

Hochschule Reutlingen Reutlingen University



Uwe Kloos, Natividad Martínez, Gabriela Tullius (Hrsg.)

Informatics Inside connect(IT);

Informatik-Konferenz an der Hochschule Reutlingen 9. Mai 2018



Impressum

Anschrift: Hochschule Reutlingen / Reutlingen University Fakultät Informatik Human-Centered Computing Alteburgstraße 150 D-72762 Reutlingen

Telefon: +49 7121 / 271-4002 Telefax: +49 7121 / 271-4042

E-Mail: infoinside@reutlingen-university.de Internet: http://infoinside.reutlingen-university.de

Organisationskomitee: Prof. Dr. Gabriela Tullius, Hochschule Reutlingen Prof. Dr. Natividad Martínez, Hochschule Reutlingen Prof. Dr. Uwe Kloos, Hochschule Reutlingen

Benjamin Batt Claudiu Bräuer Sinem Cicek Celik **Emanuel Geiger** Frauke Grieble Peter Grupp Pia Laubacher Öznur Öner Katharina Pavic Ngoc Linh Phan Claudia Ryniak Josia Scheytt Christian Steinmann Clemens Weißenberg Vanessa Willenbrock Steffen Wittig

> Copyright: © Hochschule Reutlingen, Reutlingen 2018 Herstellung und Verlag: Hochschule Reutlingen ISBN: 9 -83000-586453



Hochschule Reutlingen Reutlingen University

Inhaltsverzeichnis

Longpaper
Josia Scheytt Segmentierung von Polypen in Koloskopie-Bilddaten - ein Potentialanalyse von Deep-Lear- ning-Methoden1
Benjamin Weinert Untersuchung der Möglichkeiten und Risiken von Implantate11
Peter Grupp Untersuchung der Anforderungen an ein System zur Unterstützung der Reproduzierbarkeit von Ultraschalluntersu
Öznur Öner Digitalisierung im klinischen Umfeld zur Förderung der personalisierten Medizin am Uni- versitätsklinikum Tübingen am Fallbeispiel der molekularen Diagnostik mithilfe der MTB- Plattform
Sinem Cicek Celik Kulturwandel von ITIL zu DevOps im Unternehmen
Christian Steinmann IT-Sicherheit in Unternehmen - State of the Art, Gefahren und Trends
Steffen Wittig Social Crowd Simulation zur Belebung virtueller Welten
Janis Uttenweiler Identifizierung einer geeigneten Prototypingmethode für die multimodale Navigation mit dem E-Bike
Maic Schellig Konzeption zur Detektion von sich öffnenden Fahrzeugtüren
David Leisten Konzept einer Motion-Capture basierten Simulationsumgebung zur Untersuchung von Inter- aktionen zwischen Passanten und autonomen Fahrzeugen

Segmentierung von Polypen in Koloskopie-Bilddaten – eine Potentialanalyse von Deep-Learning-Methoden*

Josia Scheytt Reutlingen University Josia.Scheytt@Student.Reutlingen-University.DE

Abstract

Kolorektale Karzinome haben eine hohe Sterblichkeitsrate, wenn sie spät entdeckt werden. Eine frühzeitige Entfernung von bösartigen Polypen im Magen-Darm-Trakt, die deren Vorstufen bilden, bietet jedoch hohe Überlebenschancen. Bei Darmspiegelungen werden gerade kleine Polypen aber recht häufig übersehen. Zuverlässige bildverarbeitende Systeme, die Polypen in einem Koloskopie-Frame nicht nur detektieren, sondern pixelgenau segmentieren, könnten Ärzten bei Darmkrebs-Screenings helfen.

Diese Arbeit analysiert den aktuellen Stand der Segmentierung von Polypen im Gastrointestinaltrakt. Weiterführend wird untersucht, inwiefern die in letzter Zeit sehr erfolgreichen Methoden des Deep Learning hier Vorteile bieten.

Schlüsselwörter

Segmentierung, Bildverarbeitung, Kolorektalkarzinom, Deep Learning

CR-Kategorien

I.2.6 [Artificial Intelligence]: Learning; I.2.1 [Artificial Intelligence]: Applications and Expert Systems; I.4.6 [Image Processing and Computer Vision]: Segmentation

Betreuer: Prof. Dr.-Ing. Burgert Hochschule Reutlingen Burgert@Reutlingen-University.DE
Informatics Inside 2018
Wissenschaftliche Vertiefungskonferenz
09. Mai 2018, Hochschule Reutlingen
Copyright 2018 Josia Scheytt

1 Einführung

Unter den 20 häufigsten Todesursachen weltweit kommen auch 4 Krebsarten vor [1]. Von allen Krebsarten kommen Darmkarzinome am dritthäufigsten vor, und sie sind die Krebsart mit der vierthöchsten Sterblichkeitsrate [2]. Kolorektale Karzinome, also Krebs im Blinddarm oder Dickdarm, machen davon den allergrößten Anteil aus [3]. Wenn bösartige kolorektale Polypen frühzeitig entfernt werden, lässt sich die Mortalität verringern [4]. Ein Entfernen sämtlicher Polypen inklusive der gutartigen hingegen birgt Risiken wie Blutungen und Perforation des Darmes [5].

Zur genaueren Untersuchung des Dickdarms und der letzten Zentimeter des Dünndarms wird oft eine Darmspiegelung (Koloskopie) durchgeführt [6]. Dieses nichtinvasive Verfahren ermöglicht eine detaillierte visuelle Inspektion des Dickdarms über ein rektal eingeführtes Koloskop. Findet der Arzt Polypen (wie z. B. in Abbildung 1 oben), kann von dem verdächtigen Gewebe direkt eine Probe entnommen werden. Eine solche Biopsie kann dann histologisch auf Gut- oder Bösartigkeit untersucht werden.

2 Problemstellung und Motivation

Viele kleinere Polypen, die sich später zu Karzinomen entwickeln können, sind aufgrund ihrer Größe bei Koloskopien nicht leicht zu erkennen. Mithilfe von Algorithmen, die automatisiert Polypen in Videomaterial von Koloskopien auffinden, ließe sich





Abbildung 1: Polyp im Dickdarm (oben), Segmentierung des Polypen (unten)

die Arbeit des Arztes erleichtern und der Anteil an gefundenen Polypen erhöhen. Solche Systeme könnten potenziell auch weiterentwickelt werden, sodass sie eine Abschätzung bezüglich der Gut- oder Bösartigkeit des Gewebes durchführen.

Ebenso wäre auch der Einsatz bei einer Spiegelung des Dünndarms möglich. Karzinome im Dünndarm kommen zwar deutlich seltener vor als im Dickdarm [3], aber auch dort stellen Polypen ein Krebsrisiko dar. Bei Dünndarm-Spiegelungen kommen häufig Endoskopkapseln zum Einsatz (Wireless Capsule Endoscopy (WCE)), wobei eine große Menge an Videomaterial anfällt, das nachträglich vom Arzt untersucht werden muss und durch Automatisierung schneller bewertbar wäre.

In dieser Arbeit werden bisherige Forschungsarbeiten vorgestellt und verglichen, die sich mit der Segmentierung von Polypen in Koloskopie-Bilddaten beschäftigen. Dazu wird in Abschnitt 3 auf Methoden eingegangen, bei denen die Features für die Segmentierung manuell gewählt wurden. Darüber hinausgehend wird in Abschnitt 4 untersucht, inwiefern Methoden aus dem Bereich Deep Learning in diesem Anwendungsfeld einen Vorteil bieten.

Arbeiten, die nur die *Präsenz* von Polypen detektieren, werden hier nicht behandelt. Im Fokus stehen Ansätze, die pixelgenaue Segmentierungen (s. Abbildung 1) produzieren, weil nur dadurch eine präzise Lokalisierung von Polypen im Bild möglich ist.

3 Manuelle Feature-Wahl

In den folgenden Abschnitten wird aufgezeigt, welche Techniken bisher schon verwendet wurden, um Polypen zu segmentieren. Einige dieser Ansätze arbeiten auch auf Bildmaterial von Untersuchungen mit WCE. Da es sich hier um eine Problemstellung aus der Bildverarbeitung handelt, lässt sich das Problem auch in der dazugehörigen Terminologie beschreiben (s. Abbildung 2): Es ist ein Eingabe gegeben (das koloskopische Bild, wie in Abbildung 1 oben), aus diesem werden Features (markante Merkmale) extrahiert. Auf diese Weise wird der hochdimensionale Input auf die problemrelevanten Dimensionen reduziert. Anhand dieser Features wird ein Mapping angewendet, das aus den Features eine Ausgabe produziert, nämlich die Segmentierung.



Abbildung 2: Prozess der Bildverarbeitung

Bei maschinellem Lernen werden Mappings grundsätzlich datenbasiert festgelegt, also gelernt. Die im Folgenden vorgestellten Ansätze unterscheiden sich allerdings darin, ob die Wahl der Features manuell, also auf Basis von menschlichem Expertenwissen, festgelegt oder gelernt wird.

Prasath liefert in [7] ein umfassendes Review von Ansätzen im Bereich der Detektion und



Segmentierung von Polypen in Bildmaterial von WCE und teilweise auch Koloskopien. Neben über 20 verschiedensten Ansätzen zur Detektion der Präsenz von Polypen in einem Bild, auf die hier nicht eingegangen wird, sind auch einige Algorithmen aufgeführt, die entweder eine grobe Lokalisierung oder eine pixelgenaue Segmentierung erzielen.

3.1 Lokalisierung

Bei den Segmentierungs-Ansätzen bei [7], die nur eine grobe Lokalisierung vornehmen, kann diese zwar auch als pixelgenaue Maske erfolgen, aber oftmals ist die Ausgabe nur eine Ellipse oder "Bounding Box" um den/die Polypen-Kandidaten, manchmal sogar nur ein einzelner Punkt im Bild. Diese Lokalisierung basiert in der Regel auf geometrischen Annahmen, nämlich dass Polypen oft kreis- oder ellipsenförmig aussehen. Daraus leiten sich verschiedene handgemachte Features ab, die beispielsweise ein Maß für die Krümmung oder Wölbung von Strukturen im Bild berechnen und daraus ableiten, wie wahrscheinlich ein bestimmter Teil des Bildes einen Polypen enthält. Teilweise kommen zwar zusätzliche Features wie Textur und Farbe ins Spiel, aber Polypen unterscheiden sich in diesen beiden Punkten meist nur wenig von der sie umgebenden Darmschleimhaut.

Abgesehen von den in [7] untersuchten Ansätzen sind folgende Ansätze im Bereich der Lokalisierung als markante Beispiele hervorzuheben:

In [8] wird ausgehend vom Modell des Aussehens eines Polypen, das die Autoren in [9] beschrieben haben, hauptsächlich mit der Tiefe von Tälern im Verlauf der Intensitätswerte ("Depth of Valley") entlang verschiedener Richtungen gearbeitet. Zusätzlich bauen sie Mechanismen ein, die einige Kriterien zum Auffinden von Polypen forcieren; darunter Geschlossenheit, Robustheit, Kontinuität und Konkavität. Von diesen Kriterien ausgehend werden "Energy Maps" aufgebaut, deren Maximalpunkt im ursprünglichen Bild hervorgehoben wird, um die höchstwahrscheinliche Position eines Polypen markie-

ren. Damit diese Methode nicht nur wie ein naiver Konturen-Detektor funktioniert, ist eine ausführliche Vorverarbeitung nötig; ansonsten wären unter den falsch positiven Ergebnissen auch Blutgefäße und Glanzlichter. Bei [10] kommt ein Kantendetektor zum Einsatz, von dessen Ergebnis alle Nicht-Polypen-Kandidaten gefiltert werden; anschließend werden alle Kandidaten bewertet hinsichtlich der Wahrscheinlichkeit, dass sie Polypen sind. Das Filtern geschieht anhand von Features, die auf Basis der diskreten Kosinus-Transformation aus Patches des Ausgangsbildes generiert werden. Die Bewertung wiederum stützt sich besonders auf eine "Energy Map" auf Basis der Normalenrichtung des Gradienten jedes Patches. Sowohl in der Filterung als auch bei der Bewertung kommen Lernverfahren zum Einsatz, die ein überwachtes Lernen des Mappings anhand von realen Segmentierungen ermöglichen.

3.2 Segmentierung

Eine möglichst akkurate Segmentierung wird in den bei [7] vorgestellten Ansätzen fast ausschließlich durch den Einsatz einer Konturenverfolgung erzielt. Diese verlassen sich größtenteils auf die bereits erwähnten geometrischen Features. Außer den in [7] vorgestellten Ansätzen sind folgende Ansätze gute Beispiele für eine Polypen-Segmentierung basierend auf manuellen Features:

Bei [9] wird wie bei [8] mithilfe der "Depth of Valleys" gearbeitet: Nach einigen Vorverarbeitungsschritten, die besonders der Entfernung von Glanzlichtern dienen, wird eine erste Segmentierung mithilfe des Watershed-Algorithmus erstellt. Mithilfe des Depth-of-Valley-Bildes kann dann entschieden werden, welche segmentierte Region der endgültige Polypen-Kandidat ist.

Die Autoren von [11] führen Segmentierungen auf Koloskopie-Bilddaten aus. Auch ihr Algorithmus entfernt Glanzlichter in der Vorverarbeitung. Die Segmentierung wird mit dem Watershed-Algorithmus auf einer hierarchischen Konturen-Map durchgeführt.

In die segmentierten Regionen wird dann eine Ellipse eingepasst; die Region mit der größten Überlappung ist die endgültige Ausgabe.

4 Deep Learning

In diesem Abschnitt werden Ansätze aus dem Bereich der Polypen-Segmentierung vorgestellt, bei denen Features nicht von Experten entwickelt, sondern maschinell gelernt werden. Werden Features gelernt, statt dass sie durch die Wahl bestimmter Algorithmen hart kodiert werden, können diese Merkmale deutlich robuster sein. Zum anderen können aber auch ganz neue Features entdeckt werden, die kein Experte bisher betrachtet hatte. Außerdem ist die Übertragung von solchen Systemen auf neue Anwendungsgebiete, sogenanntes "Transfer Learning" oder "Domain Transfer", deutlich leichter machbar und spart dadurch Entwicklungsressourcen.

Genauer betrachtet werden in diesem Abschnitt nur solche Ansätze untersucht, die Deep Learning verwenden. Deep Learning sind künstliche Neuronale Netze (NN) mit mehr als nur einer verborgenen Schicht [12]. Sie sind bei weitem nicht die einzige Möglichkeit, Features durch maschinelles Lernen selektieren zu lassen, aber im Gegensatz zu z. B. k-Means-Clustering oder Hauptkomponentenanalyse sind sie deutlich genereller einsetzbar und dadurch auch erfolgreicher. Der Siegeszug von Deep Learning hat verschiedene Ursachen; Goodfellow et al. [12] identifizieren darunter unter anderem folgende Faktoren:

- die generelle, aber mächtige A-Priori-Annahme, dass abstraktere Konzepte hierarchisch aus immer simpleren Konzepten gelernt werden,
- die Inspiration durch neuronale Strukturen im tierischen und menschlichen Gehirn und
- die Verfügbarkeit von leistungsfähigerer Hardware für ein effizientes Trainieren von tiefen NN.

miteinander verbundenen Neuronen gebildet, inspiriert vom Aufbau natürlicher Gehirne. Jedes künstliche Neuron transformiert seine eingehenden Werte anhand einer Aktivierungsfunktion und gibt diese an die Neuronen der nächsten Schichten weiter. Das Training solcher Netze ist in der Regel mit relativ simplen Gradientenabstiegs-Methoden möglich.

Die Besonderheit bei Deep Learning ist die Tiefe der Netze: Es wird eine große Anzahl an Neuronen-Schichten hintereinander geschaltet. Diese Schichten können auch vergleichsweise "schmal" sein mit wenigen Neuronen in einer Schicht. Die Tiefe der Netze ermöglicht es ihnen, zuerst simple Konzepte zu lernen, in der nächsten Schicht abstraktere Konzepte aus diesen simplen Konzepten zu lernen und diesen Abstraktionsprozess hierarchisch immer weiter fortzuführen.

4.1 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) sind eine Variante von tiefen NN, bei denen in mindestens einer Schicht die Faltungs-Operation angewendet wird [12]. Die Kernelmatrix, mit der gefaltet wird, ist in jedem Bildfenster einer Neuronenschicht dieselbe und ihre Werte werden vom Netz gelernt. Im Gegensatz zu vollständig verbundenen Schichten (Fully Connected (FC) Layers) nutzen mehr Neuronen dieselben Parameter als bei reiner Matrixmultiplikation, was die Speicherkomplexität um mehrere Größenordnungen reduzieren kann. Bei CNNs kommen außerdem Pooling-Schichten hinzu, die dafür sorgen, dass die gelernte Repräsentation der Daten invariant gegenüber kleinen Verschiebungen des Inputs wird. Dadurch drücken CNNs eine A-Priori-Annahme aus, die gerade bei Input-Daten wie Bildern gilt und deswegen dort Vorteile bietet, nämlich dass kleine Verschiebungen innerhalb einer Bildstruktur die Semantik des Bildes nicht verändern.

Genauso wie bei Deep Learning im Allgemeinen gilt auch bei CNNs, dass eine große



18 In

Tiefe, also eine hohe Anzahl an Schichten, beim Abstrahieren und robusten Lernen hilft. Hierbei lernt ein CNN z. B. in den ersten Schichten simple Features wie Kanten und Ecken, in tieferen Schichten dann Konturen und Texturen bis hin zu Objektteilen.

Das Trainieren eines CNN erfordert in der Regel viele Trainingsdaten, damit das Netz gut generalisieren kann. Große CNNs, die erfolgreich auf großen und diversen Bild-Datensätzen mit über 1000 Klassen trainiert wurden (z. B. VGGNet [13]), werden gerne als Grundlage für neue CNNs genommen. Solches Transfer Learning wird durchgeführt, indem Teile der Architektur des vortrainierten Netzes mitsamt dessen gelernten Parametern übernommen und nur die dahinter liegenden Schichten angepasst werden.

Dieses Transfer Learning wird besonders häufig in Domänen angewandt, in denen wenig Trainingsdaten zur Verfügung stehen. Es wurde gezeigt, dass der Transfer allgemein vortrainierter CNNs auf die spezifische medizinische Domäne mindestens so gute Ergebnisse erzielt wie ein CNN, das mit zufällig initialisierten Werten trainiert wird [14]. In allen nachfolgend vorgestellten Ansätzen wird ein solches Transfer Learning angewendet.

4.2 Lokalisierung

Bei [15] wird mithilfe von Farb-Features und CNNs in Kombination mit einer Support Vector Machine (SVM) ein Fenster innerhalb des Bildes ermittelt, in dem sich ein Polyp befindet. Die Autoren stellen fest, dass Features, die von CNNs gelernt werden, gegenüber anderen Methoden robuster sind hinsichtlich 3D-Rotation und Rauschen. Als zusätzliches Merkmal nutzen sie zusätzlich noch Farb-Wavelets, die hier nicht näher erläutert werden. Da die Wavelet-Transformation mehrere Skalierungen berücksichtigt, kann mit ihren Features auch eine Skaleninvarianz erzielt werden [16].

Das zu verarbeitende Bild wird in Fenster mit fester Größe aufgeteilt, die nacheinander verarbeitet werden. Wird in einem Fenster ein Polyp gefunden, wird nicht mehr gesucht, ob noch weitere vorhanden sind; es kann also maximal ein Polyp gefunden werden. Es wird auch nicht evaluiert, wie der Einsatz von Farb-Wavelets zusammen mit CNNs gegenüber einem reinen Einsatz von CNNs abschneidet. Dadurch wird nicht ersichtlich, welchen Vorteil eine solche Symbiose bietet, da CNNs sowohl mit Farbe als auch unterschiedlichen Auflösungen umgehen können.

Lequan et al. [17] nutzen zur Polypen-Lokalisierung ein 3D-CNN, das sie zu einem Fully Convolutional Network (FCN) umbauen. 3D-CNNs ermöglichen die Verarbeitung dreidimensionaler Bilddaten wie z. B. von CT und MRT in einem CNN. In diesem Fall ist der Input allerdings eine Bildsequenz, damit nicht nur räumliche, sondern auch zeitliche Zusammenhänge gelernt werden können. Diese 3D-CNN-Architektur wird erweitert auf ein 3D-FCN.

FCNs sind eine spezielle Form von CNNs, bei denen die Faltungs-Operation nicht nur in mindestens einer Schicht zum Einsatz kommt, sondern in allen Schichten [18]. Das macht nicht nur das Training weniger komplex, sondern hebt auch die Beschränkung auf Inputs fester Größe auf: Bilddaten mit beliebiger Auflösung können jetzt verarbeitet werden. Außerdem ist jetzt keine Fensterung des Inputs mit einem "Sliding Window" mehr nötig, denn das gesamte Originalbild wird auf einmal verarbeitet. Die resultierende semantische Segmentierung wird zwar durch ein Upsampling hoch skaliert, welches das Netz selbst lernt, hat aber in der Regel eine geringere Skalierung als der Original-Input.

Das 3D-FCN von Lequan et al. gibt eine Wahrscheinlichkeits-Map aus, deren Maxima auf Polypen-Positionen im Bild hindeuten. Anhand dieser Maxima kann dann mit einer Markierung, die über das ursprüngliche Bild gelegt wird, die Position von Polypen im Bild hervorgehoben werden.

4.3 Segmentierung

FCNs kommen bereits zur Segmentierung von Polypen zum Einsatz bei Vázquez et

al. [19]. Sie präsentieren in diesem Ansatz ein neues öffentliches Dataset, damit Algorithmen zur Polypen-Segmentierung sich daran messen können. Gleichzeitig legen sie direkt eine Baseline fest, nämlich ein ohne zusätzliche Optimierung trainiertes FCN. Die Autoren zeigen, dass die semantische Multi-Klassen-Segmentierung einer koloskopischen Szene mit FCNs in einem einzigen Schritt möglich ist; es wird nicht mehr ein gesonderter Algorithmus pro Segmentierungs-Klasse benötigt. Die 2-Klassen-Segmentierung, die nur Polypen und Hintergrund unterscheidet, schneidet am besten ab.

Eine gängige Praxis bei CNNs ist der Einsatz von Data Augmentation: Hierbei werden Transformationen an den Input-Bildern durchgeführt, um einerseits eine geringe Menge an Trainingsdaten künstlich zu vervielfachen und andererseits die gelernte Repräsentation der Daten robuster zu machen. Zu diesen Transformationen können geometrische Transformationen wie Rotation und Translation zählen, aber auch künstliches Rausches und Moiré-Effekte. Die Autoren von [19] stellen fest, dass die Segmentierung der Polypen und auch die des Hintergrundes sich am meisten verbessert, wenn bei der Data Augmentation eine Kombination von Zoom, Verkrümmung, Scherung und Rotation zum Einsatz kommt.

Wichakam et al. [20] stellen eine Verbesserung dieser Baseline-Architektur vor, bei der das FCN komprimiert wird. Der Hauptvorteil gegenüber der vorherigen Architektur ist eine Beschleunigung der Inferenzzeit, also der Zeit, die bei der Live-Segmentierung eines neuen Samples benötigt wird. Diese wird hier reduziert, sodass Echtzeit-Performance möglich wird, indem die Autoren zwei der Faltungsschichten der vorherigen Architektur verwerfen.

Außerdem verändern sie die Verlustfunktion, sodass der DICE-Score nicht mehr bezogen auf die beiden Klassen "Hintergrund" und "Polyp" berechnet wird, sondern nur noch bezüglich der Polypen-Klasse. Dies rechtfertigen sie mit dem Klassen-Ungleichgewicht, das häufig in medizinischen Bilddaten auftritt: Die Klasse "Hintergrund" macht die Mehrheit der Pixel der Trainingsdaten aus. Ihre quantitativen Ergebnisse liegen weniger als 1 % unter dem der Baseline, gleichzeitig ist die Inferenzzeit im Vergleich auf unter 15 % gesunken und die Trainingszeit auf etwa ein Drittel.

4.4 Generative Netze

CNNs sind ein Beispiel für die effektive Lösung von überwachten Lernproblemen mit Deep Learning, aber auch bei den unüberwachten Lernverfahren gibt es einige interessante Entwicklungen. Eine dieser Entwicklungen ist von der Motivation getrieben, die vom Netz gelernte Repräsentation besser verstehen zu können, aber auch davon, mit einer geringeren Menge gelabelter Daten auszukommen. Sogenannte generative Netze sind in der Lage, eine Repräsentation der Daten zu lernen und dann selbstständig völlig neue Samples zu generieren, die nicht aus der Trainingsmenge stammen [12]. Die Mehrheit der Ansätze für generative Netze, auf die hier nicht genauer eingegangen wird, sind im Gegensatz zu konventionellen NN nur als ungerichtete Graphen darstellbar, und das sorgt dafür, dass das Training und teilweise sogar die Inferenz nur noch approximiert werden können. Das Training mit den dafür entwickelten Algorithmen verläuft oft zu instabil für einen praktischen Einsatz.

Die von Goodfellow et al. entwickelten Generative Adversarial Networks (GANs) hingegen können mit effizienten und für Deep Learning üblichen Methoden trainiert werden [21]. Die spieltheoretische Grundidee besteht darin, dass ein Generator-Netz und ein Diskriminator-Netz gegeneinander arbeiten. Der Diskriminator wird darauf trainiert, realistische Samples von gefälschten Samples zu unterscheiden. Gleichzeitig erzeugt der Generator so lange neue Samples, bis der Diskriminator nicht mehr unterscheiden kann, ob die generierten Samples echt oder unecht sind.

Die ursprünglich entwickelten GANs produzierten anfangs noch Bildsamples von re-



lativ niedriger Auflösung. Spätere Weiterentwicklungen bauten eine Art Laplace'sche Pyramide auf, bei der in jeder Upsampling-Stufe das generierte Sample künstlich mit Gauss-Filter hochskaliert wird und ein GAN pro Stufe eine Verfeinerung der Hochskalierung lernt [22]. Die Qualität dieser Netze wurde wiederum von den Deep Convolutional GANs (DCGANs) [23] noch übertroffen. Diese Art von GAN nutzt ausschließlich Faltungsschichten.

Das Generator-Netz in GANs generiert in seiner ursprünglichen Form ein neues Sample, indem es einen Vektor mit zufällig initialisierten Werten als Eingabe nimmt und diesen durch seine Schichten schickt. Wird hingegen ein GAN umfunktioniert, sodass der Generator statt einem Rausch-Vektor einen "sinnvollen" Input, z. B. ein Bild, als Eingabe bekommt, ergibt sich ein konditioniertes GAN. Solche Conditional Adversarial Networks (CAN) können dann für überwachte Lernprobleme genutzt werden, bei denen die Verlustfunktion nicht mehr von Hand erstellt wird, sondern von der Netzarchitektur selbst gelernt wird.

Das eröffnet ganz neue Möglichkeiten, beispielsweise für eine Image-to-Image-Translation, bei der die Domänenanpassung nur noch durch die Trainingsdaten geschieht und keine Anpassung der Verlustfunktion mehr notwendig ist [24]: Das CAN wird z. B. auf Paaren von Farbbildern und deren Graustufen-Variante trainiert und kann dann Graustufenbilder kolorieren; ebenso kann beim Training mit Paaren von Straßenszenen und deren Segmentierungen aus einer gegebenen Segmentierung eine zugehörige Straßenszene generiert werden und umgekehrt. Theoretisch wäre auch auf diese Weise eine Polypen-Segmentierung durchführbar, wenn genügend Daten zum Trainieren vorhanden sind. Allerdings merken die Autoren an, dass ihre CANs sich noch nicht besonders gut für semantische Segmentierung eignen, da sie zwar hochauflösende Ergebnisse produzieren, aber auch einige kleine Objekte "halluzinieren", die im Ausgangsbild nicht vorhanden sind.

Das hält Zisimopoulos et al. allerdings nicht davon ab, mithilfe der Image-to-Image-CANs OP-Instrumente in Simulationen von Katarakt-OPs zu segmentieren [25]. Im Vergleich zu FCNs in verschiedenen Auflösungen produziert ein auf hoher Auflösung trainiertes CAN sehr genaue Segmentierungen, die aber immer noch einige kleine falsch positive Ergebnisse enthalten. Goodfellow et al. [12] empfehlen aus ihrer Erfahrung im Erstellen tiefer NN, tendenziell ein Modell mit zu großer Kapazität zu wählen und dessen Überanpassung durch starke Regularisierung einzuschränken statt ein Modell mit geringer Kapazität zu wählen. Diese Einschätzung lässt es möglich erscheinen, dass durch eingehende Regularisierung des CAN ein Netz erzielt werden kann, das präzise Segmentierungen ohne Halluzinationen liefert.

Abgesehen von [25] nutzt der Recherche des Autors nach bisher noch niemand generative Netze oder speziell GANs auf endoskopischen Bilddaten außer Mahmood et al. [26]. Sie greifen folgenden Trend auf, der sich auch im medizinischen Deep Learning ausbreitet: Es werden synthetische Trainingsdaten und die zugehörigen Labels oder Segmentierungen automatisiert erstellt, indem Szenen in einem virtuellen OP-Simulator nachgestellt werden. Dann lernt ein Netz auf diesen Trainingsdaten und es wird angenommen, dass es dann auch auf realen Daten zuverlässig funktioniert. Dieses Vorgehen ist sehr attraktiv, weil maschinelles Lernen viele Trainingsdaten benötigt und diese gerade im medizinischen Sektor teuer zu erzeugen sind. Auch Zisimopoulos et al. [25] sind ein Beispiel für ein solches Vorgehen, allerdings bleiben die Ergebnisse auf realen Daten noch weit hinter den Erwartungen zurück.

Um diese Lücke zu schließen, gibt es bereits Ansätze, die versuchen, mithilfe von GANs synthetische Bilder an reale anzunähern, um dadurch die automatisiert erzeugten Trainingsdaten qualitativ an reale anzunähern. Mahmood et al. gehen den entgegengesetzten Weg: Sie nähern mithilfe eines "Transformer-GANs" *reale an synthetische Bilder* an, um dann mit Netzen, die auf syn-

thetischen Daten trainiert wurden, eine Tiefenkarte aus dem monokularen Koloskopiebild abzuschätzen. Eine solche Tiefenschätzung könnte dann wiederum verwendet werden, um eine Segmentierung von Polypen zu lernen, da die Form der Polypen das alles entscheidende Feature zu sein scheint.

5 Fazit

In dieser Arbeit wurde untersucht, inwiefern Deep Learning im Bereich der Polypen-Segmentierung Vorteile bringen kann. Eine Analyse des aktuellen Standes der Wissenschaft zeigt, dass bei den von Hand gewählten Features besonders die Form der Polypen aussagekräftiger ist als Farbe und Textur. Das Lernen von Features, für das Deep Learning bisher die erfolgreichste Methode ist, benötigt zwar eine große Menge Trainingsdaten, kann aber potenziell bessere Ergebnisse liefern und außerdem neuartige Features entdecken.

FCNs als vollständig faltende Erweiterung von CNNs ermöglichen eine semantische Segmentierung. Vázquez et al. [19] stellen mit ihrer Anwendung des FCN-8s [18] eine Baseline in der Polypen-Segmentierung auf, deren Performanz in komprimierter Form von Wichakam et al. [20] verbessert wird.

Sie prädizieren allerdings nur auf einzelnen Frames und betrachten den zeitlichen Kontext nicht. Das 3D-FCN von Lequan et al. [17] behandelt eine Frame-Sequenz als 3D-Volumen, erzeugt davon allerdings bisher nur eine punktgenaue Lokalisierung und keine pixelgenaue Segmentierung.

Isola et al. [24] stellen eine domänenunabhängige Image-to-Image-Translation vor, die auf konditionierten DCGANs [23] basiert und auch Segmentierungen lernen kann. Anwendungen im Kontext der Segmentierung von OP-Instrumenten [25] belegen genauere Ergebnisse als FCNs, allerdings auch eine höhere Anzahl an falsch positiven Ergebnissen.

Bei der Polypen-Segmentierung werden generative Netze bereits erfolgreich eingesetzt, um reale Koloskopie-Bilder an synthetische anzunähern, damit auf diesen monokularen Bildern eine Tiefenkarte geschätzt werden kann [26]. Die Segmentierung als nächster Schritt fehlt allerdings noch. Eine Weiterentwicklung und Fusion der be-

Eine Weiterentwicklung und Fusion der bestehenden Ansätze bietet ein großes Potenzial zur Verbesserung der Polypen-Segmentierung. Denkbare Beispiele aus Sicht des Autors sind

- eine aggressive Regularisierung der CANs von Isola et al.,
- eine Fortführung der Tiefenschätzung von Mahmood et al. (von den Autoren bereits angekündigt),
- alle bereits erwähnten Ansätze (außer 3D-FCN), erweitert um Sequenzbetrachtung – entweder durch Erweiterung des 3D-CNN wie bei Lequan et al. oder durch zusätzliche Features wie den optischen Fluss vom vorherigen Frame.

Literaturverzeichnis

- [1] R. Lozano, M. Naghavi, K. Foreman, S. Lim, K. Shibuya u. a. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*, 380:2095–2128, 9859, 2012. ISSN: 0140-6736. DOI: 10.1016/S0140-6736(12)61728-0.
- [2] J. Ferlay, H. R. Shin, F. Bray, D. Forman, C. Mathers u. a. GLOBO-CAN, Cancer incidence and mortality worldwide: IARC CancerBase No. 10 [Internet]. Lyon, France: International Agency for Research on Cancer; 2010. globocan. iarc. fr, 2012.
- [3] V. Kumar, A. K. Abbas, N. Fausto, S. L. Robbins und R. S. Cotran, Herausgeber. Robbins and Cotran pathologic basis of disease. 7. ed. Auflage. Elsevier Saunders, Philadelphia, Pa., 7. ed. Auflage, 2005. 1525 Seiten. ISBN: 0721601871.



- [4] A. G. Zauber, S. J. Winawer, M. J. O'Brien, I. Lansdorp-Vogelaar, M. van Ballegooijen u. a. Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *The New England journal of medicine*, 366:687–696, 8, 2012. DOI: 10. 1056/NEJMoa1100370.
- [5] D. K. Rex. Reducing costs of colon polyp management. *The lancet oncology*, 10:1135–1136, 12, 2009. ISSN: 1470-2045.
- [6] G. Schachschal. Praktische Koloskopie: Methodik, Leitlinien, Tipps und Tricks; 16 Tabellen. Thieme, 2010. ISBN: 9783131477415.
- [7] V. B. S. Prasath. Polyp Detection and Segmentation from Video Capsule Endoscopy: A Review. *Journal of Imaging*, 3:1, 1, 2016. DOI: 10.3390/ jimaging3010001.
- [8] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez u. a. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Computerized Medical Imaging and Graphics*, 43:99–111, 2015. ISSN: 0895-6111. DOI: 10.1016/j. compmedimag.2015.02.007.
- [9] J. Bernal, J. Sánchez und F. Vilariño. Towards automatic polyp detection with a polyp appearance model. *Pattern Recognition*, 45:3166–3182, 9, 2012. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2012.03.002.
- [10] N. Tajbakhsh, S. R. Gurudu und J. Liang. Automated Polyp Detection in Colonoscopy Videos Using Shape and Context Information. *IEEE transactions on medical imaging*, 35:630–644, 2, 2016. DOI: 10.1109/TMI.2015.2487997.
- [11] M. Ganz, Y. Xiaoyun und G. Slabaugh. Automatic segmentation of polyps in colonoscopic narrow-band imaging data. *IEEE transactions on*

bio-medical engineering, 59:2144–2151, 8, 2012. ISSN: 1558-2531. DOI: 10.1109/TBME.2012.2195314.

- [12] I. Goodfellow, Y. Bengio und A. Courville. *Deep learning*. The MIT Press, Cambridge, Massachusetts und London, England, 2016. XXII, 775 Seiten. ISBN: 978-0-262-03561-3.
- K. Simonyan und A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition, 10. Apr. 2015. URL: http://arxiv. org/pdf/1409.1556.
- [14] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall u. a. Convolutional Neural Networks for Medical Image Analysis Full Training or Fine Tuning? *IEEE transactions on medical imaging*, 35:1299–1312, 5, 2016. DOI: 10.1109/TMI. 2016.2535302.
- [15] M. Billah, S. Waheed und M. M. Rahman. An Automatic Gastrointestinal Polyp Detection System in Video Endoscopy Using Fusion of Color Wavelet and Convolutional Neural Network Features. *International journal of biomedical imaging*, 2017:9545920, 2017. ISSN: 1687-4188. DOI: 10.1155/2017/9545920.
- [16] S. G. Mallat. Multifrequency channel decompositions of images and wavelet models. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37:2091–2110, 12, 1989. ISSN: 00963518. DOI: 10.1109/29.45554.
- [17] Y. Lequan, C. Hao, D. Qi, Q. Jing und A. H. Pheng. Integrating Online and Offline Three-Dimensional Deep Learning for Automated Polyp Detection in Colonoscopy Videos. *IEEE journal of biomedical and health informatics*, 21:65–75, 1, 2017. ISSN: 2168-2208. DOI: 10.1109 / JBHI. 2016.2637004.



- [18] J. Long, E. Shelhamer und T. Darrell. Fully convolutional networks for semantic segmentation. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (Boston, MA, USA,), Seiten 3431– 3440, Piscataway, NJ. IEEE, 2015. ISBN: 978-1-4673-6964-0. DOI: 10. 1109/CVPR.2015.7298965.
- [19] D. Vázquez, J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, A. M. López u. a. A Benchmark for Endoluminal Scene Segmentation of Colonoscopy Images. *Journal of healthcare engineering*, 2017:4037190, 2017. ISSN: 2040-2295. DOI: 10.1155/2017/ 4037190.
- [20] I. Wichakam, T. Panboonyuen, C. Udomcharoenchaikit und P. Vatee-kul. Real-Time Polyps Segmentation for Colonoscopy Video Frames Using Compressed Fully Convolutional Network. In K. Schoeffmann u. a., *Multi-Media modeling*. Band 10704, Nummer 10704, Seiten 393–404. Springer, Cham, 2018. ISBN: 978-3-319-73602-0. DOI: 10.1007/978-3-319-73603-7_32.
- [21] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley u.a. Generative Adversarial Nets. In Advances in Neural Information Processing Systems, Seiten 2672–2680, 2014.
- [22] E. L. Denton, S. Chintala, a. szlam arthur und R. Fergus. Deep Generative Image Models using a Laplacian

Pyramid of Adversarial Networks. In Advances in Neural Information Processing Systems, Seiten 1486–1494, 2015.

- [23] A. Radford, L. Metz und S. Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, 2016. URL: http://arxiv.org/pdf/ 1511.06434v2.
- [24] P. Isola, J.-Y. Zhu, T. Zhou und A. A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. In herausgegeben von R. Chellappa, Z. Zhang und A. Hoogs, *30th IEEE Conference on Computer Vision and Pattern Recognition*. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (Honolulu, HI), Seiten 5967–5976, Piscataway, NJ. IEEE, 2017. ISBN: 978-1-5386-0457-1. DOI: 10.1109/ CVPR.2017.632.
- [25] O. Zisimopoulos, E. Flouty, M. Stacey, S. Muscroft, P. Giataganas u.a. Can surgical simulation be used to train detection and classification of neural networks? *Healthcare technology letters*, 4:216–222, 5, 2017. ISSN: 2053-3713. DOI: 10.1049 / htl. 2017.0064.
- [26] F. Mahmood, R. Chen und N. J. Durr. Unsupervised Reverse Domain Adaptation for Synthetic Medical Images via Adversarial Training, 29. Nov. 2017. URL: http://arxiv.org/ pdf/1711.06606.

