IPEM

Institute of Physics and
Engineering in Medicine

TOPICAL REVIEW • OPEN ACCESS

Beyond automatic medical image segmentation—the spectrum between fully manual and fully automatic delineation

View the article online for updates and enhancements.

Physics in Medicine & Biology

IPEM
Institute of Physics and
Engineering in Medicine

CrossMark

TOPICAL REVIEW

Beyond automatic medical image segmentation—the spectrum between fully manual and fully automatic delineation

Michael J Trimpl[1,2,3] , Sergey Primakov[4], Philippe Lambin[4], Eleanor P J Stride[2] , Katherine A Vallis[3] and Mark J Gooding[1]

[1] Mirada Medical Ltd, Oxford, United Kingdom
[2] Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, United Kingdom
[3] Oxford Institute for Radiation Oncology, University of Oxford, Oxford, United Kingdom
[4] The D-Lab, Department of Precision Medicine, GROW-School for Oncology, Maastricht University, Maastricht, NL, The Netherlands

E-mail: michael.trimpl@mirada-medical.com

Abstract

Semi-automatic and fully automatic contouring tools have emerged as an alternative to fully manual segmentation to reduce time spent contouring and to increase contour quality and consistency. Particularly, fully automatic segmentation has seen exceptional improvements through the use of deep learning in recent years. These fully automatic methods may not require user interactions, but the resulting contours are often not suitable to be used in clinical practice without a review by the clinician. Furthermore, they need large amounts of labelled data to be available for training. This review presents alternatives to manual or fully automatic segmentation methods along the spectrum of variable user interactivity and data availability. The challenge lies to determine how much user interaction is necessary and how this user interaction can be used most effectively. While deep learning is already widely used for fully automatic tools, interactive methods are just at the starting point to be transformed by it. Interaction between clinician and machine, via artificial intelligence, can go both ways and this review will present the avenues that are being pursued to improve medical image segmentation.

1. Introduction

Image segmentation is an integral part of many medical tasks. For instance in radiotherapy, image segmentation is essential for treatment planning (Ramkumar *et al* 2016), enabling the radiation dose delivered to different regions to be calculated and so minimise damage to healthy tissue. More generally, segmentation of anatomical structures allows for detailed volumetric analysis to facilitate various diagnostic and clinical decision-making processes (van Timmeren *et al* 2020).

Segmentation can be performed using either manual, semi-automatic or fully automatic methods. Manual contouring gives the clinician full control over the process, but it can be tedious and time-consuming work. Additionally, manual contouring is prone to inter- and intra-observer variability (Sharp *et al* 2014). In contrast, fully automatic segmentation methods do not require any human input and have the potential to be very fast and highly reproducible and ultimately, to reduce the workload for clinicians (Primakov *et al* 2022). Advances in machine learning (ML) have greatly improved fully automatic approaches (Hamidian *et al* 2017, Bakator and Radosav 2018, Hosny *et al* 2018, Jarrett *et al* 2019, Renard *et al* 2020) and ML-based methods are now being implemented in clinical systems (Lustberg *et al* 2018). Notably, contours generated using fully automatic ML tools have been shown to be indistinguishable from manual contours (Gooding *et al* 2018, Liu *et al* 2019, Primakov *et al* 2022).

Fully automatic methods still face several problems, however. In particular, a large and well-curated labelled training set is required for most ML approaches. Such data sets are not yet available for many regions-of-interest/anatomical structures. Furthermore, the variability of some structures, such as tumours, makes it very
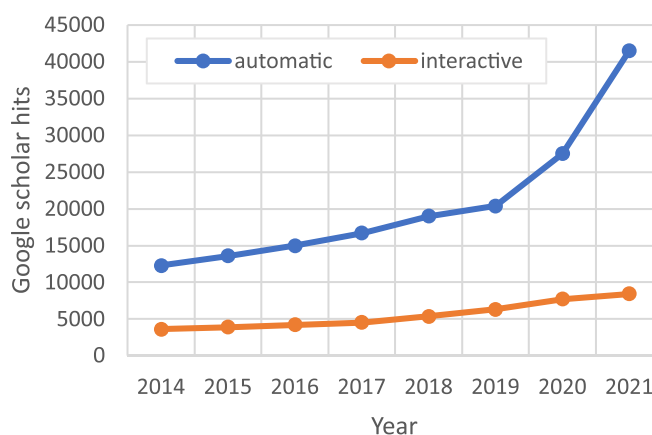
Figure 1. Google scholar search hits for keywords 'segmentation' 'deep learning' 'medical imaging' and additionally 'automatic' (blue) or 'interactive' (orange).

difficult to build a sufficiently representative dataset (Tian *et al* 2021). Technical differences between scanners and reconstruction protocols that are used for image acquisition pose another problem for fully automated ML methods, further increasing the need for large and diverse training data sets (Zhao *et al* 2014, Mackin *et al* 2015). The complexity of ML and particularly deep neural networks can also make it difficult to determine the uncertainty associated with segmentation boundaries, so that expert review by clinicians is still required (Wang *et al* 2020a, Abdar *et al* 2021). While fully automated ML-based segmentation methods are very successful for some anatomical and pathological regions, but for others remedial human intervention is required, increasing both clinical workload and inter- and intra-observer variability.
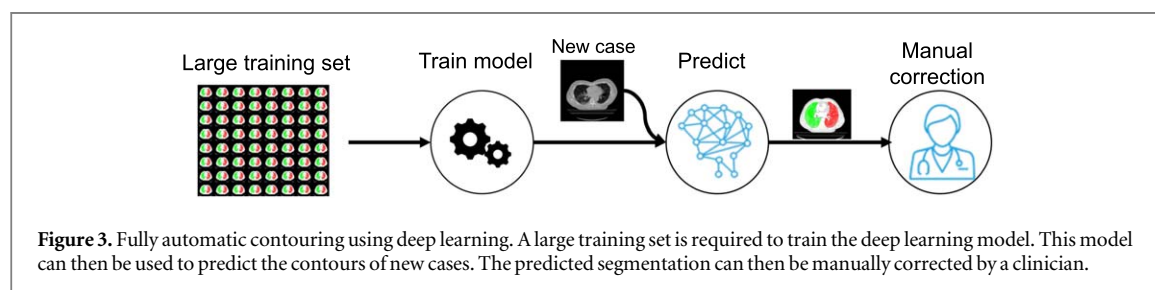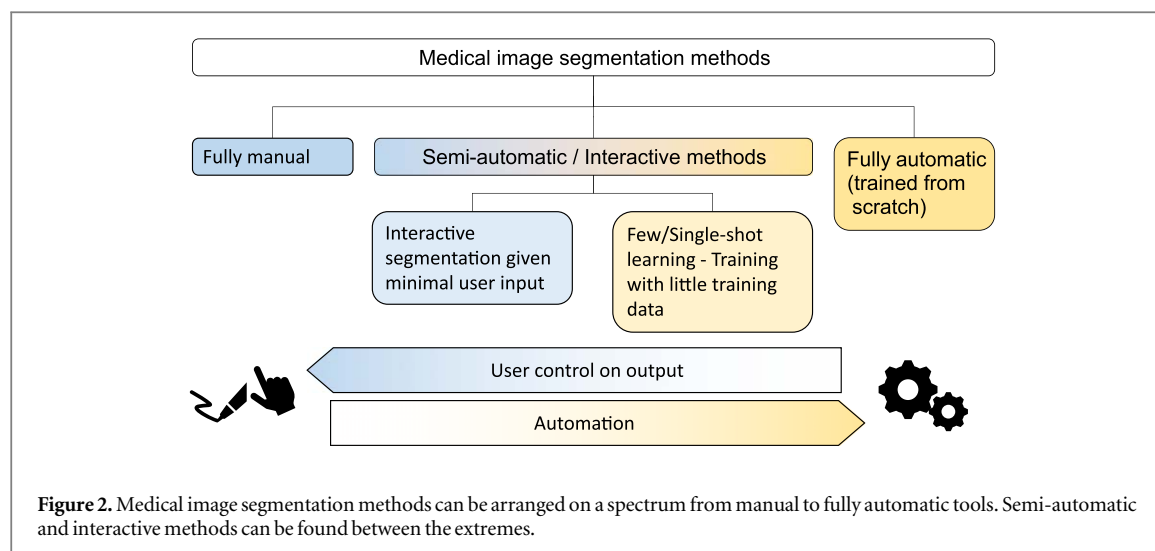
Semi-automatic or interactive segmentation methods assist clinicians in cases for which fully automatic methods fail. While interactive methods still require manual input, they can reduce the extent of user interaction by making predictions based on adjustments made by previous users. This gives more control over the contour outcome than with fully automated workflows, while still benefiting from semi-automation in achieving satisfactory segmentation. When compared to fully manual segmentation, interactive techniques have been shown to improve consistency and repeatability and reduce the time spent on contouring (Olabarriaga and Smeulders 2001, Wang *et al* 2018, Wang *et al* 2019b, Sakinis *et al* 2019). The types of user interaction can vary from a few clicks or stylus strokes to produce a contour on a single slice, to drawing selected contours in 2D to produce a 3D contour on a medical image. User interaction can be used to create contours from scratch or can be used to refine the structures of automatic methods. Interest in fully automatic deep learning methods has increased rapidly in recent years, whereas interactive methods have received comparatively little attention, as illustrated in figure 1. Consequently, there is considerable opportunity for progress in this area.

The aim of this review is to describe and evaluate current segmentation methods on a spectrum from manual to fully automatic tools, as shown in figure 2. The review starts with a synopsis of state-of-the-art fully automatic deep learning methods for segmentation, corresponding to the far-right side of the spectrum. Moving to the left, few-shot and transfer learning approaches can be trained with small quantities of annotated data. This distinguishes them from fully supervised methods that require large data sets that are difficult to obtain due to costly annotation time and data privacy regulations. Closer to manual editing, interactive methods provide some of the benefits of automation while retaining user control over the contour output. The review ends with a discussion of emerging techniques such as guiding user interaction through feedback provided by ML to achieve full interactivity.

2. Deep learning in medical image segmentation

Deep learning has become established as the method of choice for fully automated contouring. The process from creation to application of an automatic deep learning contouring tool is shown in figure 3. This section will give a brief overview of the development of deep learning methods for medical imaging, and the challenges that remain for example in interactive segmentation. A wider introduction to deep learning applied to medical imaging can be found in (Erickson *et al* 2017, Litjens *et al* 2017, Shen, Wu and Suk 2017, Suzuki 2017, Anwar *et al* 2018, Guo *et al* 2019, Kim *et al* 2019, Willemink *et al* 2020, Wang *et al* 2020b, Seo *et al* 2020, Tajbakhsh *et al* 2020).

Deep learning, as a sub-field of ML, has led to many breakthroughs in computer vision tasks. Fundamentally, it uses neural networks to extract features from the provided input data and map these to an
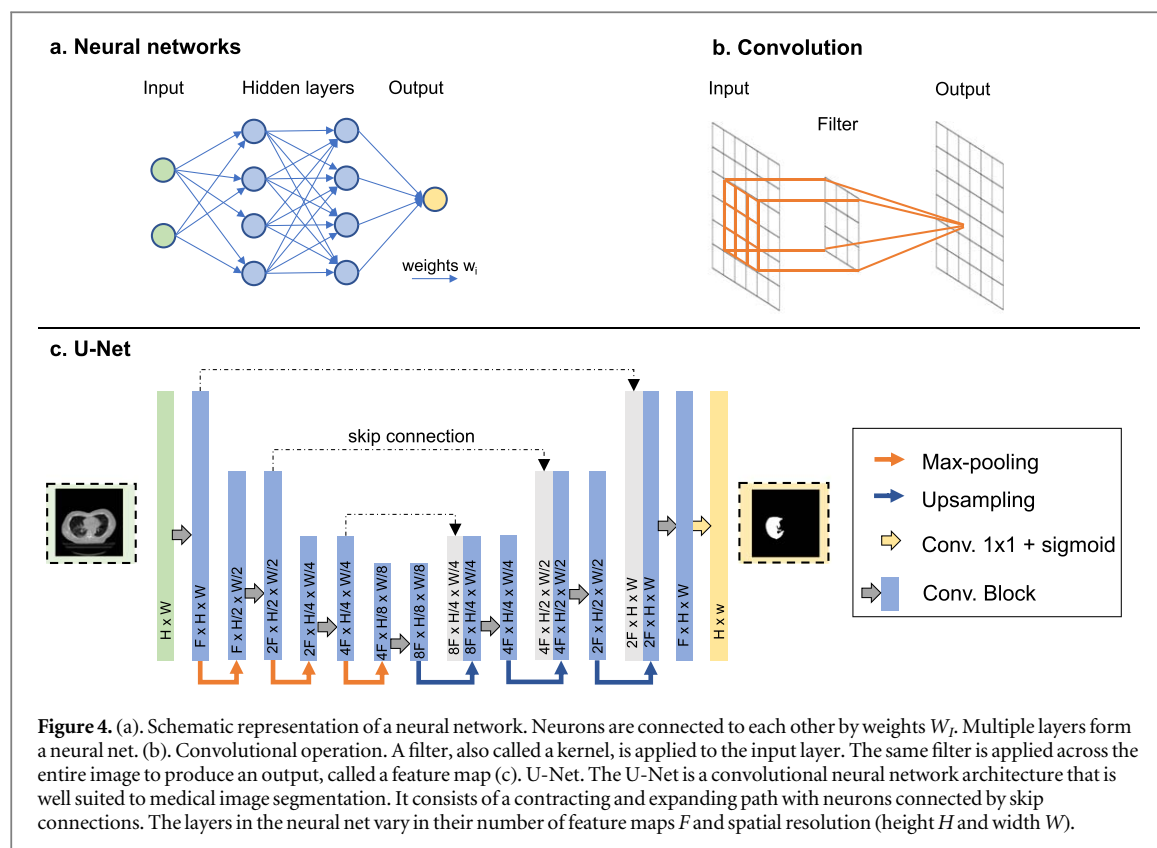
Figure 2. Medical image segmentation methods can be arranged on a spectrum from manual to fully automatic tools. Semi-automatic and interactive methods can be found between the extremes.

Figure 3. Fully automatic contouring using deep learning. A large training set is required to train the deep learning model. This model can then be used to predict the contours of new cases. The predicted segmentation can then be manually corrected by a clinician.

output. A neural network consists of information processing units called neurons, that are connected to each other to form the neural network, as shown in figure 4(a). If many layers of neurons are stacked, it is called deep learning. The recent success of deep learning has only been made possible by the availability of increased computing power that can handle the computationally expensive task of training deep neural networks (Schmidhuber 2015).

For computer vision tasks, including medical image segmentation, convolutional neural networks (CNN) are the most successful network architectures. CNNs are inspired by the hierarchical receptive field model of the visual cortex of the human brain (Hubel and Wiesel 1959). The key features of CNNs are convolutional layers that apply a filter to the input to extract features. The appearance of an object is recognized independent of its location in an image. Thus, detection can be performed using convolution across the image. A given neuron gets a weighted input from the units in the previous layer within a small receptive field. In deep learning, multiple layers are stacked to achieve an increasingly wide receptive field. By sharing the weights of the feature mapping in different positions for each layer, the number of parameters can be decreased compared to other types of neural networks. Illustration of a convolutional operation is shown in figure 4(b).

One of the first implementations of CNNs for image segmentation was a fully convolutional network (FCN) (Long *et al* 2015). This enabled the use of non-fixed input sizes, by using exclusively convolutional layers, as well as the output of a spatial segmentation map. Skip connections were used to merge upsampled feature maps from the final layers with feature maps of earlier layers. The disadvantage of this type of FCN is that the upsampling is crude and not sensitive to the details of the image. The resulting segmentation is therefore of low resolution. Additionally, in FCN, pixels are classified without fully considering spatial consistency between them.

Shortcomings of FCNs were addressed by U-Net architecture (Ronneberger *et al* 2015), which has since been widely used in medical image segmentation. In principle, the U-Net architecture consists of a contracting path to capture the spatial context and an expanding path for localization (figure 4(c)). Both pathways are connected by skip connections. Skip connections provide an alternative path for the gradient, which helps solve the vanishing gradient problem often faced in deep neural networks (Drozdzal *et al* 2016). A vanishing gradient during backpropagation can prevent the neural network from updating the weights successfully during training. Furthermore, skip connections allow the U-Net to combine high-level and low-level image information and localize these (Drozdzal *et al* 2016). The U-net architecture was first applied to 2D microscope images. Many variants of U-Net architecture have emerged to improve performance of specific tasks. For example, for 3D volumetric images, the 3D U-Net and V-Net architectures (Ciçek *et al* 2016, Milletari *et al* 2016) replace 2D

Figure 4. (a). Schematic representation of a neural network. Neurons are connected to each other by weights W_I. Multiple layers form a neural net. (b). Convolutional operation. A filter, also called a kernel, is applied to the input layer. The same filter is applied across the entire image to produce an output, called a feature map (c). U-Net. The U-Net is a convolutional neural network architecture that is well suited to medical image segmentation. It consists of a contracting and expanding path with neurons connected by skip connections. The layers in the neural net vary in their number of feature maps F and spatial resolution (height H and width W).

convolutional blocks with 3D convolutions. This enables information from neighbouring slices to be used in generation of a contour on a particular slice, which is often critical in medical applications. Other modified neural networks include the Res-UNet (He *et al* 2016, Wang *et al* 2017a, Alom *et al* 2018) which includes so called residual layers to mitigate the vanishing gradient problem, Attention U-Net (Oktay *et al* 2018, Schlemper *et al* 2019) which replaces skip connections with self-learned attention gates to determine which image regions matter most, or the hybrid densely connected U-Net (Li *et al* 2018) which fuses features from a 2D and 3D U-Net to combine intra-slice representations and inter-slice features, respectively.
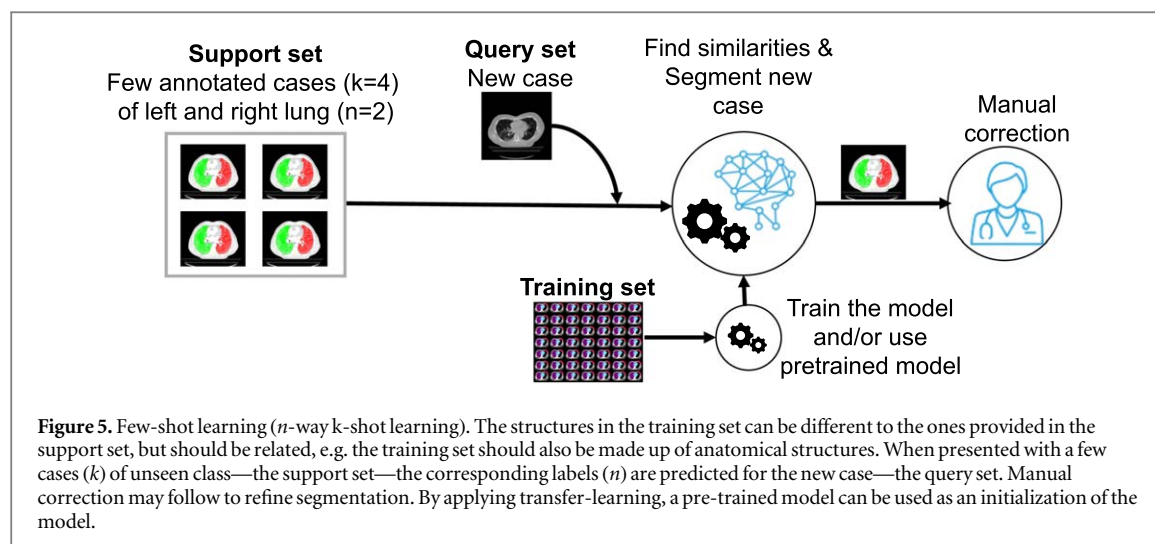
The above are just a few examples of how deep learning has been applied to fully automatic segmentation in medical imaging. These methods still face challenges that arise from working with deep learning, such as integrating interactivity into the workflow or learning on small datasets. One option to overcome the need for large quantities of data is to use alternative deep learning frameworks, such as few-shot learning or transfer learning. These will be discussed in more detail in the next section.

3. Few-shot learning, transfer learning and fine-tuning

Currently, the majority of the research on automatic segmentation for medical imaging utilizes fully-supervised ML models (Isensee *et al* 2018, Li *et al* 2018, Zhou *et al* 2018, Wei *et al* 2020). There is, however, an increasing number of articles proposing weakly-supervised methods, including low-, few-, one- and zero-shot learning approaches. An in-depth review of these methods can be found in Kadam and Vaidya (2018), Wang *et al* (2020d). Fully supervised DL methods require large and representative datasets to maximize their performance. However, in the medical imaging field, acquiring such datasets and corresponding segmentation labels can be problematic, due to restrictive personal data regulations, heterogenous clinical protocols and labelling complexity. In cases when there are only a few training images available, fully supervised methods struggle to generate correct predictions. To tackle this problem few-shot/low-shot learning methods have been proposed. Transfer learning and fine-tuning frameworks can also help when dealing with small annotated datasets of medical images (Pan and Yang 2010, Karimi *et al* 2020, Karimi *et al* 2021).

3.1. Few-shot learning methods in medical imaging segmentation
The name few-shot/low-shot usually refers to methods that use a small number of labelled images, called the 'support set', to assist in solving the segmentation task. Training the model still requires a large training set from which the support set is sampled to simulate the few-shot problem (Shaban *et al* 2017). The goal of training is not
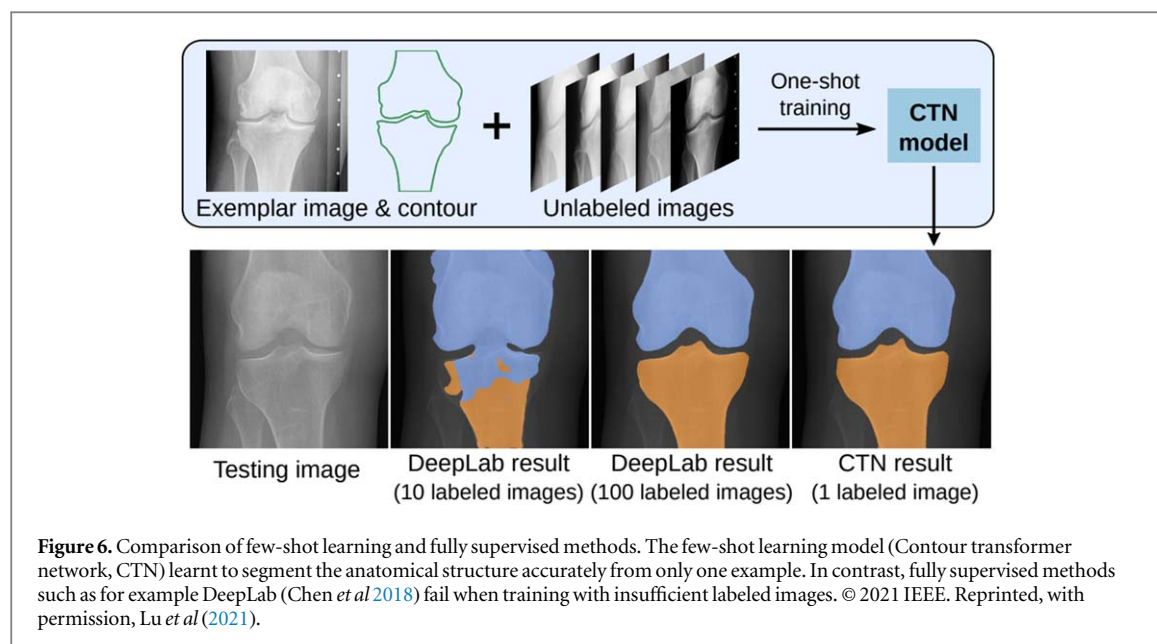
Figure 5. Few-shot learning (*n*-way k-shot learning). The structures in the training set can be different to the ones provided in the support set, but should be related, e.g. the training set should also be made up of anatomical structures. When presented with a few cases (*k*) of unseen class—the support set—the corresponding labels (*n*) are predicted for the new case—the query set. Manual correction may follow to refine segmentation. By applying transfer-learning, a pre-trained model can be used as an initialization of the model.

to know what a specific structure is, as is the case in the fully automatic approach discussed in the previous section. Instead, the goal is to learn the similarity and difference between structures. The training set may have a variety of structures in it, expanding the pool of available data for training, even if only a few cases are available for the specific structure to be segmented.

The 'query set' contains the images that are to be segmented. A method that uses a support set with k labelled images and n semantic classes would be called a n-way k-shot learning approach (Shaban *et al* 2017). The term Zero-shot Learning methods refers to methods where the target class is not present in the support set (Bucher *et al* 2019). A few annotated cases are all that are needed to achieve the corresponding segmentation on new cases, as illustrated in figure 5.

Early studies on few-shot learning were mainly focused on image classification tasks (Fei-Fei *et al* 2003, 2006, Santoro *et al* 2016, Snell *et al* 2017). However, this approach was soon adopted in natural image computer vision for segmentation tasks (Dong and Xing 2018, Zhang *et al* 2018) due to its reduced demand for supporting data. To perform one shot semantic segmentation, Shaban *et al* proposed using a model with two branches: conditioning and segmentation (Shaban *et al* 2017). The conditioning branch is used to extract the parameters from the masked support set image and the segmentation branch extracts the features from the query set image. The final segmentation mask is then obtained by performing pixel level logistic regression on the query set features using parameters from the conditioning branch (Shaban *et al* 2017). This approach has since been improved in several ways. Instead of using separate feature extractors for the support and query set, it has been proposed to use shared network weights to extract the features from both sets, reducing the number of parameters in the model (Wang *et al* 2019c). Following the feature extraction, masked averaged pooling has been shown to better extract foreground and background information from the support set. Additionally, prototype alignment regularization was introduced: When the segmentation mask was produced for the query image, this mask was used to perform the few-shot segmentation in reverse—from query to support—which allowed alignment of the prototype representations between query and support set during training (Wang *et al* 2019c).

In the medical imaging field, several investigators have tried to adopt few-shot learning, proposing various approaches. Mondal *et al* (2018) pioneered few-shot medical imaging segmentation and argued that the methods suggested by Shaban *et al* have limitations, due to the heterogeneity of medical images (Shaban *et al* 2017, Mondal *et al* 2018, Wang *et al* 2019c). Therefore, a modified 3D U-Net was proposed as a discriminator in a generative adversarial network (GAN) setting to perform few-shot infant brain MRI segmentation. This enabled the use of the unlabelled and synthetic image patches to boost the performance of the 3D U-Net (Mondal *et al* 2018). As an alternative approach, the one-shot medical imaging segmentation task can be treated as a classical atlas-based segmentation problem. For this, a VoxelMorph framework was used with a GAN for additional supervision (Wang *et al* 2020c). The proposed method takes the atlas and target images as input and predicts the correspondence map which can be applied to transfer the segmentation label (Wang *et al* 2020c). More recently, Lu *et al* presented a one-shot anatomical structure segmentation method (Contour transformer network, CTN) incorporating user intervention (Lu *et al* 2021). The framework takes only one labelled image and a set of unlabelled images as input, to generate the predicted contour. To bring this solution into clinical application user corrections were incorporated to improve segmentation performance (Lu *et al* 2021). Training the CTN model requires only one labeled image and leverages additional unlabeled data through loss functions that measure the global shape and appearance consistency of contours. The CTN uses a pre-trained network trained on ImageNet (Deng *et al* 2010) as the backbone of the encoding block. Optionally, additional labeled data

Figure 6. Comparison of few-shot learning and fully supervised methods. The few-shot learning model (Contour transformer network, CTN) learnt to segment the anatomical structure accurately from only one example. In contrast, fully supervised methods such as for example DeepLab (Chen *et al* 2018) fail when training with insufficient labeled images. © 2021 IEEE. Reprinted, with permission, Lu *et al* (2021).

or user annotation on mislabeled outputs of the network can be used to fine-tune the model in order to improve segmentation performance. An example segmentation of this approach is shown in figure 6 and compared to a fully supervised approach. The one-shot learning approach can successfully segment the structures whereas the fully supervised approach fails unless a sufficiently large training set is provided.
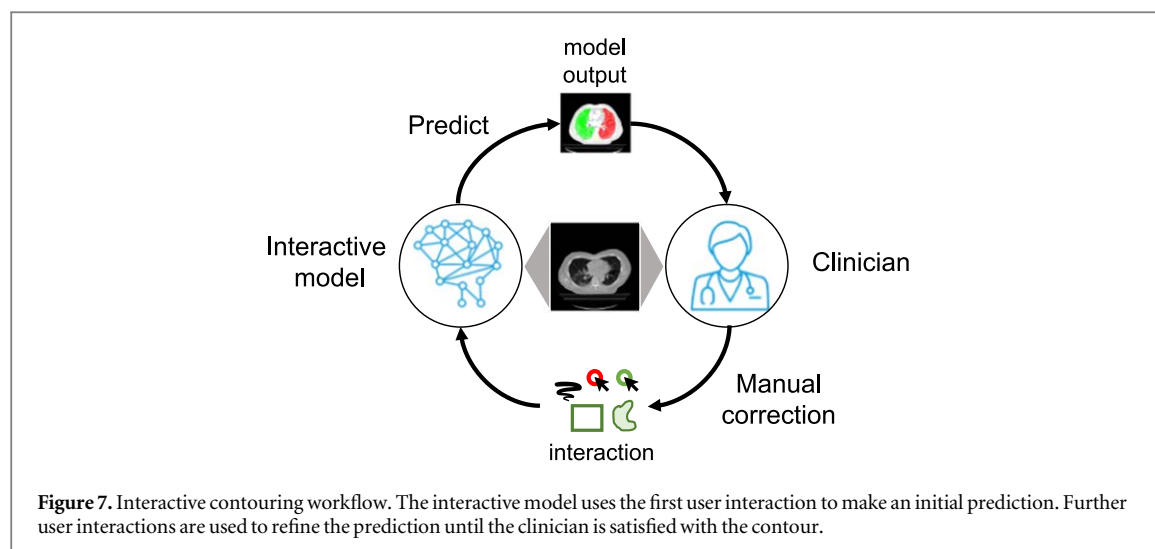
3.2. Transfer learning

Deep learning has been used for medical image segmentation with a specific task in mind. Consequently, these methods are built and trained from scratch using a task-specific dataset. However, the human brain - the inspiration for modern neural networks—does not need to be retrained in this same way. By learning how to recognize specific shapes our brain can transfer this knowledge and reuse it for solving more complex tasks (Parisi *et al* 2019). Transfer learning is the idea that previously learned neuron interaction coefficients, i.e. image features, can be used to solve new tasks instead of starting from scratch (Pan and Yang 2010).

Transfer learning can help in tackling the limited data problem by transferring knowledge from models that were trained previously on the large datasets. This can reduce the amount of data needed to train a model for the new task. Moreover, in some settings, it can drastically reduce the model training time. Compared to few-shot learning, however, a larger training set is required.

Transfer learning can be applied in a variety of settings that depend on the problem domain, model selection, and available data. Pan *et al* distinguish at least 3 settings based on the availability of the labelled data: Inductive Transfer Learning, Transductive Transfer Learning, and Unsupervised Transfer Learning (Pan and Yang 2010). Inductive Transfer Learning characterizes the condition when both, target and source tasks are different but related, and source and target domains are the same. An example in medical imaging would be segmentation of a different organ in the same image modality; or using a model trained for lung tumour classification to initialize a lung tumour segmentation model. In Transductive Transfer Learning, tasks are the same, but the source and target domains are different, e.g. segmentation of the same organ on different image modalities. In Unsupervised Transfer Learning, both tasks and domains are different but related and there are no labels available in both domains during training. For example, using clustering to identify distinguishing characteristics in lung tumour patient CTs and utilize these clusters as features in a cancer classification model (Pan and Yang 2010).

In most cases, Deep Learning models utilize an inductive transfer learning strategy, where previously trained weights could be used in two ways, i.e. without retraining as a feature extractor or being fine-tuned for a target dataset. When the pre-trained model is used as a feature extractor, the pre-trained weights are used without being updated when trained on a new task. During training on target data, only the last layer gets trained. If additional layers of the networks are retrained on the new data, it is called fine-tuning.

Fine-tuning is frequently used together with the transfer learning approach; it can include several techniques such as selective layers retraining or pruning. After transferring the weights from a pre-trained network, users can decide to re-train some of the deep layers together with a model's final fully-connected layer (Peng and Wang 2020). These deep layers may contain very specific features that are irrelevant to a new domain/problem. By retraining them on the target data these neurons will learn features specific to the target domain, therefore contributing to the performance of the model (Wang *et al* 2017b). Pruning is another technique to deal with

Figure 7. Interactive contouring workflow. The interactive model uses the first user interaction to make an initial prediction. Further user interactions are used to refine the prediction until the clinician is satisfied with the contour.

irrelevant features. Network pruning methods allow redundant neurons to be omitted during inference, resulting in reduced computing costs (Luo *et al* 2017, Liu *et al* 2018).

Transfer learning with fine-tuning has been widely adopted in the medical imaging field (Raghu *et al* 2019, Karimi *et al* 2020, Wang *et al* 2020b, Karimi *et al* 2021). In many of these works, a deep learning model that was pre-trained on a publicly available imaging dataset such as ImageNet (Deng *et al* 2010) is used. This model is then fine-tuned on the target medical imaging dataset (Rajpurkar *et al* 2017, Wang *et al* 2017c, Gulshan *et al* 2019, Liu and Chan 2021). Several studies have reported that transfer learning-based pipelines were able to achieve performance comparable to a human reader (Esteva *et al* 2017, Ding *et al* 2019). For example, the pretrained architecture (InceptionV3) fine-tuned on a PET brain dataset to predict Alzheimer's disease (Ding *et al* 2019) outperformed the user's performance on the independent testing set (Ding *et al* 2019). Using a similar approach, a classification of skin lesions, equivalent to that defined by dermatologists, was achieved on a test set consisting of 1942 images (Esteva *et al* 2017). To compare transfer learning with a fully-supervised training model with no knowledge transfer, Van Opbroek showed that when there are few data available, transfer learning can outperform common supervised methods and reduce classification errors by up to 60% (Van Opbroek *et al* 2015). Despite the success of transfer learning in medical imaging, some studies have drawn attention to the possibility of overparameterization and suggested the use of more flexible hybrid approaches to transfer learning for medical imaging tasks (Raghu *et al* 2019).

## 4. Interactive methods in medical image segmentation

In contrast to few-shot learning approaches, which move along the spectrum shown in figure 2 by seeking to reduce the need for training data for automated segmentation, interactive methods start with fully manual contouring and seek to reduce the need for manual intervention. Such interactive tools can be applied both when contouring manually and when editing a pre-existing segmented image. User interaction can be introduced in different ways for semi-automatic processing: 2D contours can be generated by using clicks (Sakinis *et al* 2019, Alemi Koohbanani *et al* 2020), scribbles (Lin *et al* 2016, Wang *et al* 2016, Boers *et al* 2020) or bounding boxes (Rajchl *et al* 2017, Wang *et al* 2018, Redekop and Chernyavskiy 2021); a contour may also be adjusted automatically in real-time while the clinician is drawing (Barrett and Mortensen 1997) ; or some 2D contours can be used to generate 3D contours of a structure (Léger *et al* 2018, Michael Trimpl *et al* 2021). The key to the success of these interactive methods is to try and find a balance between the human interaction and automation. Often this results in an iterative workflow between user interaction and processing, as illustrated in figure 7. Interactive tools have the potential to save the clinician time and effort when contouring, compared to manual annotation alone or reviewing and editing the results of fully automated segmentation.

This section discusses interactive methods for 2D or 3D segmentation. For this, approaches that do not rely on deep learning will be introduced and contrasted with deep learning methods.

4.1. Interactive methods in 2D

In this section, different methods for generating contours in 2D are discussed. An example of how user interactions can result in a contour is illustrated in figure 8. The figure shows a series of clicks that result in the segmentation of the user indicated structure. To achieve such interactive segmentation, a wide range of
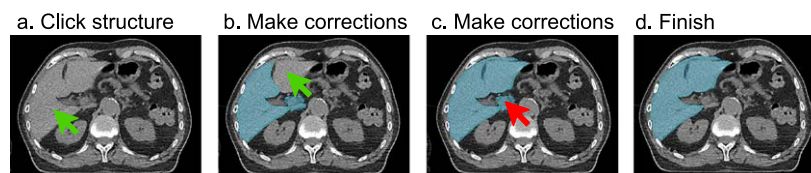
Figure 8. Example of interactive 2D segmentation via clicks. (a) With a first click the structure to be segmented is selected. This resulting initial segmentation may need further corrections which can be edited by clicking on areas that need to be (b) included (c) or excluded. Further clicks can be used until arriving at the (d) final segmentation.

interactive image segmentation approaches exist that do not rely on deep learning. Here, a selection of these methods and how they affect the development of deep learning methods will be discussed. A fuller description of non-deep learning methods may be found in Camilus and Govindan (2012), Zhao and Xie (2013).
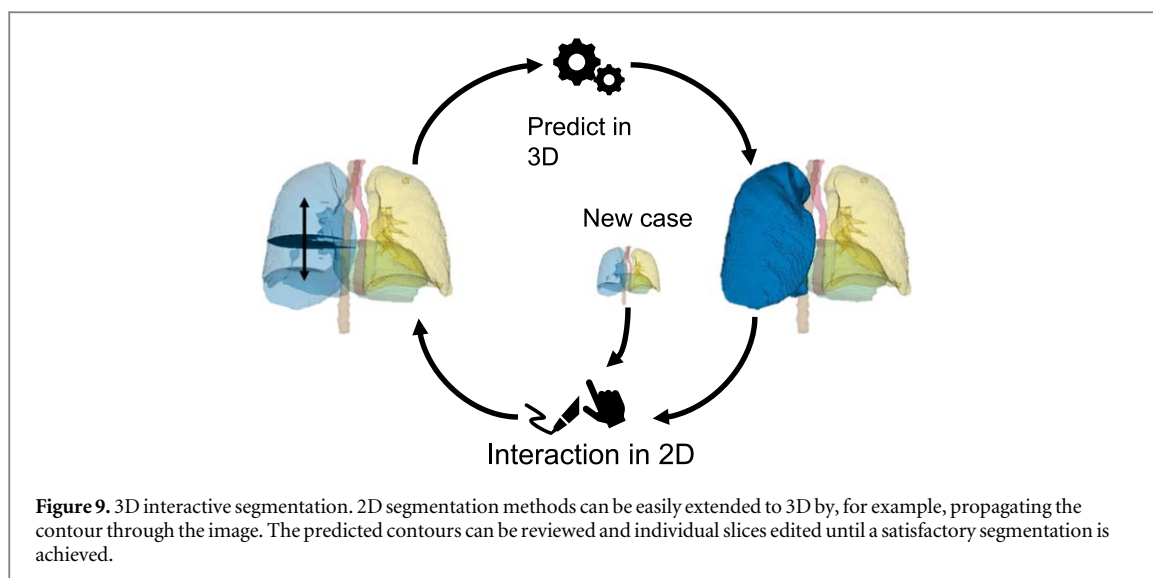
4.2. Traditional methods

Graphical energy minimisation techniques have been a popular method for image segmentation (Xu and Prince 1998, Boykov and Jolly 2001, Komodakis *et al* 2007, Zhou *et al* 2016) particularly for interactive segmentation (Boykov and Jolly 2001, Freedman and Zhang 2005, Price, Morse and Cohen 2010, Isensee *et al* 2018). For example, GrabCut sought to maximize the separation of the foreground and background classes, as initially indicated by the user, by modelling these classes using Gaussian mixture models. (Rother *et al* 2004). The segmentation was then produced by applying the graph cut method from the user-provided annotation. The model and segmentation are then iteratively refined based on additional user feedback. The strength of this method is the speed of the graph cut segmentation allowing rapid feedback to the user such that the segmentation could be refined easily by providing additional annotation. On the downside, GrabCut often resulted in shrunken structures. This was prevented by adding a topological prior (Lempitsky *et al* 2009). Furthermore, the performance was improved further by including a Conditional Random Field (CRF) (Lempitsky *et al* 2009, Cheng *et al* 2015) to encourage segmentation homogeneity. Instead of bounding boxes, geodesic distance transforms have been introduced to ensure spatial regularization and contrast-sensitivity. This method is called GeoS (Criminisi *et al* 2008).

Another example of a framework for segmentation is Livewire, a contouring tool that quickly updates the contour based on user interaction. After beginning contouring, the optimal path between the starting point and the current cursor location is calculated using a lowest cost path algorithm (Dijkstra 1959). The contour changes as the user moves the mouse. Various algorithms may be used. For example, edge detection may be implemented using a Sobel filter. In that case the lowest cost path will be that along the edges. Livewire has been extended to the Intelligent Scissor tool (Barrett and Mortensen 1997). On-the-fly training enables the contour to be applied to the specific type of edge being traced rather than just to the strongest edge in the image area. Furthermore, the Intelligent Scissors tool automatically freezes unchanging segments and inserts additional seed points to increase contouring efficiency and the accuracy of the generated contours. The cost calculation can be adjusted to fit the specific needs of a given image modality. For example, ultrasound images are still difficult to process using fully automatic segmentation due to the presence of speckle and other artefacts—which are inherent to this imaging modality (Rackham *et al* 2013). Livewire was adapted for ultrasound images by introducing two sets of costs: first, feature asymmetry to improve edge localization, and second, a weak shape constraint cost to improve boundary selection in the presence of missing information or artefacts (Rackham *et al* 2013). As a result, the fuzzy boundaries in ultrasound images can be detected by identifying structural relevance rather than intensity gradients.

Other examples for non-deep learning based methods include, level-set segmentation (Qiu *et al* 2013), random walks and random forest (Grady *et al* 2005). For example, SlicSeg uses online random forests to segment fetal MRI images (Wang *et al* 2016). Additionally, SlicSeg allows for contour estimation in the remaining image volume and interactivity for refinement following an adapted GraphCut method.

Open-source tool kits have been created based on these methods, for example ITK-SNAP (Yushkevich *et al* 2016) and trainable WEKA (Waikato Environment for Knowledge Analysis) (Arganda-Carreras *et al* 2017). Both provide an intuitive graphical user interface and are able to process various image modalities and formats. ITK-SNAP uses active contour methods to produce segmentation given an initial user input in 2D, 3D and multi-modality medical images. The trainable WEKA allows the user to train a model on a given contoured image set. The framework uses and compares any available classifier to perform image segmentation based on pixel classification. Such semi-automatic and interactive tool kits can help minimize the time and effort required for manual contouring task.

Figure 9. 3D interactive segmentation. 2D segmentation methods can be easily extended to 3D by, for example, propagating the contour through the image. The predicted contours can be reviewed and individual slices edited until a satisfactory segmentation is achieved.

4.3. Deep learning methods

More recently, user interactions have been incorporated into CNNs. Unlike standard networks as discussed in section 2, extra input is provided to the model through user interaction. Rajchl *et al* introduced DeepCut, which uses a CRF to update the parameters of a CNN model to favour segmentation homogeneity—as has been done with GrabCut before. DeepIGeoS combines CNNs, CRFs and geodesic distance transforms (G Wang *et al* 2019b), bringing deep learning to the GeoS method. This approach uses two models. The first model is used to obtain an initial segmentation. The user can then interact with the result to identify misclassification. A second network is used to refine the result given the user interaction. When applied to placenta and brain tumour segmentation of fetal MRI images, DeepIGeoS reduced user time by 66% compared with segmentation using GeoS. DeepCut performs segmentation using a bounding box input by users (Rajchl *et al* 2017). To improve performance, Deep Extreme Cut uses extreme points (edges) of the structure as an input to the CNN (Maninis *et al* 2017). The CNN learns to match the segmentation to the extreme points of the object resulting in improved performance compared to bounding boxes. To further improve performance of this method, other deep neural network frameworks, such as recurrent neural networks, have been proposed (Zheng *et al* 2015).

Most deep learning based methods that allow for user interaction build on the popular U-Net architecture. For example, Amrhen *et al* proposed UI-Net for interactive segmentation (Amrehn *et al* 2017). In addition to allowing each image slice to be contoured, as in the standard U-Net, it also allows for user 'scribbles' to be included in the model as an input. These scribbles indicate which areas should and should not be included in the segmentation. During the segmentation process, the user can continue to provide input to achieve a precise segmentation result. This system has shown superior results compared to networks without the user input channel component when applied to liver lesion segmentation. This is attributed to consistent improvement in segmentation with each user interaction. To make better use of the provided scribbles, Lin *et al* proposed ScribbleSup (Lin *et al* 2016). Instead of using the scribbles directly or applying a geodesic distance transform, a graphical model propagates the information from the scribbles to the unmarked pixels, based on spatial constraints and appearance.

Open-source tool kits, similar to ITK-SNAP and trainable WEKA, also exist for deep learning approaches. For example, RIL-Contour (Philbrick *et al* 2019) (Radiology Informatics Laboratory) supports medical image annotation using fully automated deep-learning, semi-automated methods, and manual methods. It also enables workflows for continual learning on newly annotated data provided by multiple annotators.

4.4. Interactive methods in 3D

Clinicians frequently need to contour multiple slices, for example in 3D image sets. Manual segmentation of 3D images on a slice-by-slice basis can be very time consuming. A workflow for segmentation of 3D images is illustrated in figure 9. Some of the methods discussed in the section above have been extended into 3D. For example, Livewire in 2D uses two points to define a curve. In 3D, a user needs to indicate one or more closed curves (contours) and the algorithm finds the corresponding minimum cost surface (Grady *et al* 2005). Bounding box based approaches have been extended by either using individual bounding boxes on various 2D slices or by using a bounding cuboid (Redekop and Chernyavskiy 2021).

4.5. Contour propagation

If direct extension of a 2D to 3D contouring method is not possible, then there are a number of alternative approaches. These include automatically initializing the segmentation on the adjacent slice via a seed input and performing a segmentation on this slice. Alternatively, the contour can be iteratively propagated throughout consecutive images of a scan. China *et al* introduced a volumetric segmentation approach that relies on subsequent initialization of the segmentation on adjacent image slices in ultrasound volumes. A gradient vector flow based propagation technique was used to provide an initial segmentation for the next image slice, whereas iterative random walks were used to correct the contours in subsequent steps of the algorithm (China *et al* 2019). Contour propagation and interpolation has also been achieved by using slice-to-slice registration of contours (Penney *et al* 2004). For this the deformation between two image slices is calculated and that transformation is then applied to the contour. These propagation approaches can be applied in addition to any of the 2D methods mentioned before to produce 3D segmentation based on 2D input. Alternatively, if a few slices in a 3D image have been contoured, interpolation (for example linear interpolation) between these contours, can provide an estimate for the full 3D volume of a structure.

Using CNNs, contour propagation has been applied to bladder segmentation using a single contoured image slice as the input (Léger *et al* 2018). A similar propagation approach has been used for multi-class image segmentation of the cardiac system (Zheng *et al* 2018). These methods have been shown to offer better segmentation performance than fully automatic methods. However, they are only able to segment structures included in the training set and require retraining if used for other structures. This problem is often faced in supervised deep learning, as it is highly dependent on the training set provided. One solution to this is to train CNNs on a large variety of structures simultaneously. In this way the model does not learn structure-specific features, but rather learns to predict the adjacent slice based on the context between input image and contoured slice (Michael Trimpl *et al* 2021).
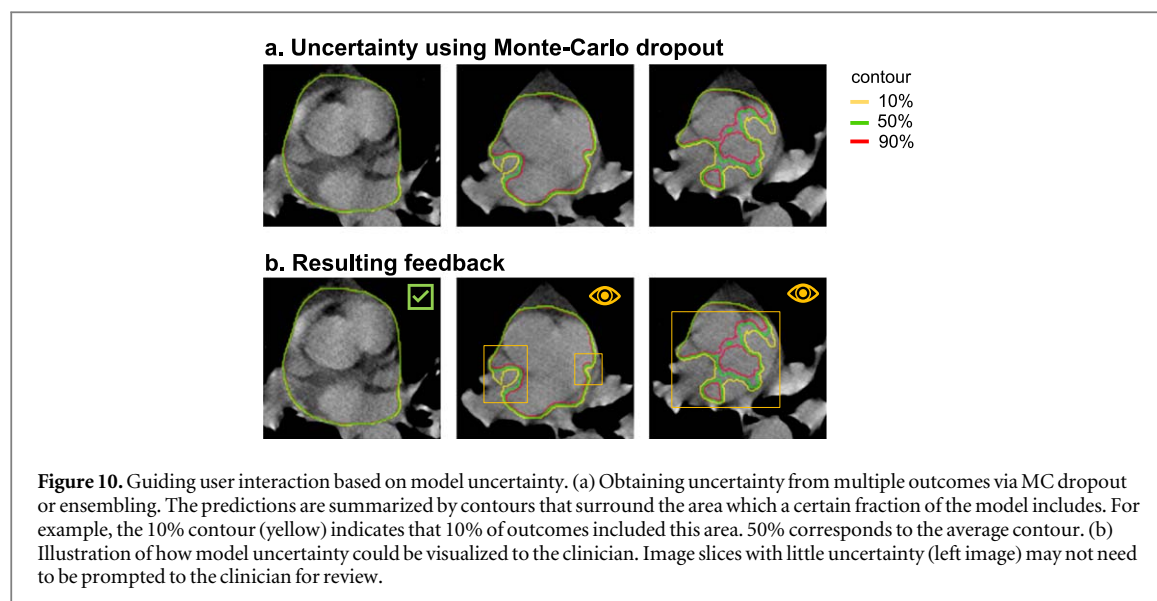
4.6. Other deep learning methods in 3D

Sparse user annotations, by providing selected 2D contours, can be used to predict the remaining slices of a 3D structure. A 3D U-Net (Ciçek *et al* 2016) has been proposed, that uses multiple sparse annotations in a 3D volume. This method can learn from just a few contoured slices of a 3D scan and complete the segmentation by using on-the-fly elastic deformations for efficient data augmentation. The image slices with a user-defined contour are used to fine-tune a deep learning model to segment the remaining non-contoured image slices of the specific scan. This approach in effect applies transfer learning on a case-by-case basis (as discussed in section 3) to enable an interactive contouring approach.

Again, as for 2D methods, user control on the segmentation output through image specific fine-tuning can be increased by providing scribbles to the network. These scribbles are used to create a weighted loss function. User-provided scribbles are associated with higher confidence than the other pixels and are therefore assigned heavier weighting. Additionally, during fine-tuning pixels with low confidence in the test image are given lower weighting. Model based uncertainty has been associated with the softmax output, where low confidence corresponds to a value close to 0.5 (Wang *et al* 2018). This is discussed further in the next section. Using scribble-based and model uncertainty-based loss function for fine-tuning has been shown to improve interactive segmentation performance in medical imaging for several structures, including placenta, brain, fetal lungs and maternal kidneys (Wang *et al* 2018). As an alternative to scribbles, seed points have been used to indicate the structure to be segmented into a 2D or 3D U-Net structure (Sakinis *et al* 2019, Pepe *et al* 2020).

The interactive 2D and 3D segmentation methods discussed in this section can synergize well with fully automatic segmentation methods. Fully automatic methods can produce a quick first estimate for a contour. Subsequently, interactive methods can be used to get the final increment in accuracy that is needed for medical image segmentation. Thus, interactive methods are an excellent tool during the contour review stage.

4.7. Guiding user interaction using ML feedback

Up to this point the review has covered how user interaction can be interpreted by a computer to assist segmentation. In this section, the possibility of a computer giving feedback to the clinician is discussed. As the performance of computer aided segmentation methods improves, the human role moves from one of active agent to a more supervisory role. In this capacity, the user's role is to check and edit the segmentation where required. Even when using interactive segmentation methods, this checking process can still be time-consuming since the image and segmentation must still be loaded and displayed, and reviewing requires the user's attention and interaction (Michael J Trimpl *et al* 2021). If a model could estimate on its own where it is certain and uncertain about a prediction, or better yet where it is accurate or likely to be inaccurate, a clinician might not have to review a full image scan but could focus on the few critical regions of the scan (Wang *et al* 2020a).

Figure 10. Guiding user interaction based on model uncertainty. (a) Obtaining uncertainty from multiple outcomes via MC dropout or ensembling. The predictions are summarized by contours that surround the area which a certain fraction of the model includes. For example, the 10% contour (yellow) indicates that 10% of outcomes included this area. 50% corresponds to the average contour. (b) Illustration of how model uncertainty could be visualized to the clinician. Image slices with little uncertainty (left image) may not need to be prompted to the clinician for review.

To effectively design a feedback system, the measure of uncertainty obtained from the model must correlate with the accuracy of the model. More generally, transparency in how the model came to make a prediction could help clinicians in their decision-making process. It has been suggested that model uncertainty could be used to estimate areas to edit for interactive segmentation (Zheng *et al* 2021). For example, an Uncertainty-Guided Refinement Framework has been applied to segmentation in motion-corrupted fetal MRI (Wang *et al* 2020a). In this framework, the clinician is asked to edit slices with the highest segmentation uncertainty following the automatic processing step. This uncertainty estimation was shown to correlate well with mis-segmentations. By guiding the clinician's attention, contouring time was reduced by 30% when compared to using the DeepIGeoS method, while achieving similar final accuracy (Wang *et al* 2020a). An example of this is shown in figure 10, where model uncertainty is determined for different image slices (figure 10(a)). Based on the model uncertainty selected image slice or image regions may be prompted to the clinician for review (figure 10(b)). Here, a brief overview of different approaches for determining uncertainty in deep learning prediction is given. It should be noted that determining the uncertainty of deep learning, is still a subject of ongoing research and a more extensive discussion may be found in (Abdar *et al* 2021).

The most straightforward way to estimate uncertainty is to use the final activation layer of a deep learning model as a proxy for confidence in a prediction. The activation layer (e.g. softmax) may output a value between 0 and 1 to predict if a pixel belongs to the background or foreground respectively. If the output is close to 0.5, the model is not sure to which class to assign the pixel. An uncertainty can then be attributed to the prediction using least confidence, marginal confidence or entropy (Shannon 1948, Settles 2009). However, these estimates on model uncertainty suffer from a tendency, common across deep learning methods, to be overconfident in the accuracy of the model predictions.

To estimate the uncertainty of a model, Gal *et al* introduced dropout in their neural networks as a Bayesian approximation to uncertainty (Gal and Ghahramani 2015). Using dropout means that different connections within the neural network are randomly dropped. This results in a slightly different model each time. Originally, dropout was used during model training to reduce overfitting and increase the robustness of the model (Srivastava *et al* 2014). Applying dropout at test time can be used to calculate the outcome of the resulting slightly different models. The resulting distribution of outcomes can then be used as a measure of uncertainty of the model predictions (Gal and Ghahramani 2015, Gal *et al* 2017, Wang *et al* 2019a).

Similarly, the uncertainty of a prediction can be estimated if an ensemble of deep learning models is used. Ensembling describes the method of using various models to make a prediction by consensus. Additionally, disagreement between the individual predictions can indicate a region of uncertainty (Settles 2009), illustrated in figure 10. A way to create an ensemble of models is by using convolutions in multiple groups (Wang *et al* 2020a). Thereby, multiple models following the same architecture can be trained in parallel. The disagreement in the prediction can then be used to obtain an uncertainty measure, which has been shown to be superior to Monte-Carlo dropout when estimating uncertainty (Wang *et al* 2020a). As above, determining uncertainty in a deep learning model remains a matter of ongoing research. However, focusing a clinician's attention on the most crucial areas and so guiding their user interaction can help make contouring more efficient and accurate (G Wang *et al* 2020a).

5. Discussion

Deep learning has become the state-of-the-art approach to medical image processing and the resulting tools are being adopted into the clinical workflow. However, the clinically implemented approaches have largely been restricted to fully automatic segmentation models. Yet, as outlined in this review, there are many approaches beyond fully automatic tools that could aid clinicians in performing their work.

Fully automatic methods are excellent when presented with a clearly defined task and where a large, labelled training data set is available. In such a scenario, these methods can potentially eliminate the need for editing by clinicians. However, while the task of segmentation might appear to be clearly defined, in reality it is difficult to find a gold standard contour for many structures. Clinicians may disagree on what should and shouldn't be included in a contour; indeed if an individual clinician is asked to outline the same case twice, they will likely produce two different sets of contours (Sharp *et al* 2014). A deep learning model trained by a single user or set of users will be biased towards the style of the clinician or centre providing the cases for the training set. Conversely, if the cases in the training set are representative of a wide range of institutions and clinicians, the model may average the different opinions in a way that satisfies no one's requirements. Furthermore, common deep learning methods require large amounts of data to be able to create robust models. The confidential nature of medical imaging data makes sharing and creating large datasets difficult. While there are more and more open access image datasets available (Yang *et al* 2017, Aerts *et al* 2019, Simpson *et al* 2019, Wee and Dekker 2019), they are restricted often to datasets that were acquired under certain conditions for a specific study. Consequently, fully automatic methods are very rigid and not always suitable for application to a general problem.

This review set out to explore possible alternatives to give users greater control over deep learning model output, whilst retaining efficiency. The first step to changing to a more user-controlled model is to provide a minimal training set for a deep learning model. The question: 'How little data can we get away with?' is essential here and this in itself is a large area of ongoing research in computer vision, covered by k-shot learning, transfer-learning and model fine tuning. These methods can produce segmentations requiring only a small number of annotated images of the target class. The performance of these methods strongly depends on the problem domain and annotations in the training/support set and the performance is typically lower than the performance of the fully supervised state of the art methods. Yet, this could give the clinician the option to—for example—define a completely new structure that a model should learn to segment. All that is needed for this is a few prior annotated cases—the support set. From these few cases an automatic contouring pipeline can be established that can then be applied on other cases. Overall, these methods take conventional deep learning and lower the barrier for model training by allowing for small data set sizes. By doing so, a support set can easily be defined by a single person that can than determine what (and how) they want a certain structure to be contoured, given them a greater amount of control on the output.

Despite the promising results of the k-shot learning application in the medical imaging domain reported in several studies (Fei-Fei *et al* 2003, 2006, Santoro *et al* 2016, Snell, Swersky and Zemel 2017, Shaban *et al* 2017, Dong and Xing 2018, Zhang *et al* 2018, Mondal, Dolz and Desrosiers 2018, Bucher *et al* 2019, Wang *et al* 2019c, Wang *et al* 2020c, Lu *et al* 2021), challenges remain. As these methods use one or few annotated images to learn how to produce the segmentation, they are more sensitive to the variation in the medical images than fully supervised methods. To overcome this problem, some methods use unlabelled images or/and GAN's to produce more training samples artificially. Another challenge for using the k-shot learning methods for medical imaging segmentation arises from the segmentation task itself. For example, segmentation of tumours is an arduous task even for fully supervised methods as tumours are very heterogeneous. Thus, tumour segmentation is very challenging for k-shot learning methods as they cannot capture tumour heterogeneity from a small number of samples.

Interactive methods for many applications can provide a good middle ground in terms of the trade-off between automation and user control. They are designed to interpret what the user wants to segment and suggest a segmentation based on that. The interactive methods discussed in this paper highlight the diverse ways that this can be done, ranging from bounding boxes and scribbles to clicks and individual contours. Regardless of the specific interaction type, these methods try to interpret the user interaction and complete the segmentation that the clinician requires. In the case of energy minimization techniques, user interaction is transformed into something the computer understands by including the user interaction as a constraint to the minimization problem. This problem is generally well understood and therefore it is possible to introduce the user interaction directly into the equation. For deep learning this becomes more complicated. The loss function, by which the model is optimized is not directly constrained by the user interaction. The loss function compares how well the prediction matches the ground truth of the training set. The parameters in the neural network are updated based on the discrepancy. With deep learning, the loss function is effectively optimized at training time, when the user interaction on a specific patient cannot be known. It remains a challenge to express the user interaction at interaction time as a clearly defined constraint, similarly as it is done in energy optimization. To make a more effective constraint when training a model, it is necessary to better understand deep learning models themselves.

Model interpretability is an ongoing research topic. A lack of understanding of how user interaction affects a model presents a challenge to its usability. If the clinician does not know what to expect as an output from any interaction, they cannot easily control the output. This review has briefly addressed how a model can communicate with clinician to guide its attention to the regions, user input is likely to be most needed. Such regions are where confidence in the prediction is low and consequently, the prediction is most likely to be incorrect.

While, to date, it remains challenging to determine uncertainty, there are other ways to improve understanding of a model and guide the clinician when contouring. Given the large number of parameters in a deep learning model it is hard to intuitively understand what is going on throughout a neural network and therefore, it is intrinsically difficult to interpret deep learning models (Reyes *et al* 2020). This understanding can be beneficial to be able to effectively use deep learning in a feedback loop, through which certain regions of a scan are brought to the attention of the clinician for review. What goes on in the model can be visualized by highlighting image-specific saliency maps - image regions that drive a model to its prediction (Selvaraju *et al* 2017).

Another way of making deep learning more accessible to human interpretation is by using 'attention' in neural networks (Oktay *et al* 2018, Schlemper *et al* 2019). Attention in a network with image input, refers to the regions in the image that receive a larger weighting due to their importance in making a prediction. They also give a more intuitive insight into what is going on inside the neural network. While convolutional filters and the resulting feature maps are very abstract, the deeper the network attention maps are easily interpreted as they simply highlight which regions of the body are weighted highly to make a prediction. While model uncertainty and interpretability are just beginning to be understood, they can be an avenue to a more fully interactive contouring system, where the model not only interprets user inputs but can guide future user interaction where it is most needed.

6. Concluding remarks

Research to date has focused on fully automatic segmentation methods. In practice, clinician review and editing of these outputs is needed. Interactive methods can support clinicians in these tasks. This review has outlined various methods that incorporate user interaction to make contouring for clinicians easier. In particular, the recent advances in integrating user interaction into deep learning have been highlighted. While this is often challenging, due to the often poor interpretability of deep learning models, these models can already leverage a clinician's expert input and provide support during the contouring workflow. Moving forward, it would be desirable to create interactive deep learning tools that can learn from previous user interactions. Future research should also focus on new methods to combine artificial intelligence and clinical expertise, instead of focusing on one or the other. A clinician can point out segmentation errors to the model, but the model may also communicate areas of high and low uncertainty to the clinician. By combining the strength of artificial intelligence and clinical expertise, patient care can be improved - by elevating contour consistency and quality, and by reducing the time taken on segmentation thus freeing up clinicians to focus on direct patient care.

Acknowledgments

ORCID iDs

Michael J Trimpl ● https://orcid.org/0000-0003-3393-0717
Eleanor P J Stride ● https://orcid.org/0000-0003-3371-5929
Mark J Gooding ● https://orcid.org/0000-0002-1177-5608

References

Abdar M *et al* 2021 A review of uncertainty quantification in deep learning: techniques, applications and challenges *Inform. Fusion* Elsevier **76** 243–97
Aerts H J W L *et al* 2019 Data from NSCLC-radiomics *The Cancer Imaging Archive* https://doi.org/10.7937/K9/TCIA.2015.PF0M9REI
Alemi Koohbanani N *et al* 2020 NuClick: a deep learning framework for interactive segmentation of microscopic images *Med. Image Anal.* **65**
Alom M Z *et al* 2018 Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation arXiv:1802.06955

Amrehn M, Gaube S, Unberath M, Schebesch F, Horz T, Strumia M, Steidl S, Kowarschik M and Maier A 2017 UI-Net: interactive artificial neural networks for iterative image segmentation based on a user model *Eurographics Workshop on Visual Computing for Biology and Medicine (2017)* pp 143–47

Anwar S M *et al* 2018 Medical image analysis using convolutional neural networks: a review *J. Med. Syst.* **42** 226

Arganda-Carreras I *et al* 2017 Trainable weka segmentation: a machine learning tool for microscopy pixel classification *Bioinform. Oxford Acad.* **33** 2424–6

Bakator M and Radosav D 2018 Deep learning and medical diagnosis: a review of literature *Multimodal Technol. Interact.* **2** 47

Barrett W A and Mortensen E N 1997 Interactive live-wire boundary extraction' *Med. Image Anal.* Elsevier **1** 331–41

Boers T G W, Hu Y, Gibson E, Barrat D C, Bonmatti E, Krdzalic J, van der Heijden F, Hermans J J and Huisman H J 2020 Interactive 3D U-net for the segmentation of the pancreas in computed tomography scans *Phys. Med. Biol.* **65** 065002

Boykov Y Y and Jolly M-P 2001 Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images *Int. Conf. Computer Vision (Vancouver, BC, Canada, 7–14 July 2001)* (Picastaway, NJ: IEEE) pp 105–12

Bucher M *et al* 2019 Zero-shot semantic segmentation arXiv:1906.00817 NeurIPS 2019 (accepted)

Camilus K S and Govindan V K 2012 A review on graph based segmentation *Graph. Signal Process.* **4** 1

Chen L-C, Zhu Y, Papandreou G, Schroff F and Adam H 2018 Encoder-decoder with atrous separable convolution for semantic image segmentation *Proc. of the European Conf. on Computer Vision* vol 11211 (*Lecture Notes in Computer Science*) (Cham: Springer) pp 833–51

Cheng M M, Prisacariu V A, Zheng S, Torr P H S and Rother C 2015 Densecut: densely connected crfs for realtime grabcut *Comput. Graph. Forum* **34** 193–201

China D *et al* 2019 Anatomical structure segmentation in ultrasound volumes using cross frame belief propagating iterative random walks *IEEE J. Biomed. Health Inform.* Institute of Electrical and Electronics Engineers Inc **23** 1110–8

Ciçek Ö *et al* 2016 3D U-net: learning dense volumetric segmentation from sparse annotation *MICCAI* 424–32 (http://lmb.informatik.uni-freiburg.de/resources/opensource/unet.en.html) (Accessed: 12 July 2020)

Criminisi A, Sharp T and Blake A 2008 GeoS: geodesic image segmentation *Computer Vision—ECCV 2008 10th European Conf. on Computer Vision* vol 5302, pp 99–112

Deng J *et al* 2010 Imagenet: a large-scale hierarchical image database *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Institute of Electrical and Electronics Engineers (IEEE)* pp 248–55

Dijkstra E W 1959 A note on two problems in connexion with graphs *Numer. Math. 1959 1:1*. Springer **1** 269–71

Ding Y *et al* 2019 A deep learning model to predict a diagnosis of alzheimer disease by using 18 F-FDG PET of the brain *Radiol. Radiol.* **290** 456–64

Dong N and Xing E P 2018 Few-shot semantic segmentation with prototype learning *British Machine Vision Conf.* http://bmvc2018.org/contents/papers/0255.pdf

Drozdzal M *et al* 2016 The importance of skip connections in biomedical image segmentation *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* (Cham: Springer) 10008 LNCS pp 179–87

Erickson B J *et al* 2017 Machine learning for medical imaging *Radiographics* **37** 505–15

Esteva A *et al* 2017 Dermatologist-level classification of skin cancer with deep neural networks *Nature* **542** 115–8

Fei-Fei L, Fergus R and Perona P 2003 A Bayesian approach to unsupervised one-shot learning of object categories *Proc. of the IEEE Int. Conf. on Computer Vision* 2, pp 1134–41

Fei-Fei L, Fergus R and Perona P 2006 One-shot learning of object categories *IEEE Trans. Pattern Anal. Mach. Intell.* **28** 594–611

Freedman D and Zhang T 2005 Interactive graph cut based segmentation with shape priors *Proc.—2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, CVPR 2005. IEEE Computer Society* I *(San Diego, CA, 20–25 June 2005)* (Picastaway, NJ: IEEE) pp 755–62

Gal Y and Ghahramani Z 2015 Dropout as a bayesian approximation: representing model uncertainty in deep learning *ICML* 16 pp 1050–9

Gal Y, Islam R and Ghahramani Z 2017 Deep bayesian active learning with image data arXiv:1703.02910

Gooding M *et al* 2018 PV-0531: multi-centre evaluation of atlas-based and deep learning contouring using a modified turing test *Radiother. Oncol.* **127** S282–3

Grady L *et al* 2005 Random walks for interactive organ segmentation in two and three dimensions: implementation and validation in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* (Berlin, Heidelberg: Springer) pp 773–80

Gulshan V *et al* 2019 Performance of a deep-learning algorithm versus manual grading for detecting diabetic retinopathy in india *JAMA Ophthalmol. Am. Med. Assoc.* **137** 987–93

Guo Z *et al* 2019 Deep learning-based image segmentation on multimodal medical imaging *IEEE Trans. Radiat. Plasma Med. Sci.* **3** 162–9

Hamidian S *et al* 2017 3D convolutional neural network for automatic detection of lung nodules in chest CT *Proc. SPIE 10134, Medical Imaging 2017: Computer-Aided Diagnosis. SPIE* vol 10134, pp 54–9

He K *et al* 2016 Deep residual learning for image recognition *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition* pp 770–8 (Accessed: 30 June 2020) (http://image-net.org/challenges/LSVRC/2015)

Hosny A *et al* 2018 Artificial intelligence in radiology *Nat. Rev. Cancer* **18** 500 NIH Public Access

Hubel D H and Wiesel T N 1959 Receptive fields of single neurones in the cat's striate cortex *J. Physiol.* Wiley-Blackwell **148** 574–91

Isensee F *et al* 2018 nnU-Net: self-adapting framework for U-Net-based medical image segmentation in *Med. Segmentation Decathlon Challenge* **2018**

Jarrett D *et al* 2019 Applications and limitations of machine learning in radiation oncology *Br. J. Radiol. Br. Inst. Radiol.* **92** 20190001

Kadam S and Vaidya V 2018 Review and analysis of zero, one and few shot learning approaches *Advances in Intellgent Systems and Computing* (Cham: Springer) vol 940, pp 100–12

Karimi D *et al* 2020 Critical assessment of transfer learning for medical image segmentation with fully convolutional neural networks Available at: (https://arxiv.org/abs/2006.00356v1) (Accessed: 2 January 2022)

Karimi D, Warfield S K and Gholipour A 2021 Transfer learning in medical image segmentation: New insights from analysis of the dynamics of model parameters and learned representations *Artif. Intell. Med.* Elsevier **116** 102078

Kim M *et al* 2019 Deep learning in medical imaging *Neurospine* **16** 657–68

Komodakis N, Tziritas G and Paragios N 2007 Fast, approximately optimal solutions for single and dynamic MRFs *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (https://doi.org/10.1109/CVPR.2007.383095)

Léger J *et al* 2018 Contour propagation in CT scans with convolutional neural networks *Int. Conf. on Advanced Concepts for Intelligent Vision Systems ACIVS 2018: Advanced Concepts for Intelligent Vision Systems* pp 380–91 Available at: (https://openreggui.org/) (Accessed: 13 November 2019) in

Lempitsky V *et al* 2009 Image segmentation with a bounding box prior *Proc. of the IEEE Int. Conf. on Computer Vision (Kyoto, Japan, 29 September–2 October, 2009)* (Picastaway, NJ: IEEE) pp 277–84

Li X *et al* 2018 H-denseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes *IEEE Trans. Med. Imaging* **37** 2663–74

Lin D *et al* 2016 Scribblesup: scribble-supervised convolutional networks for semantic segmentation *CVPR* 3159–67 (http://research.microsoft.com/en-us/um/people/jifdai/downloads/scribble sup) (Accessed: 12 July 2020)

Litjens G *et al* 2017 A survey on deep learning in medical image analysis *Med. Image Anal.* **42** 60–88

Liu A *et al* 2019 Applying the turing test to contouring: are machine-generated contours indistinguishable from human generated ones? *Int. J. Radiat. Oncol. Biol. Phys.* Elsevier **105** E136

Liu A W and Chan J H 2021 An evaluation of transfer learning with CheXNet on lung opacity detection in COVID-19 and pneumonia chest radiographs *2021 13th Int. Conf. on Information Technology and Electrical Engineering (ICITEE). IEEE (Chiang Mai, Thailand, 14–15 October, 2021)* (Picastway, NJ: IEEE) pp 137–42

Liu Z *et al* 2018 Rethinking the value of network pruning *7th Int. Conf. on Learning Representations, ICLR 2019. Int. Conf. on Learning Representations, ICLR* arXiv:1810.05270v2 (Accessed: 8 December 2021)

Long J, Shelhamer E and Darrell T 2015 Fully convolutional networks for semantic segmentation in *IEEE Conf. on Computer Vision and Pattern Recognition. IEEE Computer Society* pp 3431–40

Lu Y *et al* 2021 Contour transformer network for one-shot segmentation of anatomical structures *IEEE Trans. Med. Imaging* **40** 2672–84

Luo J H, Wu J and Lin W 2017 ThiNet: a filter level pruning method for deep neural network compression *Proc. of the IEEE Int. Conf. on Computer Vision* October 2017 (Institute of Electrical and Electronics Engineers Inc.) pp 5068–76

Lustberg T *et al* 2018 Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer *Radiotherapy and Oncology* vol 126 (Ireland: Elsevier) pp 312–7

Mackin D *et al* 2015 Measuring computed tomography scanner variability of radiomics features *Invest. Radiol.* **50** 757–65

Maninis K K *et al* 2017 Deep extreme cut: from extreme points to object segmentation *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. IEEE Computer Society* pp 616–25

Milletari F, Navab N and Ahmadi S A 2016 V-Net: fully convolutional neural networks for volumetric medical image segmentation *Proc. 2016 4th Int. Conf. on 3D Vision, 3DV 2016* (Institute of Electrical and Electronics Engineers Inc) pp 565–71

Mondal A K, Dolz J and Desrosiers C 2018 Few-shot 3D multi-modal medical image segmentation using generative adversarial learning arXiv:1810.12241v1 (Accessed: 8 December 2021)

Oktay O *et al* 2018 Attention U-Net: learning where to look for the pancreas *Medical Imaging with Deep Learning* (Amsterdam)

Olabarriaga S D and Smeulders A W M 2001 Interaction in the segmentation of medical images: a survey *Med. Image Anal.* **5** 127–42

Pan S J and Yang Q 2010 A survey on transfer learning *IEEE Trans. Knowl. Data Eng.* **22** 1345–59

Parisi G I *et al* 2019 Continual lifelong learning with neural networks: a review *Neural Netw.* **113** 54–71

Peng P and Wang J 2020 How to fine-tune deep neural networks in few-shot learning? arXiv:2012.00204v1 (Accessed: 7 January 2022)

Penney G P *et al* 2004 Registration-based interpolation *IEEE Trans. Med. Imaging* **23** 922–6

Pepe A *et al* 2020 IRIS: interactive real-time feedback image segmentation with deep learning *SPIE Med. Imaging* (https://doi.org/10.1117/12.2551354)

Philbrick K A *et al* 2019 RIL-contour: a medical imaging dataset annotation tool for and with deep learning *Journal of Digital Imaging* vol 32 (New York LLC: Springer) pp 571–81

Price B L, Morse B and Cohen S 2010 Geodesic graph cut for interactive image segmentation *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 3161–8

Primakov S *et al* 2022 Automated detection and segmentation of non-small cell lung cancer computed tomography images

Qiu W *et al* 2013 Three-dimensional prostate segmentation using level set with shape constraint based on rotational slices for 3D end-firing TRUS guided biopsy *Med. Phys.* **40**

Rackham T M *et al* 2013 Ultrasound image segmentation using feature asymmetry and shape guided live wire *Med. Imaging 2013: Image Process.* **8669** 86690P

Raghu M *et al* 2019 Transfusion: understanding transfer learning for medical imaging *Adv. Neural Inform. Process. Syst. Neural Inform. Process. Syst. found.* **32** Available at: (https://arxiv.org/abs/1902.07208v3) (Accessed: 8 December 2021)

Rajchl M *et al* 2017 DeepCut: object segmentation from bounding box annotations using convolutional neural networks *IEEE Trans. Med. Imaging* **36** 674–83

Rajpurkar P *et al* 2017 CheXNet: radiologist-level pneumonia detection on chest x-rays with deep learning Available at: (https://arxiv.org/abs/1711.05225v3) (Accessed: 8 December 2021)

Ramkumar A *et al* 2016 User interaction in semi-automatic segmentation of organs at risk: a case study in radiotherapy *J. Digit. Imaging* **29** 264–77

Redekop E and Chernyavskiy A 2021 Medical image segmentation with imperfect 3D bounding boxes *Lecture Notes Comput. Sci.* **13003** 193–200

Renard F *et al* 2020 Variability and reproducibility in deep learning for medical image segmentation *Sci. Rep.* **10** 1–16

Reyes M *et al* 2020 On the interpretability of artificial intelligence in radiology: challenges and opportunities *Radiol.: Artif. Intell.* **2** e190043

Ronneberger O, Fischer P and Brox T 2015 U-net: convolutional networks for biomedical image segmentation *Med. Image Comput. Comput.-Assist. Intervention* **9351** 234–41

Rother C, Kolmogorov V and Blake A 2004 GrabCut'- Interactive foreground extraction using iterated graph cuts in *ACM Trans. on Graphics – Proc. of ACM SIGGRAPH* vol 2004, pp 309–14

Sakinis T *et al* 2019 Interactive segmentation of medical images through fully convolutional neural networks arXiv:1903.08205

Santoro A *et al* 2016 Meta-learning with memory-augmented neural networks *Proc. of The 33rd Int. Conf. on Machine Learning. PMLR* pp 1842–50 Available at: (https://proceedings.mlr.press/v48/santoro16.html) (Accessed: 8 December 2021)

Schlemper J *et al* 2019 Attention gated networks: learning to leverage salient regions in medical images *Med. Image Anal.* **53** 197–207

Schmidhuber J 2015 Deep learning in neural networks: an overview *Neural Netw.* **61** 85–117

Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D and Batra D 2017 Grad-CAM: visual explanations from deep networks via gradient-based localization *2017 IEEE International Conference on Computer Vision (ICCV) (Venice, Italy, 22–29 October, 2017)* (Picastaway, NJ: IEEE) pp 618–26

Seo H *et al* 2020 Machine learning techniques for biomedical image segmentation: an overview of technical aspects and introduction to state-of-art applications *Med. Phys.* **47** e148–67

Settles B 2009 Active learning literature survey computer sciences technical report *Active learning literature survey computer sciences technical report Technical Report #1648* University of Wisconsin, Madison (https://research.cs.wisc.edu/techreports/2009/TR1648.pdf)

Shaban A *et al* 2017 One-shot learning for semantic segmentation *British Machine Vision Conf. 2017, BMVC 2017. BMVA Press* (https://doi.org/10.5244/C.31.167)

Shannon C E 1948 A mathematical theory of communication *Bell Syst. Tech. J.* (https://doi.org/10.1002/j.1538-7305.1948.tb01338.x)

Sharp G *et al* 2014 Vision 20/20: perspectives on automated image segmentation for radiotherapy *Med. Phys.* **41**

Shen D, Wu G and Suk H I 2017 Deep learning in medical image analysis *Annu. Rev.* **19** 221–48

Simpson A L *et al* 2019 A large annotated medical image dataset for the development and evaluation of segmentation algorithms

Snell J, Swersky K and Zemel T R 2017 Prototypical networks for few-shot learning arXiv:1703.05175

Srivastava N *et al* 2014 Dropout: a simple way to prevent neural networks from overfitting *J. Mach. Learn. Res.* **15** 1929–58

Suzuki K 2017 Overview of deep learning in medical imaging *Radiol. Phys. Technol.* **10** 257–73

Tajbakhsh N *et al* 2020 Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation *Med. Image Anal.* **63** 101693

Tian J *et al* 2021 Tumour segmentation *Radiomics and Its Clinical Application.* (Amsterdam: Elsevier) pp 1–18

Trimpl M J *et al* 2021 Interactive contouring through contextual deep learning *Med. Phys.* **48** 2951–9

Trimpl M J *et al* 2021 *PO-1164 - Clinical evaluation of an interactive deep-learning assisted contouring method for target contouring |* (ESTRO) *2021, ESTRO 2021.* Available at: (https://estro2021.estro.org/poster/media/po-1164-clinical-evaluation-interactive-deep-learning-assisted-contouring-method) (Accessed: 7 January 2022)

Van Opbroek A *et al* 2015 Transfer learning improves supervised image segmentation across imaging protocols *IEEE Trans. Med. Imaging* **34** 1018–30

Van Timmeren J E *et al* 2020 Radiomics in medical imaging—'how-to' guide and critical reflection *Insights Into Imaging* (Berlin: Springer) vol 11, pp 1–16

Wang F *et al* 2017a Residual attention network for image classification *Proc. 30th IEEE Conf. on Computer Vision and Pattern Recognition, CVPR 2017. Institute of Electrical and Electronics Engineers Inc* 6450–8 Available at: (http://arxiv.org/abs/1704.06904) (Accessed: 15 July 2020)

Wang G *et al* 2016 Slic-seg: a minimally interactive segmentation of the placenta from sparse and motion-corrupted fetal MRI in multiple views *Med. Image Anal.* **34** 137–47

Wang G *et al* 2018 Interactive medical image segmentation using deep learning with image-specific fine tuning *IEEE Trans. Med. Imaging* **37** 1562–73

Wang G *et al* 2019a Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks *Neurocomputing* (Amsterdam: Elsevier) vol 338, pp 34–45

Wang G *et al* 2020a Uncertainty-guided efficient interactive refinement of fetal brain segmentation from stacks of MRI slices *Lecture Notes Comput. Sci.* (https://doi.org/10.1007/978-3-030-59719-1_28)

Wang G *et al* 2019b DeepIGeoS: a deep interactive geodesic framework for medical image segmentation *IEEE Trans. Pattern Anal. Mach. Intell.* **41** 1559–72

Wang K *et al* 2017b Pay attention to features, transfer learn faster CNNs | *OpenReview, British Machine Vision Conf. (BMVC)* Available at: (https://openreview.net/forum?id=ryxyCeHtPB) (Accessed: 8 December 2021)

Wang K *et al* 2019c PANet: few-shot image semantic segmentation with prototype alignment *Proc. of the IEEE Int. Conf. on Computer Vision. Institute of Electrical and Electronics Engineers Inc* 2019, 9196–205

Wang R *et al* 2020b Medical image segmentation using deep learning: a survey Available at: (https://arxiv.org/abs/2009.13120v3) (Accessed: 2 January 2022)

Wang S *et al* 2020c LT-net: label transfer by learning reversible voxel-wise correspondence for one-shot medical image segmentation *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. IEEE Computer Society* 9159–68

Wang X *et al* 2017c Chest x-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases *Proc. 30th IEEE Conf. on Computer Vision and Pattern Recognition, CVPR 2017. Institute of Electrical and Electronics Engineers Inc., 2017-January* pp 3462–71

Wang Y *et al* 2020d Generalizing from a few examples *ACM Computing Surveys (CSUR).* (New York, NY, USA: ACM PUB27) vol 53 (https://doi.org/10.1145/3386252)

Wee L and Dekker A 2019 Data from head-neck-radiomics-HN1. *The Cancer Imaging Archive* (https://doi.org/10.7937/tcia.2019.8kap372n)

Wei J *et al* 2020 Genetic U-Net: automatically designed deep networks for retinal vessel segmentation using a genetic algorithm *IEEE Trans. Med. Imaging* (https://doi.org/10.1109/TMI.2021.3111679)

Willemink M J *et al* 2020 Preparing medical imaging data for machine learning *Radiology* **295** 4–15

Xu C and Prince J L 1998 Snakes, shapes, and gradient vector flow *IEEE Trans. Image Process.* **7** 359–69

Yang J *et al* 2017 Data from lung CT segmentation challenge *The Cancer Imaging Archive.* (https://doi.org/10.7937/K9/TCIA.2017.3r3fvz08)

Yushkevich P A, Gao Y and Gerig G 2016 ITK-SNAP: an interactive tool for semi-automatic segmentation of multi-modality biomedical images *Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Int. Conf.. Annu Int Conf IEEE Eng Med Biol Soc* vol 2016, pp 3342–5

Zhang X *et al* 2018 SG-one: similarity guidance network for one-shot semantic segmentation *IEEE Trans. Cybernetics* **50** 3855–65

Zhao B *et al* 2014 Exploring variability in CT characterization of tumors: a preliminary phantom study *Transl. Oncol.* **7** 88–93

Zhao F and Xie X 2013 An overview of interactive medical image segmentation *Annals of the BMVA* **2013** 1–22 (http://bmva.org/annals/2013/2013-0007.pdf)

Zheng E *et al* 2021 A continual learning framework for uncertainty-aware interactive image segmentation *AAAI Conference on Artificial Intelligence*

Zheng Q *et al* 2018 3D consistent robust segmentation of cardiac images by deep learning with spatial propagation *IEEE Trans. Med. Imaging* **37** 2137–48

Zheng S *et al* 2015 Conditional random fields as recurrent neural networks 1529–37

Zhou S *et al* 2016 Active contour model based on local and global intensity information for medical image segmentation *Neurocomputing* (The Netherlands: Elsevier Science Publishers B. V. PUB568 Amsterdam) vol 186, pp 107–18

Zhou Z *et al* 2018 UNet++: a Nested U-Net Architecture for Medical Image Segmentation *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, held in conjunction with MICCAI 2018 (Granada, Spain)* **vol 11045** p 3 Available at: (https://arxiv.org/pdf/1807.10165).pdf (Accessed: 16 May 2019)