin und die Behandlung von Patienten haben. Computergestützte Diagnosesysteme (CADs) werden den Einzug ins Krankenhaus erhalten und Ärzte bei ihrer Arbeit unterstützen. Intelligente Assistenzsysteme werden chronisch-kranken Patienten, wie z.B. Diabetikern, eine wichtige Stütze bei der täglichen Behandlung ihrer Krankheit sein. Die gesellschaftliche Diskussion über die zukünftige Rolle und Gefahren von intelligenten Systemen und Diensten in der Medizin steht jedoch noch aus.

„Künstliche Intelligenz ist allemal besser als natürliche Dummheit."

Hans Matthöfer (1925-2009)

Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields

Authoren: Patrick Ferdinand Christ, Mohamed Ezzeldin A Elshaer, Florian Ettlinger, Sunil Tatavarty, Marc Bickel, Patrick Bilic, Markus Rempfler, Marco Armbruster, Felix Hofmann, Melvin D'Anastasi, Wieland H Sommer, Seyed-Ahmad Ahmadi und Bjoern H Menze

Abstract: Automatic segmentation of the liver and its lesion is an important step towards deriving quantitative biomarkers for accurate clinical diagnosis and computer-aided decision support systems. This paper presents a method to automatically segment liver and lesions in CT abdomen images using cascaded fully convolutional neural networks (CFCNs) and dense 3D conditional random fields (CRFs). We train and cascade two FCNs for a combined segmentation of the liver and its lesions. In the first step, we train a FCN to segment the liver as ROI input for a second FCN. The second FCN solely segments lesions from the predicted liver ROIs of step 1. We refine the segmentations of the CFCN using a dense 3D CRF that accounts for both spatial coherence and appearance. CFCN models were trained in a 2-fold cross-validation on the abdominal CT dataset 3DIRCAD comprising 15 hepatic tumor volumes. Our results show that CFCN-based semantic liver and lesion segmentation achieves DICE scores over 94% for liver with computation times below 100s per volume. We experimentally demonstrate the robustness of the proposed method as a decision support system with a high accuracy and speed for usage in daily clinical routine.

Publikationsdatum: 17.10.2016
Konferenz: International Conference on Medical Image Computing and Computer-Assisted Intervention
Seiten: 415-423
Verlag: Springer International Publishing

Individuelle Leistungsbeiträge: Datenakquise und Datenaufbereitung, Konzeption und Durchführung von Experimenten, Federführende Anfertigung des Manuskripts

Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields

Patrick Ferdinand Christ[1( )], Mohamed Ezzeldin A. Elshaer[1], Florian Ettlinger[1], Sunil Tatavarty[2], Marc Bickel[1], Patrick Bilic[1], Markus Rempfler[1], Marco Armbruster[4], Felix Hofmann[4], Melvin D'Anastasi[4], Wieland H. Sommer[4], Seyed-Ahmad Ahmadi[3], and Bjoern H. Menze[1]

[1] Image-Based Biomedical Modeling Group, Technische Universität München, Arccisstrasse 21, 80333 Munich, Germany
{Patrick.Christ,Bjoern.Menze}@tum.de
[2] Chair for Data Processing, Technische Universität München, Arccisstrasse 21, 80333 Munich, Germany
[3] Department for Neurology, LMU Hospital Grosshadern, Marchioninistrasse 15, 81377 Munich, Germany
[4] Department for Clinical Radiology, LMU Hospital Grosshadern, Marchioninistrasse 15, 81377 Munich, Germany

Abstract. Automatic segmentation of the liver and its lesion is an important step towards deriving quantitative biomarkers for accurate clinical diagnosis and computer-aided decision support systems. This paper presents a method to automatically segment liver and lesions in CT abdomen images using cascaded fully convolutional neural networks (CFCNs) and dense 3D conditional random fields (CRFs). We train and cascade two FCNs for a combined segmentation of the liver and its lesions. In the first step, we train a FCN to segment the liver as ROI input for a second FCN. The second FCN solely segments lesions from the predicted liver ROIs of step 1. We refine the segmentations of the CFCN using a dense 3D CRF that accounts for both spatial coherence and appearance. CFCN models were trained in a 2-fold cross-validation on the abdominal CT dataset 3DIRCAD comprising 15 hepatic tumor volumes. Our results show that CFCN-based semantic liver and lesion segmentation achieves Dice scores over 94 % for liver with computation times below 100 s per volume. We experimentally demonstrate the robustness of the proposed method as a decision support system with a high accuracy and speed for usage in daily clinical routine.

Keywords: Liver · Lesion · Segmentation · FCN · CRF · CFCN · Deep learning

1 Introduction

Anomalies in the shape and texture of the liver and visible lesions in CT are important biomarkers for disease progression in primary and secondary hepatic

S. Ourselin et al. (Eds.): MICCAI 2016, Part II, LNCS 9901, pp. 415–423, 2016.
DOI: 10.1007/978-3-319-46723-8_48

tumor disease [9]. In clinical routine, manual or semi-manual techniques are applied. These, however, are subjective, operator-dependent and very time-consuming. In order to improve the productivity of radiologists, computer-aided methods have been developed in the past, but the challenges in automatic segmentation of combined liver and lesion remain, such as low-contrast between liver and lesion, different types of contrast levels (hyper-/hypo-intense tumors), abnormalities in tissues (metastasectomie), size and varying amount of lesions.

Nevertheless, several interactive and automatic methods have been developed to segment the liver and liver lesions in CT volumes. In 2007 and 2008, two Grand Challenges benchmarks on liver and liver lesion segmentation have been conducted [4,9]. Methods presented at the challenges were mostly based on statistical shape models. Furthermore, grey level and texture based methods have been developed [9]. Recent work on liver and lesion segmentation employs graph cut and level set techniques [15–17], sigmoid edge modeling [5] or manifold and machine learning [6,11]. However, these methods are not widely applied in clinics, due to their speed and robustness on heterogeneous, low-contrast real-life CT data. Hence, interactive methods were still developed [1,7] to overcome these weaknesses, which yet involve user interaction.

Deep Convolutional Neural Networks CNN have gained new attention in the scientific community for solving computer vision tasks such as object recognition, classification and segmentation [14,18], often out-competing state-of-the art methods. Most importantly, CNN methods have proven to be highly robust to varying image appearance, which motivates us to apply them to fully automatic liver and lesions segmentation in CT volumes.

Semantic image segmentation methods based on fully convolutional neural networks FCN were developed in [18], with impressive results in natural image segmentation competitions [3,24]. Likewise, new segmentation methods based on CNN and FCNs were developed for medical image analysis, with highly competitive results compared to state-of-the-art. [8,12,19–21,23].

In this work, we demonstrate the combined automatic segmentation of the liver and its lesions in low-contrast heterogeneous CT volumes. Our contributions are three-fold. First, we train and apply fully convolutional CNN on CT volumes of the liver for the first time, demonstrating the adaptability to challenging segmentation of hepatic liver lesions. Second, we propose to use a cascaded fully convolutional neural network (CFCN) on CT slices, which segments liver and lesions sequentially, leading to significantly higher segmentation quality. Third, we propose to combine the cascaded CNN in 2D with a 3D dense conditional random field approach (3DCRF) as a post-processing step, to achieve higher segmentation accuracy while preserving low computational cost and memory consumption. In the following sections, we will describe our proposed pipeline (Sect. 2.2) including CFCN (Sect. 2.3) and 3D CRF (Sect. 2.4), illustrate experiments on the 3DIRCADb dataset (Sect. 2) and summarize the results (Sect. 4).

2 Methods

In the following section, we denote the 3D image volume as I, the total number of voxels as N and the set of possible labels as $\mathcal{L} = \{0, 1, \ldots, l\}$. For each voxel i, we define a variable $x_i \in \mathcal{L}$ that denotes the assigned label. The probability of a voxel i belonging to label k given the image I is described by $P(x_i = k|I)$ and will be modelled by the FCN. In our particular study, we use $\mathcal{L} = \{0, 1, 2\}$ for background, liver and lesion, respectively.

2.1 3DIRCADb Dataset

For clinical routine usage, methods and algorithms have to be developed, trained and evaluated on heterogeneous real-life data. Therefore, we evaluated our proposed method on the 3DIRCADb dataset[1] [22]. In comparison to the grand challenge datasets, the 3DIRCADb dataset offers a higher variety and complexity of livers and its lesions and is publicly available. The 3DIRCADb dataset

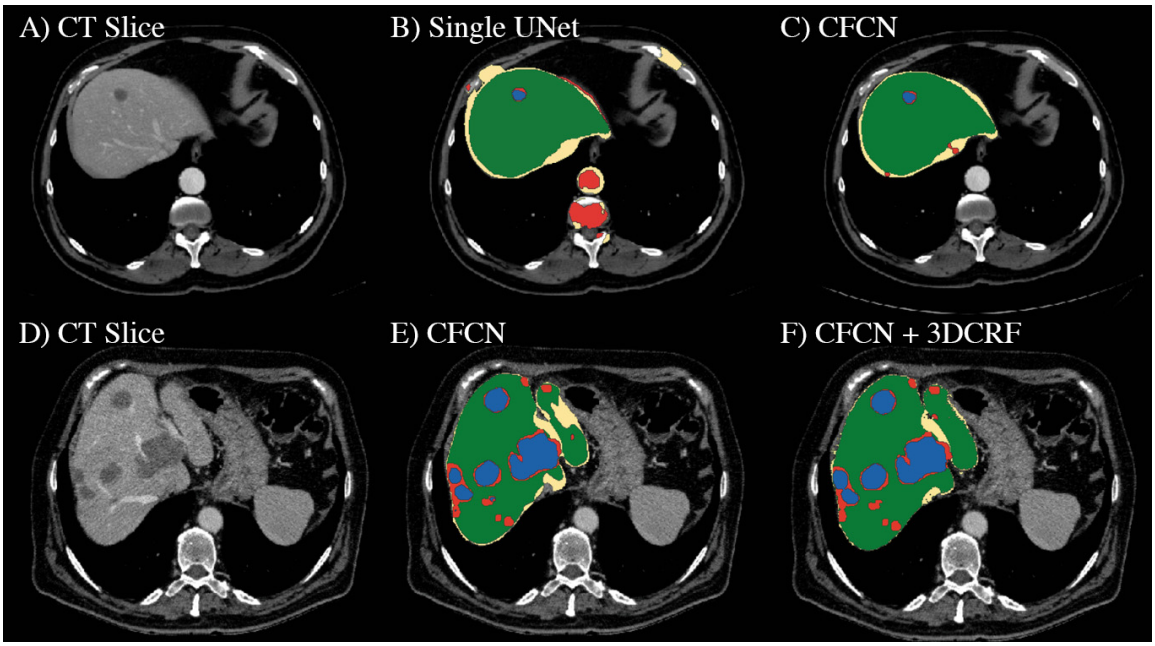


Fig. 1. Automatic liver and lesion segmentation with cascaded fully convolutional networks (CFCN) and dense conditional random fields (CRF). Green depicts correctly predicted liver segmentation, yellow for liver false negative and false positive pixels (all wrong predictions), blue shows correctly predicted lesion segmentation and red lesion false negative and false positive pixels (all wrong predictions). In the first row, the false positive lesion prediction in B of a single UNet as proposed by [20] were eliminated in C by CFCN as a result of restricting lesion segmentation to the liver ROI region. In the second row, applying the 3DCRF to CFCN in F increases both liver and lesion segmentation accuracy further, resulting in a lesion Dice score of 82.3 %.

[1] The dataset is available on http://ircad.fr/research/3d-ircadb-01.

includes 20 venous phase enhanced CT volumes from various European hospitals with different CT scanners. For our study, we trained and evaluated our models using the 15 volumes containing hepatic tumors in the liver with 2-fold cross validation. The analyzed CT volumes differ substantially in the level of contrast-enhancement, size and number of tumor lesions (1 to 42). We assessed the performance of our proposed method using the quality metrics introduced in the grand challenges for liver and lesion segmentation by [4,9].

2.2 Data Preparation, Processing and Pipeline

Pre-processing was carried out in a slice-wise fashion. First, the Hounsfield unit values were windowed in the range $[-100, 400]$ to exclude irrelevant organs and objects, then we increased contrast through histogram equalization. As in [20], to teach the network the desired invariance properties, we augmented the data by applying translation, rotation and addition of gaussian noise. Thereby resulting in an increased training dataset of 22,693 image slices, which were used to train two cascaded FCNs based on the UNet architecture [20]. The predicted segmentations are then refined using dense 3D Conditional Random Fields. The entire pipeline is depicted in Fig. 2.

2.3 Cascaded Fully Convolutional Neural Networks (CFCN)

We used the UNet architecture [20] to compute the soft label probability maps $P(x_i|I)$. The UNet architecture enables accurate pixel-wise prediction by combining spatial and contextual information in a network architecture comprising 19 convolutional layers. In our method, we trained one network to segment the liver in abdomen slices (step 1), and another network to segment the lesions, given an image of the liver (step 2). The segmented liver from step 1 is cropped and resampled to the required input size for the cascaded UNet in step 2, which further segments the lesions.

The motivation behind the cascade approach is that it has been shown that UNets and other forms of CNNs learn a hierarchical representation of the provided data. The stacked layers of convolutional filters are tailored towards the desired classification in a data-driven manner, as opposed to designing handcrafted features for separation of different tissue types. By cascading two UNets, we ensure that the UNet in step 1 learns filters that are specific for the detection and segmentation of the liver from an overall abdominal CT scan, while the UNet in step 2 arranges a set of filters for separation of lesions from the liver tissue. Furthermore, the liver ROI helps in reducing false positives for lesions.

A crucial step in training FCNs is appropriate class balancing according to the pixel-wise frequency of each class in the data. In contrast to [18], we observed that training the network to segment small structures such as lesions is not possible without class balancing, due to the high class imbalance. Therefore we introduced an additional weighting factor ω^{class} in the cross entropy loss

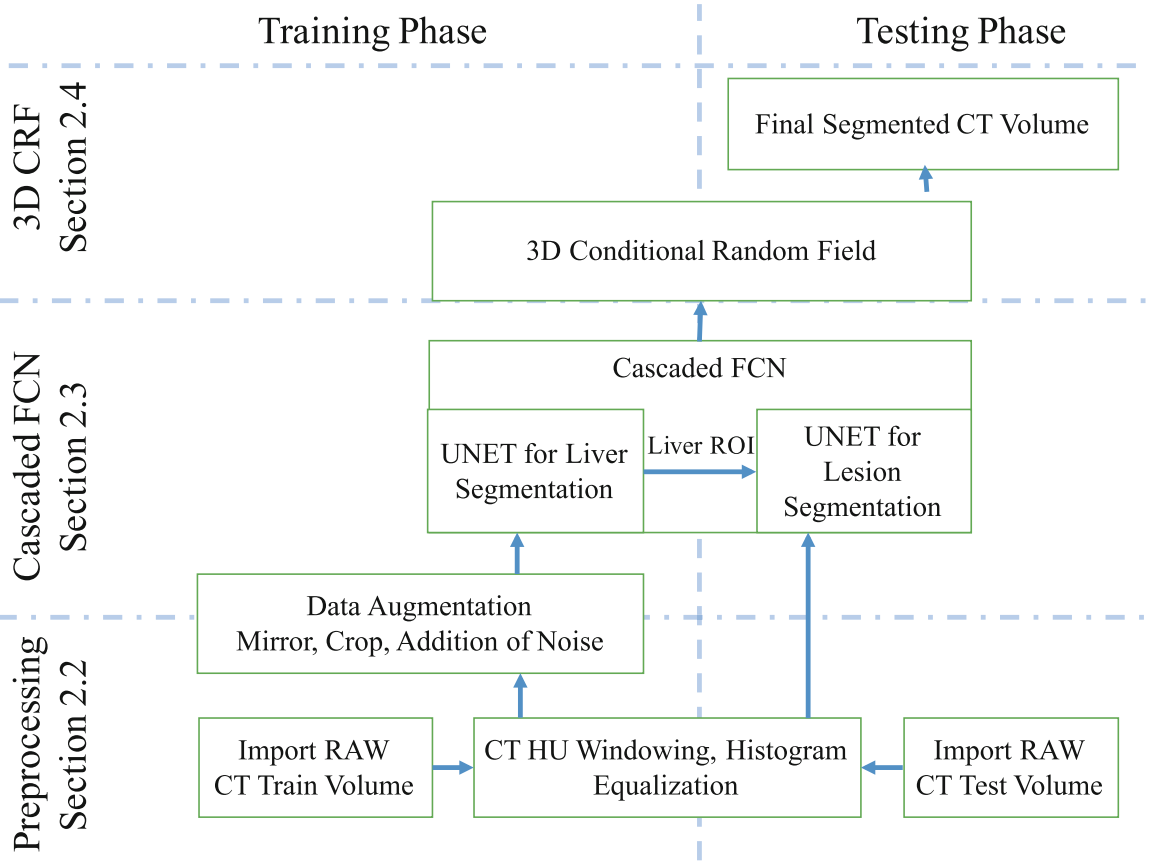


Fig. 2. Overview of the proposed image segmentation pipeline. In the training phase, the CT volumes are trained after pre-processing and data augmentation in a cascaded fully convolutional neural network (CFCN). To gain the final segmented volume, the test volume is fed-forward in the (CFCN) and refined afterwards using a 3D conditional random field 3DCRF.

function L of the FCN.

$$L = -\frac{1}{n}\sum_{i=1}^{N} \omega_i^{class} \left[\hat{P}_i \log P_i + (1-\hat{P}_i)\log(1-P_i)\right] \tag{1}$$

P_i denotes the probability of voxel i belonging to the foreground, $\hat{P}_i$ represents the ground truth. We chose ω_i^{class} to be $\frac{1}{|\text{Pixels of Class } x_i=k|}$.

The CFCNs were trained on a NVIDIA Titan X GPU, using the deep learning framework caffe [10], at a learning rate of 0.001, a momentum of 0.8 and a weight decay of 0.0005.

2.4 3D Conditional Random Field (3DCRF)

Volumetric FCN implementation with 3D convolutions is strongly limited by GPU hardware and available VRAM [19]. In addition, the anisotropic resolution of medical volumes (e.g. 0.57−0.8 mm in xy and 1.25−4 mm in z voxel dimension in 3DIRCADb) complicates the training of discriminative 3D filters. Instead, to

capitalise on the locality information across slices within the dataset, we utilize 3D dense conditional random fields CRFs as proposed by [13]. To account for 3D information, we consider all slice-wise predictions of the FCN together in the CRF applied to the entire volume at once.

We formulate the final label assignment given the soft predictions (probability maps) from the FCN as *maximum a posteriori* (MAP) inference in a dense CRF, allowing us to consider both spatial coherence and appearance.

We specify the dense CRF following [13] on the complete graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with vertices $i \in \mathcal{V}$ for each voxel in the image and edges $e_{ij} \in \mathcal{E} = \{(i,j) \ \forall i,j \in \mathcal{V} \text{ s.t. } i < j\}$ between *all* vertices. The variable vector $\mathbf{x} \in \mathcal{L}^N$ describes the label of each vertex $i \in \mathcal{V}$. The energy function that induces the according Gibbs distribution is then given as:

$$E(\mathbf{x}) = \sum_{i \in \mathcal{V}} \phi_i(x_i) + \sum_{(i,j) \in \mathcal{E}} \phi_{ij}(x_i, x_j), \tag{2}$$

where $\phi_i(x_i) = -\log P(x_i|I)$ are the unary potentials that are derived from the FCNs probabilistic output, $P(x_i|I)$. $\phi_{ij}(x_i, x_j)$ are the pairwise potentials, which we set to:

$$\begin{aligned} \phi_{ij}(x_i, x_j) = \mu(x_i, x_j) \Bigg(& w_{\text{pos}} \exp\left(-\tfrac{|p_i - p_j|^2}{2\sigma_{\text{pos}}^2}\right) \\ & + w_{\text{bil}} \exp\left(-\tfrac{|p_i - p_j|^2}{2\sigma_{\text{bil}}^2} - \tfrac{|I_i - I_j|^2}{2\sigma_{\text{int}}^2}\right) \Bigg), \end{aligned} \tag{3}$$

where $\mu(x_i, x_j) = \mathbf{1}(x_i \neq x_j)$ is the Potts function, $|p_i - p_j|$ is the spatial distance between voxels i and j and $|I_i - I_j|$ is their intensity difference in the original image. The influence of the pairwise terms can be adjusted with their weights w_{pos} and w_{bil} and their effective range is tuned with the kernel widths $\sigma_{\text{pos}}, \sigma_{\text{bil}}$ and σ_{int}.

We estimate the best labelling $\mathbf{x}^* = \arg\min_{\mathbf{x} \in \mathcal{L}^N} E(\mathbf{x})$ using the efficient mean field approximation algorithm of [13]. The weights and kernels of the CRF were chosen using a random search algorithm.

3 Results and Discussion

The qualitative results of the automatic segmentation are presented in Fig. 1. The complex and heterogeneous structure of the liver and all lesions were detected in the shown images. The cascaded FCN approach yielded an enhancement for lesions with respect to segmentation accuracy compared to a single FCN as can be seen in Fig. 1. In general, we observe significant[2] additional improvements for slice-wise Dice overlaps of liver segmentations, from mean Dice 93.1 % to 94.3 % after applying the 3D dense CRF.

Quantitative results of the proposed method are reported in Table 1. The CFCN achieves higher scores as the single FCN architecture. Applying the 3D

[2] Two-sided paired t-test with p-value $< 4 \cdot 10^{-19}$.

Table 1. Quantitative segmentation results of the liver on the 3DIRCADb dataset. Scores are reported as presented in the original papers.

Approach	VOE [%]	RVD [%]	ASD [mm]	MSD [mm]	DICE [%]
UNET as in [20]	39	87	19.4	119	72.9
Cascaded UNET	12.8	−3.3	2.3	46.7	93.1
Cascaded UNET + 3D CRF	10.7	−1.4	1.5	24.0	94.3
Li et al. [16] (liver-only)	9.2	−11.2	1.6	28.2	
Chartrand et al. [2] (semi-automatic)	6.8	1.7	1.6	24	
Li et al. [15] (liver-only)					94.5

CRF improved the segmentations results of calculated metrics further. The runtime per slice in the CFCN is $2 \cdot 0.2\,s = 0.4\,\text{s}$ without and 0.8 s with CRF.

In comparison to state-of-the-art, such as [2,5,15,16], we presented a framework, which is capable of a combined segmentation of the liver and its lesion.

4 Conclusion

Cascaded FCNs and dense 3D CRFs trained on CT volumes are suitable for automatic localization and combined volumetric segmentation of the liver and its lesions. Our proposed method competes with state-of-the-art. We provide our trained models under open-source license allowing fine-tuning for other medical applications in CT data[3]. Additionally, we introduced and evaluated dense 3D CRF as a post-processing step for deep learning-based medical image analysis. Furthermore, and in contrast to prior work such as [5,15,16], our proposed method could be generalized to segment multiple organs in medical data using multiple cascaded FCNs. All in all, heterogeneous CT volumes from different scanners and protocols as present in the 3DIRCADb dataset and in clinical trials can be segmented in under 100 s each with the proposed approach. We conclude that CFCNs and dense 3D CRFs are promising tools for automatic analysis of liver and its lesions in clinical routine.

References

1. Ben-Cohen, A., et al.: Automated method for detection and segmentation of liver metastatic lesions in follow-up CT examinations. J. Med. Imaging **3** (2015)
2. Chartrand, G., et al.: Semi-automated liver CT segmentation using Laplacian meshes. In: ISBI, pp. 641–644. IEEE (2014)
3. Chen, L.C., et al.: Semantic image segmentation with deep convolutional nets and fully connected CRFs. In: ICLR (2015)

[3] Trained models are available at https://github.com/IBBM/Cascaded-FCN.

4. Deng, X., Du, G.: Editorial: 3D segmentation in the clinic: a grand challenge ii-liver tumor segmentation. In: MICCAI Workshop (2008)
5. Foruzan, A.H., Chen, Y.W.: Improved segmentation of low-contrast lesions using sigmoid edge model. Int. J. Comput. Assist. Radiol. Surg., 1–17 (2015)
6. Freiman, M., Cooper, O., Lischinski, D., Joskowicz, L.: Liver tumors segmentation from cta images using voxels classification and affinity constraint propagation. Int. J. Comput. Assist. Radiol. Surg. **6**(2), 247–255 (2011)
7. Häme, Y., Pollari, M.: Semi-automatic liver tumor segmentation with hidden markov measure field model and non-parametric distribution estimation. Med. Image Anal. **16**(1), 140–149 (2012)
8. Havaei, M., et al.: Brain tumor segmentation with deep neural networks. Med. Image Anal. (2016)
9. Heimann, T., et al.: Comparison and evaluation of methods for liver segmentation from ct datasets. IEEE Trans. Med. Imag. **28**(8), 1251–1265 (2009)
10. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: convolutional architecture for fast feature embedding. In: Proceeding ACM International Conference Multimedia, pp. 675–678. ACM (2014)
11. Kadoury, S., Vorontsov, E., Tang, A.: Metastatic liver tumour segmentation from discriminant grassmannian manifolds. Phys. Med. Biol. **60**(16), 6459 (2015)
12. Kamnitsas, K., et. al.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation (2016). arXiv preprint arXiv:1603.05959
13. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected CRFs with gaussian edge potentials. In: NIPS, pp. 109–117 (2011)
14. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: NIPS, pp. 1097–1105 (2012)
15. Li, C., Wang, X., Eberl, S., Fulham, M., Yin, Y., Chen, J., Feng, D.D.: A likelihood and local constraint level set model for liver tumor segmentation from ct volumes. IEEE Trans. Biomed. Eng. **60**(10), 2967–2977 (2013)
16. Li, G., Chen, X., Shi, F., Zhu, W., Tian, J., Xiang, D.: Automatic liver segmentation based on shape constraints and deformable graph cut in ct images. IEEE Trans. Image Process. **24**(12), 5315–5329 (2015)
17. Linguraru, M.G., Richbourg, W.J., Liu, J., Watt, J.M., Pamulapati, V., Wang, S., Summers, R.M.: Tumor burden analysis on computed tomography by automated liver and tumor segmentation. IEEE Trans. Med. Imag. **31**(10), 1965–1976 (2012)
18. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: CVPR (2015)
19. Prasoon, A., Petersen, K., Igel, C., Lauze, F., Dam, E., Nielsen, M.: Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8150, pp. 246–253. Springer, Heidelberg (2013). doi:10.1007/978-3-642-40763-5_31
20. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Heidelberg (2015). doi:10.1007/978-3-319-24574-4_28
21. Roth, H.R., Lu, L., Farag, A., Shin, H.-C., Liu, J., Turkbey, E.B., Summers, R.M.: DeepOrgan: multi-level deep convolutional networks for automated pancreas segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9349, pp. 556–564. Springer, Heidelberg (2015). doi:10.1007/978-3-319-24553-9_68

22. Soler, L., et al.: 3D image reconstruction for comparison of algorithm database: a patient-specific anatomical and medical image database (2012)
23. Wang, J., MacKenzie, J.D., Ramachandran, R., Chen, D.Z.: Detection of Glands and Villi by collaboration of domain knowledge and deep learning. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9350, pp. 20–27. Springer, Heidelberg (2015). doi:10.1007/978-3-319-24571-3_3
24. Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.: Conditional random fields as recurrent neural networks. In: ICCV (2015)

SurvivalNet: Predicting patient survival from diffusion weighted magnetic resonance images using cascaded fully convolutional and 3D convolutional neural networks

Authoren: Patrick Ferdinand Christ, Florian Ettlinger, Georgios Kaissis, Sebastian Schlecht, Freba Ahmaddy, Felix Grün, Alexander Valentinitsch, Seyed-Ahmad Ahmadi, Rickmer Braren, Bjoern Menze

Abstract: Automatic non-invasive assessment of hepatocellular carcinoma (HCC) malignancy has the potential to substantially enhance tumor treatment strategies for HCC patients. In this work we present a novel framework to automatically characterize the malignancy of HCC lesions from DWI images. We predict HCC malignancy in two steps: As a first step we automatically segment HCC tumor lesions using cascaded fully convolutional neural networks (CFCN). A 3D neural network (SurvivalNet) then predicts the HCC lesions' malignancy from the HCC tumor segmentation. We formulate this task as a classification problem with classes being "low riskänd "high risk"represented by longer or shorter survival times than the median survival. We evaluated our method on DWI of 31 HCC patients. Our proposed framework achieves an end-to-end accuracy of 65% with a DICE score for the automatic lesion segmentation of 69% and an accuracy of 68% for tumor malignancy classification based on expert annotations. We compared the SurvivalNet to classical handcrafted features such as Histogram and Haralick and show experimentally that SurvivalNet outperforms the handcrafted features in HCC malignancy classification. End-to-end assessment of tumor malignancy based on our proposed fully automatic framework corresponds to assessment based on expert annotations with high significance ($p > 0.95$).

Publikationsdatum: 19.06.2017
Konferenz: IEEE International Symposium on Biomedical Imaging (ISBI)
Publikationsorgan: IEEE

Individuelle Leistungsbeiträge: Projektkoordination, Datenakquise und Datenaufbereitung, Konzeption und Durchführung von Experimenten, Federführende Anfertigung des Manuskripts

SURVIVALNET: PREDICTING PATIENT SURVIVAL FROM DIFFUSION WEIGHTED MAGNETIC RESONANCE IMAGES USING CASCADED FULLY CONVOLUTIONAL AND 3D CONVOLUTIONAL NEURAL NETWORKS

Patrick Ferdinand Christ [1,⋆] *Florian Ettlinger*[1,⋆] *Georgios Kaissis*[1,†]
Sebastian Schlecht[⋆] Freba Ahmaddy[†] Felix Grün[⋆] Alexander Valentinitsch[†]
Seyed-Ahmad Ahmadi[+] Rickmer Braren[2,†] Bjoern Menze[2,⋆]

[⋆] Technische Universität München, Image-Based Biomedical Modeling Group, Arccisstrasse 21, 80333 Munich
[†]Technische Universität München, Institute for diagnostic and interventional Radiology, Ismaninger Str. 22, 81675 Munich
[+] Ludwig-Maximilians-Universität, German Center for Vertigo and Balance Disorders, Feodor-Lynen-Straße 19, 81377 Munich

ABSTRACT

Automatic non-invasive assessment of hepatocellular carcinoma (HCC) malignancy has the potential to substantially enhance tumor treatment strategies for HCC patients. In this work we present a novel framework to automatically characterize the malignancy of HCC lesions from DWI images.

We predict HCC malignancy in two steps: As a first step we automatically segment HCC tumor lesions using cascaded fully convolutional neural networks (CFCN). A 3D neural network (SurvivalNet) then predicts the HCC lesions' malignancy from the HCC tumor segmentation. We formulate this task as a classification problem with classes being "low risk" and "high risk" represented by longer or shorter survival times than the median survival. We evaluated our method on DWI of 31 HCC patients. Our proposed framework achieves an end-to-end accuracy of 65% with a Dice score for the automatic lesion segmentation of 69% and an accuracy of 68% for tumor malignancy classification based on expert annotations. We compared the SurvivalNet to classical handcrafted features such as Histogram and Haralick and show experimentally that SurvivalNet outperforms the handcrafted features in HCC malignancy classification. End-to-end assessment of tumor malignancy based on our proposed fully automatic framework corresponds to assessment based on expert annotations with high significance ($p > 0.95$).

***Index Terms*—** Survival Prediction, 3D Neural Network, Fully Convolutional Neural Networks, MRI

1. INTRODUCTION

1.1. Motivation

Hepatocellular carcinoma (HCC) presents the sixth most common cancer and the third most common cause of cancer-related deaths worldwide [1]. HCC comprises a genetically and molecularly highly heterogeneous group of cancers that commonly arise in a chronically damaged liver. Importantly, HCC subtypes differ significantly in clinical outcome. The stepwise transformation to HCC is accompanied by major changes in tissue architecture including an increase in cellularity and a switch in vascular supply (i.e. arterialization). These differences provide the basis for the non-invasive detection of HCC [2]. In particular, diffusion weighted-magnetic resonance imaging (DW-MRI) detects differences in random Brownian motion, which is commonly reduced in highly cellular HCC due to an increase in cell membranes and macromolecules. The apparent diffusion coefficient (ADC) parameter value, which can be derived from two DW-MRI scans, quantifies this effect. DW-MRI imaging techniques provide a high level of sensitivity and specificity for tumor detection, the distinction of tumor subtypes requires the identification of more subtle differences. Computer aided analysis techniques allow medical image feature extraction far beyond the capabilities of the human eye and thus hold the potential for an imaging based differentiation of tumor subtypes. Non-invasive differentiation of tumor subtypes in HCC would enable pre-therapeutic patient stratification and the systematic testing of novel therapeutic strategies.

1.2. Related Works

Heid et al. (2016) have recently established a close relationship between the regional DW-MRI derived apparent diffusion coefficient (ADC) parameter value and distinct subtypes in pancreatic ductal adenocarcinoma (PDAC) [3]. Computer aided extraction of image features for tumor subtyping has previously been reported for several tumor entities. Prior work focused mostly on hand-crafted feature extraction such as histogram features [4], Gabor and Haralick features [5, 6], and grey level run length based features [7] to predict survival times for diverse tumor entities and image modalities. Recent works leveraged the discriminative power of the ap-

[1]Authors contributed equally
[2]Corresponding authors: rbraren@tum.de and bjoern.menze@tum.de

978-1-5090-1172-8/17/$31.00 ©2017 IEEE

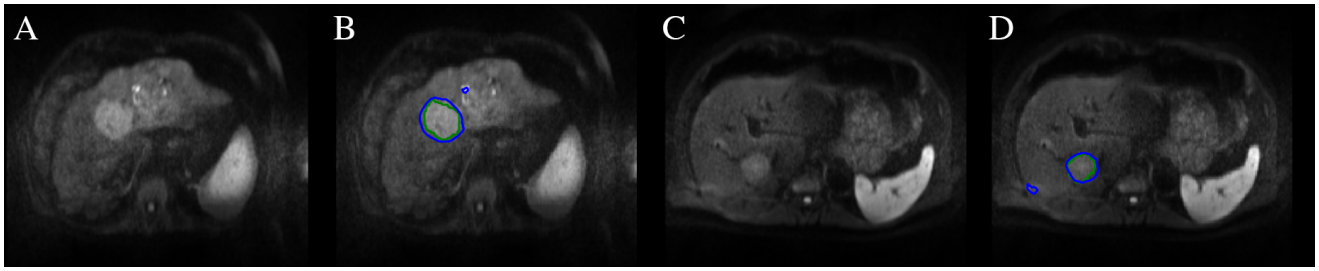


Fig. 1. Results of the automatic HCC tumor segmentation in DW-MRI: A and C show the DW-MRI slice. B and D show the ground truth label of HCC in green and the automatic segmentation using CFCN in blue. The automatic segmentation algorithms successfully segments the HCC tumor in both cases. Only two small false-positive regions and a slight inaccuracy around the edges of the tumors are visible leading to a Dice overlap score of 85 % for B and 83 % for D. The patient in A/B belongs to class "high risk" whereas C/D belongs to class "low risk". Yet, only subtle overall differences in appearance between the tumors are visible.

parent diffusion coefficient (ADC) by extracting texture features for survival or malignancy characterization [8, 9]. Zhou et al. (2016) proposed a method to characterize malignancy of HCC in contrast enhanced MRI by extracting histogram and texture based features such as grey-level co-occurrence and run length (GLRL and GLCM) of HCC lesions [10]. However, their method required manual segmentation of the lesions beforehand.

1.3. Contribution

In comparison to prior work, we developed a method to predict HCC survival from DW-MRI volumes using automatic segmentations of tumors. Our contribution in this work is three-fold. First, we developed an automatic method to detect and segment HCC tumor lesions in DW-MRI data. Second, we found and analyzed quantitative biomarkers using handcrafted and CNN-based features to predict patient survival. Third, we experimentally demonstrated a fully automatic method to predict long/short survival of HCC patients from DW-MRI images.

2. METHODS

Our proposed framework to fully automatically predict short/long survival from DW-MRI of HCC patients is depicted in figure 3.

2.1. Dataset

31 Patients underwent clinical assessment and MR imaging for the primary diagnosis of HCC. Barcelona Clinic Liver Cancer Classification was used to assess the clinical stage of the disease. Patients with a history of prior malignancy were excluded. No data with insufficient quality due to breathing artifacts, excessive banding or distortion, diffuse tumor growth or non-detectability of the lesions in the DW-MRI sequences was included in the dataset. Imaging was performed using a 1.5 T clinical MRI scanner (Avanto, Siemens) with a standard imaging protocol including axial and coronal T2w, axial T1w images before and after application of Gadolinium-DTPA contrast agent (Jenapharm Magnograf 0.5 mmol/ml per manufacturers instructions). Post-contrast T1w images were acquired in the early, mid and late arterial phases as well as in the portal venous phase. Diffusion weighted imaging was performed using a slice thickness of 5mm and a matrix size of 192 by 192. Institutional review board approval was obtained for this retrospective study.

2.2. Automatic Segmentation

To automatically detect and segment tumor lesions we applied a cascaded fully convolutional neural network to segment in step 1 the liver and in a step 2 the tumor lesions from a liver ROI volume [11]. We used the DW-MRI as input to the FCN architecture proposed by Ronneberger et al. (2015) [12]. We fine-tuned our networks using the liver and liver tumor model provided by Christ et al. (2016) [11] and applied a 5-fold cross-validation. Tumor margins were identified in the early arterial phase and in DWI images (b=600). Manual segmentation was performed by an experienced radiologist using the software TurtleSeg® and reviewed by two expert radiologists.

2.3. Survival Prediction

To predict the survival rate of HCC tumor patients we calculated different features using the detected and segmented tumor lesions applied in the ADC image sequences. We calculated handcrafted features and features trained end-to-end by a 3D Convolutional Neural Network (SurvivalNet).

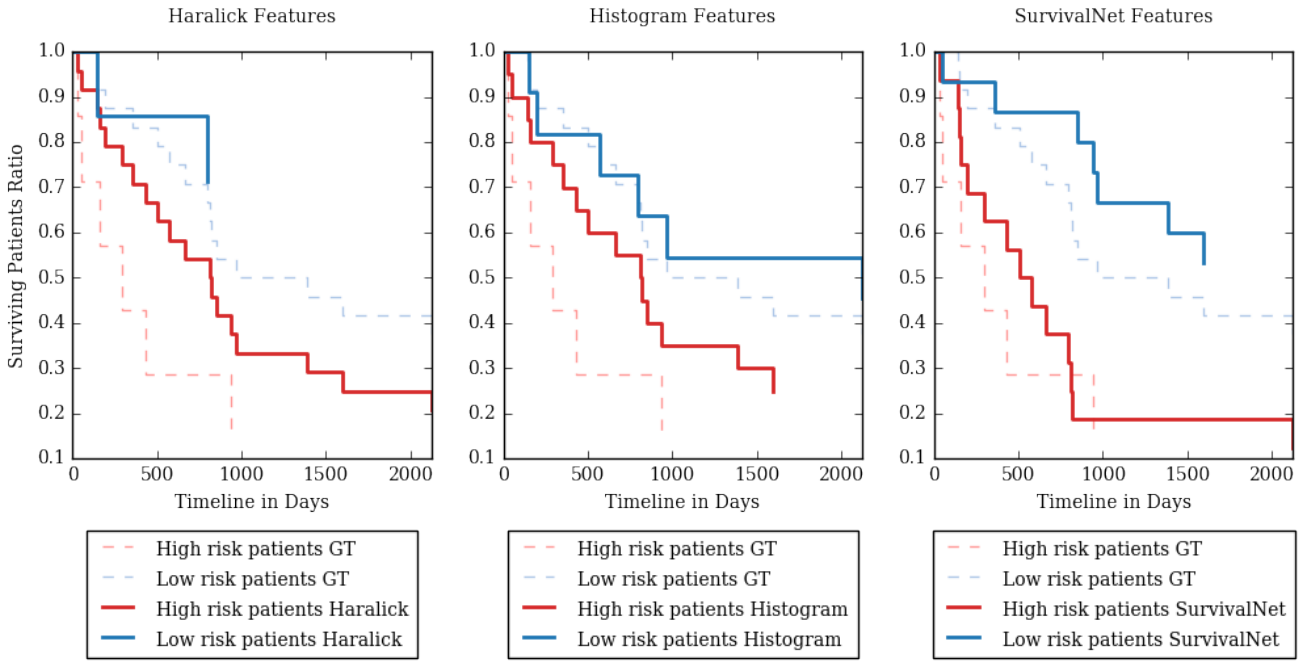


Fig. 2. Kaplan-Meier Survival Analysis for Haralick Features (left), Histogram Features (middle) and 3D SurvivalNet CNN (right). The SurvivalNet is able to split the HCC patients into high risk i.e. short survival and low risk i.e. long survival. In contrast, classical handcrafted features such as histogram and Haralick do not predict the patient survival correctly. GT stands for ground truth.

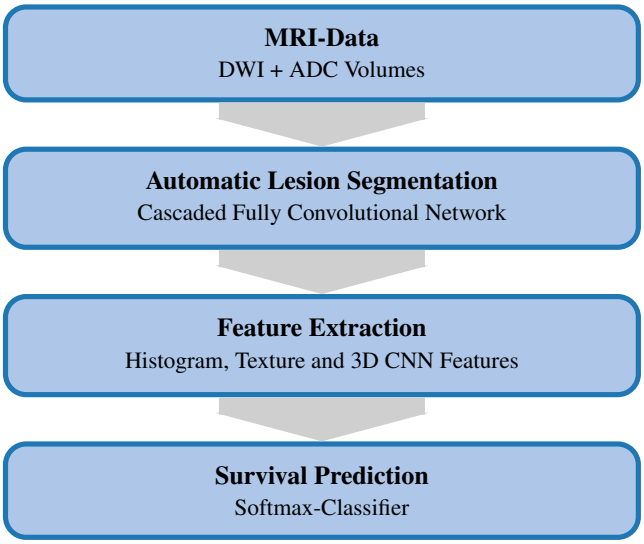


Fig. 3. Framework for lesion segmentation and survival prediction

2.3.1. *Handcrafted Features*

ADC value histograms were generated from the regions of the ADC map corresponding to the tumor ROI in the b=600 image. Histogram descriptors were obtained including mean, median, kurtosis and skewness. In addition, we extracted features representing ADC texture by calculating 3D Haralick statistics of the grey-level co-occurrence matrix [13]. We trained a k-nearest-neighbour classifier with k=4 and validated the results using 10-fold cross-validation.

2.3.2. *SurvivalNet: 3D Convolutional Neural Network*

Finally, we trained a 3D CNN to predict the survival rate in an end-to-end fashion. The SurvivalNet consists of two stacks of 3D convolution and max pooling layers, followed by 2 fully-connected layers. The 3D convolutions have 50 kernels with a kernel size of 3x3x3 pixels and a 3D spatial dropout with $p = 0.3$. The first fully-connected layer has 500 neurons. Figure 4 shows the SurvivalNet architecture. We trained the SurvivalNet from scratch using the Adadelta gradient-descent algorithm [14] at a learning rate of 0.5, $\rho = 0.95$ and $\epsilon = 1 \cdot 10^{-8}$. We employed no data-augmentation.

Table 2 shows the performance of the handcrafted features and SurvivalNet.

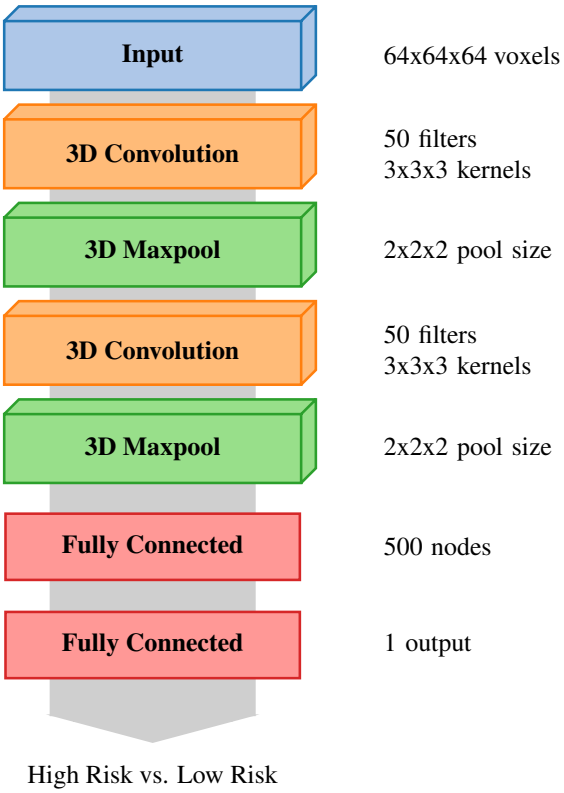


Fig. 4. The SurvivalNet is a 3D CNN that consists of two blocks of 3D convolution followed by 3D maxpooling and finally two fully connected layers.

3. RESULTS

3.1. Qualitative

The qualitative results of the automatic segmentation are depicted in figure 1. The complex and heterogeneous shape of the tumor lesions was detected and segmented in both images using our automatic segmentation algorithm. The trained model achieves a Dice overlap score of 85% in both images. The segmentation reaches a high level of specificity by classifying all lesion pixels in the image as lesions. Small false positive outliers within the liver reduce the overall accuracy and Dice score.

3.2. Quantitative

The quantitative results for our automatic segmentation method are shown in table 1. Our automatic HCC lesion segmentation algorithm achieves a Dice overlap score of 69.7% trained on DW-MRI images. The trained model is highly sensitive in recognizing HCC lesions with a Sensitivity of 91.1%, i.e. only few false negative errors occur.

Table 1. Quantitative tumor segmentation results

Method	Sensitivity [%]	Precision [%]	TNR [%]	RVD [%]	Dice [%]
Cascaded FCN on DWI	91.1	70.0	99.6	52.1	69.7

Table 2. Survival prediction results using both manual tumor segmentations and the output of the automatic tumor segmentation as inputs for the survival prediction classifier

	Features	ACC [%]	Precision [%]	Sensitivity [%]	F1-Score [%]
Manual Tumor Seg.	SurvivalNet CNN	68	69	68	65
	Histogram Features	61	62	61	60
	Texture Features: 3D Haralick	61	65	61	58
Automatic Tumor Seg.	SurvivalNet CNN	65	64	65	64
	Histogram Features	58	59	58	56
	Texture Features: 3D Haralick	61	62	62	60

Table 2 shows the quantitative results of our proposed automatic survival prediction framework. Figure 2 shows a Kaplan-Meier plot of the survival prediction results. SurvivalNet achieves higher scores on both manual and automatic segmentation compared to handcrafted features. SurvivalNet trained on manual segmentations achieves an accuracy of 68% with a Precision and Sensitivity of 69% and 68% respectively. Furthermore, SurvivalNet accomplishes a classification accuracy of 65% at a Precision and Sensitivity of 64% and 65% when trained using our automatic tumor segmentation in a fully automatic fashion.

As a final experiment, we calculated a paired Wilcoxon signed-rank test with H0: the output posterior class probabilities of SurvivalNet with manual and automatic segmentation belong to the same distribution. At $p > 0.953$, we found H0 to be confirmed, i.e. SurvivalNet produces the same results with automatic segmentation or manual segmentation.

4. CONCLUSION AND DISCUSSION

The predictive value of various imaging parameters has previously been suggested in HCC. With the growing appreciation of tumor heterogeneity as a major obstacle to treatment response, more sophisticated image analysis algorithms are required. The complexity of such data analyses, especially considering multi-parametric multimodality imaging, requires computer aided techniques. We have presented a fully automatic framework to predict survival times of HCC patients. This approach based on fully convolutional and 3D convolutional neural networks outperformed state-of-the art handcrafted features, while still achieving the same diagnostic outcome as if human expert segmentations were provided. This work may have potential applications in HCC treatment planning.

5. ACKNOWLEDGEMENT

This work was supported by the German Research Foundation (DFG) within the SFB-Initiative 824 (collaborative research center), "Imaging for Selection, Monitoring and Individualization of Cancer Therapies" (SFB824, project C6) and the BMBF project Softwarecampus. We thank NVIDIA and Amazon AWS for granting GPU and computation support.

6. REFERENCES

[1] Jacques Ferlay, Hai-Rim Shin, Freddie Bray, David Forman, Colin Mathers, and Donald Maxwell Parkin, "Estimates of worldwide burden of cancer in 2008: Globocan 2008," *International Journal of Cancer*, vol. 127, no. 12, pp. 2893–2917, 2010.

[2] European Association For The Study Of The Liver, "Easl–eortc clinical practice guidelines: management of hepatocellular carcinoma," *Journal of Hepatology*, vol. 56, no. 4, pp. 908–943, 2012.

[3] Irina Heid, Katja Steiger, Marija Trajkovic-Arsic, Marcus Settles, Manuela R Eßwein, Mert Erkan, Jorg Kleeff, Carsten Jäger, Helmut Friess, Bernhard Haller, Andreas Steingötter, Roland M Schmid, Markus Schwaiger, Ernst J Rummeny, Irene Esposito, Jens T Siveke, and Rickmer Braren, "Co-clinical assessment of tumor cellularity in pancreatic cancer," *Clinical Cancer Research*, 2016.

[4] Sang Ho Lee, Koichi Hayano, Dushyant V. Sahani, Andrew X. Zhu, and Hiroyuki Yoshida, "Kinetic textural biomarker for predicting survival of patients with advanced hepatocellular carcinoma after antiangiogenic therapy by use of baseline first-pass perfusion ct," in *Abdominal Imaging. Computational and Clinical Applications: 6th International Workshop, ABDI 2014, Held in Conjunction with MICCAI*, pp. 48–61. 2014.

[5] Jiawen Yao, Sheng Wang, Xinliang Zhu, and Junzhou Huang, "Imaging biomarker discovery for lung cancer survival prediction," in *MICCAI*, pp. 649–657. 2016.

[6] X. Zhu, J. Yao, X. Luo, G. Xiao, Y. Xie, A. Gazdar, and J. Huang, "Lung cancer survival prediction from pathological images and genetic data x2014; an integration study," in *IEEE ISBI*, 2016, pp. 1173–1176.

[7] J. Song, D. Dong, Y. Huang, Z. Liu, and J. Tian, "Association between tumor heterogeneity and overall survival in patients with non-small cell lung cancer," in *IEEE ISBI*, 2016, pp. 1249–1252.

[8] Islam Reda, Ahmed Shalaby, Mohammed Elmogy, Ahmed Aboulfotouh, Fahmi Khalifa, Mohamed Abou El-Ghar, Georgy Gimelfarb, and Ayman El-Baz, "Image-based computer-aided diagnostic system for early diagnosis of prostate cancer," in *MICCAI*, pp. 610–618. 2016.

[9] M. Shehata, F. Khalifa, A. Soliman, M. Abou El-Ghar, A. Dwyer, G. Gimel'farb, R. Keynton, and A. El-Baz, "A promising non-invasive cad system for kidney function assessment," in *MICCAI*, pp. 613–621. 2016.

[10] Wu Zhou, Lijuan Zhang, Kaixin Wang, Shuting Chen, Guangyi Wang, Zaiyi Liu, and Changhong Liang, "Malignancy characterization of hepatocellular carcinomas based on texture analysis of contrast-enhanced mr images," *Journal of Magnetic Resonance Imaging*, 2016.

[11] Patrick Ferdinand Christ, Mohamed Ezzeldin A. Elshaer, Florian Ettlinger, Sunil Tatavarty, Marc Bickel, Patrick Bilic, Markus Rempfler, Marco Armbruster, Felix Hofmann, Melvin D'Anastasi, Wieland H. Sommer, Seyed-Ahmad Ahmadi, and Bjoern H. Menze, "Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields," in *MICCAI*, pp. 415–423. 2016.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, vol. 9351, pp. 234–241. 2015.

[13] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, May 1979.

[14] Matthew D. Zeiler, "ADADELTA: an adaptive learning rate method," *CoRR*, vol. abs/1212.5701, 2012.

Human-Drone-Interaction: A Case Study to Investigate the Relation Between Autonomy and User Experience

Authoren: Patrick Ferdinand Christ, Florian Lachner, Axel Hösl, Bjoern Menze, Klaus Diepold, Andreas Butz

Abstract: Autonomous robots effectively support the human workforce in a variety of industries such as logistics or health care. With an increasing level of system autonomy humans normally have to give up control and rely on the system to react appropriately. We wanted to investigate the effects of different levels of autonomy on the User Experience (UX) and ran a case study involving autonomous flying drones. In a student competition, four teams developed four drone prototypes with varying levels of autonomy. We evaluated the resulting UX in 24 semi-structured interviews in a setting with high perceived workload (competition, autonomous vs. manual) and a non-competition setting (autonomous). The case study showed that the level of autonomy has various influences on UX, particularly in situations with high perceived workload. Based on our findings, we derive recommendations for the UX-oriented development of autonomous drones.

Publikationsdatum: 03.11.2016
Konferenz: ECCV 2016: Computer Vision – ECCV 2016 Workshops
Seiten: 238-253
Verlag: Springer International Publishing

Individuelle Leistungsbeiträge: Projektkoordination, Konzeption und Durchführung von Befragungen, Federführende Anfertigung des Manuskripts

Human-Drone-Interaction: A Case Study to Investigate the Relation Between Autonomy and User Experience

Patrick Ferdinand Christ[1,3(✉)], Florian Lachner[2,3], Axel Hösl[3], Bjoern Menze[1], Klaus Diepold[4], and Andreas Butz[2]

[1] Image-based Biomedical Modeling Group, Technical University of Munich (TUM), Munich, Germany
{patrick.christ,bjoern.menze}@tum.de
[2] Chair for Human-Computer-Interaction, University of Munich (LMU), Munich, Germany
{florian.lachner,butz}@ifi.lmu.de
[3] Center for Digital and Technology Management, TUM and LMU, Munich, Germany
axel.hoesl@ifi.lmu.de, {christ,lachner}@cdtm.de
[4] Department Electrical and Computer Engineering, Technical University of Munich (TUM), Munich, Germany
kldi@tum.de

Abstract. Autonomous robots effectively support the human workforce in a variety of industries such as logistics or health care. With an increasing level of system autonomy humans normally have to give up control and rely on the system to react appropriately. We wanted to investigate the effects of different levels of autonomy on the User Experience (UX) and ran a case study involving autonomous flying drones. In a student competition, four teams developed four drone prototypes with varying levels of autonomy. We evaluated the resulting UX in 24 semi-structured interviews in a setting with high perceived workload (competition, autonomous vs. manual) and a non-competition setting (autonomous). The case study showed that the level of autonomy has various influences on UX, particularly in situations with high perceived workload. Based on our findings, we derive recommendations for the UX-oriented development of autonomous drones.

Keywords: Human-robot interaction · Drones · Assistive technologies · User experience

1 Introduction

With an increasing technical reliability of autonomous systems, more and more human responsibilities are carried out by machines. The increasing level of autonomy shall increase the efficiency and the safety and shall simultaneously decrease

P.F. Christ and F. Lachner contributed equally.

G. Hua and H. Jégou (Eds.): ECCV 2016 Workshops, Part II, LNCS 9914, pp. 238–253, 2016.
DOI: 10.1007/978-3-319-48881-3_17

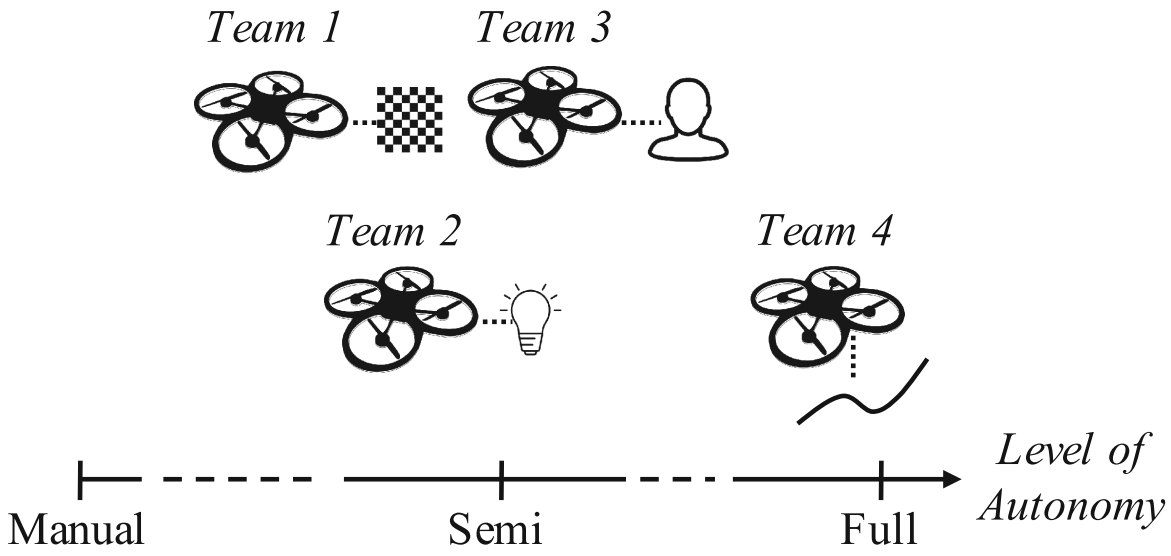


Fig. 1. Levels of autonomy of the team's drone prototypes using a checkerboard (1), a light source (2), a human face (3), and a floor marking (4) as control unit.

the human workload [1]. Traditionally, the design of autonomous systems focuses on the technical implementation aspects, especially technology-heavy disciplines such as computer vision and robotics, ranging from the system's functionality to associated sensors and software [2]. Previous research in the field of human-robot interaction, such as [3–6], already intensively analyzes the utility of computer vision technology for autonomous drones. However, these projects do not focus on the experience of interacting with vision-based drones.

In this paper, we want to take this thought further and investigate the relation between autonomy and User Experience (UX) under different levels of perceived workload. We see this consideration as a key issue in the success of future assistive technologies. To create different levels of perceived workload, we chose to conduct a case study with four teams in competitive settings using (semi)-autonomously flying drones as exemplary systems. The teams used four different control mechanisms based on state-of-the art computer vision algorithms (see Fig. 1). Hence, the research question of this study can be summarized as follows:

RQ: *"How does the user's experience when interacting with flying robots differ in situations with different perceived workload?"*

This study provides two main contributions: First, based on the analysis of four different drone prototypes, each based on an individual (semi-)autonomous interaction design, we investigated the relation between autonomy level and UX. Second, we propose concrete design recommendations for the UX-oriented design of future (semi-)autonomous flying robots.

The goal of this paper is to foster the discussion of experiences with flying robots in the computer vision community and to encourage researchers and practitioners to consider both technical and UX-related attributes when building next generation of assistive flying robots.

2 Background

The term UX is an established concept in a variety of different disciplines, ranging from ergonomics to human factors and human-computer interaction. An established approach to consolidate the variety of different perspectives is the breakdown of UX into pragmatic and hedonic product attributes [7]. However, pragmatic product attributes (i.e., the usability) of technological tools is more and more taken for granted [8]. With an increasing technological maturity researchers should put more emphasis on hedonic product attributes in order to ensure the quality of everyday actions - particularly when designing assistive technologies.

As found by Fitts [9], machines perform better than human operators in certain aspects, such as precision and efficiency, in ensuring consistent quality in repetitious tasks, or in moving heavy loads smoothly. In other aspects humans outperform machines, e.g., in improvising and using flexible procedures, in identifying visual patterns, in reasoning or in exercising judgement. Consequently, when done properly, exploiting machine benefits generally leads to a reduction of workload for users and decreased stress, fatigue, or human error. To make these benefits accessible to users, interaction with systems is necessary, yet at the same time systems need to be able to execute tasks or subtasks on their own. How independently a system can operate is generally referred to as its autonomy. The term itself, as coined in research on human-robot interaction [10], has multiple definitions in the literature [11–14] with varying characterizations. Sheridan and Verplank [14] characterized it by distinguishing ten *levels of autonomy* (LOA) ranging from 'Human does it all (1)' to'Computer acts entirely autonomously (10)' with increasing autonomy for each level. How autonomously a system can operate is determined by its design (e.g.,'Computer executes alternative if human approves (5)'). In some use cases a more or respectively less autonomous design is desirable. Therefore, flexible or adaptive autonomy approaches with a dynamically changing level of autonomy were proposed e.g., by Miller and Parasuraman [15]. Looking at the consequences for users and results when interacting with such systems, they describe an *inevitable trade-off between workload and unpredictability*: The more autonomously systems operate, the more workload[1] is taken off the user's shoulders. In consequence, however, the unpredictability of the results increases as users are no longer in control of the execution details. The more users need or want to be in control of the execution details on the other hand, the more their workload increases in turn.

Drones can serve well as a practical example in applying Sheridan and Verplanks LOA as they incorporate multiple at once. One reason for their popularity is their ease of control compared to remote controlled helicopters, for instance. This is due to their four (or more) rotor design leading to easier in-air stabilization. The stabilization is done fully autonomously by a built-in control unit (10).

[1] Hart 1988 introduced in his work NASA TLX the concept of perceived workload as a combination of mental, physical, and temporal demand as well as performance, effort, and frustration [16].

The different LOA can be used depending on usage contexts such as manual control for recording landscape or semi-autonomous tracking of and circling around a protagonist as in action sports.

3 Related Work

A range of prior work investigated interactions between humans and autonomously controlled systems (i.e., ground and aerial robots) in a variety of different settings.

With an increasing interest in the interaction between humans and autonomous systems, researchers move away from a pure analysis of input devices towards the investigation of more natural control gestures. Ende et al. [17] thereby focus on co-working tasks of technical robots, whereas Nagi et al. [18] analyze the interplay of gesture and facial recognition. Based on the analysis of human-drone interaction, Cauchard et al. [19] illustrate that natural gesture control generally lead to more personal relations to autonomous systems. The work of Ng and Sharlin [20] that examines body controls of drones inspired by falconeering gestures supports this view on natural human-drone interaction. Furthermore, Cid et al. [21], Heenan et al. [22], and Szafir et al. [23] highlight that visual feedback increases the level of empathy of human-robot interaction.

Against the background of these studies, we want to investigate how different levels of autonomy of an autonomous drone influence the interaction with the associated UX. First attempts to analyze the perception of different levels of autonomy are mentioned by Rödel et al. [24] and Hassenzahl and Klapprich [1]. These studies, however, do not comprehensively analyze the complexity of autonomous system but remain on a higher level of automation tasks (see [1]) or focus on the indication of the presumable UX of future autonomous cars (see [24]).

Based on the NASA TLX, researchers have already shown that with an increasing level of system autonomy the perceived workload decreases [16,25]. The challenge of analyzing the interaction with autonomous systems is based on the subjective interpretation of each facet of the experienced interaction, ranging from usability over workload to experience. For the course of this study we want to investigate existing measurement tools in order to derive an interview guideline that is applicable for our particular research question. The interview guideline is comprehensively explained in the next section *Methodology*.

4 Methodology

As the implementation of autonomous flying robots is still on the rise, we decided to organize a student competition in order to develop various prototypes. We chose a a Parrot AR Drone 2.0 with the goal to implement different interaction designs.

The student competition was conducted in the form of a case study. First, students from our research institution were able to sign up for a drone course.

Within this course, the students developed different prototypes. Second, the course ended in a competition, where the prototypes were put into practice. Third, we conducted interviews with the participants in order to analyze their experiences of the interaction with the drone.

4.1 Development of Prototypes

The case study was announced as a one-week student competition at our research institution[2]. The course itself consisted of two steps: Initially, the registered students were coupled in teams and had one week the develop a drone prototype. After one week, the student teams put their work into practice in three different settings, as described below.

Participants, Setting, and Task: In total, eight participants from different academic backgrounds (6x Computer Science, 1x Electrical Engineering, and 1x Communication Studies) and ages (22 to 26 $\mu = 24$) signed up for the one-week student competition without a financial reward. At the beginning of the competition, the students were randomly coupled in four teams of two. Over the course of the initial development phase, the participants were trained in python programming, image processing, computer vision, feedback control theory, state estimation, and autonomous navigation by academic and industry experts to ensure an equally distributed level of knowledge regarding the design of autonomous systems.

In the first phase of the case study the teams had to program a Parrat AR Drone 2.0 (52,5 cm x 51,5 cm), a quadcopter with an integrated HD camera, using a open-source python API[3]. The student teams were asked to process the video stream of the drone in real-time in order to fly and compete autonomously in a race at the end of the course. However, the teams were not dictated an obligatory interaction design. All four teams were told to individually develop a prototype with a desired level of autonomy at their own discretion. In the final race, each drone prototype had to pass the same predetermined track consisting of three hockey goals that were positioned in a L-shaped track. Figure 3 shows an impression of the drone race.

Prototypes: The four student teams programmed and implemented four unique types of drone interactions that cover different levels of autonomy. For the analysis in this paper, we were able to distinguish two types of autonomous interaction designs: *"Semi-autonomous"* when the drone "executes an alternative if the human approves" and *"full-autonomous"* when "the drone decides everything" - related to the LOA according to Sheridan and Verplank [14]. Figure 1 illustrates the four different drone prototypes and the associated interaction design whereas algorithm 1 exemplarily for all four teams demonstrates the algorithm of team 1 as described below.

[2] Course Information can be found at http://drones.cdtm.de.

[3] Source-code can be found at https://github.com/CDTM/Autonomous-Drones.

Team 1: Recognition of a printed checkerboard. Team 1 implemented an algorithm based on Rufli et al. [26] that enabled the drone's front camera to detect and follow the movements of a checkerboard that was printed on a piece of paper. Based on the known geometry the center of the checkerboard is found using corner and edge detection. The drone is steered and controlled as it tries to keep the centroid in the center of the image frame. Furthermore, through the identification of the outer-most square it is possible to push and pull the drone forward and backward. In order to avoid oscillation, a PID controller is used to improve the magnitude of the movement speeds. This interaction is semi-autonomous.

```
Data: Drone front camera stream
Result: Drone movement
while drone not landed do
    read current frame;
    if frame is valid then
        recognize Checkerboard;
        if Checkerboard is recognized then
            Find center of Checkerboard;
            Calculate offset of checkerboard center to camera center;
            Calculate and apply steering commands to PID Controller;
            Move Drone;
        else
            break
        end
        Drone hover;
    end
end
```

Algorithm 1. Exemplary algorithm for semi-autonomous drone interaction using a checkerboard recognition by Rufli et al. [26].

Team 2: Recognition of a color/light source. Team 2 employed an algorithm based on Comaniciu et al. [27] that allowed the drone's front camera to detect a homogeneously colored object or a light source. This mechanism had a setup phase, in which the algorithm was trained to recognize either a colored object or a light source. In the final competition, the team used a light source to control the drone. After the setup phase the drone tried to center the light-source in the image frame and follow the track of the light-source. This interaction is semi-autonomous.

Team 3: Recognition of a human face. Team 3 programmed a face detection algorithm based on Viola and Jones [28] that recognizes a human face from the drone's front camera. In this approach the drone tried to center a human face in the image frame and therefore follow the track and movements of the respective human. Furthermore, an additionally implemented emergency mode allowed the drone to keep its position through "hovering" as soon as the face recognition is interrupted. This interaction is semi-autonomous.

Team 4: Recognition of a floor marking. Team 4 implemented an algorithm based on Hart [29] that can detect and follow a colored line on the ground using the drone's bottom camera as an input device. Thereby, the drone is positioned at a certain height above the particular line. In the final race at the end of the student competition, the team used a red tape to mark the respective line on the ground. The algorithm recognized the line (i.e., the tape) and constantly tried to keep this line in the center of the bottom camera frame. As soon as the line is not centered anymore the correct angle to approach the line again is calculated. This interaction is full-autonomous.

4.2 Competition, Data Collection, and Analysis

In order to analyze the experience of interacting with flying robots in situations with differently perceived workloads we identified four suitable settings for the final race. We distinguished different perceived workloads through the setting dimensions *"Competition vs. No Competition"* and *"Manual Control vs. Autonomous Control"*. As Cauchard et al. [19] already conducted an elaborate study on manually controlled human-drone interactions in a setting without competition we concentrated on (1) *Competition/Autonomous*, (2) *No Competition/Autonomous*, and (3) *Competition/Manual Control* as described below and illustrated in Fig. 2. In all three settings, both participants of the four teams had three attempts to finish the track. As the best run of each participant counted we ended up with 24 eligible runs in total.

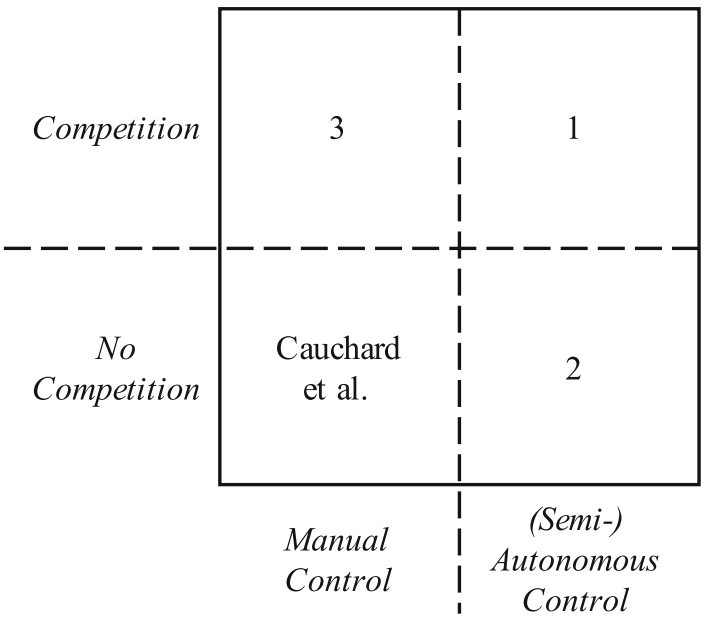


Fig. 2. Allocation of the three analyzed settings. (1) Competition/Autonomous, (2) No Competition/Autonomous and (3) Competition/Manual Control.

1. Competition/Autonomous: In the Competition/Autonomous setting the student teams had to compete in the race using the autonomous control algorithm of their drone prototype. A manual interaction would disqualify them for the current run. The best student team was rewarded with a gift.

2. No Competition/Autonomous: In the No Competition/Autonomous setting the student teams where asked to autonomously direct their drone prototype through the track. However, no time was tracked in this setting.

3. Competition/Manual Control: In the Competition/Manual Control setting the student teams had to use the official Parrot App for Smartphones to steer the drone manually through the track. The autonomous control mechanisms were deactivated in this setting. The best student team was rewarded with a gift.

After all three attempts per setting we conducted interviews with all eight participants (in total 24 interviews, each between 15 and 20 min) to analyze experience-related aspects of the interaction with the drone prototypes. Our interviews were semi-structured and audio-recorded for post-hoc analysis.

In order to meet the requirements of our research question we developed an interview guideline that served as a basis for the semi-structured interviews (see Table 1). Inspired by related work in the fields of UX, usability, and workload evaluation (as indicated in Table 1), relevant experience- and workload-related categories (e.g., "User" and "Environment") and dimensions (e.g., "Mental Demand" and "Frustration Level") and associated interview questions were developed by the first and the second author.

Post-hoc coding was conducted according to Mayring and Fenzl [30] by the second author, who has a broad experience in open-coding of interview data. Interview statements were therefore clustered according to the questions' categories (see table 1). The objective was to identify key issues across the study settings and to derive design recommendations that strengthen the linkage of computer vision and robotics and the interaction of technological tools with people.

5 Results

The next sections represent the results of our case study. First, we demonstrate the outcome of our interviews with regards to the three study settings. Thereby, we focus on the perceived workload (based on the dimension *"Competition vs. No Competition"*) as well as the participants' experiences with autonomous and manual control mechanisms (based on the dimension *"Manual Control vs. Autonomous Control"*). Second, based on these outcomes we derive design recommendations for (semi-)autonomous flying robots.

5.1 Interview Outcomes

To analyze the relation between autonomy and UX (i.e., the associated perceived workload) of vision-based drones we consolidated key findings of our interviews. These key findings allow the consequent derivation of design recommendations to understand the interaction of people with flying robots.

Table 1. Semi-structured interview guideline.

Dimension	Scale (qualitative)	Related Work
Mental Demand	Can you describe a situation that was mentally demanding for you?	[31–34]
Physical Demand	Can you describe a situation that was physically demanding for you?	
Autonomy / Independence	Can you describe a situation where you had the feeling that you directly control the drone?	
	Can you describe a situation where you had the feeling that you were not in full control over the drone?	
Competence	At what point of the course did you feel confident in performing the task?	
Enjoyment, Pleasure	Can you describe the most enjoyable aspect of performing the task?	[31–33,35,36]
Frustration Level	Can you describe the most frustrating aspect of performing the task?	
	Can you describe the most stressful aspect of performing the task?	
Perceived Ease of Use	Which part of your interaction/solution would you describe as easy to use?	
	Which part of your interaction/solution would you describe as difficult to use?	
Personal Attachment	In which way did you build a relationship to your drone?	[37]
	How did the relationship to your drone influence how you interacted with the drone?	
Performance/Outcome Satisfaction	How would you describe your performance?	[31,32,38]
	If you could optimize one thing next time, what would it be?	
Unpredictability/Error-handling	What kind of problems did you have to overcome?	[31,32,35,36,38]
	What went differently than expected and how did you handle these situations?	
Temporal Demand (hurried or rushed)	At what time of the course did you feel being rushed or hurried?	

System Feedback Enhances the Experience of interactions: All participants enjoyed interacting with their drones regardless of the respective level of autonomy. Having established a feeling of control, directing the

(semi-)autonomously controlled drone was considered as very enjoyable (setting 1 and 2). The pleasure of being in control arose either through feedback from the (semi-autonomous) drone, a tangible input device (semi-autonomous) or through a reduced workload (autonomous drone). A student from team 1 (semi-autonomous drone) mentioned: *"I think it was very enjoyable [...] that we could take very direct influence on the drone using the checkerboard. It kind of was like in the circus, where you have a tiger and you say 'jump over this' [...] and we basically made the same thing with the drone navigating it through the obstacle course" [P1].* The instant feedback from the prototype facilitated the development of a feeling of being in control. However, external factors as well as latency reduced the feeling of being in control: *"When the drone reacted to my input or my actions without much of a delay I felt confident. For example, when moving the light [source] to left or right [and] the drone also directly rotated to the left or the right, I had the feeling of complete control. So I think it is also a matter of latency" [P3].* The team that used the autonomously controlled drone (team 4), however, described the decreased workload as enjoyable, *"[The] most enjoyable moment in the race was, when the drone surprisingly went along the path without any [manual] corrections" [P7].*

Direct feedback mechanisms also positively influenced the ease of use of the prototypes: *" [The interaction] did not really need a lot of time to explain someone who has never seen this specific drone and implementation or control. You just say 'here is the checkerboard'. And even with small movements you [realize] how the drone moves and it is very easy to keep the drone on track" [P2].* However, participants from team 2 (recognition of color/light source) and team 3 (recognition of a human face) mentioned difficulties regarding the ease of use due to a lack of robustness of the associated algorithms. External factors such as different lighting, fast movements of the tracked object, and a lack of control mechanisms (e.g., a PID controller) led to difficulties in the interaction with the drone. The participants highlighted that they had difficulties *"If the background is too bright, it [did not] work [recognition of a light source]" [P4].* and while *"Holding it stable, while moving, shaking it not too much, was difficult" [P3].*

Environment Perception Influences the Feeling of control: Environmental factors played an important role in the autonomous setting with competition (setting 1). Unexpected environmental conditions, bystanders, and the orientation of the drone in space were the most prominent environmental factors as, for example, *"this direct sunlight completely misguided the drone. [...] We did not anticipate that problem" [P2].* Furthermore, *"There were a lot of faces in the room [...] and also different parts of walls were recognized as faces" [P6].* For others it was *"hard to locate where the obstacle is, relatively to the drone, because [the student] was looking at the drone and then while flying fast [one] can not really see if the path [the drone is] taking will work out or if [the drone will] touch something" [P8].* All in all, these unexpected environmental factors lowered the perceived feeling of control. In the manual setting (setting 3), the participants were less bothered by external influences. The possibility to use an additional

Fig. 3. Impressions from the autonomous drone race competition. Four student teams had to program a Parrot AR Drone 2.0 in order to fly autonomously in a drone race. In this picture a semi-autonomous interaction using face recognition is depicted.

input device even increased one student's risk tolerance: *"I would try to check whether you can even increase the speed in the setting, lower the limitations of the drone. So basically taking away safety features" [P3].*

5.2 Design Recommendations

Based on the investigation of the three different perceived workload settings, we derived three design recommendations for autonomous flying robots. The goal of these recommendations is to support a user-centered design of future autonomous flying robots and to carry on the concept of UX in the field of computer vision and robotics.

Maneuvering in 3D Space: Autonomous systems such as naval or aerial drones move in 3D space. We observed that maneuvering and interacting with a flying drone in 3D Space was mentally demanding for all participants, particularly at the beginning of each race. Adding an additional degree of freedom led to a high cognitive load, since the participants were accustomed to 2D movements, such as walking or driving a car.

Experiences from the case study: In the manual controlled setting, the participants needed a certain amount of time to familiarize themselves with the control mechanism in a 3D space. *"I think it's getting better and better the more I try. So it's really something which is dependent on my skills" [P2].* In the autonomous controlled setting, the participants reduced the complexity of the (semi-)autonomously controlled drone in 3D space by reducing the numbers of allowed movement directions. Team 2, for example, disabled the backward pitch movement of their drone to overcome the obstacles. Team 1 restricted the drone to a fixed altitude to simplify the semi-autonomous interaction. *"We lacked the controls to move the drone up and downwards. We just thought the drone will fit through the gates in the end" [P1].*

Recommendation: With an increasing number of degrees of freedom, familiarization with the control of a system becomes more time consuming. For manual

and autonomous interactions, we recommend to restrict the number of possible movements to the movements that are necessary in the respective use case. For example, one can fix or autonomously adjust the altitude of an autonomous system (e.g., of a surveillance drone) or restrict the system to one type of movement at a time. Consequently, (semi-)autonomous control mechanisms can support the handling in complex situations.

Precision, Feedback, and Latency: The interaction with autonomous systems requires a precise, direct, and instant feedback to foster the feeling of control. Latency in performing an interaction or the lack of feedback can substantially reduce the perceived feeling of control.

Experiences from the case study: We observed that for both systems, semi-autonomous and autonomous, a precise and direct feedback of the system led to a high feeling of control and consequently a positive UX. *"It was a great feeling, [...] I could feel [...] the small changes and when I changed the position of the paper [i.e., the checkerboard] it was following it" [P3].* In contrast, latency within the interaction with the drone harmed the feeling of control, although it was regained again afterwards. *"I thought it actually lost [the detection of] my face but it didn't. So again the [latency] problem solved itself by being a little bit more patient" [P7].*

Recommendation: As a conclusion, we suggest to design direct feedback mechanisms, as similarly mentioned by Cauchard et al. [19]. Moreover, the implementation of advanced and precise control procedures, such as a PID Controller, and the reduction of latency through a stable interaction design can promote a higher feeling of control and consequently a better UX.

Natural Emergency Procedures: Dealing with emergency situations is one of the key issues in designing autonomous flying robots for assistive purposes. Emergency situations are unforeseen and potentially harm people or the environment. Thus, the interaction with autonomous flying robots in an emergency situation is generally demanding. The challenge in emergency situations is based on the loss of control of the user and the consequential requirement of a suitable emergency procedure. In our case study four emergency actions were possible: direct control, immediate stop, immediate landing, and hovering (i.e., constant positioning in 3D space).

Experiences from the case study: In manual interactions we observed that in emergency situations the participants automatically used the immediate stop mechanism or the landing function. *"In the second run [of the manual competition] I first anticipated the drone's path [...] when I lost control I tried to to regain control, but then I emergency landed it" [P3].* In autonomous interactions we observed that participants resolved emergency situations initially using the hovering mechanism and later using immediate landing. *"I bumped into the goal [i.e., one of the obstacles], which was not a big problem because [...] you could just wait a few seconds, the drone hovered and you could just start again" [P1].*

Recommendation: With an increasing level of autonomy the importance of emergency considerations increases as users have to rely on the system to function correctly. Therefore, we suggest to design natural emergency handling schemes (i.e., hovering for drones) according to the level of autonomy in order to assist the user in potential breakdowns. Natural emergency procedures allow the user to realize and understand the need to interfere. Thus, a positive UX can be ensured.

6 Limitations and Future Work

This study aims to foster an multilateral discourse about autonomous systems. However, experiences and associated evaluations are subjective in nature, thus complicating generalization. Extensive and diverse studies are required to comprehensively understand users' feelings and emotions. For our case study we were able to count eight registered participants from our research institution. We asked the participants to develop an individual interaction design for a aerial robot (i.e., a flying drone) in teams of two. Thus, we ended up with four different drone prototypes, whereas the analysis of more interaction designs as well as different levels of autonomy can lead to further and more profound insights. Nevertheless, we were able to derive reasonable insights and design recommendation across all prototypes. Here, the study can serve as a basis and provide comparative data for future research.

To ensure the comparability of the experienced interactions of all participants we chose drones as the development object for all teams. As a consequence, we focused on merely one specific aspect (i.e., the relation between autonomy and UX) in our case study and neglected further peculiarities of drones, such as noise generation of the rotor blades or specific flight characteristics. Moreover, the particular study setting (i.e., participants developed the interaction design themselves) may have resulted in a higher personal attachment compared to just using the system. We therefore want to motivate other researchers to take the concept of UX-oriented, autonomous systems further to additional application domains, such as ground or naval robots.

7 Conclusion

The central issue of this study was to analyze the relation between different levels of autonomy and the associated UX. To investigate this relation, we implemented a case study in the form of a student competition and selected flying drones as exemplary autonomous systems. In the end, we were able to contrast four different human-drone interactions based on semi-structured interviews with all participants. Altogether, we derive two main contributions from this study. First, we found autonomy-specific insights on the UX of human-drone interaction. Second, we presented three design recommendations for the future design of autonomous flying robots.

In summary, we see our work as a step towards the design of UX-sensitive autonomous flying robots. We want to highlight the consideration of UX as a crucial factor and foster an ongoing discussion in the field of computer vision and robotics research.

References

1. Hassenzahl, M., Klapperich, H.: Convenient, Clean, and efficient? The experiential costs of everyday automation. In: Proceeding of NordiCHI 2014, pp. 21–30. ACM (2014)
2. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A model for types and levels of human interaction with automation. Syst. Man Cybern. Part A Syst. Hum. **30**(3), 286–297 (2000)
3. Layne, R., Hospedales, T.M., Gong, S.: Investigating open-world person re-identification using a drone. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014. LNCS, vol. 8927, pp. 225–240. Springer, Heidelberg (2015). doi:10.1007/978-3-319-16199-0_16
4. Gemert, J.C., Verschoor, C.R., Mettes, P., Epema, K., Koh, L.P., Wich, S.: Nature conservation drones for automatic localization and counting of animals. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014. LNCS, vol. 8925, pp. 255–270. Springer, Heidelberg (2015). doi:10.1007/978-3-319-16178-5_17
5. Dotenco, S., Gallwitz, F., Angelopoulou, E.: Autonomous approach and landing for a low-cost quadrotor using monocular cameras. In: Agapito, L., Bronstein, M.M., Rother, C. (eds.) ECCV 2014. LNCS, vol. 8925, pp. 209–222. Springer, Heidelberg (2015). doi:10.1007/978-3-319-16178-5_14
6. Kim, J., Lee, Y.S., Han, S.S., Kim, S.H., Lee, G.H., Ji, H.J., Choi, H.J., Choi, K.N.: Autonomous flight system using marker recognition on drone. In: 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV), pp. 1–4. IEEE (2015)
7. Hassenzahl, M.: User experience (UX): towards an experiential perspective on product quality. In: Proceeding of IHM 2008, pp. 11–15 (2008)
8. Pine, J., Gilmore, J.H.: Welcome to the experience economy. Harvard Bus. Rev. **76**(4), 97–105 (1998)
9. Fitts, P.M.: Human Engineering for an Effective Air-Navigation and Traffic-Control System. National Research Council, Division of Anthropology and Psychology, Committee on Aviation Psychology (1951)
10. Goodrich, M.A., Schultz, A.C.: Human-robot interaction: a survey. Foundations Trends Hum. Comput. Interact. **1**(3), 203–275 (2007)
11. Albus, J.S.: Outline for a theory of intelligence. IEEE Trans. Syst. Man Cybern. **21**(3), 473–509 (1991)
12. Beavers, G., Hexmoor, H.: Types and limits of agent autonomy. In: Nickles, M., Rovatsos, M., Weiss, G. (eds.) AUTONOMY 2003. LNCS (LNAI), vol. 2969, pp. 95–102. Springer, Heidelberg (2004). doi:10.1007/978-3-540-25928-2_8
13. Crandall, J.W., Goodrich, M., Olsen Jr., D.R., Nielsen, C.W., et al.: Validating human-robot interaction schemes in multitasking environments. IEEE Trans. Syst. Man Cybern. **35**(4), 438–449 (2005). others:
14. Sheridan, T.B., Verplank, W.L.: Human and Computer Control of Undersea Teleoperators. Technical report, DTIC Document (1978)
15. Miller, C.A., Parasuraman, R.: Designing for flexible interaction between humans and automation: delegation interfaces for supervisory control. Hum. Factors **49**(1), 57–75 (2007)

16. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv. Psychol. **52**, 139–183 (1988)
17. Ende, T., Haddadin, S., Parusel, S., Wüsthoff, T., Hassenzahl, M., Albu-Schäffer, A.: A human-centered approach to robot gesture based communication within collaborative working processes. In: Proceeding of IROS 2011, pp. 3367–3374 (2011)
18. Nagi, J., Giusti, A., Di Caro, G.A., Gambardella, L.M.: Human control of UAVs using face pose estimates and hand gestures. In: Proceeding of HRI 2014, pp. 1–2. ACM (2014)
19. Cauchard, J.R., Jane, L.E., Zhai, K.Y., Landay, J.A.: Drone and me: an exploration into natural human-drone interaction. In: Proceeding UbiComp 2015, pp. 361–365. ACM (2015)
20. Ng, W.S., Sharlin, E.: Collocated interaction with flying robots. In: Proceeding IEEE RO-MAN 2011, pp. 143–149 (2011)
21. Cid, F., Manso, L.J., Calderita, L.V., Sánchez, A., Nuñez, P.: Engaging human-to-robot attention using conversational gestures and lip-synchronization. J. Phys. Agents **6**(1), 3–10 (2012)
22. Heenan, B., Greenberg, S., Manesh, S.A., Sharlin, E.: Designing social greetings in human robot interaction. In: Proceeding DIS 2014, pp. 855–864. ACM (2014)
23. Szafir, D., Mutlu, B., Fong, T.: Communicating directionality in flying robots. In: Proceeding HRI 2015, vol. 2, pp. 19–26 (2015)
24. Rödel, C., Stadler, S., Meschtscherjakov, A., Tscheligi, M.: Towards Autonomous Cars: the effect of autonomy levels on acceptance and user experience. In: Proceeding AutoUI 2014, Seattle, WA, USA, pp. 1–8. ACM (2014)
25. Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., Goodrich, M.: Common metrics for human-robot interaction. In: Proceedings of the 1st ACM/IEEE International Conference on Human-Robot Interaction, pp. 33–40. ACM (2006)
26. Rufli, M., Scaramuzza, D., Siegwart, R.: Automatic detection of checkerboards on blurred and distorted images. In: Proceeding IROS 2008, pp. 3121–3126 (2008)
27. Comaniciu, D., Ramesh, V., Meer, P.: Real-time tracking of non-rigid objects using mean shift. IEEE Conf. Comput. Vis. Pattern Recogn. **2**(7), 142–149 (2000)
28. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2001)
29. Hart, P.E.: Use of the hough transformtion to detect lines. Commun. ACM **15**, 11–15 (1972)
30. Mayring, P., Fenzl, T.: Qualitative Inhaltsanalyse. Springer, Wiesbaden (2014)
31. Hart, S.G.: Nasa-Task Load Index (NASA-TLX); 20 Years Later. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 50(9) (2006)
32. Reid, G.B., Potter, S.S., Bressler, J.R.: Subjective Workload Assessment Technique (SWAT): A User's Guide. Technical report (1989)
33. Sheldon, K.M., Elliot, A.J., Kim, Y., Kasser, T.: What is satisfying about satisfying events? Testing 10 candidate psychological needs. J. Personal. Soc. Psychol. **80**(2), 325–339 (2001)
34. Brooke, J.: SUS - a quick and dirty usability scale. Usability Eval. Ind. **189**(195), 4–7 (1996)
35. Laugwitz, B., Held, T., Schrepp, M.: Construction and evaluation of a user experience questionnaire. In: Holzinger, A. (ed.) USAB 2008. LNCS, vol. 5298, pp. 63–76. Springer, Heidelberg (2008)
36. Lund, A.M.: Measuring usability with the USE questionnaire. Usability Interface **8**(2), 3–6 (2001)

37. Madsen, M., Gregor, S.: Measuring human-computer trust. In: Proceeding ACIS 2000, pp. 6–8 (2000)
38. Lewis, J.: IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int. J. Hum.-Comput. Interact. **7**(1), 57–78 (1995)

Literaturverzeichnis

[1] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D'Anastasi, W. H. Sommer, S.-A. Ahmadi, and B. H. Menze. Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields. In *Medical Image Computing and Computer-Assisted Intervention*, pages 415–423. Springer International Publishing, 2016.

[2] P. F. Christ, F. Ettlinger, G. Kaissis, S. Schlecht, F. Ahmaddy, F. Grün, A. Valentinitsch, S.-A. Ahmadi, R. Braren, and B. H. Menze. SurvivalNet: Predicting patient survival from diffusion weighted magnetic resonance images using cascaded fully convolutional and 3D convolutional neural networks. In *IEEE International Symposium on Biomedical Imaging*. IEEE, 2017.

[3] P. F. Christ, F. Lachner, A. Hösl, B. H. Menze, K. Diepold, and A. Butz. Human-Drone-Interaction: A Case Study to Investigate the Relation Between Autonomy and User Experience. In *European Conference on Computer Vision Workshops*, pages 238–253. Springer International Publishing, 2016.

[4] P. F. Christ, S. Schlecht, F. Ettlinger, S.-A. Ahmadi, K. Diepold, and B. H. Menze. Diabetes60 - Inferring Bread Units From Food Images Using Fully Convolutional Neural Networks. *Unveröffentlichtes Manuskript*, 2017.

[5] P. F. Christ, F. Ettlinger, F. Grün, M. E. A. Elshaera, J. Lipkova, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, F. Hofmann, M. D. Anastasi, S.-A. Ahmadi, G. Kaissis, J. Holch, W. Sommer, R. Braren, V. Heinemann, and B. H. Menze. Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks. *Unveröffentlichtes Manuskript*, 2017.

[6] J. Lipková, M. Rempfler, P. F. Christ, J. Lowengrub, and B. H. Menze. Automated Unsupervised Segmentation of Liver Lesions in CT scans via Cahn-Hilliard Phase Separation. *Unveröffentlichtes Manuskript*, 2017.

[7] R. Hooke. *Micrographia: or some physiological descriptions of minute bodies made by magnifying glasses, with observations and inquiries thereupon*. 1665.

[8] P. F. Christ. Moderne Methoden, Algorithmen und Simulationen für 13C metabolische Kernspinresonanz. Master thesis, Technische Universität München, 2012.

[9] W. C. Röntgen. Über eine neue Art von Strahlen. *Annalen der Physik*, 300(1):1–11, 1898.

[10] A. Haase, G. Landwehr, and E. Umbach. *Röntgen Centennial: X-rays in Natural and Life Sciences*. World Scientific, 1997.

[11] J. Radon. On the determination of functions from their integral values along certain manifolds. *IEEE Transactions on Medical Imaging*, 5(4):170–176, 1986.

[12] A. Grillenberger and E. Fritsch. *Computertomographie: Einführung in ein modernes bildgebendes Verfahren*. Facultas Universitätsverlag, 2012.

[13] W. Kalender. *Computertomographie: Grundlagen, Gerätetechnologie, Bildqualität, Anwendungen*. Publicis-MCD-Verlag, 2000.

[14] E. C. Beckmann. CT scanning the early days. *The British Journal of Radiology*, 79(937):5–8, 2006.

[15] Barmer GEK. Anzahl der Untersuchungen mit Computertomographie und Magnetresonanztomographie in Deutschland. `https://de.statista.com/statistik/daten/studie/172699/umfrage/ct-und-mrt---anzahl-der-untersuchungen-2009/`, 2009. (abgerufen am 20.05.2017).

[16] F. Bloch. Nuclear induction. *Physical review*, 70(7-8):460, 1946.

[17] E. M. Purcell, H. Torrey, and R. V. Pound. Resonance absorption by nuclear magnetic moments in a solid. *Physical review*, 69(1-2):37, 1946.

[18] R. Damadian. Tumor detection by nuclear magnetic resonance. *Science*, 171:1151–1153, 1971.

[19] P. C. Lauterbur. Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature*, 1973.

[20] P. Mansfield. Multi-planar image formation using NMR spin echoes. *Journal of Solid State Physics*, 10(3):L55, 1977.

[21] A. Haase, J. Frahm, D. Matthaei, W. Hanicke, and K.-D. Merboldt. FLASH imaging. Rapid NMR imaging using low flip-angle pulses. *Journal of Magnetic Resonance*, 67(2):258–266, 1986.

[22] R. Damadian. First MRI Scan of a human by Raymond Damadian. `http://www.two-views.com/images/ics-45.jpg`, 1977. (abgerufen am 20.05.2017).

[23] T. Heimann, B. Van Ginneken, M. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, A. Beck, C. Becker, R. Beichel, G. Bekes, F. Bello, G. Binnig, H. Bischof, A. Bornik, P. Cashman, Y. Chi, A. Córdova, B. Dawant, M. Fidrich, J. Furst, D. Furukawa, L. Grenacher, J. Hornegger, D. Kainmüller, R. Kitney, H. Kobatake, H. Lamecker, T. Lange, J. Lee, B. Lennon, R. Li, S. Li, H. Meinzer, G. Németh, D. Raicu, A. Rau,

E. Van Rikxoort, M. Rousson, L. Ruskó, K. Saddi, G. Schmidt, D. Seghers, A. Shimizu, P. Slagmolen, E. Sorantin, G. Soza, R. Susomboon, J. Waite, A. Wimmer, and I. Wolf. Comparison and evaluation of methods for liver segmentation from ct datasets. *IEEE Transactions on Medical Imaging*, 28(8):1251–1265, 2009.

[24] J. E. Niederhuber, J. O. Armitage, J. H. Doroshow, M. B. Kastan, and J. E. Tepper. *Abeloff's clinical oncology*. Elsevier Health Sciences, 2013.

[25] G. Krombach, A. Mahnken, C. Alt, U. Attenberger, and T. Franiel. *Radiologische Diagnostik Abdomen und Thorax: Bildinterpretation unter Berücksichtigung anatom. Landmarken u. klin. Symptome*. Thieme, 2015.

[26] J. Ferlay, H.-R. Shin, F. Bray, D. Forman, C. Mathers, and D. M. Parkin. Estimates of worldwide burden of cancer in 2008: Globocan 2008. *International Journal of Cancer*, 127(12):2893–2917, 2010.

[27] F. Ettlinger. Neuronale Netze zur Klassifizierung und Segmentierung von Medizinischen Bilddaten. Master thesis, Technische Universität München, 2017.

[28] European Association For The Study Of The Liver. EASL–EORTC clinical practice guidelines: management of hepatocellular carcinoma. *Journal of Hepatology*, 56(4):908–943, 2012.

[29] E. Eisenhauer, P. Therasse, J. Bogaerts, L. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, L. Rubinstein, L. Shankar, L. Dodd, R. Kaplan, D. Lacombe, and J. Verweij. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *European Journal of Cancer*, 45(2):228–247, 2009.

[30] J. H. Rothe, C. Grieser, L. Lehmkuhl, D. Schnapauff, C. P. Fernandez, M. H. Maurer, A. Mussler, B. Hamm, T. Denecke, and I. G. Steffen. Size determination and response assessment of liver metastases with computed tomography— Comparison of RECIST and volumetric algorithms . *European Journal of Radiology*, 82(11):1831 – 1839, 2013.

[31] M. L. Giger, H.-P. Chan, and J. Boone. Anniversary Paper: History and status of CAD and quantitative image analysis: The role of Medical Physics and AAPM. *Medical Physics*, 35(12):5799–5820, 2008.

[32] W. M. Wells. Medical Image Analysis – past, present, and future. *Medical Image Analysis*, 33:4 – 6, 2016.

[33] L. B. Lusted. Medical Electronics. *New England Journal of Medicine*, 252(14):580–585, 1955.

[34] G. S. Lodwick, T. E. Keats, and J. P. Dorst. The Coding of Roentgen Images for Computer Analysis as Applied to Lung Cancer. *Radiology*, 81(2):185–200, 1963.

[35] H. Becker, W. Nettleton, P. Meyers, J. Sweeney, and C. Nice. Digital computer determination of a medical diagnostic index directly from chest X-ray images. *IEEE Transactions on Biomedical Engineering*, (3):67–72, 1964.

[36] P. H. Meyers, C. M. Nice Jr, H. C. Becker, W. J. Nettleton Jr, J. W. Sweeney, and G. R. Meckstroth. Automated computer analysis of radiographic images 1. *Radiology*, 83(6):1029–1034, 1964.

[37] J.-I. Toriwaki, Y. Suenaga, T. Negoro, and T. Fukumura. Pattern recognition of chest x-ray images. *Computer Graphics and Image Processing*, 2(3-4):252–271, 1973.

[38] W. A. Kalender, W. Seissler, E. Klotz, and P. Vock. Spiral volumetric ct with single-breath-hold technique, continuous transport, and continuous scanner rotation. *Radiology*, 176(1):181–183, 1990.

[39] T. Kooi, G. Litjens, B. van Ginneken, A. Gubern-Mérida, C. I. Sánchez, R. Mann, A. den Heeten, and N. Karssemeijer. Large scale deep learning for computer aided detection of mammographic lesions. *Medical image analysis*, 35:303–312, 2017.

[40] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017.

[41] X. Deng and G. Du. Editorial: 3D segmentation in the clinic: a grand challenge II liver tumor segmentation. In *MICCAI Workshop*, 2008.

[42] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M. A. Weber, T. Arbel, B. B. Avants, N. Ayache, P. Buendia, D. L. Collins, N. Cordier, J. J. Corso, A. Criminisi, T. Das, H. Delingette, C. Demiralp, C. R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N. M. John, E. Konukoglu, D. Lashkari, J. A. Mariz, R. Meier, S. Pereira, D. Precup, S. J. Price, T. R. Raviv, S. M. S. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H. C. Shin, J. Shotton, C. A. Silva, N. Sousa, N. K. Subbanna, G. Szekely, T. J. Taylor, O. M. Thomas, N. J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D. H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, and K. V. Leemput. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 2015.

[43] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[44] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition*, pages 248–255. IEEE, 2009.

[45] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. *Computer Vision and Pattern Recognition*, 2015.

[46] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 88(2):303–338, 2010.

[47] V. Heinemann, L. F. von Weikersthal, T. Decker, A. Kiani, U. Vehling-Kaiser, S.-E. Al-Batran, T. Heintges, C. Lerchenmüller, C. Kahl, G. Seipelt, et al. FOLFIRI plus cetuximab versus FOLFIRI plus bevacizumab as first-line treatment for patients with metastatic colorectal cancer (FIRE-3): a randomised, open-label, phase 3 trial. *The Lancet Oncology*, 15(10):1065–1075, 2014.

[48] M. Ezz. Automatic Liver and lesion segmentation in 3D CT Images using Cascaded Convolutional Neural Networks. Master thesis, Technische Universität München, 2016.

[49] C. M. Bishop. Pattern recognition. *Machine Learning*, 128:1–58, 2006.

[50] I. Goodfellow, Y. Bengio, and A. Courville. Deep learning. MIT Press, 2016.

[51] S. Luo, X. Li, and J. Li. Review on the methods of automatic liver segmentation from abdominal images. *Journal of Computer and Communications*, 2(02):1, 2014.

[52] A. Choudhary, N. Moretto, F. P. Ferrarese, and G. A. Zamboni. An entropy based multi-thresholding method for semi-automatic segmentation of liver tumors. In *MICCAI Workshop*, volume 41, pages 43–49, 2008.

[53] M. Kobashi and L. G. Shapiro. Knowledge-based organ identification from CT images. *Pattern Recognition*, 28(4):475–491, 1995.

[54] M. Mancas, B. Gosselin, and B. Macq. Segmentation using a region-growing thresholding. In *Electronic Imaging 2005*, pages 388–398. International Society for Optics and Photonics, 2005.

[55] R. Adams and L. Bischof. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6):641–647, 1994.

[56] D. Wong, J. Liu, Y. Fengshou, Q. Tian, W. Xiong, J. Zhou, Y. Qi, T. Han, S. Venkatesh, and S.-c. Wang. A semi-automated method for liver tumor segmentation based on 2D region growing with knowledge-based constraints. In *MICCAI Workshop*, volume 41, page 159, 2008.

[57] L. Rusko, G. Bekes, G. Nemeth, and M. Fidrich. Fully automatic liver segmentation for contrast-enhanced CT images. In *MICCAI Workshop*, volume 2, 2007.

[58] F. Liu, B. Zhao, P. K. Kijewski, L. Wang, and L. H. Schwartz. Liver segmentation for CT images using GVF snake. *Medical physics*, 32(12):3699–3706, 2005.

[59] T. Heimann, H.-P. Meinzer, and I. Wolf. A statistical deformable model for the segmentation of liver CT volumes. *3D Segmentation in the clinic: A grand challenge*, pages 161–166, 2007.

[60] D. Kainmüller, T. Lange, and H. Lamecker. Shape constrained automatic segmentation of the liver based on a heuristic intensity model. In *MICCAI Workshop*, pages 109–116, 2007.

[61] H. Lamecker, T. Lange, and M. Seebass. *Segmentation of the liver using a 3D statistical shape model*. Konrad-Zuse-Zentrum für Informationstechnik Berlin, 2004.

[62] L. Massoptier and S. Casciaro. Fully automatic liver segmentation through graph-cut technique. In *IEEE International Conference of the Engineering in Medicine and Biology Society*, pages 5243–5246. IEEE, 2007.

[63] Y. Boykov and G. Funka-Lea. Graph cuts and efficient nd image segmentation. *International Journal of Computer Vision*, 70(2):109–131, 2006.

[64] A. Afifi and T. Nakaguchi. Liver segmentation approach using graph cuts and iteratively estimated shape and intensity constrains. In *Medical Image Computing and Computer-Assisted Intervention*, pages 395–403. Springer, 2012.

[65] D.-Y. Tsai and N. Tanahashi. Neural-network-based boundary detection of liver structure in CT images for 3D visualization. In *IEEE International Conference on Computational Intelligence*, volume 6, pages 3484–3489. IEEE, 1994.

[66] J. E. Koss, F. Newman, T. Johnson, and D. Kirch. Abdominal organ segmentation using texture transforms and a hopfield neural network. *IEEE Transactions on Medical imaging*, 18(7):640–648, 1999.

[67] D. Zikic, B. Glocker, E. Konukoglu, A. Criminisi, C. Demiralp, J. Shotton, O. Thomas, T. Das, R. Jena, and S. Price. Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR. In *Medical Image Computing and Computer-Assisted Intervention*, pages 369–376. Springer Berlin/Heidelberg, 2012.

[68] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[69] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[70] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition*, volume 1, pages 886–893. IEEE, 2005.

[71] I. Heid, K. Steiger, M. Trajkovic-Arsic, M. Settles, M. R. Eßwein, M. Erkan, J. Kleeff, C. Jäger, H. Friess, B. Haller, et al. Co-clinical assessment of tumor cellularity in pancreatic cancer. *Clinical Cancer Research*, 23(6):1461–1470, 2017.

[72] S. H. Lee, K. Hayano, D. V. Sahani, A. X. Zhu, and H. Yoshida. Kinetic Textural Biomarker for Predicting Survival of Patients with Advanced Hepatocellular Carcinoma After Antiangiogenic Therapy by Use of Baseline First-Pass Perfusion CT. In *MICCAI Workshop*, pages 48–61, 2014.

[73] J. Yao, S. Wang, X. Zhu, and J. Huang. Imaging biomarker discovery for lung cancer survival prediction. In *Medical Image Computing and Computer-Assisted Intervention*, pages 649–657, 2016.

[74] X. Zhu, J. Yao, X. Luo, G. Xiao, Y. Xie, A. Gazdar, and J. Huang. Lung cancer survival prediction from pathological images and genetic data; An integration study. In *IEEE International Symposium of Biomedical Imaging*, pages 1173–1176, 2016.

[75] W. Zhou, L. Zhang, K. Wang, S. Chen, G. Wang, Z. Liu, and C. Liang. Malignancy characterization of hepatocellular carcinomas based on texture analysis of contrast-enhanced MR images. *Journal of Magnetic Resonance Imaging*, 2016.

[76] R. M. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, 1979.

[77] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *International Conference on Computer Vision*, pages 1026–1034, 2015.

[78] W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[79] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

[80] S. Schlecht. Methods for automated visual assessment of food using deep convolutional neural networks. Master thesis, Technische Universität München, 2017.

[81] M. Minsky and S. Papert. *Perceptrons*. MIT press, 1988.

[82] P. J. Werbos. Beyond regression : new tools for prediction and analysis in the behavioral sciences. Master's thesis.

[83] K. Hornik. Approximation capabilities of multilayer feedforward networks. *Neural networks*, 4(2):251–257, 1991.

[84] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[85] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

[86] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Computer Vision and Pattern Recognition*, pages 2818–2826, 2016.

[87] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[88] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *ArXiv e-prints*, abs/1409.1556, 2014.

[89] LISA lab. Convolutional Neural Network of LeNet. `http://deeplearning.net/tutorial/_images/mylenet.png`. (abgerufen am 27.05.2017).

[90] R. Wolf and J. C. Platt. Postal address block location using a convolutional locator network. *Advances in Neural Information Processing Systems*, pages 745–745, 1994.

[91] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle. Brain tumor segmentation with deep neural networks. *Medical Image Analysis*, 35:18 – 31, 2017.

[92] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. In *Medical Image Computing and Computer-Assisted Intervention*, volume 16, pages 246–253, 2013.

[93] H. R. Roth, L. Lu, A. Seff, K. M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, and R. M. Summers. A new 2.5 d representation for lymph node detection using random sets of deep convolutional neural network observations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 520–527. Springer International Publishing, 2014.

[94] F. Milletari, S.-A. Ahmadi, C. Kroll, A. Plate, V. Rozanski, J. Maiostre, J. Levin, O. Dietrich, B. Ertl-Wagner, K. Bötzel, et al. Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound. *Computer Vision and Image Understanding*, 2017.

[95] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, volume 9351, pages 234–241, 2015.

[96] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal. The importance of skip connections in biomedical image segmentation. In *International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, pages 179–187. Springer, 2016.

[97] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3D U-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 424–432, 2016.

[98] F. Milletari, N. Navab, and S.-A. Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *International Conference on 3D Vision*, pages 565–571. IEEE, 2016.

[99] G. Li, X. Chen, F. Shi, W. Zhu, J. Tian, and D. Xiang. Automatic liver segmentation based on shape constraints and deformable graph cut in CT images. *IEEE Transactions on Image Processing*, 24(12):5315–5329, 2015.

[100] G. Chartrand, T. Cresson, R. Chav, A. Gotra, A. Tang, and J. DeGuise. Semi-automated liver CT segmentation using Laplacian meshes. In *International Journal of Computer Vision*, pages 641–644. IEEE, 2014.

[101] C. Li, X. Wang, S. Eberl, M. Fulham, Y. Yin, J. Chen, and D. D. Feng. A likelihood and local constraint level set model for liver tumor segmentation from CT volumes. *IEEE Transactions on Biomedical Engineering*, 60(10):2967–2977, 2013.

[102] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[103] I. Reda, A. Shalaby, M. Elmogy, A. Aboulfotouh, F. Khalifa, M. A. El-Ghar, G. Gimelfarb, and A. El-Baz. Image-based computer-aided diagnostic system for early diagnosis of prostate cancer. In *Medical Image Computing and Computer-Assisted Intervention*, pages 610–618, 2016.

[104] J. Song, D. Dong, Y. Huang, Z. Liu, and J. Tian. Association between tumor heterogeneity and overall survival in patients with non-small cell lung cancer. In *IEEE International Symposium of Biomedical Imaging*, pages 1249–1252, 2016.

[105] M. Shehata, F. Khalifa, A. Soliman, M. A. El-Ghar, A. Dwyer, G. Gimel'farb, R. Keynton, and A. El-Baz. A promising non-invasive cad system for kidney function assessment. In *Medical Image Computing and Computer-Assisted Intervention*, pages 613–621, 2016.

[106] L. Bossard, M. Guillaumin, and L. Van Gool. Food-101–mining discriminative components with random forests. In *European Conference on Computer Vision*, pages 446–461. Springer, 2014.

[107] M. Chen, K. Dhingra, W. Wu, L. Yang, R. Sukthankar, and J. Yang. Pfid: Pittsburgh fast-food image dataset. In *IEEE International Conference on Image Processing*, pages 289–292. IEEE, 2009.

[108] A. Meyers, N. Johnston, V. Rathod, A. Korattikara, A. Gorban, N. Silberman, S. Guadarrama, G. Papandreou, J. Huang, and K. P. Murphy. Im2calories: towards an automated mobile vision food diary. In *International Conference of Computer Vision*, pages 1233–1241, 2015.

[109] T. B. Sheridan and W. L. Verplank. Human and computer control of undersea teleoperators. Technical report, DTIC Document, 1978.

[110] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D'Anastasi, W. H. Sommer, S.-A. Ahmadi, and B. H. Menze. Quellcode zu Cascaded Fully Convolutional Neural Networks. `https://github.com/IBBM/Cascaded-FCN`, 2016. (abgerufen am 27.05.2017).

[111] L. Bi, J. Kim, A. Kumar, and D. Feng. Automatic liver lesion detection using cascaded deep residual networks. *ArXiv e-prints*, abs/1704.02703, 2017.

[112] G. Chlebus, H. Meine, J. Hendrik Moltz, and A. Schenk. Neural Network-Based Automatic Liver Tumor Segmentation With Random Forest-Based Candidate Filtering. *ArXiv e-prints*, abs/1706.00842, 2017.

[113] X. Han. Automatic liver lesion segmentation using A deep convolutional neural network method. *ArXiv e-prints*, abs/1704.07239, 2017.

[114] P. F. Christ, F. Lachner, L. Altenmüller, D. Durner, P. Fritzen, M. Frohlich, J. Gebauer, G. Christian, V. Hauzeneder, C. Helm, M. Jahnen, C. Munch, D. Paiva, M. Patz, F. Cordona, I. Poppek, L. Rambold, F. Schaule, H. Schmidtchen, M. Wiggert, A. Hösl, B. H. Menze, K. Diepold, and A. Butz. Quellcode zu Human Drone Interaction. `https://github.com/PatrickChrist/CDTM-Deep-Learning-Drones`, 2016. (abgerufen am 27.05.2017).

Anhang

Diabetes60 - Inferring Bread Units From Food Images Using Fully Convolutional Neural Networks

Authoren: Patrick Ferdinand Christ, Sebastian Schlecht, Florian Ettlinger, Felix Grün, Christoph Heinle, Sunil Tatavatry, Seyed-Ahmad Ahmadi, Klaus Diepold und Bjoern H. Menze

Abstract: In this paper we propose a challenging new computer vision task of inferring Bread Units (BUs) from food images. Assessing nutritional information and nutrient volume from a meal is an important task for diabetes patients. At the moment, diabetes patients learn the assessment of BUs on a scale of one to ten, by learning correspondence of BU and meals from textbooks. We introduce a large scale data set of around 9k different RGB-D images of 60 western dishes acquired using a Microsoft Kinect v2 sensor. We recruited 20 diabetes patients to give expert assessments of BU values to each dish based on several images. For this task, we set a challenging baseline using state-of-the-art CNNs and evaluated it against the performance of human annotators. In our work we present a CNN architecture to infer the depth from RGB-only food images to be used in BU regression such that the pipeline can operate on RGB data only and compare its performance to RGB-D input data. We show that our inferred depth maps from RGB images can replace RGB-D input data at high significance for the BU regression task. In its best configuration, our proposed method achieves a $RMSE$ of 1.53 BUs using RGB and inferred depth. Considering the variability among the raters themselves of $RMSE = 0.89$, we can show that our baseline method with depth prediction can extract reasonable nutritional information from RGB image data only.

Individuelle Leistungsbeiträge: Projektkoordination, Datenakquise und Datenaufbereitung, Konzeption und Durchführung von Experimenten, Federführende Anfertigung des Manuskripts

Unveröffentlichtes Manuskript

Diabetes60 - Inferring Bread Units From Food Images Using Fully Convolutional Neural Networks

Patrick Ferdinand Christ*†[1], Sebastian Schlecht*[2], Florian Ettlinger[1], Felix Grün[1], Christoph Heinle[2], Sunil Tatavatry[2], Seyed-Ahmad Ahmadi[3], Klaus Diepold[2], and Bjoern H. Menze[1]

[1]Department for Computer Science, Technical University of Munich, Arccistrasse 21, 80333 Munich
[2]Department for Electric Engineering, Technical University of Munich, Arccistrasse 21, 80333 Munich
[3]Department for Neurology, University Hospital Grosshadern, Marchioninistrasse 15, 81377 Munich

Abstract

In this paper we propose a challenging new computer vision task of inferring Bread Units (BUs) from food images. Assessing nutritional information and nutrient volume from a meal is an important task for diabetes patients. At the moment, diabetes patients learn the assessment of BUs on a scale of one to ten, by learning correspondence of BU and meals from textbooks. We introduce a large scale data set of around 9k different RGB-D images of 60 western dishes acquired using a Microsoft Kinect v2 sensor. We recruited 20 diabetes patients to give expert assessments of BU values to each dish based on several images. For this task, we set a challenging baseline using state-of-the-art CNNs and evaluated it against the performance of human annotators. In our work we present a CNN architecture to infer the depth from RGB-only food images to be used in BU regression such that the pipeline can operate on RGB data only and compare its performance to RGB-D input data. We show that our inferred depth maps from RGB images can replace RGB-D input data at high significance for the BU regression task. In its best configuration, our proposed method achieves a $RMSE$ of 1.53 BUs using RGB and inferred depth. Considering the variability among the raters themselves of $RMSE = 0.89$, we can show that our baseline method with depth prediction can extract reasonable nutritional information from RGB image data only.

*Authors contributed equally
†Corresponding address: patrick.christ@tum.de

1. Introduction

1.1. Motivation

Diabetes mellitus is one of the most common chronic diseases worldwide and continues to increase from 285 million today to 439 million diseased people in 2030, as changing lifestyles lead to reduced physical activity, and increased obesity [34]. For diabetic patients an accurate caloric assessment of their nutritional intake is needed to regulate their dysfunctional blood sugar cycle. Diabetologists introduced a simplified scheme: the bread units or carbohydrate units to assess the nutritional intake of a meal. One bread unit corresponds to a quantity of food containing 12-15g of digestible i.e. blood-sugar-effective carbohydrates present in different forms of sugar or starch [39]. Diabetes patients learn the assessment of bread units (BU) by learning correspondence between BU and meals from textbooks and personal experience. Apart from experience, the process of estimating one's personal caloric intake may additionally require holistic knowledge about nutrition. Yet unknown dishes' BUs may be difficult to estimate, local customs in food preparation that are not visually apparent, e.g. preparing spaghetti with butter versus sunflower oil, may lead to additional uncertainty for experienced diabetes patients. Furthermore, there is a high uncertainty and danger of miscalculation for patients new to the disease. Digital support systems can be a way to provide guidance and help in those situations. Especially children could benefit significantly, due to their initially limited knowledge about their disease and nutritional values of food. Also, around 5% of pregnancies coincide with a short-term gestational diabetes mellitus (GDM) with potential harm for the unborn baby. With such a sudden onset GDM, affected pregnant women could also highly benefit from a computer aided diabetes assessment system [13]. Even though BU estima-

tion is a task that is very specific to diabetes, estimating the amount of carbohydrates and other micro nutrients is done in many more contexts like sports or weight-loss. A healthy diet is described not only by the kind of dish and its ingredients, but also by the amount which is consumed. In those cases, a digital system which processes meal images and derives rich information could provide additional support to reduce the effort of diets and better engage users in a healthy lifestyle. In this work we want to take a step towards computer aided nutrition assessment.

1.2. Related Work

Food Computer aided assessment of food and nutritional information of meals is an uprising research field in the computer vision community. Previous work can be categorized into meal classification, segmentation and caloric assessment. Public datasets so far focused on food classification such as Food101 [3], PFID [6], UNICT-FD889 [12], VIREO172 [5] and UECFOOD-100 [23]. Meyer et al. 2015 collected a 3D food dataset for assessing calories, but did neither publish their 3D data nor food classification data [25]. In the past, classical hand crafted features have been extracted to classify meals, ingredients or restaurant-specific multi-labels [3, 14]. Recently, deep convolutional neural network based methods are gaining also popularity in food classification [5, 21]. [7, 14, 25, 5] applied deep learning based segmentation methods to segment food on plates to perform higher level vision tasks. High level vision tasks include calory assessment [42, 25, 28], cooking recipe retrieval [5] and carbohydrate estimation [29]. Many of these approaches use structure from motion information from several images to develop a 3D food model [29, 7, 18].

Food volume estimation [41] and [4] used template based matching to estimate volumes of food. Especially [27] obtained very good results in regard to volume estimation using feature matching and pose estimation, however in order to obtain an absolute scale, a reference object was needed which had to be placed next to the food item.

Depth prediction Using RGB data as a basis to generate a corresponding depth image has been researched intensely whereas the classical approach in this field is merely using stereo-imagery. Scharstein et al. [33] for example investigated a broad range of existing algorithms based on stereo matching. These algorithms however rely on stereo cameras to work. More closely related to our experiments are methods trying to generate depth information from more loosely aligned images. Sturm et al. [38] presented an effective way to obtain proper scaling for consecutive images in order to calculate 3D structure and motion information - the algorithm thus relies on a sequence of consecutive images. In a more unconstrained setting, Snavely et al. were [37] using many unstructured images from popular sites to generate a 3D view. The underlying system in this case is also based on features and keypoints which are later matched. Machine learning itself has also already been applied to stereo imagery and depth estimation as shown in [17]. Also, in [24] deep neural networks have been trained to be able to predict disparity by learning binocular filters. These systems could then be used as support for stereo setups. Very closely related to depth prediction from single still images are the works from Eigen et al. [11][10], Laina et al. [20] or Liu et al. [22] who use neural networks to infer the depth of still images.

1.3. Contribution

Our contributions in this work are fourfold.

First, we provide and formulate a new computer vision task of inferring bread units (BU) from RGB or RGB-D data by publishing 9k RGB-D image pairs rated by 20 experts.

Second, we present an automatic method for BU regression given RGB-D images using residual neural networks.

Third, we propose a new fully convolutional neural network architecture using skip connections to infer depth maps from RGB images. The architecture at hand shows very good convergence behavior and is especially suited for prediction tasks where small relative errors and local details are especially important.

Finally, we present an automatic method of regressing BUs given only RGB images in two steps by 1) predicting the depth map from the RGB image and 2) predicting the BUs from the RGB image and the predicted depth map.

2. Dataset

2.1. Data Acquisition

Our hardware setup for data-collection consists of a Microsoft Kinect v2 sensor which is connected to a laptop via USB. Since the original Kinect v2 is primarily powered by a 230V power supply, its portability is rather limited. To overcome this issue, we connected the device to a 12V battery-pack making it suitable for mobile use. The device captures depth in a 512x424 pixel frame by default while providing a 1920x1080 pixel RGB output [1]. We collected a total of about 9k RGB-D pairs of 60 different western dishes. The image-streams have been recorded from various angles and distances to capture a wide range of perspectives for each dish. Even though version 2 of the Kinect sensor improved in terms of available ranges, it is still required to maintain a certain minimum distance to the object of interest to receive valid depth values from the device. During recording, we projected the incoming depth stream onto the RGB frame, thus we only provide the projected depth-map in our dataset. Since valid depth is also only provided within certain parts of the RGB frame's spatial dimensions due to the

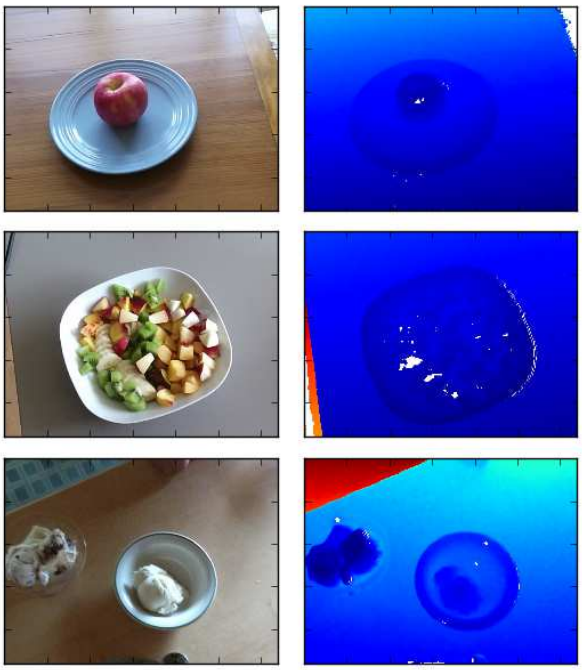


Figure 1. RGB frame (left) and registered depth frame (right) of exemplary classes *Apple*, *Fruit Salad* and *Ice cream*. This figure is best viewed in color.

smaller size of the obtained depth frame, we center crop the RGB-D pair at a size of 640x480 pixel. Since we tried to keep operating the device within a certain maximum distance, the majority of the depth measurements are between 60-80cm. In some scenes, background structures such as floors, chairs or adjacent rooms are visible. Since those pixels exhibit a depth with is mostly larger than 1.2m, they can easily be masked if necessary. Similar to [35] and [26] we experienced similar artifacts degrading the quality of the depths maps such as occlusions from specular or low albedo surfaces, as well as shadowing caused by the physical alignment of infrared emitter and camera. Especially plates, glasses, cutlery or greasy food show a frequent absence of valid depth measurements. Since the algorithm presented in section 3.2.1 has a natural ability to deal with missing depth values by neglecting them during cost computation, we did not see the necessity to fill in missing values during post-processing. From the incoming stream of data, we dumped equally spaced RGB-D pairs at a frequency of about 8-10 frames per second. Due to buffering inconsistencies with the underlying library that we used to interface the Kinect, the capture frequency may differ slightly from recording to recording. The dataset may also contain some slight noise such as blurs from camera movement or partial occlusion through other objects due to the fact that is has been recorded by a handheld device without a tripod. However, we removed unusable frames from the data. Exemplary recordings of the dataset can be seen in figure 1.

2.2. Dataset Specifics

Our dataset compromises 60 western dishes with RGB images and depth-maps (RGB-D) with a total of 8820 images, i.e. 147 images per dish on average. The 60 western dishes were chosen in such a way to cover common meal types. The dataset contains dishes from various categories like "Salads", "Traditional" or "Breakfast". The dishes have been recorded at various locations around TUM university campus, cafeteria or at home. The distribution of these categories in the dataset can be seen in figure 2.

To learn the correspondence from RGB-D to BU, we surveyed 20 long-term diabetic melitus type 1 patients to estimate the bread unit count for our 60 dishes. We showed them a RGB image of a meal and asked them to estimate the bread units. The assessment has been conducted via a proprietary web-application to which images could be uploaded and presented to annotators by sending them a link to the application. Each annotator could then browse through each of the images individually and assign a single BU value per dish. We set the maximum precision per rating to 0.5 BU.

Figure 3 shows the boxplot of the expert BU ratings. The average BU of our data is 3.49 and the averaged BU STD is 0.89 with a minimum STD of 0 BUs (all raters rated the same value) and maximum of STD of 1.99 BU where rater opinions highly disagreed.

3. BU Prediction

Since the assessment of bread units (BU) strongly depends on the volume of the food, we took the depth information of our food dataset into account. We state this problem in the following way:

$$BU(V, \rho_{Food}) = V \cdot \rho_{Food} \tag{1}$$

With the volume V of the meal and ρ_{Food} the bread unit density. The volume V of a meal can be stated as:

$$V \approx \iint_{\mathcal{F}} h d\sigma \tag{2}$$

Where h is the measured depth value of the meal taken from top view normalised such that depth values are zero outside the dish and $\mathcal{F}$ is the projected area of the dish. I.e. we make two assumptions: a) dishes do not overhang b) dishes have a homogeneous bread unit density.

We present two experiments to regress to bread units from our dataset. In all experiments we train on the images of 40 dishes and test on the other 20 dishes of our dataset, i.e. we evaluate our networks capability of predicting BU of categories of food it has never seen before. We use 3-fold cross-validation.

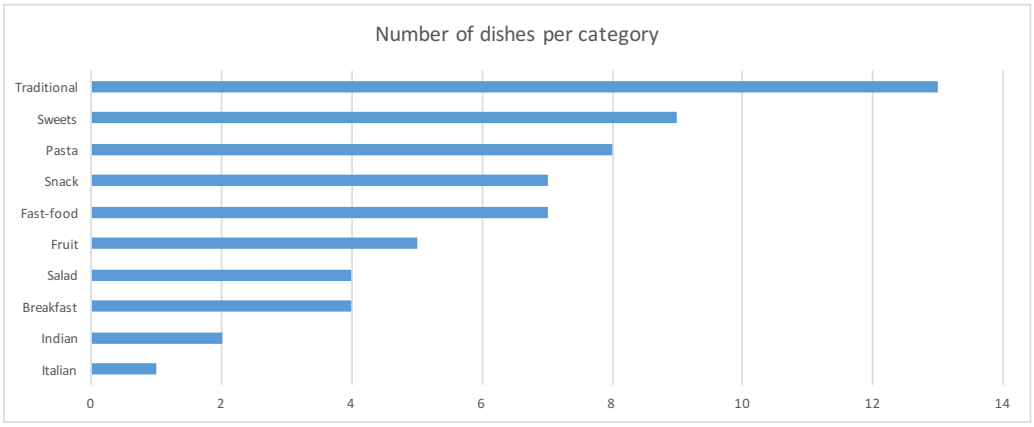


Figure 2. Distribution of different food categories present in the Diabetes60 dataset.

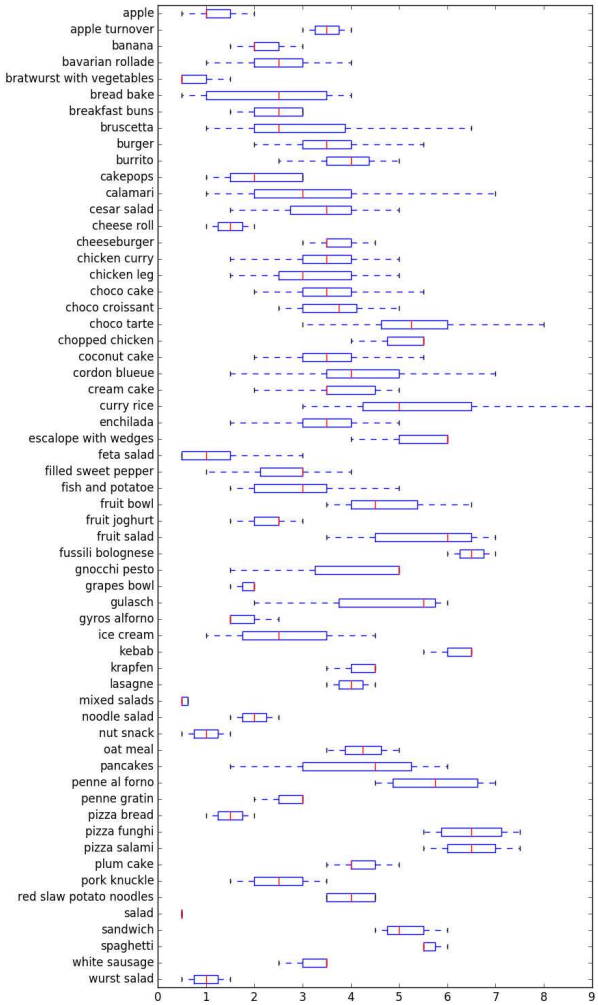


Figure 3. Boxplots of the Bread Unit (BU) estimates from diabetic patients for each class id.

In our first experiment we regress the bread units given the RGB and the corresponding ground-truth depth map obtained from the Kinect using a state-of-the-art Convolutional Neural Network architecture pretrained on the Food 101 dataset. The architecture of choice is Resnet-50 as proposed in [15]. We selected this type of data for pre-training since the task domains are similar. In both cases, images of foods are used for input. To make the network regress values instead of producing a certain class probability, we changed the cross-entropy loss to $\mathcal{L}_2$. In the second experiment, we trained a fully convolutional neural network to predict the depth map of a given RGB image to remove the necessity to have a depth camera. We fine-tuned the depth prediction model on top of the NYU Depth v2 dataset [26]. Afterwards, we trained the Resnet-50 with the predicted depth maps produced by the trained depth predictor. During test time we only provided RGB to regress the bread units. To obtain ground-truth values for the bread units, we averaged the individual ratings per dish. An overview of our conducted experiments is shown in figure 4. All our experiments were conducted on an Ubuntu workstation equipped with a 12GB NVIDIA TITAN X GPU. The neural networks were assembled using the Deep Learning framework Lasagne[8]. In our setup, we downsample all frames by a factor of 2 within the dataset, yielding spatial dimensions of 320x240. For our networks' inputs, we chose images of 304x228 in size, such that there is space for random cropping to further augment the data. In addition to random cropping we use random horizontal flips for augmentation. We also normalize all inputs $x_{i,c}$ with c being the 4 channels via simple precomputed statistics as seen in equation 3

$$x^*_{i,c} = \frac{x_{i,c} - \mu_{i,c}}{\sigma_c} \tag{3}$$

with $\mu_{i,c}$ being the pixel- and channel-wise mean for each pixel i and each channel c. σ_c denotes the standard-

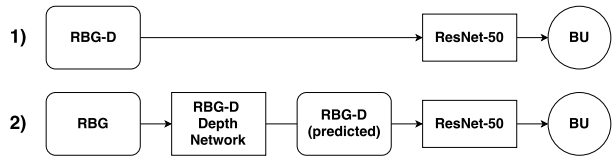


Figure 4. Pipelines of the conducted experiments. 1) BU prediction from RGB-D images, 2) BU prediction from RGB images with intermediate RGB-D prediction using the RGB-D Depth Network.

deviation for each channel c, both computed for the dataset we use for training.

3.1. BU Prediction from RGB and measured depth

We model bread unit estimation by using the depth information as an additional channel to our CNN architecture.

In this experiment, we used a pre-trained the Resnet-50 model on RGB data but tried to preserve the filter learned on the color input channels. We thus initialized that part of the weight tensor corresponding to the RGB input with the weights from the pre-trained network, whereas the part of the filters operating on the depth input channel was initialized with random values following the initialization scheme in [16].

In order to train the residual network for direct bread-unit regression, we replaced the last softmax-layer with a single neuron (ReLU activation) and corresponding L_2 loss. Initial experiments on training the network from scratch led to bad convergence behaviour and overall bad performance. Instead, initializing the weights of neural networks from related tasks often not only promotes convergence but can also lead to higher absolute performance [40]. Therefore, we pre-trained the Resnet-50 model to classify food images first. The Food101 dataset [3] features around 101k images of western foods of various categories. The network achieved a top-1 accuracy of around 70% ([3]: 50.7%). In all BU regression experiments, we used a starting learning rate of 10^{-3} and trained the network for 40 epochs using SGD with momentum and reduced the learning rate once we observed plateaus. Momentum was set to 0.95, whereas we used a weight-decay factor of 10^{-4}.

3.2. BU Prediction from RGB and Inferred Depth

The availability of depth information can provide an additional channel to derive features from, its availability is often lower compared to RGB data. Even though there are handheld devices such as Google's Tango [2] that allow for mobile depth perception, the vast majority of todays' mobile devices are solely equipped with a single RGB camera. This motivates the use of a model to predict the corresponding depth map to a given input image such that only a single RGB image is required to regress the amount of bread units for a given dish.

3.2.1 Inferring Depth from RGB

In a real-life scenario, a diabetes patient is more likely to have access to a camera equipped device such as a smartphone compared to a device equipped with a structured light sensor or a stereo camera setup. We thus want to incorporate a model into our pipeline that estimates the depth of a given scene using only RGB data from a single image. However, mapping from RGB input values to depth is a physically ill-posed problem and with only a single image, this ambiguity cannot be removed. In practice, it is however possible to find a model that can predict depth with reasonable accuracy. The reason is the fact that, apart from unlikely extremes, objects often tend to have similar dimensions in the context of particular scenes, rendering neural networks capable of finding good generalizations to map from an image to its corresponding depth.

Architecture Several works like [10] [11] [32] [22] [20] have already addressed this issue using Deep Neural Networks. In our work, we used an architecture closely related to the one proposed in [20] and performed a set alterations. This architecture has proven to be superior to architectures based on convolutions and fully connected layers such as AlexNet- [19] or VGG-based networks [36] because it is solely composed of convolutional layers while still being able to obtain a receptive field large enough to grasp the whole scene. We made small changes however by incorporating skip connections as proposed in [30]. The purpose of those connections is to provide features of small scales to the later expansive path of the network to preserve local details of the food items. In addition, works like [9] have shown that this approach can ease training and improve overall results. We observed low convergence rates when training an architecture without skip connections completely from scratch as proposed in [20]. With skip connections however, the model converged reliably fast, which allowed end-to-end training in all training cases. To implement that, we also altered the expanding path such that the spatial dimensions of the feature maps match those in the contracting path. This allows for concatenating the activations without cropping. The overall architecture for the depth prediction model is shown in figure 5. Please note that in convolution (symbol: *), no reduction of spatial dimensions takes place. Each orange-colored arrow represents a sequence of residual blocks, the length of the sequence is depicted by the number on the left side of each arrow. The first block in the sequence does always have a shortcut with a projection convolution in place. As in Resnet-50, we use bottleneck blocks. See [15] for details. Our upsampling blocks are conventional residual blocks with projection shortcuts that receive up-scaled versions of the previous feature map with a scaling factor of 2 and use convolutional

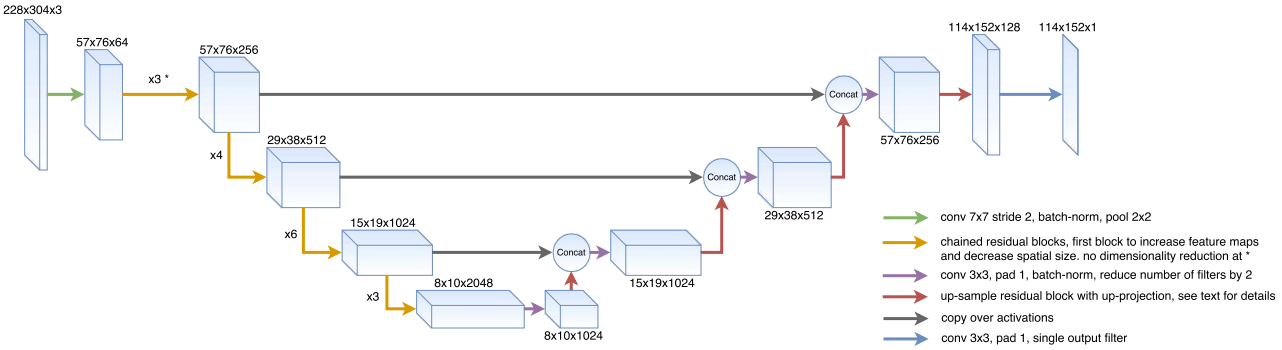


Figure 5. RGB-D Depth Network Architecture with skip connections: The network has an encoding and decoding pathway. Skip connections introduced by [30] allow spatial information exchange and promote convergence.

filters in sizes from 4x4 to 5x5 such that the spatial dimensions of the output feature maps match those obtained in the corresponding contractive counterpart to ease concatenation.

Loss Function We use the reverse-Huber loss function for depth prediction as introduced in [20]. This function is expressed in equation 4 with $d = D - D^*$, where D and D^* denote the prediction and ground-truth depth maps:

$$\mathcal{B}(d) = \begin{cases} |d|, & |d| \leq c \\ \frac{x^2+c^2}{2c}, & |d| > c \end{cases} \tag{4}$$

and c being $\frac{1}{5} max_i(|d_i|)$ for the pixel-wise residuals d_i.

Pre-training We pre-trained the model on the NYU Depth v2 dataset for indoor scene segmentation [35] to start off with a better weight configuration. In contrast to [20], [10] or [11] we did not initialize the contractive part of the network with weights. We leave two-staged pre-training open for future work (training the contractive part on a classification task first, then finetuning on NYU Depth v2, then finetuning on the target dataset).
We extracted equally spaced frames from the raw dataset to obtain a total number of around 26k RGB-D pairs which we globally shuffled afterwards. To actually verify whether the network generalizes, we used the official train/test split of the dataset. To make the data fit the network's inputs, we downsampled the frames by a factor of two using nearest-neighbour interpolation. It is important not to use a higher order interpolation method as they tend to interpolate between valid and invalid pixels. Augmentation was performed on-the-fly during training. The following methods were used to augment the data following values in [10]:

- **Random rotation** Rotating image and ground-truth in-plane for a random angle $\alpha \in [-5, 5]$
- **Zooming** Zooming the image and randomly select a part of it. The zooming factor was drawn per image within a range of [1, 1.5].
- **Random cropping** Similar to [19] we randomly crop images and ground-truth toward the desired network input size
- **Horizontal flips** Images and ground-truth are flipped horizontally with a probability of $p = 0.5$.
- **Random RGB scaling** Input images are randomly scaled with a pixel value $\beta \in [0.9, 1.1]^3$
- **Exposure** We made small changes in exposure to simulate various lighting conditions for the RGB input.

For pre-training we used a starting learning rate of 10^{-2}. In total, we extracted only abound 26k frames from NYU Depth v2 on which we trained the network for 80 epochs using SGD with momentum. We decreased the learning rate following a step-wise policy with a step-width of 20 epochs. The learning rate was decreased by a factor of $\gamma = 0.5$ per step.

Fine-tuning We fine-tuned our network on the data we collected. For training we split the 60 scenes into a training and test set using a split-factor of 0.75 resulting in about 7k frames for training and around 2k frames for test. We made sure that all frames belonging to a certain dish would end up either in the training set or in the test set. To further augment the data, we used the same processing pipeline as for the pre-training step. We trained the network for 40 epochs following a step-wise policy, starting with a learning rate of 10^{-3}, a step-width of 20 epochs while reducing the learning rate by a factor of $\gamma = 0.1$. Additionally, we mask out values larger than 1.2m, since those distances primarily belong to backgrounds in the image. Our experiments revealed also that the inclusion of masks yielded smoother gradients in the estimated depth maps.

	RMSE (lin)	RMSE (log)	rel	$\delta_1 = 1.25$	$\delta_2 = 1.25^2$	$\delta_3 = 1.25^3$
Our network	0.119	0.161	0.129	0.781	0.995	0.999

Table 1. Quantitative results for depth regression on 3D food data. For training and test we masked out all depth values larger than 1.2m in order to ignore the surfaces belonging to background.

4. Results

4.1. Depth Prediction

Our proposed depth architecture achieves state-of-the-art RMSE of 0.651m ([10]: 0.753m and 0.641m, [11]: 0.877m) on NYU Depth v2 when training completely from scratch. Since we were primarily interested in using those weights as a starting point for regressing the depth of food images, we did not extensively fine-tune the hyper parameters for this learning task or pre-trained the contractive stem on a large dataset like Imagenet [31].

Qualitative results of our model trained on the newly recorded food data are shown in figure 6. The results show that the model is able to grasp fine, local details, as seen in the example of the peaches in the bowl of fruits. We hypothesise that the skip-connections not only helped to make the model converge when pre-training, but also support to predict local structures, especially, when looking at the overall range of values. Local food structure is mostly in the range of only 1-2 centimeters whereas the overall depth ranges from about 60cm to partially up to over 1.2m. This is a result of the data being recorded handheld. A similar refinement effect has been reported by Eigen et al. in [11, 10] by using refinement stages in later stages of the network to improve the prediction. In contrast to [25], the depth maps for for food images predicted by our model feature fine details. Even though their model also has refinement stages, the resulting depth maps from our model are with spatial dimensions of 152x114 fairly large.

Furthermore, by masking out invalid depth values during loss-computation, the model also becomes inherently robust to deal with missing/invalid pixel-data from the Kinect. This makes inpainting or other filling techniques unnecessary, even though data recorded with the Kinect v2 is already less prone to contain large amounts of invalid pixels compared to earlier versions of the device. Quantitative metrics for our dataset are shown in table 1. The linear RMSE is 0.119 with a relative error of 0.129. These metrics set the baseline for the depth prediction task of our new dataset.

4.2. BU Prediction

Table 2 shows the results of the BU prediction. The CNN trained on RBG-D data yields an RMSE of 1.46. Trained end-to-end just on RBG with inferred depth achieved a RMSE of 1.53. Those results were obtained using 3-fold cross-validation. Figure 7 shows the box plot of predictions from RBG with inferred depth for all 60 dishes in our dataset. Our methods achieves for many dishes reasonable predictions within the spread of human expert ratings.

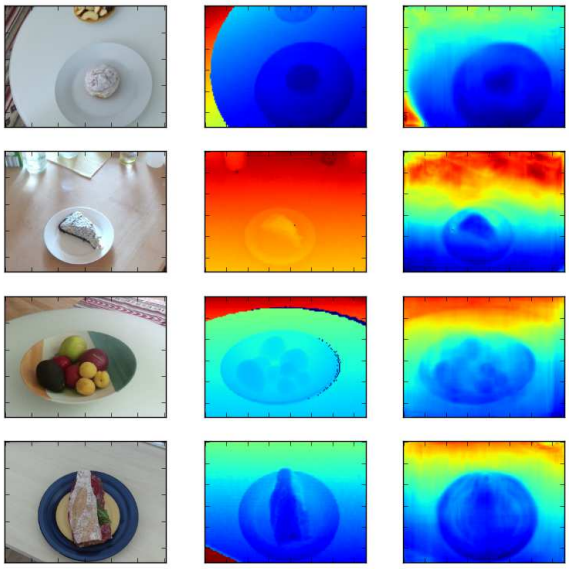


Figure 6. Qualitative results of the model on dishes of the categories *snack*, *sweets* and *fruit*. RGB input (left), ground-truth depth (middle) and predicted depth (right) are shown. The dishes shown above are part of our test-set, images are scaled individually. This figure is best viewed in color.

Approach	Root Mean Square Error (RMSE)
RGB-D Ground truth	**1.46**
RGB-D Predicted Depth	1.53

Table 2. Bread unit inference using Convolutional Neural Networks and Fully Convolutional Neural Networks.

When training the CNN to predict bread-units, the network converged quite quickly as very common in fine-tuning scenarios. This still holds when we use 4 input channels instead of 3 as we preserve the filters operating on color input by transplanting the weights. Providing predicted depth expectedly yields worse results compared to ground-truth depth input even though the margin of error is relatively small. The high accuracy of the predicted depth maps helped to obtain very close results. To further investigate this relation we calculated a Wilcoxon signed-rank test to determine whether RGB-D with ground truth or prediction do lead to the same RMSE. We found that the two approaches do produce the same output distribution with a p-value of 1.52×10^{-12}. We can conclude, that our method with predicted depth does convey the same results.

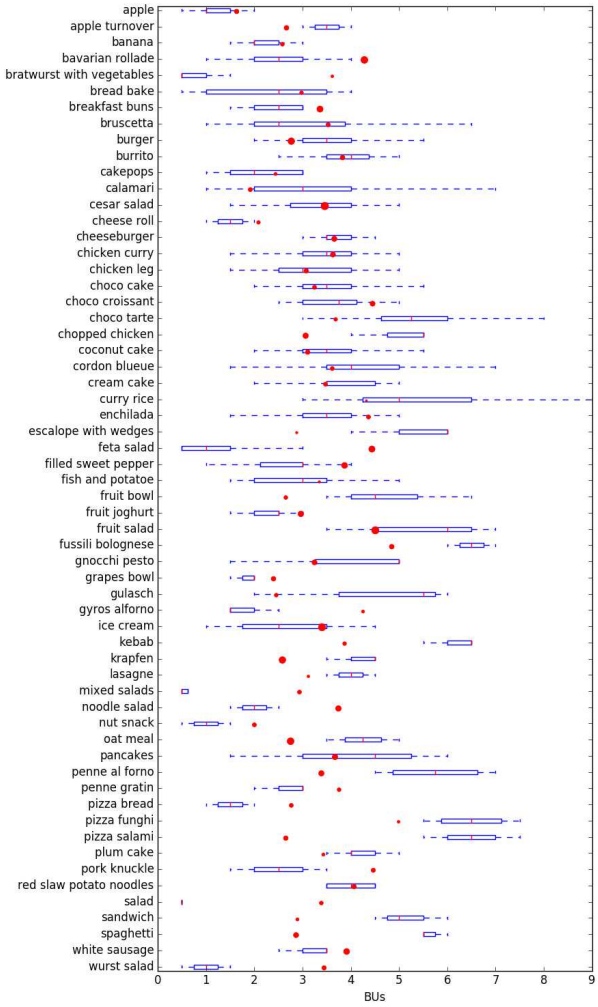


Figure 7. Box plot of ratings and the predictions given by the CNN using RGB and inferred depth. The BU ground truth by the expert annotators is shown in blue boxes and the average predicted BU in shown as a red dot.

5. Discussion and Conclusion

In this work we proposed a new computer vision task of inferring bread units from food image data. We collected RGB-D images of 60 western dishes and surveyed 20 experts to assess the bread unit count of the dishes. We demonstrated two methods of inferring bread units from RGB-D and RGB with a inferred depth map of a fully convolutional depth regression network. The high inter-rater RMSE of 0.89 shows that the task at hand is in fact very hard to solve, even to long-term Diabetes patients. Compared to human raters, our implemented methods perform automatic BU estimation at a RMSE of 1.53 for RGB + inferred depth, which sets a baseline for this task on our proposed dataset. For most dishes, our method yields reasonable BU estimates, i.e. within the standard deviation of human expert raters. However, there are also dishes with faulty BU inference outside this range, in particular pizza salami, spaghetti and salads. This highlights the challenging nature of our proposed learning task for state of the art computer vision methods.

We tried to accomplish a similar goal as [25, 4, 37, 29] in an end-to-end fashion. Calorie and bread unit assessment are closely related tasks and both rely on depth or volume of a meal, besides contextual and semantic information. We addressed the contextual and semantic information using state of the art residual neural network architectures as proposed by [15]. In our approach, to incorporate the depth and volumetric information, we neither relied on structure from motion information such as [29, 7, 18] nor on a reference object [27]. State of the art depth network architectures as proposed by [11, 20] did not converge on our dataset without pre-training. The relative error of our proposed depth network architecture on Diabetes60 is 0.129. Unfortunately the food depth data of [25] was not published. They reported a relative error of 0.18 on their food data [25]. Comparing their qualitative depth predictions (see figure 6c in [25]), our RBG depth network could reconstruct finer details of the food as shown in figure 6. Our proposed depth prediction architecture may also be useful to other high-dimensional regression tasks where pretrained weights are not available or there is a strong focus on local details. We hope that by publishing our dataset along with baseline methods and results, we provide a starting point for researchers to tackle the same or comparable problems, either in a similar end-to-end fashion or by splitting the task into several sub-tasks and solving them independently. The depth values at hand may also be useful for people working on 3D reconstruction and modeling of food items which may or may not be part of a pipeline achieving a different end goal. Public RBG-D datasets are rare and we hope to foster computer vision research in this field with our dataset contribution. Advancements in the fields of automated assessment of food intake could become highly valuable for diabetes patients or generally everyone keen on keeping track of her or his nutrition. Right now, all our models require fairly recent desktop GPUs to operate. Once deep learning becomes more adopted by smartphones or other portable devices, those models could operate on device and thus provide faster feedback. In addition, location services could be integrated tightly into the estimation process to leverage local information obtained from restaurants or food courts.

6. Acknowledgements

This work was supported by the Technical University of Munich - Institute for Advanced Study (funded by the German Excellence Initiative and the European Union Seventh

Framework Program under grant agreement n 291763), the Marie Curie COFUND program of the the European Union (Rudolf Mossbauer Tenure - Track Professorship to BHM) and the BMBF Softwarecampus project by Patrick Christ. We thank NVIDIA and Amazon AWS for granting GPU and computation support.

References

[1] Kinect V2. https://developer.microsoft.com/en-us/windows/kinect/hardware. Accessed: 2016-05-22. 2

[2] Tango, Google Developers. Accessed: 2016-05-22. 5

[3] L. Bossard, M. Guillaumin, and L. Van Gool. Food-101–mining discriminative components with random forests. In *ECCV*, pages 446–461. Springer, 2014. 2, 5

[4] J. Chae, I. Woo, S. Kim, R. Maciejewski, F. Zhu, E. J. Delp, C. J. Boushey, and D. S. Ebert. Volume estimation using food specific shape templates in mobile image-based dietary assessment. In *IS&T/SPIE Electronic Imaging*, pages 78730K–78730K. International Society for Optics and Photonics, 2011. 2, 8

[5] J. Chen and C.-W. Ngo. Deep-based ingredient recognition for cooking recipe retrieval. In *Proceedings of the 2016 ACM on Multimedia Conference*, pages 32–41. ACM, 2016. 2

[6] M. Chen, K. Dhingra, W. Wu, L. Yang, R. Sukthankar, and J. Yang. Pfid: Pittsburgh fast-food image dataset. In *ICIP*, pages 289–292. IEEE, 2009. 2

[7] J. Dehais, S. Shevchik, P. Diem, and S. G. Mougiakakou. Food volume computation for self dietary assessment applications. In *Bioinformatics and Bioengineering (BIBE), 2013 IEEE 13th International Conference on*, pages 1–4. IEEE, 2013. 2, 8

[8] S. Dieleman, J. Schlter, C. Raffel, E. Olson, S. K. Snderby, D. Nouri, D. Maturana, M. Thoma, E. Battenberg, J. Kelly, J. D. Fauw, M. Heilman, D. M. de Almeida, B. McFee, H. Weideman, G. Takcs, P. de Rivaz, J. Crall, G. Sanders, K. Rasul, C. Liu, G. French, and J. Degrave. Lasagne: First release., Aug. 2015. 4

[9] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal. The importance of skip connections in biomedical image segmentation. *CoRR*, abs/1608.04117, 2016. 5

[10] D. Eigen and R. Fergus. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In *CVPR*, pages 2650–2658, 2015. 2, 5, 6, 7

[11] D. Eigen, C. Puhrsch, and R. Fergus. Depth map prediction from a single image using a multi-scale deep network. In *NIPS*, pages 2366–2374, 2014. 2, 5, 6, 7, 8

[12] G. M. Farinella, D. Allegra, and F. Stanco. A benchmark dataset to study the representation of food images. In *ECCV*, pages 584–599. Springer, 2014. 2

[13] A. Ferrara. Increasing prevalence of gestational diabetes mellitus a public health perspective. *Diabetes care*, 30(Supplement 2):S141–S146, 2007. 1

[14] H. He, F. Kong, and J. Tan. Dietcam: Multiview food recognition using a multikernel svm. *IEEE Journal of biomedical and health informatics*, 20(3):848–855, 2016. 2

[15] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015. 4, 5, 8

[16] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *ICCV*, pages 1026–1034, 2015. 5

[17] K. Konda and R. Memisevic. Unsupervised learning of depth and motion. *arXiv preprint arXiv:1312.3429*, 2013. 2

[18] F. Kong and J. Tan. Dietcam: Automatic dietary assessment with mobile camera phones. *Pervasive and Mobile Computing*, 8(1):147–163, 2012. 2, 8

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, pages 1097–1105, 2012. 5, 6

[20] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, and N. Navab. Deeper depth prediction with fully convolutional residual networks. *arXiv preprint arXiv:1606.00373*, abs/1606.00373, 2016. 2, 5, 6, 8

[21] C. Liu, Y. Cao, Y. Luo, G. Chen, V. Vokkarane, and Y. Ma. Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment. In *International Conference on Smart Homes and Health Telematics*, pages 37–48. Springer, 2016. 2

[22] F. Liu, C. Shen, and G. Lin. Deep convolutional neural fields for depth estimation from a single image. In *CVPR*, pages 5162–5170, 2015. 2, 5

[23] Y. Matsuda, H. Hoashi, and K. Yanai. Recognition of multiple-food images by detecting candidate regions. In *International Conference on Multimedia and Expo*, pages 25–30. IEEE, 2012. 2

[24] R. Memisevic and C. Conrad. Stereopsis via deep learning. In *NIPS Workshop on Deep Learning*, volume 1, 2011. 2

[25] A. Meyers, N. Johnston, V. Rathod, A. Korattikara, A. Gorban, N. Silberman, S. Guadarrama, G. Papandreou, J. Huang, and K. P. Murphy. Im2calories: towards an automated mobile vision food diary. In *ICCV*, pages 1233–1241, 2015. 2, 7, 8

[26] P. K. Nathan Silberman, Derek Hoiem and R. Fergus. Indoor segmentation and support inference from rgbd images. In *ECCV*, 2012. 3, 4

[27] M. Puri, Z. Zhu, Q. Yu, A. Divakaran, and H. Sawhney. Recognition and volume estimation of food intake using a mobile device. In *WACV*, pages 1–8. IEEE, 2009. 2, 8

[28] D. Rav, B. Lo, and G. Z. Yang. Real-time food intake classification and energy expenditure estimation on a mobile device. In *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pages 1–6, June 2015. 2

[29] D. Rhyner, H. Loher, J. Dehais, M. Anthimopoulos, S. Shevchik, R. H. Botwey, D. Duke, C. Stettler, P. Diem, and S. Mougiakakou. Carbohydrate estimation by a mobile phone-based system versus self-estimations of individuals with type 1 diabetes mellitus: A comparative study. *Journal of medical Internet research*, 18(5), 2016. 2, 8

[30] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241. Springer, 2015. 5, 6

[31] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. 7

[32] A. Saxena, M. Sun, and A. Y. Ng. Make3d: Learning 3d scene structure from a single still image. *PAMI*, 31(5):824–840, 2009. 5

[33] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3):7–42, 2002. 2

[34] J. E. Shaw, R. A. Sicree, and P. Z. Zimmet. Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes research and clinical practice*, 87(1):4–14, 2010. 1

[35] N. Silberman and R. Fergus. Indoor scene segmentation using a structured light sensor. In *Proceedings of the International Conference on Computer Vision - Workshop on 3D Representation and Recognition*, 2011. 3, 6

[36] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 5

[37] N. Snavely, S. M. Seitz, and R. Szeliski. Photo tourism: exploring photo collections in 3d. In *ACM transactions on graphics (TOG)*, volume 25, pages 835–846. ACM, 2006. 2, 8

[38] P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In *ECCV*, pages 709–720. Springer, 1996. 2

[39] H. Warshaw and K. Kulkarni. *Complete Guide to Carb Counting: How to Take the Mystery Out of Carb Counting and Improve Your Blood Glucose Control*. American Diabetes Association, 2011. 1

[40] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. How transferable are features in deep neural networks? In *NIPS*, pages 3320–3328, 2014. 5

[41] Y. Yue, W. Jia, J. D. Fernstrom, R. J. Sclabassi, M. H. Fernstrom, N. Yao, and M. Sun. Food volume estimation using a circular reference in image-based dietary studies. In *Proceedings of the 2010 IEEE 36th Annual Northeast Bioengineering Conference (NEBEC)*, pages 1–2. IEEE, 2010. 2

[42] W. Zhang, Q. Yu, B. Siddiquie, A. Divakaran, and H. Sawhney. snap-n-eat food recognition and nutrition estimation on a smartphone. *Journal of diabetes science and technology*, 9(3):525–533, 2015. 2

Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks

Authoren: Patrick Ferdinand Christ, Florian Ettlinger, Felix Grün, Mohamed Ezzeldin A Elshaera, Jana Lipkova, Sebastian Schlecht, Freba Ahmaddy, Sunil Tatavarty, Marc Bickel, Patrick Bilic, Markus Rempfler, Felix Hofmann, Melvin D Anastasi, Seyed-Ahmad Ahmadi, Georgios Kaissis, Julian Holch, Wieland Sommer, Rickmer Braren, Volker Heinemann, Bjoern Menze

Abstract: Automatic segmentation of the liver and hepatic lesions is an important step towards deriving quantitative biomarkers for accurate clinical diagnosis and computer-aided decision support systems. This paper presents a method to automatically segment liver and lesions in CT and MRI abdomen images using cascaded fully convolutional neural networks (CFCNs) enabling the segmentation of large-scale medical trials and quantitative image analyses. We train and cascade two FCNs for the combined segmentation of the liver and its lesions. As a first step, we train an FCN to segment the liver as ROI input for a second FCN. The second FCN solely segments lesions within the predicted liver ROIs of step 1. CFCN models were trained on an abdominal CT dataset comprising 100 hepatic tumor volumes. Validation results on further datasets show that CFCN-based semantic liver and lesion segmentation achieves Dice scores over 94% for the liver with computation times below 100s per volume. We further experimentally demonstrate the robustness of the proposed method on 38 MRI liver tumor volumes and the public 3DIRCAD dataset.

Individuelle Leistungsbeiträge: Projektkoordination, Datenakquise und Datenaufbereitung, Konzeption und Durchführung von Experimenten, Federführende Anfertigung des Manuskripts

Unveröffentlichtes Manuskript

Automatic Liver and Tumor Segmentation of CT and MRI Volumes Using Cascaded Fully Convolutional Neural Networks

Patrick Ferdinand Christ[a,1], Florian Ettlinger[a,1], Felix Grün[a], Mohamed Ezzeldin A. Elshaer[a], Jana Lipková[a], Sebastian Schlecht[a], Freba Ahmaddy[a], Sunil Tatavarty[a], Marc Bickel[a], Patrick Bilic[a], Markus Rempfler[a], Felix Hofmann[b], Melvin D'Anastasi[b], Seyed-Ahmad Ahmadi[b], Georgios Kaissis[a], Julian Holch[b], Wieland Sommer[b], Rickmer Braren[a], Volker Heinemann[b], Bjoern Menze[a]

[a] *Technical University of Munich, Arccistrasse 21, 80333 Munich*
[b] *LMU Hospital Grosshadern, Marchioninistrasse 15, 81377 Munich, Germany*

Abstract

Automatic segmentation of the liver and hepatic lesions is an important step towards deriving quantitative biomarkers for accurate clinical diagnosis and computer-aided decision support systems. This paper presents a method to automatically segment liver and lesions in CT and MRI abdomen images using cascaded fully convolutional neural networks (CFCNs) enabling the segmentation of large-scale medical trials and quantitative image analyses. We train and cascade two FCNs for the combined segmentation of the liver and its lesions. As a first step, we train an FCN to segment the liver as ROI input for a second FCN. The second FCN solely segments lesions within the predicted liver ROIs of step 1. CFCN models were trained on an abdominal CT dataset comprising 100 hepatic tumor volumes. Validation results on further datasets show that CFCN-based semantic liver and lesion segmentation achieves Dice scores over 94% for the liver with computation times below 100s per volume. We further experimentally demonstrate the robustness of the proposed method on 38 MRI liver tumor volumes and the public 3DIRCAD dataset.

Keywords: Liver, Lesion, Segmentation, FCN, CRF, Deep Learning

1. Introduction

1.1. Motivation

Anomalies in the shape and texture of the liver and visible lesions in computed tomography (CT) and magnetic resonance images (MRI) images are important biomarkers for initial disease diagnosis and progression in both primary and secondary hepatic tumor disease [1].

Primary tumors such as breast, colon and pancreas cancer often spread metastases to the liver during the course of disease. Therefore, the liver and its lesions are routinely

[1] Authors contributed equally

analyzed in primary tumor staging. In addition, the liver is also a site of primary tumor disease such as Hepatocellular carcinoma (HCC). Hepatocellular carcinoma (HCC) presents the sixth-most common cancer and the third-most common cause of cancer-related deaths worldwide [2]. HCC comprises a genetically and molecularly highly heterogeneous group of cancers that commonly arise in a chronically damaged liver. Importantly, HCC subtypes differ significantly in clinical outcome. The stepwise transformation to HCC is accompanied by major changes in tissue architecture including an increase in cellularity and a switch in vascular supply (i.e. arterialization). These quantifiable changes in tissue architecture provide the basis for the non-invasive detection of HCC in imaging [3], but also lead to highly variable structures and shapes.

In clinical routine, manual or semi-manual segmentation techniques are applied to interprete CT and MRI images that have been acquired in the diagnosis of the liver. These techniques, however, are subjective, operator-dependent and very time-consuming. In order to improve the productivity of radiologists, computer-aided methods have been developed in the past. However, an automated robust segmentation of combined liver and lesion remains still an open problem because of challenges as a low-contrast between liver and lesion, different types of contrast levels (hyper-/hypo-intense tumors), abnormalities in tissues (such as after surgical resection of metastasis), size and varying number of lesions. As shown in figure 1 the heterogeneity in liver and lesion contrast is very large among subjects. Different acquisition protocols, differing contrast-agents, varying levels of contrast enhancements and dissimilar scanner resolutions lead to unpredictable intensity differences between liver and lesion tissue. This complexity of contrast differences make it difficult for intensity-based methods to generalize to unseen test cases from different clinical sites. In addition, the varying shape of lesions due to irregular tumor growth and response to treatment (i.e surgical resection) reduce efficiency of computational methods that make use of prior knowledge on lesion shape.

1.2. Related Works

Nevertheless, several interactive and automatic methods have been developed to segment the liver and liver lesions in CT volumes. In 2007 and 2008, two Grand Challenges benchmarks on liver and liver lesion segmentation have been conducted in conjunction with MICCAI conference [1, 4]. Methods presented at the challenges were mostly based on statistical shape models. Furthermore, grey level and texture based methods have been developed [1]. Recent work on liver and lesion segmentation employs graph cut and level set techniques [5, 6, 7], sigmoid edge modeling [8] or manifold and machine learning [9, 10, 11, 12]. However, these methods are not widely applied in clinics, due to their speed and robustness on heterogeneous, low-contrast real-life CT data. To overcome these weaknesses, interactive methods were still developed [13] to overcome these weaknesses.

Deep Convolutional Neural Networks (CNN) have gained significant attention in the scientific community for solving computer vision tasks such as object recognition, classification and segmentation [14, 15], often out-competing state-of-the art methods. Most importantly, CNN methods have proven to be highly robust to varying image appearance, which motivates us to apply them to fully automatic liver and lesions segmentation in CT volumes.

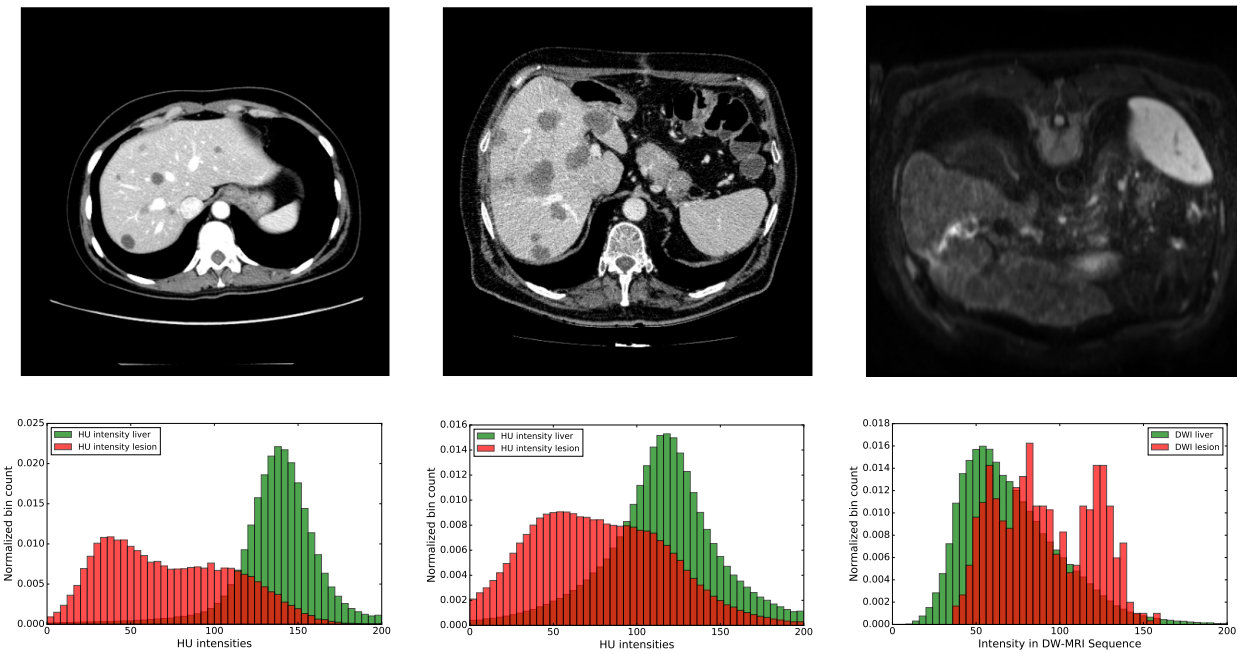


Figure 1: Liver and liver lesions slices in CT and diffusion weighted DW-MRI as well as the corresponding histogram for liver and lesions pixels in the respective modality. The shape, size and level of contrast vary for different lesions. As the histograms indicate, there is a significant overlap between liver and lesion intensities, leading to a low overall contrast.

Semantic image segmentation methods based on fully convolutional neural networks FCN were developed in [15], with impressive results in natural image segmentation competitions [16, 17]. Likewise, new segmentation methods based on CNN and FCNs were developed for medical image analysis, with highly competitive results compared to state-of-the-art. [18, 19, 20, 21, 22, 23, 24, 25].

1.3. Contribution

In this work, we demonstrate the combined automatic segmentation of the liver and its lesions in low-contrast heterogeneous medical volumes. Our contributions are threefold. First, we train and apply fully convolutional CNN on CT volumes of the liver for the first time, demonstrating the adaptability to challenging segmentation of hepatic liver lesions. Second, we propose to use a cascaded fully convolutional neural network (CFCN) on CT slices, which segments liver and lesions sequentially, leading to significantly higher segmentation quality, as demonstrated on a public challenge dataset. Third, we experimentally demonstrate the generalization and scalability of our methods to different modalities and diverse real-life datasets, including a novel diffusion weighted MRI dataset and a large multri-centric CT dataset.
A preliminary version of this work was presented in MICCAI 2016 [26] and will be presented at ISBI 2017 [27]. In this paper, we have substantially revised and extended these previous publications. The main modifications include an elaborated description of the proposed

methods, an analysis of underlying design principles and architectures as well as the application to new datasets and modalities.

In the following sections, we will describe our proposed pipeline (2.1) including CFCN (2.3) and 3D CRF (2.4). The experiments are illustrated in section (3).

2. Methods

2.1. Overview of our Proposed Segmentation Workflow

Our proposed segmentation workflow is depicted in figure 2. The workflow consists of three major steps. The first step (e.g. section 2.2) deals with data preprocessing and preparation for the neural network segmentation. In a second step (e.g. section 2.3) two cascaded fully convolutional neural networks first segment the liver and then lesions within the liver region-of-interest (ROI). In the final third step, the calculated probabilities of CFCN will be refined using a dense 3D conditional random field to produce the final segmentation result.

2.2. Data Preparation

The following section deals with data preprocessing and augmentation for CT data. Preprocessing was carried out in a slice-wise fashion. First, the Hounsfield unit values were windowed in the range $[-100, 400]$ to exclude irrelevant organs and objects. Figure 3 shows the effect of our applied preprocessing to a raw medical slice. We increased contrast through histogram equalization. Figure 3 shows also the final slice after HU-windowing and contrast-enhancement. The contrast within the liver has been enhanced to allow better differentiation of abnormal liver tissue. For DW-MRI the data preparation scheme is similar and differs in the data normalization, which additionally performs N4bias correction [28].

As in [18, 22], to teach the network the desired invariance properties, several data augmentations steps, such as elastic deformation, translation, rotation and addition of Gaussian noise with standard deviation of the current slice, have been employed to increase the training data for the CFCN. Details on the data augmentation schemes is made available in our sourcecode[2].

2.3. Cascaded Fully Convolutional Neural Networks

In the following section, we describe different state-of-the the art deep learning architecture and design choices that we evaluated for a use in our segmentation tasks. We denote the 3D image volume as I, the total number of voxels as N and the set of possible labels as $\mathcal{L} = \{0, 1, \ldots, l\}$. For each voxel i, we define a variable $x_i \in \mathcal{L}$ that denotes the assigned label. The probability of a voxel i belonging to label k given the image I is described by $P(x_i = k|I)$ and will be modelled by the FCN. In our particular study, we use $\mathcal{L} = \{0, 1, 2\}$ for background, liver and lesion, respectively.

[2]Sourcecode and models are available at https://github.com/IBBM/Cascaded-FCN

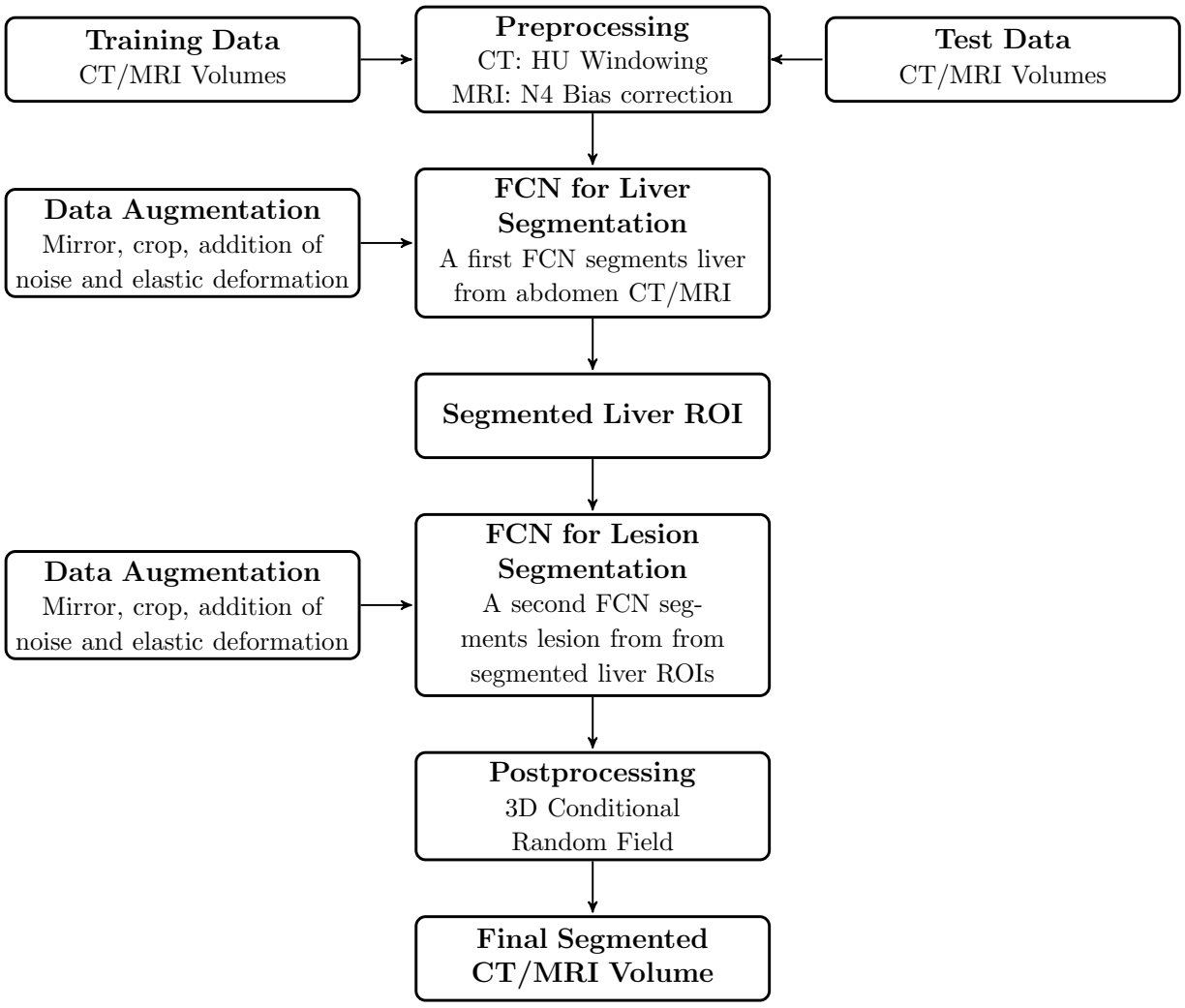


Figure 2: Overview of the proposed image segmentation workflow for training and testing. As the first step the CT/MRI volumes are preprocessed with either HU-windowing or N4 bias correction. During the training phase the training data is augmented to foster the learning of invariance against noise and deformations in medical data. The CT/MRI volumes are trained after pre-processing and data augmentation in a cascaded fully convolutional neural network (CFCN). A first FCN segments the liver from abdomen CT/MRI scans. This segmented liver region of interest ROI is the input for a second FCN, that segments lesions from the given segmented liver ROI. To gain the final segmented volume is refined afterwards using a 3D conditional random field 3D CRF.

Figure 3: Overview of the applied preprocessing steps. The raw CT slices (left) are windowed to a Hounsfield Unit range of -100 to 400 HU to neglect organs and tissues that are not of interest. The HU-windowed slice (middle) is further processed using a histogram equalization to allow further contrast enhancement of abnormal tissue (right).

2.3.1. From AlexNet to U-Net

Long et al. (2015) presented the first fully convolutional network architecture for semantic segmentation [15]. The main idea in their work is to replace the last fully connected layers of a classification network such as the AlexNet [14] with fully convolutional layers to allow dense pixel-wise predictions. The last fully convolutional layers have to be upscaled to match the input dimensions. In comparison to prior work, the AlexFCN allows pixel-wise prediction from full-sized medical slices, instead of patch-wise classification. Figures 4a and 4b show the training curves for training the AlexFCN (without class balancing) on 3DIRCAD dataset. Both training curves converged fast to a steady state in training and test Dice overlap. Both training curves show a large overfitting of the AlexFCN without class balancing, with Dice overlaps of 71%/90% in test/training data for liver, and 24%/60% for lesions. In general the lesion Dice of 24% at test time is comparable low. Long et al. (2015) explicitly stated that they did not need to apply class balancing to their natural image segmentation problem. A reason for this is that they used pretrained AlexNet weights trained on natural images, i.e. ImageNet data. However, for many medical applications it is mandatory to apply class balancing since pre-trained networks from natural images cannot be used properly and the class of interest occurs more seldomly in the dataset. Figures 4c and 4d show the importance of class balancing in medical image segmentation. The training and test Dice for both liver and lesions increases noticeably to 78% for liver and 38% for lesions. A further large improvement can be obtained by applying the U-Net Architecture proposed by Ronneberger et al. (2015) [18]. Besides the increased depth of 19 layers and learnable upscaling (up-convolution), the U-Net provides a superior design pattern of skip connections between different stages of the neural network.

In early stages of the neural network, spatial information is present in the activations of the current stage. In later stages of the neural network, spatial information gets transferred to semantic information at the cost of specific knowledge on the localization of these structures. Here, for example, the original U-Net architecture reduces an input image of

size 388x388 to a size of 28x28 in the U-Net bottleneck. Ronneberger et al. introduced skip-connections to allow utilization of spatial and semantic information at later stage, since the spatial information from earlier stage can be fused in the neural network at later stages. Thus the neural network at later stages can utilize semantic and spatial information to infer information.

2.3.2. From FCN to CFCN

We used the U-Net architecture [18] to compute the soft label probability maps $P(x_i|I)$. The U-Net architecture enables accurate pixel-wise prediction by combining spatial and contextual information in a network architecture comprising 19 convolutional layers. Figures 4e and 4f show the training curves for the U-Net on 3DIRCAD data set. The overall performance of the lesion segmentation is further increased to 53% test Dice. The U-Net learned features to discriminate liver and lesion at the same time. As one of our main contributions, we propose a cascaded training of FCNs to learn specific features for solving a segmentation task once per training, which leads to higher segmentation performance.

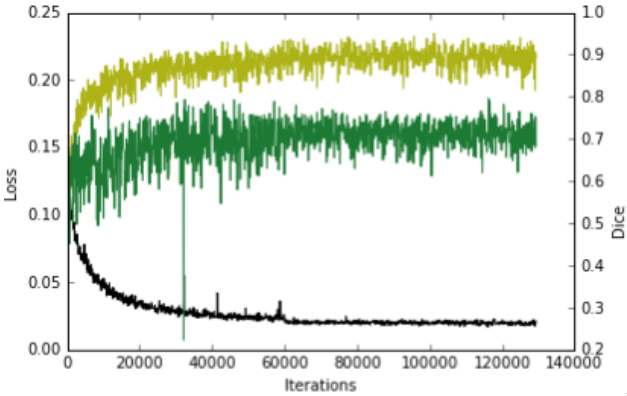
The motivation behind the cascade approach is that it has been shown that U-Nets and other forms of CNNs learn a hierarchical representation of the provided data. The stacked layers of convolutional filters are tailored towards the desired classification in a data-driven manner, as opposed to designing hand-crafted features for separation of different tissue types. By cascading two U-Nets, we ensure that the U-Net in step 1 learns filters that are specific for the detection and segmentation of the liver from an overall abdominal CT scan, while the U-Net in step 2 arranges a set of filters for separation of lesions from the liver tissue. Furthermore, the liver ROI helps in reducing false positives for lesions. Figures 5 and 6 illustrate our proposed method. We train one network to segment the liver in abdomen slices (step 1). This network can solely concentrate on learning discriminative features for liver vs. background segmentation, e.g. figure 5. After that we train another network to segment the lesions, given an image of the liver (step 2). The segmented liver from step 1 is cropped and re-sampled to the required input size for the cascaded U-Net in step 2. All non-liver regions are masked out and the second U-Net can concentrate on learning discriminative features for lesion vs. liver background segmentation.

2.3.3. Effect of Class Balancing

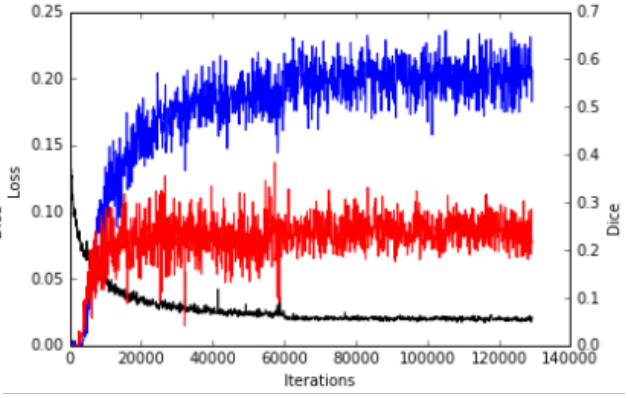
A crucial step in training FCNs is appropriate class balancing according to the pixel-wise frequency of each class in the data. In contrast to [15], we observed that training the network to segment small structures such as lesions is not possible without class balancing, due to the high class imbalance that is typically in the range of ¡1% for lesion pixels. Therefore we introduced an additional weighting factor ω^{class} in the cross entropy loss function L of the FCN:

$$L = -\frac{1}{n}\sum_{i=1}^{N} \omega_i^{class} \left[\hat{P}_i \log P_i + (1-\hat{P}_i)\log(1-P_i)\right] \tag{1}$$

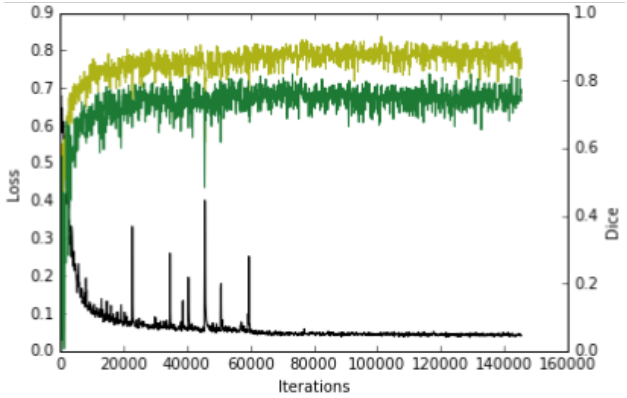
P_i denotes the probability of voxel i belonging to the foreground, $\hat{P}_i$ represents the ground truth. We chose ω_i^{class} to be $\frac{\sum_i 1-\hat{P}_i}{\sum_i \hat{P}_i}$ if $\hat{P}_i = 1$ and 1 otherwise.



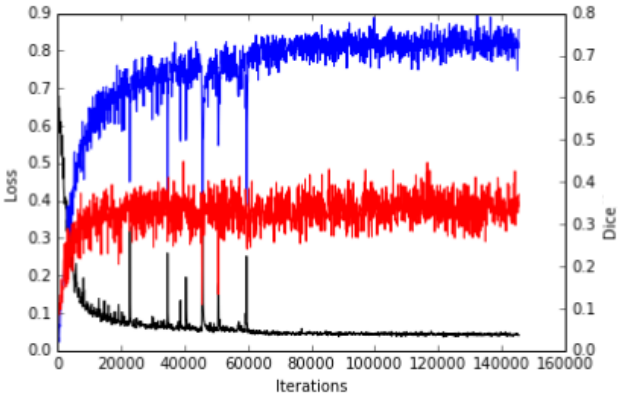
(a) AlexFCN architecture without class balancing: Loss (black), Training Dice (light green), Test Dice (dark green) of Liver



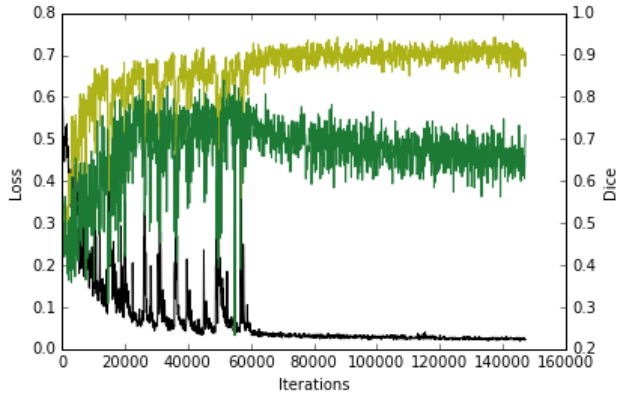
(b) AlexFCN architecture without class balancing: Loss (black), Training Dice (blue), Test Dice (red) of Lesion



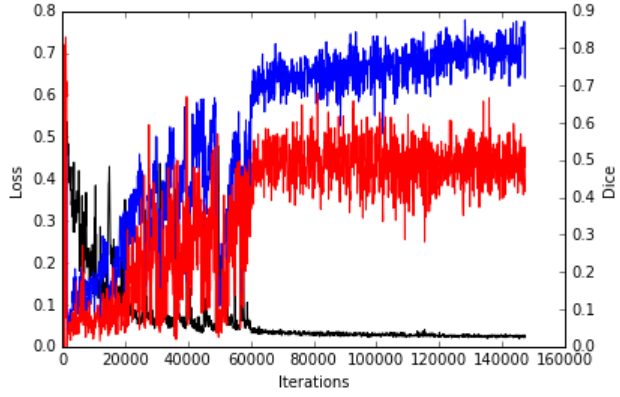
(c) AlexFCN architecture with class balancing: Loss (black), Training Dice (light green), Test Dice (dark green) of Liver



(d) AlexFCN architecture with class balancing: Loss (black), Training Dice (blue), Test Dice (red) of Lesion



(e) U-Net architecture with class balancing: Loss (black), Training Dice (light green), Test Dice (dark green) of Liver



(f) U-Net architecture with class balancing: Loss (black), Training Dice (blue), Test Dice (red) of Lesion

Figure 4: Training curves of different network architectures and training procedures of liver and lesion on 3DIRCAD dataset.

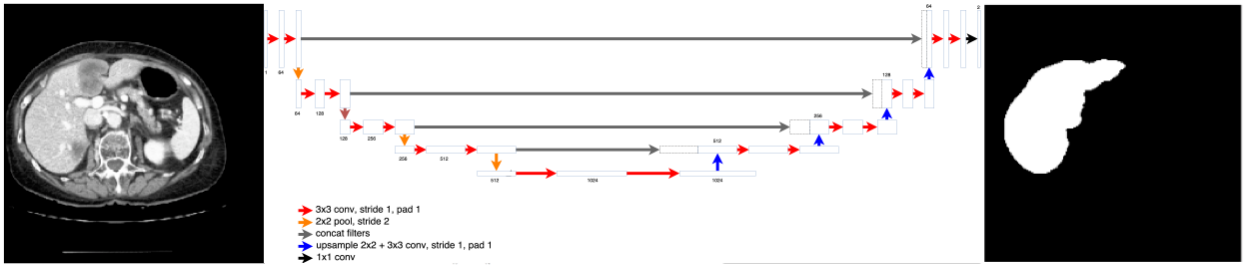


Figure 5: Step 1 of Cascaded FCN: The first U-Net learns to segment livers from a CT slice.

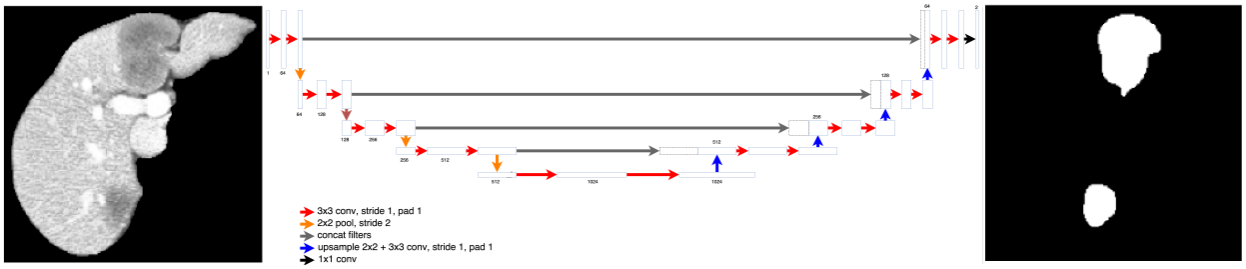


Figure 6: Step 2 of Cascaded FCN: The second U-Net learns to segment lesions from a liver segmentation mask segmented in step 1 of the cascade

2.3.4. Transfer Learning and Pretraining

A common concept in deep learning is transfer learning using pretrained neural network models. Neural networks pretrained on a other task, e.g. a natural image classification data set, can be used as initialization of the network weights when training on a new task e.g. image segmentation of medical volumes. The intuition behind this idea is, that also for other tasks or dataset the first layers of neural networks learn similar concepts to recognize basic structures such as blobs and edges. This concepts do not have be trained again from scratch when using pretrained models. For our experiments we used pretrained U-Net models provided by Ronneberger et al. (2015), which were trained on cell image segmentation data [18]. We have released our trained models on liver and lesion segmentation to allow other researcher to start their training with learned liver and lesion concepts[3].

2.4. 3D Conditional Random Field

Volumetric FCN implementation with 3D convolutions was strongly limited by GPU hardware and available VRAM [21]. Recent work such as V-Net and 3D U-Net, allow nowadays 3D FCNs at decreased resolution [29, 30]. In addition, the anisotropic resolution of medical volumes (e.g. 0.57-0.8mm in axial and 1.25-4mm in sagital/coronal voxel dimension in 3DIRCADb) complicates the training of discriminative 3D filters. Instead, to capitalise on the locality information across slices within the dataset, we utilize 3D dense conditional random fields (CRFs) as proposed by [31]. To account for 3D information, we consider all slice-wise predictions of the FCN together in the CRF applied to the entire volume at once.

[3]Sourcecode and models are available at https://github.com/IBBM/Cascaded-FCN

We formulate the final label assignment given the soft predictions (probability maps) from the FCN as *maximum a posteriori* (MAP) inference in a dense CRF, allowing us to consider both spatial coherence and appearance.

We specify the dense CRF following [31] on the complete graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with vertices $i \in \mathcal{V}$ for each voxel in the image and edges $e_{ij} \in \mathcal{E} = \{(i,j)\ \forall i,j \in \mathcal{V} \text{ s.t. } i < j\}$ between *all* vertices. The variable vector $\mathbf{x} \in \mathcal{L}^N$ describes the label of each vertex $i \in \mathcal{V}$. The energy function that induces the according Gibbs distribution is then given as:

$$E(\mathbf{x}) = \sum_{i \in \mathcal{V}} \phi_i(x_i) + \sum_{(i,j) \in \mathcal{E}} \phi_{ij}(x_i, x_j) \tag{2}$$

where $\phi_i(x_i) = -\log P(x_i|I)$ are the unary potentials that are derived from the FCNs probabilistic output, $P(x_i|I)$. $\phi_{ij}(x_i, x_j)$ are the pairwise potentials, which we set to:

$$\begin{aligned} \phi_{ij}(x_i, x_j) = \ & \mu(x_i, x_j) \Bigg(w_{\text{pos}} \exp\left(-\tfrac{|p_i - p_j|^2}{2\sigma_{\text{pos}}^2} \right) \\ & + w_{\text{bil}} \exp\left(-\tfrac{|p_i - p_j|^2}{2\sigma_{\text{bil}}^2} - \tfrac{|I_i - I_j|^2}{2\sigma_{\text{int}}^2} \right) \Bigg) \end{aligned} \tag{3}$$

where $\mu(x_i, x_j) = \mathbf{1}(x_i \neq x_j)$ is the Potts function, $|p_i - p_j|$ is the spatial distance between voxels i and j and $|I_i - I_j|$ is their intensity difference in the original image. The influence of the pairwise terms can be adjusted with their weights w_{pos} and w_{bil} and their effective range is tuned with the kernel widths $\sigma_{\text{pos}}, \sigma_{\text{bil}}$ and σ_{int}.

We estimate the best labelling $\mathbf{x}^* = \arg\min_{\mathbf{x} \in \mathcal{L}^N} E(\mathbf{x})$ using the efficient mean field approximation algorithm of [31]. The weights and kernels of the CRF were chosen using a random search algorithm adapted on the trainind data set.

2.5. Quality Measures

We assessed the performance of our proposed method using the quality metrics introduced in the grand challenges for liver and lesion segmentation by [1, 4].

Our main metric is the Dice score. Additionally we report Volume Overlap Error (VOE), Relative Volume Difference (RVD), Average Symmetric Surface Distance (ASD) and Symmetric Maximum Surface Distance (MSD). Metrics are applied to binary valued volumes, so a metric computed on the lesions for example considers only lesion objects as foreground and everything else as background. We refer to the foreground object in the ground truth as object A, and object B for the predicted object.

2.5.1. Dice Score (DICE)

The Dice score or F1 measure is evaluates as:

$$DICE(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

where the Dice score is in the interval [0,1]. A perfect segmentation yields a Dice score of 1.

2.5.2. Volume Overlap Error (VOE)

VOE is the just the complement of the Jaccard coefficient:

$$VOE(A,B) = 1 - \frac{|A \cap B|}{|A \cup B|}$$

2.5.3. Relative Volume Difference (RVD)

RVD is an asymmetric metric. It is defined as follows:

$$RVD(A,B) = \frac{|B| - |A|}{|A|}$$

2.5.4. Average Symmetric Surface Distance (ASD)

Let $S(A)$ denote the set of surface voxels of A. The shortest distance of an arbitrary voxel v to $S(A)$ is defined as:

$$d(v, S(A)) = \min_{s_A \in S(A)} ||v - s_A||$$

where $||.||$ denotes the Euclidean distance. The average symmetric surface distance is then given by:

$$ASD(A,B) = \frac{1}{|S(A)| + |S(B)|} \left(\sum_{s_A \in S(A)} d(s_A, S(B)) + \sum_{s_B \in S(B)} d(s_B, S(A)) \right)$$

2.5.5. Maximum Surface Distance (MSD)

MSD is also known as the Symmetric Hausdorff Distance. Maximum Surface Distance (MSD) is similar to ASD, except that the maximum distance is taken instead of the average.

$$MSD(A,B) = \max \left\{ \max_{s_A \in S(A)} d(s_A, S(B)), \max_{s_B \in S(B)} d(s_B, S(A)), \right\}$$

3. Experiments and Results

For clinical routine usage, methods and algorithms have to be developed, trained and evaluated on heterogeneous real-life data. In this work we want to demonstrate the robustness, generalization and scalability of our proposed method by applying it to a public dataset for comparison (section 3.1), a clinical CT dataset (section 3.2) and finally a clinical MRI dataset (section 3.3).

3.1. 3DIRCAD

3.1.1. Dataset

We evaluated our proposed method on the 3DIRCADb dataset[4] [32]. In comparison to the grand challenge datasets, the 3DIRCADb dataset offers a higher variety and complexity of livers and its lesions and is publicly available. The 3DIRCADb dataset includes 20 venous phase enhanced CT volumes from various European hospitals with different CT scanners. For our study, we trained and evaluated our models using the 15 volumes containing hepatic tumors in the liver with 2-fold cross validation. The analyzed CT volumes differ substantially in the level of contrast-enhancement, size and number of tumor lesions (1 to 42).

3.1.2. Experimental Setting

Data was prepared as described in section 2.2. Our data augmentation scheme lead to a total training data size of 22693 image slices. The CFCN were trained on a recent desktop PC with a single NVIDIA Titan X GPU with 12 GB VRAM. The neural networks were implemented and trained using the deep learning framework caffe [33] from University of Berkeley. We used stochastic gradient descent as optimizer with a learning rate of 0.001 and a momentum of 0.8. To reduce overfitting we applied a weight decay of 0.0005.

3.1.3. Effect of Class Balancing

The effect of class balancing can be seen in figure 4a - 4d. Introducing class balancing improved the segmentation Dice score on both liver and lesion, while simultaneously decreasing over-fitting. The effect is less for liver, since the percentage of liver voxels in a CT abdomen dataset is on the order of 7%, in comparison to 0.25% for lesions. For all following experiments we accounted for class imbalance by weighting the imbalanced class according to its frequency in the dataset by introducing a weight factor described in section 2.3.3.

3.1.4. Qualitative and Quantitative Results

The qualitative results of the automatic segmentation are presented in figure 7. The complex and heterogeneous structure of the liver and all lesions were detected in the shown images. The cascaded FCN approach yielded an enhancement for lesions with respect to segmentation accuracy compared to a single FCN as can be seen in figure 7. In general, we observe significant[5] additional improvements for Dice overlaps of liver segmentations, from mean Dice 93.1% to 94.3% after applying the 3D CRF. For lesions we could achieve a Dice score of 56% at a standard deviation of 26% with a 2 fold cross-validation.

3.2. Clinical Dataset CT

3.2.1. Dataset

The second dataset we evaluated is a real-life clinical CT dataset from multiple CT scanners and acquired at different centers. It compromises 100 CT scans from different patients. The examined patients were suffering from different kind of cancerous diseases

[4]The dataset is available at `http://ircad.fr/research/3d-ircadb-01`

[5]Two-sided paired t-test with p-value $< 4 \cdot 10^{-19}$

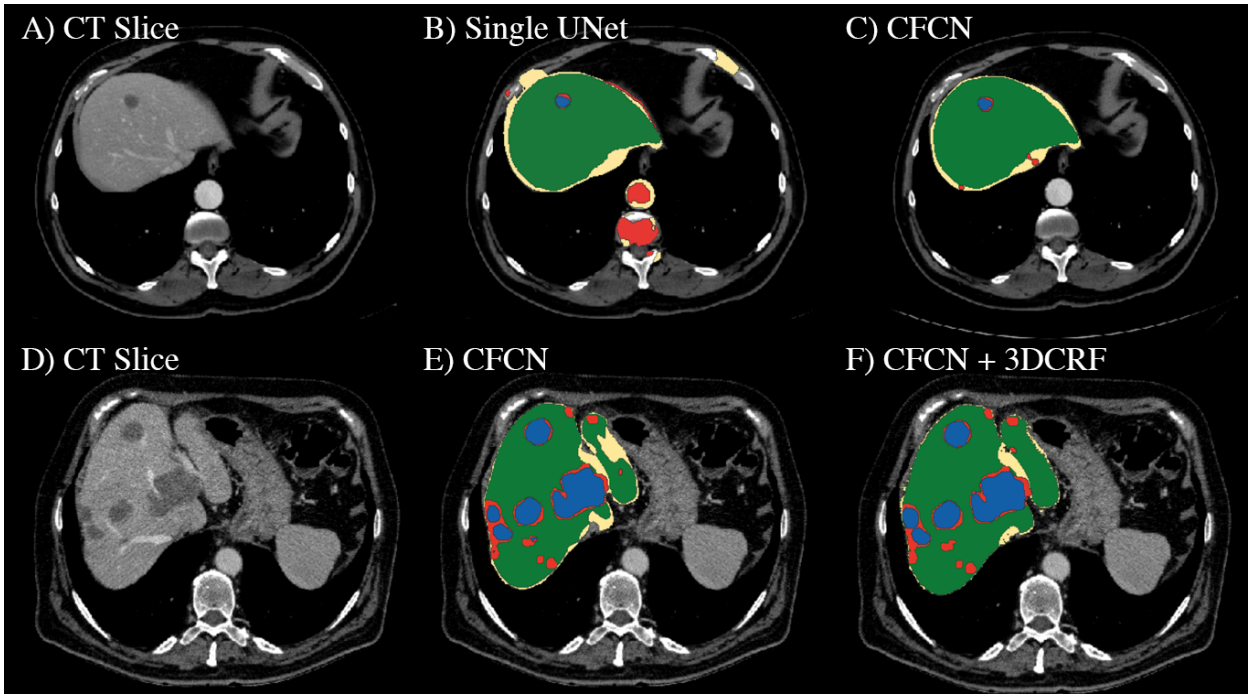


Figure 7: Automatic liver and lesion segmentation with cascaded fully convolutional networks (CFCN) and dense conditional random fields (CRF). Green depicts correctly predicted liver segmentation, yellow for liver false negative and false positive pixels (all wrong predictions), blue shows correctly predicted lesion segmentation and red lesion false negative and false positive pixels (all wrong predictions). In the first row, the false positive lesion prediction in B of a single U-Net as proposed by [18] were eliminated in C by CFCN as a result of restricting lesion segmentation to the liver ROI region. In the second row, applying the 3D CRF to CFCN in F increases both liver and lesion segmentation accuracy further, resulting in a lesion Dice score of 82.3%.

Approach	Dataset	VOE [%]	RVD [%]	ASD [mm]	MSD [mm]	DICE [%]
U-Net as in [18]	3DIRCAD	39	87	19.4	119	72.9
Cascaded U-Net	3DIRCAD	12.8	-3.3	2.3	46.7	93.1
Cascaded U-Net + 3D CRF	3DIRCAD	10.7	-1.4	1.5	24.0	94.3
Li et al. [5] (liver-only)	3DIRCAD	9.2	−11.2	1.6	28.2	
Chartrand et al. [34] (semi-automatic)	3DIRCAD	6.8	1.7	1.6	24	
Li et al. [6] (liver-only)	3DIRCAD					94.5
Cohen et al. [35] (liver-only)	Own Clinical CT					89
Cascaded U-Net	MR-DWI	23	14	5.2	135.3	87
Cascaded U-Net	Clinical CT	22	-3	9.5	165.7	88
Cascaded U-Net + 3D CRF	Clinical CT	16	-6	5.3	48.3	91

Table 1: Quantitative segmentation results of the liver on the 3DIRCADb dataset and other clinical CT am MR-DWI datasets. Scores are reported as presented in the original papers.

with different manifestations in the liver. The dataset ranges from single HCC lesions to diffusive and confluent metastatic lesions. In addition different contrast agents and therefore different levels of contrast enhancement are present in this dataset. Human rater ground truth was obtained through manual volumetric segmentation using the software TurtleSeg[6] [36, 37].

3.2.2. Experimental Setting

The clinical CT dataset was prepared and augmented in the same way as the 3DIRCAD dataset as described in 2.2. The data set was split in 60 for training, 20 for test and 20 for validation. The neural networks, where trained on the same setup and training parameters as the 3DIRCAD dataset. In this experiment, an Adam optimizer was applied with $\epsilon = 0.1$ [38].

3.2.3. Qualitative and Quantitative Results

As shown in table 1 the Cascaded FCN and Cascaded FCN + 3D CRF reach up to 88% and 91% Dice score on this dataset. An inter-rater Dice comparison among 5 training cases yielded a Dice overlap score of 95%. Considering the inter-rater Dice score, the proposed method provides remarkable segmentations. Furthermore, our proposed method achieves a Dice overlap score of $61\% \pm 25\%$ for lesions on the validation set.

3.3. Clinical Dataset MRI

3.3.1. Dataset

To demonstrate the generalization to other modalities we employed our methods to a clinical DW-MRI dataset. 31 Patients underwent clinical assessment and MR imaging for the primary diagnosis of HCC. Imaging was performed using a 1.5 T clinical MRI scanner (Avanto, Siemens) with a standard imaging protocol including axial and coronal T2w, axial T1w images before and after application of Gadolinium-DTPA contrast agent. Diffusion weighted imaging was performed using a slice thickness of 5mm and a matrix size of 192 by 192. The human rater ground truth segmentation was created for the DW-MRI sequence to allow further automatic image analysis e.g. section 3.4.

3.3.2. Experimental Setting

In comparison to the CT datasets, the DW-MRI dataset was prepared differently. The DW-MRI dataset was normalized using the N4Bias correction algorithm [28]. Afterwards the same pre-processing steps were carried out as for CT. The CFCN for the DW-MRI dataset, where trained on the same hardware and training setup. The optimizer in this experiment was an Adam optimizer with $\epsilon = 0.1$.

[6] www.turtleseg.com

3.3.3. Qualitative and Quantitative Results

As seen in figure 8, the CFCN was able to segment the liver lesion correctly. In both cases the CFCN undersegments the lesion leading to a Dice score of 85% in both cases. The quantitative segmentation results are shown in table 1. The Cascaded U-Net was able to reach a dice score for liver in MR-DWI of 87%. For lesion we found a mean dice score of 69.7%.

3.4. HCC Survival Prediction Based on Automatic Liver and Lesion Segmentation

Accurate liver and lesion segmentation are necessary for advanced medical image analysis and are meant to be input to radiomics algorithms, such as the SurvivalNet predictor [27]. In this paragraph we want to introduce a possible applications of our automatic liver and lesion segmentation algorithms in medical imaging. Survival and outcome prediction are important fields in medical image analysis. For hepatic- cellular carcinoma HCC, prior work relied on manual liver and lesion segmentation in DW-MRI to calculate features over the liver and lesion ROI in the ADC sequence to predict patient survival. In contrast to prior work, we trained a CFCN to automatically segment liver and lesion segmentation in DW-MRI to allow automatic survival predictions. We formulate this task as a classification problem with classes being "low risk" and "high risk" represented by longer or shorter survival times than the median survival. We predict HCC malignancy in two steps: As the first step we automatically segment HCC tumor lesions using our proposed method of cascaded fully convolutional neural networks (CFCN). As the second step we predict the HCC lesions' malignancy from the HCC tumor segmentation in the MR-DWI sequence using classical texture features and 3D CNN features. As one of our main contributions we found, that the accuracy of end-to-end assessment of tumor malignancy based on our proposed cascaded fully convolutional neural networks (CFCN) is equal to assessment based on expert annotations with high significance ($p > 0.95$). In other words, our automatic tumor malignancy framework performs equally as assessment based on expert annotations in terms of accuracy. Detailed information can be found in Christ, Ettlinger & Kaissis et al. (2017) [27].

4. Discussion

4.1. Combined Segmentation and Clinical Relevance

In comparison to state-of-the-art, such as [8, 6, 5, 34], we presented a framework, which is capable of a combined segmentation of the liver and its lesion. Moreover, we presented the clinical relevance of our proposed method by utilisation of our automatic segmentations to derive quantitative medical insights. Furthermore, and in contrast to prior work such as [1, 39, 40, 41], our proposed method could be generalized to segment the liver and lesion in different modalities and also multiple organs in medical data. As recent results from natural image segmentation indicate, fully convolutional networks are capable of segmenting dozens of labels with ease. By cascading the FCN architecture to smaller subregions the segmentation accuracy could be further increased. In addition with a runtime per slice

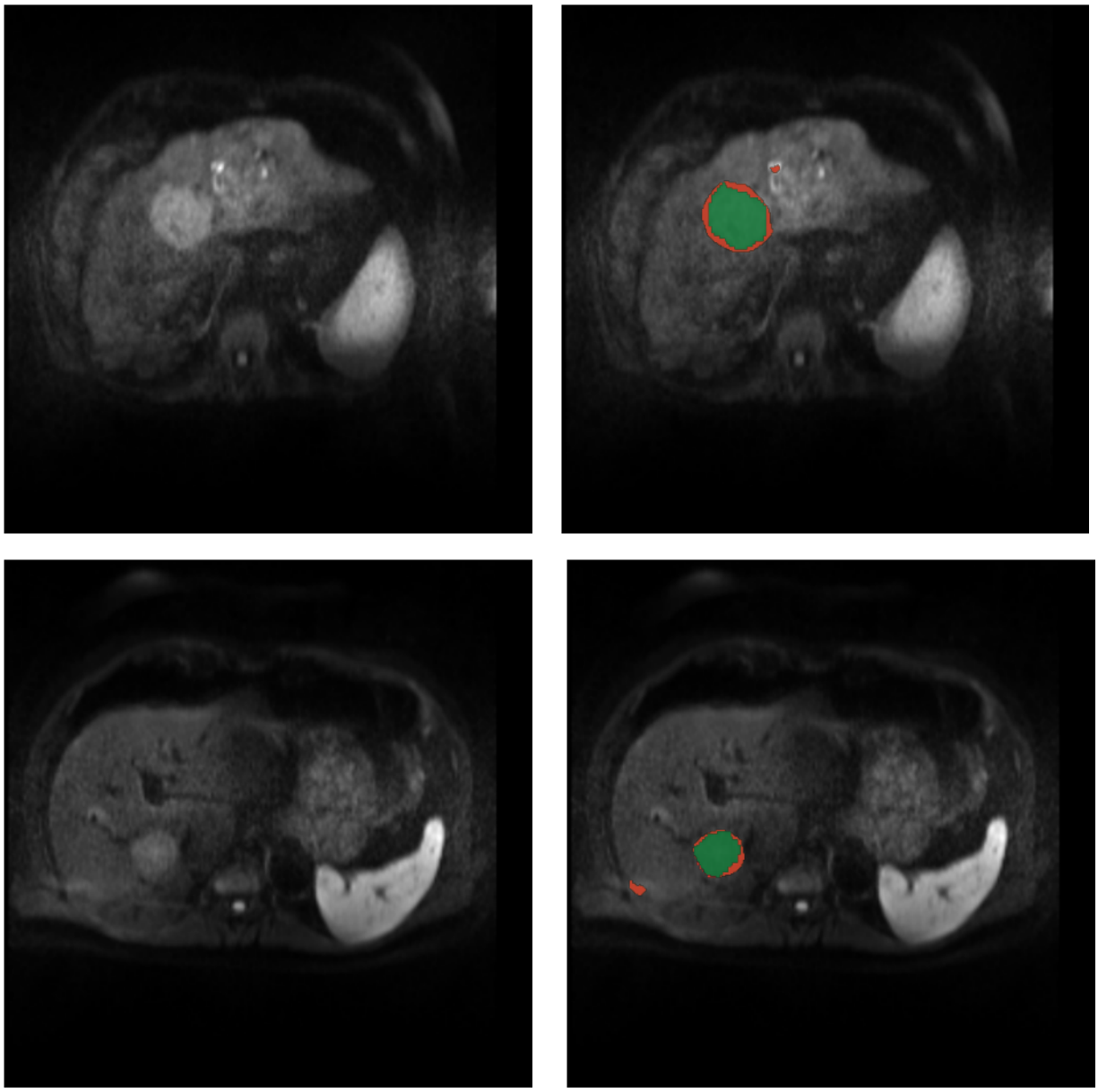


Figure 8: Automatic lesion segmentation with cascaded fully convolutional neural networks (CFCN) in DW-MRI. The raw DW-MRI slices (left), were automatically segmented with our proposed method. Green depicts correctly segmented lesion pixels. Red shows false positive and false negative, i.e. all wrong predictions, of the lesions. In both cases the proposed CFCN achieves an dice score for lesions of 85%.

of 0.19ms and 0.59ms our proposed method enables automatic segmentation of large-scale clinical trials in days and not months [7] using a single desktop PC.

4.2. 3D CNN and FCN Architectures

Recent works such as DeepMedic [22], the V-Net [29] and the 3D U-Net [30] became possible due to efficient implementations of 3D convolutions on GPUs, and they show promising results on their respective segmentation tasks. The proposed idea of cascaded FCN could also be applied to novel 3D CNN and 3D FCN architectures. The restriction of the Region of Interest ROI to relevant organs as shown for the 2D U-Net, when restricting to liver only pixels for segmenting lesions, significantly boosts the segmentation accuracy. The intuition that more specific filters for the underlying problem could be trained, when restricting the relevant regions, holds for 3D as well. Future work will show whether 3D architectures could cope with less training data available for lesion segmentation.

4.3. 3D Conditional Random Field

We showed a statistically significant improvement of segmentation quality, when applying the 3D CRF to our segmentation problem. However, tuning of hyperparameters such as those of the 3D CRF is very time-consuming and task dependent. We found that for highly heterogeneous structures in shape and appearance, such as HCC lesions, it is hard to find a hyperparameter set that generalizes to unseen cases with a random search. A similar conclusion was made in [22] when applying a 3D CRF to heterogeneous brain lesions. Recent work successfully integrated the learning of the CRF hyperparameter in the training process [17]. This approach in combination with additional pairwise terms that incorporate prior knowlegde of the problem could lead to a improvement of the CRF for this task.

5. Conclusion

Cascaded FCNs and dense 3D CRFs trained on CT volumes are suitable for automatic localization and combined volumetric segmentation of the liver and its lesions. Our proposed method competes with state-of-the-art. We provide our trained models under open-source license allowing fine-tuning for other medical applications in CT data [8]. Additionally, we introduced and evaluated dense 3D CRF as a post-processing step for deep learning-based medical image analysis. Furthermore, and in contrast to prior work such as [8, 6, 5], our proposed method could be generalized to segment multiple organs in medical data using multiple cascaded FCNs. As future work, the application of further cascaded FCNs on lesions ROIs to classify malignancy of the lesions as well as advanced techniques such as data augmentation using adversarial networks could enhance the accuracy of the segmentation further. All in all, heterogeneous CT and DW-MRI volumes from different scanners and protocols can be segmented in under 100s each with the proposed approach. We conclude that CFCNs are promising tools for automatic analysis of liver and its lesions in clinical routine and large-scale clinical trials.

[7] Estimating 3000 CT volumes for a large-scale clinical trial

[8] Trained models are available at `https://github.com/IBBM/Cascaded-FCN`

6. Acknowledgement

This work was supported by the German Research Foundation (DFG) within the SFB-Initiative 824 (collaborative research center), "Imaging for Selection, Monitoring and Individualization of Cancer Therapies" (SFB824, project C6) and the BMBF project Softwarecampus. We thank NVIDIA and Amazon AWS for granting GPU and computation support.

References

[1] T. Heimann, et al., Comparison and evaluation of methods for liver segmentation from ct datasets, IEEE Transactions on Medical Imaging 28 (8) (2009) 1251–1265. doi:10.1109/TMI.2009.2013851.

[2] J. Ferlay, H.-R. Shin, F. Bray, D. Forman, C. Mathers, D. M. Parkin, Estimates of worldwide burden of cancer in 2008: Globocan 2008, International Journal of Cancer 127 (12) (2010) 2893–2917.

[3] European Association For The Study Of The Liver, Easl–eortc clinical practice guidelines: management of hepatocellular carcinoma, Journal of Hepatology 56 (4) (2012) 908–943.

[4] X. Deng, G. Du, Editorial: 3d segmentation in the clinic: a grand challenge ii-liver tumor segmentation, in: MICCAI Workshop, 2008.

[5] G. Li, X. Chen, F. Shi, W. Zhu, J. Tian, D. Xiang, Automatic liver segmentation based on shape constraints and deformable graph cut in ct images, Image Processing, IEEE Transactions on 24 (12) (2015) 5315–5329.

[6] C. Li, X. Wang, S. Eberl, M. Fulham, Y. Yin, J. Chen, D. D. Feng, A likelihood and local constraint level set model for liver tumor segmentation from ct volumes, Biomedical Engineering, IEEE Transactions on 60 (10) (2013) 2967–2977.

[7] M. G. Linguraru, W. J. Richbourg, J. Liu, J. M. Watt, V. Pamulapati, S. Wang, R. M. Summers, Tumor burden analysis on computed tomography by automated liver and tumor segmentation, Medical Imaging, IEEE Transactions on 31 (10) (2012) 1965–1976.

[8] A. H. Foruzan, Y.-W. Chen, Improved segmentation of low-contrast lesions using sigmoid edge model, International Journal of Computer Assisted Radiology and Surgery (2015) 1–17.

[9] S. Kadoury, E. Vorontsov, A. Tang, Metastatic liver tumour segmentation from discriminant grassmannian manifolds, Physics in Medicine and Biology 60 (16) (2015) 6459.

[10] M. Freiman, O. Cooper, D. Lischinski, L. Joskowicz, Liver tumors segmentation from cta images using voxels classification and affinity constraint propagation, International Journal of Computer Assisted Radiology and Surgery 6 (2) (2011) 247–255.

[11] R. Vivanti, A. Ephrat, L. Joskowicz, N. Lev-Cohain, O. A. Karaaslan, J. Sosna, Automatic liver tumor segmentation in follow-up ct scans: Preliminary method and results, in: International Workshop on Patch-based Techniques in Medical Imaging, Springer, 2015, pp. 54–61.

[12] A. Ben-Cohen, E. Klang, I. Diamant, N. Rozendorn, M. M. Amitai, H. Greenspan, Automated method for detection and segmentation of liver metastatic lesions in follow-up ct examinations, Journal of Medical Imaging (3).

[13] Y. Häme, M. Pollari, Semi-automatic liver tumor segmentation with hidden markov measure field model and non-parametric distribution estimation, Medical Image Analysis 16 (1) (2012) 140–149.

[14] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: NIPS, 2012, pp. 1097–1105.

[15] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, CVPR.

[16] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille, Semantic image segmentation with deep convolutional nets and fully connected crfs, ICLR.

[17] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, P. Torr, Conditional random fields as recurrent neural networks, ICCV.

[18] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: MICCAI, Vol. 9351, 2015, pp. 234–241.

[19] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, H. Larochelle, Brain Tumor Segmentation with Deep Neural Networks, ArXiv e-printsarXiv:1505.03540.

[20] J. Wang, J. D. MacKenzie, R. Ramachandran, D. Z. Chen, Detection of glands and villi by collaboration of domain knowledge and deep learning, in: MICCAI, 2015, pp. 20–27.

[21] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, M. Nielsen, Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network, in: MICCAI, Vol. 16, 2013, pp. 246–253.

[22] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, B. Glocker, Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation, Medical Image Analysis 36 (2017) 61–78.

[23] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, R. M. Summers, Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation, in: MICCAI, 2015, pp. 556–564.

[24] H. Chen, Q. Dou, L. Yu, P.-A. Heng, Voxresnet: Deep voxelwise residual networks for volumetric brain segmentation, arXiv preprint arXiv:1608.05895.

[25] M. F. Stollenga, W. Byeon, M. Liwicki, J. Schmidhuber, Parallel multi-dimensional lstm, with application to fast biomedical volumetric image segmentation, in: Advances in Neural Information Processing Systems, 2015, pp. 2998–3006.

[26] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D'Anastasi, W. H. Sommer, S.-A. Ahmadi, B. H. Menze, Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields, MICCAI, Cham, 2016, pp. 415–423.

[27] P. F. Christ, F. Ettlinger, G. Kaissis, S. Schlecht, F. Ahmaddy, F. Grün, A. Valentinitsch, S.-A. Ahmadi, R. Braren, B. Menze, SurvivalNet: Predicting patient survival from diffusion weighted magnetic resonance images using cascaded fully convolutional and 3D convolutional neural networks, ArXiv e-prints 1702.05941.

[28] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, J. C. Gee, N4ITK: Improved N3 bias correction, IEEE Transactions on Medical Imaging 29 (6) (2010) 1310–1320. doi:10.1109/TMI.2010.2046908.

[29] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 3D Vision (3DV), 2016 Fourth International Conference on, IEEE, 2016, pp. 565–571.

[30] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, O. Ronneberger, 3d u-net: learning dense volumetric segmentation from sparse annotation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 424–432.

[31] P. Krähenbühl, V. Koltun, Efficient inference in fully connected crfs with gaussian edge potentials, in: NIPS, 2011, pp. 109–117.

[32] L. Soler, A. Hostettler, V. Agnus, A. Charnoz, J. Fasquel, J. Moreau, A. Osswald, M. Bouhadjar, J. Marescaux, 3d image reconstruction for comparison of algorithm database: a patient-specific anatomical and medical image database (2012).
URL http://www-sop.inria.fr/geometrica/events/wam/abstract-ircad.pdf

[33] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: Convolutional architecture for fast feature embedding, in: Proceedings of the ACM International Conference on Multimedia, ACM, 2014, pp. 675–678.

[34] G. Chartrand, T. Cresson, R. Chav, A. Gotra, A. Tang, J. DeGuise, Semi-automated liver ct segmentation using laplacian meshes, in: ISBI, IEEE, 2014, pp. 641–644.

[35] A. Ben-Cohen, I. Diamant, E. Klang, M. Amitai, H. Greenspan, Fully convolutional network for liver segmentation and lesions detection, in: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, Springer, 2016, pp. 77–85.

[36] A. Top, G. Hamarneh, R. Abugharbieh, Spotlight: Automated confidence-based user guidance for increasing efficiency in interactive 3d image segmentation, in: MICCAI, 2010, pp. 204–213.

[37] A. Top, G. Hamarneh, R. Abugharbieh, Active learning for interactive 3d image segmentation, in:

MICCAI, Vol. 6893, 2011, pp. 603–610.
[38] D. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980.
[39] M. Goryawala, M. R. Guillen, M. Cabrerizo, A. Barreto, S. Gulec, T. C. Barot, R. R. Suthar, R. N. Bhatt, A. Mcgoron, M. Adjouadi, A 3-d liver segmentation method with parallel computing for selective internal radiation therapy, Transactions on Information Technology in Biomedicine 16 (1) (2012) 62–69.
[40] F. López-Mir, P. González, V. Naranjo, E. Pareja, S. Morales, J. Solaz-Mínguez, A method for liver segmentation on computed tomography images in venous phase suitable for real environments, Journal of Medical Imaging and Health Informatics 5 (6) (2015) 1208–1216.
[41] J. Peng, Y. Wang, D. Kong, Liver segmentation with constrained convex variational model, Pattern Recognition Letters 43 (2014) 81 – 88.

Automated Unsupervised Segmentation of Liver Lesions in CT scans via Cahn-Hilliard Phase Separation

Authoren: Jana Lipková, Markus Rempfler, Patrick Ferdinand Christ, John Lowengrub, Bjoern H. Menze

Abstract: The segmentation of liver lesions is crucial for detection, diagnosis and monitoring progression of liver cancer. However, design of accurate automated methods remains challenging due to high noise in CT scans, low contrast between liver and lesions, as well as large lesion variability. We propose a 3D automatic, unsupervised method for liver lesions segmentation using a phase separation approach. It is assumed that liver is a mixture of two phases: healthy liver and lesions, represented by different image intensities polluted by noise. The Cahn-Hilliard equation is used to remove the noise and separate the mixture into two distinct phases with well-defined interfaces. This simplifies the lesion detection and segmentation task drastically and enables to segment liver lesions by thresholding the Cahn-Hilliard solution. The method was tested on 3Dircadb and LITS dataset.

Individuelle Leistungsbeiträge: Datenakquise und Datenaufbereitung, Revisionen des Manuskripts

Unveröffentlichtes Manuskript

Automated Unsupervised Segmentation of Liver Lesions in CT scans via Cahn-Hilliard Phase Separation

Jana Lipková[1], Markus Rempfler[1], Patrick Christ[1], John Lowengrub[2], Bjoern H. Menze[1] jana.lipkova@tum.de

[1] Department of Informatics & Institute for Advanced Study, Technical University of Munich, Germany

[2] Departments of Mathematics, & Center for Complex Biological Systems & Chao Family Comprehensive Cancer Center, University of California, Irvine, USA

Abstract. The segmentation of liver lesions is crucial for detection, diagnosis and monitoring progression of liver cancer. However, design of accurate automated methods remains challenging due to high noise in CT scans, low contrast between liver and lesions, as well as large lesion variability. We propose a 3D automatic, unsupervised method for liver lesions segmentation using a phase separation approach. It is assumed that liver is a mixture of two phases: healthy liver and lesions, represented by different image intensities polluted by noise. The Cahn-Hilliard equation is used to remove the noise and separate the mixture into two distinct phases with well-defined interfaces. This simplifies the lesion detection and segmentation task drastically and enables to segment liver lesions by thresholding the Cahn-Hilliard solution. The method was tested on 3Dircadb and LITS dataset.

1 Introduction

Liver is one of the most common cancer sites, including primary tumours like hepatocellular carcinoma and metastatic tumours that have spread from the breast, colon and prostate. Computer tomography (CT) is routinely used to detect and evaluate treatment response of liver lesions. In clinical practise, liver lesions are segmented by manual or semi-manual methods. However, these are time consuming and subjective, with an intra- and interobserver variability up to 11 % in volume difference on liver CT scans [1]. To overcome these difficulties, several semi-automated and automated methods were proposed. Semi-automated methods include support vector machine with affinity constrains propagation [2], hidden Markov fields [3], level set methods [4], sigmoid edge modelling [5] and mathematical morphology [6]. Automated methods include k-means classification [7], object-based image analysis [8] and convolutional neural networks [9]. An advantage of automated over semiautomated methods is their reproducibility, since they do not require human interactions. Despite significant efforts, the performance of automatic methods remains relatively poor, especially in comparison with segmentation methods for other lesion sites. The main challenges

of liver lesion segmentation include high levels of noise, low liver-lesion contrast and variations of image intensities caused by different acquisition protocols, tissue abnormalities such as surgical resection, metal implants and changes due to treatment. For instance, the mean liver CT values of 3Dircadb [10] datasets vary by an order of magnitude, which complicates the use of intensity based methods. Furthermore, a significant variation in lesions shape and structure compromise efficiency of supervised methods.

We propose a novel automated unsupervised method for the enhancement and segmentation of hypointense lesions in liver CT scans via phase field separation. In chemistry, phase separation is a mechanism in which a mixture of two components separates into distinct phases with different chemical compositions. The Cahn-Hilliard equation is a partial differential equation that describes phase separation driven by gradients in chemical potentials [11]. We consider liver CT as a mixture of two phases, healthy liver and lesions, represented by different image intensities. The Cahn-Hilliard equation is used to remove the noise and separate the mixture into two distinct phases with well defined interface separating the phases. The lesions are then segmented by thresholding the Cahn-Hilliard solution. This approach has several desirable properties: it is 3D, edge preserving and robust to noise, variation of intensities and lesions diversity. In comparison to other edge preserving smoothing methods, including bilateral and image guided filtering or anisotropic diffusion, phase separation is an energy minimisation problem which can be applied to data with different noise or image intensities.

2 Method

The Cahn-Hilliard equation describes the spatiotemporal evolution of phase separation in a mixed system. Let us assume a system with two phases, A and B. The state of the system at spacial location $(x, y, z) \in \mathbb{R}^3$ and time t can be represented by a phase field function (pff) $\psi := \psi(x, y, z, t) \in [0, 1]$, with $\psi = 1$ and $\psi = 0$ indicating domains of the separated phases. The free energy of the system in a domain Ω can be modelled as [11]

$$E_\varepsilon(\psi) = \int_\Omega f(\psi) + \frac{\varepsilon^2}{2} |\nabla\psi|^2 \, dV, \tag{1}$$

where $f(\psi)$ is the bulk free-energy density in phases A and B, ε is the prescribed interface thickness and $\frac{\varepsilon^2}{2}|\nabla\psi|^2$ is the additional free-energy density at the interfaces between the phases. To ensure a separation in two distinguish phases, it is assumed that $f(\psi)$ is a double-well potential, which can be modelled as $f(\psi) = \frac{1}{4}\psi^2(1-\psi)^2$. Then phase separation of the system driven by the difference in chemical potentials between the phases can be modelled by the Cahn-Hilliard equations:

$$\frac{\partial\psi}{\partial t} = \nabla \cdot (M(\psi)\nabla\mu), \quad \in \Omega, \tag{2}$$

$$\mu = \frac{\delta E_\varepsilon(\psi)}{\delta\psi} := \frac{df(\psi)}{d\psi} - \varepsilon^2 \Delta\psi, \tag{3}$$

where μ is the chemical potential of the system, defined as the variational derivation of the systems free energy. Taking the mobility term $M(\psi) = \sqrt{4f(\psi)}$ inhibits long-range diffusion and tends to preserve the volumes of the individual lesions. The Cahn-Hilliard equation describes the evolution of a system with high energy, represented by mixed phases, to a system with lower energy characterised by the separated phases. In contrast to ill-posted anisotropic diffusion problem [12], the existence and uniqueness of the Cahn-Hilliard solution is guaranteed by the existence of the free energy (Lyapunov) functional. Details on the derivation and properties of the Cahn-Hilliard equation can be found in [13].
The equation (2) is discretised by finite differences in space and forward Euler in time. It is implemented in a multi-resolution adapted grid solver, a 3D extension of the 2D solver presented in [14]. The adaptive grid approach enables fast evaluation, with typical simulation time around 7 minutes per liver volume with 4 Opteron6174 cores.

3 Experiments and results

The Cahn-Hilliard separation (CHS) method consists of three main steps: data preprocessing, phase separation and lesions segmentation. The workflow of the CHS method is depicted in Fig.1.

Data and data preprocessing: We conducted experiments on training dataset from the Liver Tumor Segmentation Challenge (LITS) [15], which also includes 3Dircadb data. The LITS datase contains abdominal CT scans acquired at various centres, with different acquisition protocols and resolution. All data are preprocessed in the following way. First, the abdominal CT scan is cropped into a box Ω containing a liver mask χ, (Fig. 1 A). Liver intensities are then clipped to a range $[0, 200]$ HU, to exclude atypical liver values like metal implants. To account for specific image intensities, a 95% credibility interval $[a, b]$ of the clipped liver CT is computed. The clipped liver CT is then clip one more time to the range $[0, b]$. Afterwards, the liver CT is normalised to take values in $[0, 1]$ and this is used to initialise the pff ψ within χ (Fig. 1 B). To prevent border artefacts, the background of the liver CT is considered as a healthy liver. This is achieved by assigning a liver-like intensity to the phase field function outside the liver, i.e. it is assumed that $\psi = 0.55$ in $\Omega \setminus \chi$.

Phase separation: The thickness of the liver-lesion interface ε is set to 6 voxels, i.e 3 voxels of smoothening per phase, which is sufficient to smooth out noise but preserve small lesions. Eq. (2), with ψ defined above and no-flux boundary conditions on $\partial\Omega$ is evolved in time until the systems energy E_ε approaches its minimum (Fig. 2). In all tests, 700 times steps (iterations) are found to be sufficient to capture changes in the energy. Figure 2 shows, that the solution of the systems does not change significantly by evolving the system longer in time. The solution of the Cahn-Hilliard equation, with initial CT scan on Fig.

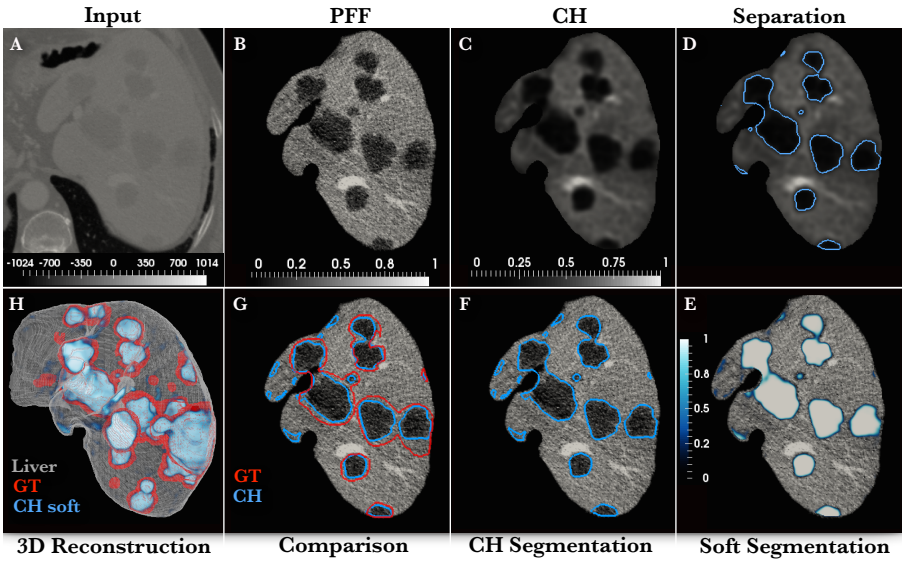


Fig. 1. A-H: A workflow of phase separation and lesions segmentation. A) cropped liver CT scan, converted into a phase field function (PFF) B), is used as initial condition for the Cahn-Hilliard (CH) equation, with the solution shown on C). D) the hard separation of the phases (blue line), is converted into a soft probabilistic segmentation E). F) final hard segmentation (blue) in comparison with the ground truth (GT) (red) F). H) comparison of the GT (red) and soft segmentation (white-blue colormap) in 3D representation.

1 A), is shown on Fig. 1 C). The phase separation dynamics removed noise and enhanced the liver-lesion contrast, while preserving the interface. The separation of the phases is also apparent from the image intensity histogram (IIH) of the normalised liver CT before and after the separation (Fig. 3). The histogram is divided into 255 bins corresponding to the gray scale levels. The spikes visible on the original liver CT histogram are caused by anisotropic data resolution. After the CHS, the originally unimodal liver CT histogram separates into two modes, one for each phase, allowing lesions segmentation by histogram thresholding. In the case of binary system, the separation of the phases is give by $\psi = 0.5$. However, this is not the case for the heterogeneous liver scan.

Lesions segmentation: Several methods have been proposed for automated histogram separation including the Otsu, Triangle and Isodata methods. However, these methods failed to detect the separation in the case of small lesions. Instead, we propose to compute the separation by detecting local maxima (peaks) of the IIH. The i-th element of the image histogram $iih(i)$, is defined as a peak, if $iih(i+1) - 2\,iih(i) + iih(i-1) < 0$. Let $\mathbf{p}$ be a vector of the detected peak locations and $\mathbf{I(p)}$ the corresponding image intensities. Let p_j be the global max-

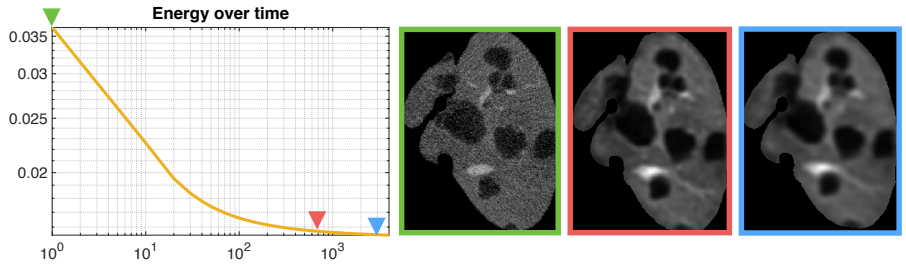


Fig. 2. Log-log plot showing evolution of the system from high energy state, caused by mixed liver-lesion interface, to low energy state with separated phases.

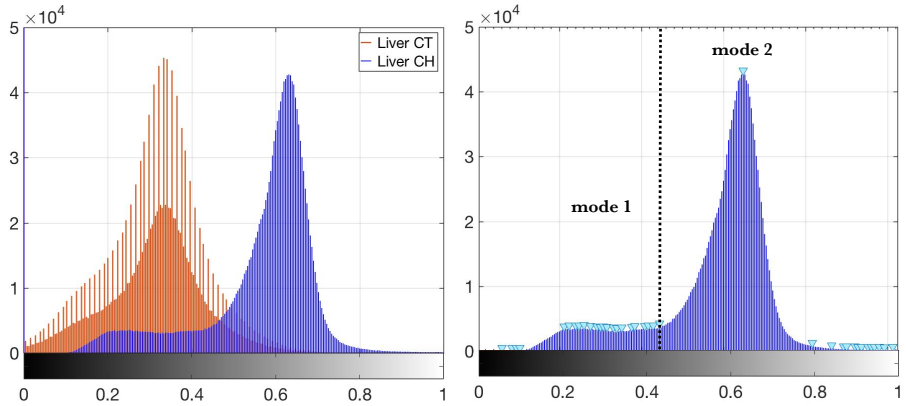


Fig. 3. Left: comparison of image intensity histogram (IIH) of the normalised liver CT before (red) and after the separation (blue). Right: IIH after the phase separation, with the detected local maxima indicated by the triangles. The dashed line indicates hard separation between lesions (mode 1) and liver (mode 2).

imum of $\mathbf{p}$ with intensity $I(p_j)$. Then the peak p_k indicating separation between liver-lesion modes is identified by the Algorithm 1. The while-loop in the algorithm ensures a correct histogram separation even if multiple peaks are detected within the liver mode. The separation of lesion (mode 1) and liver (mode 2) is shown in Fig. 3 (right) and the corresponding separation of the Cahn-Hilliard solution at the intensity $I_0 = I(p_k)$ is shown on Fig. 1 D). However, a single iso-value I_0 might not be optimal for all lesions, especially small lesions might be under segmented. To overcome this issue, the hard interface separation I_0 can be translated into a soft probabilistic one as follows. The phase interface after CHS can be approximated by hyperbolic tangent

$$\psi_{soft}(I) = \frac{1}{2}\left[1 + \tanh\left(\frac{(I_0 - I)}{2\sqrt{2}\varepsilon}\right)\right] = \frac{1}{1 + \exp\left(-\frac{(I_0 - I)}{\sqrt{2}\varepsilon}\right)}. \tag{4}$$

Algorithm 1: Histogram separation by detecting local maxima.

$k = j - 1$
while $I(p_k) > 0.75 \times I(p_j)$ **do**
  $k = k - 1$;
end
if $(k = 0)$ **then**
  $I(p_k) := 0$;
end

The soft probabilistic segmentation (Fig. 1 E) is thresholded to $[0.15, 1]$ range to obtain the final segmentation (Fig. 1 F). Figure 1 (G and D) shows a comparison of CHS and the ground truth (GT) segmentation. The interface between liver and lesions is preserved in CHS, which leads to a better lesions delineation compare to the manual segmentation, which tends to over-segment some lesions.

3.1 Qualitative results

Figure 4 A) shows the capability of the method to enhance and detect small lesions. In this case the lesion mode in the IIH is less pronounced, nevertheless the correct separation is still detected. Figure 4 B) shows a liver volume with metal implants. Since CHS depends only on difference in image intensities, not the absolute values, presence of the metal artefacts does not influence the segmentation. Furthermore, using the metal implants as landmarks, it can be seen that the liver-lesion interface is preserved. However, segmentation based only on intensity thresholding can not distinguish between lesions and other artefacts with similar intensities, which might appear at the liver border or in regions of liver folding (Fig. 4 C). Furthermore, the method is not able to detect very small lesions in low resolution data, i.e. small lesions that appear only in 1-2 slices.

3.2 Quantitative results

The CHS method was tested on the hypointense lesion from the LITS training set 2. For comparison purposes, the set was divided into two groups: 1) 3Dircadb dataset and 2) the rest of the set, referred to as LITS-hypo. Table 1 shows results of the CHS method in comparison with other automatic methods. High detection rate illustrates the capability of the method to enhance and separate lesions. On the 3Dircadb dataset, the CHS method performed better than the convolutional neural networks [9]. The Dice scores on LITS-hypo test are lower than on 3Dircadb set for two reasons. First, the set contains several very small lesions present only in 1-2 slices. Second, the CHS method depends on the quality of the liver segmentation. Liver foldings and shadows at the liver borders tends to increase the number of false positives. A comparison with other methods on LITS training set is currently not possible, however this set helped to identify weak points of the CHS method. These weaknesses could be addressed by applying a classifier trained to distinguish between lesions and other artefacts.

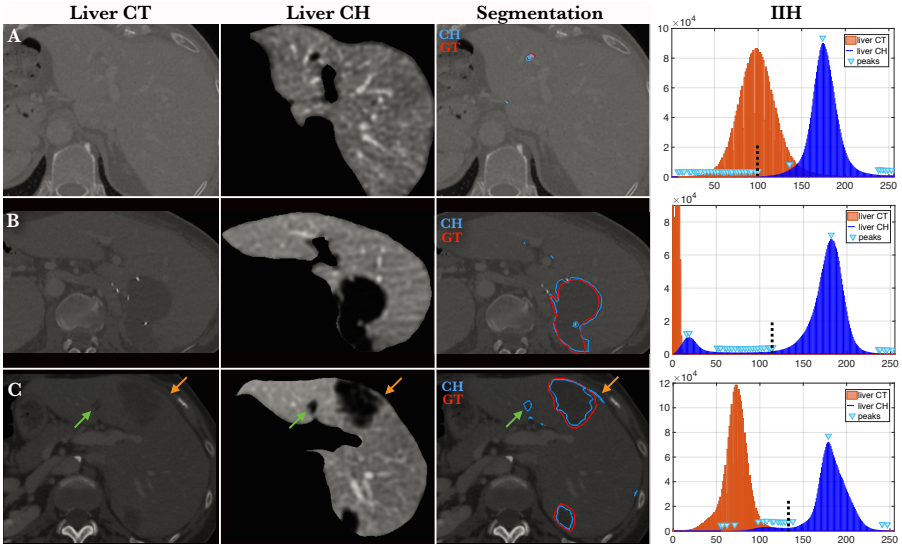


Fig. 4. Qualitative results showing original liver CT scan, liver after Cahn-Hilliard (CH) separation and comparison of the ground truth (GT) (red) and CH (blue) segmentation for three cases: A) case with small lesion, B) case with metal implants, C) case with segmentations artefacts at the liver border (orange arrow) and at region of liver folding (green arrow). Last column shows image intensity histogram (IIH) of the normalised liver CT before (orange) and after separation (blue). Cyan triangles indicate detected peaks, while the liver-lesion separation is marked by the dashed line.

Table 1. Quantitative results of automatic liver lesions segmentation methods. Scores are reported as presented in the original papers.

Approach	Dataset	Dice	Sensitivity	Specificity	Precision	Detection
CHS	3Dircadb	0.61 ± 0.22	0.64 ± 0.18	0.99 ± 0.01	0.65 ± 0.27	0.73 ± 0.25
CHS	LITS-hypo	0.53 ± 0.27	0.70 ± 0.21	0.98 ± 0.02	0.52 ± 0.30	0.85 ± 0.20
Christ [9]	3Dircadb	0.56 ± 0.27	-	-	-	-
Schweir [8]	private	-	-	-	0.53	0.77
Massoptier [7]	private	-	0.82	0.87	-	-

4 Conclusion

We have presented a novel automated and unsupervised method for segmentation of lesions in liver CT scans. The ability of the CHS method to enhance and detect lesions, allows to reach state-of-the-art results with simple thresholding of the Chan-Hilliard solution. We expect that combining the CHS with more discriminative learning approaches will enhance the quality of the segmentations. Application of the CHS method is not limited only to liver lesions and similar structures as lesions in spleen or ultrasound images. By modification of the

chemical potential, the CHS method can be used for separation of multiple phases, making it a promising tool for image preprocessing.

References

1. Bellon, E., Feron, M., Maes, F., Hoe, L.V., Delaere, D. et al.: Evaluation of manual vs semi-automated delineation of liver lesions on ct images. European Radiology 7(3), 432–438 (1997)
2. Freiman, M., Cooper, O., Lischinski, D., Joskowicz, L.: Liver tumors segmentation from cta images using voxels classification and affinity constraint propagation. International journal of computer assisted radiology and surgery 6(2), 247–255 (2011)
3. Häme, Y., Pollari, M.: Semi-automatic liver tumor segmentation with hidden markov measure field model and non-parametric distribution estimation. Medical image analysis 16(1), 140–149 (2012)
4. Li, C., Wang, X., Eberl, S., Fulham, M., Yin, et al.: A likelihood and local constraint level set model for liver tumor segmentation from ct volumes. IEEE Transactions on Biomedical Engineering 60(10), 2967–2977 (2013)
5. Foruzan, A.H., Chen, Y.W.: Improved segmentation of low-contrast lesions using sigmoid edge model. International journal of computer assisted radiology and surgery 11(7), 1267–1283 (2016)
6. Belgherbi, A., Hadjidj, I., Bessaid, A.: A semi-automated method for the liver lesion extraction from a ct images based on mathematical morphology. Journal of Biomedical Sciences 2(2) (2014)
7. Massoptier, L., Casciaro, S.: A new fully automatic and robust algorithm for fast segmentation of liver tissue and tumors from ct scans. European radiology 18(8), 1658 (2008)
8. Schwier, M., Moltz, J.H., Peitgen, H.O.: Object-based analysis of ct images for automatic detection and segmentation of hypodense liver lesions. International journal of computer assisted radiology and surgery 6(6), 737 (2011)
9. Christ, P.F., Ettlinger, F., Grün, F., Elshaera, M.E.A., Lipkova, J. et al. Automatic Liver and Tumor Segmentation of CT and MRI Volumes using Cascaded Fully Convolutional Neural Networks. ArXiv e-prints (Feb 2017)
10. Ircad Dataset. www.ircad.fr/research/3dircadb.
11. Cahn, J.W., Hilliard, J.E.: Free energy of a nonuniform system. i. interfacial free energy. The Journal of chemical physics 28(2), 258–267 (1958)
12. Guidotti, P.: Anisotropic diffusions of image processing from perona-malik on. Advanced Studies in Pure Mathematics 99, 20XX (2015)
13. Lee, D., Huh, J.Y., Jeong, D., Shin, J., Yun, A., Kim, J.: Physical, mathematical, and numerical derivations of the cahn–hilliard equation. Computational Materials Science 81, 216–225 (2014)
14. Rossinelli, D., Hejazialhosseini, B., Rees, W. et al.: Mrag-i2d: Multi-resolution adapted grids for remeshed vortex methods on multicore architectures. Journal of Computational Physics 288, 1–18 (2015)
15. LITS Dataset, https://competitions.codalab.org/competitions/15595