Informatik aktuell

H. Handels T. M. Deserno
A. Maier K. H. Maier-Hein
C. Palm T. Tolxdorff
(Hrsg.)

Bildverarbeitung für die Medizin 2019

Algorithmen
Systeme
Anwendungen

Proceedings des Workshops vom 17. bis 19. März 2019 in Lübeck

Springer Vieweg

Informatik aktuell

Herausgegeben
im Auftrag der Gesellschaft für Informatik (GI)

Heinz Handels · Thomas M. Deserno
Andreas Maier · Klaus H. Maier-Hein
Christoph Palm · Thomas Tolxdorff
Herausgeber

Bildverarbeitung für die Medizin 2019

Algorithmen – Systeme – Anwendungen

Proceedings des Workshops
vom 17. bis 19. März 2019 in Lübeck

Springer Vieweg

Herausgeber

Heinz Handels
Universität zu Lübeck
Institut für Medizinische Informatik
Ratzeburger Allee 160, 23562 Lübeck

Thomas M. Deserno, geb. Lehmann
Technische Universität Braunschweig
und Medizinische Hochschule Hannover
Peter L. Reichertz Institut für Medizinische Informatik
Mühlenpfordtstr. 23, 38106 Braunschweig

Andreas Maier
Friedrich-Alexander-Universität
Erlangen-Nürnberg
Lehrstuhl für Mustererkennung
Martensstr. 3, 91058 Erlangen

Klaus H. Maier-Hein, geb. Fritzsche
Deutsches Krebsforschungszentrum (DKFZ)
Medical Image Computing E230
Im Neuenheimer Feld 581, 69120 Heidelberg

Christoph Palm
Ostbayerische Technische Hochschule
Regensburg
Fakultät für Informatik und Mathematik
Galgenbergstr. 32, 93053 Regensburg

Thomas Tolxdorff
Charité – Universitätsmedizin Berlin
Institut für Medizinische Informatik
Hindenburgdamm 30, 12200 Berlin

ISSN 1431-472X
Informatik aktuell
ISBN 978-3-658-25325-7 ISBN 978-3-658-25326-4 (eBook)
https://doi.org/10.1007/978-3-658-25326-4

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

CR Subject Classification (1998): A.0, H.3, I.4, I.5, J.3, H.3.1, I.2.10, I.3.3, I.3.5, I.3.7, I.3.8, I.6.3

Springer Vieweg

© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2019

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von allgemein beschreibenden Bezeichnungen, Marken, Unternehmensnamen etc. in diesem Werk bedeutet nicht, dass diese frei durch jedermann benutzt werden dürfen. Die Berechtigung zur Benutzung unterliegt, auch ohne gesonderten Hinweis hierzu, den Regeln des Markenrechts. Die Rechte des jeweiligen Zeicheninhabers sind zu beachten.

Der Verlag, die Autoren und die Herausgeber gehen davon aus, dass die Angaben und Informationen in diesem Werk zum Zeitpunkt der Veröffentlichung vollständig und korrekt sind. Weder der Verlag noch die Autoren oder die Herausgeber übernehmen, ausdrücklich oder implizit, Gewähr für den Inhalt des Werkes, etwaige Fehler oder Äußerungen. Der Verlag bleibt im Hinblick auf geografische Zuordnungen und Gebietsbezeichnungen in veröffentlichten Karten und Institutionsadressen neutral.

Springer Vieweg ist Teil von Springer Nature
Springer Vieweg ist ein Imprint der eingetragenen Gesellschaft Springer Fachmedien Wiesbaden GmbH und ist ein Teil von Springer Nature.
Die Anschrift der Gesellschaft ist: Abraham-Lincoln-Str. 46, 65189 Wiesbaden, Germany

Bildverarbeitung für die Medizin 2019

Veranstalter

IMI	Institut für Medizinische Informatik, Universität zu Lübeck

Unterstützende Fachgesellschaften

BVMI	Berufsverband Medizinischer Informatiker
CURAC	Computer- und Roboterassistierte Chirurgie
DAGM	Deutsche Arbeitsgemeinschaft für Mustererkennung
DGBMT	Fachgruppe Medizinische Informatik der Deutschen Gesellschaft für Biomedizinische Technik im Verband Deutscher Elektrotechniker
GI	Gesellschaft für Informatik – Fachbereich Informatik in den Lebenswissenschaften
GMDS	Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie
IEEE	Joint Chapter Engineering in Medicine and Biology, German Section

Tagungsvorsitz

Prof. Dr. rer. nat. habil. Heinz Handels
Institut für Medizinische Informatik, Universität zu Lübeck

Tagungssekretariat

Susanne Petersen
Institut für Medizinische Informatik
Ratzeburger Allee 160, Gebäude 64
23562 Lübeck
Tel: +49 451 3101 5601
Fax: +49 451 3101 5604
Email: vm2019@imi.uni-luebeck.de
Web: http://bvm-workshop.org

Lokale BVM-Organisation

Dr. Jan Ehrhardt
Prof. Dr. Heinz Handels (Leitung)
Prof. Dr. Mattias Heinrich
Susanne Petersen
Dr. Jan Wrage
und weitere Mitarbeiterinnen und Mitarbeiter des Instituts für Medizinische Informatik der Universität zu Lübeck

Verteilte BVM-Organisation

Prof. Dr. Thomas M. Deserno, Sven Neumann, Aaron Wiora, Jamie-Céline Heinzig, Noah Maxen, Björn Bajor, Peter Reichert, Aleksej Hecht, Hendrik Griesche – Peter L. Reichertz Institut für Medizinische Informatik, Technische Universität Braunschweig (Tagungsband)
Prof. Dr. Heinz Handels, Dr. Jan-Hinrich Wrage – Institut für Medizinische Informatik, Universität zu Lübeck (Beitragsbegutachtung)
Prof. Dr. Andreas Maier – Pattern Recognition Lab, Technische Fakultät der FAU – Nürnberg (Social Media, Special Issue)
Priv. – Doz. Dr. Klaus Maier – Hein, Jens Petersen – Division of Medical Image Computing, Deutsches Krebsforschungszentrum (DKFZ) (Anmeldung, Mailingliste)
Prof. Dr. Christoph Palm, Dr. Alexander Leis, Leonard Klausmann, Sümeyye R. Yildiran – Regensburg Medical Image Computing (ReMIC), OTH Regensburg (Internetpräsenz, Newsletter, Social Media)
Prof. Dr. Thomas Tolxdorff, Dr. Thorsten Schaaf – Institut für Medizinische Informatik, Charité – Universitätsmedizin Berlin (Internetpräsenz)

BVM-Komitee

Prof. Dr. Thomas M. Deserno, Peter L. Reichertz Institut für Medizinische Informatik, Technische Universität Braunschweig
Prof. Dr. Heinz Handels, Institut für Medizinische Informatik, Universität zu Lübeck
Prof. Dr. Andreas Maier, Pattern Recognition Lab, FAU Erlangen-Nürnberg
Priv.-Doz. Dr. Klaus Maier-Hein, Division of Medical Image Computing, DKFZ
Prof. Dr. Christoph Palm, Regensburg Medical Image Computing, OTH Regensburg
Prof. Dr. Thomas Tolxdorff, Institut für Medizinische Informatik,
Charité – Universitätsmedizin Berlin

Programmkommitee

Prof. Dr. Jörg Barkhausen, Universitätsklinikum Schleswig-Holstein Campus Lübeck
Priv.-Doz. Dr. Jürgen Braun, Charité-Universitätsmedizin Berlin
Prof. Dr. Thorsten Buzug, Universität zu Lübeck
Dr. Stefanie Demirci, TU München
Prof. Dr. Thomas Deserno, TU Braunschweig
Prof. Dr. Hartmut Dickhaus, Universität Heidelberg
Dr. Jan Ehrhardt, Universität zu Lübeck
Dr. Ralf Floca, DKFZ Heidelberg
Prof. Dr. Nils Forkert, University of Calgary, Canada
Prof. Dr. Horst Hahn, Fraunhofer MEVIS, Bremen

Prof. Dr. Heinz Handels, Universität zu Lübeck
Dr. Tobias Heimann, Siemens, Erlangen
Prof. Dr. Mattias Heinrich, Universität zu Lübeck
Prof. Dr. Ron Kikinis, Fraunhofer MEVIS, Bremen und Harvard Medical School, Boston, USA
Prof. Dr. Andreas Maier, Universität Erlangen-Nürnberg
Priv.-Doz. Dr. Klaus Maier-Hein, DKFZ Heidelberg
Prof. Dr. Lena Maier-Hein, DKFZ Heidelberg
Prof. Dr. Thomas Martinetz, Universität zu Lübeck
Priv.-Doz. Dr. Andre Mastmeyer, Universität zu Lübeck
Prof. Dr. Hans-Peter Meinzer, DKFZ Heidelberg
Prof. Dr. Dorit Merhof, RWTH Aachen
Prof. Dr. Alfred Mertins, Universität zu Lübeck
Prof. Dr. Jan Modersitzki, Fraunhofer MEVIS Lübeck und Universität zu Lübeck
Prof. Dr. Heinrich Müller, TU Dortmund
Prof. Dr. Nassir Navab, TU München
Dr. Marco Nolden, DKFZ Heidelberg
Prof. Dr. Christoph Palm, OTH Regensburg
Prof. Dr. Bernhard Preim, Universität Magdeburg
Prof. Dr. Martin Reuter, Universität Bonn
Prof. Dr. Karl Rohr, Universität Heidelberg
Prof. Dr. Sylvia Saalfeld, Universität Magdeburg
Prof. Dr. Dennis Säring, FH Wedel
Prof. Dr. Stefanie Speidel, NCT Dresden
Prof. Dr. Thomas Tolxdorff, Charité-Universitätsmedizin Berlin
Prof. Dr. Klaus Tönnies, Universität Magdeburg
Dr. Gudrun Wagenknecht, Forschungszentrum Jülich
Dr. Rene Werner, Universitätsklinikum Hamburg-Eppendorf
Dr. Stefan Wesarg, Fraunhofer IGD Darmstadt
Priv.-Doz. Dr. Thomas Wittenberg, Fraunhofer IIS, Erlangen
Prof. Dr. Ivo Wolf, Hochschule Mannheim
Priv.-Doz. Dr. Stefan Wörz, Universität Heidelberg

Sponsoren des Workshops BVM 2019

Die BVM wäre ohne die finanzielle Unterstützung der Industrie in ihrer erfolgreichen Konzeption nicht durchführbar. Deshalb freuen wir uns sehr über langjährige kontinuierliche Unterstützung mancher Firmen sowie auch über das neue Engagement anderer:

Gold-Sponsoren

Euroimmun AG
Seekamp 31, 23560 Lübeck
www.euroimmun.de

ID GmbH & Co KGaA
Platz vor dem Neuen Tor 2, 10115 Berlin
www.id-berlin.de

VisiConsult X-ray Systems & Solutions GmbH
Brandenbrooker Weg 2-4, 23617 Stockelsdorf
visiconsult.de

Silber-Sponsoren

Agfa HealthCare GmbH
Konrad-Zuse-Platz 1-3, 53185 Bonn
global.agfahealthcare.com

Haption GmbH
Dennewartstraße 25, 52068 Aachen
www.haption.de

Ibeo Automotive Systems GmbH
Merkurring 60-62, 22143 Hamburg
www.ibeo-es.com

Sponsoren

Medneo GmbH
Hausvogteiplatz 12, 10117 Berlin
www.medneo.com

MiE medical imaging electronics GmbH
Hauptstraße 112, 23845 Seth
mie-scintron.com

Beste wissenschaftliche Arbeiten

1. Maximilian Blendowski (Universität zu Lübeck)
   Blendowski M, Heinrich MP: 3D-CNNs for Deep Binary Descriptor Learning in Medical Volume Data

2. Hristina Uzunova (Universität zu Lübeck)
   Uzunova H, Handels H, Ehrhardt J: Unsupervised Pathology Detection in Medical Images using Learning-based Methods

3. Maike Stöve (Friedrich-Alexander-Universität Erlangen)
   Stoeve M, Aubreville M, Oetter N, Knipfer C, Neumann H, Stelzle F, Maier A: Motion Artifact Detection in Confocal Laser Endomicroscopy Images

Beste Präsentationen:

Weilin Fu (Friedrich-Alexander-Universität Erlangen)
Fu W, Breininger K, Schaffert R, Ravikumar N, Würfl T, Fujimoto J, Moult E, Maier A: Frangi-Net

Bestes Poster

André Klein (DKFZ Heidelberg)
Klein A, Warszawski J, Hillengaß J, Maier-Hein KH: Towards Whole-body CT Bone Segmentation

Vorwort

In diesem Jahr wird die Tagung Bildverarbeitung für die Medizin (BVM 2019) vom Institut für Medizinische Informatik an der Universität zu Lübeck ausgerichtet. Nach der erfolgreichen Durchführung der BVM 2001, 2011 und 2015 findet diese zentrale Tagung zu neuen Entwicklungen in der Medizinischen Bildverarbeitung in Deutschland nun zum vierten Mal in der traditionsreichen Hansestadt Lübeck statt.

Die medizinische Bildverarbeitung ist eine Schlüsseltechnologie in verschiedenen medizinischen Bereichen wie der Diagnoseunterstützung, der OP-Planung sowie der bildgeführten Chirurgie und Strahlentherapie. Methodisch haben hierbei in den letzten Jahren insbesondere Deep Neural Networks deutliche Fortschritte in Bezug auf Genauigkeit und Geschwindigkeit der Bildverarbeitungsverfahren ermöglicht, wobei das Potenzial maschineller Lernverfahren und Methoden der künstlichen Intelligenz im Bereich der Medizinischen Bildverarbeitung bei weitem noch nicht ausgeschöpft ist.

An der Universität zu Lübeck bilden die Medizinische Bildgebung und Bildverarbeitung einen zentralen universitären Forschungsschwerpunkt, der in den letzten Jahren systematisch ausgebaut wurde. Zudem bildet die Medizinische Bildverarbeitung in den Bachelor- und Masterstudiengängen Medizinische Informatik, Medizinische Ingenieurwissenschaften und Mathematik in Medizin und Lebenswissenschaften eine wichtige Vertiefungsrichtung. Vor diesem Hintergrund ist es eine besondere Freude, die BVM 2019 in Lübeck ausrichten zu dürfen.

Die BVM hat sich als ein zentrales interdisziplinäres Forum für die Präsentation und Diskussion von Methoden, Systemen und Anwendungen im Bereich der Medizinischen Bildverarbeitung etabliert. Ziel der Tagung ist die Darstellung aktueller Forschungsergebnisse und die Vertiefung der Gespräche zwischen Wissenschaftlern, Industrie und Anwendern. Die BVM richtet sich ausdrücklich auch an Nachwuchswissenschaftler, die über ihre Bachelor-, Master-, Promotions- und Habilitationsprojekte berichten wollen.

Die BVM 2019 wird unter der Federführung von Prof. Dr. rer. nat. habil. Heinz Handels, Direktor des Instituts für Medizinische Informatik der Universität zu Lübeck, ausgerichtet. Die Organisation ist wie in den letzten Jahren auf Fachkollegen aus Berlin, Braunschweig, Erlangen, Heidelberg, Lübeck und Regensburg verteilt, so dass die Organisatoren der vergangenen Jahre ihre Erfahrungen hier mit einfließen lassen können.

Anhand anonymisierter Bewertungen durch jeweils drei Fachgutachter wurden aus 87 eingereichten Beiträgen 28 Vorträge, 45 Poster und 2 Softwaredemonstrationen zur Präsentation ausgewählt. Die Qualität der eingereichten Arbeiten war insgesamt sehr hoch. Die besten Arbeiten werden auch in diesem Jahr mit BVM-Preisen ausgezeichnet. Die schriftlichen Langfassungen der Beiträge sind im Tagungsband zusammengefasst, der auch dieses Jahr wieder im Springer Verlag in der Reihe Informatik aktuell zur BVM erscheint. Das Programm wird durch eingeladene Gastvorträge zu aktuellen Themen des Deep

Learnings in der Medizinischen Bildverarbeitung sowie zur Beleuchtung und Diskussion der Sicht des Radiologen auf die aktuellen Entwicklungen abgerundet.

Die Internetseiten des Workshops bieten ausführliche Informationen über das Programm und organisatorische Details rund um die BVM 2019. Sie sind abrufbar unter der Adresse:

http://www.bvm-workshop.org

Am Tag vor dem wissenschaftlichen Programm werden drei Tutorials angeboten, bei denen in diesem Jahr verschiedene Aspekte der Deep Learnings in der Medizinischen Bildverarbeitung beleuchtet werden: Prof. Dr.-Ing. habil. Andreas Maier von der Friedrich-Alexander-Universität Erlangen-Nürnberg hält gemeinsam mit seinen Mitarbeiterinnen und Mitarbeitern ein Tutorial zum Thema „Deep Learning: Fundamentals“ ab. Hier wird eine Einführung in die grundsätzlichen Methoden des Deep Learnings und Ihre Anwendung auf medizinische Bilder gegeben. Fortgeschrittene Methoden des Deep Learnings in der Medizinischen Bildverarbeitung stehen im zweiten Tutorial mit dem Titel „Advanced Deep Learning Methods“ im Vordergrund, das von PD Dr. Klaus Maier-Hein und seinem Team vom DKFZ Heidelberg durchgeführt wird. Ergänzt wird dieses Angebot durch das dritte Tutorial „Hands-on Deep Learning in Pytorch“, das von Prof. Dr. Mattias Heinrich von der Universität zu Lübeck und seinem Team durchgeführt wird. Hier erhalten die Teilnehmenden Anleitungen zum praktischen Einsatz von neuesten Deep Learning Netzwerken und zur Handhabung der hierzu benötigten Softwarewerkzeuge.

Die Herausgeber dieser Proceedings möchten allen herzlich danken, die zum Gelingen der BVM 2019 beigetragen haben. Den Autoren für die rechtzeitige und formgerechte Einreichung ihrer qualitativ hochwertigen Arbeiten, dem Programmkomitee für die gründliche Begutachtung, den Gastrednern und den Referenten der Tutorials für Ihre aktive Mitgestaltung und inhaltliche Bereicherung der BVM 2019. Unser besonderer Dank gilt dem lokalen Organisationsteam in Lübeck, bestehend aus Dr. Jan Ehrhardt, Prof. Dr. Heinz Handels, Prof. Dr. Mattias Heinrich, Susanne Petersen und Dr. Jan Wrage, sowie den übrigen Mitarbeiterinnen und Mitarbeitern des Instituts für Medizinische Informatik in Lübeck, die durch ihren engagierten Einsatz die Organisation und Durchführung der BVM 2019 in der vorliegenden Form erst möglich gemacht haben. Weiterhin möchten wir den Helferinnen und Helfern an den Instituten in Berlin, Braunschweig, Erlangen, Heidelberg und Regensburg für Ihre Unterstützung bei der Organisation der BVM 2019 in Lübeck danken. Für die finanzielle Unterstützung bedanken wir uns bei den Fachgesellschaften und der Industrie.

Wir wünschen allen Teilnehmerinnen und Teilnehmern der BVM 2019 lehrreiche Tutorials, viele anregende Vorträge, Gespräche an den Postern und in der

Industrieausstellung sowie interessante neue Kontakte zu Kolleginnen und Kollegen aus dem Bereich der Medizinischen Bildverarbeitung.

Januar 2019

Heinz Handels (Lübeck)
Thomas Deserno (Braunschweig)
Andreas Maier (Erlangen)
Klaus Maier-Hein (Heidelberg)
Christoph Palm (Regensburg)
Thomas Tolxdorff (Berlin)

Inhaltsverzeichnis

Die fortlaufende Nummer am linken Seitenrand entspricht den Beitragsnummern, wie sie im endgültigen Programm des Workshops zu finden sind. Dabei steht V für Vortrag, P für Poster und S für Softwaredemonstration.

Session 1: Segmentation and Prediction

Session 2: Deep Learning: Learning Strategies and Adversarial Models

Postersession 1:

Segmentation (Poster)

Denoising and Imange Enhancement (Poster)

Registration and Motion Correction (Poster)

Software-Demonstrationen

Session 3: Image Recontruction and Intra-operative Navigation

Postersession 2:

Classification and Detection (Poster)

Visualization and Virtual Reality (Poster)

Imaging and Intra-operative Tracking (Poster)

Session 4: Virtual Reality and 3D Modeling

Session 5: Registration and Motion Models

Session 6: Visible Light

Abstract: Anchor-Constrained Plausibility

A Novel Concept for Assessing Tractography and Reducing False-Positives

Peter F. Neher[1], Bram Stieltjes[2], Klaus H. Maier-Hein[1,3]

[1]Division of Medical Image Computing, German Cancer Research Center (DKFZ), Heidelberg, Germany
[2]University Hospital Basel, Radiology & Nuclear Medicine Clinic, Switzerland
[3]Section for Automated Image Analysis, Heidelberg University Hospital, Germany
p.neher@dkfz.de

The problem of false positives in fiber tractography is one of the grand challenges in the research area of diffusion-weighted magnetic resonance imaging (dMRI). Facing fundamental ambiguities especially in bottleneck situations, tractography generates huge numbers of theoretically possible candidate tracts. Only a fraction of these candidates is likely to correspond to the true fiber configuration, posing a difficult sensitivity-specificity trade-off. Current methods address this issue either by focusing exclusively on well-known fiber bundles using prior knowledge or by using tract filtering techniques based on the image signal. Currently, the link between these two choices of purely data driven and prior knowledge based approaches is missing.

We propose a novel concept that rigorously exploits prior knowledge about the existence of anatomically known tracts (anchor tracts) to reduce the degrees of freedom of a successive data-driven filtering of the remaining candidate tracts: anchor-constrained plausibility (ACP). This approach is based on the hypothesis that information about the presence or absence of each anchor influences the plausibility of the candidates and thereby reduces the ambiguities in the problem.

We demonstrate the potential of this concept in a series of phantom experiments: ACP significantly improved the tractography sensitivity-specificity trade-off in such controlled settings (AUC 0.91). The direct assessment of false-positive reduction rates requires a ground truth, which does not exist *in vivo*. *In vivo*, we therefore concentrated on assessing the capabilities of ACP in a structured and objective tractogram analysis of 110 subjects of the Human Connectome Project (HCP) young adult study, providing detailed data-driven insights into what we might be missing when focusing only on anatomically known tracts. This work has previously been published at MICCAI 2018 [1].

References

1. Neher PF, et al. Anchor-constrained plausibility (ACP): a novel concept for assessing tractography and reducing false-positives. Proc MICCAI. 2018;11072:20–27.

Automatic Detection of Blood Vessels in Optical Coherence Tomography Scans

Julia Hofmann[1], Melanie Böge[1], Szymon Gladysz[1], Boris Jutzi[2]

[1]Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB), Ettlingen
[2]Institute of Photogrammetry and Remote Sensing (IPF), KIT Karlsruhe
julia.hofmann@iosb.fraunhofer.de

Abstract. The aim of this research is to develop a new automated blood vessel (BV) detection algorithm for optical coherence tomography (OCT) scans and corresponding fundus images. The algorithm provides a robust method to detect BV shadows (BVSs) using Radon transformation and other supporting image processing methods. The position of the BVSs is determined in OCT scans and the BV thickness is measured in the fundus images. Additionally, the correlation between BVS thickness and retinal nerve fiber layer (RNFL) thickness is determined. This correlation is of great interest since glaucoma, for example, can be identified by a loss of RNFL thickness.

1 Introduction

Since optical coherence tomography (OCT) offers a noninvasive method for an ophthalmology diagnosis in the fundus area of the eye, this imaging method is of increasing importance. Glaucoma, for example, can be identified by a loss of retinal nerve fiber layer (RNFL) thickness, visible in OCT scans.

The aim of this research is to develop an automated blood vessel (BV) detection algorithm for OCT scans and corresponding fundus images. Recent researches showed reliable results using shadowgraphs to find lateral position and diameter of BVs in 2D OCT-scans. By adding Doppler information, 3D orientation was also obtained [1]. Supervised pixel classification [2] enabled lateral BV detection in OCT 3D volumes. Except manual parameter setting, unsupervised segmentation [3] offered a fully automated segmentation algorithm. Model-based approaches [4, 5] then extended the detection to axial BVs. Efficient automated detection was demonstrated using a deep learning algorithm [6] trained on a specific training data set.

In this research a new automated BV detection algorithm is described without the need of an additional model, training or supplementary data. BVs are visible as vertical shadows in OCT images. This is caused by light absorption through a BV during the OCT procedure. Since the Radon transformation is based on line integrals, vertical lines can be easily reinforced with this method. Together with supporting image processing methods, the approach offers a robust, automated BV shadow (BVS) detection in OCT images. For medical interest a

correlation between BV thickness and RNFL thickness is determined to define a new glaucoma metric.

2 Materials and methods

The data consists of OCT images and corresponding fundus images for each eye. They are recorded as circular and linear scans. Besides the recording mode (circular/linear), the recording density of the single scans (B-scan) can be varied. In this study circular scans and sparse linear OCT images (with 37 B-scans) from 16 participants and dense linear OCT images (with 193 B-scans in the recording area and a density of 30 μm) from 12 participants are included.

2.1 Methodology

The different data types are processed with the same operational sequence (Fig. 1) which was developed for 2D data. The processing steps are described in this section. Afterwards, a description of the process in case of 3D data is given.

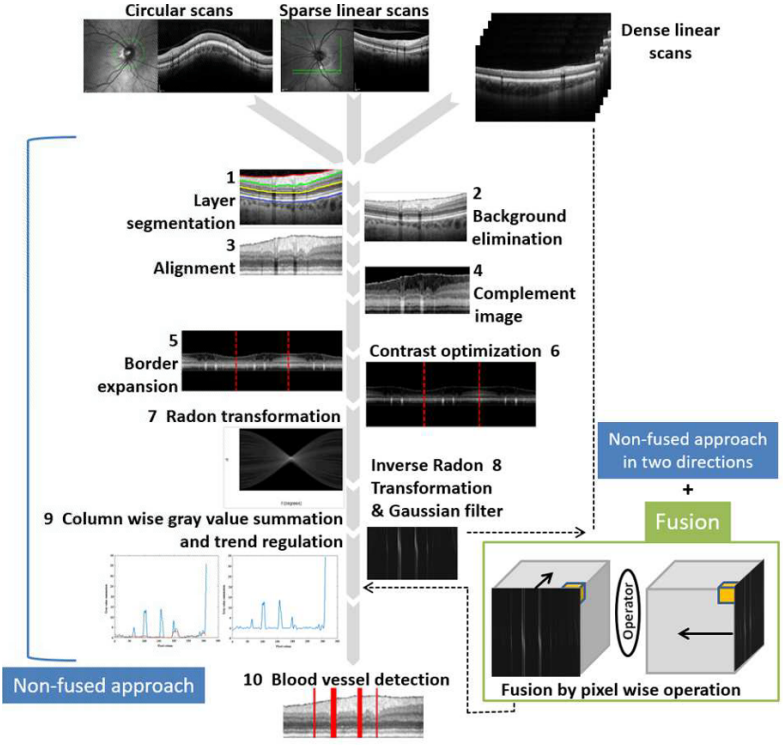


Fig. 1. Flowchart of the proposed approach.

Preparatory steps for the radon transformation The first process step is the layer segmentation (1). For background elimination (2) all values above the top layer, Inner Limiting Membrane (ILM), are set to the maximum gray value to ensure maximum contrast for the BVSs. For alignment (3) each column (A-scan) in the OCT image is shifted up or down according to the Retinal Pigment Epithelium layer (RPE) as baseline. This step simplifies cropping of the image to the area of interest, which suppresses the influence of noise from surrounding areas. Layer segmentation and alignment are applied with functions created by Mayer et al. [7]. At this point in the process BVSs have low gray values which are hard to reinforce. For this reason, a negative image (4) is produced to enable an amplification of the bright BVSs. Border expansion (5) and contrast optimization (6) are part of the boundary problem step. They will be explained after the Radon transformation.

Radon transformation Since the BVSs in OCT scans are vertical lines, a line detection algorithm is beneficial for this application. The Radon transformation (7) is a linear integral transformation defined by Johann Radon [8]. For the transformation to Radon space, each pixel will be represented in polar coordinates (ρ, θ). The rotation and translation starts from the origin of the image coordinate system. For each angle θ and each distance ρ the intensities of the image are summed up. The result is $r(\rho, \theta)$, consisting of the column gray value summations in all orientations. We assume that the BVSs are the only vertical lines in the OCT images. They can be seen in Radon space in the zero degree column. For easier and more robust BVS detection, the BVSs are reinforced. To enhance the contrast of BVSs in the original image, all values in the first column in Radon space are squared. The following columns are attenuated by multiplying all values with a sloping function. BVSs might not be exactly vertical and therefore to maintain slightly slant shadow areas, small angles are less attenuated than line integrals of higher angles. We found that the sloping exponential function was most suitable for reinforcement. It falls very steeply which causes stronger attenuation in the columns of higher angles. After the relevant columns in Radon space are amplified, the image is transformed back to the image format with the inverse Radon transformation.

Boundary problem The inverse Radon transformation causes artifacts at the border areas of the image which leads to false BVS detections. The origin of this can be explained using the concept of spatial frequency Fourier transformation. Frequencies are expected to be infinite and the border areas of the finite image can therefore not be reconstructed with the inverse Radon transformation. To avoid this, a border expansion (5) by flipping the whole image on both sides is applied. Also, a threshold is used to optimize contrast (6) and get rid of disturbing insignificant values. Darker values on the right boundary arise during the OCT recording and cause falsely detected BVSs. To remove this trend (9) a 15 x 15 pixel window is shifted along the graph. The mean values of the window values are subtracted from the original summation values.

Blood vessel shadow detection After these image processing steps, the BVSs can be detected in the trend-regulated gray value summation graph. For the BVS detection a quantile threshold is calculated from the graph.

Thickness measurement The BVSs visible in the OCT data do not necessarily correspond to the real BV thickness. Often the scan corresponds to a diagonal cut through the BV. The position of a BVS is therefore transferred to the fundus image. Here, a diameter measurement is performed by generating a BV-filling circle around the detected position.

Pseudo-3D processing - fusion after bidirectional processing The acquisition density of the B-scans corresponding to the dense linear data is too sparse to allow a true 3D processing. For that reason, a pseudo 3D processing (fusion) is applied to approximate a volumetric processing and gain an anisotropic detection. For a volumetric approximation all dense linear scans consecutively are approximated as a cube (Fig. 2). The dense linear scans are concatenated and A-sheets are generated. The A-sheets consist of the same A-scan column in each B-scan. In the x-, y-, z-coordinate system (Fig. 2), they consist of one y-z plane for each x. The pixel size varies between B-scans (4×4 μm) and A-sheets (30×4 μm) according to the acquisition density of 30 μm. After the generation of the A-sheets, the whole non-fused approach is performed separately for all B-scans and all A-sheets. This way the processing is performed for two directions giving two resultant cubes. The cubes are fused with pixel-wise averaging of the pixel values. After the fusion, the BVS detection is applied similarly to the non-fused approach (Fig. 1).

Evaluation A comparison of BVS detections to ground truth is presented. Manual detections by an expert ophthalmologist are used as ground truth in all evaluations. The consistency of the BVS thickness measurements between ground truth and the developed approach was validated using the root mean square error (RSME). This metric defines the deviation between the expected

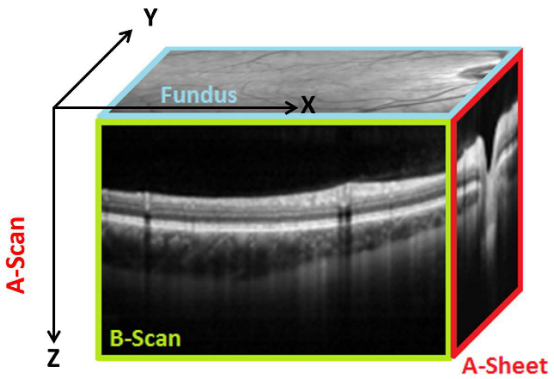


Fig. 2. Coordinate system of the OCT images.

Table 1. Table of results and comparison values.

	Radon approach (all data)	Radon approach (dense scans)		OCTSeg (all data)	OCTSeg (dense scans)
Detection rate	90%	90%	89%	90%	85%
RMSE (pixel)	5.16	5.32	5.44	3.25	8.24

value and the measurement. It provides information about the accuracy of the measurements.

Glaucoma causes RNFL thinning. It might also effect BV thickness. For this reason the correlation of the layer thickness and BV thickness is of interest for ophthalmology. It is tested if the correlation generated from patients suffering from glaucoma shows significant differences in comparison to the correlation obtained from healthy eyes. Only ground truth layer segmentation and BVS detections of sparse OCT scans are used for the correlation estimation to avoid influence of false detections. Correlation can only be determined for detected BVSs without transfer to the fundus image (Sec. Thickness measurement), since for glaucomatous eyes, only OCT scans and not fundus images are available in the data set.

3 Results

The experiments involved sparse linear scans and circular scans from 16 participants and 241 dense linear scans.

The detection rates and thickness accuracies are given in Tab. 1. All detections were compared to the approach of Mayer [7] (OCTSeg). The proposed algorithm is able to achieve the same detection rate as OCTSeg. On the dense linear scans the Radon approach even outperforms the fusion approach and OCTSeg. Regarding all data, OCTSeg achieves the best thickness measurement accuracy. On the dense linear scans the Radon approach performs best. As shown in Tab. 3, the RNFL thickness for the healthy eyes is immensely higher than for glaucomatous eyes. As expected, the RNFL thickness is a significant criteria for glaucoma identification. The measured BV thickness maxima on the other hand only differ 3 μm. The BV thickness or the correlation between BV thickness and RNFL thickness is therefore not beneficial as glaucoma metric.

	Maximum RNFL thickness (μm)	Maximum BV thickness (μm)
Glaucomatous eyes	72	38
Healthy eyes	252	41

Table 2. Correlation of RNFL thickness and BV thickness.

4 Discussion

The advantage of the proposed BV detection algorithm is that no model, training or supplementary data is needed. The results described in this paper show that a sufficient BVS detection rate is enabled with the non-fused, and also the fusion approach. Since the resolution varies according to the recording direction, the fusion detection is not as dense as the radon approach and therefore achieves a lower detection rate. With the transfer to the fundus image (Sec. Thickness measurement) a reliable BV thickness measurement is demonstrated. A metric for glaucomatous eyes could be found, even if the RNFL thickness is more significant for glaucoma than the BV thickness. Here, an additional data acquisition of OCT scans with corresponding fundus images on glaucomatous eyes should be performed to enable a meaningful correlation of layer thickness and BV thickness. Layer segmentation in the approach influences the detection. In the future, a greater independence from the layer segmentation approach (here taken of OCTSeg) will be sought. In further research, the approach could be compared with OCT Angiography to additionally prove the effectiveness.

References

1. Wehbe H, Ruggeri M, Jiao S, et al. Automatic retinal blood vessel parameters calculation in spectral domain optical coherence tomography. Procs SPIE. 2007;6429:6429–1–7.
2. Niemeijer M, Garvin MK, van B Ginneken, et al. Vessel segmentation in 3D spectral OCT scans of the retina. Procs SPIE. 2008;6914:6914–1–8.
3. Annunziata R, Garzelli A, Ballerini L, et al. Leveraging multiscale hessian-based enhancement with a novel exudate inpainting technique for retinal vessel segmentation. IEEE J Biomed Health Inform. 2016;20(4):1129–38.
4. Lee KM. Segmentations of the Intraretinal Surfaces, Optic Disc and Retinal Blood Vessels in 3D-OCT Scans. Iowa: Univ Iowa; 2009.
5. Pilch M, Wenner Y, Strohmayr E, et al. Automated segmentation of retinal blood vessels in spectral domain optical coherence tomography scans. Biomed Opt Expr. 2012;3(7):1478–91.
6. Li Q, Fengi B, Xie L, et al. A cross-modality learning approach for vessel segmentation in retinal images. IEEE Trans Med Imaging. 2016;35(1):109–18.
7. Mayer MA, Hornegger J, Mardin C, et al. Retinal nerve fiber layer segmentation on FD-OCT scans of normal subjects and glaucoma patients. Biomed Opt Expr. 2010;1(5):1358–83.
8. Radon J. Über die Bestimmung von Funktionen durch Ihre Integralwerte Längs Geweisser Mannigfaltigkeiten. Berichte Sächsische Acad Wissenschaft Math Phys, Klass. 1917;69:262.

Prediction of Liver Function Based on DCE-CT

Oliver Rippel[1], Daniel Truhn[2], Johannes Thüring[2], Christoph Haarburger[1], Christiane K. Kuhl[2], Dorit Merhof[1]

[1]Institute of Imaging & Computer Vision, RWTH Aachen University, Germany
[2]Department of Diagnostic and Interventional Radiology, University Hospital Aachen, Germany
oliver.rippel@lfb.rwth-aachen.de

Abstract. Liver function analysis is crucial for staging and treating chronic liver diseases (CLD). Despite CLD being one of the most prevalent diseases of our time, research regarding liver in the Medical Image Computing community is often focused on diagnosing and treating CLD's long term effects such as the occurance of malignancies, e.g. hepatocellular carcinoma. The Child-Pugh (CP) score is a surrogate for liver function used to quantify liver cirrhosis, a common CLD, and consists of 3 disease progression stages *A*, *B* and *C*. While a correlation between CP and liver specific contrast agent uptake for dynamic conrast enhanced (DCE)-MRI has been found, no such correlation has been shown for DCE-CT scans, which are more commonly used in clinical practice. Using a transfer learning approach, we train a CNN for prediction of CP based on DCE-CT images of the liver alone. Agreement between the achieved CNN based scoring and ground truth CP scores is statistically significant, and a rank correlation of 0.43, similar to what is reported for DCE-MRI, was found. Subsequently, a statistically significant CP classifier with an overall accuracy of 0.57 was formed by employing clinically used cutoff values.

1 Introduction

Assessing liver function is crucial for staging and treating chronic liver diseases (CLD) [1]. Due to its various functions, there exist a multitude of tests to assess liver state [2], some of them based on imaging. A very common clinical scoring system of liver function is the Child-Pugh score (CP) [3]. It scores several important indicators such as e.g. ascites and subsequently aggregates them. CP classes are then gained by applying the following thresholds: CP *A* 5-6 points, CP *B* 7-9 points, and CP *C* 10-15 points. The CP score is clinically used to assess the prognosis of liver cirrhosis, a CLD responsible for more than 1 million deaths annually [4], and monitor its transition to the end stage.

In the Medical Image Computing community, research on CP so far has been focused on Dynamic Contrast Enhanced-MRI (DCE-MRI). Here, Motosugi et al. [5] have shown an association between CP score and accumulation of liver specific contrast agent, as confirmed by further literature [6, 7]. Moreover, a successful prediction of liver fibrosis was performed by Yasaka et al. [8], again

based on the accumulation of liver specific contrast agent in DCE-MRI. To the best of our knowledge, successful evaluation of liver function based on DCE-CT has not been reported yet, irrespective of the widespread use of DCE-CT in clinical practice as a routine examination.

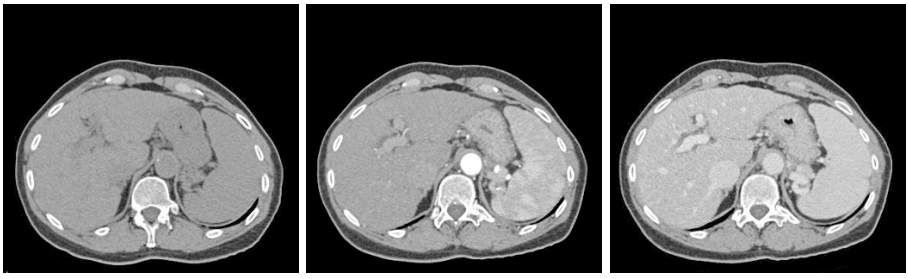
The main contribution of our work is an approach for predicting CP scores based on DCE-CT imaging alone, using a combination of state-of-the-art convolutional neural networks (CNN) with transfer learning.

2 Methods

2.1 Dataset

In total, the dataset comprises 259 subjects (76 CP score A, 120 CP B, 63 CP C). For each subject, a radiologist with more than 3 years of experience in abdominal imaging reviewed automatic liver delineation in the venous phase generated by Philips Intellispace.

CT imaging was performed by using helical CT scanners (Somatom Definition Flash and Somatom Definiton AS, Siemens Medical Sysems, Forchheim, Germany). The scans were acquired in a craniocaudal direction by using a detector configuration of 128 or 40 x 0.6 mm, a tube current of 120 kVp, quality reference of 240 mAs, and online dose modulation in all phases (pitch 1.0), during a single breath-hold helical acquisition of roughly 10 seconds (slightly varying due to the differing liver sizes). For all imaging, the gantry rotation speed was 2 Hz. The contrast-enhanced images were created with a weight-adjusted application of iodinated contrast material (1.5 mL per kilogram of body weight; Iopramide 370 mg/mL, Ultravist, Shering, Germany) administered at a rate of 3 mL/s by power injector. Subsequently the non-enhanced (native) as well as arterial and venous phases were acquired. The acquisition of the arterial phase started 6 seconds after the automatic detection of peak aortic enhancement at the level of the coeliac trunk with a threshhold of 140 HU; portal venous phase was scanned 55 seconds after the start of the contrast injection. Image reconstruction was performed with axial 1-mm images, an increment of 0.7 mm, and a B30f convolutional kernel for all phases (Fig. 1, representative axial slices).



(a) non-contrast enhanced (b) arterial enhancement (c) venous enhancement

Fig. 1. Representative images of a contrast enhanced liver CT-scan. Images were reconstructed with a B30f kernel in soft-tissue-window.

2.2 Pre-processing

First, native and arterial phases were registered to the venous phases with a rigid registration algorithm under the assumption that liver shape would be constant in all three phases. Registration itself was performed using SimpleElastix [9] with default parameters. Subsequently, voxel intensities were linearly mapped to a soft tissue window (center 40 HU, width 400 HU).

Next, axial patches of 224x224 in-plane dimension were extracted around the centerpoint of the liver along 20 % to 80 % of its craniocaudal extension. This approach has the advantage of incorporating the context around the liver in a patch, and may thus capture effects such as ascites. Furthermore, it reduces the need for resizing, as axial patch dimensions are concordant with the input shape expected by the pretrained model. In total, 12492 patches were extracted in this manner to be used for model finetuning.

2.3 Model architecture and training

For the model architecture, a ResNet18 [10] pretrained on ImageNet [11] was used in a transfer learning approach [12]. At its core, ResNet consists of residual blocks, where deviations from an identity mapping are learned by the model. This has been shown to sucessfully tackle the problem of vanishing gradients inherent to deep CNNs. While model depth of a ResNet architecture can be arbitrary, we use a depth of 18, minimizing the number of trainable parameters and therefore risk of overfitting.

The output of the pretrained ResNet18 model was adapted to our ordering problem, giving a single continous value for every slice. By stacking the three phases of the CT scan, the number of input channels satisfy the number of channels as required by the pretrained model. While modifications to the axial dimensions are not necessary, on-the-fly data augmentation was used to reduce overfitting of trained models. These consisted of rotation, scaling, as well as elastic deformation and were performed by the batchgenerators framework[1]. Model finetuning on the CT images itself was performed using an L2-regularized Adam optimizer with initial learning rate of 0.0005 and a decay rate of 0.5 every 10 epochs and L2-penalty of 0.001. All layers were trained simultaneously, employing the MeanSquareError (MSE) metric for training and the accuracy metric for validation. MSE was chosen over a classification loss function, such as e.g. CrossEntropy, to reflect the ordinal nature of the CP score. For this, class labels were assigned based on the clinically used thresholds: [5, 6] for CP *A*, [7, 9] for CP *B* and [10, 15] for CP *C*. As the output of the model is continuous, values are rounded to the nearest integer to yield CP class predictions.

The model was trained using 10 fold cross-validation. For each fold, patches are split into training, validation and test set such that patches from a single patient are exclusively included in either training, validation or test set. The splitting ratios were 0.63, 0.27 and 0.1 for training, validation and test sets,

[1] https://github.com/MIC-DKFZ/batchgenerators

respectively. Total number of training epochs were 20 epochs per fold, and validation was performed after every epoch. The model state with highest validation score was subsequently used to predict the test set. All experiments were implemented using PyTorch on a workstation equipped with Intel i7-7700K processor and Nvidia GTX 1080 Ti GPU.

2.4 Prediction generation and statistical evaluation

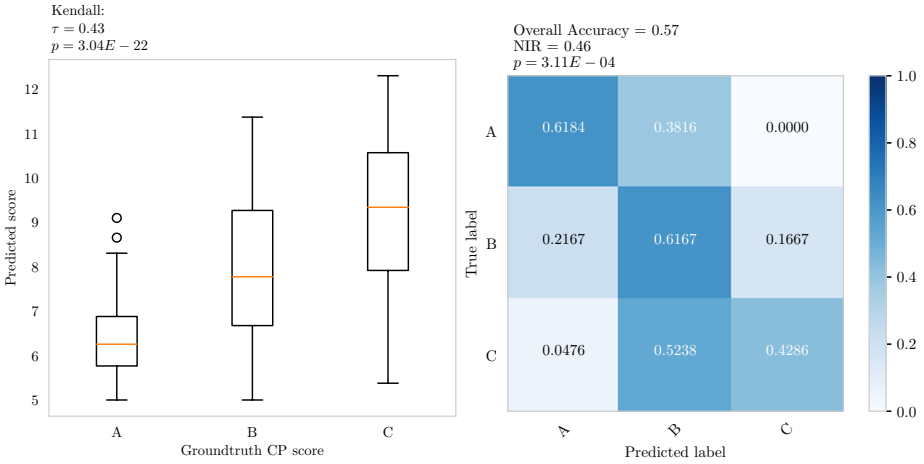
As the CP score is an ordinal score, Kendall's τ statistic was used to quantify rank correlation betweeen CNN-based CP scoring and groundtruth values [13]. Agreement was considered to be statistically significant when $p < 0.05$.

Although the CP score is ordinal in nature, its clinical use corresponds to a classification problem. To yield class predictions, the same clinical thresholds were used for the test set as during model validation. To quantify performance of the classifier, comparison against the No Information Rate (NIR) was performed with a one-sided binomial test. Here, again $p < 0.05$ was considered significnant.

3 Results

To obtain subject-level classification results from slice-level predictions, slice-level prediction was performed and subsequently averaged. Continuous CP prediction results over all 10 splits are given in Fig. 2a. The computation time for training a single fold was 15 min. Inference for a single subject can be performed in 1.5 s.

In total, a statistically significant agreement could be seen between model and groundtruth values for CP ranking ($p = 3.23 \cdot 10^{-16}$). Correlation denoted



(a) Continuous CP score generated by the CNN.

(b) CP classification gained by thresholding continous score.

Fig. 2. Cross validated prediction of the CP score with a CNN.

by τ was 0.43 (Fig. 2a). Using the cutoff values, an overall classification accuracy of 0.57 was achieved (Fig. 2b). When compared to NIR (0.46), this classification performance is statistically significant ($p = 3.11 \cdot 10^{-4}$).

4 Discussion

We developed and presented a transfer learning based approach to predict CP based on DCE-CT.

When assessing model performance in predicting CP values based on DCE-CT images, a statistically significant agreement was found between CNN predictions and groundtruth scores with a rank correlation of 0.43 (Fig. 2a). Moreover, overall achieved classification accuracy was statistically significant compared to the NIR (Fig. 2b). Motosugi et al. [5] performed a similar analysis in DCE-MRI, but found a statistically significant correlation with insufficient predictive value between CP and contrast agent enrichment in liver parenchym. While predictive models between DCE-MRI and CP have not been established in literature, correlations are well reported [6, 7]. Apart from the connection between CP score and liver CTs/MRIs, CNNs have been used to perform liver fibrosis staging, again based on DCE-MRI [8]. In this study, a Spearman rank correlation of 0.63 was reported between predicted and true fibrosis scores. The authors, however, explicitly state that their model cannot be used to perform liver function analysis, as performed by CP scoring assessed in our work.

Therefore, achieved values are within reason, but further research is required to facilitate clinical use. Nonetheless, to the best of our knowledge, we are the first to report such a correlation in DCE-CT images, with CT being the more frequently used routine diagnostic tool that is more readily available compared to DCE-MRI.

To further increase performance, the next steps would be to sample contrast agent kinetics in finer detail, similar to other fields such as tumor classification [14]. Additionally, an increase in study population size may further improve generalizability of generated models. This would also enable the use of 3D-CNNs, which require larger datasets than 2D-CNNs. Also, use of multimodal models comprising both clinical as well as image features should be assessed in future.

5 Conclusion

In this work, we investigated whether a predictive relationship between DCE-CT image features and CP score can be established. To adress the limitation of small datasets that is often encoutered when dealing with medical images, our CNN was pretrained on natural images and only fine tuned on our dataset in a transfer learning approach. Experiments revealed statistically significant correlation of rankings generated by the model and groundtruth CP scores, which were subsequently used to form a statistically significant classifier of CP score. While the classifier is overall significant, further research is needed to improve discrimination between individual CP classes.

References

1. Suk KT, Kim MY, Baik SK. Alcoholic liver disease: treatment. World J Gastroenterol. 2014;20(36):12934–12944.
2. Kortgen A, Recknagel P, Bauer M. How to assess liver function? Curr Opin Crit Care. 2010;16(2):136–141.
3. Pugh R, Murray-Lyon I, Dawson J, et al. Transection of the oesophagus for bleeding oesophageal varices. Br J Surg. 1973;60(8):646–649.
4. Rowe IA. Lessons from epidemiology: the burden of liver disease. Dig Dis. 2017;35(4):304–309.
5. Motosugi U, Ichikawa T, Sou H, et al. Liver parenchymal enhancement of hepatocyte-phase images in Gd-EOB-DTPA-enhanced MR imaging: which biological markers of the liver function affect the enhancement? J Magn Reson Imaging. 2009;30(5):1042–1046.
6. Verloh N, Haimerl M, Rennert J, et al. Impact of liver cirrhosis on liver enhancement at Gd-EOB-DTPA enhanced MRI at 3 tesla. Eur J Radiol. 2013;82(10):1710–1715.
7. Tamada T, Ito K, Higaki A, et al. Gd-EOB-DTPA-enhanced MR imaging: evaluation of hepatic enhancement effects in normal and cirrhotic livers. Eur J Radiol. 2011;80(3):e311–e316.
8. Yasaka K, Akai H, Kunimatsu A, et al. Liver fibrosis: deep convolutional neural network for staging by using gadoxetic acid-enhanced hepatobiliary phase MR images. Radiol. 2017;287(1):146–155.
9. Marstal K, Berendsen F, Staring M, et al. SimpleElastix: a user-friendly, multilingual library for medical image registration. Proc CVPR. 2016;.
10. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. Proc CVPR. 2016; p. 770–778.
11. Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database. Proc CVPR. 2009; p. 248–255.
12. Tajbakhsh N, Shin JY, Gurudu SR, et al. Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans Med Imaging. 2016;35(5):1299–1312.
13. Kendall MG. The treatment of ties in ranking problems. Biometrika. 1945;33(3):239–251.
14. Haarburger C, Langenberg P, Truhn D, et al. Transfer learning for breast cancer malignancy classification based on dynamic contrast-enhanced MR images. Proc BVM. 2018; p. 216–221.

Abstract: Adversarial Examples as Benchmark for Medical Imaging Neural Networks

Magdalini Paschali[1], Sailesh Conjeti[2], Fernando Navarro[1], Nassir Navab[1,3]

[1]Computer Aided Medical Procedures, Technische Universität München, Germany
[2] Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), Bonn, Germany
[3]Computer Aided Medical Procedures, Johns Hopkins University, USA
magda.paschali@tum.de

Deep learning has been widely adopted as the solution of choice for a plethora of medical imaging applications, due to its state-of-the-art performance and fast deployment. Traditionally, the performance of a deep learning model is evaluated on a test dataset, originating from the same distribution as the training set. This evaluation method provides insight regarding the generalization ability of a model. However, in medical imaging scenarios, especially in cases when a deep learning framework is utilized by a physician for a real-world application, the samples forwarded into the model might belong to a distribution different from the one of the training dataset, or might suffer from noise which cannot usually be modelled by a known distribution, thus raising the need for an evaluation scheme that investigates the robustness of a model, i.e. its performance on data originating from a manifold different from the training one.

To this end, we recently proposed [1] to utilize adversarial examples [2], images that look imperceptibly different from the originals but are consistently missclassified by deep neural networks, as surrogates for extreme test case scenarios, like the ones mentioned above. Extensive evaluation was performed on state-of-the-art classification and segmentation deep neural networks, for the challenging tasks of fine-grained skin lesion classification and whole brain segmentation, leveraging a variety of methods to generate adversarial examples. The results showcased the significant difference in the performance of the utilized networks on clean and on adversarial images. Specifically, networks that performed equally well regarding their generalizability had an astounding 20% difference in robustness, highlighting the need for the proposed, more thorough evaluation technique to uncover which neural network was able to grasp a deeper understanding of the training data and when deployed in real-world applications can showcase a higher robustness to out-of-distribution test samples.

References

1. Paschali M, Conjeti S, Navarro F, et al. Generalizability vs. robustness: investigating medical imaging networks using adversarial examples. Proc MICCAI. 2018; p. 493–501.
2. Szegedy C, Zaremba W, Sutskever I, et al. Intriguing properties of neural networks. Int Conf Learn Representations. 2014;Available from: http://arxiv.org/abs/1312.6199.

Evaluation of Image Processing Methods for Clinical Applications

Mimicking Clinical Data Using Conditional GANs

Hristina Uzunova, Sandra Schultz, Heinz Handels, Jan Ehrhardt

Institut für Medizinische Informatik, Universität zu Lübeck
uzunova@imi.uni-luebeck.de

Abstract. While developing medical image applications, their accuracy is usually evaluated on a validation dataset, that generally differs from the real clinical data. Since clinical data does not contain ground truth annotations, it is impossible to approximate the real accuracy of the method. In this work, a cGAN-based method to generate realistically looking clinical data preserving the topology and thus ground truth of the validation set is presented. On the example of image registration of brain MRIs, we emphasize the necessity for the method and show that it enables evaluation of the accuracy on a clinical dataset. Furthermore, the topology preserving and realistic appearance of the generated images are evaluated and considered to be sufficient.

1 Introduction

The validation of medical image processing algorithms relies on the availability of datasets with a dedicated ground truth, but translation into clinical practice is often significantly hindered by the fact that available validation datasets differ from real clinical data in the acquisition parameters and presence of pathologies, artifacts or noise. For example, a validation using images of healthy subjects possibly underestimates image registration errors for clinical images containing pathologies. On the other hand, the generation of ground truth data is tedious and costly and therefore not feasible for all kinds of clinical data. However, reliable error estimates are crucial for the application of automated image processing in many medical applications, e.g. radiotherapy or surgical planning.

Therefore, we propose to automatically generate realistically looking clinical data with ground truth annotations, based on the validation dataset at hand. With the recent development of image generation methods, especially GANs [1], their application for various medical image processing tasks gets considered more frequently, like image domain translation and denoising [2] or unsupervised detection of anomalies [3]. Works like [4, 5] show that GANs can also be used to translate between image domains. In this work we lay our focus on paired style transfer based on the pix2pix network [5], used to generate realistically looking clinical data constrained on the topology of the validation data and thus preserving ground truth segmentations, achieving the possibility to evaluate the error of algorithms applied on clinical images

2 Materials and methods

GANs are generative models that learn to map a random noise vector $\mathbf{z}$ to an output image y using a generator function $G : \mathbf{z} \rightarrow y$ [1]. An extension of regular GANs are the conditional GANs, that learn the mapping from an observed image x additionally, $G : \{x, \mathbf{z}\} \rightarrow y$ To ensure that the generator produces realistically looking images that cannot be distinguished from real ones, an adversarial discriminator, D, is enclosed in the training process, aiming to perfectly distinguish between real images and generator's fakes.

2.1 Style transfer using conditional generative adversarial networks

One possible application of cGANs is style transfer. In this case the generator takes an image x as input and trains to generate its corresponding style-transferred image $G(x) \approx y$ The discriminator takes a pair of images as input: x and $G(x)$ (fake) or x and y (real) and is trained to determine whether the pair is real of fake. Thus the objective of the fully conditional GAN can be expressed as

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,\mathbf{z}}[\log(1 - D(x, G(x, \mathbf{z})))] \tag{1}$$

where G tries to minimize this objective and an adversarial D tries to maximize it. Also to encourage the generator to produce realistic images more directly, it is beneficial to use an L1 loss: $\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,\mathbf{z}}[||y - G(x, \mathbf{z})||_1]$ The final objective is then

$$G^* = \arg\min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G) \tag{2}$$

One popular style transfer representative is pix2pix [5] and it requires strictly paired data for training. In [5] the authors show that they are able to transfer contours of an object to the photographic image of the object itself. Also if trained on a certain domain A the cGAN will generate images with the style of A even if the contours belong to a different image domain B, however the topology of the contours of B remains. Those properties are interesting for our work, since we seek contour-to-gray-value topology-preserving translation of medical images.

2.2 Medical image style transfer

Assume there is a validation dataset of healthy patients images with ground truth segmentations or landmarks of the anatomically significant parts, V, and a clinical dataset of possibly pathological images containing only segmentations of the pathologies, C. Since the focus of our work lies on applying image registration, pathologies would lead to strongly decreased registration accuracy and ground truth anatomical annotations are crucial for its evaluation. However, the presented method is not restricted to this application.

Aiming to generate realistically looking data from the clinical domain, but preserving the topology and thus the segmentation masks of the healthy validation data, the cGAN described in 2.1 is used as follows. G is trained to generate

clinical images $c_i \in C$ conditioned on the edges extracted from the images e.g. by using a Canny filter, such that $G(x_i, \mathbf{z}) \approx c_i \in C$ where $x_i = \text{edges}(c_i)$ Then in test phase only the edges of the validation images $\text{edges}(v_j)$ are inputed and $G(\text{edges}(v_j)) \in C$ outputs an image looking like clinical data but preserving the topology of the validation image v_j and thus the ground truth segmentations apply. Still, $\text{edges}(v_j)$ does not contain any pathologies and the generator would most likely generate a healthy image. However, one can explicitly simulate pathologies on the images, since we assume that their segmentations are available in C.

2.3 Network architecture

Contrary to pix2pix [5], here ResNet blocks [6] are used for the generator, as in our experience, they show better reconstruction abilities. The generator downscales the input image first by using three 2D strided convolutions (conv2D) each followed by batch normalization (BN) and a ReLU, resulting in 256 channels. Then nine ResNet blocks with 256 channels each are applied, and the image is upsampled to the original size using two transposed conv2Ds each followed by a BN and a leaky ReLU (lReLU), and at last a conv2D layer followed by the Tanh function.

The discriminator takes as input a two-channel image composed of a gray value image and its corresponding contour image. The input is then send through the architecture: conv2D $\rightarrow$ lReLU $\rightarrow$ (conv2D $\rightarrow$ BN $\rightarrow$ lReLU)$\times 3$ $\rightarrow$ lReLU $\rightarrow$ conv2D, that first iteratively downscales the image and produces 512 channels and the last convolution reduces the overall size to one neuron (real/fake).

3 Results

3.1 Data and setup

Our experiments simulate an atlas-to-patient registration scenario for brain MRI images. The LPBA40 data [7] containing 40 healthy whole-head T1 MRIs with 56 labeled anatomic regions is chosen as validation dataset. The clinical data is represented by a subset of the BRATS 2015 challenge data [8], which contains 220 brain MRI T1c images with high-grade glioblastomas. The BRATS data differs from LPBA40 by different gray value ranges as well as the presence of skull stripping artifacts and pathologies (Fig. 1(left)). For both datasets 2D slices on the same positions are extracted and the BRATS images are cropped around the center to a size of 181×217 matching the size and approximate alignment of the LPBA40 images.

The network described in Sec. 2.3 is trained to generate images in the style of BRATS from their corresponding edge images. The edge images are generated using a Canny filter. Our experience showed that better results are achieved when using gradient magnitude weighted edges rather than binary ones. Furthermore, we want to explicitly integrate the tumor structures in the training process, to

prevent the network from hallucinating pathologies [9]. Therefore, segmentation masks of the pathologies available to the BRATS data are combined with the edge images. For the test phase, the extracted LPBA40 edge images are combined with 4 different tumor masks picked at random from the BRATS dataset (affine pre-registration undertaken to ensure that the masks are placed inside the brain) resulting in 5 generated images pro input image (4 with tumors and 1 without). The generated images then have the appearance of BRATS with predefined pathology availability but follow the contours of the LPBA40 images.

To validate the atlas-to-patient registration, one subject of the LPBA40 data is selected as atlas and registered to the remaining 39 subjects using the variational registration method presented in [10]. Label overlap measures (Dice) are used as surrogate for registration accuracy. The cGAN-based style transfer now allows for the replicated validation using the generated data.

3.2 Experiments and results

Visual evaluation In first place it is important to determine whether the images generated by the cGAN are realistic for the particular domain at all. Fig. 1 shows a few generated images and in our experience they generally have a realistic appearance.

Topology preservation An important point of the method presented here, is the assumption that the topology of the LPBA40 image is being preserved, allowing for the segmentation labels to be transferred directly (Fig. 1 (right)). To evaluate this quantitatively, the contours of the input LPBA40 images and the contours of the corresponding generated BRATS images are extracted (here

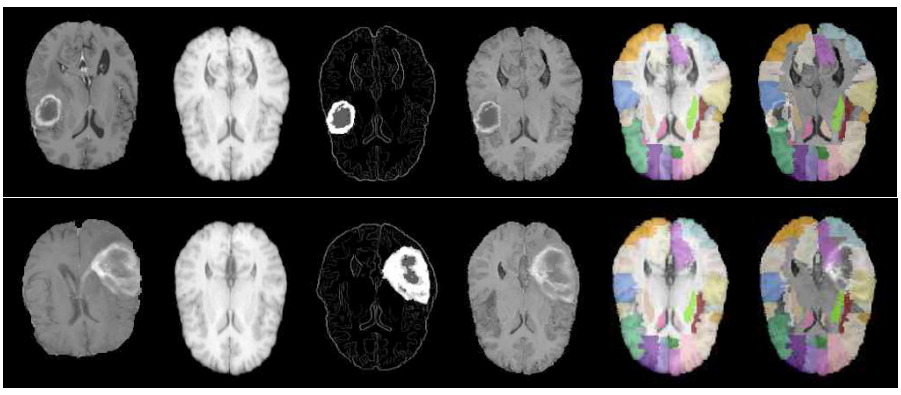


Fig. 1. Examples of two generated images. From left to right: Real BRATS image containing the tumor; LPBA image; Contours of LPBA image and tumor mask serving as input in test phase; Generated image; LPBA labels overlayed over the real LPBA; LPBA labels overlayed over the generated image (best viewed in color).