


# Medical Image Segmentation with Efficient Use of Annotations and Resources

**Doctoral Thesis**

**Author(s):**

Ozdemir, Firat 

**Publication date:**

2020

**Permanent link:**

<https://doi.org/https://doi.org/10.3929/ethz-b-000404970>

**Rights / license:**

[In Copyright - Non-Commercial Use Permitted](#)

DISS. ETH NO. 26639

# Medical Image Segmentation with Efficient Use of Annotations and Resources

A dissertation submitted to attain the degree of  
DOCTOR OF SCIENCES of ETH ZURICH  
(Dr. sc. ETH Zürich)

presented by

**Firat Özdemir**

Master of Science in Electrical and Electronics Engineering  
EPFL, Switzerland  
born on 18. February 1991  
citizen of Turkey

accepted on the recommendation of

Prof. Dr. Orçun Göksel, examiner  
Prof. Dr. Philippe Cattin, co-examiner  
Dr. Christine Tanner, co-examiner

2020

# Abstract

Correct diagnosis for any pathology is at the core of successful treatment. Except for a small subset of conditions which are readily visible to the eye or touch, additional tools for observation are critical to have a precise understanding. Today, there are a multitude of medical imaging technologies, each with their own merits and pitfalls. Some are commonly used together in order to complement information from each other for a more precise observation of the patient and diagnosis of pathology. As a result, doctors can plan and analyze treatment options for an optimal recovery. This in return allows for increased life quality and life expectancy in the population. Consequently, number of patients per doctor continues to grow. Besides the resources needed for an optimal medical assessment and treatment, the increase in population per doctor demands for assistive tools to reduce time per expert and cost per patient care, while maintaining top quality in healthcare.

Semantic labeling (segmentation) is a field of research that aims to partition an image by associating each voxel with a semantic label, which could be the tissue type, organ, pathological state, functional purpose, etc. Segmentation is a bottleneck for applying a multitude of new technologies for patient diagnosis and treatment, such as radiotherapy and surgical planning. Unfortunately, manual labeling requires substantial amounts of time from medical experts for potentially large images and volumes. Today, a lot of clinical procedures are simplified, as personalized and patient-specific procedures that would require patient segmentation masks cannot be afforded for each patient.

Segmentation tools can substantially alleviate this manual burden. For both quality assurance and liability purposes, medical expert approval or interference is important for a segmentation tool which is to be clinically used. The *interactive segmentation* field develops methodologies that aim to produce reasonable segmentations with minimal time cost for the medical expert. Ideally, the user interaction routines should facilitate expert inspection of automatically generated segmentation predictions, while allowing for on-the-fly fine adjustments of the segmentation proposals when necessary for the procedure.

Traditional automatic segmentation pipelines used to require engineered features, which are now shifting towards data-driven features with *deep learning* that necessitate large corpuses of relevant annotated datasets for supervision. Annotation efforts can be prohibitively expensive, especially in the medical field, where annotations need unique expertise and the resources are costly and limited. *Active learning* techniques aim to optimally use the available resources in order to maximize expected task performance through informed selection of data samples to be manually annotated. This ensures the best allocation of cost-intensive and difficult-to-procure manual resources.

Different medical applications can rely on images with similar or overlapping field-of-views, e.g., interventions for tear of different muscles in the shoulder, or applications

relating to different tissues in the knee such as cartilage and ligaments. Leveraging the information from earlier applications, solutions, or annotations is crucial for affordable and swift extension to newer targets and applications. The scenario where a framework is expected to solve new objectives in addition to earlier ones when subjected to streams of datasets becoming sequentially available is explored under *class-incremental learning* field. This means that a framework initially optimized for a set of anatomical structures should be able to extend to segment both the earlier and the incremental set of anatomical structures, with minimal amount of additional expert labor, i.e., manual annotations.

Accordingly, different aspects of the challenges described above are studied in this thesis. An interactive segmentation methodology based on random walker method on a multi-scale 3D graph is proposed to efficiently incorporate user input to reduce manual annotation time. New metrics in active learning are studied for selecting images to annotate, in order to reduce cumulative annotation time. In order to reduce annotation needs with neural networks based class-incremental learning, optimal sample selection for segmentation is also explored in this thesis, in particular for limited sample storage scenarios. Additionally, an automatic segmentation method by leveraging imaging physics with graphical models is presented to require minimal annotations.

# Zusammenfassung

Die richtige Diagnose für jede Pathologie ist der Kern einer erfolgreichen Behandlung. Abgesehen von einer kleinen Teilmenge von Erkrankungen, die von bloßem Auge sichtbar oder durch Berührung fühlbar sind, sind zusätzliche Hilfsmittel zur Beobachtung entscheidend, um ein genaues Verständnis zu erhalten. Heute gibt es eine Vielzahl von medizinischen Bildgebungstechnologien, wobei jede ihre eigenen Vorzüge und Tücken aufweist. Einige werden häufig zusammen verwendet um Informationen voneinander zu ergänzen. Dies ermöglicht eine genauere Beobachtung des Patienten sowie eine präzisere Diagnose der Pathologie. Dadurch können Ärzte die Behandlungsmöglichkeiten für eine optimale Genesung planen und analysieren. Im Gegenzug entsteht daraus eine erhöhte Lebensqualität und Lebenserwartung in der Bevölkerung, wodurch die Zahl der Patienten pro Arzt weiter ansteigt. Neben den Ressourcen, die für eine optimale medizinische Beurteilung und Behandlung benötigt werden, erfordert die Zunahme der Bevölkerung pro Arzt unterstützende Hilfsmittel, um die Zeit pro Experte und die Kosten pro Patient zu reduzieren und gleichzeitig die Qualität der Versorgung auf höchstem Niveau zu halten.

Die semantische Markierung (Segmentierung) ist ein Forschungsgebiet das darauf abzielt, jedem Voxel im Bild ein semantisches Label zuzuordnen. Dieses kann den Gewebetyp, das Organ, den pathologischen Zustand, den funktionellen Zweck, usw., betreffen. Sie ist ein Flaschenhals für die Anwendung einer Vielzahl neuer Technologien für die Diagnose und Behandlung von Patienten, wie z.B. Strahlentherapie und Operationsplanung. Da ein manuelles Labelling bei potenziell großen Bildern und Volumen einen erheblichen Zeitaufwand für das medizinische Personal bedeutet, werden heute viele klinische Verfahren vereinfacht; personalisierte und patientenindividuelle Verfahren, die Patienten-Segmentierungsmasken erfordern würden, können aus Zeitgründen nicht geleistet werden.

Mit Hilfe von automatischen Segmentierungswerkzeugen kann dieser manuelle Aufwand erheblich reduziert werden. Aus Gründen der Qualitätssicherung, sowie auch der Haftung, ist die Zustimmung oder Miteinbeziehung medizinischer Experten für ein Segmentierungswerkzeug, das klinisch eingesetzt werden soll, notwendig. Im Bereich der *interaktiven Segmentierung* werden Methoden mit dem Ziel entwickelt, sinnvolle Segmentierungen mit minimalem Zeitaufwand für den medizinischen Experten zu erstellen. Im Idealfall sollten die Benutzerinteraktionsroutinen die fachkundige Überprüfung der automatisch generierten Segmentierungsvorhersagen erleichtern und gleichzeitig eine on-the-fly Feinjustierung der Segmentierungsvorschläge ermöglichen, wenn dies für das Verfahren notwendig ist.

Während traditionelle automatische Segmentierungs-Pipelines auf menschengemachten Features beruhen, werden solche, für die Segmentierung relevanten Features, mehr und mehr automatisch aus dem vorliegenden Bildmaterial berechnet. Dies geschieht mit Hilfe

von tiefem Lernen, oder *deep learning*. Allerdings benötigt dies eine grosse Menge manuell annotierter Datensätze. Der Aufwand für diese Annotation kann unerschwinglich sein, insbesondere im medizinischen Bereich, wo dazu einzigartiges Fachwissen notwendig ist und die Ressourcen kostspielig und begrenzt sind. *Aktive Lerntechniken* zielen darauf ab, die verfügbaren Ressourcen optimal zu nutzen um die notwendige Durchführung der Aufgabe durch eine sachkundige Auswahl der manuell zu annotierenden Datenproben zu maximieren. Dies gewährleistet die beste Allokation der kostenintensiven und schwer zu beschaffenden manuellen Ressourcen.

Mehrere medizinische Anwendungen können auf Bilder mit ähnlichen oder überlappenden Bildfeldern zurückgreifen, z.B. Eingriffe bei Rissen verschiedener Muskeln an der Schulter oder Anwendungen, die sich auf verschiedene Gewebe im Knie wie Knorpel und Bänder beziehen. Die Nutzung von Informationen aus früheren Anwendungen, Lösungen oder Annotationen ist entscheidend für eine kostengünstige und schnelle Erweiterung auf neuere Ziele und Anwendungen. Das Szenario, bei dem von einem Framework erwartet wird, dass es zusätzlich zu den früheren Aufgaben neue Aufgaben löst, wenn die Datenströme nacheinander verfügbar werden, wird im Bereich des *klasseninkrementellem Lernens* untersucht. Dies bedeutet, dass ein Framework, das zunächst für einen Satz anatomischer Strukturen optimiert wurde, in der Lage sein sollte, sowohl den früheren als auch einen inkrementellen Satz neuer anatomischer Strukturen mit minimalem zusätzlichem Arbeitsaufwand für Experten, d.h. manuelle Annotationen, zu segmentieren.

Dementsprechend werden in dieser Arbeit verschiedene Aspekte der oben beschriebenen Herausforderungen untersucht. Eine interaktive Segmentierungsmethodik, die auf der Random-Walker-Methode in einem mehrskaligen 3D-Graphen basiert, wird vorgeschlagen, um Benutzereingaben effizient einzubeziehen und so die manuelle Annotationszeit zu reduzieren. Neue Metriken im aktiven Lernen werden für die Auswahl der zu annotierenden Bilder untersucht, um die kumulative Annotationszeit zu reduzieren. Um den Annotationsbedarf bei klasseninkrementellem Lernen auf Basis neuronaler Netze zu reduzieren, wird in dieser Arbeit auch die optimale Sampleauswahl für die Segmentierung untersucht, insbesondere für begrenzte Samplespeicherszenarien. Zusätzlich wird eine automatische Segmentierungsmethode vorgestellt, die die bildgebende Physik graphischen Modellen kombiniert, um den Annotationsbedarf zu minimieren.