ELSEVIER

Seminars in
RADIATION
ONCOLOGY

Advances in Auto-Segmentation

Carlos E. Cardenas, PhD, Jinzhong Yang, PhD, Brian M. Anderson, MS, Laurence E. Court, PhD, and Kristy B. Brock, PhD

Manual image segmentation is a time-consuming task routinely performed in radiotherapy to identify each patient's targets and anatomical structures. The efficacy and safety of the radiotherapy plan requires accurate segmentations as these regions of interest are generally used to optimize and assess the quality of the plan. However, reports have shown that this process can be subject to significant inter- and intraobserver variability. Furthermore, the quality of the radiotherapy treatment, and subsequent analyses (ie, radiomics, dosimetric), can be subject to the accuracy of these manual segmentations. Automatic segmentation (or auto-segmentation) of targets and normal tissues is, therefore, preferable as it would address these challenges. Previously, auto-segmentation techniques have been clustered into 3 generations of algorithms, with multiatlas based and hybrid techniques (third generation) being considered the state-of-the-art. More recently, however, the field of medical image segmentation has seen accelerated growth driven by advances in computer vision, particularly through the application of deep learning algorithms, suggesting we have entered the fourth generation of auto-segmentation algorithm development. In this paper, the authors review traditional (nondeep learning) algorithms particularly relevant for applications in radiotherapy. Concepts from deep learning are introduced focusing on convolutional neural networks and fully-convolutional networks which are generally used for segmentation tasks. Furthermore, the authors provide a summary of deep learning auto-segmentation radiotherapy applications reported in the literature. Lastly, considerations for clinical deployment (commissioning and QA) of auto-segmentation software are provided.
Semin Radiat Oncol 29:185–197

Introduction

Image segmentation is an important task routinely performed in radiotherapy (RT) to identify treatment targets and anatomical structures (organs-at-risk). In a typical clinical workflow, a radiation oncologist manually segments these regions of interest (ROI) on an RT simulation scan. Traditionally, computed tomography (CT) scans have been used for RT simulation; however, with the advent of magnetic resonance (MR)-guided RT, MR simulation scans are being adopted for RT planning in some clinics. The manual segmentation of these ROIs is a time consuming process with some studies reporting several hours of physician time per patient.[1-3] This could lead to significant delays in start of RT treatment, particularly in clinics with limited resources, which has been correlated to worse locoregional control and overall survival rates.[4,5] Furthermore, the significant time commitment required to segment each patient's ROIs has been regarded as a hindrance, or rate-limiting step, for adaptive RT, as it is necessary for the ROIs to be segmented on new imaging reflecting patient's anatomical changes to ensure accurate dose accumulation estimates for the RT treatment.

The efficacy and safety of the RT plan requires accurate segmentations as these regions of interest are generally used to optimize and assess the quality of the plan. However, inconsistencies in target and organs-at-risk segmentations have been reported for both inter- and intraobserver segmentation variability.[1,6-8] These inconsistencies may arise from the fact that the segmentation task is subjective in nature as the expert performing the segmentations evaluates the available imaging and then makes the decision, based on prior-knowledge and/or experience, of what voxels to include as part of the ROI being segmented. Subsequently, the inherent variability observed in manual segmentations could have a significant impact on quantitative[9-13] (eg radiomics) and dosimetric analyses.[1,14-16] Automatic segmentation (or auto-segmentation) is, therefore, preferable as it would address these challenges.

Auto-segmentation algorithms have to overcome several image-related problems to ensure accurate predictions.

Conflict of interest: Laurence E. Court is partially supported by the NCI and Varian.
Address reprint requests to Carlos E. Cardenas. E-mail: CECardenas@mdanderson.org

https://doi.org/10.1016/j.semradonc.2019.02.001
1053-4296/

First, medical images are subject to noise that can affect the intensity of a voxel. Secondly, tissues within a patient typically exhibit intensity nonuniformity, meaning that voxel intensities within a single tissue may gradually vary over the extent of the image. Lastly, medical images are reconstructed during acquisition to have a predefined voxel size which leads to partial volume averaging. Limited by a finite image resolution, voxels may contain more than one tissue such that the voxel intensity may not be representative of either tissue class. Furthermore, there are imaging modality-related challenges that may be specific to individual modalities. While MR scans provide exquisite soft-tissue contrast, image intensities tend to vary between acquisitions due to magnetic susceptibility artifacts. These problems, along with the large anatomical presentation and tissue distribution among different individuals in a population, suggest that some degree of uncertainty is expected for both manual and auto-segmentations.

The field of medical image auto-segmentation has rapidly evolved over the past 2 decades. Previously, auto-segmentation segmentation techniques have been grouped into first, second, and third generation algorithms, representing a new standard in algorithm development.[17] However, more recently, deep learning based auto-segmentation techniques have shown to provide significant improvements over more traditional approaches suggesting we have entered the fourth generation of auto-segmentation algorithm development (Fig. 1).

The field of deep learning became more mainstream after the seminal paper by Krizhevsky et al. (2012) showed that using a deep convolutional neural network architecture (AlexNet) could significantly improve predictions in image classification and recognition tasks.[18] In their work, the authors employed graphical processing units (GPU) to perform convolutional computations significantly reducing the time to train their model. Shortly after, research showed that using convolutional neural networks (CNN) for image segmentation tasks could outperform previously preferred algorithms, resulting in the swift adaptation of these architectures for medical image auto-segmentation.

This review provides a brief overview of traditional (pre-Deep Learning era) auto-segmentation techniques, introduces concepts behind deep learning-based auto-segmentation algorithms and commonly used architectures, and presents considerations for clinical implementation of auto-segmentation.

Traditional Auto-Segmentation Techniques

The development of auto-segmentation algorithms is alongside with the capability of the algorithms in using prior knowledge for new segmentation tasks. In an early stage, limited by the computer power and the availability of segmented data, most segmentation techniques used no or little prior-knowledge, referred to as low-level segmentation approaches. These include intensity thresholding, region growing, and heuristic edge detection algorithms.

More advanced techniques were developed in an attempt to avoid heuristic approaches leading to the introduction of uncertainty models and optimization methods. Region-based techniques, such as active contours, level-sets, graph cuts, and watershed algorithm, have been used in medical imaging auto-segmentation. Active contours and level-set algorithms are considered deformable models as they use closed surfaces that are able to contract or expand to conform to distinct image features within an image; whereas graph cuts and the watershed algorithm employ principles behind graph theory to maximize interconnections between image voxels. Probability-based auto-segmentation techniques, such as Gaussian mixture models, clustering, k-nearest neighbor, Bayesian classifiers, and shallow artificial neural networks, rose in popularity with the turn of the century thanks to advances in the statistical community and the availability of higher computing power. These approaches are characterized by their ability to classify individual voxels in an image as belonging to one of a known set of classes; however, these typically lack contextual information from neighboring voxels (as voxels are classified independently) which is often mediated by implementing hidden Markov random fields.[19]

In the last 2 decades, a large amount of exploratory work has been invested in making use of prior knowledge. An example is the use of shape and appearance characteristics of anatomical structures to compensate for insufficient soft tissue contrast of CT data, which prevents accurate definition of the anatomical boundary. Depending on how much prior knowledge is used in the algorithms, the approaches can be grouped as (multi)atlas-based segmentation, model-based segmentation, and machine learning-based segmentation.[20]

Single atlas based segmentation uses one reference image with structures of interest already segmented, referred to as atlas, as prior knowledge for new segmentation tasks.[21] The segmentation of a new image relies on deformable registration finding the optimal transformation between the atlas and the new image to map the atlas contours onto the new image. Varied deformable registration algorithms have been used for this purpose[22-27] and most of them are intensity-based algorithms in order to achieve full automation. The segmentation performance solely depends on the performance of deformable registration, which is influenced by the similarity of the morphology of organs of interest between atlas and the new image. To achieve good segmentation results, varied atlas selection strategies have been proposed.[28-34] Alternatively, using an atlas that reflects an average patient anatomy can potentially improve segmentation performance.[35,36]

Atlas-based segmentation is often impacted by intersubject variability. Instead of using a single atlas, multiatlas approaches use a number of atlases (usually around 10) as prior knowledge for new segmentation tasks.[37-42] Similar to single atlas-based approaches, deformable registration is the enabling technique to map individual atlas contours to the new image. An additional step, frequently referred as to label/contour fusion, is performed to combine the individual segmentations from multiple atlases to produce a final segmentation that is the best estimate of the true

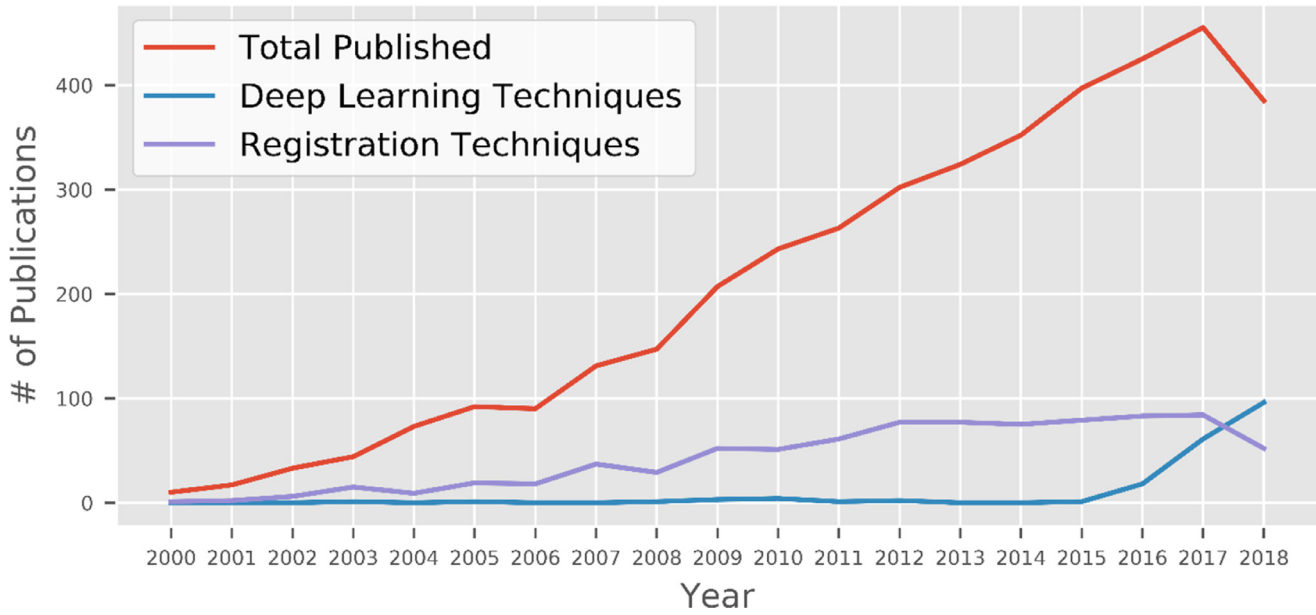

Figure 1 Number* of medical image segmentation peer-reviewed papers published since the turn of the century.

*Data was collected using Scopus with the following search criteria: Total Published – ((TITLE-ABS-KEY((segmentation OR delineation) AND (organ OR target OR tumor))) AND (ALL("automatic segmentation" OR "auto-segmentation" OR "auto-delineation" OR "automatic delineation" OR auto-segmentation OR auto-delineation OR "automated segmentation" OR "automated delineation")) AND(PUBYEAR > 1999) AND (ALL((ct OR mr OR "computed tomography" OR "magnetic resonance")))), Deep Learning Techniques – ((TITLE-ABS-KEY((segmentation OR delineation) AND ("deep learning" OR cnn OR convolution OR fcn) AND (organ OR target OR tumor))) AND(ALL("automatic segmentation" OR "auto-segmentation" OR "auto-delineation" OR "automatic delineation" OR auto-segmentation OR auto-delineation OR "automated segmentation" OR "automated delineation")) AND(PUBYEAR > 1999) AND(ALL((ct OR mr OR "computed tomography" OR "magnetic resonance")))), RegistrationTechniques – ((TITLE-ABS-KEY((segmentation OR delineation) AND (registration OR atlas-based OR atlas OR multi-atlas) AND (organ OR target OR tumor))) AND(ALL("automatic segmentation" OR "auto-segmentation" OR "auto-delineation" OR "automatic delineation" OR auto-segmentation OR autodelineation OR "automated segmentation" OR "automated delineation")) AND (PUBYEAR > 1999) AND(ALL((ct OR mr OR "computed tomography" OR "magnetic resonance")))).

segmentation.[34,43-46] Multiatlas segmentation has been shown to minimize the effects of intersubject variability and improve segmentation accuracy from single atlas approaches. In the past decade, multiatlas segmentation has been shown as one of the most effective segmentation approaches in several grand challenges.[47-49] This approach has been validated for clinical radiation oncology applications in contouring head and neck normal tissue,[50] cardiac substructures,[51] brachial plexus,[39] and among others. Commercial implementation of multiatlas segmentation is also available, eg Elekta ABAS (Elekta Oncology Systems, Crawley, UK) and RayStation (RaySearch Laboratories, Stockholm, Sweden) multiatlas segmentation.

When more contoured images are available, characteristic variations of shape or appearance of structures of interest could be used to train statistical shape models or statistical appearance models for auto-segmentation. These approaches can restrict the final segmentation results to anatomically plausible shapes described by the models.[52] However, model-based segmentation is less flexible due to the limitation of specific shapes characterized by the statistical models. Also, size and content of the training data limit the segmentation performance. In radiation oncology applications, model-based segmentation is mostly used for the segmentation of structures in the pelvic region.[53-55]

On the other hand, when more contoured images are available, machine learning approaches can aid in segmentation by learning appropriate priors for structures and organs or image context and tissue appearance for voxel classification.[56-58] Support vector machines[59-63] and tree ensemble (ie random forests)[64-70] algorithms have shown promising results in thoracic, abdominal, and pelvic tumor and normal tissue segmentation. These generally employ human-engineered features, usually derived from the image intensity histograms, from a large patient database as inputs to train the segmentation model.

Deep Learning Auto-Segmentation

Deep learning is part of the broader field of machine learning where algorithms are able to learn data representations on their own. More specifically, deep learning uses deep neural network architectures with multiple (2 or more) hidden layers (those between input and output layers) to learn features from a dataset by modeling complex nonlinear relationships. Using

deeper networks allows for improved generalization (prediction) on unseen data. Previously, deep architectures were prone to model overfitting; however, algorithmic advances over the past decade has allowed for the use of very deep architectures (100+ layers) to achieve "superhuman" performance in some tasks. Furthermore, the application of GPUs to speed up computations has fueled the field forward.

CNN are of particular interest in computer vision tasks (ie segmentation, detection, classification) as these learn the filters or kernels that were previously engineered for use in traditional approaches. These architectures are usually formed by stacking several types of layers (convolutional layers, pooling layers, fully-connected layers, etc.) that transform the input (image) into the desired output (Fig. 2). An important feature of convolutional layers is that they provide local connectivity between neurons of adjacent layers exploiting spatially local correlations. This allows the networks to learn features both globally and locally allowing the network to detect subtle variations in the input data.

Previously, CNNs allowed for the classification of each individual pixel in the image; however, this becomes computationally expensive as the same convolutions are computed several times due to the large overlap between input patches from neighboring pixels. Fully-convolutional networks (FCNs) were introduced by Long et al. to overcome the loss of spatial information resulting from the implementation of fully-connected layers as final layers of classification CNNs.[71] FCNs have both an encoding (as in traditional CNNs) and decoding paths (Fig. 2). The encoding path takes the input image and generates a high-dimensional feature vector which learns high layer (coarse) and low layer (high) feature representations of the input data. The decoding path replaced the fully-connected layers of CNNs to apply a learned up-sampling, through the pixelwise loss, to produce accurate segmentations.

FCNs are generally trained using supervised learning meaning that each input image has a corresponding labeled output. The learning process is made possible by the back-propagation algorithm[72] which can be summarized as follows: First, the data is passed through the network (forward-propagation) and an output prediction (inference) is made. Then a loss function is used as an error metric to compare

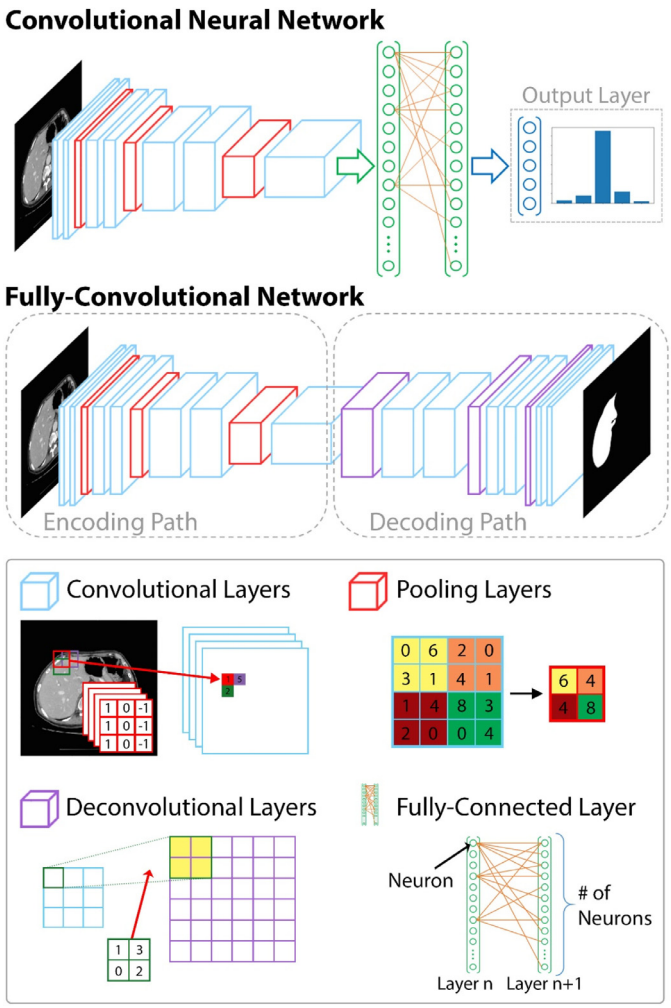

Figure 2 Illustration of convolutional neural network (CNN) and fully-convolutional network (FCN). These differ in that the FCN has a decoding path that brings the output of the network back to the original input size. Individual components (convolutional, pooling, deconvolutional, and fully-connected layers) are demonstrated in the bottom panel.

the predicted and labeled (known) output. The objective of the backpropagation algorithm is to minimize this loss. The backpropagation algorithm propagates the errors backwards through the network, calculating the error signal at each node, and updating each node's weights at each training iteration. Traditionally, a small portion of the input data, commonly referred as the batch size, is used at each iteration. These networks may require many epochs, or pass through the full training set, to converge on a combination of weights that produce a satisfactory level of accuracy.

Most FCNs used for medical image segmentation are based on 2D or 3D variants of successful methods adapted from computer vision. Improvements in 3D convolution computation efficiency and hardware, in particular the fast increase in available GPU memory, have promoted the extension of these methods to 3D imaging. Patch-wise architectures, those using 2D (N_x x $N_{y)}$ or 3D ($N_x \times N_y \times N_z$) patches centered around the voxels in an image, were introduced to address these bottlenecks. In this simple approach, patches extracted from the whole image, along with their corresponding label maps, are used to train the segmentation network. Several approaches (shift-and-stitch, fusion, etc.) are being used to combine individual patch segmentation probability maps to create dense outputs.[71,73,74] Some results have suggested that the performance of patch-wise architectures can be improved by using multiscale inputs (multiple inputs with different patch sizes) which provide the network with global and local context.[75,76]

It remains to be determined what the minimum number of patient scans is needed to produce clinically acceptable results; however, one could argue that a robust dataset (one that includes large variability in patient anatomy approximating population variability) would suffice for this task (it has been shown that reasonable results can be achieved with as little as 12 patient scans[77]). In addition, determining this number of scans could be task dependent as some anatomies may be easier to identify than others (ie lungs vs esophagus). Furthermore, it should be noted that the quality of the data (both the images and segmentations) could be closely correlated with algorithm results; therefore, one should visually inspect patient data prior to training a model. The use of different contouring guidelines between institutions could have significant impact on the performance of an algorithm when tested on a new dataset.

Unfortunately, there are very few publicly-available datasets to train and test these algorithms. Due to this limitation, it is very important to follow proper resampling methods to prevent over-fitting on a test dataset.[78] While bootstrap and k-fold cross-validation are usually the preferred methods used in statistical learning to optimize model parameters during training,[79] these are computationally expensive and therefore not commonly used for medical image segmentation. The hold-out method is the most commonly used method for image segmentation. In this approach the data is divided into training, cross-validation, and test sets. It is well-understood that the training set performance tends to overestimate the test set accuracy, therefore it is advantageous to use a cross-validation set to evaluate the trained model's performance during hyper-parameter search and model optimization. This prevents from "peeking" into the test set data, which should only be used after the final model is determined. Evaluating the training model's performance on the test set prior to identifying the final model could introduce bias and result in poor generalization on unseen datasets. There are many possible ways to split the original dataset (ie 80/10/10 meaning 80% training, 10% cross-validation, and 10% test) and this task could be dependent on data availability. Using a 56/24/20 split (20% test, and 70% train, and 30% cross-validation from the remaining 80%) has been shown to produce good generalization from the cross-validation to test set accuracy.[80]

The literature on architectures used for medical image segmentation is already very broad with many applications being investigated for the majority of anatomical regions (thoracic, abdominal, pelvic, head-and-neck, brain, etc.) and across different imaging modalities (CT, MRI, FDG-PET, etc.). A comprehensive summary of deep learning auto-segmentation radiotherapy applications by anatomical region is provided in Table 1.

Popular Architectures Used in Medical Imaging Auto-Segmentation

The most popular medical image segmentation FCN architecture is the U-net.[81] While previous works had already proposed the use of encoding and decoding paths to create dense outputs, Ronneberger et al[81] combined this approach with skip-connections, which concatenate features from the encoding to the decoding layers, to provide the architecture with higher resolution features from the encoding path with the up-sampled features from the decoding path to better localize and learn representations from the input image. Furthermore, the U-net allowed for efficient end-to-end training, meaning that it did not require a pretrained network as others had previously proposed, and showed that the network could be trained to produce accurate segmentations with very little labeled training data.[81] The original 2D application of the U-net was extended by Cicek et al[82] to allow the use of 3D images to train this network. Other groups have introduced variants of the original U-net architecture. Milletari et al[83] proposed the V-net, a 3D version of the U-net architecture that introduced the use of a Dice coefficient loss function and implemented residual learning[84] at each resolution stage. Kamnitsas and collaborators[75] introduced the DeepMedic architecture which used multiscale 3D CNNs with fully-connected conditional random fields[85] for brain lesion segmentation. Their dual pathway architecture provided the network with local and more global context from the input images by using image patches at multiple scales simultaneously. Currently, there are two open-source platforms, Deep Learning Tool Kit[86] and NiftyNet,[87] which provide code to train and test these architectures.

Table 1 Summary of Deep Learning Auto-Segmentation Applications by Anatomical Region

	Architecture Type	Image Modality	Test Cases	Results per ROI	
Thoracic					
	2.5D FCN	CT	10	Aorta (DSC = 0.77), Bladder (DSC = 0.84), Esophagus (DSC = 0.43), Gallbladder (DSC = 0.65), Heart (DSC = 0.93), Inferior Vena Cava (DSC = 0.68), Left Lung (DSC = 0.94), Left Kidney (DSC = 0.91), Liver (DSC = 0.95), Pancreas (DSC = 0.62), Prostate (DSC = 0.52), Right Lung (DSC = 0.93), Right Kidney (DSC = 0.92), Spleen (DSC = 0.92), Stomach and Duodenum Lumen (DSC = 0.76), Uterus (DSC = 0.17)	96
	2D FCN	MR	600	Left ventricle cavity (DSC = 0.94, MSD = 1.04) Left ventricle myocardium (DSC = 0.88, MSD = 1.14) Right ventricle cavity (DSC = 0.90, MSD = 1.78)	97
	2D U-net	MR	66	Breast (DSC = 0.944, MSD = 2.9) Fibroglandular tissue (DSC = 0.811, MSD = 3.1)	98
	3D FCN	CT	30	Esophagus (DSC = 0.76, HD = 10.68)	99
	2D FCN + CRF	CT	30 (CV)	Aorta (DSC = 0.89), Esophagus (DSC = 0.69), Heart (DSC = 0.90), Trachea (DSC = 0.87)	100
	2.5D CNN	CT	893[†]	Lung Node (DSC = 0.78, MSD = 0.24)	101
	2D CNN	CT	20	Left and Right Lungs (DSC∼0.98[+]), Esophagus (DSC∼0.72[+]), Heart (DSC∼0.89[+]), Mediastinum (DSC∼0.93[+]), Spinal Cord (DSC∼0.82[+])	102
	2D FCN	CT	800 (CV)	Breast cancer CTV (DSC = 0.91, HD = 10.5)	103
Abdomen					
	2D FCN	CT	43	Liver (DSC = 0.89)	104
	3D FCN	CT	32	Liver (DSC = 0.97, MSD = 0.84)	105
	2.5D CNN	CT	73	Portal Vein (DSC = 0.7, MSD = 2.94)	106
	2.5D CNN & 3D FCN	CT	36	Aorta (DSC = 0.77), Bladder (DSC = 0.70), Esophagus (DSC = 0.12), Gallbladder (DSC = 0.65), Heart (DSC = 0.91), Inferior Vena Cava (DSC = 0.58), Left Lung (DSC = 0.95), Left Kidney (DSC = 0.96), Liver (DSC = 0.94), Pancreas (DSC = 0.53), Prostate (DSC = 0.47), Right Lung (DSC = 0.96), Right Kidney (DSC = 0.87), Spleen (DSC = 0.87), Stomach and Duodenum Lumen (DSC = 0.59), Uterus (DSC = 0.34)	107
	2D FCN	CT	12	Liver (DSC = 0.90), Right Kidney(DSC = 0.89), Left Kidney (DSC = 0.89), Spleen (DSC = 0.89)	77
	2D CNN	CT	90[†]	Liver Lesions (DSC = 0.77, MSD = 1.6)	108
	2D FCN	CT	70	Liver (DSC = 0.96, MSD = 1.1) Liver Lesion (DSC = 0.66, MSD = 1.2)	109
	3D FCN	CT	140	Liver (DSC = 0.96, MSD = 1.3) Spleen (DSC = 0.94, MSD = 1.2) Kidneys (DSC = 0.95, MSD = 1.0)	110
	3D FCN	CT	30	Liver (MSD = 0.91)	111
	2D/3D FCN	CT/MR	78	CT – Liver (DSC = 0.94, MSD = 1.5) MR – Liver (DSC = 0.91, MSD = 5.3)	112
	2D CNN	CT	100 (CV)	Liver (DSC = 0.97, MSD = 1.77)	113
	3D CNN	MR	10	Bowel (DSC = 0.866, HD = 5.9), Duodenum (DSC = 0.655, HD = 7.99), Liver (DSC = 0.953, HD = 5.41), Kidneys (DSC = 0.931, HD = 6.23), Stomach (DSC = 0.85, HD = 6.88)	114

(*continued on next page*)

Table 1 (Continued)

	Architecture Type	Image Modality	Test Cases	Results per ROI	
	3D V-net	CT	90 (CV)	Duodenum (DSC = 0.63, MSD = 4.1), Esophagus (DSC = 0.71, MSD = 1.7), Gallbladder (DSC = 0.73, MSD = 1.6), Left Kidney (DSC = 0.93, MSD = 0.9), Liver (DSC = 0.95, MSD = 1.6), Pancreas (DSC = 0.75, MSD = 1.9), Spleen (DSC = 0.95, MSD = 0.8), Stomach (DSC = 0.87, MSD = 2.5)	115
	2D FCN	MR	9	Left Kidney (DSC = 0.73), Liver (DSC = 0.91), Right Kidney (DSC = 0.78), Spleen (DSC = 0.93), Stomach (DSC = 0.56)	116
	2D CNN	CT	10	Liver (DSC = 0.97,MSD = 1.48)	117
	3D FCN	CT	150	Liver (DSC = 0.95) Spleen (DSC = 0.93) Pancreas (DSC = 0.82)	76
	2.5D CNN	CT	20	Pancreas (DSC = 0.81, MSD = 0.42)	118
Pelvis					
	3D V-net	MR	30	Prostate (DSC = 0.87, HD = 5.7)	119
	2D CNN	CT	126[†]	Bladder tumor (DSC = 0.53, MSD = 4.7)	120
	AE	MR	21	Prostate (Sensitivity = 91.51%, Specificity = 88.47%)	74
	AE + DM	MR	66	Prostate (DSC = 0.878, MSD = 1.59)	121
	2D CNN	MR	250	Prostate (DSC = 0.8977, MSD = 0.16 tmm)	122
	2D FCN	CT	60	Bladder (DSC = 0.93), Colon (DSC = 0.62), Intestine (DSC = 0.65), Left femoral head (DSC = 0.92), Right femoral head (DSC = 0.92), Rectal CTV (DSC = 0.88)	123
	2D FCN	MR	70	Rectal Tumor (DSC = 0.69)	124
	3D CNN	MR	30	Prostate(DSC = 0.89)	125
	3D CNN	MR	26	Prostate(DSC = 0.88)	126
Head and Neck					
	2.5D CNN	CT	50 (CV)	Eye Left (DSC = 0.88), Eye Right (DSC = 0.88), Larynx (DSC = 0.86), Mandible (DSC = 0.99), Optic_Nerve_L (DSC = 0.639), Optic_Nerve_R (DSC = 0.645), Optic_Chiasm (DSC = 0.374), Parotid_L (DSC = 0.766), Parotid_R (DSC = 0.78), Pharynx (DSC = 0.693), Submandibular_L (DSC = 0.697), Submandibular_R (DSC = 0.730), Spinal Cord (DSC = 0.87)	127
	2D CNN	CT	46	Nasopharynx CTV (DSC = 0.83, HD = 6.9) Nasopharynx GTV-primary (DSC = 0.81, HD = 5.1) Nasopharynx GTV-nodal (DSC = 0.62, HD = 25.8)	128
	Multi-scale 3D CNN	CT	48	Chiasm(DSC = 0.58, 95HD = 2.81) Optic_nerve_L(DSC = 0.72, 95HD = 2.3) Optic_nerve_R(DSC = 0.70, 95HD = 2.1)	129
	3D FCN	CT	15	Parotids (DSC[*median*] = 0.87)	130
	3D FCN	CT	20	Brain_Stem (DSC = 0.92, MSD = 0.84), Cochlea_L (DSC = 0.75, MSD = 0.34), Cochlea_R (DSC = 0.73, MSD = 0.41), Esophagus_Upper (DSC = 0.35, MSD = 7.7), Lyranx_Glottic (DSC = 0.39, MSD = 2.4), Larynx_Supraglottic (DSC = 0.71, MSD = 2.2), Mandible (DSC = 0.96, MSD = 0.60), Oral_Cavity (DSC = 0.84, MSD = 10.1), Parotid_L (DSC = 0.86, MSD = 1.4), Parotid_R (DSC = 0.90, MSD = 1.1), Pharynx_Cons_Inf (DSC = 0.58, MSD = 2.0), Pharynx_Cons_Mid (DSC = 0.61, MSD = 2.0), Pharynx_Cons_Sup (DSC = 0.46, MSD = 2.1), Submandibular_L (DSC = 0.79, MSD = 1.5), Submandibular_R (DSC = 0.88, MSD = 0.83), Spinal_Cord (DSC = 0.96, MSD = 0.39)	131

(*continued on next page*)

Table 1 (Continued)

	Architecture Type	Image Modality	Test Cases	Results per ROI	
	3D CNN	CT	46	Brain_Stem (DSC = 0.87), Chiasm (DSC = 0.53), Mandible (DSC = 0.93), Optic_Nerve_L (DSC = 0.72), Optic_Nerve_R (DSC = 0.71), Parotid_L (DSC = 0.88), Parotid_R(DSC = 0.87), Submandibular_L (DSC = 0.81), Submandibular_R (DSC = 0.81)	132
	3D FCN	CT	75	Brain (DSC = 0.99), Brain_Stem (DSC = 0.88), Cochlea_L (DSC = 0.63), Cochlea_R (DSC = 0.75), Lacrimal_L (DSC = 0.69), Lacrimal_R (DSC = 0.70), Lens_L (DSC = 0.81), Lens_R(DSC = 0.80), Lung_L (DSC = 0.99), Lung_R (DSC = 0.99), Mandible (DSC = 0.96), Optic_nerve_L (DSC = 0.76), Optic_nerve_R (DSC = 0.77), Oribit_L (DSC = 0.95), Orbit_R (DSC = 0.95), Parotid_L (DSC = 0.85), Parotid_R (DSC = 0.85), Spinal_Canal (DSC = 0.95), Spinal_Cord (DSC = 0.88), Submandibular_L (DSC = 0.85), Submandibular_R (DSC=0.85)	133
	AE	CT	52 (CV)	Oropharyngeal Cancer high-risk CTV (DSC = 0.81, MSD = 2.9)	134
	3D U-net	CT	75	Oropharyngeal Cancer low-risk CTV (DSC = 0.82, MSD = 3.3)	135
Brain					
	3D CNN	MRI	220 (CV)	Tumor (DSC = 0.90)	75
	2D CNN	MRI	460 (CV)	Tumor_Core (DSC = 0.71)*	136
				Tumor (DSC = 0.82)	137
	2D CNN	MRI	44	Tumor (DSC = 0.72)*	138
	2D CNN	MRI	244	Tumor (DSC = 0.86)*	139

DSC, dice similarity coefficient, MSD, mean surface distance (in mm), HD, Hausdorff distance (in mm).

† Number of lung/liver/bladder tumor/nodules.

+ Mean value estimated from figure.

* Metric's average over multiple dataset's results.

Progress Through Publicly Available Image Datasets

It could be argued that the availability of publicly available datasets with ground-truth labeled segmentations has promoted advances in deep learning segmentation algorithms. These datasets are generally published as part of "grand challenges" that are usually hosted by organizations such as the American Association of Physicists in Medicine, the Medical Image Computing and Computer-Assisted Intervention Society, and the International Society for Optics and Photonics. These segmentation challenges allow participants to evaluate their algorithm's performance on a common benchmark image dataset. Recent challenges have reported on the superior segmentation performance of deep learning based algorithms when compared to previous generation techniques.[49,88] Yang and colleagues reported on the results from the American Association of Physicists in Medicine 2017 Thoracic segmentation challenge.[49] In this challenge, the top-3 performing algorithms were deep learning based (4th place was multiatlas based). While deep learning and atlas-based algorithms performed similarly for some normal tissues, there was a significant improvement in esophagus auto-segmentation when using deep learning approaches. Bernard et all showed how deep learning auto-segmentations in the "Automatic Cardiac Diagnosis Challenge" produced overlap and distance metrics that were well within the inter- and intraobserver variability scores suggesting that the state-of-the-art deep learning techniques' results may reach a plateau in performance due to the inherent variability observed in the input manual segmentations.[88]

Segmentation Software Commissioning and Quality Assurance

When contours are used in the radiotherapy process, any errors in the segmentation can have a serious impact on the patient treatment. Depending on the location and extent of the error, normal tissues (eg cord) could receive unintended doses, or targets could be under-treated. Thus, it is important to perform appropriate commissioning of the system, routine procedural maintenance, and patient-specific verification of the auto-segmentation.

The commissioning process involves testing of the functions of a given piece of software and documentation of its different capabilities. The most obvious test for segmentation software is an evaluation of the accuracy of the segmentation, probably by comparison with manually drawn contours using overlap and distance metrics. Examples for many anatomical sites have been reported above. Commissioning for clinical use, however, involves a more comprehensive evaluation than these analyses. The commissioning process should include extensive testing with patient data from the local institution, to ensure that the software works as expected for their range of image types, patient anatomies, etc. Additionally, it is important to ensure that segmentations created within one software tool are exported/imported properly to other systems, with all segmentations information being transferred consistently and accurately to the treatment planning system. If the segmentation does not work sufficiently accurately or reliably for any of these combinations, then this limitation should be clearly documented so that the users are aware and vendors can address these issues.

Once the commissioning process is complete, and the user has verified that the segmentation software works sufficiently well in their environment and any limitations have been identified, the software can then be released for clinical use. Some routine maintenance is necessary after that to ensure that the software continues to perform in a consistent manner. The focus on quality is, however, now performed on an individual patient-by-patient basis. All segmentations should be carefully reviewed and approved by the local clinical staff (eg radiation oncologists) before use in a treatment plan. During the initial stages of deployment, the output of the automatic segmentation software should be treated as if a trainee had performed the contouring — that is, it is probably a reasonable starting point, but careful review is essential. The benefits of peer-review assessment through quality assurance contouring rounds have been previously reported,[89-91] and establishing similar practices to assess auto-segmentation results, even for algorithms that have been shown to give excellent segmentations, could ensure the overall safety of the radiotherapy treatment.

Automatic quality assurance of auto-segmentations has also been investigated.[92-94] These measure ROI specific characteristics (centroid, volume, shape, etc.) and use statistical approaches to determine any large deviations in segmented volumes. Another suggested approach, for example, could use the results of a primary segmentation algorithm and compare these to a secondary verification algorithm.[95] This approach requires the two algorithms to be independent, as the assumption is that they will fail in different ways. Although this approach does not replace the need for careful review of contours by the attending physician, it may help flag cases that require extra attention.

While deep learning auto-segmentation techniques have been shown to be very promising, comprehensive commissioning and QA of deep learning auto-segmentation software — or any other auto-segmentation technique — is critical prior and during clinical deployment to ensure patient safety. Furthermore, we recommend any auto-segmentation software to be used as a decision support tool which requires expert visual inspection and approval during clinical deployment.

Current Limitations of Deep Learning Auto-Segmentation

There are several important limitations to deep learning auto-segmentation. One significant limitation is that deep learning algorithms provide very little interpretability

(black-box algorithm) to understand how and which features (anatomical and/or image intensity-based) affected the trained network during segmentation prediction. This hinders the ability to fully understand and identify the cause behind inaccurate segmentations. Many data-related challenges are presented for deep learning applications, especially the requirement of high-quality segmented datasets. Deep learning approaches depend on the quality of the segmentations (prior-knowledge) used to train a model. This limitation could be addressed through standardization of manual contours via the adoption of established international consensus guidelines. A reduction in inter- and intraobserver contouring variability could further improve the prediction accuracy of an existing model. Lastly, variations in image acquisition protocols could potentially affect the performance of a deep learning algorithm; however, this limitation is applicable to all auto-segmentation techniques.

Conclusion

Deep learning auto-segmentation algorithms have quickly become the state-of-the-art in medical image segmentation suggesting we have entered the fourth generation of algorithm development. These algorithms have been applied to auto-segment targets and normal tissues in many anatomical sites including the thorax, abdomen, pelvis, head and neck, and brain with some applications producing better results than the measured inter- and intraobserver contouring variability. Over the next few years we expect increased availability (commercial and open-source) of deep learning-based auto-segmentation tools for radiotherapy treatment planning, as well as increased acceptance and implementation of auto-segmentation tools in clinical practice. While these tools have been shown to be very promising, commissioning and periodic QA of these systems should be performed to ensure patient safety.

References

1. Hong TS, Tome WA, Harari PM: Heterogeneity in head and neck IMRT target design and clinical practice. Radiother Oncol 103:92-98, 2012
2. Multi-Institutional Target Delineation in Oncology. Human – Computer Interaction in Radiotherapy Target Volume Delineation : A Prospective, Multi-Institutional Comparison of User Input Devices, D: 794-803, 2011
3. Harari PM, Song S, Tomé WA: Emphasizing conformal avoidance versus target definition for IMRT planning in head-and-neck cancer. Int J Radiat Oncol Biol Phys 77:950-958, 2010
4. Chen Z, King W, Pearcey R, et al: The relationship between waiting time for radiotherapy and clinical outcomes: A systematic review of the literature. Radiother Oncol 87:3-16, 2008
5. Stefoski Mikeljevic J, Haward R, Johnston C, et al: Trends in postoperative radiotherapy delay and the effect on survival in breast cancer patients treated with conservation surgery. Br J Cancer 90:1343-1348, 2004
6. Li XA, Ph D, Tai A, et al: Variability of target and normal structure delineation for breast- cancer radiotherapy: A RTOG multi-institutional and multi-observer study. Int J Radiat Oncol Biol Phys 73:944-951, 2009
7. Eminowicz G, McCormack M: Variability of clinical target volume delineation for definitive radiotherapy in cervix cancer. Radiother Oncol 117:542-547, 2015
8. Ng SP, Dyer BA, Kalpathy-Cramer J, et al: A prospective in silico analysis of interdisciplinary and interobserver spatial variability in postoperative target delineation of high-risk oral cavity cancers: Does physician specialty matter? Clin Transl Radiat Oncol 12:40-46, 2018
9. Owens CA, Peterson CB, Tang C, et al: Lung Tumor Segmentation Methods : Impact on the Uncertainty of Radiomics Features for Non-Small Cell Lung Cancer 1-23, 2018
10. Parmar C, Velazquez ER, Leijenaar R, et al: Robust radiomics feature quantification using semiautomatic volumetric segmentation. PLoS One 9:1-8, 2014
11. Balagurunathan Y, Gu Y, Wang H, et al: Reproducibility and prognosis of quantitative features extracted from CT images. Transl Oncol 7: 72-87, 2014
12. Lee M, Woo B, Kuo MD, et al: Quality of radiomic features in glioblastoma multiforme: Impact of semi-automated tumor segmentation software. Korean J Radiol 18:498-509, 2017
13. Kalpathy-cramer J, Mamomov A, Zhao B, et al: Radiomics of lung nodules: A multi-institutional study of robustness and agreement of quantitative imaging features. Tomography 2:430-437, 2016
14. Rasch C, Steenbakkers R, Van Herk M: Target definition in prostate, head, and neck. Semin Radiat Oncol 15:136-145, 2005
15. Weiss E, Hess CF: The Impact of Gross Tumor Volume (GTV) and Clinical Target Volume (CTV) Definition on the Total Accuracy in Radiotherapy Theoretical Aspects and Practical Experiences, 1;21-30, 2003
16. Saarnak AE, Boersma M, Van Bunningen BNFM, et al: Inter-observer variation in delineation of bladder and rectum contours for brachytherapy of cervical cancer. Radiother Oncol 56:37-42, 2000
17. Withey DJ, Koles XJ: Medical image segmentation: Methods and software. In: Proc. NFSI ICFBI 2007;140-143, 2017
18. Krizhevsky A, Sutskever I, Hinton GE: ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25:1-9, 2012
19. Zhang Y, Brady M, Smith S: Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. IEEE Trans Med Imaging 20:45-57, 2001
20. Sharp G, Fritscher KD, Pekar V, et al: Vision 20 / 20 : Perspectives on automated image segmentation for radiotherapy. Med Phys 41:1-13, 2014
21. Rohlfing T, Brandt R, Menzel R, et al: Quo vadis, atlas-based segmentation? Handbook of Biomedical Image Analysis. Boston, MA: Springer US;435-486, April 2005
22. Thirion J-P: Image matching as a diffusion process: An analogy with Maxwell's demons. Med Image Anal 2:243-260, Sep. 1998
23. Rueckert D, Sonoda LI, Hayes C, et al: Nonrigid registration using free-form deformations: Application to breast MR images. IEEE Trans Med Imaging 18:712-721, 1999
24. Qazi AA, Pekar V, Kim J, et al: Auto-segmentation of normal and target structures in head and neck CT images: A feature-driven model-based approach. Med Phys 38:6160-6170, 2011
25. Han X, Hoogeman MS, Levendag PC, et al: Atlas-based auto-segmentation of head and neck CT images. Lect Notes Comput Sci 5242:434-441, 2008. LNCS
26. Klein S, van der Heide UA, Lips IM, et al: Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. Med Phys 35:1407-1417, Mar. 2008
27. Wang H, Dong L, Lii MF, et al: Implementation and validation of a three-dimensional deformable registration algorithm for targeted prostate cancer radiotherapy. Int J Radiat Oncol 61:725-735, Mar. 2005
28. Commowick O, Malandain G: Efficient Selection of the Most Similar Image in a Database for Critical Structures Segmentation, Berlin, Heidelberg: Springer Berlin Heidelberg 4792(Pt 2), 2007
29. Rohlfing T, Brandt R, Menzel R, et al: Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. Neuroimage 21:1428-1442, Apr. 2004
30. Blezek DJ, Miller JV: Atlas stratification. Med Image Anal 11:443-457, Oct. 2007
31. Aljabar P, Heckemann RA, Hammers A, et al: Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. Neuroimage 46:726-738, Jul. 2009

32. Wu M, Rosano C, Lopez-Garcia P, et al: Optimum template selection for atlas-based segmentation. Neuroimage 34:1612-1618, Feb. 2007
33. Jia H, Wu G, Wang Q, et al: ABSORB: Atlas building by self-organized registration and bundling. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 51; 2785- 2790
34. Yang J, Haas B, Fang R, et al: Atlas ranking and selection for automatic segmentation of the esophagus from CT scans. Phys Med Biol 62:9140-9158, 2017
35. Commowick O, Warfield SK, Malandain G: Using Frankenstein's creature paradigm to build a patient specific atlas. *Lect Notes Comput Sci* 5762 LNCS:993-1000, 2009
36. Yang J, Zhang Y, Zhang L, et al: Automatic segmentation of parotids from CT scans using multiple atlases. Med Image Anal Clin A Gd Chall 323-330, 2010
37. Iglesias JE, Sabuncu MR: Multi-atlas segmentation of biomedical images: A survey. Med Image Anal 24:205-219, Aug. 2015
38. Chen A, Niermann KJ, Deeley MA, et al: Evaluation of multiple-atlas-based strategies for segmentation of the thyroid gland in head and neck CT images for IMRT. Phys Med Biol 57:93-111, Jan. 2012
39. Yang J, Amini A, Williamson R, et al: Automatic contouring of brachial plexus using a multi-atlas approach for lung cancer radiation therapy. Pract Radiat Oncol 3:e139-e147, Oct. 2013
40. Sjöberg C, Lundmark M, Granberg C, et al: Clinical evaluation of multi-atlas based segmentation of lymph node regions in head and neck and prostate cancer patients. Radiat Oncol 8:1-7, 2013
41. Kirişli HA, Schaap M, Klein S, et al: Evaluation of a multi-atlas based method for segmentation of cardiac CTA data: A large-scale, multicenter, and multivendor study. Med Phys 37:6279-6291, Nov. 2010
42. Isgum I, Staring M, Rutten A, et al: Multi-atlas-based segmentation with local decision fusion—application to cardiac and aortic segmentation in CT scans. IEEE Trans Med Imaging 28:1000-1010, Jul. 2009
43. Sabuncu MR, Yeo BTT, Van Leemput K, et al: A generative model for image segmentation based on label fusion. IEEE Trans Med Imaging 29:1714-1729, 2010
44. Warfield SK, Zou KH, Wells WM: Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation, 23:903–921, 2004
45. Langerak TR, van der Heide UA, Kotte ANTJ, et al: Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE). IEEE Trans Med Imaging 29:2000-2008, 2010
46. Ramus L, Malandain G: Multi-atlas based segmentation: Application to the head and neck region for radiotherapy planning. Medical Image Analysis for the Clinic: A Grand Challenge - Workshop Proceedings from the 13th International Conference on Medical Image Computing and Computer Assisted Intervention. 281-288, 2010
47. Pekar V, Allaire S, Qazi A: Head and Neck Auto-Segmentation Challenge: Segmentation of the Parotid Glands. MICCAI 2010 A Grand Challenge Clinics; 2010. p. 273-280
48. Raudaschl PF, Zaffino P, Sharp GC, et al: Evaluation of segmentation methods on head and neck CT: Auto-segmentation challenge 2015. Med Phys 44:2020-2036, 2017
49. Yang J, Veeraraghavan H, Armato SG, et al: Autosegmentation for thoracic radiation treatment planning: A grand challenge at AAPM 2017. Med Phys 45:4568-4581, Sep. 2018
50. McCarroll RE, Beadle BM, Balter PA. et al: Retrospective validation and clinical implementation of automated contouring of organs at risk in the head and neck: A step toward automated radiation treatment planning for low- and middle-income countries. J Glob Oncol 4:1-11, 2018
51. Zhou R, Liao Z, Pan T, et al: Cardiac atlas development and validation for automatic segmentation of cardiac substructures. Radiother Oncol 122:66-71, Jan. 2017
52. Heimann T, Meinzer H-P: Statistical shape models for 3D medical image segmentation: A review. Med Image Anal 13:543-563, Aug. 2009
53. Pekar V, McNutt TR, Kaus MR: Automated model-based organ delineation for radiotherapy planning in prostatic region. Int J Radiat Oncol 60:973-980, Nov. 2004
54. Freedman D, Radke RJ, Zhang T, et al: Model-based segmentation of medical imagery by matching distributions. IEEE Trans Med Imaging 24:281-292, Mar. 2005
55. Feng Q, Foskey M, Chen W, et al: Segmenting CT prostate images using population and patient-specific statistics for radiotherapy. Med Phys 37:4121-4132, Jul. 2010
56. Geremia E, Clatz O, Menze BH, et al: Spatial decision forests for MS lesion segmentation in multi-channel magnetic resonance images. Neuroimage 57:378-390, Jul. 2011
57. Criminisi A, Shotton J, Konukoglu E: Decision forests: A Unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Found Trends® Comput Graph Vis 7:81-227, 2011
58. Li W, Liao S, Feng Q, et al: Learning image context for segmentation of prostate in CT-guided radiotherapy. Med Image Comput Comput Assist Interv 14:570-578, 2011
59. Bauer S, Nolte L-P, Reyes M: Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. In: Medical Image Computing And Computer-Assisted Intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention, 14; 2011. p. 354-361
60. Vaishnavee KB, Amshakala K: An automated MRI brain image segmentation and tumor detection using SOM-clustering and Proximal support vector machine classifier. In: 2015 IEEE International Conference on Engineering and Technology (ICETECH); 2015. p. 1-6
61. Lu J, Wang D, Shi L, et al: Automatic liver segmentation in CT images based on Support Vector Machine. In: Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics, 25; 2012. p. 333-336
62. Zhang X, Tian J, Xiang D, et al: Interactive liver tumor segmentation from CT scans using support vector classification with watershed. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2011; 2011. p. 6005-6008
63. Rendon-Gonzalez E, Ponomaryov V: Automatic lung nodule segmentation and classification in CT images based on SVM. In: 2016 9th International Kharkiv Symposium on Physics and Engineering of Microwaves, Millimeter and Submillimeter Waves (MSMW); 2016. p. 1-4
64. Mahapatra D: Automatic cardiac segmentation using semantic information from random forests. J Digit Imaging 27:794-804, 2014
65. Pereira S, Pinto A, Oliveira J, et al: Automatic brain tissue segmentation in MR images using random forests and conditional random fields. J Neurosci Methods 270:111-123, 2016
66. Jin C, Shi F, Xiang D, et al: 3D fast automatic segmentation of kidney based on modified AAM and random forest. IEEE Trans Med Imaging 35:1395-1407, 2016
67. Chang KW, Summers RM, Narayanan D, et al: Automated segmentation of the thyroid gland on thoracic CT scans by multiatlas label fusion and random forest classification random forest classification. Med Imaging 3: 2017
68. Gao Y: Accurate segmentation of CT pelvic organs via incremental cascade learning and regression-based deformable models. ProQuest Diss Theses 35:153, 2016
69. Liu J, Hoffman J, Zhao J, et al: Mediastinal lymph node detection and station mapping on chest CT using spatial priors and random forest. Med Phys 43:4362-4374, 2016
70. Serag A, Wilkinson AG, Telford EJ, et al: SEGMA: An automatic SEGMentation approach for human brain MRI using sliding window and random forests. Front Neuroinf 11:1-11, 2017
71. Long J, Shelhamer E, Darrell T: Fully convolutional networks for semantic segmentation. In: Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit, 07–12–June; 2015. p. 3431-3440
72. Rumelhart DE, Hinton GE, Williams RJ: Learning representations by back-propagating errors. Nature 323:533-536, Oct. 1986
73. Liu H, Yan M, Song E, et al: Label fusion method based on sparse patch representation for the brain MRI image segmentation. IET Image Process 11:502-511, Jul. 2017

74. Zhu Y, Wang L, Liu M, et al: MRI-based prostate cancer detection with high-level representation and hierarchical classification. Med Phys 44:1028-1039, Mar., 2017
75. Kamnitsas K, Ferrante E, Parisot S, et al: DeepMedic for brain tumor segmentation. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2016; 2016. p. 138-149
76. Roth HR, Oda H, Zhou X, et al: An application of cascaded 3D fully convolutional networks for medical image segmentation. Comput Med Imaging Graph 66:90-99, Jun. 2018
77. Yang Y, Jiang H, Sun Q: A multiorgan segmentation model for CT volumes via full convolution-deconvolution network. Biomed Res Int 2017:1-9, 2017
78. James G, Witten D, Hastie T, et al: An Introduction to Statistical Learning, 103. New York, NY: Springer New York; 2013
79. Borra S, Di Ciaccio A: Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods. Comput Stat Data Anal 54:2976-2989, Dec. 2010
80. Shahin MA, Maier HR, Jaksa MB: Data division for developing neural networks applied to geotechnical engineering. J Comput Civil Eng 18:105-114, Apr. 2004
81. Ronneberger O, Fischer P, Brox T: U-Net: Convolutional Networks for Biomedical Image Segmentation. 1-8, 2015
82. Çiçek Ö, Abdulkadir A, Lienkamp SS, et al: 3D U-net: Learning dense volumetric segmentation from sparse annotation. Lect Notes Comput Sci 9901:424-432, 2016. LNCS
83. Milletari F, Navab N, Ahmadi S-A: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 4th International Conference on 3D Vision (3DV); 2016. p. 565-571
84. He K, Zhang X, Ren S, et al: Deep Residual Learning for Image Recognition. 2015
85. Krähenbühl P, Koltun V: Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials.1–9, 2012
86. Pawlowski N, Ktena SI, Lee MCH, et al: DLTK: State of the Art Reference Implementations for Deep Learning on Medical Images. 1–4, 2017
87. Gibson E, Li W, Sudre C, et al: NiftyNet: A deep-learning platform for medical imaging. Comput Methods Prog Biomed 158:113-122, May 2018
88. Bernard O, Lalande A, Zotti C, et al: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved? IEEE Trans Med Imaging 0062: 1-1, 2018
89. Cardenas CE, Mohamed ASR, Tao R, et al: Prospective qualitative and quantitative analysis of real-time peer review quality assurance rounds incorporating direct physical examination for head and neck cancer radiation therapy. Int J Radiat Oncol Biol Phys 98:532-540, 2017
90. Marks LB, Adams RD, Pawlicki T, et al: Enhancing the role of case-oriented peer review to improve quality and safety in radiation oncology: Executive summary. Pract Radiat Oncol 3:149-156, 2013
91. Cox BW, Kapur A, Sharma A, et al: Prospective contouring rounds: A novel, high-impact tool for optimizing quality assurance. Pract Radiat Oncol 5:e431-e436, Sep. 2015
92. Chen HC, Tan J, Dolly S, et al: Automated contouring error detection based on supervised geometric attribute distribution models for radiation therapy: A general strategy. Med Phys 42:1048-1059, 2015
93. McCarroll R, Yang J, Cardenas CE, et al: Machine learning for the prediction of physician edits to clinical auto-contours in the head-and-neck. Med Phys 44:3160, 2017
94. Hui CB, Nourzadeh H, Watkins WT, et al: Quality assurance tool for organ at risk delineation in radiation therapy using a parametric statistical approach. Med Phys 45:2089-2096, 2018
95. Court LE, Kisling K, McCarroll R, et al: Radiation planning assistant - a streamlined, fully automated radiotherapy treatment planning system. J Vis Exp 2018: 2018
96. Zhou X, Ito T, Takayama R, et al: Three-dimensional CT image segmentation by combining 2D fully convolutional network with 3D majority voting. In: Carneiro G, Mateus D, Peter L, Bradley A, Tavares JMRS, Belagiannis V, Papa JP, Nascimento JC, Loog M, Lu Z, Cardoso JS, Cornebise J, eds. Deep Learning and Data Labeling for Medical Applications. DLMIA 2016, LABELS 2016. Lecture Notes in Computer Science, Cham: Springer International Publishing; 2016:111-120
97. Bai W, Sinclair G, Tarroni G, et al: Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. J Cardiovasc Magn Reson 20:1-12, 2017
98. Dalmış MU, Litjens G, Holland K, et al: Using deep learning to segment breast and fibroglandular tissue in MRI volumes. Med Phys 44:533-546, Feb. 2017
99. Fechter T, Adebahr S, Baltas D, et al: Esophagus segmentation in CT via 3D fully convolutional neural network and random walk. Med Phys 44:6341-6352, 2017
100. Trullo R, Petitjean C, Nie D, et al: Joint segmentation of multiple thoracic organs in CT images with two collaborative deep architectures. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA 2017, ML-CDS 2017. Lecture Notes in Computer Science, 10553; 2017. p. 21-29
101. Wang S, Zhou M, Gevaert O, et al: A multi-view deep convolutional neural networks for lung nodule segmentation. In: 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); 2017. p. 1752-1755
102. Lustberg T, Van Soest J, Gooding M, et al: Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. Radiother Oncol 126:312-317, Feb. 2018
103. Men K, Zhang T, Chen X, et al: Fully automatic and robust segmentation of the clinical target volume for radiotherapy of breast cancer using big data and deep learning. Phys Med 50:13-19, 2018
104. Ben-Cohen A, Diamant I, Klang E, et al: Fully convolutional network for liver segmentation and lesions detection. In: Carneiro G, Mateus D, Peter L, Bradley A, Tavares JMRS, Belagiannis V, Papa JP, Nascimento JC, Loog M, Lu Z, Cardoso JS, Cornebise J, eds. Deep Learning and Data Labeling for Medical Applications, 10008, Cham: Springer International Publishing; 2016:77-85
105. Hu P, Wu F, Peng J, et al: Automatic 3D liver segmentation based on deep learning and globally optimized surface evolution. Phys Med Biol 61:8676-8698, 2016
106. Ibragimov B, Toesca D, Chang D, et al: Combining deep learning with anatomical analysis for segmentation of the portal vein for liver SBRT planning. Phys Med Biol 62:8943-8958, 2017
107. Zhou X, Takayama R, Wang S: Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method. Med Phys 44:5221-5233, 2017
108. Vorontsov E, Tang A, Roy D, et al: Metastatic liver tumour segmentation with a neural network-guided 3D deformable model. Med Biol Eng Comput 55:127-139, Jan. 2017
109. Yuan Y: Hierarchical Convolutional-Deconvolutional Neural Networks for Automatic Liver and Tumor Segmentation, i; 2017. p. 3–6, 2017
110. Hu P, Wu F, Peng J, et al: Automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets. Int J Comput Assist Radiol Surg 12:399-411, Mar. 2017
111. Lu F, Wu F, Hu P, et al: Automatic 3D liver location and segmentation via convolutional neural network and graph cut. Int J Comput Assist Radiol Surg 12:171-182, 2017
112. Christ PF, Ettlinger F, Grün F, et al: Automatic Liver and Tumor Segmentation of CT and MRI Volumes Using Cascaded Fully Convolutional Neural Networks. 1–20, 2017
113. Qin W, Wu J, Han F, et al: Superpixel-based and boundary-sensitive convolutional neural network for automated liver segmentation. Phys Med Biol 63:1-14, 2018
114. Fu Y, Mazur TR, Wu X, et al: A novel MRI segmentation method using CNN based correction network for MRI guided adaptive radiotherapy. Med Phys 2018
115. Gibson E, Giganti F, Hu Y, et al: Automatic multi-organ segmentation on abdominal CT with dense V-networks. IEEE Trans Med Imaging 37:1822-1834, 2018
116. Bobo MF, Huo Y, Yao Y, et al: Fully convolutional neural networks improve abdominal organ segmentation. In: Proc. SPIE 10574, Medical Imaging 2018: Image Processing, 105742 V; 2018
117. Liu X, Guo S, Yang B, et al: Automatic organ segmentation for CT scans based on super-pixel and convolutional neural networks. J Digit Imaging 31:748-760, 2018

118. Roth HR, Lu L, Lay N, et al: Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation. Med Image Anal 45:94-107, 2018
119. Milletari F, Navab N, Ahmadi S-A: V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 1–11, 2016
120. Cha KH, Hadjiiski LM, Samala RK, et al: Bladder cancer segmentation in CT for treatment response assessment: Application of deep-learning convolution neural network—a pilot study. Tomography 2:421-429, Dec. 2016
121. Guo Y, Gao Y, Shen D: Deformable MR prostate segmentation via deep feature learning and sparse patch matching. In: Deep Learning for Medical Image Analysis, Elsevier; 2017:197-222
122. Cheng R, Roth HR, Lay N, et al: Automatic magnetic resonance prostate segmentation by deep learning with holistically nested networks. J Med Imaging 4:1, 2017
123. Men K, Dai J, Li Y: Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks. Med Phys 44:6377-6389, 2017
124. Trebeschi S, Van Griethuysen JJM, Lambregts DMJ, et al: Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric MR. Sci Rep 7:1-9, Dec. 2017
125. To MNN, Vu DQ, Turkbey B, et al: Deep dense multi-path neural network for prostate segmentation in magnetic resonance imaging. Int J Comput Assist Radiol Surg 13:1687-1696, 2018
126. Karimi D, Samei G, Kesch C, et al: Prostate segmentation in MRI using a convolutional neural network architecture and training strategy based on statistical shape models. Int J Comput Assist Radiol Surg 13:1211-1219, 2018
127. Ibragimov B, Xing L: Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. Med Phys 44:547-557, Feb. 2017
128. Men K, Chen X, Zhang Y, et al: Deep deconvolutional neural network for target segmentation of nasopharyngeal cancer in planning computed tomography images. Front Oncol 7:1-9, 2017
129. Ren X, Xiang L, Nie D, et al: Interleaved 3D-CNNs for joint segmentation of small-volume structures in head and neck CT images. Med Phys 45:2063-2075, May 2018
130. Hänsch A, Schwier M, Morgas T, et al: Comparison of different deep learning approaches for parotid gland segmentation from CT images. In: Proc. SPIE 10575, Med. Imaging 2018 Comput. Diagnosis; 2018. 1057519
131. Willems S, Crijns W, La Greca Saint-Esteven A, et al: Clinical Implementation of DeepVoxNet for auto-delineation of organs at risk in head and neck cancer patients in radiotherapy. In: OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis, Springer International Publishing; 2018:223-232
132. Zhu W, Huang Y, Tang H, et al: AnatomyNet: Deep 3D Squeeze-and-Excitation U-Nets for Fast and Fully Automated Whole-Volume Anatomical Segmentation.1–14, 2018
133. Nikolov S, Blackwell S, Mendes R, et al: Deep Learning to Achieve Clinically Applicable Segmentation of Head and Neck Anatomy for Radiotherapy. 1–31, 2018
134. Cardenas CE, McCarroll RE, Court LE, et al: Deep learning algorithm for auto-delineation of high-risk oropharyngeal clinical target volumes with built-in dice similarity coefficient parameter optimization function. Int J Radiat Oncol Biol Phys 101:468-478, 2018
135. Cardenas CE, Anderson BM, Aristophanous M, et al: Auto-delineation of oropharyngeal clinical target volumes using 3D convolutional neural networks. Phys Med Biol 63: Nov. 2018:215026
136. Liu Y, Stojadinovic S, Hrycushko B, et al: A deep convolutional neural network-based automatic delineation strategy for multiple brain metastases stereotactic radiosurgery. PLoS One 12: Oct. 2017:e0185844
137. Zhao X, Wu Y, Song G, et al: A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. Med Image Anal 43:98-111, Jan. 2018
138. AlBadawy EA, Saha A, Mazurowski MA: Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing. Med Phys 45:1150-1158, Mar. 2018
139. Hoseini F, Shahbahrami A, Bayat P: AdaptAhead optimization algorithm for learning deep CNN applied to MRI segmentation. J Digit Imaging 32:104-115, Jul. 2018